

Figure 22-42 Energy-dependent binding change mechanism for ATP synthesis by proton-translocating ATP synthase. F_1 has three chemically identical but conformationally distinct interacting $\alpha\beta$ protomers: O, the open conformation, has very low affinity for ligands and is catalytically inactive; L has loose binding for ligands and is catalytically inactive; T has tight binding for ligands and is catalytically active. ATP synthesis occurs in three steps. (1) Binding of ADP and P_i to site L. (2) Energy-dependent conformational change converting

the ATP-containing T site to an “open” (O) site and convert the O site to an L site.

3. ATP is synthesized at the T site on one subunit while ATP dissociates from the O site on another subunit. On the surface of the active site, the formation of ATP from ADP and P_i entails little free energy change, that is, the reaction is essentially at equilibrium. Consequently, the free energy supplied by the proton flow primarily facilitates the release of the newly synthesized ATP from the enzyme; that is, it drives the $T \rightarrow O$ transition, thereby disrupting the enzyme–ATP interactions that had previously promoted the spontaneous formation of ATP from ADP + P_i in the T site.

How is the free energy of proton transfer coupled to the synthesis of ATP? Boyer proposed that *the binding changes are driven by the rotation of the catalytic assembly, $\alpha_3\beta_3$ with respect to other portions of the F_1F_0 -ATPase*. This hypothesis is supported by the X-ray structure of F_1 . Thus, the closely fitting nearly circular arrangement of the α and β subunits’ inner surface about the γ subunit’s helical C-terminus is reminiscent of a cylindrical bearing rotating in a sleeve (Fig. 22-38c). Indeed, the contacting hydrophobic surfaces in this assembly are devoid of the hydrogen bonding and ionic interactions that would interfere with their free rotation; that is, the bearing and sleeve appear to be “lubricated.” Moreover, the central cavity in the $\alpha_3\beta_3$ assembly (Fig. 22-38a) would permit the passage of the γ subunit’s N-terminal helix within the core of this particle during rotation. Finally, the conformational differences between F_1 ’s three catalytic sites appear to be correlated with the rotational position of the γ subunit. *Apparently the γ subunit, which is thought to rotate within the fixed $\alpha_3\beta_3$ assembly, acts as a molecular cam shaft in linking the proton-motive force-driven rotational motor to the conformational changes in the catalytic sites of F_1 .* This concept is also supported by molecular dynamics simulations (Section 9-4) by Leslie, Walker, and Martin Karplus, which indicate that the conformational changes in the β subunits arise from both steric and electrostatic interactions with the rotating γ subunit.

binding site L to T, T to O, and O to L. (3) Synthesis of ATP at site T and release of ATP from site O. The enzyme returns to its initial state after two more passes of this reaction sequence. The energy that drives the conformational change is apparently transmitted to the catalytic $\alpha_3\beta_3$ assembly via the rotation of the $\gamma\epsilon$ assembly (in *E. coli*; $\gamma\delta\epsilon$ in mitochondria), here represented by the centrally located asymmetric pointer (green). [After Cross, R.L., *Annu. Rev. Biochem.* **50**, 687 (1980).]  **See the Animated Figures**

Rotating assemblies are not unprecedented in biological systems. Bacterial flagella, which function as propellers, had previously been shown to be membrane-mounted rotary engines that are driven by the discharge of a proton gradient (Section 35-3Ib).

e. The F_1F_0 -ATPase Is a Rotary Engine

In the F_1F_0 -ATPase, the rotor is proposed to be an assembly of the c ring and its associated γ and (*E. coli*) ϵ subunits, whereas the ab_2 unit and the (*E. coli*) δ subunit together with the $\alpha_3\beta_3$ spheroid form the stator (Fig. 22-41). The rotation of the c ring in the membrane relative to the stationary a subunit is driven by the migration of protons from the outside to the inside, as we discuss below. The peripheral arm ($b_2\delta$) presumably functions to hold the $\alpha_3\beta_3$ spheroid in place while the γ subunit rotates inside it.

The rotation of the *E. coli* $\gamma\epsilon$ - c -ring rotor with respect to the $ab_2\delta$ - $\alpha_3\beta_3$ stator has been ingeniously demonstrated by Masamitsu Futai using techniques developed by Kazuhiko Kinoshita Jr. and Masasuke Yoshida (Fig. 22-43a). The $\alpha_3\beta_3$ spheroid of *E. coli* F_1F_0 -ATPase was fixed, head down, to a glass surface as follows: Six consecutive His residues (a so-called **His Tag**; Section 6-3Dg) were mutagenically appended to the N-terminus of the α subunit, which is located at the top of the $\alpha_3\beta_3$ spheroid as it is drawn in Fig. 22-38a. The His-tagged assembly was applied to a glass surface coated with horseradish peroxidase (which, like most proteins, sticks to glass) conjugated with Ni^{2+} -**nitroacetic acid** [$N(CH_2COOH)_3$], which tightly binds His tags], thereby binding the F_1F_0 -ATPase with its F_0 side facing away from the surface. The Glu 2 residues of this assembly’s c subunits, which are located on the side of the c ring facing away from F_1 , had been mutagenically replaced by Cys residues, which were then covalently linked to **biotin** (a coenzyme that normally participates in carboxylation reactions; Section 23-1Ab). A fluorescently labeled and biotinylated (at one end) filament of the muscle protein **actin** (Section 35-3Ac) was then attached to the c subunit through the addition of a bridging molecule of **streptavidin**, a protein that avidly binds biotin to each of four binding sites (Cys 193 of

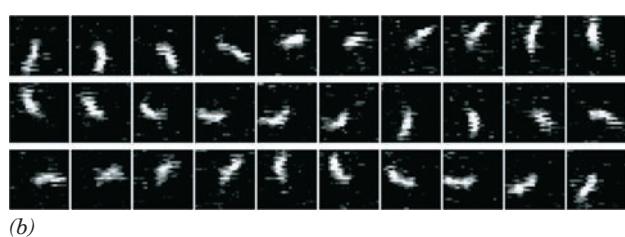
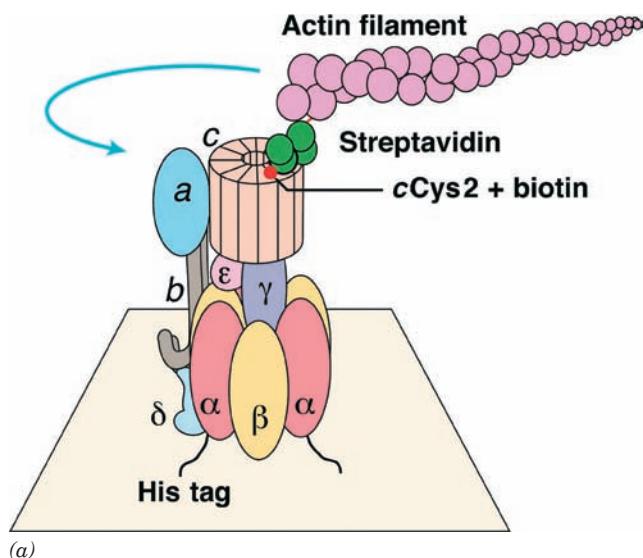


Figure 22-43 Rotation of the c ring in *E. coli* F_1F_0 -ATPase. (a) The experimental system used to observe the rotation. See the text for details. The blue arrow indicates the observed direction of rotation of the fluorescently labeled actin filament that was linked to the c ring. (b) The rotation of a 3.6- μm -long actin filament in the presence of 5 mM MgATP as seen in successive video images taken through a fluorescence microscope. [Courtesy of Masamitsu Futai, Osaka University, Osaka, Japan.]

the γ subunit, the only other Cys residue in the rotor, was mutagenically replaced by Ala to prevent it from being linked to an actin filament).

E. coli F_1F_0 -ATPase can work in reverse, that is, it can pump protons from the inside (cytoplasm) to the outside (periplasm) at the expense of ATP hydrolysis (this enables the bacterium to maintain its proton gradient under anaerobic conditions, which it uses to drive various processes such as flagellar rotation). Thus, the foregoing preparation was observed under a fluorescence microscope as a 5 mM MgATP solution was infused over it. Many of the actin filaments were seen to rotate (Fig. 22-43b), and always in a counterclockwise direction when viewed looking down on the glass surface (from the outside). This would permit the γ subunit to sequentially interact with the β subunits in the direction



(Figs. 22-38b and 22-42), the direction expected for ATP hydrolysis.

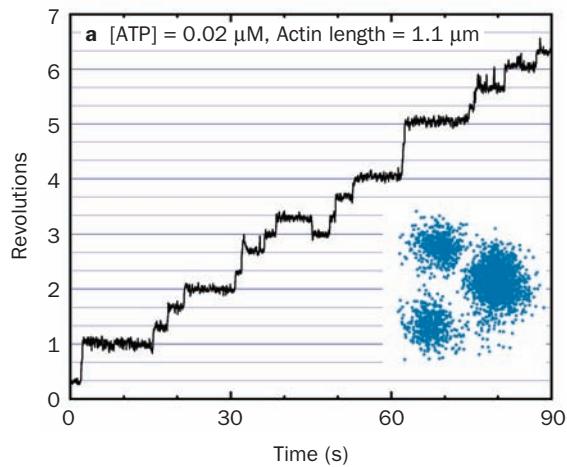


Figure 22-44 Stepwise rotation of the γ subunit of F_1 relative to an immobilized $\alpha_3\beta_3$ unit at low ATP concentration as observed by fluorescence microscopy. The graph plots the cumulative number of rotations made by a fluorescently labeled actin filament that was linked at one end to the γ subunit in a preparation similar to that diagrammed in Fig. 22-43a (but lacking F_0 , δ , and ϵ). Note that the actin filament rotates in increments of 120°. This is also evident in the inset, which shows the superposition of the centers of the actin images (the $\delta_3\beta_3\gamma$ assembly is fixed in the center). [Courtesy of Kazuhiko Kinosita, Jr., Keio University, Yokohama, Japan.]

In a variation of the above experiment, the γ subunit of the $\alpha_3\beta_3\gamma$ complex was directly cross-linked, via its Cys 193, to a fluorescently labeled actin filament and, in this case, the β subunits were immobilized by appended His tags. At very low ATP concentrations (e.g., 0.02 μM), video images (Fig. 22-44) revealed that the fluorescent actin filament rotated counterclockwise in discrete steps of 120°, as the binding change mechanism predicts. Moreover, the calculated frictional work done in each rotational step is very nearly equal to the energy available from the hydrolysis of one ATP molecule, that is, the F_1F_0 -ATPase converts chemical to mechanical energy with nearly 100% efficiency.

The foregoing system also works in reverse. An ~0.7- μm -diameter magnetic bead that was coated with streptavidin was attached to the biotinylated γ subunit of an immobilized $\alpha_3\beta_3\gamma$ complex. When the resulting assembly was placed in a rotating magnetic field in the presence of ADP and P_i , ATP was produced when the magnetic field rotated in the clockwise direction but hydrolyzed when it rotated in the counterclockwise direction. This further demonstrates that F_1 is a device that interconverts mechanical and chemical energy.

f. c-Ring Rotation Is Impelled by H^+ -Induced Conformational Changes

The foregoing structural and biochemical information has led to the model for proton-driven rotation of the F_0 subunit that is diagrammed in Fig. 22-45. Protons from the outside enter a hydrophilic channel between the a subunit and the c ring, where they bind to a c subunit. The c ring

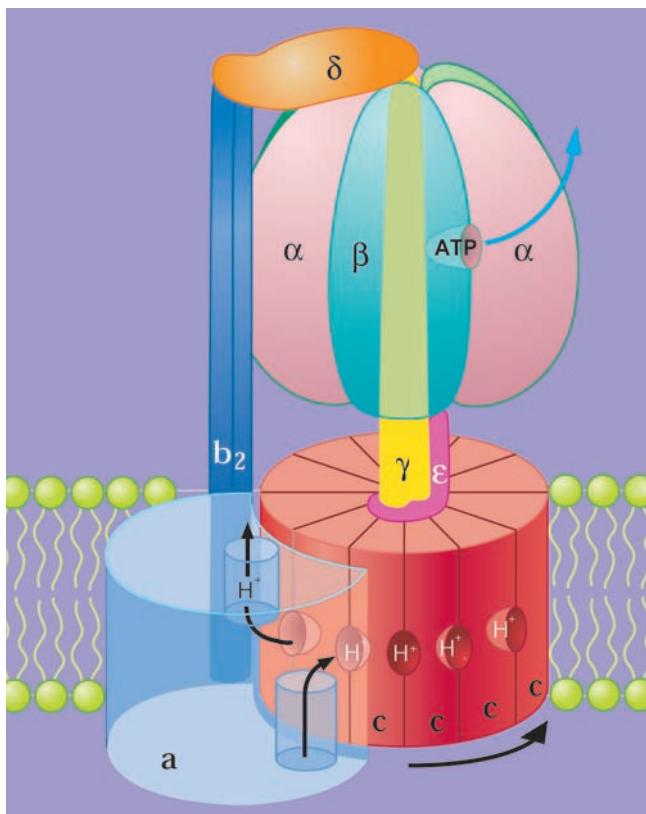


Figure 22-45 Schematic diagram of the action of the *E. coli*

F₁F₀-ATPase. The $\gamma\epsilon-c_{12}$ ring complex is the rotor and the $ab_2-\alpha_3\beta_3\delta$ complex is the stator. Rotational motion is imparted to the rotor by the passage of protons from the outside (periplasm) to the inside (cytoplasm). Protons entering from the outside bind to a *c* subunit where it interacts with the *a* subunit, and exit to the inside after the *c* subunit has made a nearly full rotation as indicated (black arrows), so that the *c* subunit again contacts the *a* subunit. The *b*₂*δ* complex presumably functions to prevent the $\alpha_3\beta_3$ assembly from rotating with the γ subunit. [Courtesy of Richard Cross, State University of New York, Syracuse, New York.]

then rotates nearly a full turn (while protons bind to successive *c* subunits as they pass this input channel) until the subunit reaches a second hydrophilic channel between the *a* subunit and the *c*-ring that opens into the inside. There the proton is released. Thus, *E. coli* F₁F₀-ATPase, which has 12 *c* subunits in its F₀ assembly and generates 3ATP per turn, ideally forms $3/12 = 0.25$ ATP for every proton it passes from the periplasmic space (outside) to the cytosol (inside). Organisms with more/less *c* subunits in their *c*-rotor tend to have lesser/greater values of proton-motive force across their membranes and hence have less/more impetus imparted to their *c*-rotor per proton passed. Hence it takes the passage of proportionately more/less protons to generate each ATP, as the first law of thermodynamics requires.

How does the passage of protons through this system induce the rotation of the *c* ring and hence the synthesis of ATP? The mutation of the *c* subunit's conserved Asp 61 to Asn inactivates *E. coli* F₁F₀-ATPase. The *a* subunit's invariant Arg 210 (*E. coli* numbering) has been similarly impli-

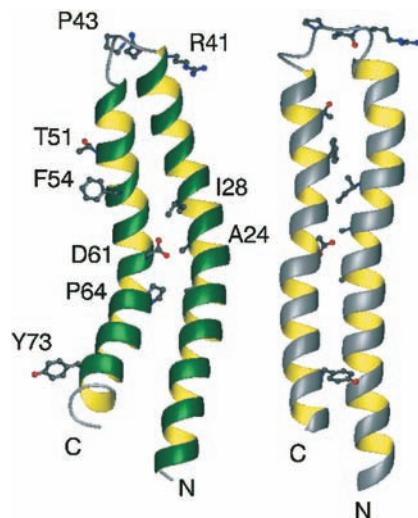


Figure 22-46 NMR structures of the *c* subunit of *E. coli*

F₁F₀-ATPase. The structures, which closely resemble that in Fig. 22-39, were determined in chloroform-methanol-water (4:4:1) solution at (a) pH 8 (at which D61 is deprotonated) and (b) pH 5 (at which D61 is protonated). Selected side chains are shown to aid in the comparison of the two structures. Note that the C-terminal helix in the pH 8 structure has rotated by 140° clockwise, as viewed from the top of the drawing, relative to that in the pH 5 structure. [Courtesy of Mark Girvin, Albert Einstein College of Medicine. PDBIDs (a) 1C99 and (b) 1C0V.]

cated in proton translocation. Through the mutagenic conversion of selected residues on the *a* and *c* subunits to Cys, Robert Fillingame has shown that the outer (C-terminal) helix of *E. coli* subunit *c* (Fig. 22-39), which contains Asp 61, can be disulfide-cross-linked to the putative fourth helix of subunit *a*, which contains Arg 210. Evidently, these helices are juxtaposed at some point in the *c* ring's rotation cycle. Thus, it is postulated that the protonation of Asp 61 releases its attraction to Arg 210, thereby permitting the *c* ring to rotate.

Comparison of the NMR structures of subunit *c* at pH 8 and pH 5 (Fig. 22-46), at which Asp 61 is, respectively, deprotonated and protonated, reveals that its main conformational change on protonation is an ~140° clockwise rotation (as viewed from F₁) of its Asp 61-containing C-terminal helix with respect to its N-terminal helix. Since the C-terminal helix is the *c* ring's outer helix (Fig. 22-39b), this suggests that, on protonation, the rotation of the C-terminal helix mechanically pushes against the juxtaposed *a* subunit so as to rotate the *c* ring in the direction indicated in Fig. 22-45.

D. Uncoupling of Oxidative Phosphorylation

Electron transport (the oxidation of NADH and FADH₂ by O₂) and oxidative phosphorylation (the synthesis of ATP) are normally tightly coupled due to the impermeability of the inner mitochondrial membrane to the passage of protons. Thus the only way for H⁺ to reenter the matrix is through the F₀ portion of the proton-translocating ATP

synthase. In the resting state, when oxidative phosphorylation is minimal, the proton-motive force across the inner mitochondrial membrane builds up to the extent that the free energy to pump additional protons is greater than the electron-transport chain can muster, thereby inhibiting further electron transport. However, many compounds, including **2,4-dinitrophenol (DNP)** and **carbonylcyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP)**, have been found to “uncouple” these processes. The chemiosmotic hypothesis has provided a rationale for understanding the mechanism by which these uncouplers act.

The presence in the inner mitochondrial membrane of an agent that renders it permeable to H^+ uncouples oxidative phosphorylation from electron transport by providing a route for the dissipation of the proton-motive force that does not require ATP synthesis. Uncoupling therefore allows electron transport to proceed unchecked even when ATP synthesis is inhibited. DNP and FCCP are lipophilic weak acids that therefore readily pass through membranes. In a pH gradient, they bind protons on the acidic side of the membrane, diffuse through, and release them on the alkaline side, thereby dissipating the gradient (Fig. 22-47). Thus, *such uncouplers are proton-transporting ionophores* (Section 20-2C).

Even before the mechanism of uncoupling was known, it was recognized that metabolic rates were increased by such compounds. Studies at Stanford University in the

early part of the twentieth century documented an increase in respiration and weight loss caused by DNP. The compound was even used as a “diet pill” for several years. In the words of Efraim Racker (*A New Look at Mechanisms in Bioenergetics*, p. 155):

*In spite of warnings from the Stanford scientists, some enterprising physicians started to administer dinitrophenol to obese patients without proper precautions. The results were striking. Unfortunately in some cases the treatment eliminated not only the fat but also the patients, and several fatalities were reported in the *Journal of the American Medical Association* in 1929. This discouraged physicians for a while. . . .*

a. Hormonally Controlled Uncoupling in Brown Adipose Tissue Functions to Generate Heat

The dissipation of an electrochemical H^+ gradient, which is generated by electron transport and uncoupled from ATP synthesis, produces heat. Heat generation is the physiological function of **brown adipose tissue (brown fat)**. This tissue is unlike typical (white) adipose tissue in that, besides containing large amounts of triacylglycerols, it contains numerous mitochondria whose cytochromes color it brown. Newborn mammals that lack fur, such as humans, as well as hibernating mammals, contain brown fat in their neck and upper back that functions in **nonshivering thermogenesis**, that is, as a “biological heating pad.” (The ATP hydrolysis that occurs during the muscle contractions of shivering—or

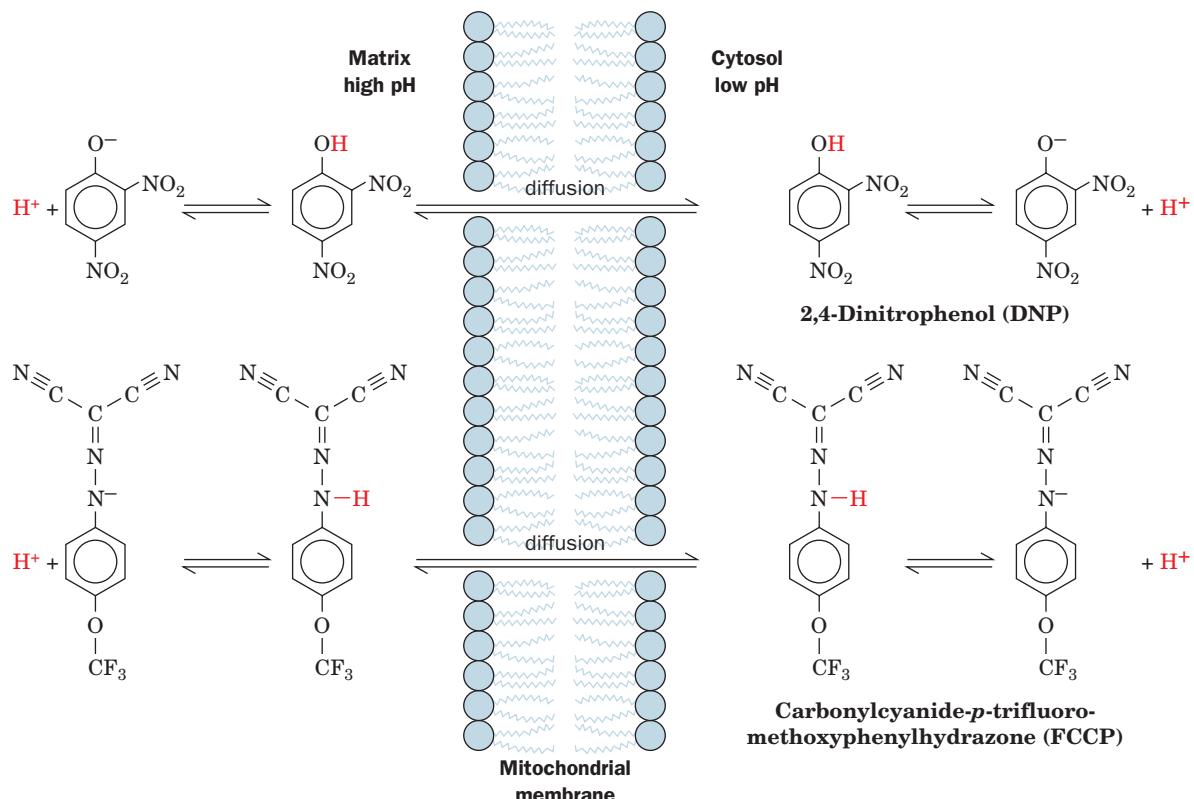


Figure 22-47 Uncoupling of oxidative phosphorylation. The proton-transporting ionophores DNP and FCCP uncouple oxidative phosphorylation from electron transport by

discharging the electrochemical proton gradient generated by electron transport.

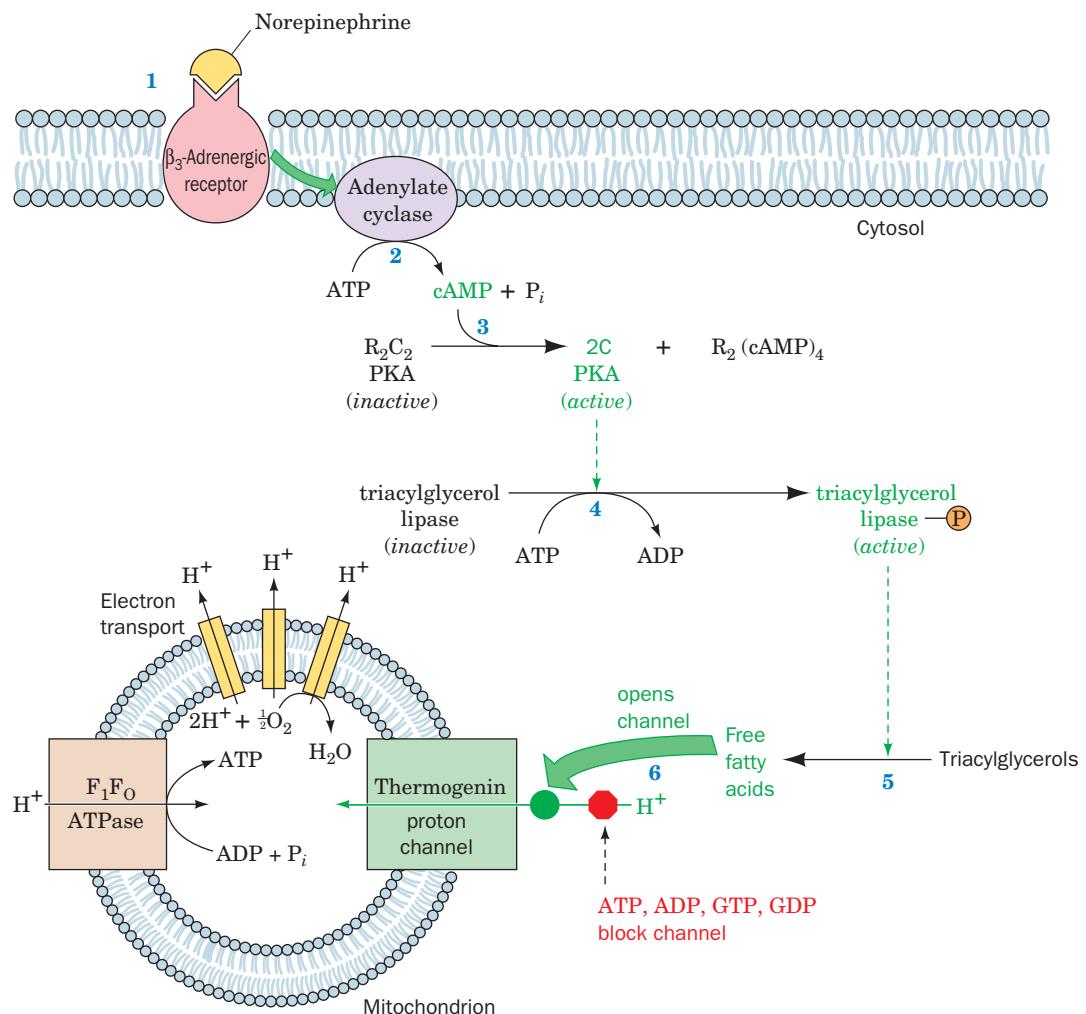


Figure 22-48 Mechanism of hormonally induced uncoupling of oxidative phosphorylation in brown fat mitochondria.

(1) Norepinephrine binds to a β_3 -adrenergic receptor. (2) This stimulates the associated heterotrimeric G protein to activate adenylate cyclase (upper green arrow) to synthesize cAMP. (3) cAMP binding activates protein kinase A (PKA). (4) PKA

phosphorylates hormone-sensitive triacylglycerol lipase, thereby activating it. (5) Triacylglycerols are hydrolyzed, yielding free fatty acids. (6) Free fatty acids overcome the purine nucleotide block of thermogenin's proton channel (lower green arrow), allowing H^+ to enter the mitochondrion uncoupled from ATP synthesis.

any other movement—also produces heat. Nonshivering thermogenesis through substrate cycling is discussed in Section 17-4F.)

The mechanism of heat generation in brown fat involves the regulated uncoupling of oxidative phosphorylation in their mitochondria. These mitochondria contain the protein **thermogenin** [also called **uncoupling protein (UCP)**], a transmembrane homodimer of 307-residue subunits that acts as a channel to control the permeability of the inner mitochondrial membrane to protons. In cold-adapted animals, thermogenin constitutes up to 15% of brown fat inner mitochondrial membrane proteins. The flow of protons through this channel protein is inhibited by physiological concentrations of purine nucleotides (ADP, ATP, GDP, GTP), but this inhibition can be overcome by free fatty acids. The components of this system interact under hormonal control.

Thermogenesis in brown fat mitochondria is activated by free fatty acids. These counteract the inhibitory effects of purine nucleotides, thereby stimulating the flux through the proton channel and uncoupling electron transport from oxidative phosphorylation. *The concentration of fatty acids in brown adipose tissue is controlled by the adrenal hormone norepinephrine (noradrenaline; Section 18-3E) with cAMP acting as a second messenger (Section 18-3).* Norepinephrine binds to the β_3 -adrenergic receptor, a G-protein coupled receptor (GPCR) that, via an associated heterotrimeric G protein, stimulates adenylate cyclase to synthesize cAMP (Fig. 22-48), as described in Section 19-2. The cAMP, in turn, activates protein kinase A (PKA), which activates **hormone-sensitive triacylglycerol lipase** by phosphorylating it (Section 25-5). Finally, the activated lipase hydrolyzes triacylglycerols to yield the free fatty acids that open thermogenin's proton channel. The transcription of

the gene encoding thermogenin is stimulated by the thyroid hormone triiodothyronine (T3; Section 19-1D).

b. Other Tissues Contain UCP Homologs

Although it originally seemed that only brown fat mitochondria contain an uncoupling protein, it is now apparent that other tissues contain homologs of UCP1. Thus, **UCP2** is expressed in many tissues including white adipose tissue, whereas **UCP3** occurs in both brown and white adipose tissues as well as in muscle. These proteins may help regulate metabolic rates, and variations in UCP levels or activity might explain why some people seem to have a “fast” or “slow” metabolism (Section 27-3E). UCPs are being studied as targets for treating obesity, since increasing the activity of UCPs could uncouple respiration from ATP synthesis, thus permitting stored metabolic fuels (especially fat) to be metabolized. The recent discovery that adult humans have small depots of brown fat that are activated by cold has made this an attractive weight loss strategy, although it is possible that stimulating UCPs will cause a compensatory increase in appetite.

Uncoupling proteins are not limited to animals. Some plants express uncoupling proteins in response to cold stress or to increase flower temperature, possibly to enhance the vaporization of scent to attract pollinators.

4 CONTROL OF ATP PRODUCTION

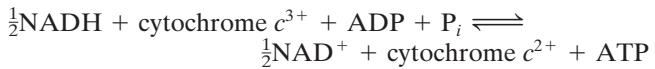
A typical adult woman requires some 1500 to 1800 kcal (6300–7500 kJ) of metabolic energy per day. This corresponds to the free energy of hydrolysis of over 200 mol of ATP to ADP and P_i . Yet the total amount of ATP present in the body at any one time is <0.1 mol; obviously, this sparse supply of ATP must be continually recycled. As we have seen, when carbohydrates serve as the energy supply and aerobic conditions prevail, this recycling involves glycogenolysis, glycolysis, the citric acid cycle, and oxidative phosphorylation.

Of course the need for ATP is not constant. There is a 100-fold change in ATP utilization between sleep and vigorous activity. *The activities of the pathways that produce ATP are under strict coordinated control so that ATP is never produced more rapidly than necessary.* We have already discussed the control mechanisms of glycolysis, glycogenolysis, and the citric acid cycle (Sections 17-4, 18-3, and 21-4). In this section we discuss the mechanisms through which oxidative phosphorylation is controlled and observe how all four systems are synchronized to produce ATP at precisely the rate required at any particular moment.

A. Control of Oxidative Phosphorylation

In our discussion of the control of glycolysis, we saw that most of the reactions in a metabolic pathway function close to equilibrium. *The few irreversible reactions constitute the potential control points of the pathway and usually are catalyzed by regulatory enzymes that are under allosteric control.*

In the case of oxidative phosphorylation, the pathway from NADH to cytochrome *c* functions near equilibrium ($\Delta G \approx 0$):



for which

$$K_{\text{eq}} = \left(\frac{[\text{NAD}^+]}{[\text{NADH}]} \right)^{1/2} \frac{[\text{c}^{2+}]}{[\text{c}^{3+}]} \frac{[\text{ATP}]}{[\text{ADP}][P_i]} \quad [22.2]$$

This pathway is therefore readily reversed by the addition of ATP. *In the cytochrome *c* oxidase reaction, however, the terminal step of the electron-transport chain is irreversible and is thus one of the important regulatory sites of the pathway.* Cytochrome *c* oxidase, in contrast to most regulatory enzyme systems, appears to be controlled exclusively by the availability of one of its substrates, reduced cytochrome *c* (c^{2+}). Since this substrate is in equilibrium with the rest of the coupled oxidative phosphorylation system (Eq. [22.2]), its concentration ultimately depends on the intramitochondrial $[\text{NADH}]/[\text{NAD}^+]$ ratio and the **ATP mass action ratio** ($[\text{ATP}]/[\text{ADP}][P_i]$). By rearranging Eq. [22.2], the ratio of reduced to oxidized cytochrome *c* is expressed

$$\frac{[\text{c}^{2+}]}{[\text{c}^{3+}]} = \left(\frac{[\text{NADH}]}{[\text{NAD}^+]} \right)^{1/2} \left(\frac{[\text{ADP}][P_i]}{[\text{ATP}]} \right) K_{\text{eq}} \quad [22.3]$$

Consequently, the higher the $[\text{NADH}]/[\text{NAD}^+]$ ratio and the lower the ATP mass action ratio, the higher is the $[c^{2+}]$ (reduced cytochrome *c*) and thus the higher is the cytochrome *c* oxidase activity.

How is this system affected by changes in physical activity? In an individual at rest, ATP hydrolysis to ADP and P_i is minimal and the ATP mass action ratio is high; the concentration of reduced cytochrome *c* is therefore low and oxidative phosphorylation is minimal. Increased activity results in hydrolysis of ATP to ADP and P_i , thereby decreasing the ATP mass action ratio and increasing the concentration of reduced cytochrome *c*. This results in an increase in the electron-transport rate and its coupled phosphorylation. Such control of oxidative phosphorylation by the ATP mass action ratio is called **acceptor control** because the rate of oxidative phosphorylation increases with the concentration of ADP, the phosphoryl group acceptor. In terms of a supply–demand system (Section 17-4D), acceptor control is understood as control by the demand block.

The compartmentalization of the cell into mitochondria, where ATP is synthesized, and cytoplasm, where ATP is utilized, presents an interesting control problem: Is it the ATP mass action ratio in the cytosol or in the mitochondrial matrix that ultimately controls oxidative phosphorylation? Clearly the ATP mass action ratio that exerts direct control must be that of the mitochondrial matrix where ATP is synthesized. However, the inner mitochondrial membrane, which is impermeable to adenine nucleotides and P_i , depends on specific transport systems to maintain communication between the two compartments (Section 20-4C). This organization makes it possible for the transport

of adenine nucleotides or P_i to participate in the control of oxidative phosphorylation.

Considerable research effort has been aimed at determining how oxidative phosphorylation is controlled in terms of metabolic control analysis. For example, Hans Westerhoff and Martin Kushmerick employed ^{31}P NMR to measure the ATP/ADP ratios in human forearm muscle at rest and during twitch contractions caused by external electrical stimulation (the ^{31}P NMR spectrum of ATP is shown in Fig. 16-15). Under conditions of low to moderate ATP demand, the cytosolic mass action ratio as controlled by the demand block of the system appears to be the major control factor for mitochondrial oxidation. However, as other laboratories have shown, as the demand for ATP increases, the ADP-ATP translocator exerts greater control until finally, when the demand for ATP is high, control shifts to the supply block of the system, oxidative phosphorylation itself.

B. Coordinated Control of ATP Production

Glycolysis, the citric acid cycle, and oxidative phosphorylation constitute the major pathways for cellular ATP production. Control of oxidative phosphorylation by the ATP mass action ratio depends, of course, on an adequate supply of electrons to fuel the electron-transport chain. This aspect of the system's control is, in turn, dependent on the $[NADH]/[NAD^+]$ ratio (Eq. [22.3]), which is maintained high by the combined action of glycolysis and the citric acid cycle in converting 10 molecules of NAD^+ to NADH per molecule of glucose oxidized (Fig. 22-1). It is clear, therefore, that coordinated control is necessary for the three processes. This is provided by the regulation of each of the control points of glycolysis [hexokinase, phosphofructokinase (PFK), and pyruvate kinase] and the citric acid cycle (pyruvate dehydrogenase, citrate synthase, isocitrate dehydrogenase, and α -ketoglutarate dehydrogenase) by adenine nucleotides or NADH or both as well as by certain metabolites (Fig. 22-49).

a. Citrate Inhibits Glycolysis

The main control points of glycolysis and the citric acid cycle are regulated by several effectors besides adenine nucleotides or NADH (Fig. 22-49). This is an extremely complex system with complex demands. Its many effectors, which are involved in various aspects of metabolism, increase its regulatory sensitivity. One particularly interesting regulatory effect is the inhibition of PFK by citrate. When demand for ATP decreases, [ATP] increases and [ADP] decreases. The

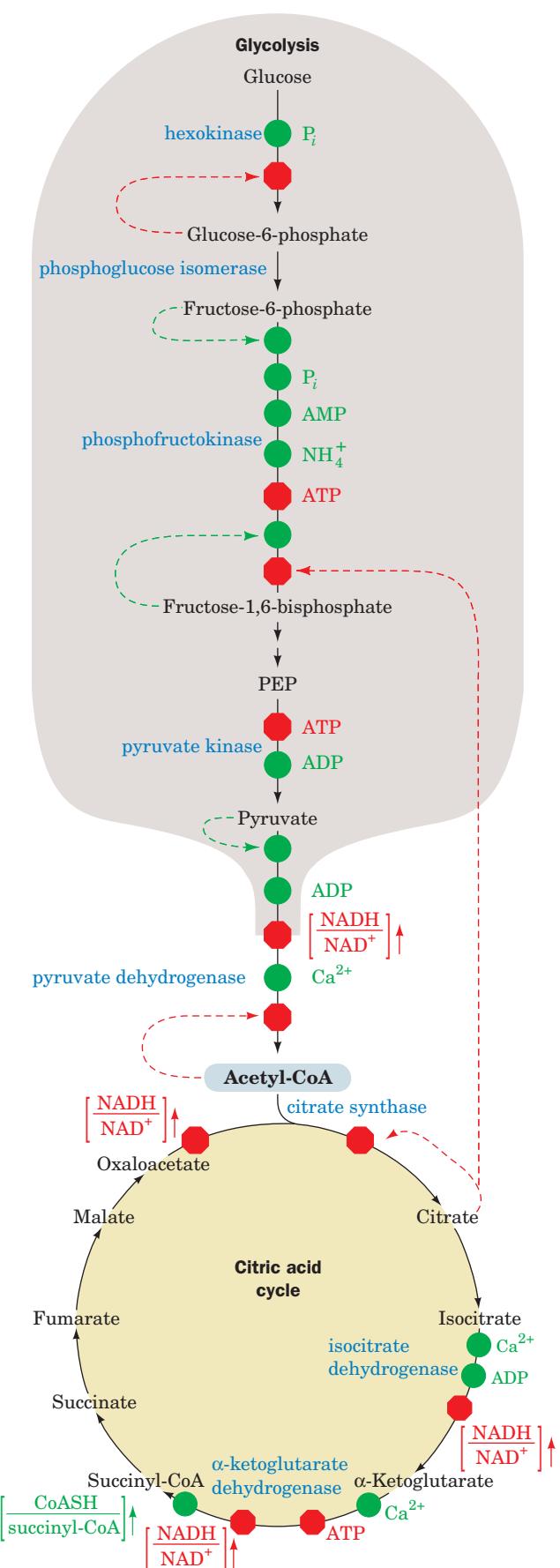


Figure 22-49 Schematic diagram depicting the coordinated control of glycolysis and the citric acid cycle by ATP, ADP, AMP, P_i , Ca^{2+} , and the $[NADH]/[NAD^+]$ ratio (the vertical arrows indicate increases in this ratio). Here a green dot signifies activation and a red octagon represents inhibition. [After Newsholme, E.A. and Leech, A.R., *Biochemistry for the Medical Sciences*, pp. 316 and 320, Wiley (1983).] 

citric acid cycle slows down at its isocitrate dehydrogenase (activated by ADP) and α -ketoglutarate dehydrogenase (inhibited by ATP) steps, thereby causing the citrate concentration to build up. Citrate can leave the mitochondrion via a specific transport system and, *once in the cytosol, acts to restrain further carbohydrate breakdown by inhibiting PFK*.

b. Fatty Acid Oxidation Inhibits Glycolysis

As we shall see in Section 25-2, the oxidation of fatty acids is an aerobic process that produces acetyl-CoA, which enters the citric acid cycle, thereby increasing both the mitochondrial and cytoplasmic concentrations of citrate. The increased [acetyl-CoA] inhibits the pyruvate dehydrogenase complex, whereas the increased [citrate] inhibits phosphofructokinase, leading to a buildup of glucose-6-phosphate, which inhibits hexokinase (Fig. 22-49). This inhibition of glycolysis by fatty acid oxidation is called the **glucose–fatty acid cycle** or **Randle cycle** (after its discoverer, Philip Randle), although it is not, in fact, a cycle. The Randle cycle allows fatty acids to be utilized as the major fuel for oxidative metabolism in heart muscle, while conserving glucose for organs such as the brain, which require it.

C. Physiological Implications of Aerobic versus Anaerobic Metabolism

In 1861, Louis Pasteur observed that *when yeast are exposed to aerobic conditions, their glucose consumption and ethanol production drop precipitously* (the **Pasteur effect**; alcoholic fermentation in yeast to produce ATP, CO_2 , and ethanol are discussed in Section 17-3B). An analogous effect is observed in mammalian muscle; the concentration of lactic acid, the anaerobic product of muscle glycolysis, drops dramatically when cells switch to aerobic metabolism.

a. Hypoxia Causes an Increase in Glycolysis

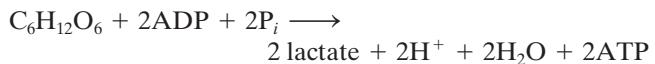
In the presence of sufficient oxygen, oxidative phosphorylation supplies most of the body's ATP needs. However, during **hypoxia** (when oxygen is limiting), glycolysis must be stimulated (with its inherent increased rate of glucose consumption; the reverse of the Pasteur effect) to supply the necessary ATP. F2,6P, the most potent activator of PFK-1, also participates in this process. The concentration of F2,6P, as we have seen (Section 18-3Fc), is regulated by the bifunctional enzyme PFK-2/FBPase-2. In its heart isozyme, the PFK-2 activity is stimulated by phosphorylation at its Ser 466. Among the enzymes that do so is **AMP-activated protein kinase** (AMPK; Sections 25-4Ba, 25-5, and 27-1). When oxygen deficiency prevents oxidative phosphorylation from providing sufficient ATP for heart function, as occurs in **ischemia** (insufficient blood flow), the resulting increased [AMP] activates AMPK. The consequent phosphorylation and hence activation of PFK-2 results in an increase of [F2,6P], thereby activating PFK-1 and thus glycolysis.

b. Aerobic ATP Production Is Far More Efficient than Anaerobic ATP Production

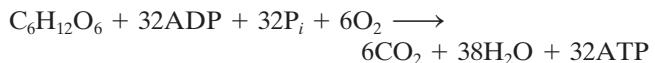
One reason for the decrease in glucose consumption on switching from anaerobic to aerobic metabolism is clear

from an examination of the stoichiometries of anaerobic and aerobic breakdown of glucose ($\text{C}_6\text{H}_{12}\text{O}_6$).

Anaerobic glycolysis:



Aerobic metabolism of glucose:



(2.5 ATP for each of the 10 NADH generated per glucose oxidized, 1.5 ATP for each of the 2 FADH_2 generated, 2 ATP produced in glycolysis, and $2\text{GTP} \rightleftharpoons 2\text{ATP}$ produced in the citric acid cycle.) Thus *aerobic metabolism is 16 times more efficient than anaerobic glycolysis in producing ATP*. The switch to aerobic metabolism therefore rapidly increases the ATP mass action ratio. As the ATP mass action ratio increases, the rate of electron transport decreases, which has the effect of increasing the $[\text{NADH}]/[\text{NAD}^+]$ ratio. The increases in [ATP] and [NADH] inhibit their target enzymes in the citric acid cycle and in the glycolytic pathway. *The activity of PFK, which is citrate- and adenine nucleotide-regulated and one of the rate-controlling enzymes of glycolysis, decreases manyfold on switching from anaerobic to aerobic metabolism. This accounts for the dramatic decrease in glycolysis.*

c. Anaerobic Glycolysis Has Advantages as Well as Limitations

Animals can sustain anaerobic glycolysis for only short periods of time. This is because PFK, which cannot function effectively much below pH 7, is inhibited by the acidification arising from lactic acid production. Despite this limitation and the low efficiency of glycolytic ATP production, *the enzymes of glycolysis are present in such great concentrations that when they are not inhibited, ATP can be produced much more rapidly than through oxidative phosphorylation.*

The different characteristics of aerobic and anaerobic metabolism permit us to understand certain aspects of cancer cell metabolism and cardiovascular disease.

d. Cancer Cell Metabolism

As Warburg first noted in 1926, certain cancer cells produce more lactic acid under aerobic conditions than do normal cells. This is because the glycolytic pathway in these cells produces pyruvate more rapidly than the citric acid cycle can accommodate. How can this happen given the interlocking controls on the system? One explanation is that these controls have broken down in cancer cells. Another is that their ATP utilization occurs at rates too rapid to be replenished by oxidative phosphorylation. This would alter the ratios of adenine nucleotides so as to relieve the inhibition of PFK-1. In addition, many cancer cell lines have a much larger [F2,6P] than do normal cells. These cells contain an inducible isozyme of PFK-2/FBPase-2 that has an AMPK-phosphorylatable site for activating PFK-2. Consequently, an [AMP] increase in these cells results in an increase in their [F2,6P], which further activates PFK-1 and

glycolysis. Efforts to understand the metabolic differences between cancer cells and normal cells may eventually lead to a treatment of certain forms of this devastating disease.

e. Cardiovascular Disease

Oxygen deprivation of certain tissues resulting from cardiovascular disease is of major medical concern. For example, two of the most common causes of human death, **myocardial infarction** (heart attack) and **stroke**, are caused by interruption of the blood (O_2) supply to a portion of the heart or the brain, respectively. It seems obvious why this should result in a cessation of cellular activity, but why does it cause cell death?

In the absence of O_2 , a cell, which must then rely only on glycolysis for ATP production, rapidly depletes its stores of phosphocreatine (a source of rapid ATP production; Section 16-4Cd) and glycogen. As the rate of ATP production falls below the level required by membrane ion pumps for the maintenance of proper intracellular ionic concentrations, the osmotic balance of the system is disrupted, so that the cell and its membrane-enveloped organelles begin to swell. The resulting overstretched membranes become permeable, thereby leaking their enclosed contents. [In fact, a useful diagnostic criterion for myocardial infarction is the presence in the blood of heart-specific enzymes, such as the H-type isozyme of lactate dehydrogenase (vs the M-type isozyme, which predominates in skeletal muscle; Section 17-3A), which leak out of necrotic (dead) heart tissue.] Moreover, the decreased intracellular pH that accompanies anaerobic glycolysis (because of lactic acid production; Section 17-3A) permits the released lysosomal enzymes (which are active only at acidic pH's) to degrade the cell contents. Thus, the cessation of metabolic activity results in irreversible cell damage. Rapidly respiring tissues, such as those of heart and brain, are particularly susceptible to such damage.

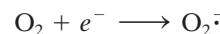
f. IF₁ Inhibits F₁F₀-ATPase during Hypoxia

Under hypoxic conditions, the proton-motive force across the inner mitochondrial membrane is reduced to the point that F₁F₀-ATPase would switch from the synthesis of ATP to its hydrolysis, resulting in a catastrophic loss of ATP. This is prevented through the interaction of the F₁F₀-ATPase with an 84-residue regulatory protein named **IF₁**. Under normal physiological conditions, IF₁ forms inactive tetramers and higher order oligomers. However, when the pH drops below 6.5, which occurs under anaerobic conditions due to lactic acid production, IF₁ forms dimers in which its almost entirely α helical subunits associate via an antiparallel coiled coil involving its residues 48 to 84. The X-ray structure of F₁ in complex with AMPPNP and IF₁, determined by Leslie and Walker, reveals that each N-terminal segment of the IF₁ dimer has bound to the α_{DP} - β_{DP} interface (Fig. 22-38b) of a separate F₁. This traps AMPPNP and presumably ATP in the β_{DP} binding site, which would prevent it from hydrolyzing ATP (which, since AMPPNP rather than ADP is bound to β_{DP} , suggests that this structure is that of a prehydrolysis step in the catalytic reaction). When oxygen becomes available, the cell

re-energizes and its pH increases, thereby causing IF₁ to dissociate from F₁F₀, which then commences synthesizing ATP.

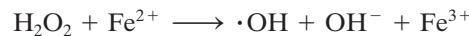
g. Partial Oxygen Reduction Produces Reactive Oxygen Species (ROS)

Although the four-electron reduction of O_2 by cytochrome c oxidase normally goes to completion, the enzyme infrequently releases partially reduced reactive oxygen species (ROS) that readily react with a variety of cellular components. The best known ROS is the superoxide radical, O_2^- . It is also produced by the occasional leakage of electrons from Complexes I and III:



Its production is enhanced under hypoxic conditions.

Superoxide radical is a precursor of other reactive species. Protonation of O_2^- yields **HO₂[•]**, a much stronger oxidant than O_2^- . The most potent oxygen species in biological systems is probably the **hydroxyl radical**, which forms from the relatively harmless hydrogen peroxide (H_2O_2):



The hydroxyl radical also forms through the reaction of superoxide with H_2O_2 :



ROS readily extract electrons from other molecules, converting them to free radicals and thereby initiating a chain reaction.

The random nature of ROS attacks makes it difficult to characterize their reaction products, but all classes of biological molecules are susceptible to oxidative damage caused by free radicals. The oxidation of polyunsaturated lipids in cells may disrupt the structures of membranes, and oxidative damage to DNA may result in point mutations. Enzyme function may also be compromised through radical reactions with amino acid side chains. Because the mitochondrion is the site of the bulk of the cell's oxidative metabolism, its lipids, DNA, and proteins bear the brunt of free radical-related damage.

Several degenerative diseases, including Parkinson's, Alzheimer's, and Huntington's diseases, are associated with oxidative damage to mitochondria. Such observations have led to the free-radical theory of aging, which holds that *free-radical reactions arising during the course of normal oxidative metabolism are at least partially responsible for the aging process*. In fact, individuals with congenital defects in their mitochondrial DNA suffer from a variety of symptoms typical of old age, including neuromotor difficulties, deafness, and dementia. These genetic defects may increase the susceptibility of mitochondria to ROS-generated damage.

h. Cells Are Equipped with Antioxidant Mechanisms

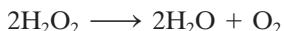
Antioxidants eliminate oxidative free radicals such as O_2^- and $\cdot OH$. In 1969, Irwin Fridovich discovered that the

enzyme **superoxide dismutase (SOD)**, which is present in nearly all cells, catalyzes the conversion of O_2^- to H_2O_2 :



Mitochondrial and bacterial SOD are both Mn^{2+} -containing tetramers; eukaryotic cytosolic SOD is a dimer that contains both Cu^{2+} and Zn^{2+} ions. Although the rate of nonenzymatic superoxide breakdown is $\sim 2 \times 10^5 M^{-1} \cdot s^{-1}$, that of the Cu, Zn -SOD-catalyzed reaction is $\sim 2 \times 10^9 M^{-1} \cdot s^{-1}$, close to the diffusion-controlled limit (Section 14-2Bb). This is apparently accomplished by electrostatic guidance of the negatively charged superoxide substrate into the enzyme's active site (Fig. 14-10).

H_2O_2 is degraded to water and oxygen by enzymes such as **catalase**, which catalyzes the reaction

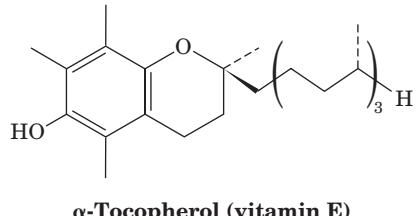


and **glutathione peroxidase**, which uses glutathione (GSH; Section 21-2Ba) as a reducing agent:



The latter enzyme also catalyzes the breakdown of organic hydroperoxides. Some types of glutathione peroxidase require Se for activity, which is one reason why Se appears to have antioxidant activity.

Other potential antioxidants are plant-derived compounds such as ascorbic acid (vitamin C; Section 11-1Cb) and **vitamin E**, a group of compounds whose most prominent member is **α -tocopherol**.



α -Tocopherol (vitamin E)

These compounds may help protect plants from oxidative damage during photosynthesis, a process in which H_2O is oxidized to O_2 (Section 24-2). However, clinical trials indicate that their use does not contribute to longevity in humans.

CHAPTER SUMMARY

1 The Mitochondrion Oxidative phosphorylation is the process through which the NADH and $FADH_2$ produced by nutrient oxidation are oxidized with the concomitant formation of ATP. The process takes place in the mitochondrion, an ellipsoidal organelle that is bounded by a permeable outer membrane and contains an impermeable and highly invaginated inner membrane that encloses the matrix. Enzymes of oxidative phosphorylation are embedded in the inner mitochondrial membrane. P_i is imported into the mitochondrion by a specific transport protein. Ca^{2+} import and Ca^{2+} export proteins operate to maintain a constant cytosolic $[Ca^{2+}]$. NADH's electrons are imported into the mitochondrion by shuttle systems such as the glycerophosphate shuttle and the malate-aspartate shuttle.

2 Electron Transport The standard free energy change for the oxidation of NADH by O_2 is $\Delta G^\circ = -218 \text{ kJ} \cdot \text{mol}^{-1}$, whereas that for the synthesis of ATP from ADP and P_i is $\Delta G^\circ = 30.5 \text{ kJ} \cdot \text{mol}^{-1}$. Consequently, the molar free energy of oxidation of NADH by O_2 is sufficient to power the synthesis of several moles of ATP under standard conditions. The electrons generated by oxidation of NADH and $FADH_2$ pass through four protein complexes, the electron-transport chain, with the coupled synthesis of ATP. Complexes I, III, and IV participate in the oxidation of NADH, producing ~ 2.5 ATPs per NADH, whereas $FADH_2$ oxidation, which involves Complexes II, III, and IV, produces only ~ 1.5 ATPs per $FADH_2$. Thus, the ratio of moles of ATP produced per mole of coenzyme oxidized by O_2 , the P/O ratio, is ~ 2.5 for NADH oxidation and ~ 1.5 for $FADH_2$ oxidation. The route taken by electrons through the electron-transport chain was elucidated, in part, through the use of electron-transport inhibitors. Rotenone and amytaf inhibit Complex I, antimycin inhibits Complex III, and CN^- inhibits Complex IV. Also involved were measurements of the reduction potentials of the electron-carrying prosthetic groups contained in the electron-transport complexes.

Complex I contains FMN and nine iron-sulfur clusters in a 45-subunit (in mammals) transmembrane protein complex. This L-shaped complex passes electrons from NADH to CoQ, a nonpolar small molecule that diffuses freely within the membrane. Complex II, which is also the citric acid cycle enzyme succinate dehydrogenase, also passes electrons to CoQ, in this case from succinate through FAD and three iron-sulfur clusters. The X-ray structure of Complex II indicates that its redox cofactors are arranged in a linear chain. $CoQH_2$ passes electrons to Complex III (cytochrome bc_1), a homodimeric complex whose protomers each contain two *b*-type hemes bound to a cytochrome *b* subunit, a Rieske iron-sulfur protein (ISP), and a cytochrome c_1 . An electron from cytochrome c_1 of Complex III is passed to the Cu_A center of Complex IV (cytochrome *c* oxidase) via the peripheral membrane protein cytochrome *c*. This electron is then passed to cytochrome *a*, which, in turn, passes it to a binuclear center composed of heme a_3 and Cu_B , which reduces O_2 to H_2O . This process occurs in four 1-electron steps that pump four protons from the mitochondrial matrix/bacterial cytoplasm to the intermembrane space/periplasm. Complexes I, III, and IV form supercomplexes that increase the efficiency of electron transport.

3 Oxidative Phosphorylation The mechanism by which the free energy released by the electron-transport chain is stored and utilized in ATP synthesis is described by the chemiosmotic hypothesis. This hypothesis states that the free energy released by electron transport is conserved by the generation of an electrochemical proton gradient across the inner mitochondrial membrane (bacterial cell membrane; outside positive and acidic), which is harnessed to synthesize ATP. The proton gradient is created and maintained by the obligatory outward translocation of H^+ across the inner mitochondrial membrane as electrons travel through Complexes I, III, and IV.

Complex III pumps protons via a redox loop mechanism called the Q cycle, a bifurcated double cycle in which one

molecule of CoQH_2 is oxidized to CoQ and then is rereduced to CoQH_2 by a second molecule of CoQH_2 in a process that collectively transfers four protons from the inside to the outside while oxidizing one molecule of CoQH_2 to CoQ . Electrons are transferred between the two CoQ 's, which are bound at different sites, Q_o and Q_i , as well as between the CoQH_2 bound at Q_o and cytochrome c_1 via the ISP, which undergoes a conformational change in doing so. Complex IV contains no ($\text{H}^+ + e^-$) carriers such as CoQH_2 and hence translocates proteins via a proton pump mechanism. Bacteriorhodopsin, the best characterized proton pump, translocates protons in a light driven process. This involves a trans to cis isomerization of bacteriorhodopsin's retinal prosthetic group on absorbing a photon, followed by the translocation of a proton through the hydrophilic central channel of this transmembrane protein via a process that involves conformational and $\text{p}K$ changes of the polar groups lining the channel as the retinal relaxes to its ground state. Complex IV is thought to pump protons via a similar mechanism that is driven by the changes in the redox state of its heme $a_3\text{-Cu}_B$ binuclear center as it reduces O_2 to H_2O .

The energy stored in the electrochemical proton gradient is utilized by proton-translocating ATP synthase (Complex V, F_1F_0 -ATPase) in the synthesis of ATP via the binding change mechanism, by coupling this process to the exergonic transport of H^+ back to the inside. Mitochondrial proton-translocating ATP synthase consists of two oligomeric components: F_1 ($\alpha_3\beta_3\gamma\delta\epsilon$), a peripheral membrane protein that appears as "lolipops" in electron micrographs of the inner mitochondrial membrane, and F_0 (ab_2c_{12} in *E. coli*), an integral membrane

protein that contains the proton channel. The conformational changes that promote the synthesis of ATP from $\text{ADP} + \text{P}_i$ arise through the demonstrated rotation of the γ subunit relative to the catalytic $\alpha_3\beta_3$ assembly that contains the enzyme's three active sites. The γ subunit is attached to a ring of c subunits in F_0 , whose rotation is driven by the passage of protons between it and the α subunit.

Compounds such as 2,4-dinitrophenol are uncouplers of oxidative phosphorylation because they carry H^+ across the mitochondrial membrane, thereby dissipating the proton gradient and allowing electron transport to continue without concomitant ATP synthesis. Brown fat mitochondria contain a regulated uncoupling system that, under hormonal control, generates heat instead of ATP.

4 Control of ATP Production Under aerobic conditions, the rate of ATP synthesis by oxidative phosphorylation is regulated, in a phenomenon known as acceptor control, by the ATP mass action ratio. ATP synthesis is tightly coupled to the oxidation of NADH and FADH_2 by the electron-transport chain. Glycolysis and the citric acid cycle are coordinately controlled so as to produce NADH and FADH_2 only at a rate required to meet the system's demand for ATP. IF_1 inhibits the ATP hydrolysis by mitochondrial F_1F_0 -ATPase that would otherwise occur under hypoxic conditions. Incomplete and side reactions of Complexes I, III, and IV produce damaging reactive oxygen species (ROS) that are largely eliminated through the actions of several cellular enzymes, most notably superoxide dismutase.

REFERENCES

Historical Overview

- Ernster, L. and Schatz, G., Mitochondria: A historical review, *J. Cell Biol.* **91**, 227s–255s (1981).
 Fruton, J.S., *Molecules and Life*, pp. 262–396, Wiley–Interscience (1972).
 Krebs, H., *Otto Warburg, Cell Physiologist, Biochemist, and Eccentric*, Clarendon Press (1981). [A biography of one of the pioneers in the biochemical study of respiration, by a distinguished student.]
 Prebble, J., Peter Mitchell and the ox phos wars, *Trends Biochem. Sci.* **27**, 209–212 (2002).
 Racker, E., *A New Look at Mechanisms in Bioenergetics*, Academic Press (1976). [A fascinating personal account by one of the outstanding contributors to the field.]

General

- Nicholls, D.G. and Ferguson, S.J., *Bioenergetics* (3rd ed.), Academic Press (2002). [An authoritative monograph devoted almost entirely to the mechanism of oxidative phosphorylation and the techniques used to elucidate it.]
 Schäfer, G. and Penefsky, H. (Eds.), *Bioenergetics*, Springer (2008).
 Schultz, B.E. and Chan, S.I., Structures and proton-pumping strategies of mitochondrial respiratory enzymes, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 23–65 (2001).

Mitochondria

- Goodsell, D.S., Mitochondrion, *Biochem. Mol. Biol. Educ.* **38**, 134–140 (2010). [An illustrated guide to the mitochondrion.]
 Frey, T.G. and Mannella, C.A., The internal structure of mitochondria, *Trends Biochem. Sci.* **25**, 319–324 (2000).

- Frey, T.G., Perkins, G.A., and Ellisman, M.H., Electron tomography of membrane-bound cellular organelles, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 199–224 (2006).
 Logan, D.C., The mitochondrial compartment, *J. Exp. Botany* **57**, 1225–1243 (2006).
 Scheffler, I.E., *Mitochondria* (2nd ed.), Wiley-Liss (2008).

Electron Transport

- Beinert, H., Holm, R.H., and Münck, E., Iron-sulfur clusters: Nature's modular, multipurpose structures, *Science* **277**, 653–659 (1997).
 Belevich, I. and Verkhovsky, M.I., Molecular mechanism of proton translocation by cytochrome c oxidase, *Antioxidants Redox Signaling* **10**, 1–29 (2008). [A comprehensive review.]
 Brandt, U., Energy converting NADH:quinone oxidoreductase (Complex I), *Annu. Rev. Biochem.* **75**, 69–92 (2006).
 Collman, J.P., Rapta, M., Bröring, M., Raptova, L., Schwenninger, R., Boitrel, B., Fu, L., and L'Her, M., Close structural analogues of the cytochrome c oxidase $\text{Fe}_{a3}/\text{Cu}_B$ center show clean $4e^-$ electroreduction of O_2 to H_2O at physiological pH, *J. Am. Chem. Soc.* **121**, 1387–1388 (1999).
 Crofts, A.R., The cytochrome bc_1 complex: Function in the context of structure, *Annu. Rev. Physiol.* **66**, 689–733 (2004).
 Efremov, R.G., Baradaran, R., and Sazanov, L.A., The architecture of respiratory complex I, *Nature* **465**, 441–445 (2010).
 Hinkle, P.C., P/O ratios of mitochondrial oxidative phosphorylation, *Biochim. Biophys. Acta* **1706**, 1–11 (2005).
 Hosler, J.P., Ferguson-Miller, S., and Mills, D.A., Energy transduction: Proton transfer through the respiratory complexes, *Annu. Rev.*

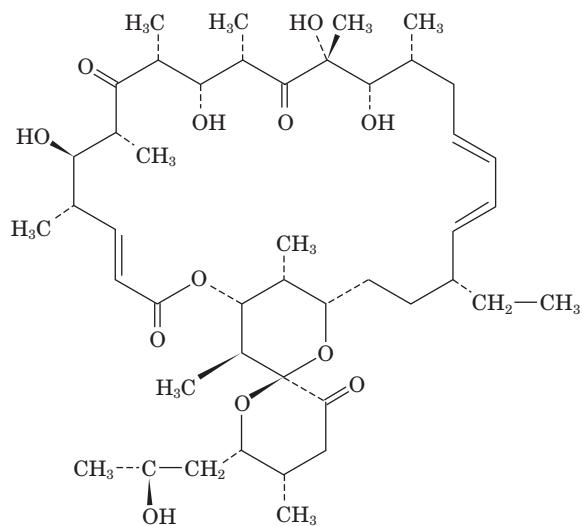
- Biochem.* **75**, 165–187 (2006). [A review that focuses on cytochrome *c* oxidase.]
- Huang, L., Sun, G., Cobessi, D., Wang, A.C., Shen, J.T., Tung, E.Y., Anderson, V.E., and Berry, E.A., 3-Nitropropionic acid is a suicide inhibitor of mitochondrial respiration that, upon oxidation by Complex II, forms a covalent adduct with a catalytic base arginine in the active site of the enzyme, *J. Biol. Chem.* **281**, 5965–5972 (2006). [X-ray structure of chicken Complex II.]
- Hunte, C., Koepke, J., Lange, C., Rossmanith, T., and Michel, H., Structure at 2.3 Å resolution of the cytochrome *bc*₁ complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment, *Structure* **8**, 669–684 (2000).
- Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H., Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*, *Nature* **376**, 660–669 (1995).
- Iwata, S., Lee, J.W., Okada, K., Lee, J.K., Iwata, M., Rasmussen, B., Link, T.A., Ramaswamy, S., and Jap, B.K., Complete structure of the 11-subunit bovine mitochondrial cytochrome *bc*₁ complex, *Science* **281**, 64–71 (1998).
- Johnson, D.C., Dean, D.R., Smith, A.D., and Johnson, M.K., Structure, function, and formation of biological iron–sulfur clusters, *Annu. Rev. Biochem.* **74**, 247–281 (2005).
- Lenaz, G., Fato, R., Genova, M.L., Bergamini, C., Bianchi, C., and Biondi, A., Mitochondrial complex I: Structural and functional aspects, *Biochim. Biophys. Acta* **1757**, 1406–1420 (2006).
- Michel, H., Behr, J., Harrenga, A., and Kannt, A., Cytochrome *c* oxidase: Structure and spectroscopy, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 329–356 (1998).
- Moser, C.C., Keske, J.M., Warncke, K., Farid, R.S., and Dutton, L.S., Nature of biological electron transfer, *Nature* **355**, 796–802 (1992).
- Osyczka, A., Moser, C.C., and Dutton, P.L., Fixing the Q cycle, *Trends Biochem. Sci.* **30**, 176–182 (2005).
- Radermacher, M., Ruiz, T., Clason, T., Benjamin, S., Brandt, U., and Zickerman, V., The three dimensional structure of complex I from *Yarrowia lipolytica*: A highly dynamic enzyme, *J. Struct. Biol.* **154**, 269–279 (2006).
- Sazanov, L.A. and Hinchliffe, P., Structure of the hydrophilic domain of respiratory Complex I from *Thermus thermophilus*, *Science* **311**, 1430–1436 (2006); and Sazanov, L.A., Respiratory Complex I: Mechanistic and structural insights provided by the crystal structure of the hydrophilic domain, *Biochemistry* **46**, 2275–2288 (2007).
- Schäfer, E., Dencher, N.A., Vonck, J., and Parcej, D.N., Three dimensional structure of the respiratory supercomplex I₁III₂IV₁ from bovine heart mitochondria, *Biochemistry* **46**, 12579–12585 (2007); and Vonck, J., and Schäfer, E., Supramolecular organization of protein complexes in the mitochondrial inner membrane, *Biochim. Biophys. Acta* **1793**, 117–124 (2009).
- Solmaz, S.R.N. and Hunte, C., Structure of Complex III with bound cytochrome *c* in reduced state and definition of a minimal core interface for electron transfer, *J. Biol. Chem.* **283**, 17542–17549 (2008); and Lange, C. and Hunte, C., Crystal structure of the yeast cytochrome *bc*₁ complex with its bound substrate cytochrome *c*, *Proc. Natl. Acad. Sci.* **99**, 2800–2805 (2002).
- Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., Bartlam, M., and Rao, Z., Crystal structure of mitochondrial respiratory protein Complex II, *Cell* **121**, 1043–1057 (2005).
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S., The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å, *Science* **272**, 1136–1144 (1996).
- Xia, D., Yu, C.-A., Kim, H., Xia, J.-Z., Kachurin, A.M., Zhang, L., Yu, L., and Deisenhofer, J., Crystal structure of the cytochrome *bc*₁ complex from heart mitochondria, *Science* **277**, 60–66 (1997).
- Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Légar, C., Byrne, B., Cecchini, G., and Iwata, S., Architecture of succinate dehydrogenase and reactive oxygen species generation, *Science* **299**, 700–704 (2003).
- Yoshikawa, S., et al., Redox-coupled crystal structural changes in bovine heart cytochrome *c* oxidase, *Science* **280**, 1723–1729 (1998); and Yoshikawa, S., Beef heart cytochrome *c* oxidase, *Curr. Opin. Struct. Biol.* **7**, 574–579 (1997).
- Zhang, Z., Huang, L., Shulmeister, V.M., Chi, Y.-I., Kim, K.K., Huang, L.-W., Crofts, A.R., Berry, E.A., and Kim, S.H., Electron transfer by domain movement in cytochrome *bc*₁, *Nature* **392**, 677–684 (1998).
- Bacteriorhodopsin**
- Heberle, J., Proton transfer reactions across bacteriorhodopsin and along the membrane, *Biochim. Biophys. Acta* **1458**, 135–147 (2000).
- Kühlbrandt, W., Bacteriorhodopsin—the movie, *Nature* **406**, 569–570 (2000).
- Lanyi, J.K., Bacteriorhodopsin, *Annu. Rev. Physiol.* **66**, 665–688 (2004).
- Oxidative Phosphorylation**
- Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E., Structure at 2.8 Å resolution of F₁-ATPase from bovine heart mitochondria, *Nature* **370**, 621–628 (1994).
- Boyer, P.D., The binding change mechanism for ATP synthase—some probabilities and possibilities, *Biochim. Biophys. Acta* **1140**, 215–250 (1993).
- Boyer, P.D., The ATP synthase—a splendid molecular machine, *Annu. Rev. Biochem.* **66**, 717–749 (1997).
- Capaldi, R. and Aggeler, R., Mechanism of F₁F₀-type ATP synthase, a biological rotary motor, *Trends Biochem. Sci.* **27**, 154–160 (2002).
- Dickson, V.K., Silvester, J.A., Fearnley, I.M., Leslie, A.G.W., and Walker, J.E., On the structure of the stator of the mitochondrial ATP synthase, *EMBO J.* **25**, 2911–2918 (2006).
- Gibbons, C., Montgomery, M.G., Leslie, A.G.W., and Walker, J.E., The structure of the central stalk in bovine F₁-ATPase at 2.4 Å resolution, *Nature Struct. Biol.* **7**, 1055–1061 (2000).
- Hausrath, A.C., Capaldi, R.A., and Matthews, B.M., The conformation of the ε- and γ-subunits of *Escherichia coli* F₁ ATPase, *J. Biol. Chem.* **276**, 47227–47232 (2001). [The X-ray structure of the *E. coli* F₁.]
- Itoh, H., Takahashi, A., Adachi, K., Noji, H., Yasuda, R., Yoshida, M., and Kinoshita, K., Jr., Mechanically driven ATP synthesis by F₁-ATPase, *Nature* **427**, 465–468 (2004).
- Klingenberg, M., Mechanism and evolution of the uncoupling protein of brown adipose tissue, *Trends Biochem. Sci.* **15**, 108–112 (1990).
- Ma, J., Flynn, T.C., Cui, Q., Leslie, A.G.W., Walker, J.E., and Karplus, M., A dynamic analysis of the rotation mechanism for the conformational change in F₁-ATPase, *Structure* **10**, 921–931 (2002).
- Meier, T., Polzer, P., Diederichs, K., Welte, W., and Dimroth, P., Structure of the rotor ring of F-type Na⁺-ATPase from *Illyobacter tartaricus*, *Science* **308**, 659–662 (2005).
- Mitchell, P., Vectorial chemistry and the molecular mechanics of chemiosmotic coupling: Power transmission by proticity, *Biochem. Soc. Trans.* **4**, 398–430 (1976).
- Nicholls, D.G. and Rial, E., Brown fat mitochondria, *Trends Biochem. Sci.* **9**, 489–491 (1984).

- Noji, H. and Yoshida, M., The rotary engine in cell ATP synthase, *J. Biol. Chem.* **276**, 1665–1668 (2001).
- Rastogi, V.K. and Girvin, M.E., Structural changes linked to proton translocation by subunit c of the ATP synthase, *Nature* **402**, 262–268 (1999); and Girvin, M.E., Rastogi, V.K., Abildgaard, F., Markley, J.L., and Fillingame, R.H., Solution structure of the transmembrane H⁺-transporting subunit c of the ATP synthase, *Biochemistry* **37**, 8817–8824 (1998).
- Rubinstein, J.L., Walker, J.E., and Henderson, R., Structure of the mitochondrial ATP synthase by electron cryomicroscopy, *EMBO J.* **22**, 6182–6192 (2003).
- Sambongi, Y., Iko, Y., Tanabe, M., Omote, H., Iwamoto-Kihara, A., Ueda, I., Yanagida, T., Wada, Y., and Futai, M., Mechanical rotation of the c subunit oligomer in ATP synthase (F₀F₁): Direct observation, *Science* **286**, 1722–1724 (1999).
- Stock, D., Gibbons, C., Arechaga, I., Leslie, A.G.W., and Walker, J.E., The rotary mechanism of ATP synthase, *Curr. Opin. Struct. Biol.* **10**, 672–679 (2000).
- Verkhovsky, M.I., Jasaitis, A., Verkhovskaya, M.L., Morgan, J.E., and Wikström, M., Proton translocation by cytochrome c oxidase, *Nature* **400**, 480–483 (1999).
- von Ballamoos, C., Wiedenmann, A., and Dimroth, P., Essentials for ATP synthase by F₁F₀ ATP synthases, *Annu. Rev. Biochem.* **78**, 649–672 (2009); and von Ballamoos, C., Cook, G.M., and Dimroth, P., Unique rotary ATP synthase and its biological diversity, *Annu. Rev. Biophys.* **37**, 43–64 (2008).
- Wilkins, S., Rotary molecular motors, *Adv. Prot. Chem.* **71**, 345–382 (2005).
- Yasuda, R., Noji, H., Kinoshita, K., Jr., and Yoshida, M., F₁-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps, *Cell* **93**, 1117–1124 (1998).
- Yoshida, M., Muneyuki, E., and Hisabori, T., ATP synthase—a marvelous rotary engine of the cell, *Nature Rev. Mol. Cell Biol.* **2**, 669–677 (2001); and Noji, H. and Yoshida, M., The rotary machine in the cell ATP synthase, *J. Biol. Chem.* **276**, 1665–1668 (2001).
- Control of ATP Production**
- Brown, G.C., Control of respiration and ATP synthesis in mammalian mitochondria and cells, *Biochem. J.* **284**, 1–13 (1992).
- Cabezon, E., Montgomery, M.G., Leslie, A.G.W., and Walker, J.E., The structure of bovine F₁-ATPase in complex with its regulatory protein IF₁, *Nature Struct. Biol.* **10**, 744–750 (2003).
- Celi, F.S., Brown adipose tissue—when it pays to be inefficient, *New Engl. J. Med.* **360**, 1553–1556 (2009).
- Chesney J., Mitchell, R., Benigni, F., Bacher, M., Spiegel, L., Al-Abed, Y., Han, J.H., Metz, C., and Bucala, R., An inducible gene product for 6-phosphofructo-2-kinase with an AU-rich instability element: role in tumor cell glycolysis and the Warburg effect, *Proc. Natl. Acad. Sci.* **96**, 3047–3052 (2000).
- Harris, D.A. and Das, A.M., Control of mitochondrial ATP synthesis in the heart, *Biochem. J.* **280**, 561–573 (1991).
- Jeneson, J.A.L., Westerhoff, H.V., and Kushmerick, M.J., A metabolic control analysis of kinetic controls in ATP free energy metabolism in contracting skeletal muscle, *Am. J. Physiol. Cell Physiol.* **279**, C813–C832 (2000).
- Marsin, A.-S., Bertrand, L., Rider, M.H., Deprez, J., Beauloye, C., Vincent, M.F., Van den Berghe, G., Carling, D., and Hue, L., Phosphorylation and activation of heart PFK-2 by AMPK has a role in the stimulation of glycolysis during ischaemia, *Curr. Biol.* **10**, 1247–1255 (2000).
- Marsin, A.-S., Bouzin, C., Bertrand, L., and Hue, L., The stimulation of glycolysis by hypoxia in activated monocytes is mediated by AMP-activated protein kinase and inducible 6-phosphofructo-3-kinase, *J. Biol. Chem.* **277**, 30778–30783 (2002).
- Randle, P. J., Regulatory interactions between lipids and carbohydrates: The glucose fatty acid cycle after 35 years, *Diabetes/Metab. Rev.* **14**, 263–283 (1998).
- Ricquier, D. and Bouillaud, F., The mitochondrial uncoupling protein: Structural and genetic studies, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 83–108 (1997).

PROBLEMS

- 1.** Rank the following redox-active coenzymes and prosthetic groups of the electron-transport chain in order of increasing affinity for electrons: cytochrome *a*, CoQ, FAD, cytochrome *c*, NAD⁺.
- 2.** Why is the oxidation of succinate to fumarate only associated with the production of two ATPs during oxidative phosphorylation, whereas the oxidation of malate to oxaloacetate is associated with the production of three ATPs?
- 3.** What is the thermodynamic efficiency of oxidizing FADH₂ so as to synthesize two ATPs under standard biochemical conditions?
- 4.** Sublethal cyanide poisoning may be reversed by the administration of nitrites. These substances oxidize hemoglobin, which has a relatively low affinity for CN[−], to methemoglobin, which has a relatively high affinity for CN[−]. Why is this treatment effective?
- 5.** Match the compound with its behavior: (1) rotenone, (2) dinitrophenol, and (3) antimycin. (a) Inhibits oxidative phosphorylation when the substrate is pyruvate but not when the substrate is succinate. (b) Inhibits oxidative phosphorylation when the substrate is either pyruvate or succinate. (c) Allows pyruvate to be oxidized by mitochondria even in the absence of ADP.
- 6.** **Nigericin** is an ionophore (Section 20-2C) that exchanges K⁺ for H⁺ across membranes. Explain how the treatment of functioning mitochondria with nigericin uncouples electron transport from oxidative phosphorylation. Does valinomycin, an ionophore that transports K⁺ but not H⁺, do the same? Explain.
- 7.** Why is it possible for electrons in an electron-transfer complex to flow from a redox center to one with a lesser value of \mathcal{E}° ?
- 8.** How do the P/O ratios for NADH differ in ATP synthases that contain 10 and 15 c subunits?
- 9.** The difference in pH between the internal and external surfaces of the inner mitochondrial membrane is 1.4 pH units (external side acidic). If the membrane potential is 0.06 V (inside negative), what is the free energy released on transporting 1 mol of protons back across the membrane? How many protons must be transported to provide enough free energy for the synthesis of 1 mol of ATP (assume standard biochemical conditions)?
- *10.** (a) A simplistic interpretation of the Q cycle would predict that the proton pumping efficiency of cytochrome *bc*₁ would be reduced by no more than 50% in the presence of saturating amounts of antimycin. Explain. (b) Indicate why cytochrome *bc*₁ is nearly 100% inhibited by antimycin.

11. The antibiotic oligomycin B



Oligomycin B

binds to the F_0 subunit of the mitochondrial F_1F_0 -ATPase and thereby prevents it from synthesizing ATP [note that oligomycin-sensitivity conferral protein (OSCP), the mitochondrial counterpart of the *E. coli* δ subunit (Fig. 22-41), does not bind oligomycin

B.] Explain why: (a) Submitochondrial particles from which F_1 has been removed are permeable to protons. (b) Addition of oligomycin B to F_1 -depleted submitochondrial particles decreases this permeability severalfold.

12. Oligomycin B (see Problem 11) and cyanide both inhibit oxidative phosphorylation when the substrate is either pyruvate or succinate. Dinitrophenol can be used to distinguish between these inhibitors. Explain.

13. The *E. coli* F_1F_0 -ATPase cannot synthesize ATP when Met 23 of its γ subunit is mutated to Lys. Yet the F_1 component of this complex still exhibits rotation of its γ subunit relative to its $\alpha_3\beta_3$ spheroid when it is supplied with ATP. Suggest a reason for these effects.

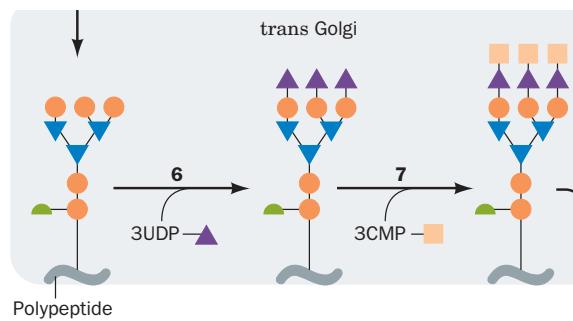
14. For the oxidation of a given amount of glucose, does nonshivering thermogenesis by brown fat or shivering thermogenesis by muscle produce more heat?

15. What is the advantage of hormones activating a lipase to stimulate nonshivering thermogenesis in brown fat rather than activating UCP1 directly?

16. How does atroctyloside affect mitochondrial respiration? (Hint: See Section 20-4C.)

17. Certain unscrupulous operators offer, for a fee, to freeze recently deceased individuals in liquid nitrogen until medical science can cure the disease from which they died. What is the biochemical fallacy of this procedure?

Other Pathways of Carbohydrate Metabolism



CHAPTER 23

1 Gluconeogenesis

- A. The Gluconeogenesis Pathway
- B. Regulation of Gluconeogenesis
- C. The Cori Cycle

2 The Glyoxylate Cycle

3 Biosynthesis of Oligosaccharides and Glycoproteins

- A. Lactose Synthesis
- B. Glycoprotein Synthesis

4 The Pentose Phosphate Pathway

- A. Oxidative Reactions of NADPH Production
- B. Isomerization and Epimerization of Ribulose-5-Phosphate
- C. Carbon–Carbon Bond Cleavage and Formation Reactions
- D. Control of the Pentose Phosphate Pathway
- E. Glucose-6-Phosphate Dehydrogenase Deficiency

This chapter completes our study of carbohydrate metabolism in animals; photosynthesis, which occurs only in plants and certain bacteria, is the subject of Chapter 24.

1 GLUCONEOGENESIS

Glucose occupies a central role in metabolism, both as a fuel and as a precursor of essential structural carbohydrates and other biomolecules. The brain and red blood cells are almost completely dependent on glucose as an energy source. Yet the liver's capacity to store glycogen is only sufficient to supply the brain with glucose for about half a day under fasting or starvation conditions. Thus, *when fasting, most of the body's glucose needs must be met by gluconeogenesis (literally, new glucose synthesis), the biosynthesis of glucose from noncarbohydrate precursors.* Indeed, isotopic labeling studies determining the source of glucose in the blood during a fast showed that gluconeogenesis is responsible for 64% of total glucose production over the first 22 hours of the fast and accounts for almost all the glucose production by 46 hours. Thus, gluconeogenesis provides a substantial fraction of the glucose produced in fasting humans, even after a few hours' fast. Gluconeogenesis occurs in liver and, to a smaller extent, in kidney.

The noncarbohydrate precursors that can be converted to glucose include the glycolysis products lactate and pyruvate, citric acid cycle intermediates, and the carbon skeletons of most amino acids. First, however, all these substances must be converted to oxaloacetate, the starting material for gluconeogenesis (Fig. 23-1). The only amino acids that cannot be converted to oxaloacetate in animals are leucine and lysine because their breakdown yields only acetyl-CoA (Section 26-3F). There is no pathway in animals for the net conversion of acetyl-CoA to oxaloacetate. Likewise, fatty acids cannot serve as glucose precursors in animals because most fatty acids are degraded completely to acetyl-CoA (Section 25-2C). Unlike animals, however, plants do contain a pathway for the conversion of acetyl-CoA to oxaloacetate, the **glyoxylate cycle** (Section 23-2), so that fatty acids can serve as a plant cell's only carbon source. Glycerol, a triacylglycerol breakdown product, is converted to glucose via synthesis of the glycolytic intermediate dihydroxyacetone phosphate, as described in Section 25-1.

Heretofore, we have dealt with many aspects of carbohydrate metabolism. We have seen how the free energy of glucose oxidation is sequestered in ATP through glycolysis, the citric acid cycle, and oxidative phosphorylation. We have also studied the mechanism by which glucose is stored as glycogen for future use and how glycogen metabolism is controlled in response to the needs of the organism. In this chapter, we examine several other carbohydrate metabolism pathways of importance:

1. Gluconeogenesis, through which noncarbohydrate precursors such as lactate, pyruvate, glycerol, and amino acids are converted to glucose.

2. The glyoxylate cycle, through which plants convert acetyl-CoA to glucose.

3. Oligosaccharide and glycoprotein biosynthesis, through which oligosaccharides are synthesized and added to specific amino acid residues of proteins.

4. The pentose phosphate pathway, an alternate pathway of glucose degradation, which generates **NADPH**, the source of reducing equivalents in reductive biosynthesis, and **ribose-5-phosphate**, the sugar precursor of the nucleic acids.

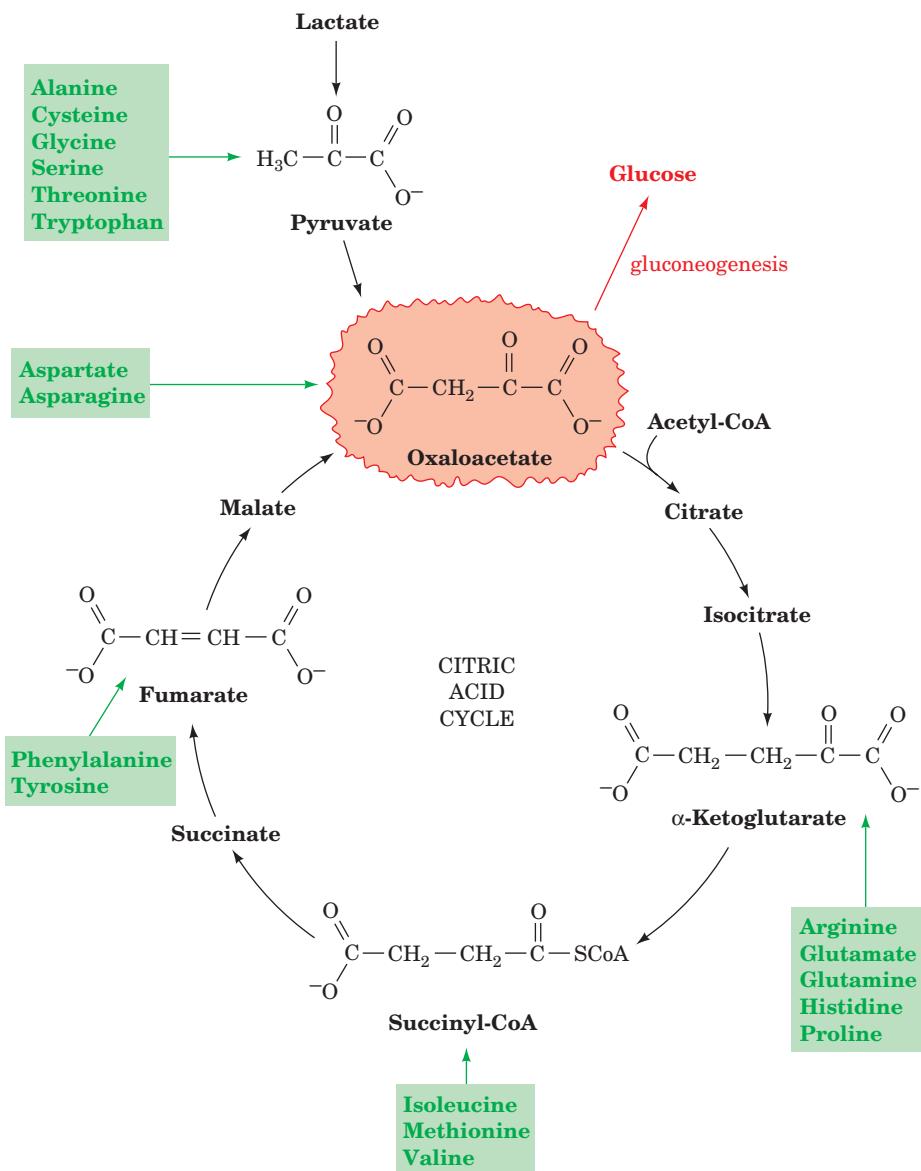


Figure 23-1 Pathways converting lactate, pyruvate, and citric acid cycle intermediates to oxaloacetate. The carbon skeletons of

all amino acids but leucine and lysine may be, at least in part, converted to oxaloacetate and thus to glucose by these reactions.

A. The Gluconeogenesis Pathway

Gluconeogenesis utilizes glycolytic enzymes. Yet three of these enzymes, hexokinase, phosphofructokinase (PFK), and pyruvate kinase, catalyze reactions with large negative free energy changes in the direction of glycolysis. These reactions must therefore be replaced in gluconeogenesis by reactions that make glucose synthesis thermodynamically favorable. Here, as in glycogen metabolism (Section 18-1D), we see the recurrent theme that *biosynthetic and degradative pathways differ in at least one reaction. This not only permits both directions to be thermodynamically favorable under the same physiological conditions but allows the pathways to be independently controlled so that one direction can be activated while the other is inhibited.*

a. Pyruvate Is Converted to Oxaloacetate before Conversion to Phosphoenolpyruvate

The formation of phosphoenolpyruvate (PEP) from pyruvate, the reverse of the pyruvate kinase reaction, is endergonic and therefore requires free energy input. This is accomplished by first converting the pyruvate to oxaloacetate. Oxaloacetate is a “high-energy” intermediate whose exergonic decarboxylation provides the free energy necessary for PEP synthesis. The process requires the participation of two enzymes (Fig. 23-2):

1. **Pyruvate carboxylase** catalyzes the ATP-driven formation of oxaloacetate from pyruvate and HCO_3^- .
2. **PEP carboxykinase (PEPCK)** converts oxaloacetate

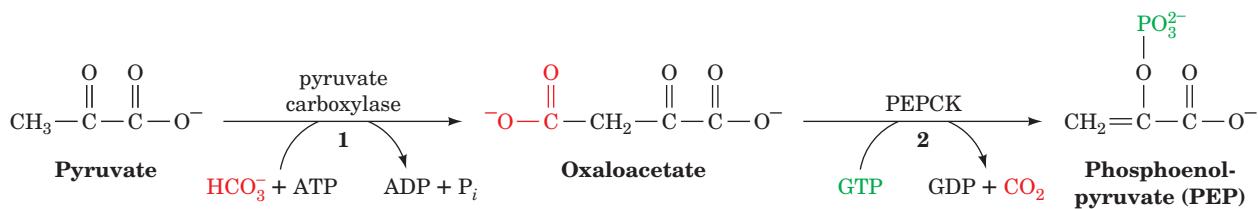


Figure 23-2 Conversion of pyruvate to oxaloacetate and then to phosphoenolpyruvate. The enzymes involved are (1) pyruvate carboxylase and (2) PEP carboxykinase (PEPCK).

to PEP in a reaction that uses GTP as a phosphorylating agent.

b. Pyruvate Carboxylase Has a Biotin Prosthetic Group

Pyruvate carboxylase, discovered in 1959 by Merton Utter, is a tetrameric protein of identical ~ 130 -kD subunits, each of which has a **biotin** prosthetic group. Biotin (Fig. 23-3a) functions as a CO_2 carrier by acquiring a carboxyl substituent at its **ureido group** (Fig. 23-3b). Biotin is covalently bound to the enzyme by an amide linkage between the carboxyl group of its valerate side chain and the ϵ -amino group of an enzyme Lys residue to form a **biocytin**

(alternatively, **biotinyllysine**) residue (Fig. 23-3b). The biotin ring system is therefore at the end of a 16-Å-long flexible arm, much like that of the lipoic acid prosthetic group in the pyruvate dehydrogenase multienzyme complex (Section 21-2Ac).

Biotin, which was first identified in 1935 as a growth factor in yeast, is an essential human nutrient. Its nutritional deficiency is rare, however, because it occurs in many foods and is synthesized by intestinal bacteria. Human biotin deficiency almost always results from the consumption of large quantities of raw eggs. This is because egg whites contain a protein, **avidin**, that binds biotin so tightly (dissociation constant, $K = 10^{-15} \text{ M}$) as to prevent its intestinal absorption (cooked eggs do not cause this problem because cooking denatures avidin). The presence of avidin in eggs is thought to inhibit the growth of microorganisms in this highly nutritious environment. The avidin homolog **streptavidin**, which is secreted by *Streptomyces avidinii*, is used as a linking agent in numerous biotechnological applications (e.g., Section 22-3Ce) because of its particularly high affinity for biotin.

c. The Pyruvate Carboxylase Reaction

The pyruvate carboxylase reaction occurs in two phases (Fig. 23-4):

Phase I Biotin is carboxylated at its N1 atom by bicarbonate ion in a three-step reaction in which the hydrolysis of ATP to ADP + P_i functions, via the intermediate formation of **carboxyphosphate**, to dehydrate bicarbonate. This yields free CO_2 , which has sufficient free energy to carboxylate biotin. The resulting carboxyl group is activated relative to bicarbonate (ΔG° for its cleavage is $-19.7 \text{ kJ} \cdot \text{mol}^{-1}$) and can therefore be transferred without further free energy input.

Phase II The activated carboxyl group is transferred from carboxybiotin to pyruvate in a three-step reaction to form oxaloacetate.

These two reaction phases occur on different active sites of the same enzyme.

d. Acetyl-CoA Regulates Pyruvate Carboxylase

Oxaloacetate synthesis is an anaplerotic (filling up) reaction that increases citric acid cycle activity (Section 21-5b). Accumulation of the citric acid cycle substrate acetyl-CoA is therefore indicative of the need for more oxaloacetate.

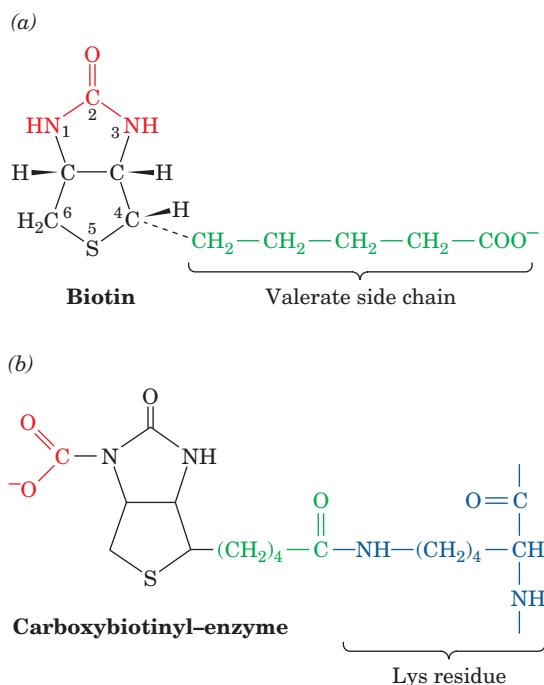


Figure 23-3 Biotin and carboxybiotinyl-enzyme.

(a) Biotin consists of an imidazoline ring that is cis-fused to a tetrahydrothiophene ring bearing a valerate side chain. The chirality at each of its three asymmetric centers is indicated. Positions 1, 2, and 3 constitute a ureido group. (b) In carboxybiotinyl-enzyme, N1 of the biotin ureido group is the carboxylation site. Biotin is covalently attached to carboxylases by an amide linkage between its valeryl carboxyl group and the ϵ -amino group of an enzyme Lys side chain to form biocytin.

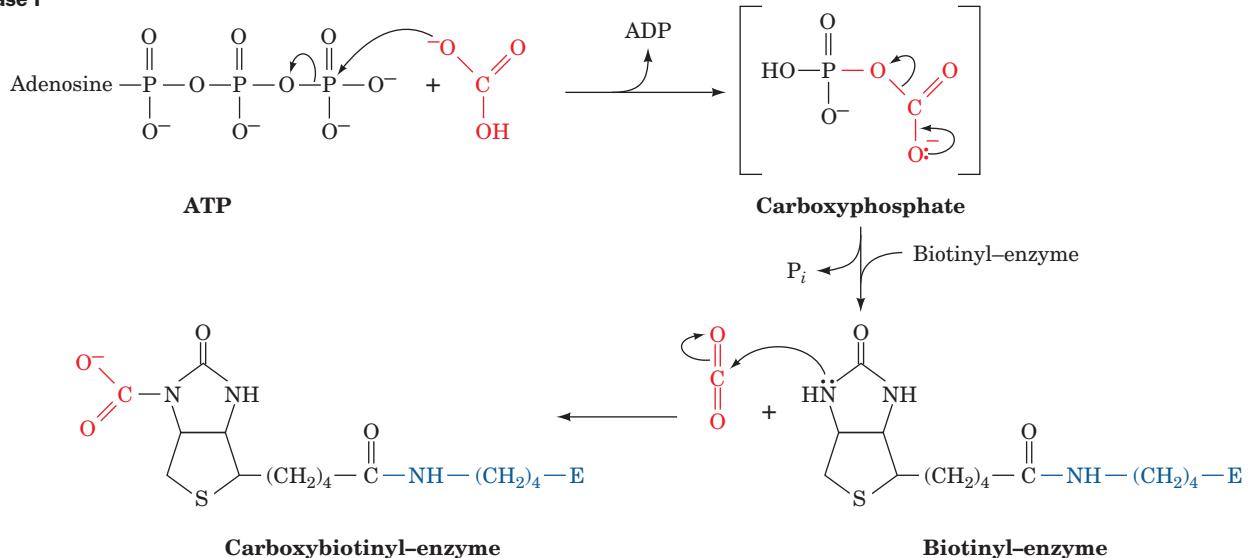
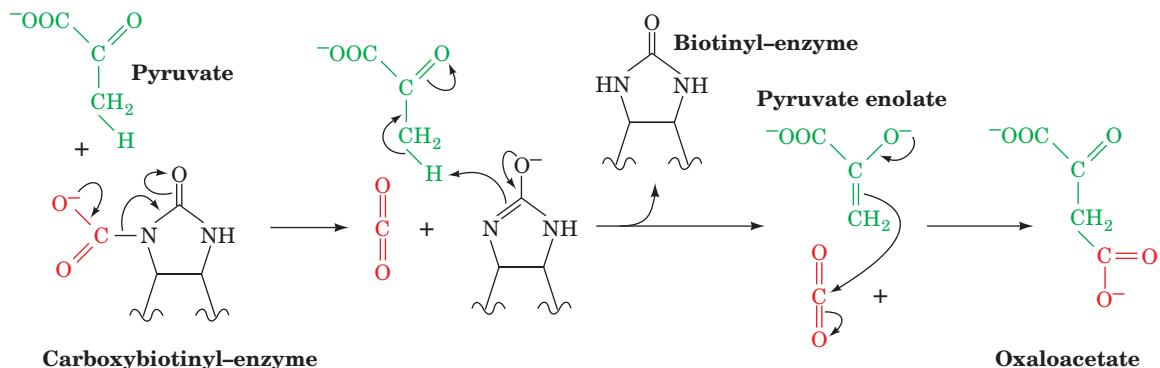
Phase I**Phase II**

Figure 23-4 Two-phase reaction mechanism of pyruvate carboxylase. **Phase I** is a three-step reaction in which carboxyphosphate is formed from bicarbonate and ATP, followed by the generation of CO_2 on the enzyme, which then carboxylates biotin. **Phase II** is a three-step reaction in which

CO_2 is produced at the active site via the elimination of the biotinyl enzyme, which accepts a proton from pyruvate to generate pyruvate enolate. This, in turn, nucleophilically attacks the CO_2 , yielding oxaloacetate. [After Knowles, J.R., *Annu. Rev. Biochem.* **58**, 217 (1989).]

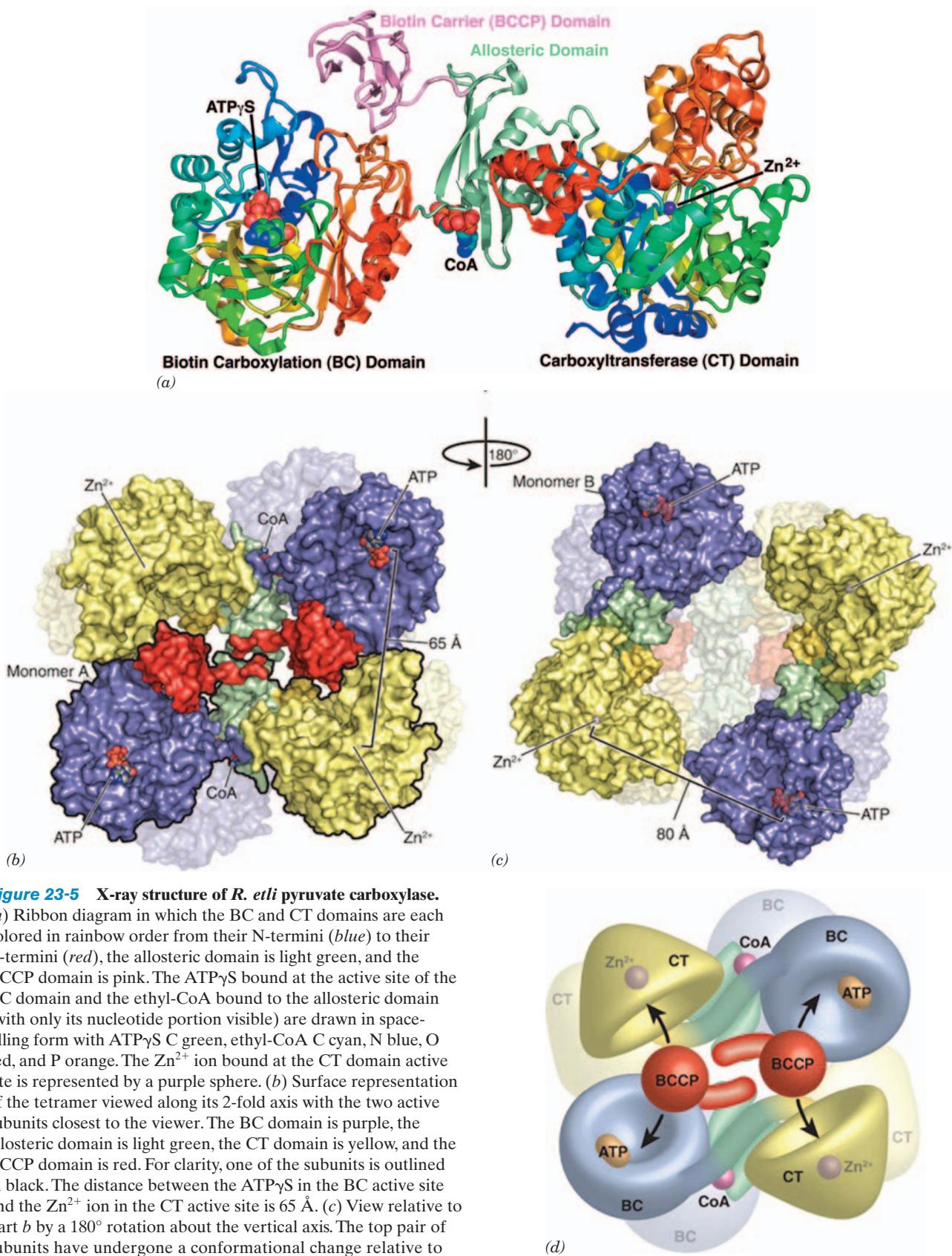
Indeed, acetyl-CoA is a powerful allosteric activator of pyruvate carboxylase; the enzyme is all but inactive without bound acetyl-CoA. If, however, the citric acid cycle is inhibited (by ATP and NADH, whose presence in high concentrations indicates a satisfied demand for oxidative phosphorylation; Section 21-4), oxaloacetate instead undergoes gluconeogenesis.

e. X-Ray Structure of Pyruvate Carboxylase Suggests a Mechanism for Carboxyl Group Transfer

The X-ray structure of pyruvate carboxylase from the soil bacterium *Rhizobium etli* in complex with ATP γ S and ethyl-CoA (acetyl-CoA with an ethyl group in place of its acetyl group), determined by Ivan Rayment, reveals that each subunit of this homotetrameric, 1154-residue protein consists of four domains (Fig. 23-5a): a biotin carboxylation (BC) domain (residues 1–465) that carries out Phase I of

the pyruvate carboxylase reaction (Fig. 23-4); an allosteric domain (residues 466–510 and 1001–1073) that binds acetyl-CoA; a carboxyltransferase (CT) domain (residues 511–1000) that catalyzes Phase II of the pyruvate carboxylase reaction (Fig. 23-4); and a domain named the biotin carboxyl carrier protein (BCCP; residues 1074–1154) to which the enzyme's biotin prosthetic group is covalently linked via Lys 1119 (Fig. 23-3). The active site of the BC domain is marked by its bound ATP γ S, the CT domain is mainly an α/β barrel (viewed from the side in Fig. 23-5a) whose active site is located in the mouth of the barrel and is marked by its bound Zn^{2+} ion, and the allosteric domain binds the ethyl-CoA, which like acetyl-CoA, activates the enzyme.

The biotin prosthetic group on the BCCP domain is disordered and hence not visible. Nevertheless it is clear that the $\sim 80\text{-}\text{\AA}$ distance between a subunit's active sites is too



domain of its neighboring subunit. The view and domain colors are the same as in Part b. [Part a based on an X-ray structure by and Parts b, c, and d courtesy of Ivan Rayment, University of Wisconsin, Madison, Wisconsin. PDBid 2QF7.]

large to be bridged by the 16-Å-long carboxybiotinyl arm (Fig. 23-3b). However, the BCCP domain is attached to the enzyme by a flexible polypeptide linker that is 34 Å long, much like that linking the lipoyl domain(s) to each dihydrolipoyl transacetylase (E2) subunit of the pyruvate dehydrogenase complex (Section 21-2Ae). Even so, it would require a dramatic movement of the entire BCCP domain to transfer substrate between the two active sites of a single subunit. How, then, does a BCCP domain translocate its carboxybiotin group between the active sites of a BC domain and a CT domain?

Pyruvate carboxylase's homotetrameric structure is required for its enzymatic activity; isolated subunits are catalytically inactive. However, the tetramer has only 2-fold symmetry because the top pair of subunits in Fig. 23-5b differ in conformation from the top pair in Fig. 23-5c by a 40° rotation and a 40-Å translocation of the BC domain relative to the CT domain of the same subunit. Indeed, the BCCP domains in the top pair of subunits in Fig. 23-5c are disordered, probably because the allosteric domains on these subunits do not bind ethyl-CoA. Consequently, the distance between active sites from adjacent subunits is 65 Å for the top pair in Fig. 23-5b, whereas it is 80 Å for the top pair in Fig. 23-5c. This suggests the model drawn in Fig. 23-5d in which each BCCP domain at the top of Fig. 23-5b shuttles CO₂ in the form of carboxybiotin from the active site of the BC domain on the same subunit to the CT domain on the adjacent subunit, whereas the other two subunits are inactive. This is an unusual example of allosteric activation coupled with negative cooperativity. It may permit pyruvate carboxylase to carry out efficient catalysis in association with other metabolic enzymes.

The foregoing model is supported by experiments involving two mutant forms of pyruvate carboxylase: K1119Q, which eliminates the biotinylation of the BC domain; and K718Q, which impairs the Phase II reaction. Tetramers of each of these mutant subunits exhibited 0.1% and 4% of the wild-type enzymatic activity, respectively. However, mixed tetramers exhibit 20% activity, thus indicating the formation of neighboring pairs of functional BC and CT domains.

f. PEP Carboxykinase

PEPCK, a monomeric ~630-residue enzyme, catalyzes the GTP-driven decarboxylation of oxaloacetate to form PEP and GDP (Fig. 23-6). Note that the CO₂ that carboxylates pyruvate to yield oxaloacetate is eliminated in the formation of PEP. Oxaloacetate may therefore be considered to be “activated” pyruvate, with CO₂ and biotin facilitating the activation at the expense of ATP hydrolysis. Acetyl-CoA is similarly activated for fatty acid biosynthesis through such a carboxylation-decarboxylation process (forming malonyl-CoA; Section 25-4B). In general, β-keto acids may be considered “high-energy” compounds because of the high free energy of decarboxylation of the β-carboxyl group. The enolates they generate are used to form carbon–carbon bonds in fatty acid biosynthesis or phosphoenolpyruvate here in gluconeogenesis.

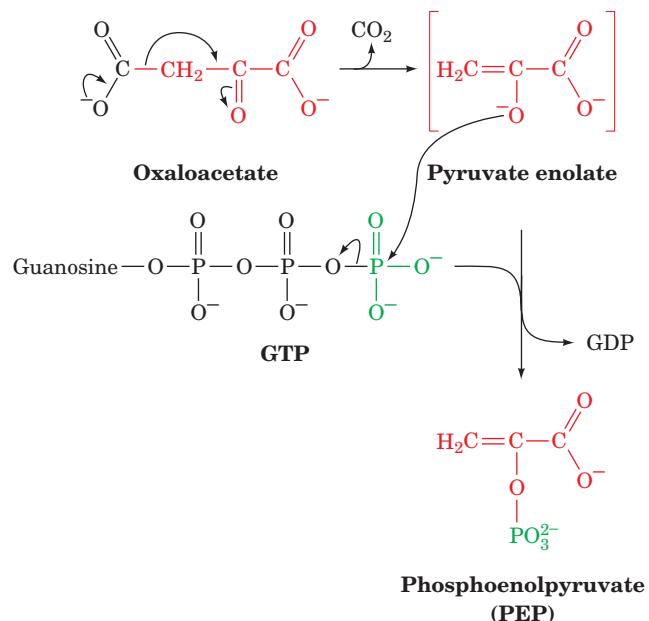


Figure 23-6 The PEPCK mechanism. Decarboxylation of oxaloacetate (a β-keto acid) forms a resonance-stabilized enolate anion whose oxygen atom attacks the γ phosphoryl group of GTP forming PEP and GDP.

g. Gluconeogenesis Requires Metabolite Transport between Mitochondria and Cytosol

The generation of oxaloacetate from pyruvate or citric acid cycle intermediates occurs only in the mitochondrion, whereas the enzymes that convert PEP to glucose are cytosolic. The cellular location of PEPCK varies with the species. In mouse and rat liver it is located almost exclusively in the cytosol, in pigeon and rabbit liver it is mitochondrial, and in guinea pig and humans it is more or less equally distributed between both compartments. In order for gluconeogenesis to occur, either oxaloacetate must leave the mitochondrion for conversion to PEP or the PEP formed there must enter the cytosol.

PEP is transported across the mitochondrial membrane by specific membrane transport proteins. There is, however, no such transport system for oxaloacetate. It must first be converted either to aspartate (Fig. 23-7, Route 1) or to malate (Fig. 23-7, Route 2), for which mitochondrial transport systems exist (Section 22-1B). The difference between these two routes involves the transport of NADH reducing equivalents. The **malate dehydrogenase** route (Route 2) results in the transport of reducing equivalents from the mitochondrion to the cytosol, since it utilizes mitochondrial NADH and produces cytosolic NADH. The **aspartate aminotransferase** route (Route 1) does not involve NADH. Cytosolic NADH is required for gluconeogenesis so, under most conditions, the route through malate is a necessity. If the gluconeogenic precursor is lactate, however (Section 23-1C), its oxidation to pyruvate generates cytosolic NADH, so that either transport route may then be used. Of course, as we have seen, during oxidative metabolism

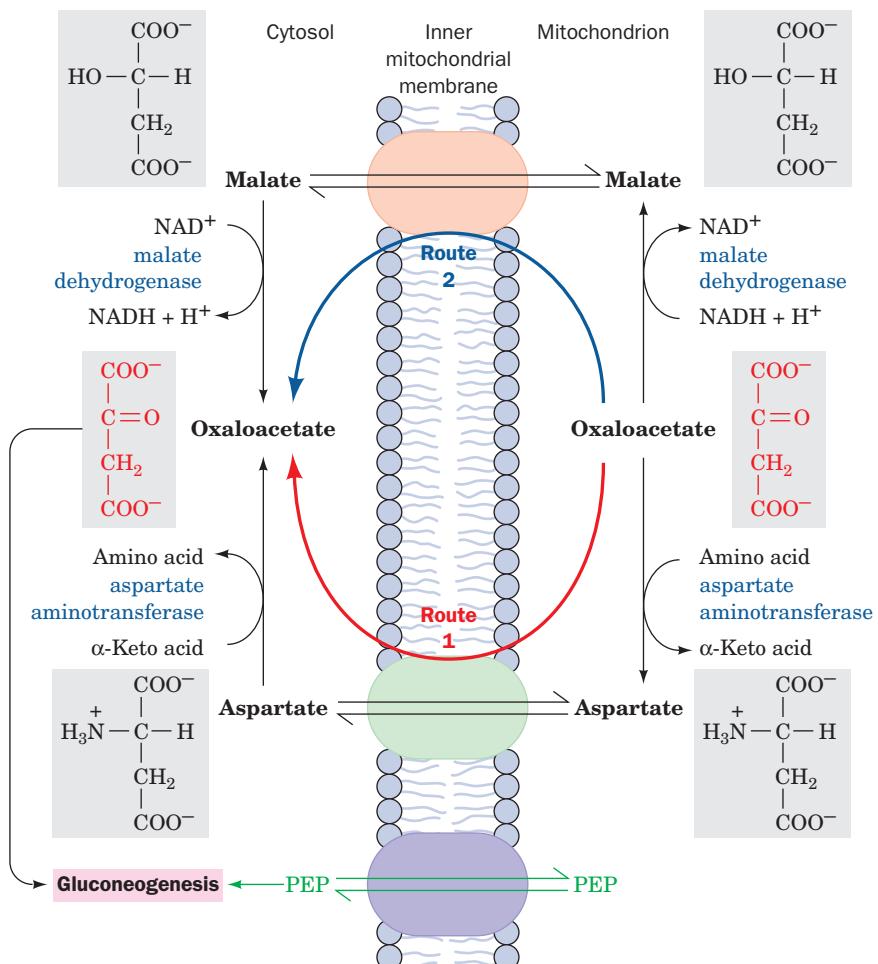


Figure 23-7 Transport of PEP and oxaloacetate from the mitochondrion to the cytosol. PEP is directly transported between these compartments. Oxaloacetate, however, must first be converted to either aspartate through the action of **aspartate aminotransferase** (Route 1) or to malate by malate

dehydrogenase (Route 2). Route 2 involves the mitochondrial oxidation of NADH followed by the cytosolic reduction of NAD^+ and therefore also transfers NADH reducing equivalents from the mitochondrion to the cytosol.  **See the Animated Figures**

the two routes may also alternate (with Route 2 reversed) to form the malate–aspartate shuttle, which transports NADH reducing equivalents into the mitochondrion (Section 22-1Bc).

In the liver, where the urea cycle occurs (Section 26-2), a third route, a modification of Route 1, may be followed for transporting oxaloacetate into the cytosol. The aspartate that enters the cytosol by Route 1 may be converted to fumarate as part of the urea cycle (Fig. 26-8), instead of being transaminated. Fumarate is then hydrated to malate and dehydrogenated to oxaloacetate by cytosolic equivalents of citric acid cycle enzymes. This third route generates cytosolic NADH in the same way as does Route 2.

h. Hydrolytic Reactions Bypass PFK and Hexokinase

The opposing pathways of gluconeogenesis and glycolysis utilize many of the same enzymes (Fig. 23-8). However, the free energy change is highly unfavorable in the gluconeogenic

direction at two other points in the pathway in addition to the pyruvate kinase reaction: the PFK reaction and the hexokinase reaction. At these points, instead of generating ATP by reversing the glycolytic reactions, FBP and G6P are hydrolyzed, releasing P_i in exergonic processes catalyzed by **fructose-1,6-bisphosphatase (FBPase)** and **glucose-6-phosphatase**, respectively. *Glucose-6-phosphatase is unique to liver and kidney, permitting them to supply glucose to other tissues.*

Because of the presence of separate gluconeogenic enzymes at the three irreversible steps in the glycolytic conversion of glucose to pyruvate, both glycolysis and gluconeogenesis are rendered thermodynamically favorable. This is accomplished at the expense of the free energy of hydrolysis of two molecules each of ATP and GTP per molecule of glucose synthesized by gluconeogenesis in addition to that which would be consumed by the direct reversal of glycolysis.

Glycolysis:**Gluconeogenesis:****Overall:**

Such free energy losses in a cyclic process are thermodynamically inescapable. They are the price that must be paid to maintain independent regulation of the two pathways.

B. Regulation of Gluconeogenesis

If both glycolysis and gluconeogenesis were to proceed in an uncontrolled manner, the net effect would be a futile cycle wastefully hydrolyzing ATP and GTP. This does not occur. Rather, *these pathways are reciprocally regulated so as to meet the needs of the organism*. In the fed state, when the blood glucose level is high, the liver is geared toward fuel conservation: Glycogen is synthesized and the glycolytic pathway and pyruvate dehydrogenase are activated, breaking down glucose to acetyl-CoA for fatty acid biosynthesis and fat storage. In the fasted state, however, the liver maintains the blood glucose level both by glycogen breakdown and by reversing the flux through glycolysis toward gluconeogenesis [using mainly protein degradation products via the **glucose-alanine cycle** (Section 26-1Ad) and glycerol from triacylglycerol hydrolysis (Section 25-1e)].

a. Glycolysis and Gluconeogenesis Are Controlled by Allosteric Interactions and Covalent Modifications

The rate and direction of glycolysis and gluconeogenesis are controlled at the points in these pathways where the forward and reverse directions can be independently regulated: the reactions catalyzed by (1) hexokinase/glucose-6-phosphatase, (2) PFK/FBPase, and (3) pyruvate kinase/pyruvate carboxylase-PEPCK (Fig. 23-8). Table 23-1 lists these regulatory enzymes and their regulators. The dominant mechanisms are allosteric interactions and cAMP-dependent covalent modifications (phosphorylation/dephosphorylation; Section 18-3). cAMP-dependent covalent modification renders this system sensitive to control by glucagon and other hormones that alter cAMP levels.

One of the most important allosteric effectors involved in the regulation of glycolysis and gluconeogenesis is fructose-2,6-bisphosphate (F2,6P), which activates PFK and inhibits FBPase (Section 18-3F). The concentration of F2,6P is controlled by its rates of synthesis and breakdown by phosphofructokinase-2 (PFK-2) and fructose bisphosphatase-2 (FBPase-2), respectively. Control of the activities of PFK-2 and FBPase-2 is therefore an important aspect of gluconeogenic regulation even though these

enzymes do not catalyze reactions of the pathway. PFK-2 and FBPase-2 activities, which occur on separate domains of the same bifunctional enzyme, are subject to allosteric

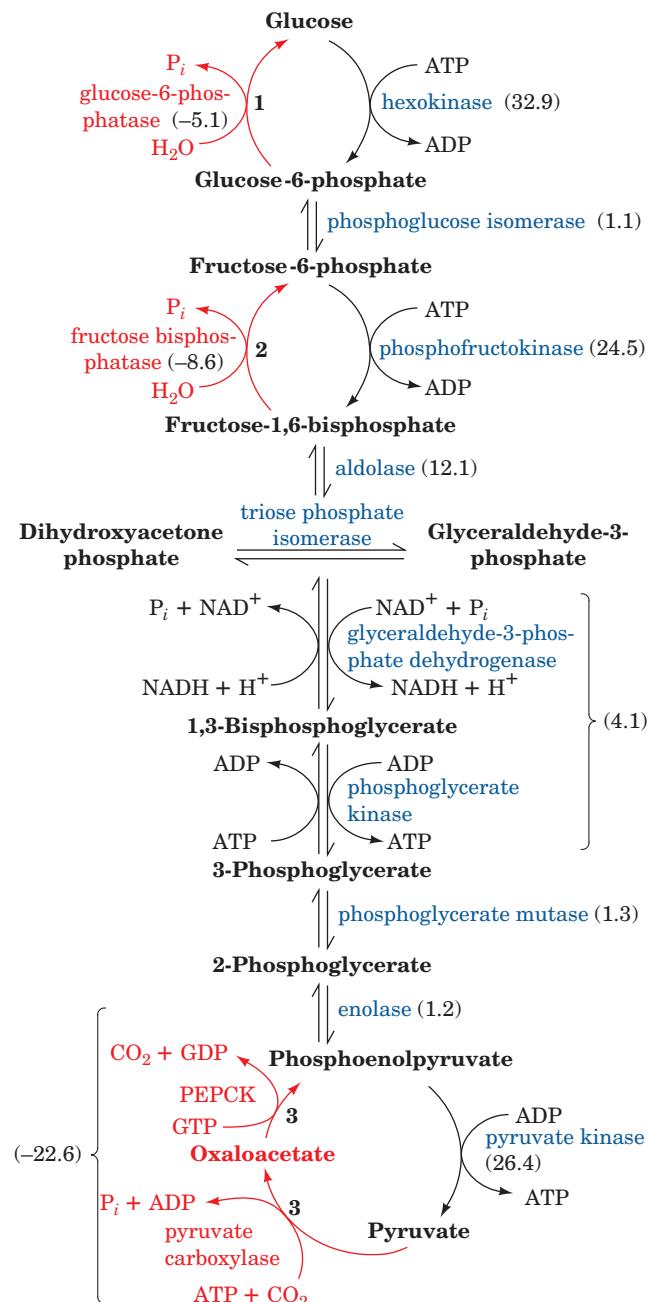


Figure 23-8 Pathways of gluconeogenesis and glycolysis. The three numbered steps, which are catalyzed by different enzymes in gluconeogenesis, have red arrows. The ΔG 's for the reactions in the direction of gluconeogenesis under physiological conditions in liver are given in parentheses in $\text{kJ} \cdot \text{mol}^{-1}$. [ΔG 's obtained from Newsholme, E.A. and Leech, A.R., *Biochemistry for the Medical Sciences*, p. 448, Wiley (1983).] 

Table 23-1 Regulators of Gluconeogenic Enzyme Activity

Enzyme	Allosteric Inhibitors	Allosteric Activators	Enzyme Phosphorylation	Protein Synthesis
PFK	ATP, citrate	AMP, F2,6P		
FBPase	AMP, F2,6P			
Pyruvate kinase	Alanine	F1,6P	Inactivates	
Pyruvate carboxylase		Acetyl-CoA		
PEPCK				Stimulated by glucagon, thyroid hormone, and glucocorticoids, and inhibited by insulin
PFK-2	Citrate	AMP, F6P, P_i	Inactivates	
FBPase-2	F6P	Glycerol-3-P	Activates	

regulation as well as control by covalent modifications (Table 23-1). Low levels of blood glucose result in hormonal activation of gluconeogenesis through regulation of [F2,6P] (Fig. 23-9).

Activation of gluconeogenesis in liver also involves inhibition of glycolysis at the level of pyruvate kinase. *Liver pyruvate kinase is inhibited both allosterically by alanine (a pyruvate precursor; Section 26-1Ad) and by phosphorylation.* Glycogen breakdown, in contrast, is stimulated by phosphorylation (Section 18-3C). Both pathways then flow toward G6P, which is converted to glucose for export to muscle and brain. Muscle pyruvate kinase, an isozyme of the liver enzyme, is not subject to these controls. Indeed, such controls would be counterproductive in muscle since this tissue lacks glucose-6-phosphatase and thus the ability to synthesize glucose via gluconeogenesis.

b. PEPCK Concentration Is Transcriptionally Controlled

PEPCK is the enzyme that catalyzes the first committed reaction of gluconeogenesis. It is therefore of interest (Table 23-1) that PEPCK's activity is controlled solely through the transcriptional regulation of the gene encoding it (transcriptional regulation is outlined in Section 5-4Aa and discussed in detail in Sections 31-3 and 34-3). In particular, the transcription of the PEPCK gene is stimulated by glucagon, glucocorticoids, and thyroid hormones, and is inhibited by insulin. For instance, the cAMP that is produced in response to stimulation of the liver by glucagon, in addition to its initiation of phosphorylation cascades (Section 18-3), induces the transcription of the PEPCK gene. Richard Hanson has shown that this occurs because the PEPCK gene **promoter** (a control region that precedes the transcriptional initiation site of genes encoding proteins; Section 5-4Aa) contains a specific DNA sequence called the **cAMP response element (CRE)** that is bound by a **transcription factor** named **CRE binding protein (CREB)**, but only when CREB is also binding cAMP (recall that a transcription factor is a protein that binds to a specific segment of its target promoter and, in doing so, activates RNA polymerase to initiate the transcription of the associated gene; Section 5-4Aa). However, the PEPCK gene promoter contains numerous other binding sites for specific transcription factors. Among them are the **thyroid hormone response element (TRE)**, which is bound by thyroid hormone receptor in complex with thyroid hormone (Section 19-1D), and the **glucocorticoid hormone response element (GRE)**, which is bound by the **glucocorticoid receptor** in complex with a glucocorticoid hormone (Sections 19-1G and 34-3Bn). In contrast, PEPCK gene transcription is strongly repressed by protein factors phosphorylated by the PI3K signaling cascade (Section 19-4D) initiated by the binding of insulin to the insulin receptor (these protein factors may repress transcription by interfering with the binding of the above transcription factors; the mechanism of insulin signaling is discussed in Sections 19-3Ac, 19-3Cg, and 19-4F). The rate of PEPCK mRNA production is determined by the integration of these various interactions and hence of the signals that caused them.

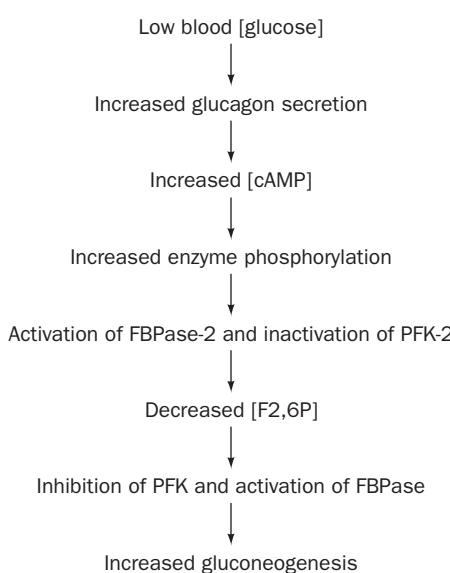


Figure 23-9 Hormonal regulation of [F2,6P]. This process activates gluconeogenesis in liver in response to low blood [glucose].

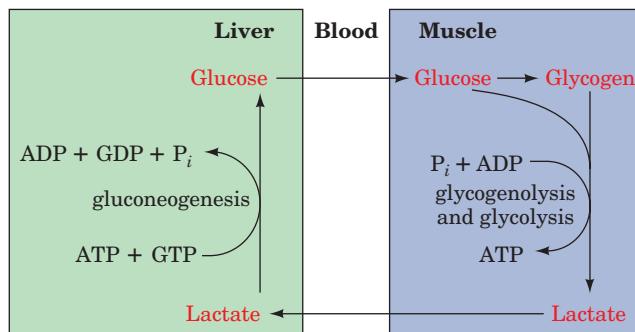


Figure 23-10 The Cori cycle. Lactate produced by muscle glycolysis is transported by the bloodstream to the liver, where it is converted to glucose by gluconeogenesis. The bloodstream carries the glucose back to the muscles, where it may be stored as glycogen. 

C. The Cori Cycle

Muscle contraction is powered by hydrolysis of ATP, which is then regenerated through oxidative phosphorylation in the mitochondria of slow-twitch (red) muscle fibers and by glycolysis yielding lactate in fast-twitch (white) muscle fibers. Slow-twitch fibers also produce lactate when ATP demand exceeds oxidative flux. The lactate is transferred, via the bloodstream, to the liver, where it is reconverted to pyruvate by lactate dehydrogenase and then to glucose by gluconeogenesis. Thus, through the intermediacy of the bloodstream, liver and muscle participate in a metabolic cycle known as the **Cori cycle** (Fig. 23-10) in honor of Carl and Gerty Cori, who first described it. This is the same ATP-consuming glycolysis/gluconeogenesis “futile cycle” we discussed above. Here, however, instead of occurring in the same cell, the two pathways occur in different organs. Liver ATP is used to resynthesize glucose from lactate produced in muscle. The resynthesized glucose is returned to the muscle, where it is stored as glycogen and used, on demand, to generate ATP for muscle contraction. The ATP utilized by the liver for this process is regenerated by oxidative phosphorylation. After vigorous exertion, it often takes at least 30 min for all of the lactate so produced to be converted to glycogen and the oxygen consumption rate to return to its resting level, a phenomenon known as **oxygen debt**.

2 THE GLYOXYLATE CYCLE

Plants, but not animals, possess enzymes that mediate the net conversion of acetyl-CoA to succinate, which is then converted, via malate, to oxaloacetate. This is accomplished via the **glyoxylate cycle** (Fig. 23-11), a pathway involving enzymes of the **glyoxysome** (a membranous plant organelle; Section 1-2Ad). The glyoxylate cycle involves five enzymes, three of which also participate in the citric acid cycle: citrate synthase, aconitase, and malate dehydrogenase.

The two other enzymes, isocitrate lyase and malate synthase, are unique to the cycle.

The glyoxylate cycle consists of five reactions (Fig. 23-11):

Reactions 1 and 2. Glyoxysomal oxaloacetate is condensed with acetyl-CoA to form citrate, which is isomerized to isocitrate as in the citric acid cycle. Since the glyoxysome contains no aconitase, Reaction 2 presumably takes place in the cytosol.

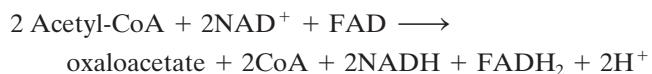
Reaction 3. Glyoxysomal **isocitrate lyase** cleaves the isocitrate to succinate and **glyoxylate** (hence the cycle’s name).

Reaction 4. Malate synthase, a glyoxysomal enzyme, condenses glyoxylate with a second molecule of acetyl-CoA to form malate.

Reaction 5. Glyoxysomal malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate by NAD^+ , thereby completing the cycle.

The glyoxylate cycle therefore results in the net conversion of two acetyl-CoA to succinate instead of to four molecules of CO_2 , as would occur in the citric acid cycle. The succinate produced in Reaction 3 is transported to the mitochondrion, where it enters the citric acid cycle and is converted to malate, which has two alternative fates: (1) It can be converted to oxaloacetate in the mitochondrion, continuing the citric acid cycle and thereby making the glyoxylate pathway an anaplerotic process (Section 21-5b); or (2) it can be transported to the cytosol, where it is converted to oxaloacetate for entry into gluconeogenesis.

The overall reaction of the glyoxylate cycle can be considered to be the formation of oxaloacetate from two molecules of acetyl-CoA.



Isocitrate lyase and malate synthase, the only enzymes of the glyoxylate pathway unique to plants, enable germinating seeds to convert their stored triacylglycerols, through acetyl-CoA, to glucose. It had long been assumed that this was a requirement of germination. However, a mutant of *Arabidopsis thaliana* (an oilseed plant) lacking isocitrate lyase, and hence unable to convert lipids to carbohydrate, nevertheless germinated. This process was only inhibited when the mutant plants were subjected to low light conditions. It therefore appears that the glyoxylate cycle’s importance in seedling growth is its anaplerotic function in providing 4-carbon units to the citric acid cycle, which can then oxidize the triacylglycerol-derived acetyl-CoA.

3 BIOSYNTHESIS OF OLIGOSACCHARIDES AND GLYCOPROTEINS

Oligosaccharides consist of monosaccharide units joined together by glycosidic bonds (linkages between C1, the anomeric carbon, of one unit and an OH group of a second

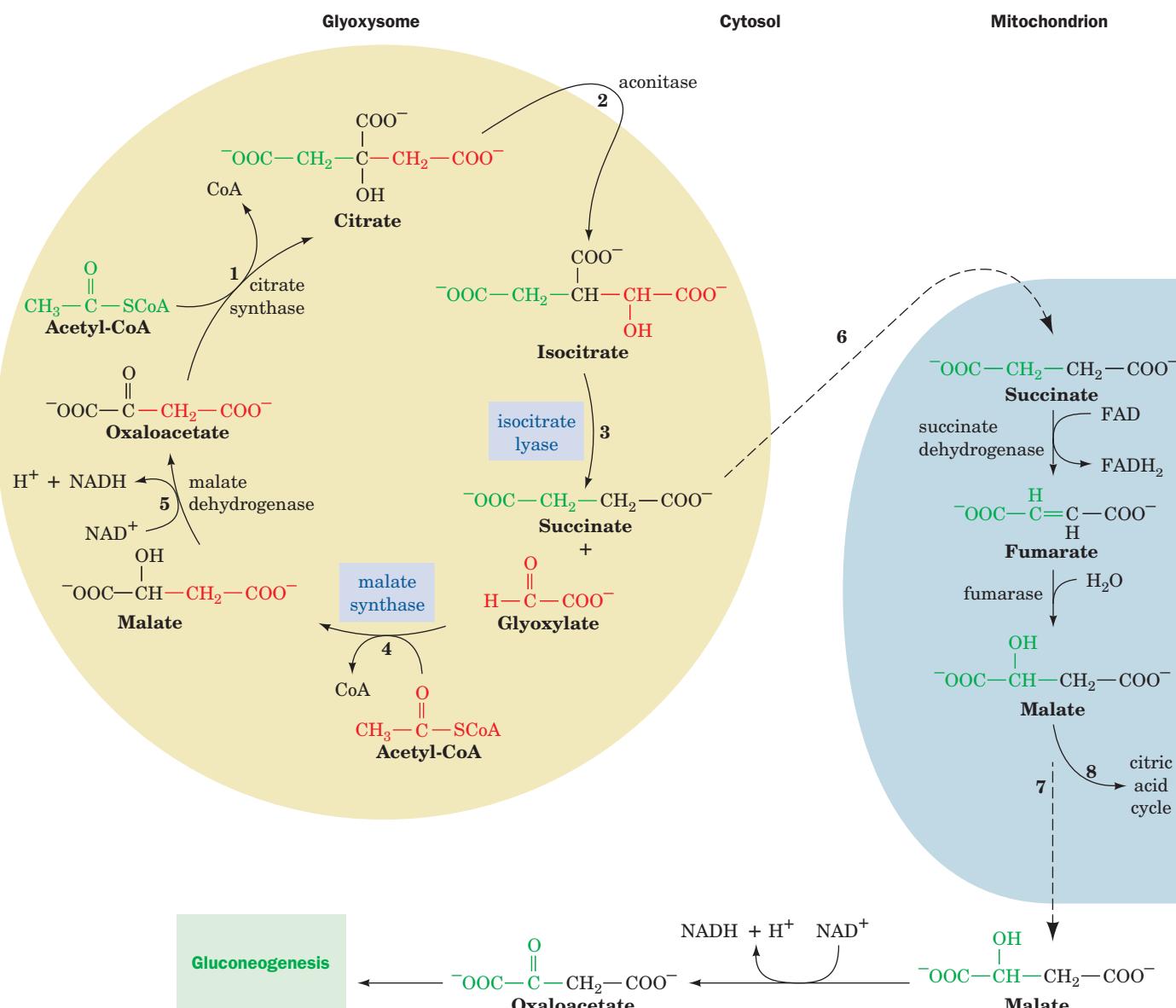


Figure 23-11 The glyoxylate cycle. The cycle results in the net conversion of two acetyl-CoA to succinate in the glyoxysome, which can be converted to malate in the mitochondrion for use in gluconeogenesis. Isocitrate lyase and malate synthase, enzymes unique to glyoxysomes (which occur only in plants), are boxed in blue. (1) Glyoxysomal citrate synthase catalyzes the condensation of oxaloacetate with acetyl-CoA to form citrate. (2) Cytosolic aconitase catalyzes the conversion of citrate to isocitrate. (3) Isocitrate lyase catalyzes the cleavage of isocitrate to succinate and glyoxylate. (4) Malate synthase catalyzes the

condensation of glyoxylate with acetyl-CoA to form malate. (5) Glyoxysomal malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate, completing the cycle. (6) Succinate is transported to the mitochondrion, where it is converted to malate via the citric acid cycle. (7) Malate is transported to the cytosol, where malate dehydrogenase catalyzes its oxidation to oxaloacetate, which can then be used in gluconeogenesis. (8) Alternatively, malate can continue in the citric acid cycle, making the glyoxylate cycle anaplerotic.

unit; Section 11-1C). About 80 different kinds of naturally occurring glycosidic linkages are known, most of which involve mannose, *N*-acetylglucosamine, *N*-acetylmuramic acid, glucose, galactose, fucose (6-deoxygalactose), *N*-acetylneurameric acid (sialic acid), and *N*-acetylgalactosamine (Section 11-1C). Glycosidic linkages also occur to lipids (e.g., glycosphingolipids; Section 12-1D) and proteins (glycoproteins; Section 11-3C).

Glycosidic bond formation requires free energy input under physiological conditions ($\Delta G^\circ = 16 \text{ kJ} \cdot \text{mol}^{-1}$). This free energy, as we have seen in the case of glycogen synthesis (Section 18-2B), is acquired through the conversion of monosaccharide units to nucleotide sugars. A nucleotide at a sugar's anomeric carbon atom is a good leaving group and thereby facilitates formation of a glycosidic bond to a second sugar unit via reactions catalyzed by **glycosyltransferases**.

(Fig. 23-12). The nucleotides that participate in monosaccharide transfers are UDP, GDP, and CMP; a given sugar is associated with only one of these nucleotides (Table 23-2).

A. Lactose Synthesis

Several disaccharides are synthesized for future use as metabolic fuels. In plants, the major fuel disaccharide is sucrose (Section 11-2B), whose synthesis is discussed in Section 24-3Ad. Typical of mammalian disaccharides is lactose [β -galactosyl-(1 \rightarrow 4)-glucose; milk sugar], which is synthesized in the mammary gland by **lactose synthase** (Fig. 23-13). The donor sugar is UDP-galactose, which is formed by epimerization of UDP-glucose (Section 17-5B). The acceptor sugar is glucose.

Lactose synthase consists of two subunits:

1. Galactosyltransferase, the catalytic subunit, which occurs in many tissues, where it catalyzes the reaction of UDP-galactose and *N*-acetylglucosamine to yield *N*-acetyllactosamine, a constituent of many complex oligosaccharides (see, e.g., Fig. 23-20, Reaction 6).

2. α -Lactalbumin, a mammary gland protein with no catalytic activity, which alters the specificity of galactosyltransferase such that it utilizes glucose as an acceptor, rather than *N*-acetylglucosamine, to form lactose instead of *N*-acetyllactosamine.

B. Glycoprotein Synthesis

Eukaryotic proteins destined for secretion, incorporation into membranes, or localization inside membranous organelles contain carbohydrates and are therefore classified as glycoproteins. *Glycosylation and oligosaccharide processing play an indispensable role in the sorting and the distribution of these proteins to their proper cellular destinations.* Their polypeptide components are ribosomally synthesized and processed by addition and modification of oligosaccharides.

The oligosaccharide portions of glycoproteins, as we have seen in Sections 11-3C and 12-3Bc, are classified into three groups:

1. *N*-Linked oligosaccharides, which are attached to their polypeptide chain by a β -*N*-glycosidic bond to the side chain N of an Asn residue in the sequence Asn-X-Ser

Table 23-2 Sugar Nucleotides and Their Corresponding Monosaccharides in Glycosyltransferase Reactions

UDP	GDP	CMP
<i>N</i> -Acetylgalactosamine	Fucose	Sialic acid
<i>N</i> -Acetylglucosamine	Mannose	
<i>N</i> -Acetylmuramic acid		
Galactose		
Glucose		
Glucuronic acid		
Xylose		

or Asn-X-Thr, where X is any amino acid residue except Pro (Fig. 23-14a).

2. *O*-Linked oligosaccharides, which are attached to their polypeptide chain through an α -*O*-glycosidic bond to the side chain O of a Ser or Thr residue (Fig. 23-14b) or, only in collagens (Section 8-2Bb), to that of a 5-hydroxylysine (Hyl) residue (Fig. 23-14c).

3. Glycosylphosphatidylinositol (GPI) membrane anchors, which are attached to their polypeptide chain through an amide bond between mannose-6-phosphoethanolamine and the C-terminal carboxyl group (Fig. 23-14d).

We shall consider the synthesis of these three types of oligosaccharides in turn.

a. *N*-Linked Glycoproteins Are Synthesized in Four Stages

N-Linked glycoproteins are formed in the endoplasmic reticulum and further processed in the Golgi apparatus. Synthesis of their carbohydrate moieties occurs in four stages:

1. Synthesis of a lipid-linked oligosaccharide precursor.
2. Transfer of this precursor to the side chain N of an Asn residue on a growing polypeptide.
3. Removal of some of the precursor's sugar units.
4. Addition of sugar residues to the remaining core oligosaccharide.

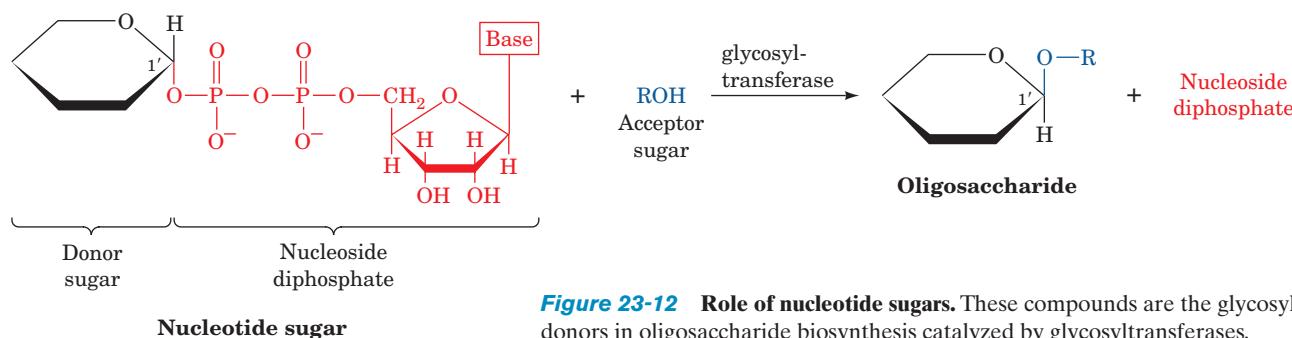


Figure 23-12 Role of nucleotide sugars. These compounds are the glycosyl donors in oligosaccharide biosynthesis catalyzed by glycosyltransferases.

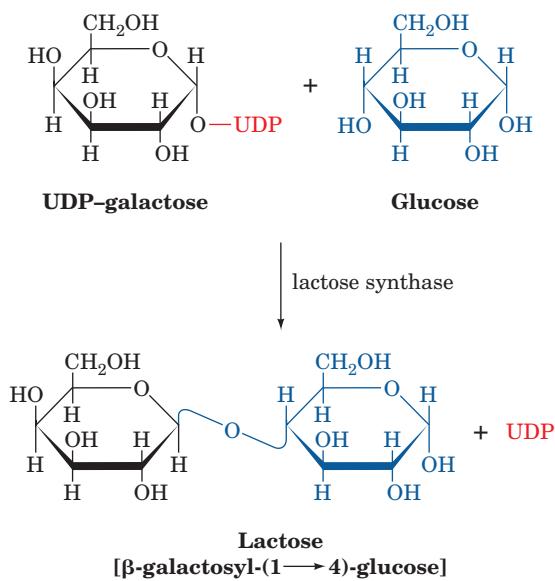


Figure 23-13 Lactose synthase. This enzyme catalyzes the formation of lactose from UDP-galactose and glucose.

We shall discuss these stages in order.

b. N-Linked Oligosaccharides Are Constructed on Dolichol Carriers

N-Linked oligosaccharides are initially synthesized as lipid-linked precursors. The lipid component in this process is **dolichol**, a long-chain polyisoprenol of 14 to 24 isoprene units (17–21 units in animals and 14–24 units in fungi and plants; isoprene units are C₅ units with the carbon skeleton of isoprene; Section 25-6A), which is linked to the oligosaccharide precursor via a pyrophosphate bridge (Fig. 23-15). Dolichol apparently anchors the growing oligosaccharide to the endoplasmic reticulum membrane. Involvement of lipid-linked oligosaccharides in *N*-linked glycoprotein synthesis was first demonstrated in 1972 by Armando Parodi and Luis Leloir, who showed that, when a lipid-linked oligosaccharide containing [¹⁴C]glucose is incubated with rat liver **microsomes** (vesicular fragments of isolated endoplasmic reticulum), the radioactivity becomes associated with protein.

c. N-Linked Glycoproteins Have a Common Oligosaccharide Core

The pathway of dolichol-PP-oligosaccharide synthesis involves stepwise addition of monosaccharide units to the

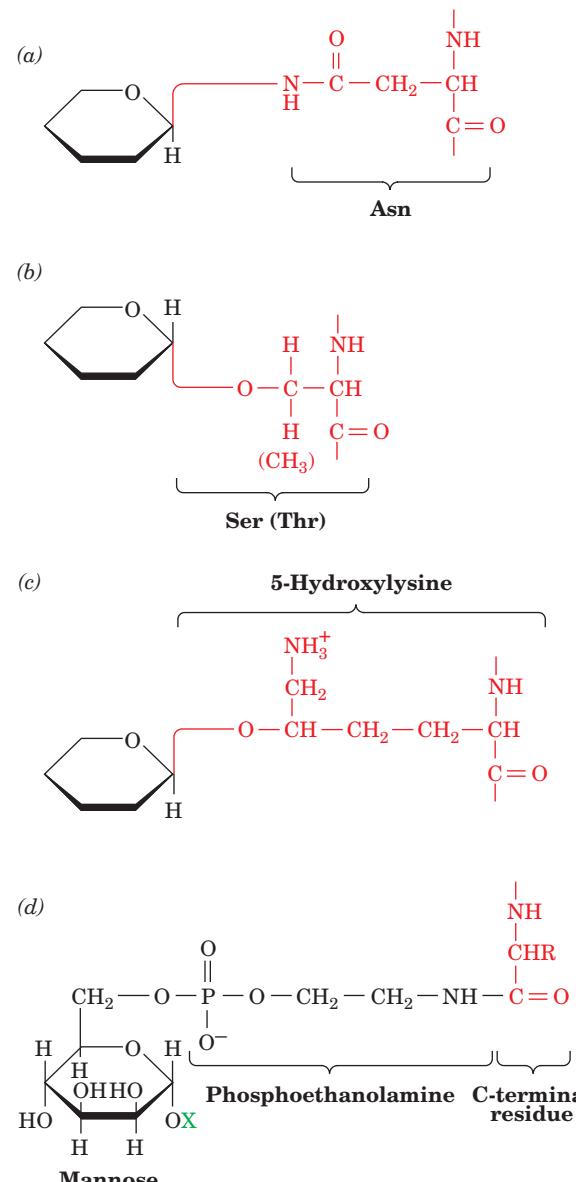


Figure 23-14 Types of saccharide-polypeptide linkages in glycoproteins. (a) An *N*-linked glycosidic bond to an Asn residue in the sequence Asn-X-Ser/Thr. (b) An *O*-linked glycosidic bond to a Ser (or Thr) residue. (c) An *O*-linked glycosidic bond to a 5-hydroxylysine residue in collagen. (d) An amide bond between the C-terminal amino acid of a protein and the phosphoethanolamine bridge to the 6 position of mannose in the glycoprophosphatidylinositol (GPI) anchor. The X group (green) denotes the rest of the GPI anchor (Fig. 12-30).

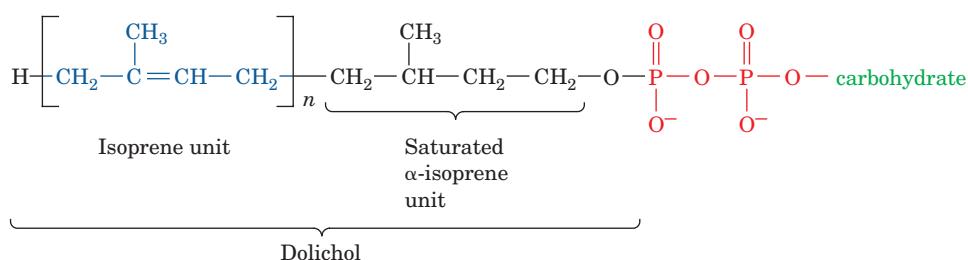


Figure 23-15 Dolichol pyrophosphate glycoside. The carbohydrate precursors of *N*-linked glycosides are synthesized as dolichol pyrophosphate glycosides. Dolichols are long-chain polyisoprenols ($n = 14$ –24) in which the α-isoprene unit is saturated.

growing glycolipid by specific glycosyltransferases to form a common “core” structure. Each monosaccharide unit is added by a unique glycosyltransferase (Fig. 23-16). For example, in Reaction 2 of Fig. 23-16, five mannose units are added through the action of five different mannosyltransferases, each with a different oligosaccharide-acceptor specificity. The oligosaccharide core, the product of Reaction 9 in Fig. 23-16, has the composition $(N\text{-acetylglucosamine})_2(\text{mannose})_5(\text{glucose})_3$.

Although nucleotide sugars are the most common monosaccharide donors in glycosyltransferase reactions, *several mannose and glucosyl residues are transferred to*

the growing dolichol-PP-oligosaccharide from their corresponding dolichol-P derivatives. This requirement for **dolichol-P-mannose** was discovered by Stuart Kornfeld, who found that mutant mouse lymphoma cells (lymphoma is a type of cancer) that are unable to synthesize normal lipid-linked oligosaccharides formed a defective, smaller glycolipid. These cells contain all the requisite glycosyltransferases but are unable to synthesize dolichol-P-mannose (Reaction 4 in Fig. 23-16 is blocked). When this substance is supplied to the mutant cells, mannose units are added to the defective dolichol-PP-oligosaccharide.

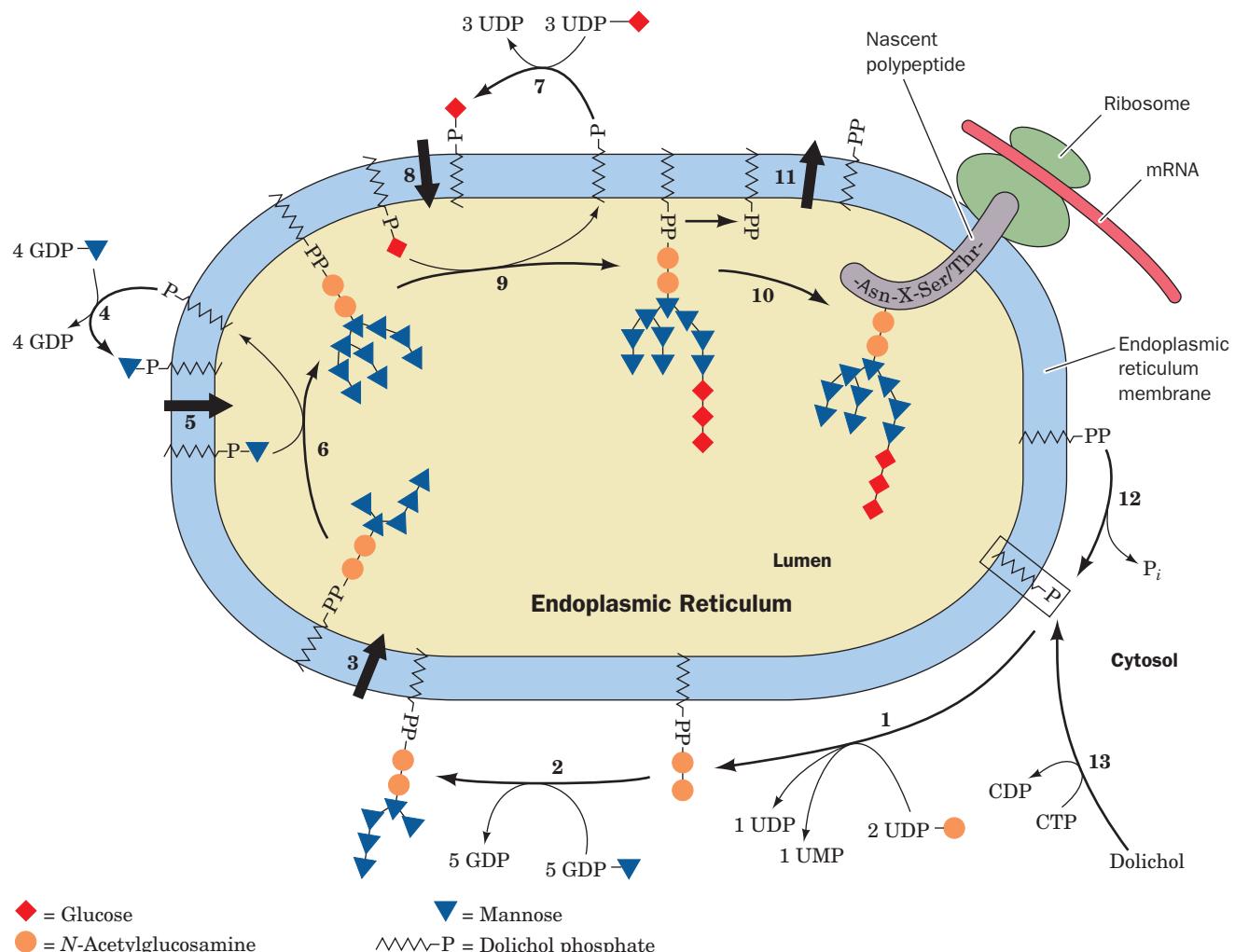


Figure 23-16 Pathway of dolichol-PP-oligosaccharide synthesis. (1) Addition of *N*-acetylglucosamine-1-P and a second *N*-acetylglucosamine to dolichol-P. (2) Addition of five mannose residues from GDP-mannose in reactions catalyzed by five different mannosyltransferases. (3) Membrane translocation of dolichol-PP- $(N\text{-acetylglucosamine})_2(\text{mannose})_5$ to the lumen of the endoplasmic reticulum (ER). (4) Cytosolic synthesis of dolichol-P-mannose from GDP-mannose and dolichol-P. (5) Membrane translocation of dolichol-P-mannose to the lumen of the ER. (6) Addition of four mannose residues from dolichol-P-mannose in reactions catalyzed by four different mannosyltransferases. (7) Cytosolic synthesis of dolichol-P-

glucose from UDP-glucose and dolichol-P. (8) Membrane translocation of dolichol-P-glucose to the lumen of the ER. (9) Addition of three glucosyl residues from dolichol-P-glucose. (10) Transfer of the oligosaccharide from dolichol-PP to the polypeptide chain at an Asn residue in the sequence Asn-X-Ser/Thr, releasing dolichol-PP. (11) Translocation of dolichol-PP to the cytosolic surface of the ER membrane. (12) Hydrolysis of dolichol-PP to dolichol-P. (13) Dolichol-P can also be formed by phosphorylation of dolichol by CTP. [Modified from Abeijon, C. and Hirschberg, C.B., *Trends Biochem. Sci.* **17**, 34 (1992).]

See the Animated Figures

d. Dolichol-PP-Oligosaccharide Synthesis Involves Topological Changes of the Intermediates

Reactions 1, 2, 4, and 7 of Fig. 23-16 all occur on the cytoplasmic side of the endoplasmic reticulum (ER) membrane. This was determined by using “right-side-out” rough ER vesicles and showing that various membrane-impermeant reagents can disrupt one or another of these reactions. Reactions 6, 9, and 10 occur in the lumen of the ER as judged by the inability of concanavalin A, a **lectin** (carbohydrate-binding protein), to bind to the products of these reactions until the membrane is permeabilized. The $(\text{mannose})_5$ $(N\text{-acetylglucosamine})_2$ -PP-dolichol product of Reaction 2, the dolichol-P-mannose product of Reaction 4, and the dolichol-P-glucose product of Reaction 7 must therefore be translocated across the ER membrane (Reactions 3, 5, and 8) such that they extend from its luminal surface in order for the synthesis of *N*-linked oligosaccharides to continue. The translocations are mediated by specific ATP-independent flippases.

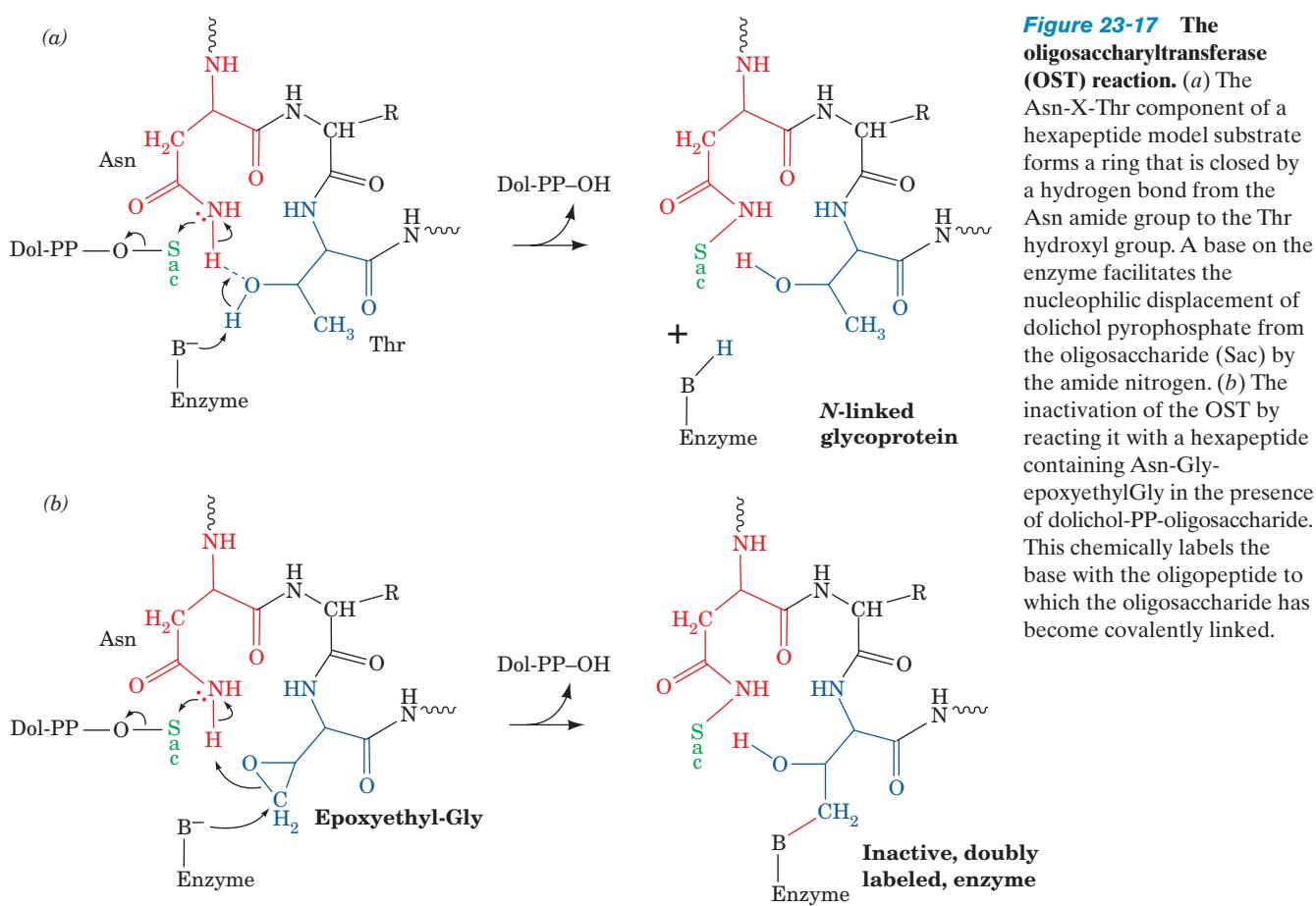
e. *N*-Linked Oligosaccharides Are Cotranslationally Added to Proteins

Vesicular stomatitis virus (VSV), which infects cattle, producing influenza-like symptoms, provides an excellent model system for studying *N*-linked glycoprotein processing. The VSV coat consists of host-cell membrane in which a single

viral glycoprotein, the **VSV G-protein** (not to be confused with the GTPases involved in signal transduction; Chapter 19), is embedded. Since a viral infection almost totally usurps an infected cell’s protein synthesizing machinery, a VSV-infected cell’s Golgi apparatus, which normally contains hundreds of different types of glycoproteins, contains virtually no other glycoprotein but G-protein. Consequently, the maturation of the G-protein is relatively easy to follow.

Of the Asn-X-Ser/Thr sites in mature eukaryotic proteins, 70 to 90% are *N*-glycosylated. Studies of VSV-infected cells indicate that the *transfer of the lipid-linked oligosaccharide to a polypeptide chain occurs while the polypeptide chain is still being synthesized*. Structural predictions (Section 9-3A), together with glycosylation studies of model polypeptides, suggest that the amino acid sequences flanking known *N*-glycosylation sites occur at β turns or loops in which Asn’s backbone N—H group is hydrogen bonded to the Ser/Thr hydroxyl O atom (Fig. 23-17a). This explains why Pro cannot occupy the X position; it would prevent Asn-X-Ser/Thr from assuming the putative required hydrogen bonded conformation.

VSV G-protein is *N*-glycosylated by **oligosaccharyltransferase (OST)**, a membrane-bound, $\sim 300\text{-kD}$, 8-subunit enzyme that recognizes the amino acid sequence Asn-X-Ser/Thr (Fig. 23-16, Reaction 10). Ernst Bause has proposed a catalytic mechanism for OST in which an enzyme base



abstracts a proton from the Ser/Thr hydroxyl group, which in turn abstracts a proton from the Asn NH₂ group, thereby promoting its nucleophilic attack on the oligosaccharide (Sac), which then displaces the dolichol pyrophosphate (Fig. 23-17a). This mechanism is supported by the observation that reacting the OST with dolichol-PP-oligosaccharide and a hexapeptide model substrate containing the sequence Asn-X-epoxyethylGly (rather than Asn-X-Ser/Thr) irreversibly inactivates the enzyme by covalently linking it to the now glycosylated hexapeptide (Fig. 23-17b).

f. The Calnexin/Calreticulin Cycle Facilitates Glycoprotein Folding

The processing of an N-linked core oligosaccharide begins in the endoplasmic reticulum by the enzymatic trimming (removal) of its three glucose residues (Fig. 23-18, Reactions 2 and 3) and one of its mannose residues (Fig. 23-18, Reaction 4) before the protein has folded to its native conformation. This is not a straightforward process, however, because **UDP-glucose:glycoprotein glucosyltransferase (GT)**, a 1513-residue soluble protein, reglucosylates the oligosaccharides of partially folded glycoproteins, a reaction that reverses the removal of the last of the three glucose residues by **glucosidase II** (Fig. 23-18, Reaction 3). This futile cycle (most glycoproteins undergo reglucosylation at least

once) is part of a chaperone-mediated glycoprotein folding process called the **calnexin/calreticulin cycle**. **Calnexin (CNX; ~570 residues)**, which is membrane bound, and **calreticulin (CRT; ~400 residues)**, its soluble homolog, are ER-resident lectins that bind partially folded glycoproteins bearing a monoglycosylated oligosaccharide in a way that protects the glycoprotein from degradation and premature transfer to the Golgi apparatus. If the glycoprotein is released and deglycosylated before it has correctly folded, GT, which recognizes only non-native glycoproteins, reglucosylates it so that the CNX/CRT cycle can repeat. CNX and CRT both also bind **ERp57**, a 481-residue thiol oxidoreductase homologous to protein disulfide isomerase (PDI; Section 9-2A). While the partially folded glycoprotein is bound to the complex, ERp57 catalyzes disulfide interchange reactions to facilitate the formation of the correctly paired disulfide bonds. The CNX/ERp57 and CRT/ERp57 complexes are therefore responsible for the correct folding and disulfide bond formation of the glycoproteins in the ER. The importance of this process is demonstrated by the observation that knockout mice lacking the gene for CRT die *in utero*.

The X-ray structure of the luminal domain of calnexin (residues 61–458), determined by Miroslaw Cygler, reveals a most unusual structure (Fig. 23-19): a compact globular domain (residues 61–262 and 415–458) from which extends

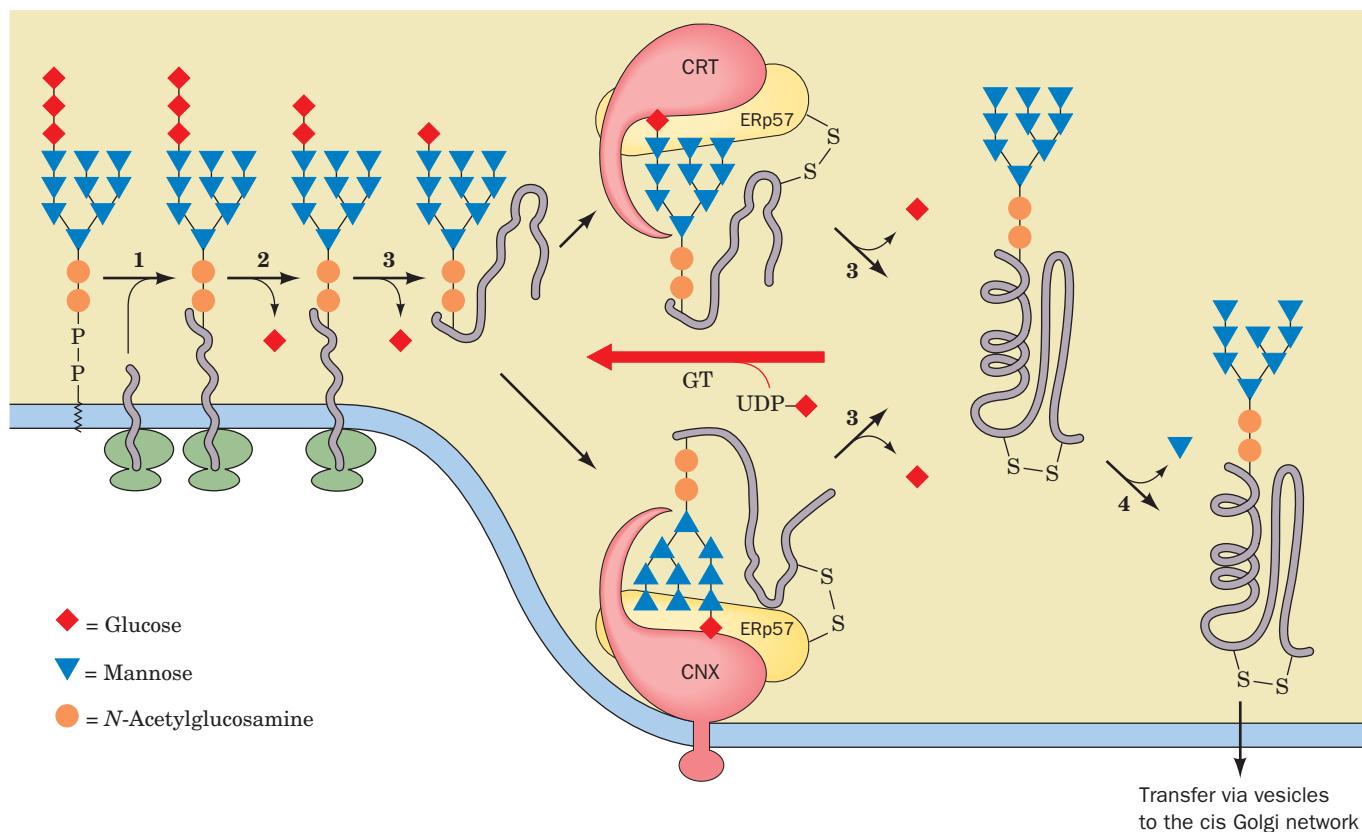


Figure 23-18 The calnexin/calreticulin cycle for glycoprotein folding in the endoplasmic reticulum. The reactions are catalyzed by: (1) oligosaccharyltransferase (OST); (2) α -glucosidase I; (3) α -glucosidase II; UDP-glucose:glycoprotein glucosyltransferase

(GT), calreticulin (CRT), calnexin (CNX), and the thiol oxidoreductase ERp57; and (4) ER α -1,2-mannosidase. [After Helenius, A. and Aebi, M., *Science* **291**, 2367 (2001).]

a 145-Å-long arm (residues 270–414). The globular domain forms a sandwich of a 6-stranded and a 7-stranded antiparallel β sheet that binds a Ca^{2+} ion and which resembles legume lectins such as concanavalin A (Fig. 8-40). This domain binds glucose on its concave (*blue*) surface, which is lined by hydrogen bonding groups that model building suggests binds the $(\text{glucose})_1(\text{mannose})_3$ portion of calnexin's natural $(\text{glucose})_1(\text{mannose})_9$ substrate. The long arm, which consists of an extended hairpin, is known as the P domain because it has four copies each of two different Pro-rich motifs arranged in the sequence 11112222, with each ~18-residue motif 1 in antiparallel association with an ~14-residue motif 2 on the opposite strand of the hairpin. Each of these motif pairs has a similar structure, with its conserved residues maintaining identical interactions in each pair. The P domain has been shown to form the binding site for ERp57 in both calnexin and calreticulin.

g. Glycoprotein Processing is Completed in the Golgi Apparatus

Once a glycoprotein has folded to its native conformation and ER α -1,2-mannosidase has removed one of its

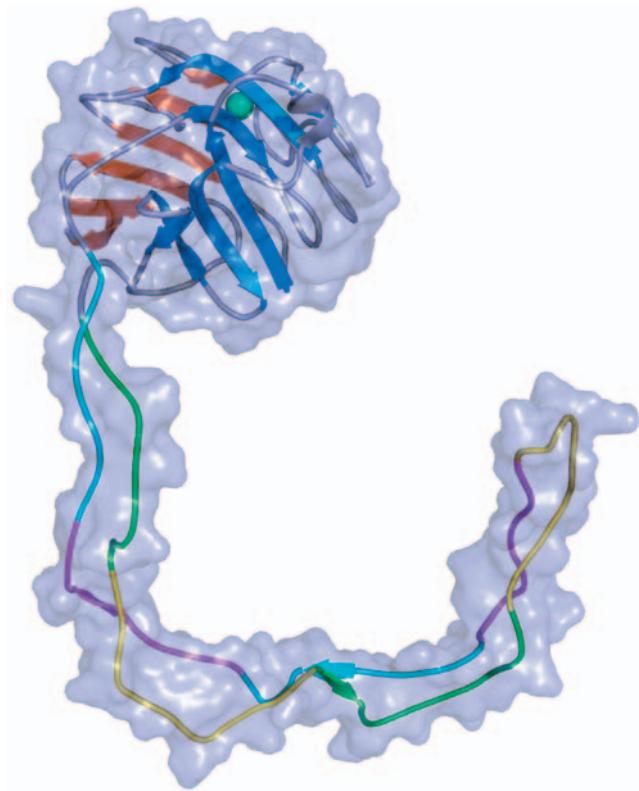


Figure 23-19 X-ray structure of the luminal portion of canine calnexin. The protein is drawn in ribbon form embedded in its semitransparent molecular surface. The 6- and 7-stranded antiparallel β sheets of its globular domain are colored orange and blue, with its remaining portions gray and its bound Ca^{2+} ion represented by a blue-green sphere. In the P domain, motifs 1 are alternately colored green and yellow and motifs 2 are alternately colored magenta and cyan. [Based on an X-ray structure by Miroslaw Cygler, Biotechnology Research Institute, NRC, Montreal, Quebec, Canada. PDBid 1JHN.]

mannosyl residues (Fig. 23-18, Step 4), the glycoprotein is transported, in membranous vesicles, to the Golgi apparatus, where it is further processed (Fig. 23-20). The Golgi apparatus (Fig. 12-58), as we discussed in Section 12-4C, consists of, from opposite the ER outward, the *cis* Golgi

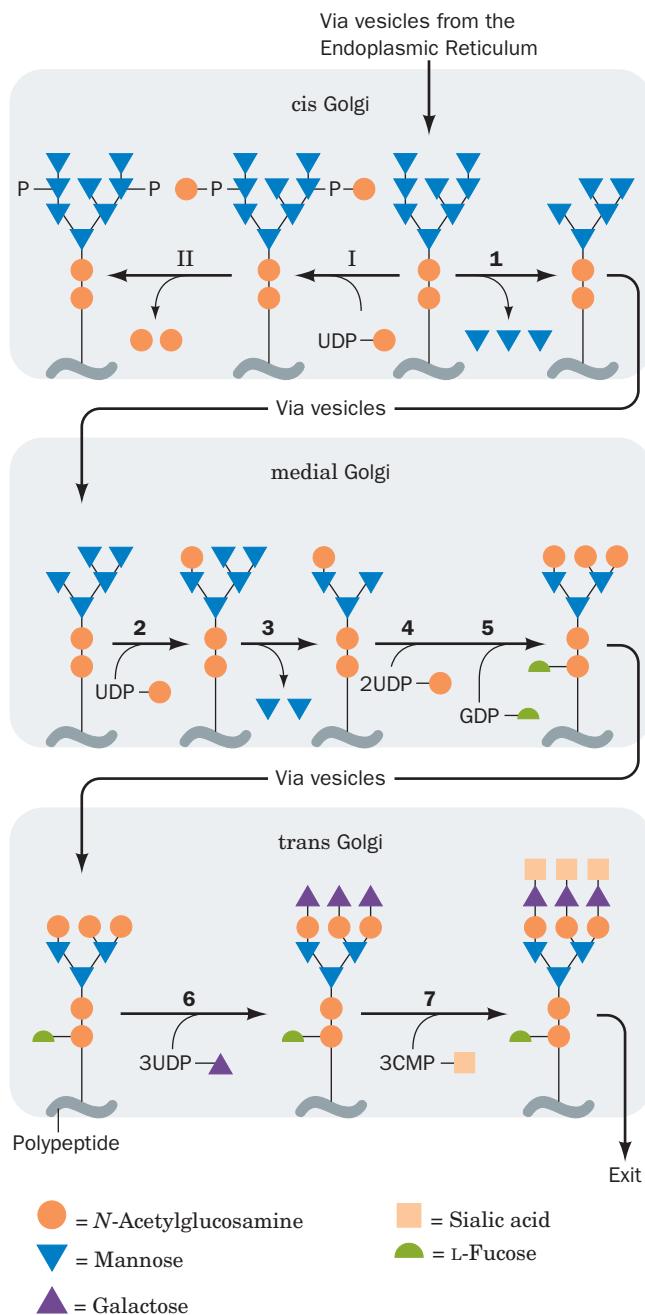


Figure 23-20 Oligosaccharide processing of VSV G-protein in the Golgi network. The reactions are catalyzed by: (1) Golgi α -mannosidase I, (2) *N*-acetylglucosaminyltransferase I, (3) Golgi α -mannosidase II, (4) *N*-acetylglucosaminyltransferase II, (5) fucosyltransferase, (6) galactosyltransferase, and (7) sialyltransferase. Lysosomal proteins are modified by: (1) *N*-acetylglucosaminyl phosphotransferase and (II) *N*-acetylglucosamine-1-phosphodiester α -*N*-acetylglucosaminidase. [Modified from Kornfeld, R. and Kornfeld, S., *Annu. Rev. Biochem.* **54**, 640 (1985).]

network, through which glycoproteins enter the Golgi apparatus; a stack of at least three different types of sacs, the cis, medial, and trans cisternae; and the trans Golgi network, through which proteins exit the Golgi apparatus. Glycoproteins traverse the Golgi stack, from the cis to the medial to the trans cisternae, each of which, as shown by James Rothman and Kornfeld, contains different sets of glycoprotein processing enzymes. As this occurs, mannose residues are trimmed from each oligosaccharide group and *N*-acetylglucosamine, galactose, fucose, and/or sialic acid residues are added to complete the processing of the glycoprotein (Fig. 23-20; Reactions 1–7). The glycoproteins are then sorted in the trans Golgi network for transport to their respective cellular destinations via membranous vesicles (Sections 12-4C and 12-4D).

There is enormous diversity among the different oligosaccharides of *N*-linked glycoproteins, as is indicated, for example, in Fig. 11-32c. Indeed, *even glycoproteins with a given polypeptide chain exhibit considerable microheterogeneity* (Section 11-3C), presumably as a consequence of incomplete glycosylation and lack of absolute specificity on the part of glycosyltransferases and glycosylases.

The processing of all *N*-linked oligosaccharides is identical through Reaction 4 of Fig. 23-18, so that all of them have a common (*N*-acetylglucosamine)₂(mannose)₃ core (five “noncore” mannose residues are subsequently trimmed from VSV G-protein; Fig. 23-20, Reactions 1 and 3). The diversity of the *N*-linked oligosaccharides therefore arises through divergence from this sequence after Fig. 23-20, Reaction 3. The resulting oligosaccharides are classified into three groups:

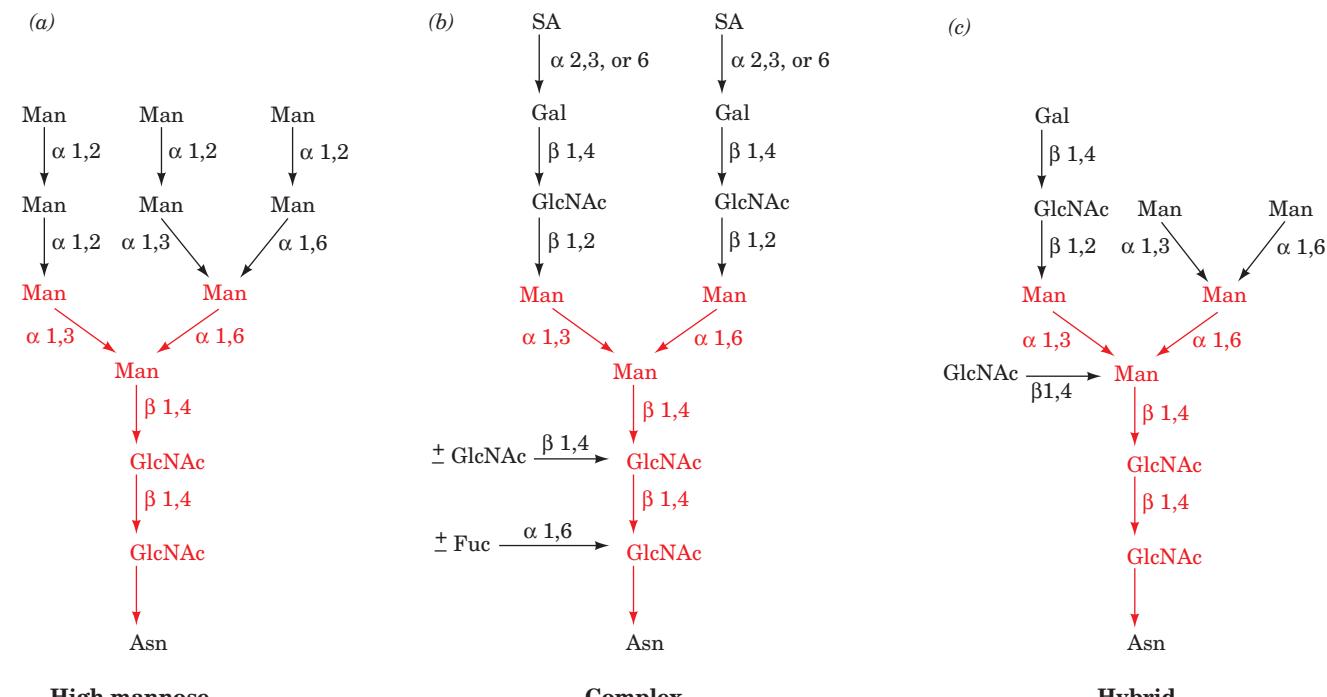


Figure 23-21 Types of *N*-linked oligosaccharides. Typical primary structures of (a) high-mannose, (b) complex, and (c) hybrid *N*-linked oligosaccharides. The pentasaccharide core

1. High-mannose oligosaccharides (Fig. 23-21a), which contain 2 to 9 mannose residues appended to the common pentasaccharide core (red residues in Fig. 23-21).

2. Complex oligosaccharides (Fig. 23-21b), which contain variable numbers of *N*-acetyllactosamine units as well as sialic acid and/or fucose residues linked to the core.

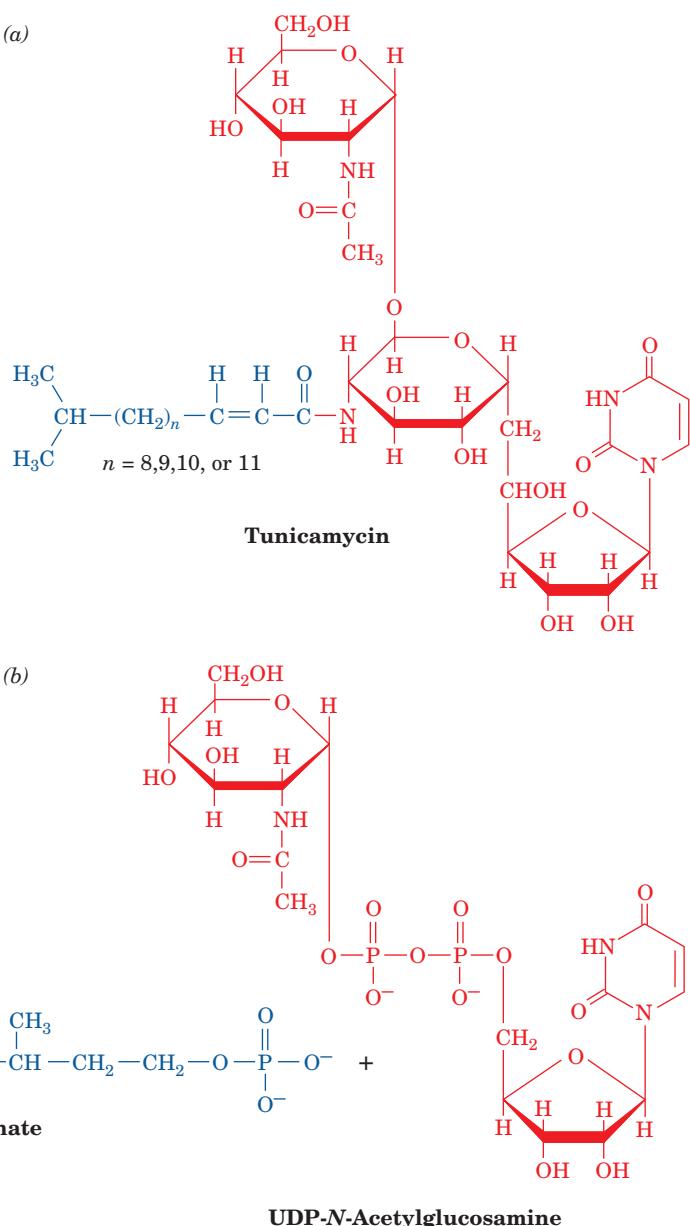
3. Hybrid oligosaccharides (Fig. 23-21c), which contain elements of both high-mannose and complex chains.

It is unclear how different types of oligosaccharides are related to the functions and/or final cellular locations of their glycoproteins. Lysosomal glycoproteins, however, appear to be of the high-mannose variety.

h. Inhibitors Have Aided the Study of *N*-Linked Glycosylation

Elucidation of the events in the glycosylation process has been greatly facilitated through the use of inhibitors that block specific glycosylation enzymes. Two of the most useful are the antibiotics **tunicamycin** (Fig. 23-22a), a hydrophobic analog of UDP-*N*-acetylglucosamine, and **bacitracin** (Fig. 23-23), a cyclic polypeptide. Both were discovered because of their ability to inhibit bacterial cell wall biosynthesis, a process that also involves the participation of lipid-linked oligosaccharides. Tunicamycin blocks the formation of dolichol-PP-oligosaccharides by inhibiting the synthesis of dolichol-PP-*N*-acetylglucosamine from dolichol-P and UDP-*N*-acetylglucosamine (Fig. 23-16, Reaction 1). Tunicamycin resembles an adduct of these reactants (Fig. 23-22b) and, in fact, binds to the enzyme with a dissociation constant of $7 \times 10^{-9} M$.

Figure 23-22 Chemical structure of tunicamycin. The structure of (a) the glycosylation inhibitor tunicamycin is compared to that of (b) dolichol-P + UDP-*N*-acetylglucosamine.



Bacitracin forms a complex with dolichol-PP that inhibits its dephosphorylation (Fig. 23-16, Reaction 12), thereby preventing glycoprotein synthesis from lipid-linked oligosaccharide precursors. Bacitracin is clinically

useful because it destroys bacterial cell walls but does not affect animal cells because it cannot cross cell membranes (bacterial cell wall biosynthesis is an extracellular process).

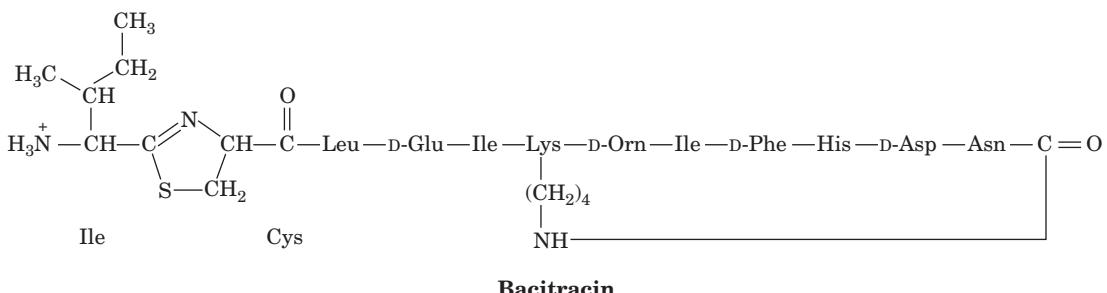


Figure 23-23 Chemical structure of bacitracin. Note that this dodecapeptide has four D-amino acid residues and two unusual

intrachain linkages. “Orn” represents the nonstandard amino acid residue ornithine (Fig. 26-7).

i. O-Linked Oligosaccharides Are Post-Translationally Formed

The study of the biosynthesis of **mucin**, an *O*-linked glycoprotein secreted by the submaxillary salivary gland, indicates that *O*-linked oligosaccharides are synthesized in the Golgi apparatus by serial addition of monosaccharide units to a completed polypeptide chain (Fig. 23-24). Synthesis starts with the transfer of *N*-acetylglucosamine (GalNAc) from UDP-GalNAc to a Ser or Thr residue on the polypeptide by **GalNAc transferase**. In contrast to *N*-linked oligosaccharides, which are transferred to an Asn in a specific amino acid sequence, the *O*-glycosylated Ser and Thr residues are not members of any common sequence. Rather, it appears that the location of glycosylation sites is specified only by the secondary or tertiary structure of the polypeptide. Glycosylation continues with stepwise addition of galactose, sialic acid, *N*-acetylglucosamine, and/or fucose by the corresponding glycosyltransferases.

j. Oligosaccharides on Glycoproteins Act as Recognition Sites

Glycoproteins that are synthesized in the endoplasmic reticulum and processed in the Golgi apparatus are targeted for secretion, insertion into cell membranes, or incorporation into cellular organelles such as lysosomes. This suggests that *oligosaccharides serve as recognition markers for this sorting process*. For example, the study of I-cell disease (Section 12-4Cg) demonstrated that in glycoprotein enzymes destined for the lysosome, a mannose residue is converted to mannose-6-phosphate (M6P) in the *cis* cisternae of the Golgi. The process involves two enzymes (Fig. 23-20, Reactions I and II), which are thought to recognize lysosomal protein precursors by certain structural features on these proteins rather than a specific amino acid sequence. In the trans Golgi network, M6P-bearing glycoproteins are sorted into lysosome-bound coated vesicles through their specific binding to one of two M6P receptors, one of which is a 275-kD membrane glycoprotein called the **M6P/IGF-II receptor** (because it has been found that this M6P receptor and the **insulinlike growth factor II receptor** are the same protein). Individuals with I-cell disease lack the enzyme catalyzing mannose phosphorylation (Fig. 23-20, Reaction I), resulting in the secretion of the normally lysosome-resident enzymes.

ABO blood group antigens (Section 12-3E) are *O*-linked glycoproteins. Their characteristic oligosaccharides are components of both cell-surface lipids and of proteins that occur in various secretions such as saliva. These oligosaccharides form antibody recognition sites.

Glycoproteins are believed to mediate cell-cell recognition. For example, an *O*-linked oligosaccharide on a glycoprotein that coats the mouse ovum surface (zona pellucida) acts as the sperm receptor. Even when this oligosaccharide is separated from its protein, it retains the ability to bind mouse sperm.

k. GPI-Linked Proteins

Glycosylphosphatidylinositol (GPI) groups function to anchor a wide variety of proteins to the exterior surface of the eukaryotic plasma membrane, thus providing an alternative to transmembrane polypeptide domains (Section

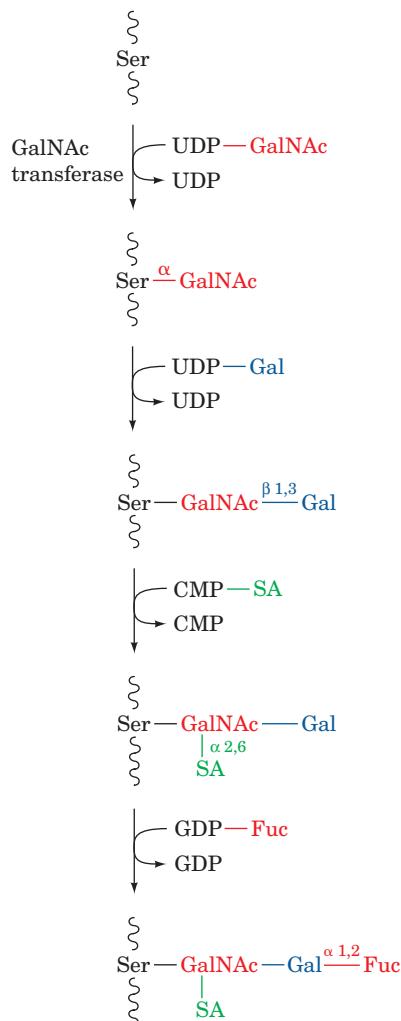
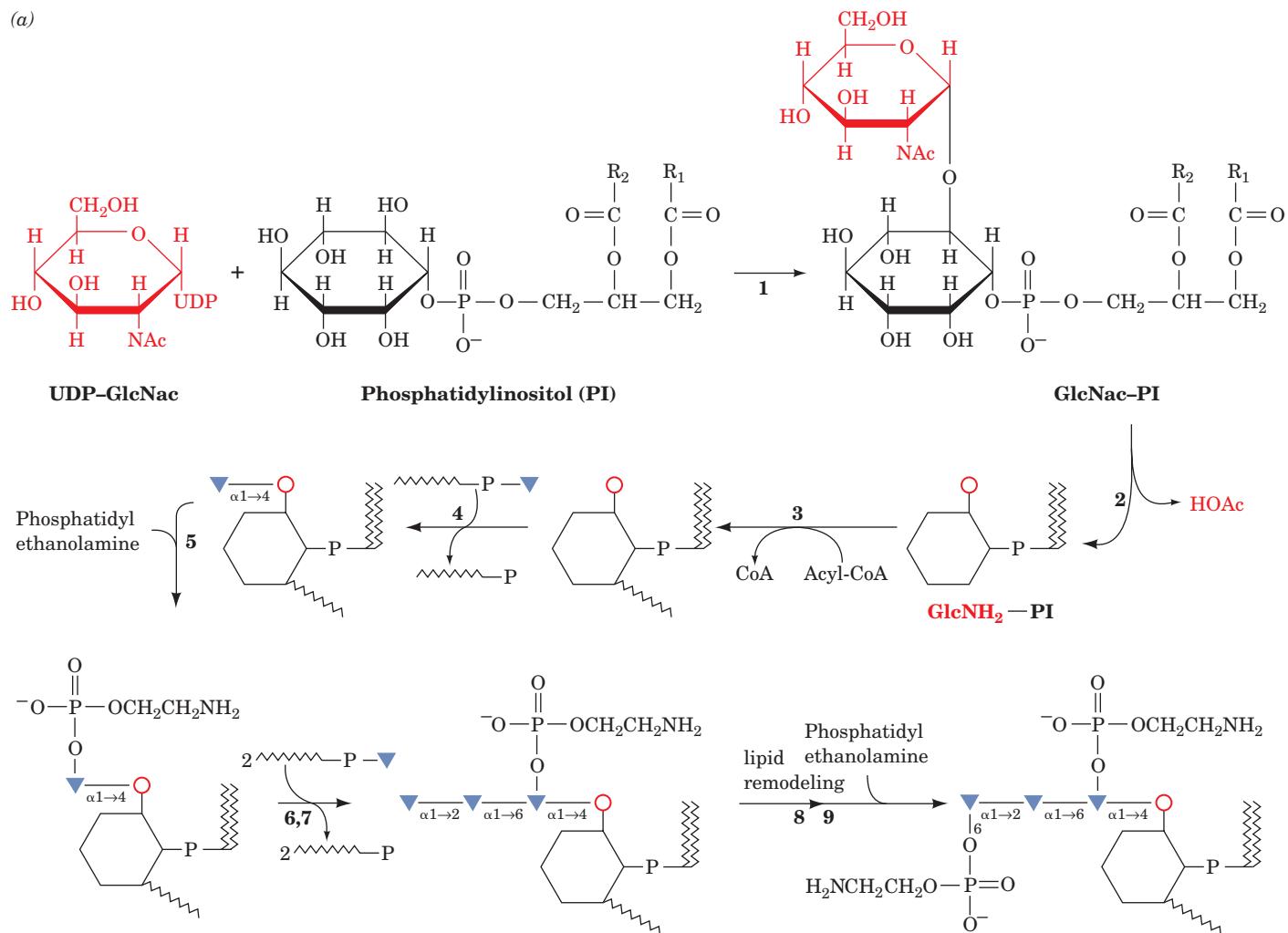


Figure 23-24 Proposed synthesis pathway for the carbohydrate moiety of an *O*-linked oligosaccharide chain of canine submaxillary mucin. SA and Fuc represent sialic acid and fucose.

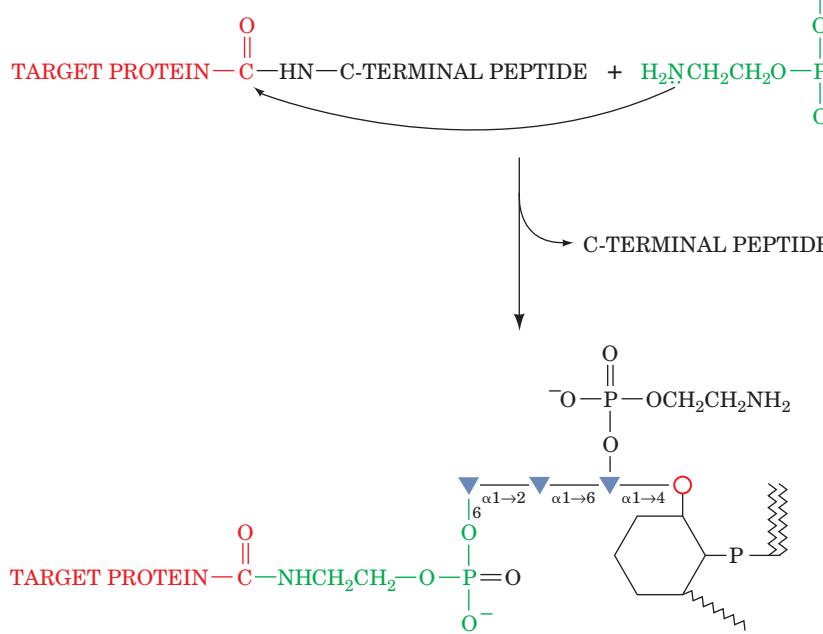
12-3Bc; Fig. 12-30). This anchoring results from transamidation of a preformed GPI glycolipid within 1 min of the synthesis and transfer of a target protein to the ER. Biosynthesis of the GPI core structure (Fig. 23-25a) begins on the cytoplasmic side of the ER with the transfer of

Figure 23-25 **GPI anchors.** (Opposite) (a) The pathway of synthesis of the tetrasaccharide core of glycosphingolipid (GPI). The following enzymes and steps are involved: (1) UDP-GlcNAc:PI $\alpha 1 \rightarrow 6$ *N*-acetylglucosaminyltransferase complex, (2) GlcNAc-PI de-*N*-acetylase, (3) inositol acyltransferase, (4) Dol-P-Man:GlcN-PI/GlcN-(acyl)PI $\alpha 1 \rightarrow 4$ mannosyltransferase (MT-I), (5) an ethanolamine phosphotransferase, (6) Dol-P-Man:Man₁GlcN-(acyl)PI $\alpha 1 \rightarrow 6$ mannosyltransferase (MT-II), (7) Dol-P-Man:Man₂GlcN-(acyl)PI $\alpha 1 \rightarrow 2$ mannosyltransferase (MT-III), (8) lipid remodeling (replacement of the fatty acyl groups on PI), and (9) transfer of phosphoethanolamine to the 6-hydroxyl group of the terminal mannose residue of the core tetrasaccharide by an ethanolamine phosphotransferase. (b) Transamidation of the target protein, resulting in a C-terminal amide link to the GPI anchor.

(a)



(b)



$\sim\sim\sim\sim\sim\text{P}$	Dolichol phosphate
\blacktriangle	mannose
○	Glcosamine
$\sim\sim\sim\sim\sim$	Phosphatidyl inositol (PI)
$\sim\sim\sim\sim$	Acyl group

N-acetylglucosamine from UDP-*N*-acetylglucosamine (UDP-GlcNAc) to the 6 hydroxyl of the inositol of phosphatidylinositol, followed by the removal of the acetyl group. The mammalian pathway then continues with the 2-acylation of inositol, translocation to the luminal side of the ER membrane, and the addition of mannose from dolichol-P-mannose (Dol-P-Man; Fig. 23-16) and phosphoethanolamine from phosphatidylethanolamine (Table 12-2), as indicated in Fig. 23-25a. This core is modified with a variety of additional sugar residues, depending on the species and the protein to which it is attached. There is considerable diversity in the fatty acid residues of GPI anchors due to the extensive lipid remodeling that occurs during anchor synthesis. Target proteins become anchored to the membrane surface when the amino group of the GPI phosphoethanolamine nucleophilically attacks a specific amino acyl group of the protein near its C-terminus, resulting in a transamidation that releases a 20- to 30-residue hydrophobic C-terminal signal peptide (Fig. 23-25b). Since GPI groups are appended to proteins on the luminal surface of the RER, GPI-anchored proteins occur on the exterior surface of the plasma membrane (Fig. 12-60). However, they are distributed unevenly in the outer leaflet of the plasma membrane because they prefer to associate with sphingolipid-cholesterol rafts (Section 12-3Cb).

The core GPI structure is evolutionarily conserved among all eukaryotes, although there are differences between species in its synthesis. For example, the cell surface of the trypanosomes that cause African sleeping sickness (a debilitating and often fatal disease that afflicts millions of people in sub-Saharan Africa) has a dense coating of **variant surface glycoprotein (VSG)** that is GPI-anchored to its plasma membrane. The VSG coating conceals the trypanosome's plasma membrane from the host's immune system although it recognizes and attacks the VSG itself. The parasite is nevertheless able to evade the host's immunological defenses because it has a genetic repertoire of about a thousand immunologically distinct VSGs. An individual trypanosome expresses only one of its VSG genes and hence the host can mount an effective immunological attack against the prevailing population of VSGs, a process that takes around 1 week (Section 35-2A). However, by switching VSG genes, a new population of trypanosomes arises that replicates unchecked until the host can mount a new immune response, a cycle that repeats until the death of the host. The comparison of the GPI biosynthetic pathway in trypanosomes with that in mammalian systems has revealed several differences in the pathway order. For example, Steps 3 and 4 of Fig. 23-25a are reversed in trypanosomes. This and other differences in the substrate specificities of the enzymes catalyzing this pathway have brought to light several promising drug targets for the treatment of African sleeping sickness.

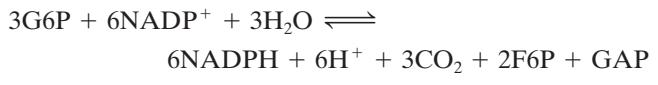
4 THE PENTOSE PHOSPHATE PATHWAY

ATP is the cell's "energy currency"; its exergonic hydrolysis is coupled to many otherwise endergonic cell functions. *Cells have a second currency, reducing power.* Many endergonic reactions, notably the reductive biosynthesis of fatty acids

(Section 25-4) and cholesterol (Section 25-6A), as well as photosynthesis (Section 24-3A), require NADPH in addition to ATP. Despite their close chemical resemblance, **NADPH and NADH are not metabolically interchangeable** (recall that these coenzymes differ only by a phosphate group at the 2'-OH group of NADPH's adenosine moiety; Fig. 13-2). Whereas NADH participates in utilizing the free energy of metabolite oxidation to synthesize ATP (oxidative phosphorylation), **NADPH is involved in utilizing the free energy of metabolite oxidation for otherwise endergonic reductive biosynthesis**. This differentiation is possible because the dehydrogenase enzymes involved in oxidative and reductive metabolism exhibit a high degree of specificity toward their respective coenzymes. Indeed, cells normally maintain their $[NAD^+]/[NADH]$ ratio near 1000, which favors metabolite oxidation, while keeping their $[NADP^+]/[NADPH]$ ratio near 0.01, which favors metabolite reduction.

NADPH is generated by the oxidation of G6P via an alternative pathway to glycolysis, the pentose phosphate pathway [also called the hexose monophosphate (HMP) shunt and the phosphogluconate pathway; Fig. 23-26]. The pathway also produces ribose-5-phosphate (R5P), an essential precursor in nucleotide biosynthesis (Sections 28-1, 28-2, and 28-5). The first evidence of this pathway's existence was obtained in the 1930s by Otto Warburg, who discovered NADP⁺ through his studies on the oxidation of G6P to 6-phosphogluconate. Further indications came from the observation that tissues continue to respire in the presence of high concentrations of fluoride ion, which, it will be recalled, blocks glycolysis by inhibiting enolase (Section 17-2I). It was not until the 1950s, however, that the pentose phosphate pathway was elucidated by Frank Dickens, Bernard Horecker, Fritz Lipmann, and Efraim Racker. Tissues most heavily involved in fatty acid and cholesterol biosynthesis (liver, mammary gland, adipose tissue, and adrenal cortex) are rich in pentose phosphate pathway enzymes. Indeed, some 30% of the glucose oxidation in liver occurs via the pentose phosphate pathway.

The overall reaction of the pentose phosphate pathway is



However, the pathway may be considered to have three stages:

1. Oxidative reactions (Fig. 23-26, Reactions 1-3), which yield NADPH and **ribulose-5-phosphate (Ru5P).**



2. Isomerization and epimerization reactions (Fig. 23-26, Reactions 4 and 5), which transform Ru5P either to **ribose-5-phosphate (R5P) or to **xylulose-5-phosphate (Xu5P)**.**



3. A series of C—C bond cleavage and formation reactions (Fig. 23-26, Reactions 6–8) that convert two molecules of Xu5P and one molecule of R5P to two molecules

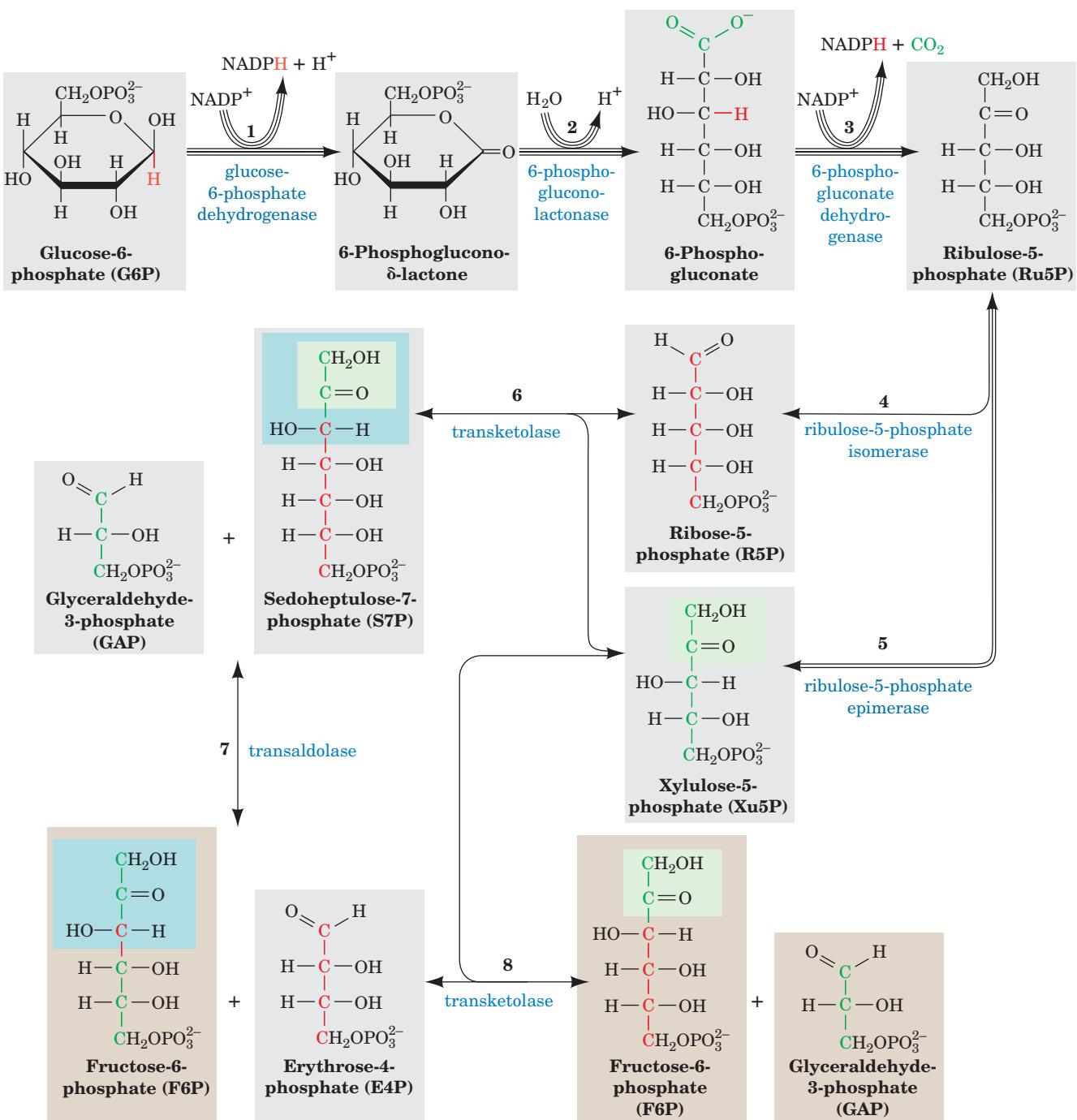


Figure 23-26 The pentose phosphate pathway. The number of lines in an arrow represents the number of molecules reacting in one turn of the pathway so as to convert three G6P to three CO_2 , two F6P, and one GAP. For the sake of clarity, sugars from Reaction 3 onward are shown in their linear forms. The carbon

skeleton of R5P and the atoms derived from it are drawn in red and those from Xu5P are drawn in green. The C_2 units transferred by transketolase are shaded in green and the C_3 units transferred by transaldolase are shaded in blue. Double-headed arrows indicate reversible reactions.

of fructose-6-phosphate (F6P) and one of glyceraldehyde-3-phosphate (GAP).



The reactions of Stages 2 and 3 are freely reversible so that the products of the pathway vary with the needs of the cell.

For example, when R5P is required for nucleotide biosynthesis, Stage 3 works in reverse, producing R5P from F6P and GAP nonoxidatively. In this section, we discuss the three stages of the pentose phosphate pathway and how this pathway is controlled. We close by considering the consequences of one of its abnormalities.

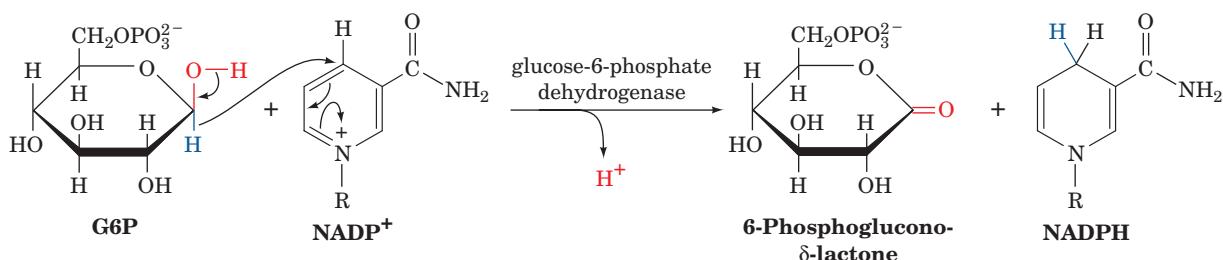


Figure 23-27 The glucose-6-phosphate dehydrogenase reaction.

A. Oxidative Reactions of NADPH Production

Only the first three reactions of the pentose phosphate pathway are involved in NADPH production.

1. Glucose-6-phosphate dehydrogenase (G6PD) catalyzes net transfer of a hydride ion to NADP^+ from C1 of G6P to form **6-phosphoglucono- δ -lactone** (Fig. 23-27). G6P, a cyclic hemiacetal with C1 in the aldehyde oxidation state, is thereby oxidized to a cyclic ester (lactone). The enzyme is specific for NADP^+ and is strongly inhibited by NADPH.

2. 6-Phosphogluconolactonase increases the rate of hydrolysis of 6-phosphoglucono- δ -lactone to **6-phosphogluconate** (the nonenzymatic reaction occurs at a significant rate), the substrate of the next oxidative enzyme in the pathway.

3. 6-Phosphogluconate dehydrogenase catalyzes the oxidative decarboxylation of 6-phosphogluconate, a β -hydroxy acid, to Ru5P and CO_2 (Fig. 23-28). The reaction is similar to that catalyzed by the citric acid cycle enzyme isocitrate dehydrogenase (Section 21-3C).

Formation of Ru5P completes the oxidative portion of the pentose phosphate pathway. *It generates two molecules of NADPH for each molecule of G6P that enters the pathway.* The product Ru5P must subsequently be converted to R5P or Xu5P for further use.

B. Isomerization and Epimerization of Ribulose-5-Phosphate

Ru5P is converted to R5P by **ribulose-5-phosphate isomerase** (Fig. 23-26, Reaction 4) and to Xu5P by **ribulose-**

5-phosphate epimerase (Fig. 23-26, Reaction 5). These isomerization and epimerization reactions, as discussed in Section 16-2Db, are both thought to occur via enediolate intermediates (Fig. 23-29).

R5P is an essential precursor in the biosynthesis of nucleotides (Sections 28-1, 28-2, and 28-5). If, however, more R5P is formed than the cell needs, the excess, along with Xu5P, is converted to the glycolytic intermediates F6P and GAP as described below.

C. Carbon–Carbon Bond Cleavage and Formation Reactions

The conversion of three C_5 sugars to two C_6 sugars and one C_3 sugar involves a remarkable “juggling act” catalyzed by two enzymes, **transaldolase** and **transketolase**. As we discussed in Section 16-2E, enzymatic reactions that make or break carbon–carbon bonds usually have mechanisms that involve generation of a stabilized carbanion and its addition to an electrophilic center such as an aldehyde. This is the dominant theme of both the transaldolase and the transketolase reactions.

a. Transketolase Catalyzes the Transfer of C_2 Units

Transketolase, which has a thiamine pyrophosphate cofactor (TPP; Section 17-3Ba), catalyzes the transfer of a C_2 unit from Xu5P to R5P, yielding GAP and **sedoheptulose-7-phosphate (S7P)** (Fig. 23-26, Reaction 6). The reaction involves the intermediate formation of a covalent adduct between Xu5P and TPP (Fig. 23-30). The X-ray structure of this homodimeric enzyme shows that the TPP binds in a deep cleft between the subunits such that residues from both

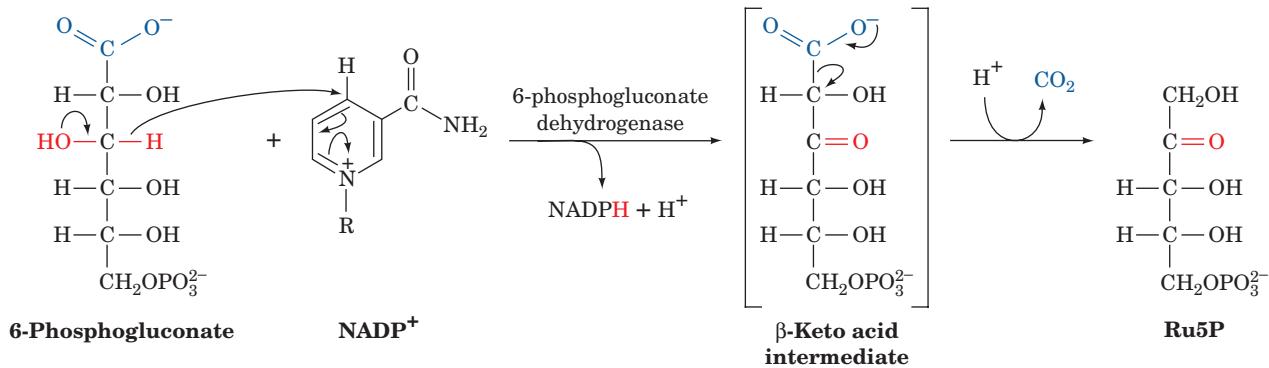


Figure 23-28 The phosphogluconate dehydrogenase reaction. Oxidation of the OH group forms an easily decarboxylated

β -keto acid (although the proposed intermediate has not been isolated).

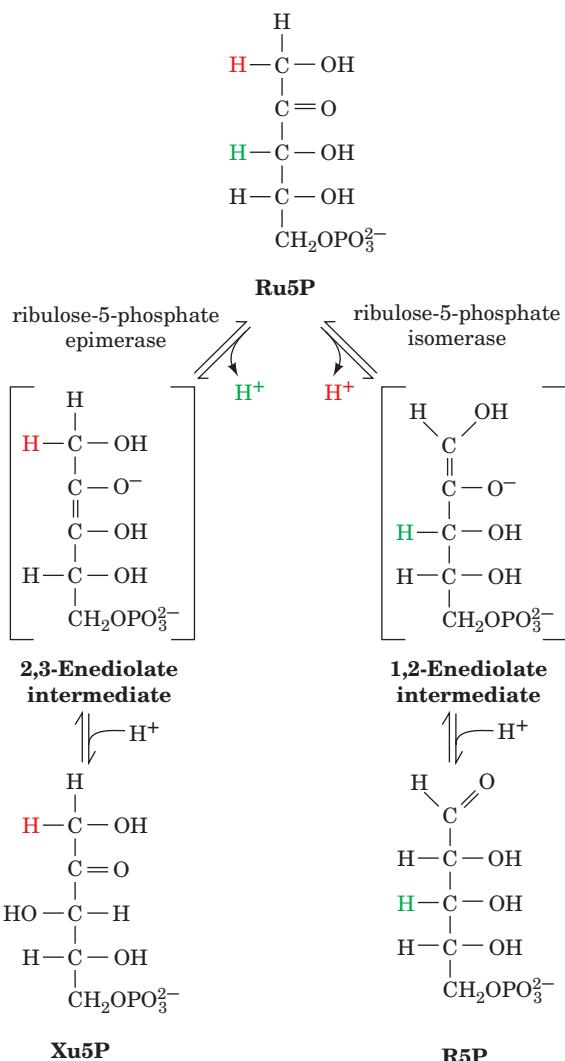


Figure 23-29 Ribulose-5-phosphate isomerase and ribulose-5-phosphate epimerase. The reactions catalyzed by both these enzymes involve enediolate intermediates. In the isomerase reaction (right), a base on the enzyme removes a proton from C1 of Ru5P to form a 1,2-enediolate and then adds a proton at C2 to form R5P. In the epimerase reaction (left), a base on the enzyme removes a C3 proton to form a 2,3-enediolate. A proton is then added to the same carbon atom but with inversion of configuration to yield Xu5P.

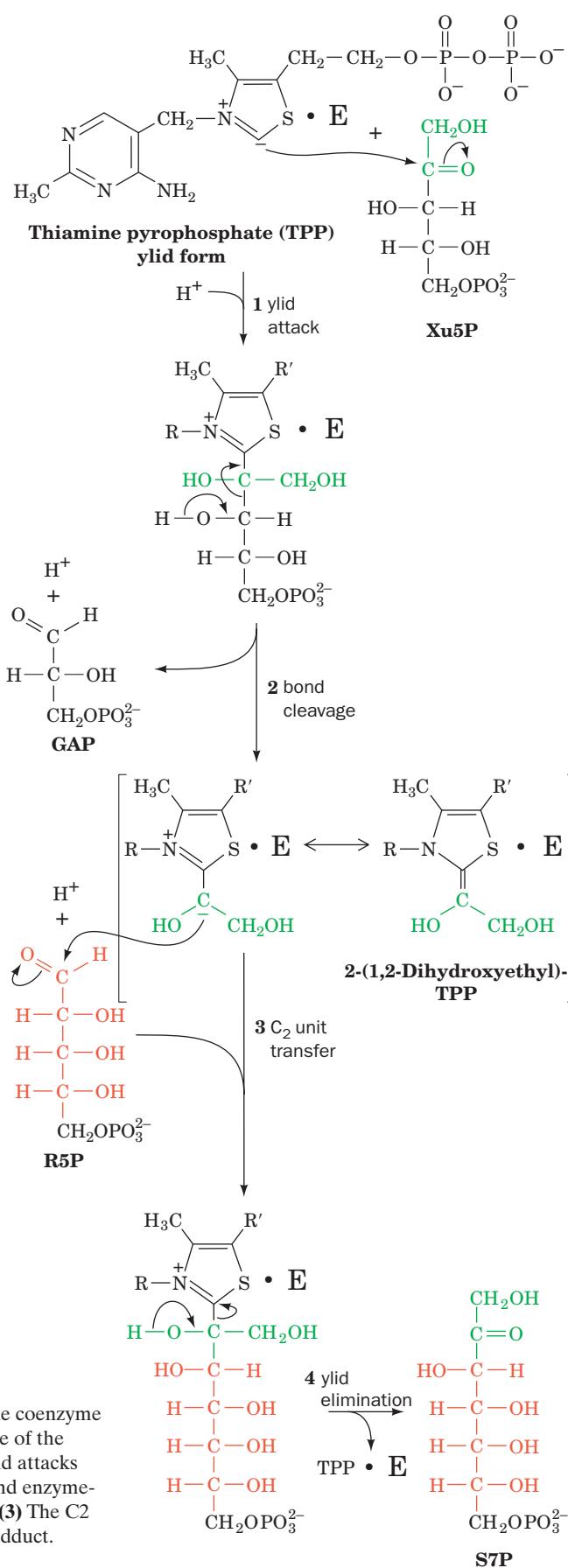


Figure 23-30 Mechanism of transketolase. Transketolase utilizes the coenzyme thiamine pyrophosphate to stabilize the carbanion formed on cleavage of the C2—C3 bond of Xu5P. The reaction occurs as follows: (1) The TPP ylid attacks the carbonyl group of Xu5P. (2) C2—C3 bond cleavage yields GAP and enzyme-bound 2-(1,2-dihydroxyethyl)-TPP, a resonance-stabilized carbanion. (3) The C2 carbanion attacks the aldehyde carbon of R5P forming an S7P-TPP adduct. (4) TPP is eliminated yielding S7P and the regenerated TPP-enzyme.

subunits participate in its binding, just as in pyruvate decarboxylase (another TPP-requiring enzyme; Figure 17-28). In fact the structures are so similar that it is likely that they diverged from a common ancestor.

b. Transaldolase Catalyzes the Transfer of C₃ Units

Transaldolase catalyzes the transfer of a C₃ unit from S7P to GAP, yielding **erythrose-4-phosphate (E4P)** and F6P (Fig. 23-26, Reaction 7). The reaction occurs by aldol cleavage, which begins with the formation of a Schiff base between an ϵ -amino group of an essential enzyme Lys residue and the carbonyl group of S7P (Fig. 23-31). Transaldolase and Class I aldolase (Section 17-2Da) share a common reaction mechanism and may also share a common ancestor, despite their lack of significant sequence identity. Both are α/β barrel proteins (Section 8-3Bh), but while the Schiff base-forming Lys is on $\beta 4$ (the fourth β strand from the N-terminus) of transaldolase, it is on $\beta 6$ of Class I aldolase. Superimposing the barrel structures of these two enzymes while maintaining the alignment of the β strands bearing the Schiff base-forming Lys residues results in a significantly better fit than doing so while maintaining the alignment of their entire α/β barrels. Moreover, five of the pairs of matched active site residues in the former superposition are identical. This suggests that, during evolution, the DNA sequence for two α/β units was transferred from the N-terminus to the C-terminus of the evolving Class I aldolase, moving the active site Lys from $\beta 6$ to $\beta 4$. Such a circular permutation of an α/β barrel's structural elements does not greatly change its structure.

c. A Second Transketolase Reaction Yields GAP and a Second F6P Molecule

In a second transketolase reaction, a C₂ unit is transferred from a second molecule of Xu5P to E4P to form GAP and another molecule of F6P (Fig. 23-26, Reaction 8). The third phase of the pentose phosphate pathway thus transforms two molecules of Xu5P and one of R5P to two molecules of F6P and one molecule of GAP. These carbon skeleton transformations (Fig. 23-26, Reactions 6–8) are summarized in Fig. 23-32.

D. Control of the Pentose Phosphate Pathway

The principal products of the pentose phosphate pathway are R5P and NADPH. The transaldolase and transketolase reactions serve to convert excess R5P to glycolytic intermediates

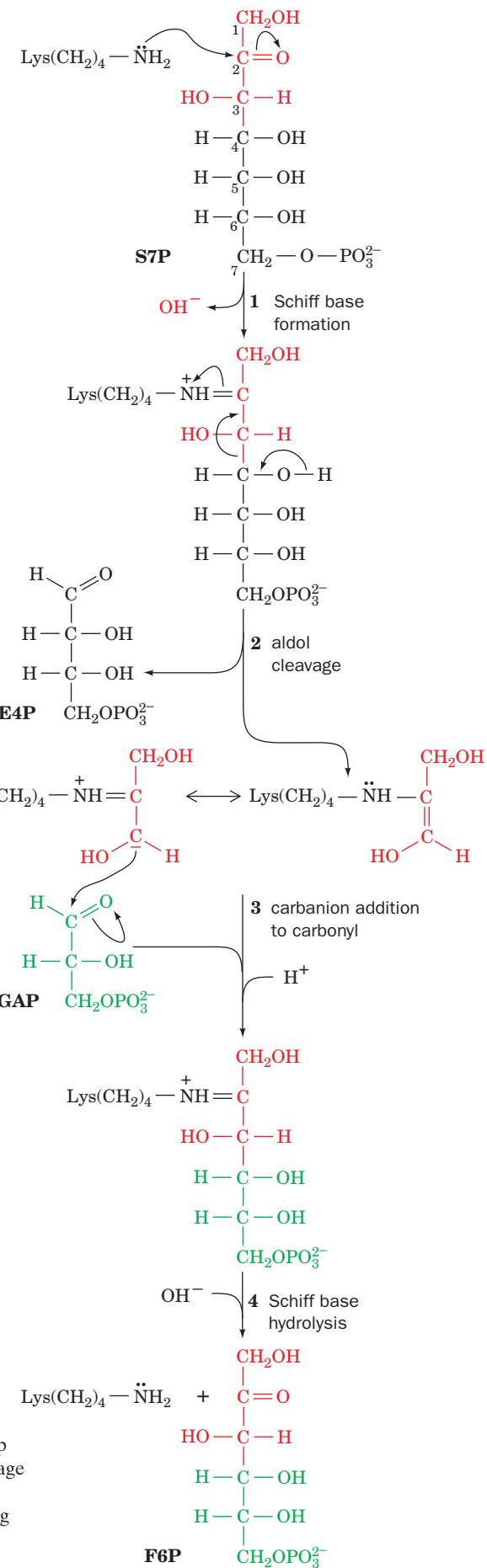


Figure 23-31 Mechanism of transaldolase. Transaldolase contains an essential Lys residue that forms a Schiff base with S7P to facilitate an aldol cleavage reaction. The reaction occurs as follows: (1) The ϵ -amino group of an essential Lys residue forms a Schiff base with the carbonyl group of S7P. (2) A Schiff base-stabilized C₃ carbanion is formed in an aldol cleavage reaction between C₃ and C₄ that eliminates E4P. (3) The enzyme-bound resonance-stabilized carbanion adds to the carbonyl C atom of GAP, forming F6P linked to the enzyme via a Schiff base. (4) The Schiff base hydrolyzes, regenerating active enzyme and releasing F6P.

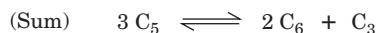
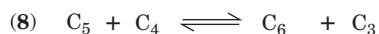
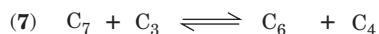


Figure 23-32 Summary of carbon skeleton rearrangements in the pentose phosphate pathway. A series of carbon–carbon bond formations and cleavages convert three C_5 sugars to two C_6 and one C_3 sugar. The number to the left of each reaction is keyed to the corresponding reaction in Fig. 23-26.

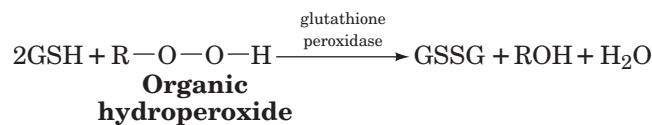
when the metabolic need for NADPH exceeds that of R5P in nucleotide biosynthesis. The resulting GAP and F6P can be consumed through glycolysis and oxidative phosphorylation or recycled by gluconeogenesis to form G6P. *In the latter case, 1 molecule of G6P can be converted, via six cycles of the pentose phosphate pathway and gluconeogenesis, to 6 CO_2 molecules with the concomitant generation of 12 NADPH molecules.* When the need for R5P outstrips that for NADPH, F6P and GAP can be diverted from the glycolytic pathway for use in the synthesis of R5P by reversal of the transaldolase and transketolase reactions. In fact, mass spectral analysis of the ^{13}C -labeled carbons from [1,2- ^{13}C]glucose incorporated into RNA in rapidly proliferating cancer cells has shown that more than ~70% of the *de novo* ribose synthesis arises through this nonoxidative reversal of the pentose phosphate pathway (rather than its forward direction).

Flux through the oxidative pentose phosphate pathway and thus the rate of NADPH production is controlled by the rate of the glucose-6-phosphate dehydrogenase reaction (Fig. 23-26, Reaction 1). The activity of this enzyme, which catalyzes the pentose phosphate pathway's first committed step ($\Delta G = -17.6 \text{ kJ} \cdot \text{mol}^{-1}$ in liver), is regulated by the $NADP^+$ concentration (substrate availability). When the cell consumes NADPH, the $NADP^+$ concentration rises, increasing the rate of the glucose-6-phosphate dehydrogenase reaction, thereby stimulating NADPH regeneration.

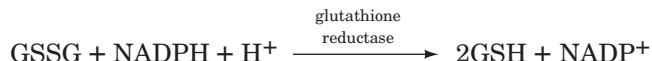
E. Glucose-6-Phosphate Dehydrogenase Deficiency

NADPH is required for several reductive processes in addition to biosynthesis. For example, erythrocyte membrane integrity requires a plentiful supply of reduced glutathione (GSH), a Cys-containing tripeptide (Sections 21-2Ba and 26-4C). A major function of GSH in the erythrocyte is to eliminate H_2O_2 and organic hydroperoxides. H_2O_2 , a toxic product of various oxidative processes (Section 22-4Cg), reacts with double bonds in the fatty acid residues of the erythrocyte cell membrane to form organic hydroperoxides. These, in turn, react to cleave fatty acid C–C bonds, thereby damaging the membrane. In erythrocytes, the unchecked buildup of peroxides results in premature cell lysis. Peroxides are eliminated through the action of **glutathione peroxidase**, one of the handful of

enzymes with a selenium cofactor, yielding glutathione disulfide (GSSG).



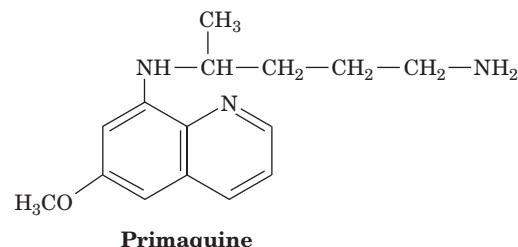
GSH is subsequently regenerated by the NADPH reduction of GSSG catalyzed by glutathione reductase (Section 21-2Ba).



A steady supply of NADPH is therefore vital for erythrocyte integrity.

a. Primaquine Causes Hemolytic Anemia in Glucose-6-Phosphate Dehydrogenase Mutants

A genetic defect, common in African, Asian, and Mediterranean populations, results in severe hemolytic anemia on infection or on the administration of certain drugs including the antimalarial agent **primaquine**.



Similar effects, which go by the name of **favism**, occur when individuals bearing this trait eat **fava beans (broad beans, *Vicia faba*)**, a staple Middle Eastern vegetable that contains small quantities of toxic glycosides (the Greek philosopher and mathematician Pythagoras, who lived in the sixth century BCE, forbade his followers from eating fava beans, possibly because of their deleterious effects). This trait has been traced to an altered gene for glucose-6-phosphate dehydrogenase (G6PD). Under most conditions, mutant erythrocytes have sufficient enzyme activity for normal function. Agents such as primaquine and fava beans, however, stimulate peroxide formation, thereby increasing the demand for NADPH to a level that mutant cells cannot meet.

The major reason for low enzymatic activity in affected cells appears to be an accelerated rate of breakdown of the mutant enzyme (protein degradation is discussed in Section 32-6). This explains why patients with G6PD deficiency react to primaquine with hemolytic anemia but recover within a week despite continued primaquine treatment. Mature erythrocytes lack a nucleus and protein synthesizing machinery and therefore cannot synthesize new enzyme molecules to replace degraded ones (they likewise cannot synthesize new membrane components, which is why they are so sensitive to membrane damage in the first place). The initial primaquine treatments result in

the lysis of old red blood cells whose defective G6PD has been largely degraded. Lysis products stimulate the release of young cells that contain more enzyme and are therefore better able to cope with primaquine stress.

It is estimated that over 400 million people are deficient in G6PD, which makes this condition the most common human enzymopathy. Indeed, ~400 G6PD variants have been reported and at least 140 of them have been characterized at the molecular level. G6PD is active in a dimer–tetramer equilibrium. Many of the mutation sites in individuals with the most severe G6PD deficiency are at the dimer interface, shifting the equilibrium toward the inactive and unstable monomer.

Several G6PD variants occur with high incidence. For example, the so-called type A[−] deficiency, which exhibits ~10% of the normal G6PD activity, has an incidence of 11% among African Americans. This variant is also the most common form of G6PD deficiency in sub-Saharan Africa. The variant “Mediterranean” is found throughout the Mediterranean and Middle East regions, and occurs in

65% of Kurdish Jews, the population with the highest known incidence of this trait. The high prevalence of defective G6PD in malarial areas of the world suggests that such mutations confer resistance to the malarial parasite, *Plasmodium falciparum* (as we likewise saw to be the case for the sickle-cell trait; Section 7-3Ab). Indeed, two epidemiological studies involving over 2000 African children with A[−] G6PD deficiency indicate that this form is associated with an ~50% reduction in the risk of severe malaria for both female heterozygotes and male hemizygotes (G6PD deficiency is an X-linked trait).

In vitro studies indicate that erythrocytes with G6PD deficiency are less suitable hosts for plasmodia than are normal cells. This is presumably because the parasite requires the products of the pentose phosphate pathway and/or because the erythrocyte is lysed before the parasite has had a chance to mature. Thus, like the sickle-cell trait, *a defective G6PD confers a selective advantage on individuals living where malaria is endemic*.

CHAPTER SUMMARY

1 Gluconeogenesis Lactate, pyruvate, citric acid cycle intermediates, and many amino acids may be converted, by gluconeogenesis, to glucose via the formation of oxaloacetate. For this to occur, the three irreversible steps of glycolysis must be bypassed. The pyruvate kinase reaction is bypassed by converting pyruvate to oxaloacetate in an ATP-driven reaction catalyzed by the biotinyl-containing enzyme pyruvate carboxylase. The two phases of the pyruvate decarboxylase reaction are catalyzed on different active sites of the homotetrameric enzyme, which translocates its covalently linked carboxybiotinyl group from its BC domain to the CT domain of a neighboring subunit. The oxaloacetate is subsequently decarboxylated and phosphorylated by GTP to form PEP in a reaction catalyzed by PEPCK. For this to happen in species in which PEPCK is a cytosolic enzyme, the oxaloacetate must be transported from the mitochondrion to the cytosol via its interim conversion to either malate or aspartate. Conversion to malate concomitantly transports reducing equivalents to the cytosol in the form of NADH. The two other irreversible steps of glycolysis, the PFK reaction and the hexokinase reaction, are bypassed by simply hydrolyzing their products, FBP and G6P, by FBPase and glucose-6-phosphatase, respectively. A glucose molecule may therefore be synthesized from pyruvate at the expense of four ATPs more than are generated by the reverse process.

Glycolysis and gluconeogenesis are reciprocally regulated so as to consume glucose when the demand for ATP is high and synthesize it when the demand is low. The control points in these processes are at pyruvate kinase/pyruvate carboxylase–PEPCK, PFK/FBPase, and hexokinase/glucose-6-phosphatase. Regulation of these enzymes is exerted largely through allosteric interactions, cAMP-dependent enzyme modifications, and, for PEPCK, gene expression. Muscle, which is incapable of gluconeogenesis, transfers much of the lactate it produces to the liver via the blood for conversion to glucose and return to the muscle. This Cori cycle shifts the metabolic burden of oxidative ATP generation for gluconeogenesis from muscle to liver.

2 The Glyoxylate Cycle Animals cannot convert fatty acids to glucose because they lack the enzymes necessary to synthesize oxaloacetate from acetyl-CoA. Plants, however, can do so via the glyoxylate cycle, a glyoxysomal process that converts two molecules of acetyl-CoA to one molecule of succinate via the intermediate formation of glyoxylate. Succinate is converted to oxaloacetate for use in gluconeogenesis or the citric acid cycle.

3 Biosynthesis of Oligosaccharides and Glycoproteins Glycosidic bonds are formed by transfer of the monosaccharide unit of a sugar nucleotide to a second sugar unit. Such reactions occur in the synthesis of disaccharides such as lactose and in the synthesis of the carbohydrate components of glycoproteins. In N-linked glycoproteins, the carbohydrate component is attached to the protein via an *N*-glycosidic bond to an Asn residue in the sequence Asn-X-Ser/Thr. In *O*-linked glycoproteins, the carbohydrate attachment is an *O*-glycosidic bond to Ser or Thr or, in collagens, to 5-hydroxylysine. In GPI-anchored proteins a glycosylphosphatidylinositol group is linked to the protein through an intermediary phosphoethanolamine bridge, which forms an amide bond to the protein’s C-terminal amino acid residue.

Synthesis of *N*-linked oligosaccharides begins in the endoplasmic reticulum with the multistep formation of a lipid-linked precursor consisting of dolichol pyrophosphate bonded to a common 14-residue core oligosaccharide. The carbohydrate is then transferred to an Asn residue of a growing polypeptide chain. The correct folding of the immature *N*-linked glycoprotein is assisted via the calnexin/calreticulin cycle and it is subsequently transferred, via a membranous vesicle, to the *cis* Golgi network of the Golgi apparatus. Processing is completed by the trimming of mannose residues followed by attachment of a variety of other monosaccharides as catalyzed by specific enzymes in the *cis*, medial, and *trans* Golgi cisternae. Completed *N*-linked glycoproteins are sorted in the *trans* Golgi network according to the identities of their carbohydrate components for transport, via membranous vesicles, to their final cellular destinations. Three

major types of *N*-linked oligosaccharides have been identified, high mannose, complex, and hybrid oligosaccharides, all of which contain a common pentasaccharide core. Studies of glycoprotein formation have been facilitated by the use of antibiotics, such as tunicamycin and bacitracin, which inhibit specific enzymes involved in the synthesis of these oligosaccharides.

O-Linked oligosaccharides are synthesized in the Golgi apparatus by sequential attachments of specific monosaccharide units to certain Ser or Thr residues. Carbohydrate components of glycoproteins are thought to act as recognition markers for the transport of glycoproteins to their proper cellular destinations and for cell-cell and antibody recognition. The GPI membrane anchor is appended to proteins on the luminal surface of the endoplasmic reticulum, thereby targeting GPI-anchored proteins to the external surface of the plasma membrane.

4 The Pentose Phosphate Pathway The cell uses NAD⁺ in oxidative reactions and employs NADPH in reductive biosynthesis. NADPH is synthesized by the pentose phosphate pathway, an alternate mode of glucose oxidation. This pathway also synthesizes R5P for use in nucleotide biosynthesis. The first three reactions of the pentose phosphate pathway

involve oxidation of G6P to Ru5P with release of CO₂ and formation of two NADPH molecules. This is followed by reactions that either isomerize Ru5P to R5P or epimerize it to Xu5P. Each molecule of R5P not required for nucleotide biosynthesis, together with two Xu5P, is converted to two molecules of F6P and one molecule of GAP via the sequential actions of transketolase, transaldolase, and, again, transketolase. The products of the pentose phosphate pathway depend on the needs of the cell. The F6P and GAP may be metabolized through glycolysis and the citric acid cycle or recycled via gluconeogenesis. If NADPH is in excess, the latter portion of the pentose phosphate pathway may be reversed to synthesize R5P from glycolytic intermediates. The pentose phosphate pathway is controlled at its first committed step, the glucose-6-phosphate dehydrogenase reaction, by the NADP⁺ concentration. A genetic deficiency in glucose-6-phosphate dehydrogenase leads to hemolytic anemia on administration of the antimalarial drug primaquine. This X-linked deficiency, which results from the accelerated degradation of the mutant enzyme, provides resistance against severe malaria to female heterozygotes and male hemizygotes for this sex-linked trait.

REFERENCES

Gluconeogenesis

- Croniger, C.M., Olswang, Y., Reshef, L., Kalhan, S.C., Tilghman, S.M., and Hanson, R.W., Phosphoenolpyruvate carboxykinase revisited. Insights into its metabolic role, *Biochem. Mol. Biol. Educ.* **30**, 14–20 (2002); and Croniger, C.M., Chakravarty, K., Olswang, Y., Cassuto, H., Reshef, L., and Hanson, R.W., Phosphoenolpyruvate carboxykinase revisited. II. Control of PEPCK-C gene expression, *Biochem. Mol. Biol. Educ.* **30**, 353–362 (2002). Knowles, J.R., The mechanism of biotin-dependent enzymes, *Annu. Rev. Biochem.* **58**, 195–221 (1989). Matte, A., Tari, L.W., Goldie, H., and Delbaere, T.J., Structure and mechanism of phosphoenolpyruvate carboxykinase, *J. Biol. Chem.* **272**, 8105–8108 (1997). Pilkis, S.J., Mahgrabi, M.R., and Claus, T.H., Hormonal regulation of hepatic gluconeogenesis and glycolysis, *Annu. Rev. Biochem.* **57**, 755–783 (1988). Rothman, D.L., Magnusson, I., Katz, L.D., Shulman, R.G., and Shulman, G.I., Quantitation of hepatic gluconeogenesis in fasting humans with ¹³C NMR, *Science* **254**, 573–576 (1991). St. Maurice, M., Reinhardt, L., Surinya, K.H., Attwood, P.V., Wallace, J.C., Cleland, W.W., and Rayment, I., Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme, *Science* **317**, 1076–1079 (2007). Van Schaftingen, E., and Gerin, I., The glucose-6-phosphatase system, *Biochem. J.* **362**, 513–532 (2002). Yang, J., Kalhan, S.C., and Hanson, R.W., What is the metabolic role of phosphoenolpyruvate carboxykinase? and Yang, J., Rashef, L., Cassuto, H., Aleman, G., and Hanson, R.W., Aspects of control of phosphoenolpyruvate carboxykinase gene transcription, *J. Biol. Chem.* **284**, 27025–27029 and 27031–27035 (2009).

The Glyoxylate Cycle

- Eastmond, P.J. and Graham, I.A., Re-examining the role of the glyoxylate cycle in oilseeds, *Trends Plant Sci.* **6**, 72–77 (2001).

Oligosaccharide Biosynthesis

- Abeijon, C. and Hirschberg, C.B., Topography of glycosylation reactions in the endoplasmic reticulum, *Trends Biochem. Sci.* **17**, 32–36 (1992).

- Aebi, C., Bernasconi, R., Clerc, S., and Molinari, M., N-glycan structures: recognition and processing in the ER, *Trends Biochem. Sci.* **35**, 74–82 (2010). Bause, E., Wesemann, M., Bartoschek, A., and Breuer, W., Epoxyethylglycyl peptides as inhibitors of oligosaccharyltransferase: double-labeling of the active site, *Biochem. J.* **322**, 95–102 (1997). Burda, P. and Aebi, M., The dolichol pathway of N-linked glycosylation, *Biochim. Biophys. Acta* **1426**, 239–257 (1999). Elbein, A.D., Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains, *Annu. Rev. Biochem.* **56**, 497–534 (1987). Englund, P.T., The structure and biosynthesis of glycosyl phosphatidylinositol protein anchors, *Annu. Rev. Biochem.* **62**, 65–100 (1993). Ferguson, M.A.J., Brimacombe, J.S., Brown, J.R., Crossman, A., Dix, A., Field, R.A., Guther, M.L.S., Milne, K.G., Sharma, D.K., and Smith, T.K., The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness, *Biochim. Biophys. Acta* **1455**, 327–340 (1999). Florman, H.M. and Wasserman, P.M., O-Linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity, *Cell* **41**, 313–324 (1985). Helenius, A. and Aebi, M., Intracellular functions of N-linked glycans, *Science* **291**, 2364–2369 (2001). Helenius, A., Trombetta, E.S., Hebert, J.N., and Simons, J.F., Calnexin, calreticulin and the folding of glycoproteins, *Trends Cell Biol.* **7**, 193–200 (1997). Hirschberg, C.B. and Snider, M.D., Topography of glycosylation in the rough endoplasmic reticulum and the Golgi apparatus, *Annu. Rev. Biochem.* **56**, 63–87 (1987). Kornfeld, R. and Kornfeld, S., Assembly of asparagine-linked oligosaccharides, *Annu. Rev. Biochem.* **54**, 631–664 (1985). Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G., Glycosyltransferases: Structures, functions, and mechanisms, *Annu. Rev. Biochem.* **77**, 521–555 (2008). Maeda, Y., Watanabe, R., Harris, C.L., Hong, Y., Ohishi, K., Kinoshita, K., and Kinoshita, T., PIG-M transfers the first

- mannose to glycosylphosphatidylinositol on the luminal side of the ER, *EMBO J.* **20**, 250–261 (2001).
- Parodi, A.J., Role of *N*-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation, *Biochem. J.* **348**, 1–13 (2000); and Protein glycosylation and its role in protein folding, *Annu. Rev. Biochem.* **69**, 69–93 (2000).
- Sanyal, S., Frank, C.G., and Menon, A.K., Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum, *Biochemistry* **47**, 7937–7946 (2008).
- Schachter, H., Enzymes associated with glycosylation, *Curr. Opin. Struct. Biol.* **1**, 755–765 (1991).
- Schrag, J.D., Bereron, J.M., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M., The structure of calnexin, an ER chaperone involved in quality control of protein folding, *Mol. Cell* **8**, 633–644 (2001).
- Shaper, J.H. and Shaper, N.L., Enzymes associated with glycosylation, *Curr. Opin. Struct. Biol.* **2**, 701–709 (1992).
- Tartakoff, A.M. and Singh, N., How to make a glycoinositol phospholipid anchor, *Trends Biochem. Sci.* **17**, 470–473 (1992).
- Taylor, M.E. and Drickamer, K., *Introduction to Glycobiology* (2nd ed.), Oxford University Press (2006).
- von Figura, K. and Hasilik, A., Lysosomal enzymes and their receptors, *Annu. Rev. Biochem.* **55**, 167–193 (1986).

The Pentose Phosphate Pathway

- Adams, M.J., Ellis, G.H., Gover, S., Naylor, C.E., and Phillips, C., Crystallographic study of coenzyme, coenzyme analogue and substrate binding in 6-phosphogluconate dehydrogenase: Implications for NADP specificity and enzyme mechanism, *Structure* **2**, 651–668 (1994).

- Au, S.W.N., Gover, S., Lam, V.M.S., and Adams, M.J., Human glucose-6-phosphate dehydrogenase: The crystal structure reveals a structural NADP⁺ molecule and provides insights into enzyme deficiency, *Structure* **8**, 293–303 (2000).
- Beutler, E., The molecular biology of G6PD variants and other red cell enzyme defects, *Annu. Rev. Med.* **43**, 47–59 (1992).
- Cappellini, M.D. and Fiorelli, G., Glucose-6-phosphate dehydrogenase deficiency, *Lancet* **375**, 64–74 (2008).
- Jia, J., Huang, W., Schörken, U., Sahm, H., Sprenger, G.A., Lindqvist, Y., and Schneider, G., Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family, *Structure* **4**, 715–724 (1996).
- Lindqvist, Y. and Schneider, G., Thiamin diphosphate dependent enzymes: transketolase, pyruvate oxidase and pyruvate decarboxylase, *Curr. Opin. Struct. Biol.* **3**, 896–901 (1993); and Muller, Y.A., Lindqvist, Y., Furey, W., Schulz, G.E., Jordan, F., and Schneider, G., A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase, *Structure* **1**, 95–103 (1993).
- Luzzato, L., Mehta, A., and Vulliamy, T., Glucose-6-phosphate dehydrogenase deficiency, Chap. 138 in Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease* <http://www.ommbid.com/>
- Ruwende, C., et al., Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria, *Nature* **376**, 246–249 (1995).
- Wood, T., *The Pentose Phosphate Pathway*, Academic Press (1985).

PROBLEMS

1. Compare the relative energetic efficiencies, in ATPs per mole of glucose oxidized, of glucose oxidation via glycolysis + the citric acid cycle versus glucose oxidation via the pentose phosphate pathway + gluconeogenesis. Assume that NADH and NADPH are each energetically equivalent to 2.5 ATP.

2. Although animals cannot synthesize glucose from acetyl-CoA, if a rat is fed ¹⁴C-labeled acetate, some of the label will appear in the glycogen extracted from its muscles. Explain.

3. Substances that inhibit specific trimming steps in the processing of *N*-linked glycoproteins have been useful tools in elucidating the pathway of this process. Explain.

4. Through clever genetic engineering you have developed an unregulatable enzyme that can interchangeably use NAD⁺ or NADP⁺ in a redox reaction. What would be the physiological consequence(s) on an organism of having such an enzyme?

5. What is the free energy change of the reaction



under physiological conditions? Assume that $\Delta G^\circ = 0$ for this reaction and that $T = 37^\circ\text{C}$.

6. If G6P is ¹⁴C-labeled at its C2 position, what is the distribution of the radioactive label in the products of the pentose phosphate pathway after one turnover of the pathway? What is the distribution of the label after passage of these products through gluconeogenesis followed by a second round of the pentose phosphate pathway?

7. After feeding rapidly growing and proliferating cells [1,2-¹³C]glucose and isolating the RNA, you find that both the C1 and C2 atoms of the ribosyl units are labeled. Show, using chemical structures and the appropriate enzymes, how the pentose phosphate pathway can yield this distribution of the label.

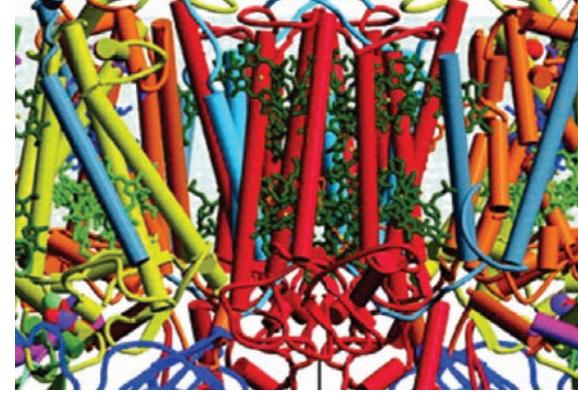
8. The relative metabolic activities in an organism of glycolysis + the citric acid cycle versus the pentose phosphate pathway + gluconeogenesis can be measured by comparing the rates of ¹⁴CO₂ generation on administration of glucose labeled with ¹⁴C at C1 with that of glucose labeled at C6. Explain.

9. (a) Describe the lengths of the products of the transketolase reaction when the two substrates are both five-carbon sugars. (b) Describe the products of the reaction when the substrates are a five-carbon aldose and a six-carbon ketose. Does it matter which of the substrates binds to the enzyme first?

10. In light of the finding that an otherwise benign or even advantageous mutation leads to abnormal primaquine sensitivity combined with the fact that human beings have enormous genetic complexity, comment on the possibility of developing drugs that exhibit no atypical side effects in any individual.

11. Glucose-6-phosphatase is located inside the endoplasmic reticulum. Describe the probable symptoms of a defect in G6P transport across the endoplasmic reticulum membrane.

Photosynthesis



CHAPTER 24

1 Chloroplasts

2 Light Reactions

- A. Absorption of Light
- B. Electron Transport in Purple Photosynthetic Bacteria
- C. Two-Center Electron Transport
- D. Photophosphorylation

3 Dark Reactions

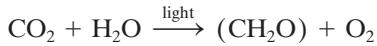
- A. The Calvin Cycle
- B. Control of the Calvin Cycle
- C. Photorespiration and the C₄ Cycle

The first indication that plants produce oxygen was found by the English clergyman and pioneering chemist Joseph Priestley, who reported:

Finding that candles burn very well in air in which plants had grown a long time, and having some reason to think, that there was something attending vegetation, which restored air that had been injured by respiration, I thought it was possible that the same process might also restore the air that had been injured by the burning of candles. Accordingly, on the 17th of August, 1771, I put a sprig of mint into a quantity of air, in which a wax candle had burned out, and found that, on the 27th of the same month, another candle burned perfectly well in it.

Although Priestley later discovered oxygen, which he named “dephlogisticated air,” it was Antoine Lavoisier who elucidated its role in combustion and respiration. Nevertheless, Priestley’s work inspired the Dutch physician Jan Ingenhousz, who in 1779 demonstrated that the “purifying” power of plants resides in the influence of sunlight on their green parts. In 1782, the Swiss pastor Jean Senebier showed that CO₂, which he called “fixed air,” is taken up during photosynthesis. His compatriot Nicolas-Théodore de Saussure found, in 1804, that the combined weights of the organic matter produced by plants and the oxygen they evolve is greater than the weight of the CO₂ they consume. He therefore concluded that water, the only other substance he added to his system, was also necessary for photosynthesis. The final ingredient in the overall photosynthetic recipe was established in 1842 by the German physiologist Robert Mayer, one of the formulators of the first law of thermodynamics, who concluded that plants convert light energy to chemical energy.

Life on Earth depends on the sun. *Plants and cyanobacteria chemically sequester light energy through photosynthesis, a light-driven process in which CO₂ is “fixed” to yield carbohydrates (CH₂O).*

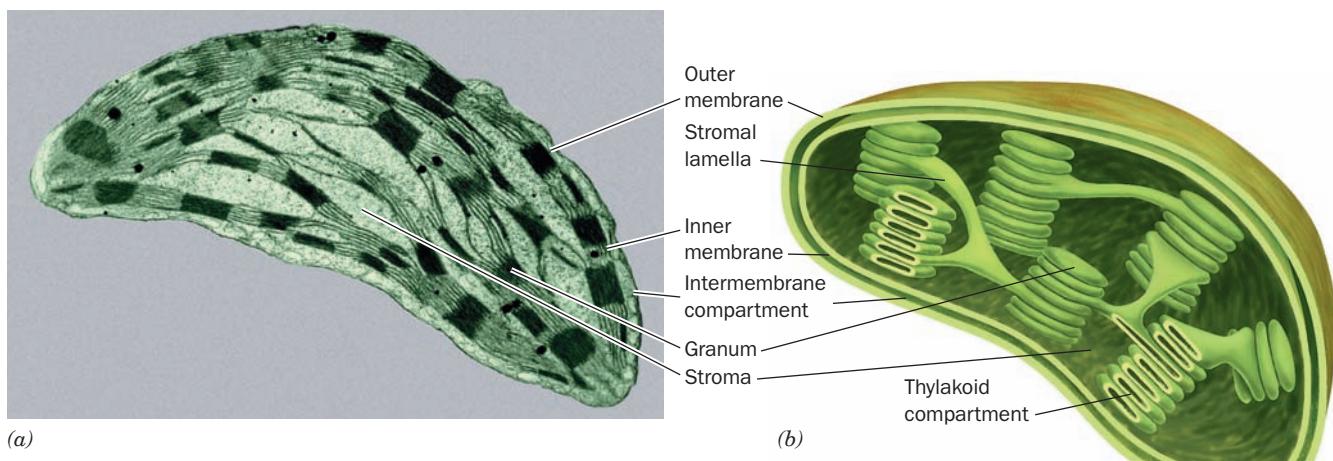


This process, in which CO₂ is reduced and H₂O is oxidized to yield carbohydrates and O₂, is essentially the reverse of oxidative carbohydrate metabolism. Photosynthetically produced carbohydrates therefore serve as an energy source for the organism that produced them as well as for nonphotosynthetic organisms that directly or indirectly consume photosynthetic organisms. In fact, even modern industry is highly dependent on the products of photosynthesis because coal, oil, and gas (the so-called fossil fuels) are thought to be the remains of ancient organisms. It is estimated that photosynthesis annually fixes $\sim 10^{11}$ tons of carbon, which represents the storage of over 10¹⁸ kJ of energy. Moreover, photosynthesis, over the eons, has produced the O₂ in Earth’s atmosphere (Section 1-5Cb).

The notion that plants obtain nourishment from such insubstantial things as light and air took nearly two centuries to develop. In 1648, the Flemish physician Jean Baptiste van Helmont reported that growing a potted willow tree from a shoot caused an insignificant change in the weight of the soil in which the tree had been rooted. Although another century was to pass before the law of conservation of matter was formulated, van Helmont attributed the tree’s weight gain to the water it had taken up. This idea was extended in 1727 by Stephen Hales, who proposed that plants extract some of their matter from the air.

1 CHLOROPLASTS

The site of photosynthesis in eukaryotes (algae and higher plants) is the **chloroplast** (Section 1-2Ag), a member of the membranous subcellular organelles peculiar to plants known as **plastids**. The first indication that chloroplasts have a photosynthetic function was Theodor Englemann’s observation, in 1882, that small, motile, O₂-seeking bacteria congregate at the surface of the alga *Spirogyra*, overlying its single chloroplast, but only while the chloroplast is illuminated. Chloroplasts must therefore be the site of light-induced O₂ evolution, that is, photosynthesis. Chloroplasts, of which there are 1 to 1000 per cell, vary considerably in size and



(a)

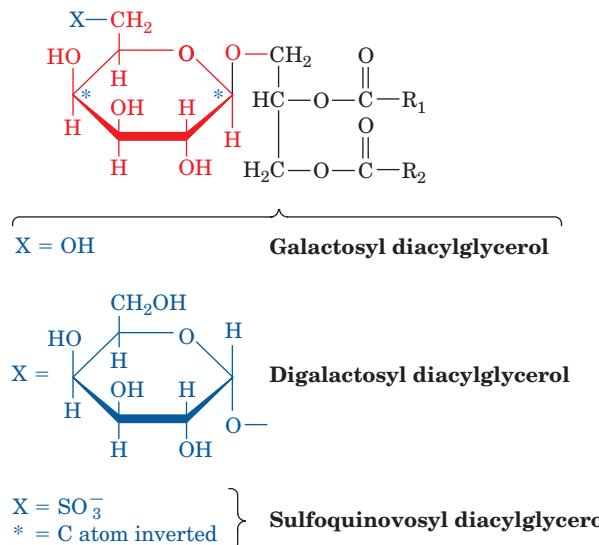
(b)

Figure 24-1 Chloroplast from corn. (a) An electron micrograph. (b) Schematic diagram.

[Electron micrograph courtesy of Lester Shumway, College of Eastern Utah.]

shape but are typically $\sim 5\text{-}\mu\text{m}$ -long ellipsoids. Like mitochondria, which they resemble in many ways, chloroplasts have a highly permeable outer membrane and a nearly impermeable inner membrane separated by a narrow intermembrane space (Fig. 24-1). The inner membrane encloses the **stroma**, a concentrated solution of enzymes much like the mitochondrial matrix, that also contains the DNA, RNA, and ribosomes involved in the synthesis of several chloroplast proteins. The stroma, in turn, surrounds a third membranous compartment, the **thylakoid** (Greek: *thylakos*, a sac or pouch). The thylakoid is probably a single highly folded vesicle, although in most organisms it appears to consist of stacks of disklike sacs named **grana**, which are interconnected by unstacked **stromal lamellae**. A chloroplast usually contains 10 to 100 grana. Thylakoid membranes arise from invaginations in the inner membrane of developing chloroplasts and therefore resemble mitochondrial cristae.

The lipids of the thylakoid membrane have a distinctive composition. They consist of only $\sim 10\%$ phospholipids; the majority, $\sim 80\%$, are uncharged **mono-** and **digalactosyl diacylglycerols**, and the remaining $\sim 10\%$ are the sulfolipids **sulfoquinovosyl diacylglycerols** (**quinovose** is 6-deoxyglucose):



The acyl chains of these lipids have a high degree of unsaturation, which gives the thylakoid membrane a highly fluid character.

Photosynthesis occurs in two distinct phases:

1. The **light reactions**, which use light energy to generate NADPH and ATP.

2. The **dark reactions**, actually light-independent reactions, which use NADPH and ATP to drive the synthesis of carbohydrate from CO_2 and H_2O .

The light reactions occur in the thylakoid membrane and involve processes that resemble mitochondrial electron transport and oxidative phosphorylation (Sections 22-2 and 22-3). In photosynthetic prokaryotes, which lack chloroplasts, the light reactions take place in the cell's plasma (inner) membrane or in highly invaginated structures derived from it called **chromatophores** (e.g., Fig. 24-2; recall that chloroplasts evolved from cyanobacteria that assumed a symbiotic relationship with a nonphotosynthetic

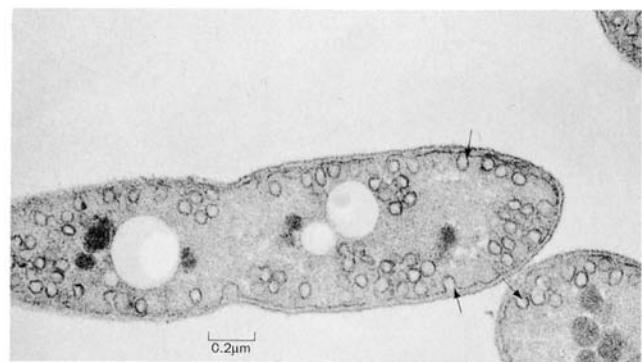


Figure 24-2 Electron micrograph of a section through the purple photosynthetic bacterium *Rhodobacter sphaeroides*. Its plasma membrane invaginates to form externally connected tubules known as chromatophores (arrows; seen here in circular cross section) that are the sites of photosynthesis. [Courtesy of Gerald A. Peters, Virginia Commonwealth University.]

eukaryote; Section 1-2Ag). In eukaryotes, the dark reactions occur in the stroma through a cyclic series of enzyme-catalyzed reactions. In the following sections, we consider the light and dark reactions in detail.

2 LIGHT REACTIONS

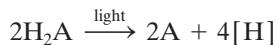
In the first decades of the twentieth century, it was generally assumed that light, as absorbed by photosynthetic pigments, directly reduced CO_2 , which, in turn, combined with water to form carbohydrate. In this view, CO_2 is the source of the O_2 generated by photosynthesis. In 1931, however, Cornelis van Niel showed that green photosynthetic bacteria, anaerobes that use H_2S in photosynthesis, generate sulfur:



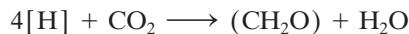
The chemical similarity between H_2S and H_2O led van Niel to propose that the general photosynthetic reaction is



where H_2A is H_2O in green plants and cyanobacteria and H_2S in photosynthetic sulfur bacteria. This suggests that photosynthesis is a two-stage process in which light energy is harnessed to oxidize H_2A (the light reactions):



and the resulting reducing agent $[\text{H}]$ subsequently reduces CO_2 (the dark reactions):



Thus, in aerobic photosynthesis, H_2O , not CO_2 , is photolyzed (split by light).

The validity of van Niel's hypothesis was established unequivocally by two experiments. In 1937, Robert Hill discovered that when isolated chloroplasts that lack CO_2 are illuminated in the presence of an artificial electron acceptor such as ferricyanide $[\text{Fe}(\text{CN})_6]^{3-}$, O_2 is evolved with concomitant reduction of the acceptor [to ferrocyanide, $\text{Fe}(\text{CN})_6^{4-}$, in our example]. This so-called **Hill reaction** demonstrates that CO_2 does not participate directly in the O_2 -producing reaction. It was discovered eventually that the natural photosynthetic electron acceptor is NADP^+ (Fig. 13-2), whose reduction product, NADPH , is utilized in the dark reactions to reduce CO_2 to carbohydrate (Section 24-3A). In 1941, when the oxygen isotope ^{18}O became available, Samuel Ruben and Martin Kamen directly demonstrated that the source of the O_2 formed in photosynthesis is H_2O :



This section discusses the major aspects of the light reactions.

A. Absorption of Light

The principal photoreceptor in photosynthesis is **chlorophyll**. This cyclic tetrapyrrole, like the heme group of globins and cytochromes (Sections 10-1A and 22-2C), is derived biosynthetically from protoporphyrin IX. Chlorophyll, however, differs from heme in four major respects (Fig. 24-3):

1. Its central metal ion is Mg^{2+} rather than $\text{Fe}(\text{II})$ or $\text{Fe}(\text{III})$.
2. It has a cyclopentenone ring, Ring V, fused to pyrrole Ring III.
3. Pyrrole Ring IV is partially reduced in **chlorophyll a** (Chl *a*) and **chlorophyll b** (Chl *b*), the two major chlorophyll varieties in eukaryotes and cyanobacteria, whereas in **bacteriochlorophyll a** (BChl *a*) and **bacteriochlorophyll b** (BChl *b*), the principal chlorophylls of photosynthetic bacteria, Rings II and IV are partially reduced.
4. The propionyl side chain of Ring IV is esterified to a tetraisoprenoid alcohol. In Chl *a* and *b* as well as in BChl *b* it is **phytol** but in BChl *a* it is either **phytol** or **geranylgeranol**, depending on the bacterial species.

In addition, Chl *b* has a formyl group in place of the methyl substituent to atom C3 of Ring II of Chl *a*. Similarly, BChl *a* and BChl *b* have different substituents to atom C4.

a. Light and Matter Interact in Complex Ways

As photosynthesis is a light-driven process, it is worthwhile reviewing how light and matter interact. Electromagnetic radiation is propagated as discrete **quanta (photons)** whose energy E is given by **Planck's law**:

$$E = h\nu = \frac{hc}{\lambda} \quad [24.1]$$

where h is **Planck's constant** ($6.626 \times 10^{-34} \text{ J} \cdot \text{s}$), c is the speed of light ($2.998 \times 10^8 \text{ m} \cdot \text{s}^{-1}$ in a vacuum), ν is the frequency of the radiation, and λ is its wavelength (visible light ranges in wavelength from 400 to 700 nm). Thus red light with $\lambda = 680 \text{ nm}$ has an energy of $176 \text{ kJ} \cdot \text{einstein}^{-1}$ (an **einstein** is a mole of photons).

Molecules, like atoms, have numerous electronic quantum states of differing energies. Moreover, because molecules contain more than one nucleus, each of their electronic states has an associated series of vibrational and rotational substates that are closely spaced in energy (Fig. 24-4). Absorption of light by a molecule usually occurs through the promotion of an electron from its ground (lowest energy) state molecular orbital to one of higher energy. However, *a given molecule can only absorb photons of certain wavelengths because, as is required by the law of conservation of energy, the energy difference between the two states must exactly match the energy of the absorbed photon.*

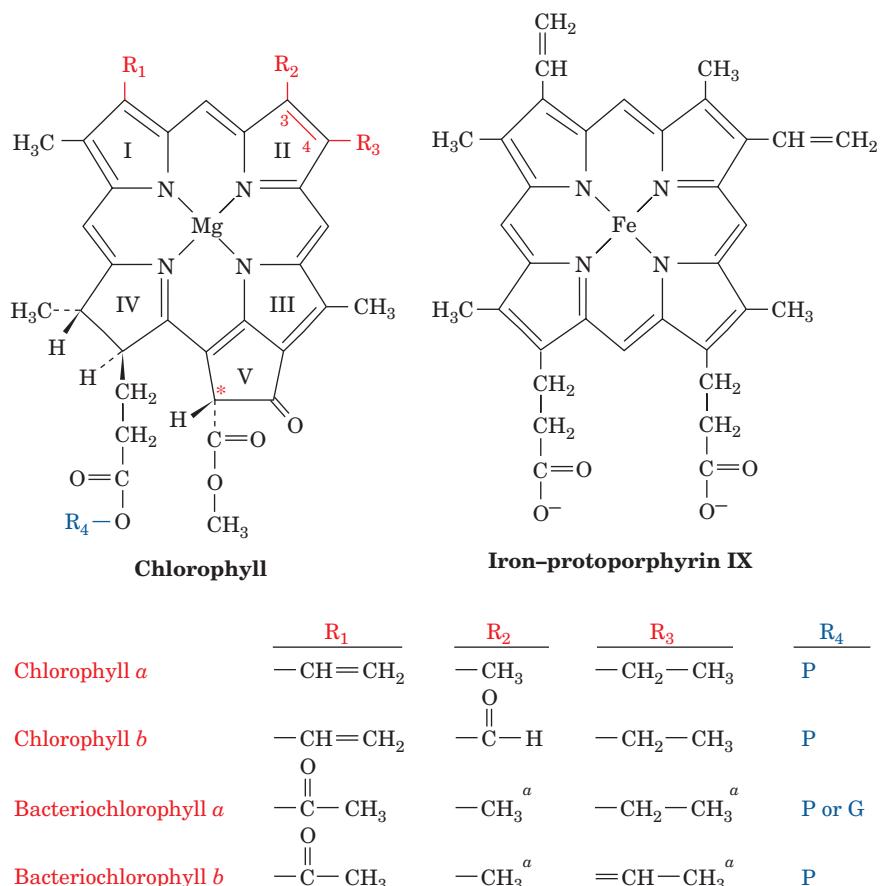
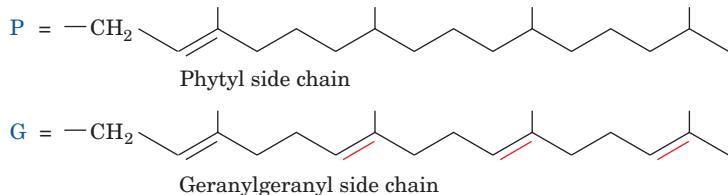


Figure 24-3 Chlorophyll structures. The molecular formulas of chlorophylls *a* and *b* and bacteriochlorophylls *a* and *b* are compared to that of iron protoporphyrin IX (heme). The starred atom has the opposite stereochemistry in **chlorophyll *a'* (Chl *a'*)**. The isoprenoid phytyl and geranylgeranyl tails presumably increase the chlorophylls' solubility in nonpolar media.

^a No double bond between positions C3 and C4.



The various chlorophylls are highly conjugated molecules (Fig. 24-3). It is just such molecules that strongly absorb visible light (the spectral band in which the solar radiation reaching Earth's surface is of peak intensity). In fact, the peak molar extinction coefficients of the various chlorophylls, $>10^5 \text{ M}^{-1} \cdot \text{cm}^{-1}$, are among the highest known for organic molecules (Fig. 24-5; absorbance spectra are discussed in Section 5-3Ca). Yet the relatively small chemical differences among the various chlorophylls greatly affect their absorption spectra. These spectral differences, as we shall see, are functionally significant.

An electronically excited molecule can dissipate its excitation energy in many ways. Those modes with the greatest photosynthetic significance are as follows (Fig. 24-4):

1. Internal conversion, a common mode of decay in which electronic energy is converted to the kinetic energy of molecular motion, that is, to heat. This process occurs

very rapidly, being complete in $<10^{-11} \text{ s}$. Many molecules relax in this manner to their ground states. Chlorophyll molecules, however, usually relax only to their lowest excited states. Therefore, *the photosynthetically applicable excitation energy of a chlorophyll molecule that has absorbed a photon in its short-wavelength band, which corresponds to its second excited state, is no different than if it had absorbed a photon in its less energetic long-wavelength band*.

2. Fluorescence, in which an electronically excited molecule decays to its ground state by emitting a photon. Such a process requires $\sim 10^{-8} \text{ s}$, so it occurs much more slowly than internal conversion. Consequently, a fluorescently emitted photon generally has a longer wavelength (lower energy) than that initially absorbed. Fluorescence accounts for the dissipation of only 3 to 6% of the light energy absorbed by living plants. However, chlorophyll in solution,

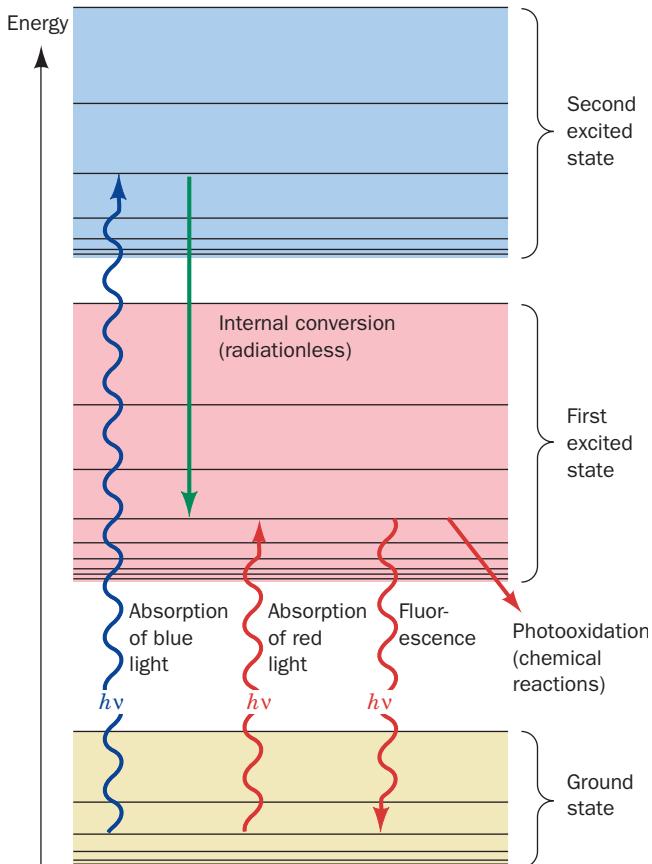


Figure 24-4 Energy diagram indicating the electronic states of chlorophyll and their most important modes of interconversion. The thin black lines denote different vibrational and rotational substates of each electronic state. The wiggly arrows represent the absorption of photons or their fluorescent emission. Excitation energy may also be dissipated in radiationless processes such as internal conversion (heat production) or chemical reactions.

 See the Animated Figures

where of course the photosynthetic uptake of this energy cannot occur, has an intense red fluorescence.

3. Exciton transfer (also known as **resonance energy transfer**), in which an excited molecule directly transfers its excitation energy to nearby unexcited molecules with similar electronic properties (the basis of FRET; Section 9-1Cd). This process occurs through interactions between the molecular orbitals of the participating molecules in a manner analogous to the interactions between mechanically coupled pendulums of similar frequencies. An exciton (excitation) may be serially transferred between members of a group of molecules or, if their electronic coupling is strong enough, the entire group may act as a single excited “supermolecule.” We shall see that *exciton transfer is of particular importance in funneling light energy to photosynthetic reaction centers*.

4. Photooxidation, in which a light-excited donor molecule is oxidized by transferring an electron to an acceptor molecule, which is thereby reduced. This process occurs because the transferred electron is less tightly bound to the donor in its excited state than it is in the ground state. In photosynthesis, excited chlorophyll (Chl^*) is such a donor. *The energy of the absorbed photon is thereby chemically transferred to the photosynthetic reaction system.* Photooxidized chlorophyll, Chl^+ , a cationic free radical, eventually returns to its ground state by oxidizing some other molecule.

b. Light Absorbed by Antenna Chlorophylls Is Transferred to Photosynthetic Reaction Centers

The primary reactions of photosynthesis, as is explained in Sections 24-2B and 24-2C, take place at **photosynthetic reaction centers (RCs)**. Yet *photosynthetic organelles contain far more chlorophyll molecules than RCs*. This was

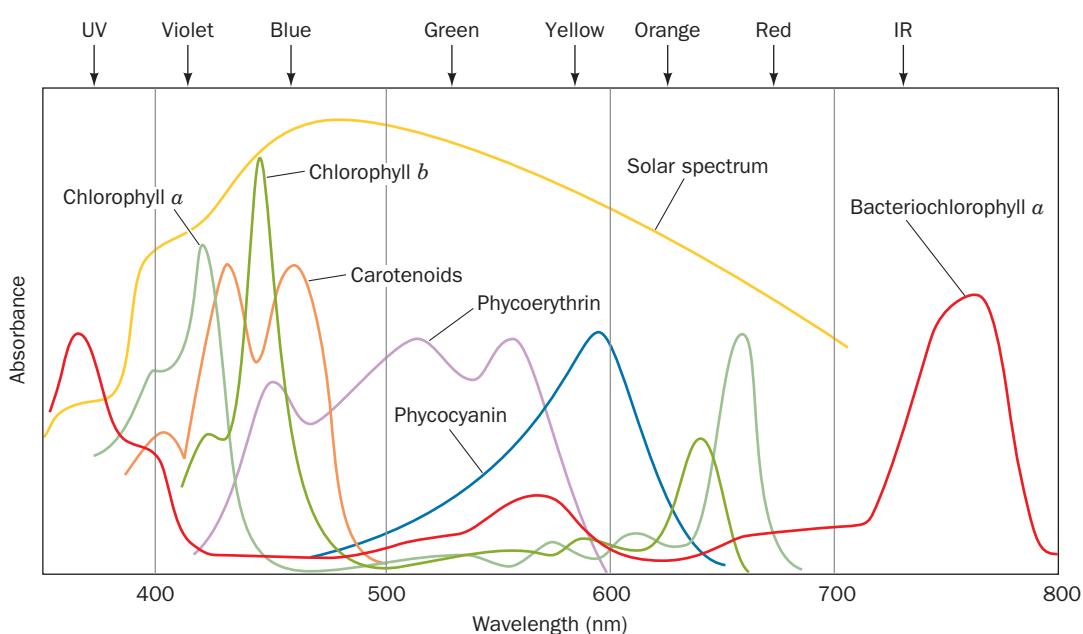


Figure 24-5 Absorption spectra of various photosynthetic pigments. The chlorophylls each have two absorption bands, one in the red and one in the blue. Phycoerythrin absorbs blue and

green light, whereas phycocyanin absorbs yellow light. Together, these pigments absorb most of the visible light in the solar spectrum. [After a drawing by Govindjee, University of Illinois.]

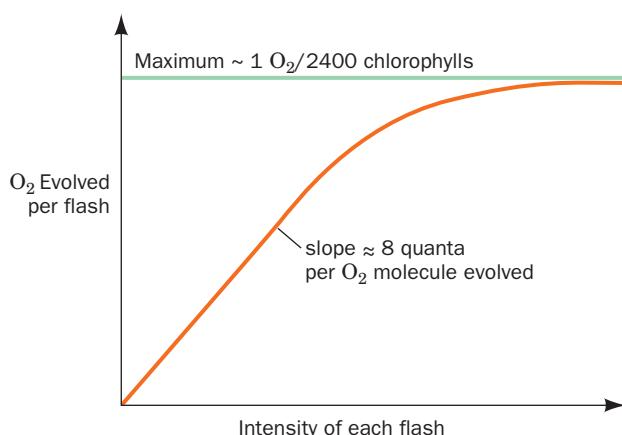


Figure 24-6 The amount of O_2 evolved by *Chlorella* algae versus the intensity of light flashes. Flashes are separated by dark intervals of >20 ms.

demonstrated in 1932 by Robert Emerson and William Arnold in their studies of O_2 production by the green alga *Chlorella* (a favorite experimental subject), which had been exposed to repeated brief ($10\text{-}\mu\text{s}$) flashes of light. The amount of O_2 generated per flash was maximal when the interval between flashes was at least 20 ms. Evidently, this is the time required for a single turnover of the photosynthetic reaction cycle. Emerson and Arnold then measured the variation of O_2 yield with flash intensity when the flash interval was the optimal 20 ms. With weak flashes, the O_2 increased linearly with flash intensity such that about one molecule of O_2 was generated per eight photons absorbed (Fig. 24-6). With increasing flash intensity the efficiency of this process fell off, no doubt because the number of pho-

tons began to approach the number of photochemical units. What was unanticipated, however, was that each flash of saturating intensity produced only one molecule of O_2 per ~ 2400 molecules of chlorophyll present. Since at least eight photons must be sequentially absorbed to liberate one O_2 molecule (Section 24-2C), these results suggest that the photosynthetic apparatus contains $\sim 2400/8 = 300$ chlorophyll molecules per RC.

With such a great excess of chlorophyll molecules per RC, it seems unlikely that all participate directly in photochemical reactions. Rather, as subsequent experiments have shown, *most chlorophylls function to gather light; that is, they act as light-harvesting antennas*. These **antenna chlorophylls** pass the energy of an absorbed photon, by exciton transfer, from molecule to molecule until the excitation reaches an RC (Fig. 24-7a). There, the excitation is trapped because RC chlorophylls, although chemically identical to antenna chlorophylls, have slightly lower excited state energies because of their different environments (Fig. 24-7b).

Transfer of energy from the antenna system to an RC occurs in $<10^{-10}$ s with an efficiency of $>90\%$. This high efficiency depends on the chlorophyll molecules having appropriate spacings and relative orientations. Even in bright sunlight, an RC intercepts only ~ 1 photon per second, a metabolically insignificant rate, and hence, these **light-harvesting complexes (LHCs)** serve an essential function.

c. The LHCs of Purple Photosynthetic Bacteria Contain Multiple Symmetrically Arranged Light-Absorbing Molecules

Most **purple photosynthetic bacteria**, which are among the simplest photosynthetic organisms, have two types of LHCs, **LH1** and **LH2**, that are transmembrane proteins but have different spectral and biochemical properties. LH2, which absorbs light at shorter wavelengths than LH1, rapidly passes the energy from the photons it absorbs to

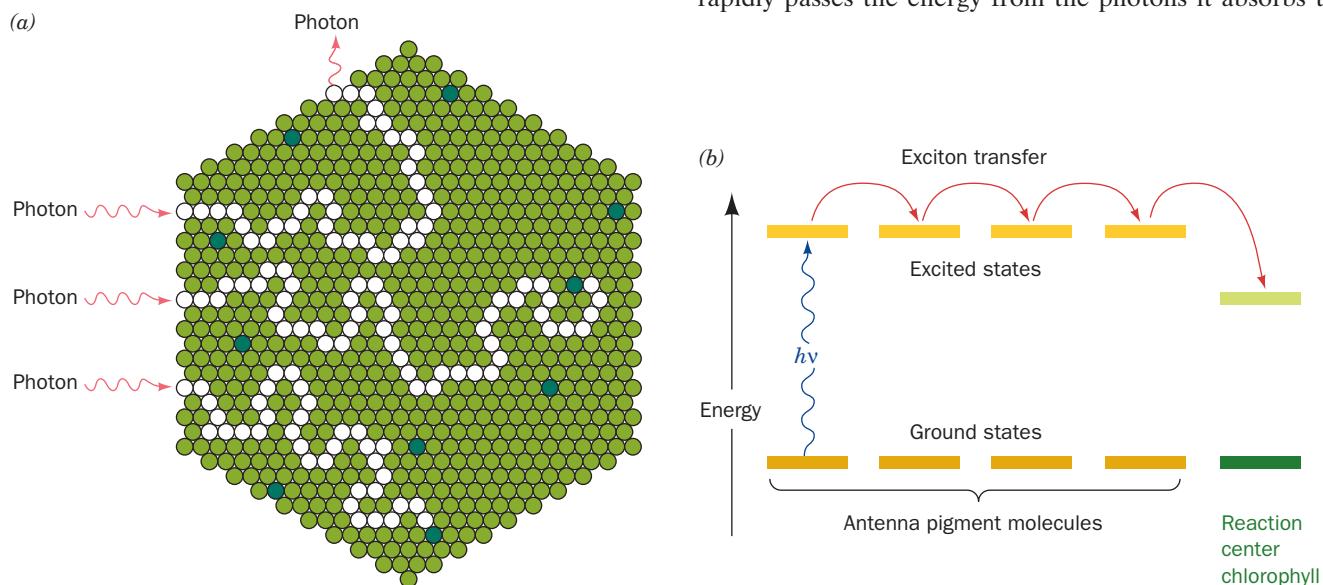
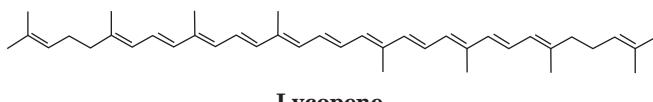


Figure 24-7 Flow of energy through a photosynthetic antenna complex. (a) The excitation resulting from photon absorption randomly migrates by exciton transfer among the molecules of the antenna complex (light green circles) until it is either trapped

by an RC chlorophyll (dark green circles) or, less frequently, fluorescently reemitted. (b) The excitation is trapped by the RC chlorophyll because its lowest excited state has a lower energy than those of the antenna pigment molecules.

LH1, which, in turn, passes it to the RC. The X-ray structure of LH2 from the purple photosynthetic bacterium *Rhodospirillum (Rs.) molischianum* (Fig. 24-8), determined by Hartmut Michel, reveals that this protein is an 8-fold rotationally symmetric $\alpha_8\beta_8$ 16-mer that binds 24 bacteriochlorophyll *a* (BChl *a*) molecules and 8 lycopene molecules (a **carotenoid**; see below):



Lycopene

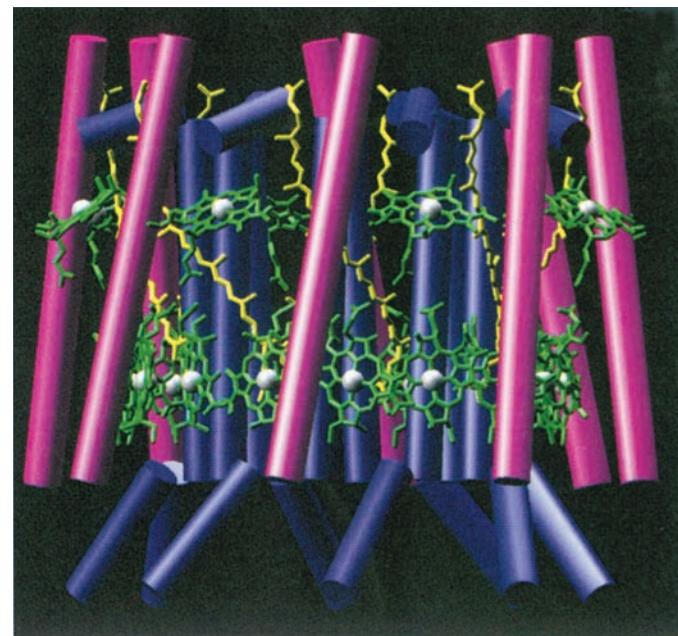
The α and β subunits (56 and 45 residues, respectively) both consist largely of single helices that are aligned nearly perpendicularly to the plane of the membrane in which they are embedded. The eight α subunits pack side by side to form a hollow cylinder of diameter ~ 31 Å (as measured between helix axes). Each of the eight β subunits occupies a position radially outward from an α subunit to form a concentric cylinder of diameter ~ 62 Å. Sixteen of the BChl *a* molecules are packed between these rings of helices in an arrangement resembling a 16-bladed turbine: Successive nearly parallel BChl *a* ring systems are in partial van der Waals contact (their Mg²⁺ ions are ~ 9 Å apart) with their planes perpendicular to the plane of the membrane. Their Mg²⁺ atoms are each singly axially liganded by His side

chains [much like the Fe(II) in deoxyhemoglobin] that alternately extend from an α and a β subunit around the lower end of the cylinder. The remaining eight BChl *a* molecules, which are each singly axially liganded by a side chain of Asp 6 α near the upper end of the cylinder, are arranged in an 8-fold symmetric ring between successive β subunit helices and are oriented with the planes of their ring systems tilted by $\sim 35^\circ$ relative to the plane of the membrane. The eight lycopene molecules are sandwiched between the α and β subunits and extend along much of their lengths, thereby contacting both sets of BChl *a* molecules. The LH2 from *Rhodopseudomonas (Rps.) acidophila*, another purple photosynthetic bacterium, is an $\alpha_9\beta_9$ 18-mer but otherwise has a similar structure in its transmembrane region to that of *Rs. molischianum*, even though their α and β subunits are only 26 and 31% identical.

Spectroscopic measurements indicate that an LH2's His-ligated and closely associated BChl *a* molecules maximally absorb radiation at a wavelength of 850 nm (and hence are called B850) and are strongly coupled, that is, they absorb radiation almost as a unit. The other, more loosely associated BChl *a* molecules (B800) maximally absorb radiation at 800 nm, largely as individual molecules (BChl *a*'s local environment in the protein alters its spectrum from that in solution; Fig. 24-5). When a B800 BChl *a* absorbs a photon, the excitation is rapidly [in ~ 700 femtoseconds (fs); 1 fs = 10^{-15} s] transferred to a lower energy B850 BChl *a* (which may independently absorb a photon),



(a)



(b)

Figure 24-8 X-ray structure of LH2 from *Rs. molischianum*. The α subunits are purple and the β subunits are pink. The bound chromophores are drawn in stick form with the BChl *a*'s green and the lycopenes yellow. The phytol tails of the BChl *a*'s have been truncated for clarity. (a) View perpendicular to the bacterial membrane from the cytoplasm. The polypeptide chains are drawn in worm form. (b) View parallel to the membrane with

the cytoplasm above. The protein subunits are represented by only their helices, which are shown as cylinders. The Mg²⁺ ions are represented by white spheres. [Courtesy of Juergen Koepke and Hartmut Michel, Max-Planck-Institut für Biochemie, Frankfurt, Germany. PDBID 1LGH.] See Interactive

Exercise 20

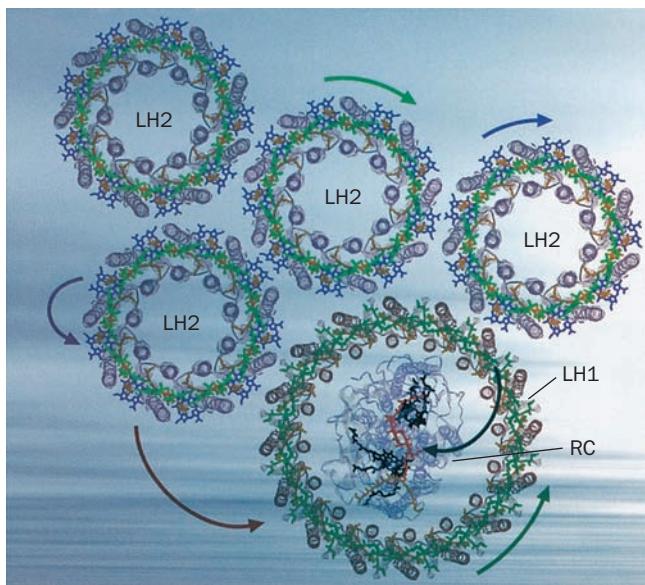


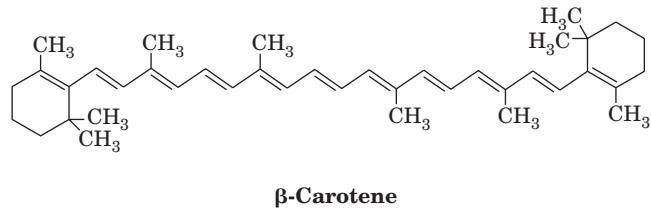
Figure 24-9 Model of the light-absorbing antenna system of purple photosynthetic bacteria. Several LH2s associate with each other and with LH1, which surrounds the photosynthetic reaction center (RC). The BChl *a*'s of LH2 B850 and LH1 are green, those of LH2 B800 are purple, and the light-absorbing pigments of the RC (see below) are red and black. Light absorbed by the BChl *a* and lycopene molecules of an LH2 is rapidly transferred (curved arrows), often via other contacting LH2s, to LH1, which, in turn, transfers the excitation to its enclosed RC. [From Bhattacharjee, Y., *Nature* **412**, 474 (2001).]

which even more rapidly (in \sim 100 fs) exchanges the excitation among the other B850 BChl *a* molecules. Hence, the B850 system acts as a kind of energy storage ring that delocalizes the excitation over a large region. The carotenoid molecules in this system absorb visible light (<800 nm) and may also facilitate the transmission of excitation between the rather distantly separated (19 Å between Mg atoms) nearest-neighbor B850 and B800 BChl *a* molecules.

LH1, like LH2, has α and β subunits of \sim 50 residues each. The low (8.5-Å) resolution structure of LH1 from *Rs. rubrum*, as determined by electron crystallography, reveals that it resembles LH2 but with 16-fold rotational symmetry, and forms a 116-Å-diameter cylinder with a 68-Å-diameter hole down its center. This hole is of sufficient size to contain an RC (see below), as electron microscopy studies indicate is, in fact, the case (Fig. 24-9). LH1's BChl *a* molecules absorb radiation at a longer wavelength than those of LH2 and consequently, when these two assemblies are in contact, excitation is rapidly [in 1–5 picoseconds (ps); 1 ps = 10^{-12} s] transferred from LH2 to LH1 and then (in 20–40 ps) to LH1's enclosed RC. Excitations may also be rapidly exchanged between contacting LH2s. Thus, this antenna system transfers virtually all of the radiation energy it absorbs to the RC in far less than the few nanoseconds (ns; 1 ns = 10^{-9} s) over which these excitations would otherwise decay. It should be noted that this complicated arrangement of **chromophores** (light-absorbing molecules) is among the simplest known; those of the light-harvesting systems of plants are even more elaborate (see below).

d. LHCs Contain Accessory Pigments

Most LHCs contain organized arrays of other light-absorbing substances in addition to chlorophyll. These **accessory pigments** function to fill in the absorption spectra of the antenna complexes in spectral regions where chlorophylls do not absorb strongly (Fig. 24-5). **Carotenoids**, which are C₄₀, largely linear polyenes such as lycopene and β -carotene,



are components of all green plants and many photosynthetic bacteria and are therefore the most common accessory pigments. They are largely responsible for the brilliant fall colors of deciduous trees as well as for the orange color of carrots (after which carotenoids are named).



[Ross M. Horowitz/The Image Bank/Getty Images.]

The light-harvesting protein **LHC-II** of green plants comprises \sim 30% of the protein in chloroplast membranes, which makes it the most abundant membrane protein in nature. Each subunit of this highly conserved, 232-residue, trimeric protein binds eight Chl *a*'s, six Chl *b*'s, and four carotenoids (Fig. 24-10), thereby accounting for around half the chlorophyll in the biosphere. The orientations of the 42 chlorophylls in each LHC-II trimer evenly sample nearly all directions in space, thus maximizing the efficiency of light harvesting.

Carotenoids serve an additional function besides that of light-gathering antennas: Through electronic interactions, they prevent their associated light-excited chlorophyll molecules from transferring this excitation to O₂, which would otherwise yield highly destructive reactive oxygen species (ROS; Section 22-4Cg). This is particularly important under full sunlight, when the rate that light energy is absorbed exceeds the rate that it can be used in photosynthesis. Then, the excess energy is dissipated as heat through internal

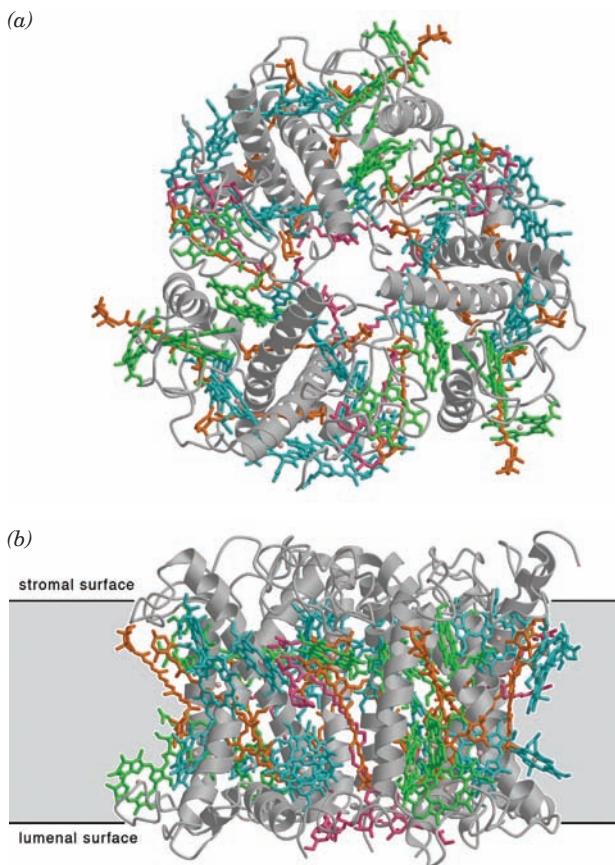
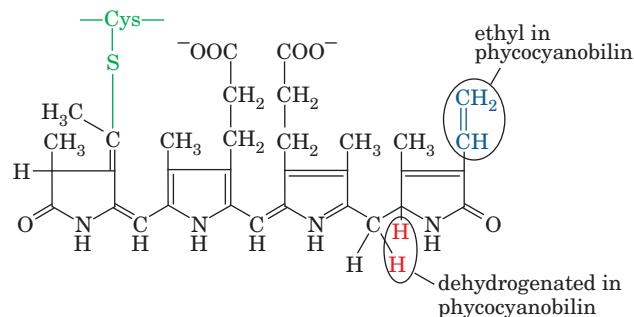


Figure 24-10 X-ray structure of the homotrimeric protein LHC-II from pea chloroplasts. The protein is drawn in ribbon form (gray) viewed (a) perpendicular to the thylakoid membrane from the stroma and along its 3-fold axis; and (b) parallel to the membrane (gray band) with the stroma above. Its bound carotenoids and chlorophylls are drawn in stick form with Chl *a* cyan, Chl *b* green, carotenoids orange, and lipids magenta. Mg²⁺ ions are represented by light pink spheres. Each subunit has three transmembrane helices oriented with its N-terminus on the stromal surface. Note the unusually high density of cofactors: Nearly 40% of this protein's nonhydrogen atoms comprise its chlorophylls and carotenoids. [Courtesy of Werner Kühlbrandt, Max Planck Institute of Biophysics, Frankfurt, Germany. PDBID 2BHW.]

conversion by carotenoids, thereby minimizing the irreversible damage to the photosynthetic system that would otherwise occur. In fact, the acidification of the thylakoid lumen resulting from high photosynthetic activity (see below) induces a conformational change in LHC-II. This converts it to a dissipative state by twisting one of its carotenoids and hence changing its electronic properties.

Aquatic photosynthetic organisms, which are responsible for nearly half of the photosynthesis on Earth, additionally contain other types of accessory pigments. This is because light outside the wavelengths 450 to 550 nm (blue and green light) is absorbed almost completely by passage through more than 10 m of water. In red algae and cyanobacteria, Chl *a* is therefore replaced as an antenna pigment by a

series of linear tetrapyrroles, notably the red **phycoerythrobilin** and the blue **phycocyanobilin**:



Peptide-linked phycoerythrobilin and phycocyanobilin

The lowest excited states of these so-called **bilins** have higher energies than those of the chlorophylls, thereby facilitating energy transfer to the RC. The bilins are covalently linked via Cys S atoms to **phycobiliproteins** to form **phycoerythrin** and **phycocyanin** (spectra in Fig. 24-5). These, in turn, are organized in high molecular mass particles called **phycobilisomes** that are bound to the outer faces of photosynthetic membranes so as to funnel excitation energy to RCs over long distances with >90% efficiency.

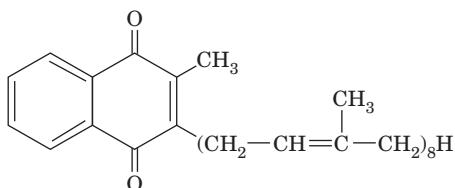
B. Electron Transport in Purple Photosynthetic Bacteria

Photosynthesis is a process in which electrons from excited chlorophyll molecules are passed through a series of acceptors that convert electronic energy to chemical energy. Thus two questions arise: (1) What is the mechanism of energy transduction; and (2) how do photooxidized chlorophyll molecules regain their lost electrons? We shall see that photosynthetic bacteria solve these problems somewhat differently from cyanobacteria and plants. We first discuss these mechanisms in photosynthetic bacteria, where they are simpler and better understood. Electron transport in cyanobacteria and plants is the subject of Section 24-2C.

a. The Photosynthetic Reaction Center Is a Transmembrane Protein Containing a Variety of Chromophores

The first indication that chlorophyll undergoes direct photooxidation during photosynthesis was obtained by Louis Duysens in 1952. He observed that illumination of membrane preparations from the purple photosynthetic bacterium *Rs. rubrum* caused a slight (~2%) bleaching of their absorbance at 870 nm, which returned to their original levels in the dark. Duysens suggested that this bleaching is caused by photooxidation of a bacteriochlorophyll complex that he named **P870** (P for pigment and 870 nm for the position of the major long-wavelength absorption band of BChl *a*; photosynthetic bacteria tend to inhabit murky stagnant ponds, so that they require an infrared-absorbing species of chlorophyll). The ability to detect the presence of P870 eventually led to the purification and characterization of the RC to which it is bound.

RC particles from several species of purple photosynthetic bacteria (**PbRCs**) have similar compositions. That from *Rps. viridis* consists of three hydrophobic subunits: H (258 residues), L (273 residues), and M (323 residues). The L and M subunits of this membrane-spanning protein collectively bind four molecules of BChl *b* (which maximally absorbs light at 960 nm), two molecules of **bacteriopheophytin *b*** (**BPho *b***; BChl *b* in which the Mg²⁺ is replaced by two protons), one nonheme/non-Fe–S Fe(II) ion, one molecule of the redox coenzyme ubiquinone (Fig. 22-17*b*), and one molecule of the related **menaquinone**



Menaquinone

(**vitamin K₂**, a substance required for proper blood clotting; Section 35-1Ba). In many PbRCs, however, the BChl *b*, BPho *b*, and menaquinone are replaced by BChl *a*, **BPho *a***, and a second ubiquinone, respectively.

The RC of *Rps. viridis*, whose X-ray structure was determined by Johann Deisenhofer, Robert Huber, and Hartmut Michel in 1984, was the first transmembrane protein to be described in atomic detail (Fig. 12-26). *The protein's transmembrane portion consists of 11 α helices that form a 45-Å-long flattened cylinder with the expected hydrophobic surface.* A *c*-type cytochrome containing four hemes, which is an integral constituent of the PbRC complex in only some photosynthetic bacteria, binds to the PbRC on the external side of the plasma membrane. In fact, the PbRC from another bacterial species, *Rhodobacter (Rb.) sphaeroides*, whose X-ray structure (Fig. 24-11) was independently determined by Marianne Schiffer and by Douglas Rees and George Feher, is nearly identical to that of *Rps. viridis* but lacks such a bound cytochrome.

b. Two BChl Molecules Form a “Special Pair”

The most striking aspect of the PbRC is that its chromophoric prosthetic groups are arranged with nearly perfect 2-fold symmetry (Fig. 24-12*a*). This symmetry arises because the L and M subunits, with which these prosthetic groups are exclusively associated, have homologous sequences and similar folds. Two of the BChl *b* molecules in the *Rps. viridis* PbRC, the so-called **special pair**, are closely associated; they are nearly parallel and have an Mg–Mg distance of ~7 Å. The special pair occupies a predominantly hydrophobic region of the protein and each of its Mg²⁺ ions has a His side chain as a fifth ligand. Each member of the special pair is in contact with another His-ligated BChl *b* molecule, which, in turn, is associated with a BPho *b* molecule. The menaquinone is in close association with the L subunit BPho *b* (Fig. 24-12*a*, right), whereas the ubiquinone, which is but loosely bound to the protein,

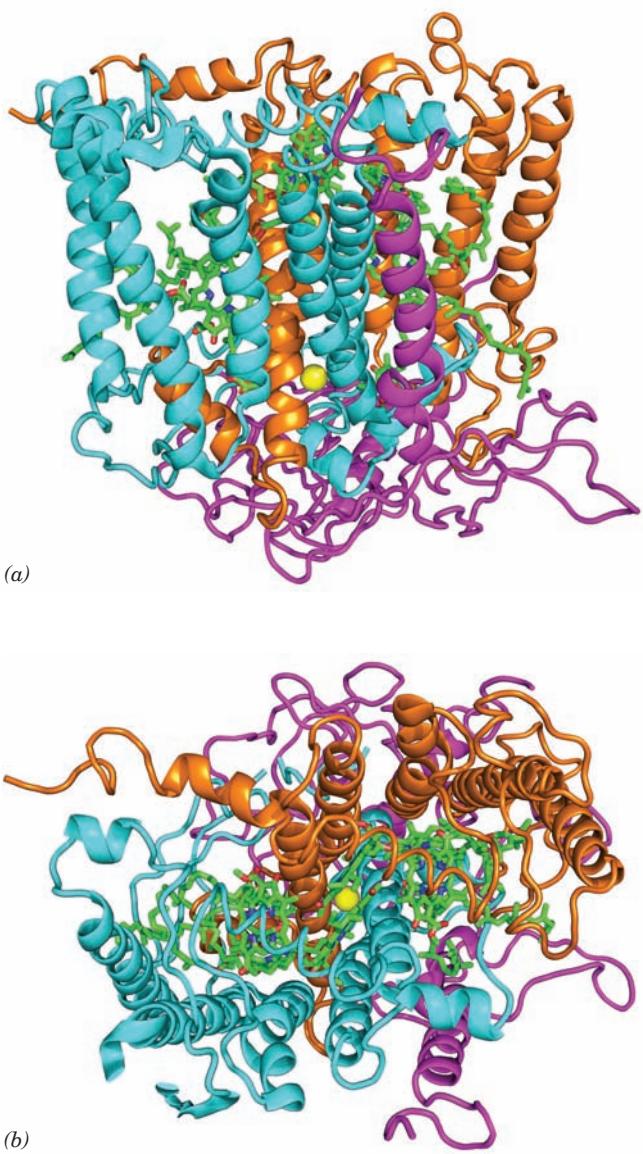


Figure 24-11 A ribbon diagram of the photosynthetic reaction center (RC) from *Rb. sphaeroides*. (a) The H, M, and L subunits, as viewed from within the plane of the plasma membrane with the cytoplasm below, are magenta, cyan, and orange, respectively. The prosthetic groups are drawn in stick form with C green, N blue, and O red. The Fe(II) atom is represented by a yellow sphere. The 11 largely vertical helices that form the central portion of the protein constitute its transmembrane region. Compare this structure with that of the RC from *Rps. viridis* (Fig. 12-26), whose H, M, and L subunits are 39, 50, and 59% identical to those of *Rb. sphaeroides*. Note that the *Rb. sphaeroides* protein lacks the four-heme *c*-type cytochrome (green in Fig. 12-26) on its periplasmic surface and that the Q_A prosthetic group, whose quinone ring lies to the right of the Fe(II), is ubiquinone in *Rb. sphaeroides* but menaquinone in *Rps. viridis*. (b) View from the extracellular side of the membrane. Note how the transmembrane portions of the M and L subunits are related by a pseudo-2-fold axis passing through the Fe(II) ion and that the prosthetic groups are sandwiched between these two subunits. [Based on an X-ray structure by Marianne Schiffer, Argonne National Laboratory. PDBid 2RCR.]  See [Interactive Exercise 21 and Kinemage Exercise 8-2](#)

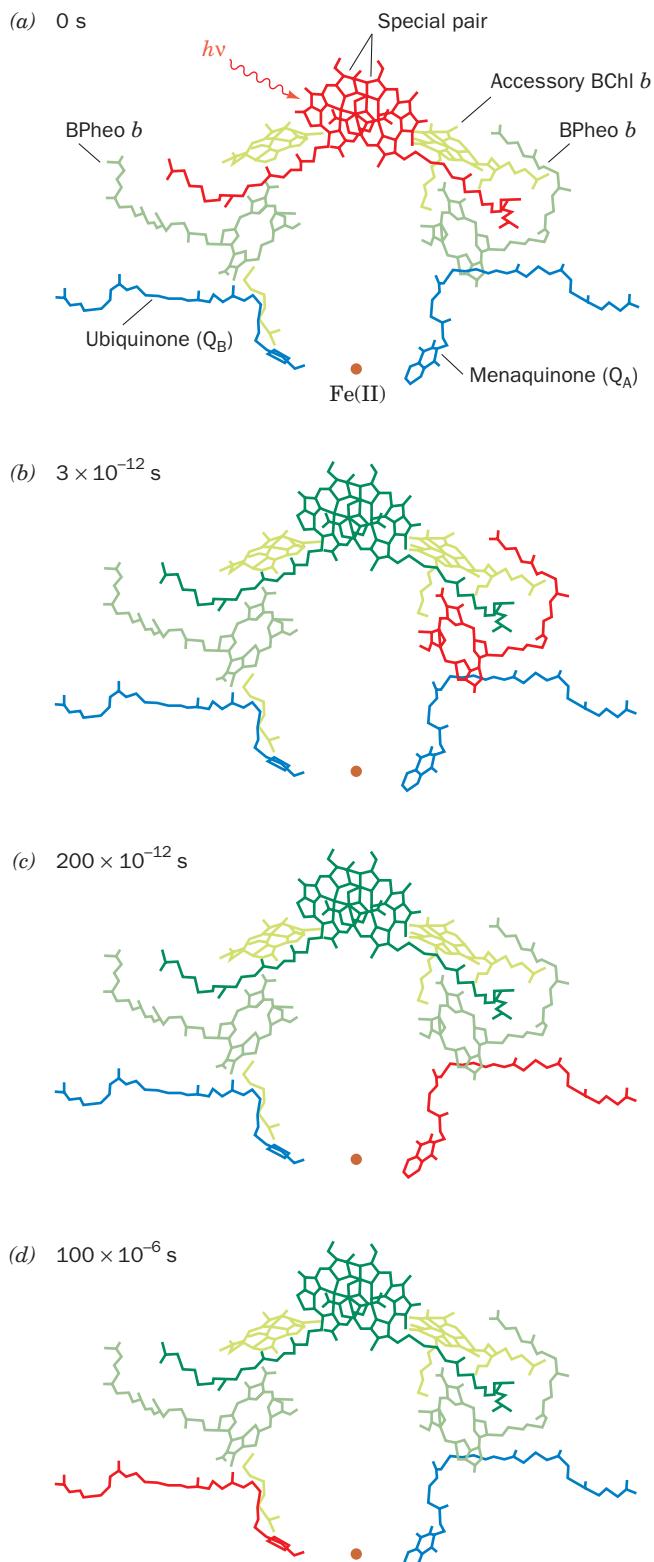


Figure 24-12 Sequence of excitations in the bacterial RC of *Rps. viridis*. The RC chromophores are shown in the same view as in Fig. 12-26a, which resembles that in Fig. 24-11a. Note that their rings, but not their aliphatic side chains, are arranged with close to 2-fold symmetry. (a) At zero time, a photon is absorbed by the “special pair” of BChl *b* molecules, thereby collectively raising them to an excited state [in each step, the excited molecule(s) is shown in red]. (b) Within 3 ps, an excited electron has passed to the BPheo *b* of the L subunit (right arm of the system) without becoming closely associated with the accessory BChl *b*. The special pair is thereby left with a positive charge. (c) Some 200 ps later, the excited electron has transferred to the menaquinone (Q_A , which is ubiquinone in *Rb. sphaeroides*). (d) Within the next 100 μ s, the special pair has been reduced (via an electron-transport chain discussed in the text), thereby eliminating its positive charge, while the excited electron migrates to the ubiquinone (Q_B). After a second such electron has been transferred to Q_B , it picks up two protons from solution and exchanges with the membrane-bound ubiquinone pool.

See Kinemage Exercise 8-2

described below. The Fe(II) is positioned between the menaquinone and ubiquinone rings and is octahedrally liganded by four His side chains and the two carboxyl oxygen atoms of a Glu side chain. Curiously, the two symmetry related groups of chromophores are not functionally equivalent; electrons, as we shall see, are almost exclusively transferred through the L subunit (the right sides of Figs. 24-11 and 24-12). This effect is generally attributed to subtle structural and electronic differences between the L and M subunits.

c. The Electronic States of Molecules Undergoing Fast Reactions Can Be Monitored by EPR and Laser Spectroscopy Techniques

The turnover time of a photosynthetic reaction cycle, as we have seen, is only a few milliseconds. Its sequence of reactions can therefore only be traced by measurements that can follow extremely rapid electronic changes in molecules. Two techniques are well suited to this task:

1. Electron paramagnetic resonance (EPR) spectroscopy [also called **electron spin resonance (ESR) spectroscopy**], which detects the spins of unpaired electrons in a manner analogous to the detection of nuclear spins in NMR spectroscopy. A molecular species with unpaired electrons, such as an organic radical or a transition metal ion, has a characteristic EPR spectrum because its unpaired electrons interact with the magnetic fields generated by the nuclei and the other electrons of the molecule. Paramagnetic species as short lived as 10 ps can exhibit definitive EPR spectra.

2. Optical spectroscopy using pulsed lasers. Laser flashes as brief as 20 attoseconds (as; 1 as = 10^{-18} s) have been generated. By monitoring the bleaching (disappearance) of certain absorption bands and the emergence of others, laser spectroscopy can track the time course of a fast reaction process.

associates with the M subunit BPheo *b* (Fig. 24-12a, *left*). These various chromophores are closely associated with a number of protein aromatic rings, which are therefore also thought to participate in the electron-transfer process

d. Photon Absorption Rapidly Photooxidizes the Special Pair

The sequence of photochemical events mediated by the photosynthetic reaction center is diagrammed in Fig. 24-12:

(a) The primary photochemical event of bacterial photosynthesis is absorption of a photon by the special pair (P870 or **P960** depending on whether it consists of BChl *a* or *b*; here, for argument's sake, we assume it to be P960). This event is nearly instantaneous; it occupies the ~3-fs oscillation time of a light wave. EPR measurements established that P960 is, in fact, a pair of BChl *b* molecules and indicated that the excited electron is delocalized over both of them.

(b) P960*, the excited state of P960, has but a fleeting existence. Laser spectroscopy has demonstrated that within ~3 ps after its formation, P960* has transferred an electron to the BPheo *b* on the right in Fig. 24-12*b* to yield P960⁺ BPheo *b*⁻. In forming this radical pair, the transferred electron must pass near but seems not to reduce the intervening BChl *b* (which is therefore termed an accessory chlorophyll), although its position strongly suggests that it has an important role in conveying electrons.

(c) By some 200 ps later, the electron has further migrated to the menaquinone (or, in many species, the second ubiquinone), designated Q_A, to form the anionic semiquinone radical Q_A⁻. All these electron transfers, as diagrammed in Fig. 24-13, are to progressively lower energy states, which makes this process all but irreversible.

Rapid removal of the excited electron from the vicinity of P960⁺ is an essential feature of the PbRC; this prevents back reactions that would return the electron to P960⁺ so as to provide the time required for the wasteful internal conversion of its excitation energy to heat. In fact, *this sequence of electron transfers is so efficient that its overall quantum yield (ratio of molecules reacted to photons absorbed) is virtually 100%*. No man-made device has yet approached this level of efficiency.

e. Electrons Are Returned to the Photooxidized Special Pair via an Electron-Transport Chain

The remainder of the photosynthetic electron-transport process occurs on a much slower timescale. Within ~100 μ s after its formation, Q_A⁻, which occupies a hydrophobic pocket in the protein, transfers its excited electron to the more solvent-exposed ubiquinone, Q_B, to form Q_B⁺ (Fig. 24-12*d*). The nonheme Fe(II) is not reduced in this process and, in fact, its removal only slightly affects the electron transfer rate, so that the Fe(II) probably functions to fine-tune the PbRC's electronic character. Q_A never becomes fully reduced; it shuttles between its oxidized and semiquinone forms. Moreover, the lifetime of Q_A⁻ is so short that it never becomes protonated. In contrast, once the PbRC again becomes excited, it transfers a second electron to Q_B⁺ to form the fully reduced Q_B²⁻. This anionic quinol takes up two protons from the solution on the cytoplasmic side of the plasma membrane to form Q_BH₂. Thus Q_B is a molecular transducer that converts two

light-driven one-electron excitations to a two-electron chemical reduction.

The electrons taken up by Q_BH₂ are eventually returned to P960⁺ via a complex electron-transport chain (Fig. 24-13). The details of this process are more species dependent than the preceding and are not so well understood. The available redox carriers include a membrane-bound pool of ubiquinone molecules, **cytochrome bc₁**, and **cytochrome c₂**. Cytochrome bc₁ is a transmembrane protein complex composed of a [2Fe-2S] cluster-containing subunit; a heme *c*-containing cytochrome *c*₁; a cytochrome *b* that contains two functionally inequivalent heme *b*'s, *b*_H and *b*_L (H and L for high and low potential); and, in some species, a fourth subunit. Note that cytochrome bc₁ is strikingly similar to the proton-translocating Complex III of mitochondria (Section 22-2C3a), which is also called cytochrome bc₁. The electron-transport pathway leads from Q_BH₂ on the cytoplasmic side of the plasma membrane, through the ubiquinone pool, with which Q_BH₂ exchanges, to cytochrome bc₁, and then to cytochrome c₂ on the external (periplasmic) side of the plasma membrane. The reduced cytochrome c₂, which, as its name implies, closely resembles mitochondrial cytochrome *c*, diffuses along the external membrane surface until it reacts with the membrane-spanning PbRC to transfer an electron to P960⁺ (the structures of several *c*-type cytochromes, including that of cytochrome c₂ from *Rs. rubrum*, are diagrammed in Fig. 9-41). In *Rps. viridis*, the four-heme *c*-type cytochrome bound to the PbRC complex on the external side of the plasma membrane (Fig. 12-26) is interposed between cytochrome c₂ and P960⁺. Note that one of this *c*-type cytochrome's hemes is positioned to reduce the photooxidized special pair. The PbRC is thereby prepared to absorb another photon.

f. Photosynthetic Electron Transport Drives the Formation of a Proton Gradient

Since electron transport in PbRCs is a cyclic process (Fig. 24-13), *it results in no net oxidation-reduction. Rather, it functions to translocate the cytoplasmic protons acquired by Q_BH₂ across the plasma membrane, thereby making the cell alkaline relative to its environment.* The mechanism of this process is essentially identical to that of proton transport in mitochondrial Complex III (Section 22-3Be); that is, in addition to the translocation of the two H⁺ resulting from the two-electron reduction of Q_B to QH₂, a Q cycle mediated by cytochrome bc₁ translocates two H⁺ for a total of four H⁺ translocated per two photons absorbed (Fig. 24-13*a*; also see Fig. 22-31). *Synthesis of ATP, a process known as photophosphorylation, is driven by the dissipation of the resulting pH gradient in a manner that closely resembles ATP synthesis in oxidative phosphorylation (Section 22-3C).* We further discuss the mechanism of photophosphorylation in Section 24-2D.

Photosynthetic bacteria use photophosphorylation-generated ATP to drive their various endergonic processes. However, unlike cyanobacteria and plants, which generate their required reducing equivalents by the light-driven oxidation of H₂O (see below), photosynthetic bacteria must obtain their reducing equivalents from the environment.

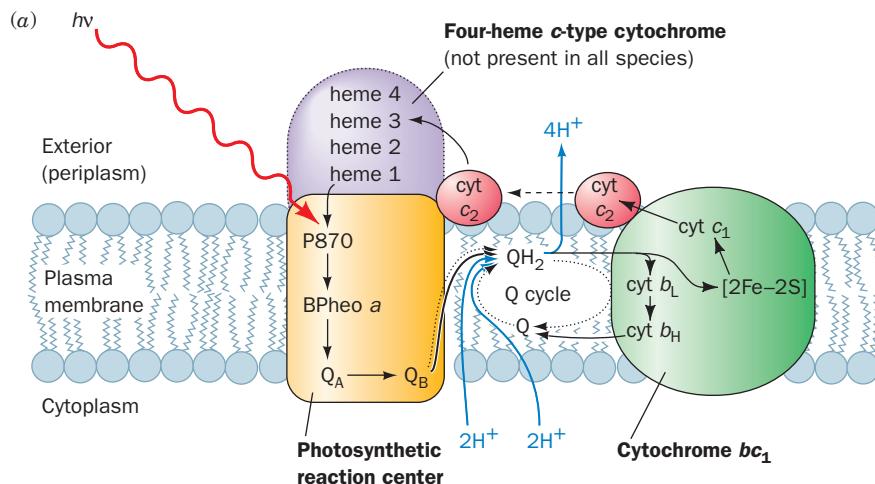
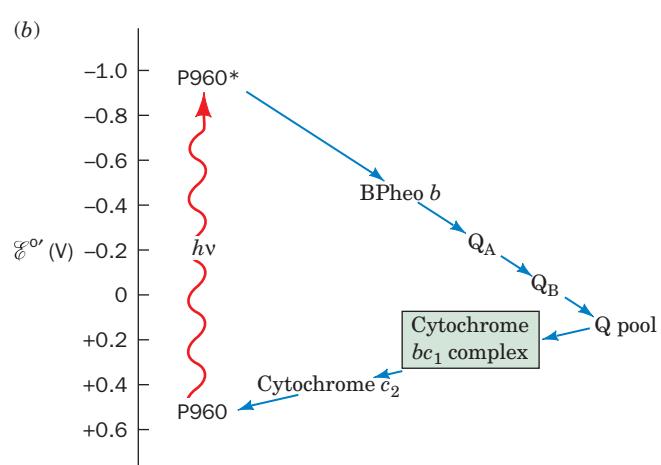


Figure 24-13 Photosynthetic electron-transport system of purple photosynthetic bacteria. (a) Schematic diagram indicating the arrangement of the system components in the bacterial plasma membrane and the flows of electrons (black arrows) and protons (blue arrows) that photon ($h\nu$) absorption promotes through them. The system contains two protein complexes, the RC and cytochrome bc_1 . Two electrons liberated from the special pair, here P870 (as in *Rb. sphaeroides*), by the consecutive absorption of two photons are taken up by ubiquinone (Q_B) together with two protons from the cytoplasm to yield ubiquinol (QH_2). The QH_2 is released from the RC and diffuses (dotted arrows) through the membrane to cytochrome bc_1 , which, in a two-electron reaction, oxidizes it to ubiquinone with the concomitant liberation of its two protons to the external medium. One of the two electrons is passed, via the $[2\text{Fe}-2\text{S}]$ cluster and cytochrome c_1 , to cytochrome c_2 , a peripheral membrane protein that then diffuses across the external surface of the membrane so as to return the electron to P870 of the RC. The second electron from QH_2 passes, via a Q cycle, through hemes b_L and b_H of cytochrome bc_1 and then contributes to the reduction of a molecule of ubiquinone (Q) with the concomitant uptake of two more cytoplasmic protons (two rounds of a Q cycle are required for the reduction of one molecule of Q to QH_2 ; Fig. 22-31). The resulting QH_2 diffuses back to cytochrome bc_1 . There it is again oxidized, with the liberation of its two protons to the exterior and the return of one of its two electrons, via cytochrome c_2 , to P870, thereby completing the electrical circuit. Note that in every turn of a Q cycle, half the electrons liberated by the oxidation of

Various substances, such as H_2S , S , $\text{S}_2\text{O}_3^{2-}$, H_2 , and many organic compounds, function in this capacity depending on the bacterial species.

Modern photosynthetic bacteria are thought to resemble the original photosynthetic organisms. These presumably arose very early in the history of cellular life when environmentally supplied sources of “high-energy” compounds were dwindling but reducing agents were still plentiful (Section 1-5Cb). During this era, photosynthetic bacteria were no doubt the dominant form of life. However, their very success eventually caused them to exhaust the available reductive resources. The ancestors of modern cyanobacteria adapted to this situation by evolving a photosynthetic sys-



QH_2 to Q are used to reduce Q to QH_2 , so that, after a large number of turns, an electron that enters the Q cycle, on average, passes through it twice before being returned to P870. Thus, the net result of the absorption of two photons by the RC is the translocation of four H^+ from the cytoplasm to the external medium. (b) The approximate standard reduction potentials of the photosynthetic electron-transport system’s various components. The overall process is essentially irreversible because electrons are transferred to progressively lower energy states (more positive standard reduction potentials).

tem with sufficient electromotive force to abstract electrons from H_2O . The gradual accumulation of the resulting toxic waste product, O_2 , forced photosynthetic bacteria, which cannot photosynthesize in the presence of O_2 (although some species have evolved the ability to respire), into the narrow ecological niches to which they are presently confined (Section 1-1Ab).

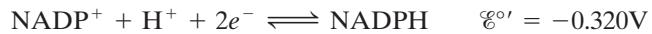
C. Two-Center Electron Transport

See Guided Exploration 22: Two-center photosynthesis (Z-scheme) overview Plants and cyanobacteria use the reducing power generated by the light-driven oxidation of H_2O to produce

NADPH. The component half-reactions of this process, together with their standard reduction potentials, are



and



Hence, the overall four-electron reaction and its standard redox potential is



This latter quantity corresponds (Eq. [16.5]) to a standard free energy change of $\Delta G^{\circ'} = 438 \text{ kJ} \cdot \text{mol}^{-1}$, which Eq. [24.1] indicates is the energy of one einstein of 223-nm photons (UV light). Clearly, *even if photosynthesis were 100% efficient, which it is not, it would require more than one photon of visible light to generate a molecule of O₂. In fact, experimental measurements indicate that algae minimally require 8 to 10 photons of visible light to produce one molecule of O₂.* In the following subsections, we discuss how plants and cyanobacteria manage this multiphoton process.

a. Photosynthetic O₂ Production Requires Two Sequential Photosystems

Two seminal observations led to the elucidation of the basic mechanism of photosynthesis in plants:

1. The quantum yield for O₂ evolution by *Chlorella pyrenoidosa* varies little with the wavelength of the illuminating light between 400 and 675 nm, but decreases precipitously above 680 nm (Fig. 24-14, *lower curve*). This phenomenon, the “red drop,” was unexpected because Chl *a* absorbs such far-red light (Fig. 24-5).

2. Shorter wavelength light, such as yellow-green light, enhances the photosynthetic efficiency of 700-nm light well in excess of the energy content of the shorter wavelength light; that is, *the rate of O₂ evolution by both lights is greater than the sum of the rates for each light acting alone* (Fig. 24-14, *upper curve*). Moreover, this enhancement still occurs if the yellow-green light is switched off several seconds before the red light is turned on and vice versa.

These observations clearly indicate that two processes are involved. They are explained by a mechanistic model, the **Z-scheme**, which postulates that O₂-producing photosynthesis occurs through the actions of two photosynthetic RCs that are connected essentially in series (Fig. 24-15).

1. **Photosystem I (PSI)** generates a strong reductant capable of reducing NADP⁺ and, concomitantly, a weak oxidant.

2. **Photosystem II (PSII)** generates a strong oxidant capable of oxidizing H₂O and, concomitantly, a weak reductant.

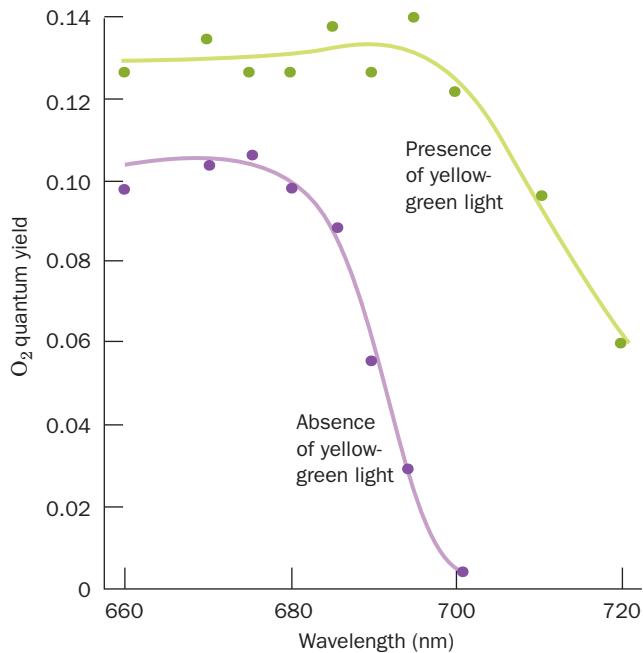


Figure 24-14 Quantum yield for O₂ production by *Chlorella* algae as a function of the wavelength of the incident light. The experiment was conducted in the absence (lower curve) and the presence (upper curve) of supplementary yellow-green light. The upper curve has been corrected for the amount of O₂ production stimulated by the supplementary light alone. Note that the lower curve falls off precipitously above 680 nm (the red drop). However, the supplementary light greatly increases the quantum yield in the wavelength range above 680 nm (far-red) in which the algae absorb light. [After Emerson, R., Chalmers, R., and Cederstrand, C., *Proc. Natl. Acad. Sci.* **49**, 137 (1957).]

The weak reductant reduces the weak oxidant, so that *PSI* and *PSII* form a two-stage electron “energizer.” Both photosystems must therefore function for photosynthesis (electron transfer from H₂O to NADP⁺, forming O₂ and NADPH) to occur.

The red drop is explained in terms of the Z-scheme by the observation that PSII is only poorly activated by 680-nm light. In the presence of only this far-red light, PSI is activated but is unable to obtain more than a few of the electrons it is capable of energizing. Yellow-green light, however, efficiently stimulates PSII to supply these electrons. The observation that the far-red and yellow-green lights can be alternated indicates that both photosystems remain activated for a time after the light is switched off.

The validity of the Z-scheme was established as follows. The oxidation state of **cytochrome f**, a *c*-type cytochrome of the electron-transport chain connecting PSI and PSII (see below), can be spectroscopically monitored. Illumination of algae with 680-nm (far-red) light results in the oxidation of cytochrome *f* (Fig. 24-16). However, the additional imposition of a 562-nm (yellow-green) light results in this

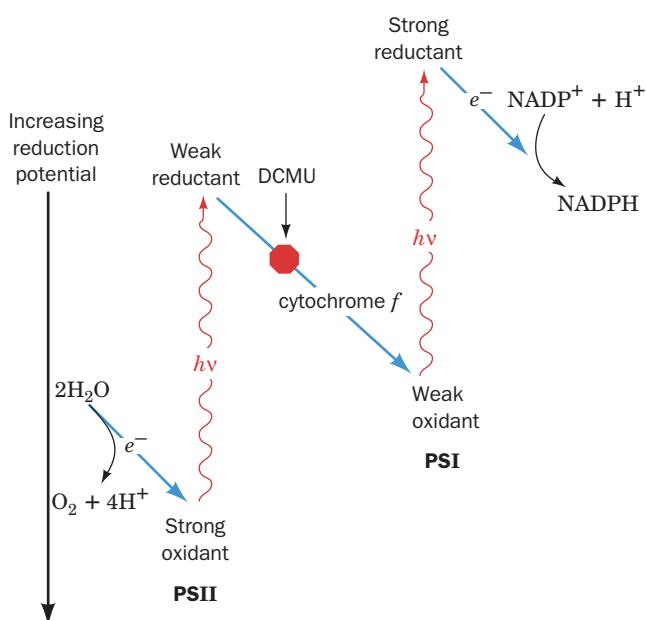


Figure 24-15 The Z-scheme for photosynthesis in plants and cyanobacteria. Two photosystems, PSI and PSII, function to drive electrons from H_2O to NADPH. The reduction potential increases downward so that electron flow occurs spontaneously in this direction. The herbicide DCMU (see text) blocks photosynthetic electron transport from PSII to cytochrome *f*.

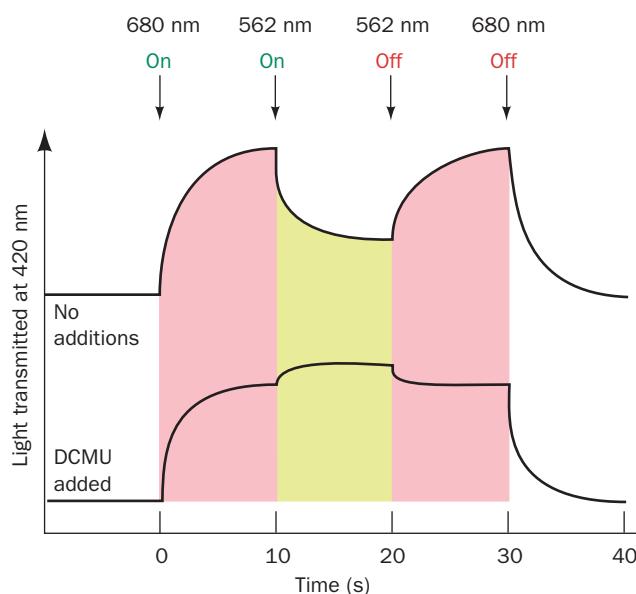
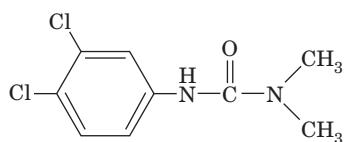


Figure 24-16 The oxidation state of cytochrome *f* in *Porphyridium cruentum* algae as monitored by a weak beam of 420-nm (blue-violet) light. An increase in the transmitted light signals the oxidation of cytochrome *f*. In the upper curve, strong light at 680 nm (far-red) causes the oxidation of the cytochrome *f* but the superposition of 562-nm (yellow-green) light causes its partial re-reduction. In the lower curve, the presence of the herbicide DCMU, which inhibits photosynthetic electron transport, causes 562-nm light to further oxidize, rather than reduce, the cytochrome *f*.

protein's partial re-reduction. In the presence of the herbicide **3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU)**,



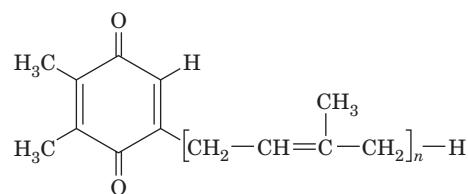
3-(3,4-Dichlorophenyl)-1,1-dimethylurea (DCMU)

which abolishes photosynthetic oxygen production, 680-nm light still oxidizes cytochrome *f* but simultaneous 562-nm light only oxidizes it further. The explanation for these effects is that 680-nm light, which efficiently activates only PSI, causes it to withdraw electrons from (oxidize) cytochrome *f*. The 562-nm light also activates PSII, which thereby transfers electrons to (reduces) cytochrome *f*. DCMU blocks electron flow from PSII to cytochrome *f* (Fig. 24-15), so an increased intensity of light, whatever its wavelength, only serves to activate PSI further.

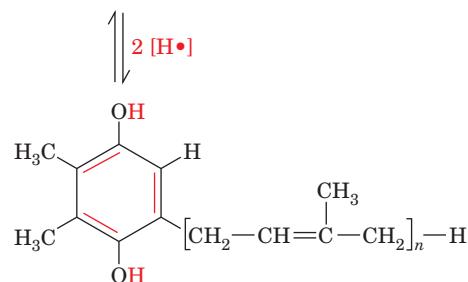
b. O_2 -Producing Photosynthesis Is Mediated by Three Transmembrane Protein Complexes Linked by Mobile Electron Carriers

The components of the Z-scheme, which mediate electron transport from H_2O to NADPH, are largely organized into three thylakoid membrane-bound particles (Fig. 24-17): (1)

PSII, (2) the cytochrome *b*₆*f* complex, and (3) PSI. As in oxidative phosphorylation, electrons are transferred between these complexes via mobile electron carriers. The ubiquinone analog **plastoquinone (Q)**, via its reduction to **plastoquinol (QH₂)**,



Plastoquinone



Plastoquinol

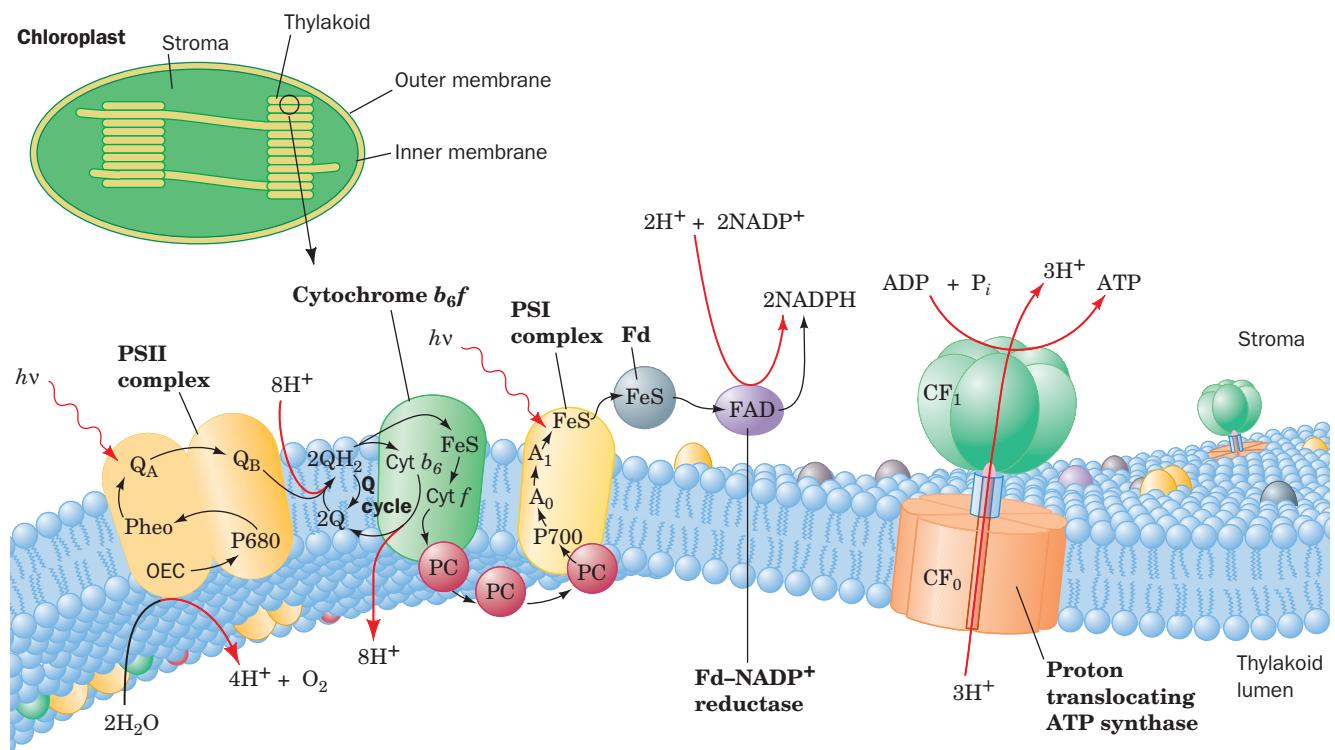


Figure 24-17 Schematic representation of the thylakoid membrane showing the components of its electron-transport chain. The system consists of three protein complexes: PSII, the cytochrome b_6f complex, and PSI, which are electrically “connected” by the diffusion of the electron carriers plastoquinol (Q) and plastocyanin (PC). Light-driven transport of electrons (black arrows) from H_2O to $NADP^+$ forming $NADPH$ motivates the transport of protons (red arrows) into the thylakoid space (Fd is ferredoxin). Additional protons are split off from water by

the oxygen-evolving complex (OEC), yielding O_2 . The resulting proton gradient powers the synthesis of ATP by the CF_1CF_0 proton-translocating ATP synthase [CF_1 and CF_0 are chloroplast (C) analogs of mitochondrial F_1 and F_0]. The membrane also contains light-harvesting complexes whose component chlorophylls and other chromophores transfer their excitations to PSI and PSII. [After Ort, D.R. and Good, N.E., *Trends Biochem. Sci.* **13**, 469 (1988).]

links PSII to the cytochrome b_6f complex, which, in turn, interacts with PSI through the mobile Cu-containing redox protein **plastocyanin (PC)**. In what follows, we trace the electron pathway through this chloroplast system from H_2O to $NADP^+$ (Fig. 24-18).

c. PSII Resembles the PbRC

PSII from the thermophilic cyanobacterium *Thermosynechococcus elongatus* consists of 20 subunits, 14 of which occupy the photosynthetic membrane. These transmembrane subunits include the reaction center proteins **D1 (PsbA)** and **D2 (PsbD)**, the chlorophyll-containing inner-antenna subunits **CP43 (PsbC)** and **CP47 (PsbB)**, and **cytochrome b_{559}** . The X-ray structure of this PSII (Fig. 24-19), independently determined by James Barber and So Iwata and by Wolfram Saenger, reveals that this ~340-kD protein is a symmetric dimer, whose protomeric units each contain 35 transmembrane helices. Each protomer, which has pseudo-2-fold symmetry, binds 36 Chl a ’s, 2 **pheophytin a’s (Pheo a’s; Chl a with its Mg^{2+} replaced by two protons)**, one heme b , one heme c , 2 plastochinones, one nonheme

Fe, 12 all-trans carotenoids presumed to be β -carotene, one HCO_3^- ion, and one Mn_4CaO_4 complex known as the **oxygen-evolving center (OEC; alternatively, the water-oxidizing complex (WOC))**. In higher plants, the PSII protomer contains ~25 subunits and forms an ~1000-kD transmembrane supercomplex with several antenna proteins. The arrangement of the 5 transmembrane helices in both D1 and D2 resembles that in the L and M subunits of the PbRC (Fig. 24-11). Indeed, these two sets of subunits have similar sequences, thereby indicating that they arose from a common ancestor.

The cofactors of PSII’s RC (Fig. 24-20) are organized similarly to those of the bacterial system (Fig. 24-12): They have essentially the same components (with Chl a , Pheo a , and plastochinone replacing BChl b , BPheo b , and menaquinone, respectively) and are symmetrically organized along the complex’s pseudo-2-fold axis. The two Chl a rings labeled P_{D1} and P_{D2} in Fig. 24-20 are positioned analogously to the BChl b ’s of P960’s “special pair” and are therefore presumed to form PSII’s primary electron donor, **P680** (named after the wavelength at which its absorbance

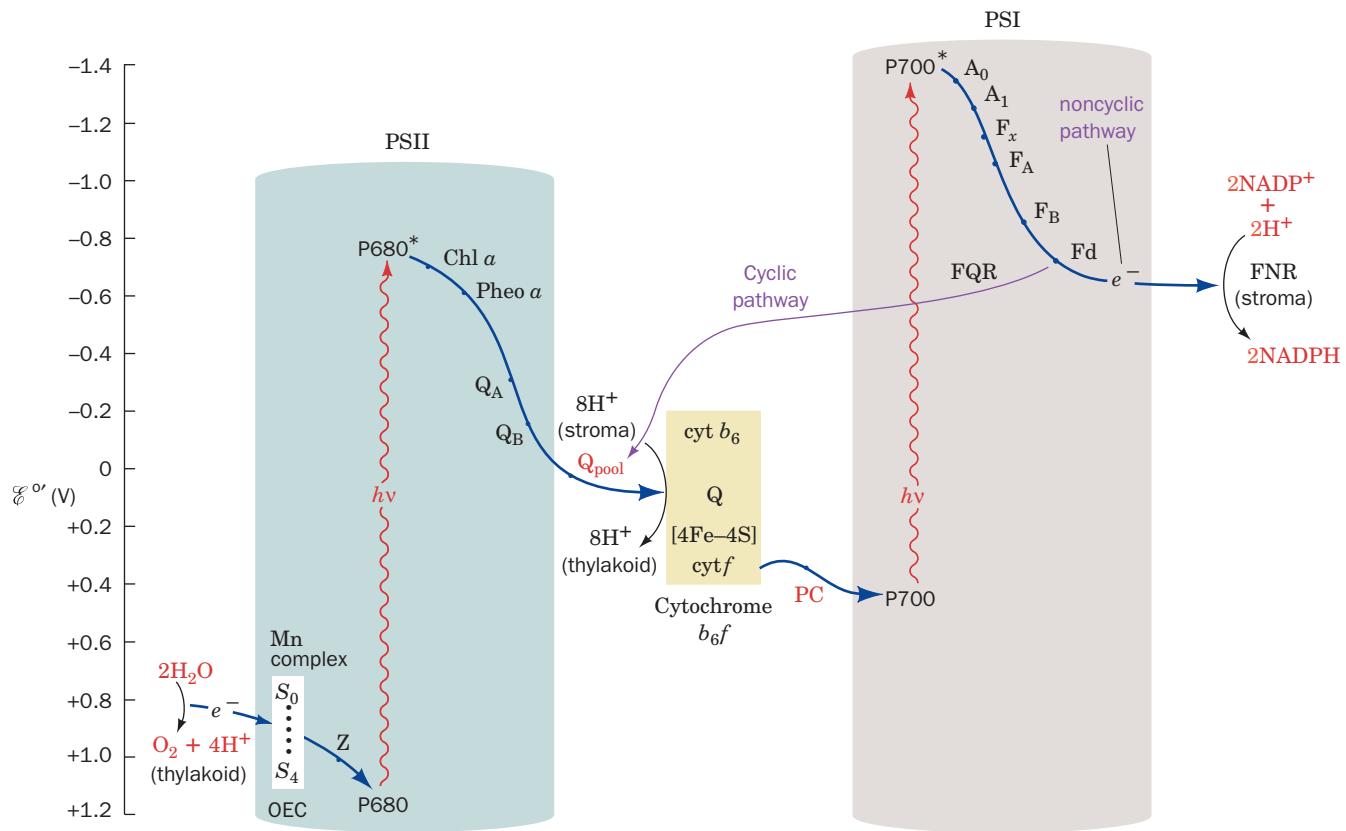


Figure 24-18 Detailed diagram of the Z-scheme of photosynthesis. Electrons ejected from P680 by the absorption of photons are replaced with electrons abstracted from H₂O by an Mn complex (OEC), thereby forming O₂ and four H⁺. Each ejected electron is passed through a chain of electron carriers to a pool of plastoquinone molecules (Q). The resulting plastoquinol, in turn, reduces the cytochrome *b*₆*f* particle (yellow box) that transfers electrons with the concomitant translocation of protons, via a Q cycle, into the thylakoid lumen. Cytochrome

*b*₆*f* then transfers the electrons to plastocyanin (PC). The plastocyanin regenerates photooxidized P700. The electron ejected from P700, through the intermediacy of a chain of electron carriers (A₀, A₁, F_x, F_A, F_B, and Fd), reduces NADP⁺ to NADPH in noncyclic electron transport. Alternatively, the electron may be returned to the cytochrome *b*₆*f* complex in a cyclic process that only translocates protons into the thylakoid lumen.

maximally decreases on photooxidation). The electron ejected from P680 follows a similar asymmetric course as that in the PbRC even though the two systems operate over different ranges of reduction potential (compare Figs. 24-13b and 24-18). As indicated in the central part of Fig. 24-18, the electron is transferred to a molecule of Pheo *a* (Pheo_{D1} in Fig. 24-20), probably via a Chl *a* molecule (Chl_{D1}), and then to a bound plastoquinone (Q_A). The electron is subsequently transferred to a second plastoquinone molecule, Q_B, which after it receives a second electron in a like manner, takes up two protons at the stromal (cytosolic in cyanobacteria) surface of the thylakoid membrane. The resulting plastoquinol, Q_BH₂, then exchanges with a membrane-bound pool of plastoquinone molecules. DCMU as well as many other commonly used herbicides compete with plastoquinone for the Q_B-binding site on PSII, which explains how they inhibit photosynthesis.

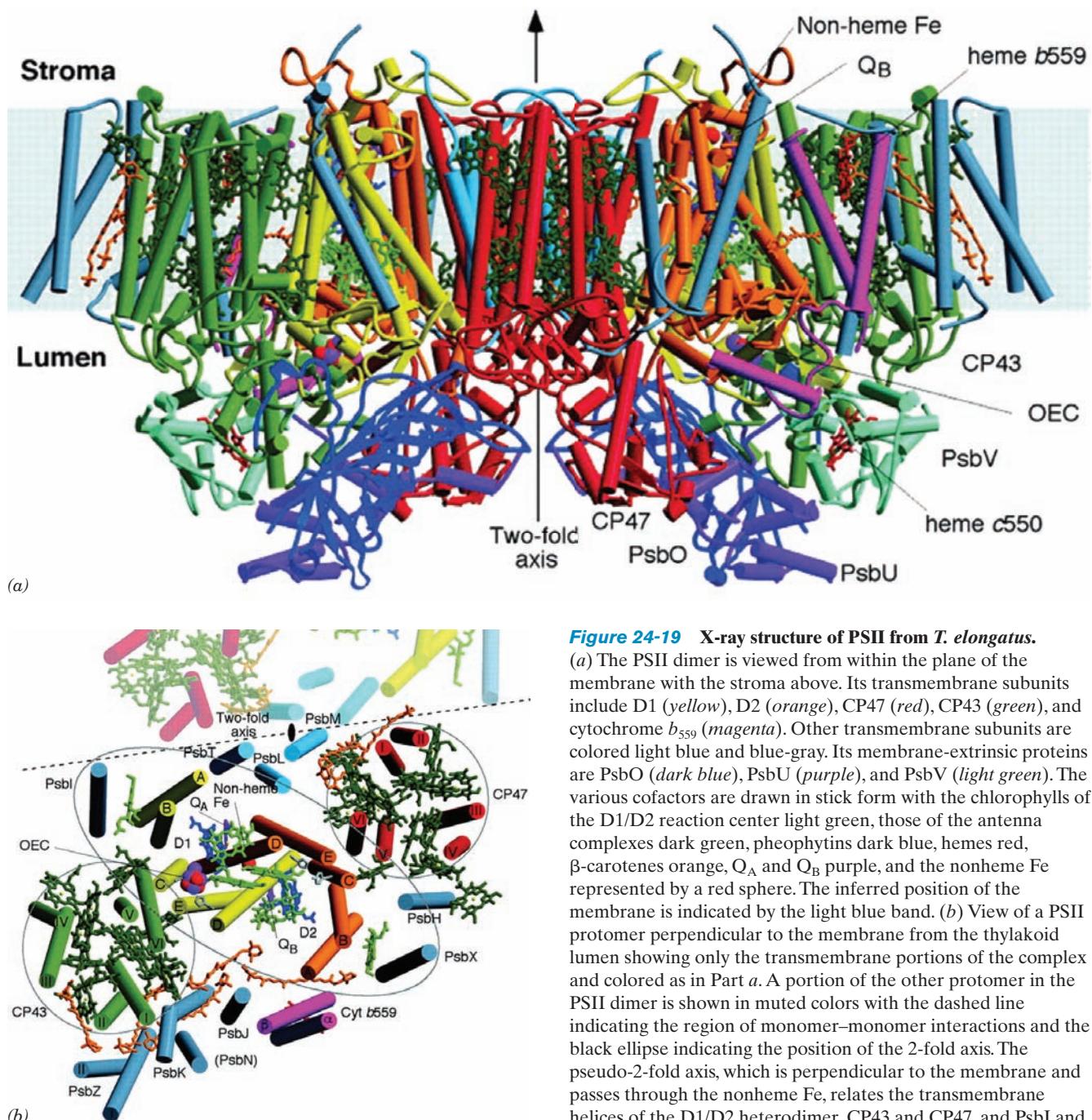
Two “extra” Chl *a* molecules, Chl_{D1} and Chl_{D2}, lie on the periphery of the RC, where they are postulated to func-

tion in the transfer of excitation from the antenna systems to P680. Cytochrome *b*₅₅₉, whose function is unclear, breaks the pseudosymmetry of the PSII protomer as does the Mn cluster, whose function we now discuss.

d. O₂ Is Generated in a Five-Stage Water-Splitting Reaction Mediated by an Mn-Containing Protein Complex

The oxidation by the OEC of two molecules of H₂O to form one molecule of O₂ requires four electrons. Since transfer of a single electron from H₂O to NADP⁺ requires two photochemical events, this accounts for the observed minimum of 8 to 10 photons absorbed per molecule of O₂ produced.

Must the four electrons necessary to produce a given O₂ molecule be removed by a single photosystem or can they be extracted by several different photosystems? Pierre Joliet and Bessel Kok answered this question by analyzing the rate at which dark-adapted chloroplasts produce O₂



when exposed to a series of short flashes. O_2 was evolved with a peculiar oscillatory pattern (Fig. 24-21). There is virtually no O_2 evolved by the first two flashes. The third flash results in the maximal O_2 yield. Thereafter, the amount of O_2 produced peaks with every fourth flash until the oscillations damp out to a steady state. This periodicity indicates that each OEC cycles through five different states, S_0 through S_4 (Fig. 24-22). Each of the transitions between S_0 and S_4 is a photon-driven redox reaction; that from S_4 to S_0

results in the release of O_2 . Thus, *each O_2 molecule must be produced by a single photosystem*. The observation that O_2 evolution peaks at the third rather than the fourth flash indicates that the OEC's resting state is predominantly S_1 rather than S_0 . The oscillations gradually damp out because a small fraction of the RCs fail to be excited or become doubly excited by a given flash of light, so that they eventually lose synchrony. The five reaction steps release a total of four water-derived protons into the inner thylakoid space

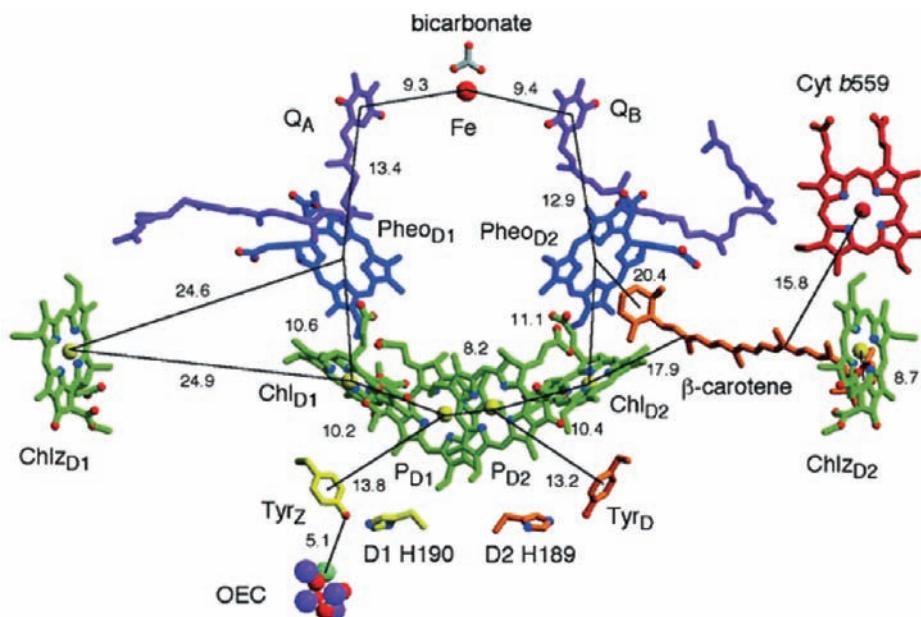


Figure 24-20 The arrangement of electron-transfer cofactors in PSII from *T. elongatus*. The complex is viewed along the membrane plane with the thylakoid lumen below. The cofactors are colored as in Fig. 24-19 but with Mg^{2+} yellow, N blue, and O red. The phytol tails of the chlorophylls and pheophytins have been removed for clarity. The side chain C atoms of Tyr_Z (D1 Tyr 161) and D1 His 190 are yellow, and those of Tyr_D (D2 Tyr 160)

and D2 His 189 are orange. The OEC is drawn in space-filling form with Mn purple, Ca^{2+} cyan, and O red. The numbers indicate the center-to-center distances, in angstroms, between the cofactors spanned by the accompanying thin black lines. Compare this figure to Fig. 24-12 (which is drawn upside down relative to this figure). [Courtesy of James Barber and So Iwata, Imperial College London, U.K. PDBid 1S5L.]

(lumen) in a stepwise manner (Fig. 24-22). These protons contribute to the transmembrane proton gradient.

Since the OEC abstracts electrons from H_2O , its five states must have extraordinarily high reduction potentials (recall from Table 22-1 that the O_2/H_2O half-reaction has a standard reduction potential of 0.815 V). PSII must also stabilize the highly reactive intermediates for extended periods (as much as minutes) in close proximity to water.

The OEC, which is located at the luminal surface of the D1 subunit (Fig. 24-20), is a Mn_4CaO_4 or Mn_4CaO_5 complex in which the O atoms bridge neighboring Mn atoms. The structure of the OEC remains elusive due to PSII's relatively poorly resolved X-ray structures and the observation

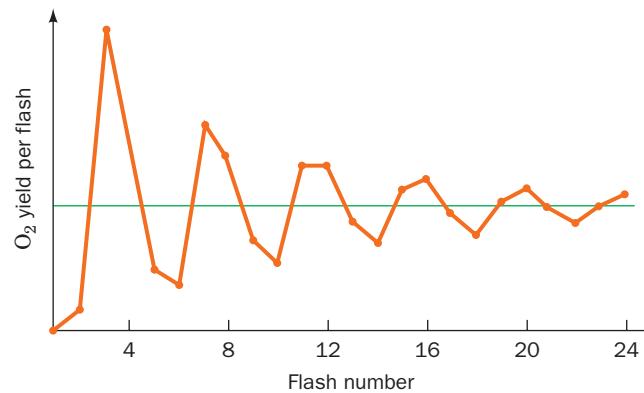


Figure 24-21 The O_2 yield per flash in dark-adapted spinach chloroplasts. Note that the yield peaks on the third flash and then on every fourth flash thereafter until the curve eventually damps out to its average value. [After Forbush, B., Kok, B., and McGloin, M.P., *Photochem. Photobiol.* **14**, 309 (1971).]

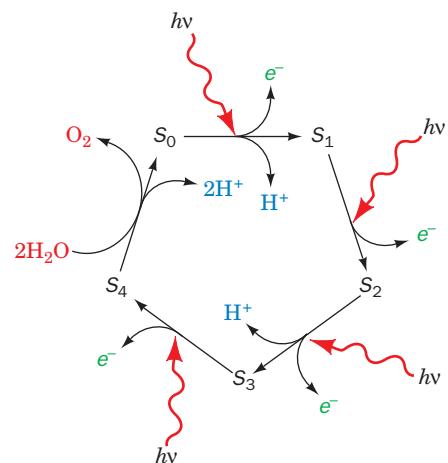


Figure 24-22 Schematic mechanism of O_2 generation in chloroplasts. Four electrons are stripped, one at a time in light-driven reactions ($S_0 \rightarrow S_4$), from two bound H_2O molecules. In the recovery step ($S_4 \rightarrow S_0$), which is light independent, O_2 is released and two more H_2O molecules are bound. Three of these five steps release protons into the thylakoid lumen.

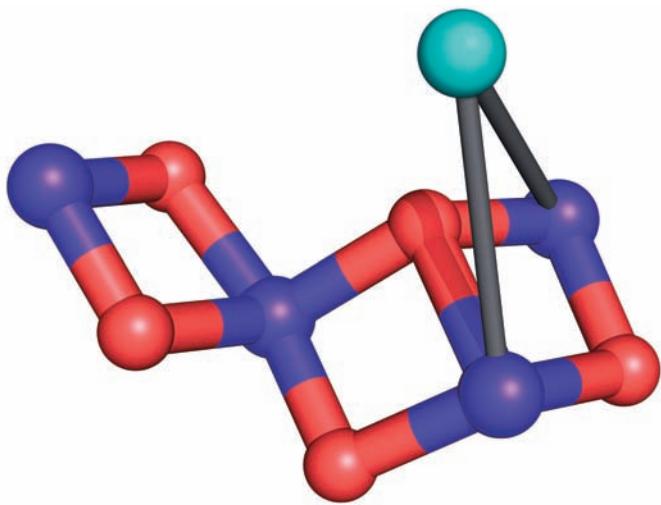


Figure 24-23 A model of the OEC. This Mn_4CaO_5 complex is shown in ball-and-stick form with Mn ions purple, the Ca^{2+} ion cyan, and O red. The bonds between the Ca and Mn ions are drawn in gray to indicate that the position of the Ca ion is relatively poorly defined. Presumably, numerous protein side chains and water molecules ligand the Ca and Mn ions. Several related models are also compatible with the structural data. [Based on a model by Vittal Yachandra, Lawrence Berkeley National Laboratory, Berkeley, California.]

that the OEC decomposes when illuminated with X-rays at the intensities used in X-ray structure determinations. However, the use of X-ray spectroscopy techniques of lower intensity that are sensitive to bond lengths have led to the formulation of several related models for the OEC that are compatible with the X-ray structure of PSII. One of these models is shown in Fig. 24-23.

The water-splitting reaction is driven by the excitation of the PSII RC. A variety of evidence indicates that the Mn ions in the OEC's various *S* states (Fig. 24-22) cycle through specific combinations of Mn(II), Mn(III), Mn(IV), and Mn(V) while abstracting protons and electrons from two H_2O molecules to yield O_2 , which is released into the thylakoid lumen. However, the mechanism whereby this occurs, that is, the nature of the five *S* states, remains unknown due to the lack of structural information concerning these states.

The next link in the PSII electron transport chain is an entity, originally named Z (Fig. 24-18), which relays electrons from the OEC to P680. The existence of Z is signaled by a transient EPR spectrum of illuminated chloroplasts that parallels the *S*-state transitions. The change in this spectrum on supplying deuterated tyrosine to cyanobacteria indicates that Z^+ is a tyrosyl radical ($TyrO\cdot$; EPR spectra reflect the nuclear spins of the atoms with which the unpaired electrons interact). It has been identified as Tyr_Z in PSII (Fig. 24-20) due to its position between the Mn cluster and P680's chlorophyll P_{D1} . Recall that a tyrosyl radical has also been implicated in the reduction of O_2 to $2 H_2O$ by cytochrome *c* oxidase (Complex IV) in the respiratory electron-transport chain (Section 22-2C5c).

e. Electron Transport through the Cytochrome b_6f Complex Generates a Proton Gradient

From the plastoquinone pool, electrons pass through the cytochrome b_6f complex. This integral membrane assembly resembles cytochrome bc_1 , its purple bacterial counterpart (Section 24-2Be), as well as Complex III of the mitochondrial electron-transport chain (also called cytochrome bc_1 ; Section 22-2C3a). Electron flow through the cytochrome b_6f complex occurs through a Q cycle (Fig. 22-31) in which plastoquinone is the $(H^+ + e^-)$ carrier. Accordingly, two protons are translocated across the thylakoid membrane for every electron transported. The four electrons abstracted from $2 H_2O$ by the OEC therefore lead to the translocation of eight H^+ from the stroma to the thylakoid lumen. *Electron transport via the cytochrome b_6f complex generates much of the electrochemical proton gradient that drives the synthesis of ATP in chloroplasts.*

The X-ray structure of cytochrome b_6f (Fig. 24-24) was independently determined by Janet Smith and William Cramer and by Jean-Luc Popot and Daniel Picot. Cytochrome b_6f is a dimer of ~ 109 -kD protomers, each containing four large subunits (18–32 kD) that have counterparts in cytochrome bc_1 : **cytochrome b_6** , a homolog of the N-terminal half of cytochrome *b*; **subunit IV**, a homolog of the C-terminal half of cytochrome *b*; a Rieske iron–sulfur protein (ISP), which is also present in cytochrome bc_1 ; and **cytochrome f** (*f* for *feuille*, French for leaf), a *c*-type cytochrome that is a functional analog of cytochrome c_1 , although the two are unrelated in structure or sequence. In fact, cytochrome *f* is an elongated, two-domain protein that is dominated by β sheets and hence has an entirely different fold from those of other *c*-type cytochromes of known structure. Cytochrome *f*'s single heme *c* is, nevertheless, covalently linked to the protein's larger domain via the two Cys residues in a Cys-X-Y-Cys-His sequence that is characteristic of *c*-type cytochromes (Fig. 9-41) and whose His residue forms one of the Fe ion's two axial ligands (Fig. 22-21). Intriguingly, however, the second axial ligand is not a Met S atom, as occurs in most *c*-type cytochromes, but instead is the protein's N-terminal amino group, a group that had previously not been observed to be a heme ligand.

In addition, cytochrome b_6f has four small hydrophobic subunits that have no equivalents in cytochrome bc_1 . Each protomer contains 13 transmembrane helices, four in cytochrome b_6 , three in subunit IV, and one each in the remaining subunits. Cytochrome b_6f binds cofactors that are the equivalents of all of those in cytochrome bc_1 : **heme f** , a *c*-type heme bound by cytochrome *f*; a [2Fe–2S] cluster bound by the ISP; hemes b_H and b_L ; a plastoquinone molecule that occupies either the Q_i site (the quinone-binding site at which fully reduced quinone is regenerated during the Q cycle; Section 22-3Be) or the Q_o site. In addition, cytochrome b_6f binds several cofactors that have no counterparts in cytochrome bc_1 : a Chl *a*, a β -carotene, and, unexpectedly, a novel heme named **heme x** (alternatively, **heme c_i**), which is covalently linked to the protein via a single thioether bond to Cys 35 of cytochrome b_6 , and whose only axial ligand is a water molecule (compare with hemes *a*, *b*, and *c*; Fig. 22-21).

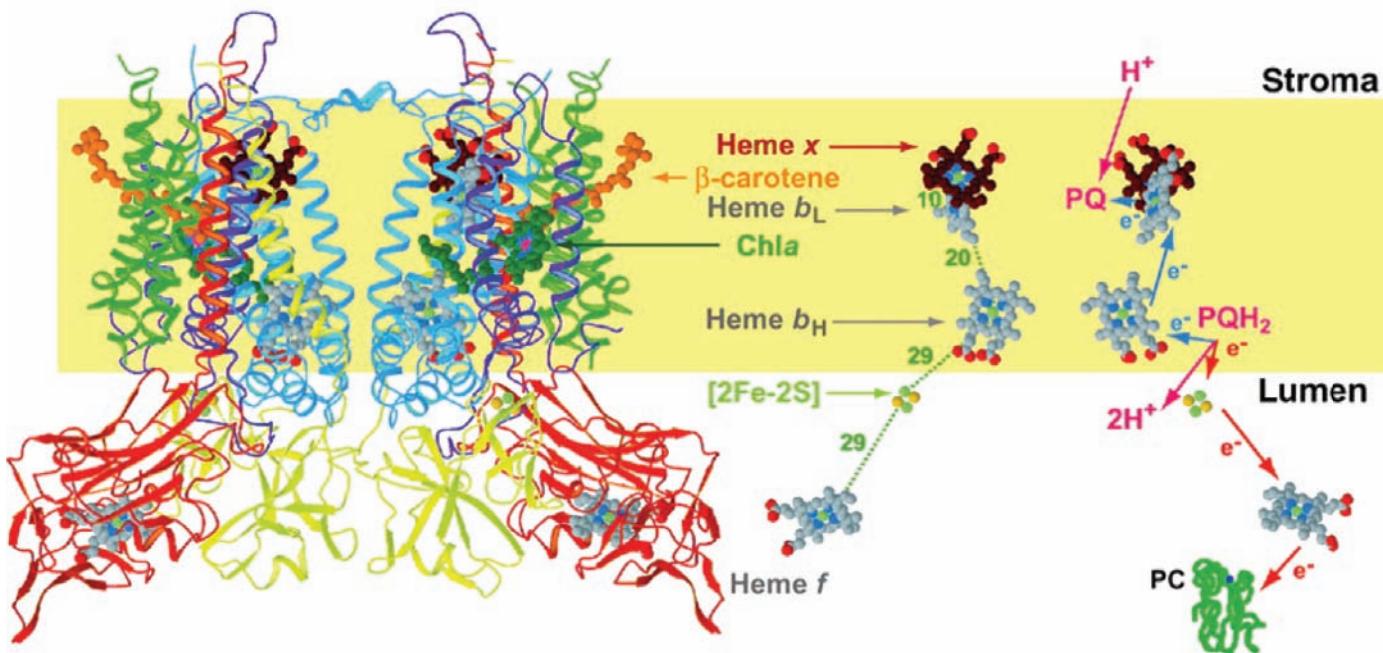


Figure 24-24 X-ray structure of the cytochrome b_6f complex from the thermophilic cyanobacterium *Mastigocladus laminosus*. A ribbon diagram of the dimeric complex is drawn on the left with cytochrome b_6 blue, subunit IV purple, cytochrome f red, the iron-sulfur protein (ISP) yellow, and the other subunits green. The inferred position of the lipid bilayer is

indicated by a yellow band. Compare this figure to Fig. 22-23 (which is upside down relative to this figure). The paths of electron and proton transfer through the complex and the distances, in angstroms, between redox centers are shown on the right. [Modified from a drawing by William A. Cramer and Janet Smith, Purdue University. PDBid 1UM3.]

f. Plastocyanin Transports Electrons from Cytochrome b_6f to PSI

Electron transfer between cytochrome f , the terminal electron carrier of the cytochrome b_6f complex, and PSI is mediated by **plastocyanin (PC)**, a 99-residue, monomeric, Cu-containing, peripheral membrane protein located on the thylakoid luminal surface (Fig. 24-17). Thus PC is the functional analog of cytochrome c , which transfers electrons from Complex III to Complex IV in the mitochondrial electron-transport chain (Section 22-2C4).

PC's redox center cycles between its Cu(I) and Cu(II) oxidation states. The X-ray structure of PC from poplar leaves, determined by Hans Freeman, shows that its single

Cu atom is coordinated with distorted tetrahedral geometry by a Cys, a Met, and two His residues (Fig. 24-25). Cu(II) complexes with four ligands normally adopt a square planar coordination geometry, whereas those of Cu(I) are generally tetrahedral. Evidently, the strain of Cu(II)'s protein-imposed tetrahedral coordination in PC promotes its

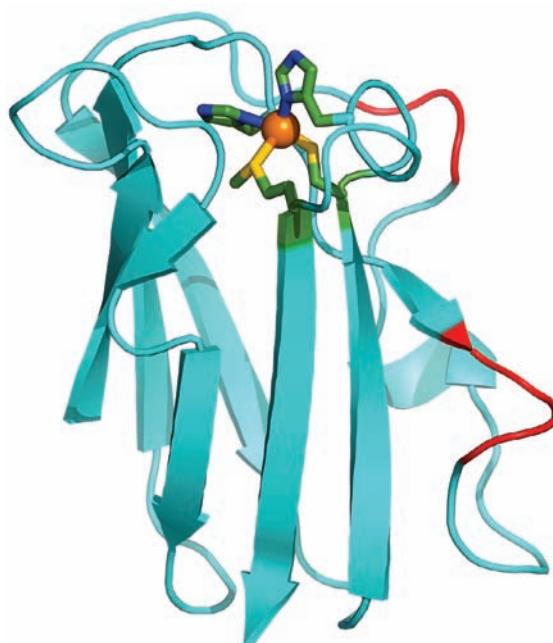


Figure 24-25 X-ray structure of plastocyanin (PC) from poplar leaves. This 99-residue monomeric protein, a member of the family of **blue copper proteins** (as is the globular domain of Complex IV's Subunit II, which binds the Cu_A center; Section 22-2C5a), folds into a β sandwich. Its Cu atom (orange sphere), which alternates between its Cu(I) and Cu(II) oxidation states, is tetrahedrally coordinated by the side chains of His 37, Cys 84, His 87, and Met 92, which are shown in stick form with their C, N, and S atoms green, blue, and yellow. Seven conserved Asp and Glu residues (red) form a negatively charged patch on the surface of PC that has been implicated in electrostatically binding to a positively charged patch on the surface of cytochrome f formed by five Lys and Arg residues. [Based on an X-ray structure by Mitchell Guss and Hans Freeman, University of Sydney, Australia. PDBid 1PLC.]

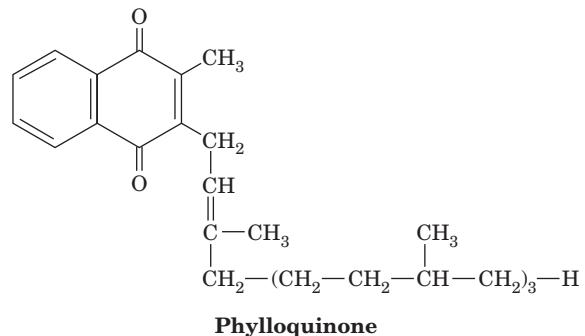
reduction to Cu(I). This hypothesis accounts for PC's high standard reduction potential (0.370 V) compared to that of the normal Cu(II)/Cu(I) half-reaction (0.158 V). This is an example of how proteins modulate the reduction potentials of their redox centers so as to match them to their function—in the case of plastocyanin, the efficient transfer of electrons from the cytochrome b_6f complex to PSI.

g. PSI Resembles Both PSII and the PbRC

Cyanobacterial PSIs are trimers of protomers that each consist of at least 11 different protein subunits coordinating >100 cofactors. The X-ray structure of PSI from *T. elongatus* (Fig. 24-26), determined at 2.5-Å resolution by Norbert Krauss, Saenger, and Petra Fromme, reveals that each of its 356-kD protomers contains nine transmembrane subunits (**PsaA**, **PsbB**, **PsaF**, **PsaI–M**, and **PsaX**) and three stromal (cytoplasmic) subunits (**PsaC–E**), which collectively bind 127 cofactors that comprise 30% of PSI's mass. The cofactors forming the PSI RC are all bound by the homologous subunits PsaA (755 residues) and PsaB (740 residues), whose 11 transmembrane helices each are arranged in a manner resembling those in the L and M subunits of the PbRC (Fig. 24-11) and the D1 and D2 subunits of PSII (Fig. 24-19), thus supporting the hypothesis that all

RCs arose from a common ancestor. PsaA and PsaB, together with other transmembrane subunits, also bind the cofactors of the core antenna system (see below).

Figure 24-27 indicates that PSI's RC consists of six molecules of chlorophyll and two molecules of **phylloquinone (vitamin K₁)**; note that it has the same phytol side chain as chlorophylls (Fig. 24-3),



all arranged in two pseudosymmetrically related branches, followed by three [4Fe–4S] clusters. The primary electron donor of this system, **P700**, consists of a Chl *a*' (Fig. 24-3) and a Chl *a* (A1 and B1, respectively), whose rings are

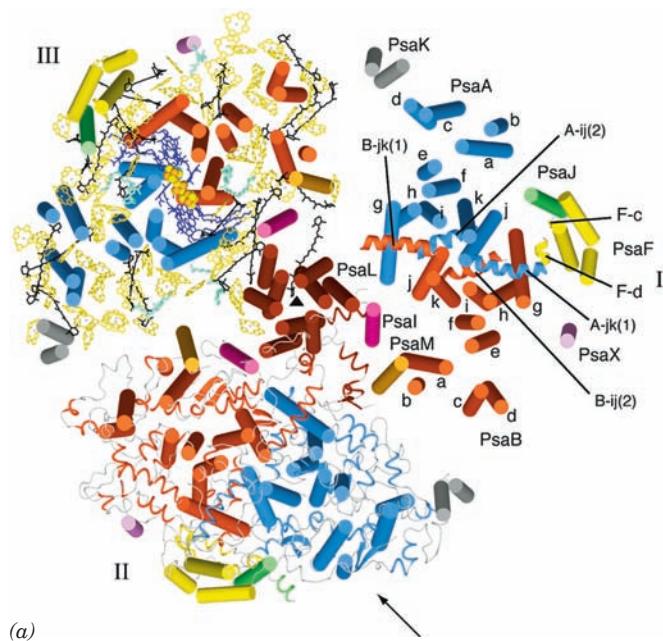
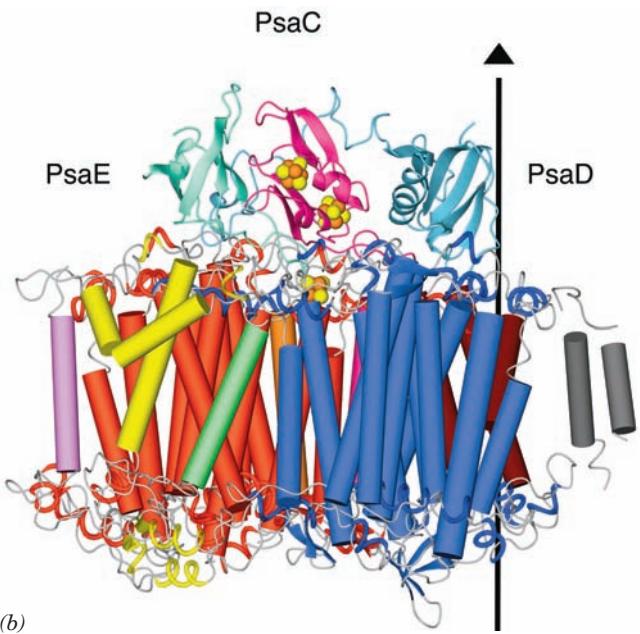


Figure 24-26 X-ray structure of PSI from *T. elongatus*.

(a) View of the trimer perpendicular to the membrane from its stromal side. The stromal subunits have been removed for clarity. PSI's 3-fold axis of symmetry is represented by the small black triangle. Different structural elements are shown for each of the three protomers (I, II, and III). I shows the arrangement of transmembrane helices (cylinders), which are differently colored for each subunit. The transmembrane helices of both PsaA (blue) and PsaB (red) are named *a* through *k* from their N- to C-termini. The six helices in extramembranous loop regions are drawn as spirals. II shows the transmembrane helices as cylinders with the stromal and luminal loop regions drawn in ribbon form. III shows the transmembrane helices as cylinders together with



all cofactors. The RC Chl *a*'s and quinones, drawn in stick form, are purple, the Fe and S atoms of the [4Fe–4S] clusters are drawn as orange and yellow spheres, the antenna system Chl *a*'s (whose side chains have been removed for clarity) are yellow, the carotenoids are black, and the bound lipids are light green.

(b) One protomer as viewed parallel to the membrane along the arrow in Part *a* with the stroma above. The transmembrane subunits are colored as in Part *a* with the stromal subunits PsaC, PsaD, and PsaE pink, cyan, and light green. The vertical line and triangle mark the trimer's 3-fold axis of symmetry. [Courtesy of Wolfram Saenger, Freie Universität Berlin, Germany. PDBID 1JB0.]

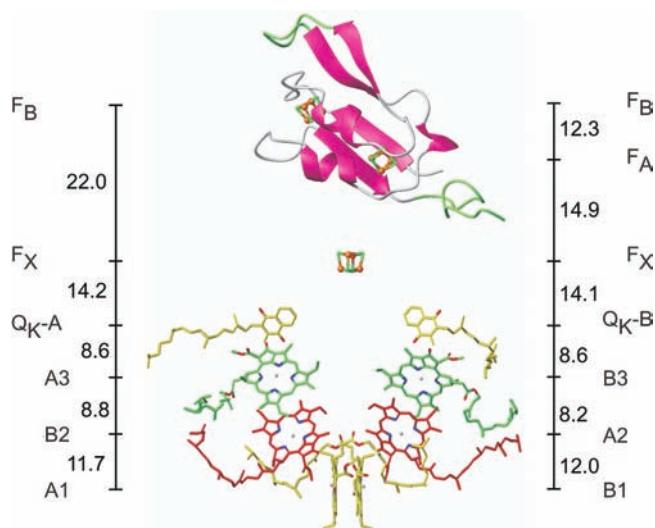


Figure 24-27 Cofactors of the PSI RC and PsaC. The structure is viewed parallel to the membrane plane with the stroma above. The Chl *a* and phylloquinone molecules are arranged in two branches that are related by PSI's 2-fold axis of pseudosymmetry, which is vertical in this drawing. The Chl *a*'s are labeled A or B to indicate that their Mg²⁺ ions are liganded by the side chains of PsaA or PsaB, respectively, and, from the luminal side upward, by different colors and numbers, 1 to 3. The phylloquinones are named Q_K-A and Q_K-B. PsaC is shown in ribbon form with those portions resembling segments in bacterial 2[4Fe-4S] ferredoxins pink and with insertions and extensions green. The three [4Fe-4S] clusters are shown in ball-and-stick form and labeled according to their spectroscopic identities F_X, F_A, and F_B. The center-to-center distances between cofactors (vertical black lines) are given in angstroms. Compare this figure with Figs. 24-20 and 24-12. [Courtesy of Wolfram Saenger, Freie Universität Berlin, Germany. PDBid 1JB0.]

parallel and whose Mg²⁺ ions are separated by 6.3 Å. Thus P700 resembles the special pair in the PbRC. However, EPR studies indicate that ~80% of the unpaired electron associated with photooxidized P700⁺ resides on Chl *a* B1. A1 is followed in the left branch of Fig. 24-27 by two more Chl *a* rings, B2 and A3, and B1 is followed by A2 and B3 in the right branch. One or both of the third pair of Chl *a* molecules, A3 and B3, probably form the spectroscopically identified primary electron acceptor A₀ (right side of Fig. 24-18). The Mg²⁺ ions of A3 and B3 are each axially liganded by the S atom of a Met residue rather than by a His side chain (thereby forming the only known biological examples of Mg²⁺—S coordination). All of the residues involved in Mg²⁺ coordination and hydrogen bonding to these second and third Chl *a*'s are strictly conserved in PSI's, from cyanobacteria to higher plants, thereby suggesting that all of these interactions are important for fine-tuning their redox potentials. Electrons are passed from A3 and B3 to the phylloquinones, Q_K-A and Q_K-B, which almost certainly correspond to the spectroscopically identified electron acceptor A₁. Spectroscopically based kinetic investigations indicate that, in contrast to the case for the PbRC, electrons pass through both branches of the PSI

RC, although at different rates: 35×10^6 s⁻¹ for the branch ending in Q_K-B and 4.4×10^6 s⁻¹ for that ending in Q_K-A. Indeed, the PSI RC is most closely related to the RC of **green sulfur bacteria** (a second class of photosynthetic bacteria), which is a true homodimer.

Up until this point, PSI's RC resembles those of PSII and purple photosynthetic bacteria. However, rather than the reduced forms of either Q_K-A or Q_K-B dissociating from PSI, both of these quinones directly pass their photoexcited electron to a chain of three spectroscopically identified [4Fe-4S] clusters designated F_X, F_A, and F_B (right side of Fig. 24-18). F_X, which lies on the pseudo-2-fold axis relating PsaA and PsaB, is coordinated by two Cys residues from each of these subunits. F_A and F_B are bound to the stromal subunit PsaC, which structurally resembles bacterial 2[4Fe-4S] ferredoxins (e.g., Fig. 22-16). Mutational studies on the Cys residues of PsaC that coordinate its two [4Fe-4S] clusters indicate that the cluster that lies closer to F_X is F_A and the more distant cluster is F_B (Fig. 24-27). The observation that both branches of PSI's electron-transfer pathways are active, in contrast to only one active branch in PSII and the PbRC, is rationalized by the fact that the two quinones at the ends of each branch are functionally equivalent in PSI but functionally different in PSII and the PbRC.

PSI's core antenna system consists of 90 Chl *a* molecules and 22 carotenoids (Fig. 24-26a). The Mg²⁺ ions of 79 of these Chl *a* molecules are axially liganded by residues of PsaA and PsaB (mostly His side chains or protein-bound water molecules), whereas the remaining 11 are so liganded by the smaller subunits PsaJ through M and PsaX. The spatial distribution of these antenna Chl *a*'s resembles that in the core antenna subunits CP43 and CP47 of PSII. Indeed, the N-terminal domains of PsaA and PsaB are similar in sequence to those of CP43 and CP47 and fold into similar structures containing six transmembrane helices each. The carotenoids, which are mostly β-carotenes, are deeply buried in the membrane, where they are in van der Waals contact with Chl *a* rings. This permits efficient energy transfer from photoexcited carotenoids to Chl *a* as well as protects PSI from photooxidative damage. PSI also tightly binds four lipid molecules such that their fatty acyl groups are embedded among the complex's transmembrane helices. This strongly suggests that these lipids have specific structural and/or functional roles rather than being artifacts of preparation. Indeed, the head group of one of them, a phospholipid, coordinates the Mg²⁺ of an antenna Chl *a*, an unprecedented interaction.

PSIs from higher plants are monomers rather than trimers as are cyanobacterial PSIs. Nevertheless, the X-ray structure of PSI from peas, determined by Nathan Nelson, reveals that the positions and orientations of the chlorophylls in both species of PSIs are nearly identical, a remarkable finding considering the >1 billion years since chloroplasts diverged from their cyanobacterial ancestors. However, pea PSI has four antenna proteins not present in cyanobacterial PSI that are arranged in a crescent-shaped transmembrane belt around one side of its RC and which collectively bind 56 chlorophyll molecules.

h. PSI-Activated Electrons May Reduce NADP⁺ or Motivate Proton Gradient Formation

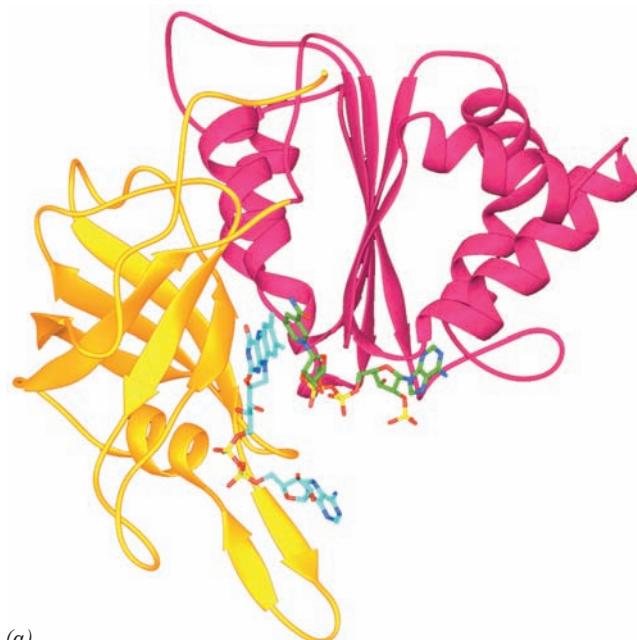
Electrons ejected from F_B in PSI may follow either of two alternative pathways (Fig. 24-18):

1. Most electrons follow a noncyclic pathway by passing to an ~100-residue, single [2Fe-2S]-containing, soluble ferredoxin (**Fd**) that is located in the stroma. Reduced Fd, in turn, reduces NADP⁺ in a reaction mediated by the ~310-residue, monomeric, FAD-containing **ferredoxin-NADP⁺ reductase** (**FNR**, Fig. 24-28a), to yield the final product of the chloroplast light reaction, NADPH. Two reduced Fd molecules successively deliver one electron each to the FAD of FNR, which thereby sequentially assumes the neutral semiquinone and fully reduced states before transferring the two electrons and a proton to the NADP⁺ via what is formally a hydride ion transfer. The X-ray structure of the complex between Fd and FNR from maize leaf (Fig. 24-28b), determined by Genji Kurisu, reveals that the shortest interatomic approach between Fd's [2Fe-2S] cluster and FNR's FAD is the 6.0 Å between an Fe atom and FAD atom C8a (the methyl C closest to its ribitol residue; Fig. 16-8). This is

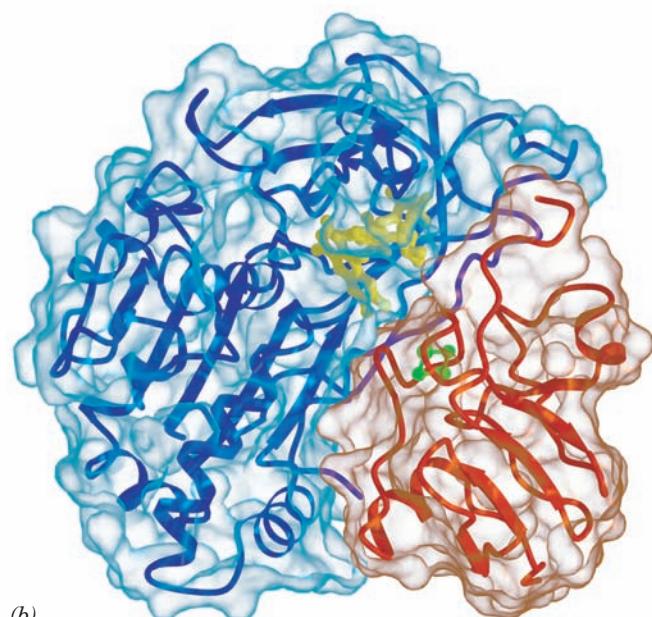
sufficiently close for direct electron transfer through space between these prosthetic groups. The complex is stabilized by five salt bridges, as similarly appears to be the case for the interaction between cytochrome *f* and PC.

2. Some electrons are returned from PSI, via cytochrome *b*₆, to the plastoquinone pool, thereby traversing a cyclic pathway that translocates protons across the thylakoid membrane. A mechanism that has been proposed for this process is that Fd transfers an electron to heme *x* of cytochrome *b*₆ (Fig. 24-24) rather than to FNR. Since heme *x* contacts heme *b*_L at the periphery of cytochrome *b*₆'s Q_i site, an electron injected into heme *x* would be expected to reduce plastoquinone via a Q cycle-like mechanism (Fig. 22-31). Note that the cyclic pathway is independent of the action of PSII and hence does not result in the evolution of O₂. This accounts for the observation that chloroplasts absorb more than eight photons per O₂ molecule evolved.

The cyclic electron flow presumably functions to increase the amount of ATP produced relative to that of



(a)



(b)

Figure 24-28 Ferredoxin-NADP⁺ reductase. (a) The X-ray structure of the Y308S mutant form of pea ferredoxin-NADP⁺ reductase (FNR) in complex with FAD and NADP⁺. This 308-residue protein has two domains: The N-terminal domain (gold), which forms the FAD binding site, folds into an antiparallel β barrel, whereas the C-terminal domain (magenta), which provides the NADP⁺ binding site, forms a dinucleotide-binding fold (Section 8-3Bi). The FAD and NADP⁺ are shown in stick form with NADP⁺ C green, FAD C cyan, N blue, O red, and P yellow. The flavin and nicotinamide rings are in opposition with C4 of the nicotinamide ring and C5 of the flavin ring 3.0 Å apart, an arrangement that is consistent with direct hydride transfer as also occurs in glutathione reductase and dihydrolipooyl dehydrogenase (Section 21-2Ba). However, in contrast to these

latter enzymes, whose bound flavin and nicotinamide rings are parallel, those in FNR are inclined by ~30°, a heretofore unobserved binding mode. [Based on an X-ray structure by Andrew Karplus, Cornell University. PDBid 1QFY.] (b) The X-ray structure of the complex between Fd (red) and FNR (blue) from maize leaf with both proteins drawn in ribbon form embedded in their transparent solvent-accessible surfaces. The [2Fe-2S] cluster of Fd (green) and the FAD of FNR (yellow) are drawn in ball-and-stick form. The Fd binds in a hollow between FNR's two domains (Part a) such that the line joining the two Fe's of the [2Fe-2S] cluster lies roughly in the plane of the flavin ring. [Courtesy of Genji Kurisu, Osaka University, Osaka, Japan. PDBid 1GAQ.]  See Interactive Exercise 22

NADPH and thus permits the cell to adjust the relative amounts of these two substances produced according to its needs. However, the mechanism that apportions electrons between the cyclic and noncyclic pathways is unknown.

i. PSI and PSII Occupy Different Parts of the Thylakoid Membrane

Freeze-fracture electron microscopy (Section 12-3Ca) revealed that the protein complexes of the thylakoid membrane have characteristic distributions (Fig. 24-29):

1. PSI occurs mainly in the unstacked stroma lamellae, in contact with the stroma, where it has access to NADP⁺.
2. PSII is located almost exclusively between the closely stacked grana, out of direct contact with the stroma.
3. Cytochrome *b*₆*f* is uniformly distributed throughout the membrane.

The high mobilities of plastoquinone and plastocyanin, the electron carriers that shuttle electrons between these complexes, permits photosynthesis to proceed at a reasonable rate.

What function is served by the segregation of PSI and PSII, which are typically present in chloroplasts in equimolar amounts? If these two photosystems were in close proximity, the higher excitation energy of PSII (P680 vs P700) would cause it to pass a large fraction of its absorbed photons to PSI via exciton transfer; that is, PSII would act as a light-harvesting antenna for PSI (Fig. 24-7b). The separation of these particles by around 100 Å eliminates this difficulty.

The physical separation of PSI and PSII also permits the chloroplast to respond to changes in illumination. The relative amounts of light absorbed by the two photosystems vary with how the light-harvesting complexes (LHCs) are distributed between the stacked and un-

stacked portions of the thylakoid membrane. Under high illumination (normally direct sunlight, which contains a high proportion of short-wavelength blue light), all else being equal, PSII absorbs more light than PSI. PSI is then unable to take up electrons as fast as PSII can supply them, so the plastoquinone is predominantly in its reduced state. The reduced plastoquinone activates a protein kinase to phosphorylate specific Thr residues of the LHCs, which, in response, migrate to the unstacked regions of the thylakoid membrane, where they bind to PSI. A greater fraction of the incident light is thereby funneled to PSI. Under low illumination (normally shady light, which contains a high proportion of long-wavelength red light), PSI takes up electrons faster than PSII can provide them so that plastoquinone predominantly assumes its oxidized form. The LHCs are consequently dephosphorylated and migrate to the stacked portions of the thylakoid membrane, where they drive PSII. The chloroplast therefore maintains the balance between its two photosystems by a light-activated feedback mechanism.

D. Photophosphorylation

Chloroplasts generate ATP in much the same way as mitochondria, that is, by coupling the dissipation of a proton gradient to the enzymatic synthesis of ATP (Section 22-3C). This was clearly demonstrated by the imposition of an artificially produced pH gradient across the thylakoid membrane. Chloroplasts were soaked, in the dark, for several hours in a succinic acid solution at pH 4 so as to bring the thylakoid lumen to this pH (the thylakoid membrane is permeable to unionized succinic acid). The abrupt transfer of these chloroplasts to an ADP + P_i-containing buffer at pH 8 resulted in an impressive burst of ATP synthesis: About 100 ATPs were synthesized per molecule of cytochrome *f* present. Moreover, the amount of ATP

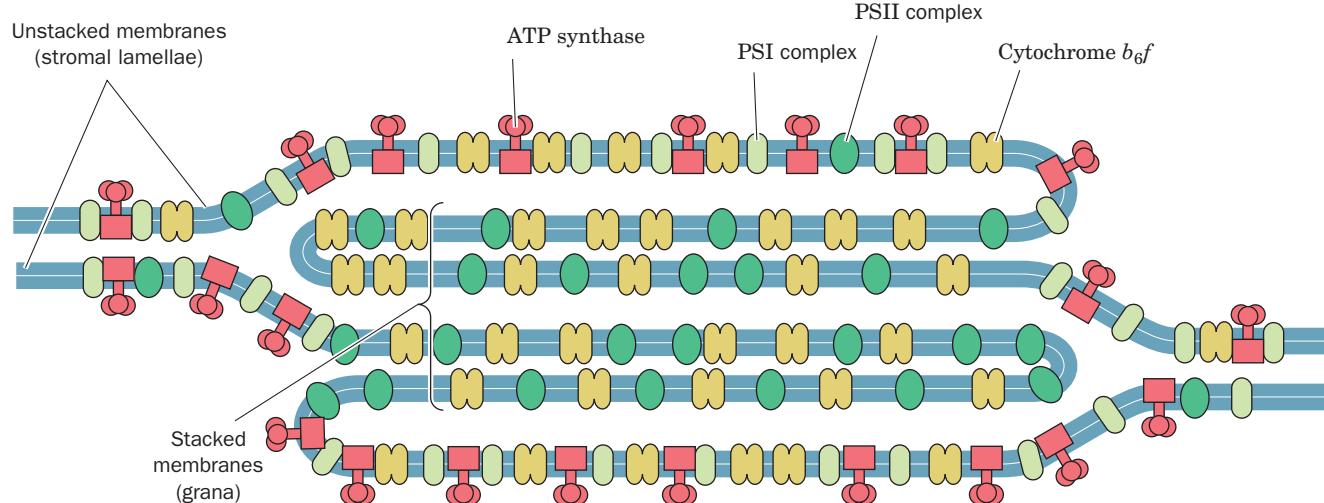


Figure 24-29 Segregation of PSI and PSII. The distribution of photosynthetic protein complexes between the stacked (grana) and the unstacked (stroma exposed) regions of the thylakoid

membrane is shown. [After Anderson, J.M. and Anderson, B., *Trends Biochem. Sci.* 7, 291 (1982).]

synthesized was unaffected by the presence of electron-transport inhibitors such as DCMU. This, together with the observations that photophosphorylation requires an intact thylakoid membrane and that proton translocators such as 2,4-dinitrophenol (Section 22-3D) uncouple photophosphorylation from light-driven electron transport, provides convincing evidence favoring the chemiosmotic hypothesis (Section 22-3A).

a. Chloroplast Proton-Translocating ATP Synthase Resembles That of Mitochondria

Electron micrographs of thylakoid membrane stromal surfaces and bacterial plasma membrane inner surfaces reveal lollipop-shaped structures (Fig. 24-30). These closely resemble the F_1 units of the proton-translocating ATP synthase studding the matrix surfaces of inner mitochondrial membranes (Fig. 22-36a). In fact, the chloroplast ATP synthase, which is termed the **CF_1CF_0 complex** (C for chloroplast), has remarkably similar properties to the mitochondrial F_1F_0 complex (Section 22-3C). For example,

1. Both F_0 and CF_0 units are hydrophobic transmembrane proteins that contain a proton-translocating channel.
2. Both F_1 and CF_1 are hydrophilic peripheral membrane proteins of subunit composition $\alpha_3\beta_3\gamma\delta\epsilon$, of which β is a reversible ATPase.
3. Both ATP synthases are inhibited by oligomycin.
4. Chloroplast ATP synthase translocates protons out of the thylakoid lumen into the stroma (Fig. 24-17), and mitochondrial ATP synthase conducts them into the matrix space (the mitochondrial equivalent of the stroma) from the intermembrane space (Section 22-3A).

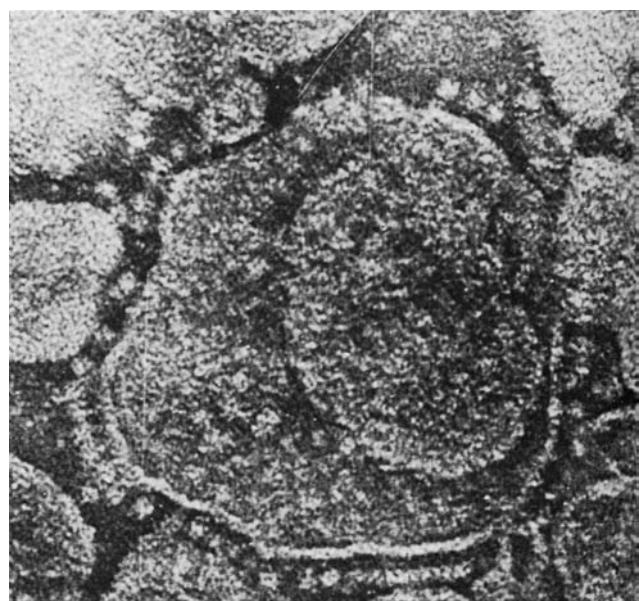


Figure 24-30 Electron micrograph of thylakoids. The CF_1 “lollipops” of their ATP synthases project from their stromal surfaces. Compare this with Fig. 22-36a. [Courtesy of Peter Hinkle, Cornell University.]

Clearly, proton-translocating ATP synthases must have evolved very early in the history of cellular life. Chloroplast ATP synthase is located in the unstacked portions of the thylakoid membrane, in contact with the stroma, where there is room for the bulky CF_1 globule and access to ADP (Fig. 24-29).

b. Photosynthesis with Noncyclic Electron Transport Produces Around 1.25 ATP Equivalents per Absorbed Photon

At saturating light intensities, chloroplasts generate proton gradients of ~ 3.5 pH units across their thylakoid membranes. This, as we have seen (Figs. 24-17 and 24-18), arises from two sources:

1. The evolution of a molecule of O_2 from two H_2O molecules releases four protons into the thylakoid lumen. These protons should be considered as being supplied from the stroma by the protons and H atoms taken up in the synthesis of NADPH.
2. The transport of the liberated four electrons through the cytochrome b_6f complex occurs with the translocation of what is estimated to be eight protons from the stroma to the thylakoid lumen.

Altogether ~ 12 protons are translocated per molecule of O_2 produced by noncyclic electron transport.

The thylakoid membrane, in contrast to the inner mitochondrial membrane, is permeable to ions such as Mg^{2+} and Cl^- . Translocation of protons and electrons across the thylakoid membrane is consequently accompanied by the passage of these ions so as to maintain electrical neutrality (Mg^{2+} out and Cl^- in). This all but eliminates the membrane potential, $\Delta\Psi$ (Eq. [22.1]). *The electrochemical gradient in chloroplasts is therefore almost entirely a result of the pH gradient.*

Chloroplast ATP synthase, according to most estimates, produces one ATP for every three protons it transports out of the thylakoid lumen. Noncyclic electron transport in chloroplasts therefore results in the production of $\sim 12/3 = 4$ molecules of ATP per molecule of O_2 evolved (although this quantity is subject to revision) or around half an ATP per photon absorbed. Cyclic electron transport is a more productive ATP generator since it yields two-thirds of an ATP (two protons transported) per absorbed photon. The noncyclic process, of course, also yields NADPH, each molecule of which has the free energy to produce three ATPs (Section 22-2A; although this does not normally occur), for a total of six more ATP equivalents per O_2 produced. Consequently, the energetic efficiency of the noncyclic process is $4/8 + 6/8 = 1.25$ ATP equivalents per absorbed photon.

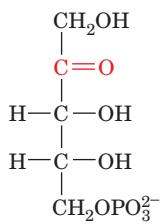
3 DARK REACTIONS

In the previous section we saw how light energy is harnessed to generate ATP and NADPH. In this section we discuss how these products are used to synthesize carbohydrates and other substances from CO_2 .

A. The Calvin Cycle

The metabolic pathway by which plants incorporate CO_2 into carbohydrates was elucidated between 1946 and 1953 by Melvin Calvin, James Bassham, and Andrew Benson. They did so by tracing the metabolic fate of the radioactive label from $^{14}\text{CO}_2$ as it passed through a series of photosynthetic intermediates. The basic experimental strategy they used was to expose growing cultures of algae, such as *Chlorella*, to $^{14}\text{CO}_2$ for varying times and under differing illumination conditions and then to drop the cells into boiling alcohol so as to disrupt them while preserving their labeling pattern. The radioactive products were subsequently separated and identified (an often difficult task) through the use of the then recently developed technique of two-dimensional paper chromatography (Section 6-3Dc) coupled with autoradiography. The overall pathway, diagrammed in Fig. 24-31, is known as the **Calvin cycle** or the **reductive pentose phosphate cycle**.

Some of Calvin's earliest experiments indicated that algae exposed to $^{14}\text{CO}_2$ for a minute or more had synthesized a complex mixture of labeled metabolic products, including sugars and amino acids. By inactivating the algae within 5 s of their exposure to $^{14}\text{CO}_2$, however, it was shown that *the first stable radioactively labeled compound formed is 3-phosphoglycerate (3PG), which is initially labeled only in its carboxyl group*. This result immediately suggested, in analogy with most biochemical experience, that the 3PG was formed by the carboxylation of a C_2 compound. Yet the failure to find any such precursor eventually forced this hypothesis to be abandoned. The actual carboxylation reaction was discovered through an experiment in which illuminated algae had been exposed to $^{14}\text{CO}_2$ for ~ 10 min so that the levels of their labeled photosynthetic intermediates had reached a steady state. The CO_2 was then withdrawn. As expected, the carboxylation product, 3PG, decreased in concentration (Fig. 24-32) because it was depleted by reactions farther along the pathway. The concentration of **ribulose-5-phosphate (Ru5P)**,



Ribulose-5-phosphate (Ru5P)

however, simultaneously increased. Evidently, Ru5P is the Calvin cycle's carboxylation substrate. If so, the resulting C_6 carboxylation product must split into two C_3 compounds, one of which is 3PG (Fig. 24-31, Reaction 2). A consideration of the oxidation states of Ru5P and CO_2 indicates that, in fact, both C_3 compounds must be 3PG and that the carboxylation reaction requires no external redox source.

While the search for the carboxylation substrate was going on, several other photosynthetic intermediates had

been identified and, through chemical degradation studies, their labeling patterns had been elucidated. For example, the hexose fructose-1,6-bisphosphate (FBP) is initially labeled only at its C3 and C4 positions (Fig. 24-31) but later becomes labeled to a lesser degree at its other atoms. Similarly, a series of tetrose, pentose, hexose, and heptose phosphates were isolated that had the identities and initial labeling patterns indicated in Fig. 24-31. A consideration of the flow of the labeled atoms through these various intermediates led, in what was a milestone of metabolic biochemistry, to the deduction of the Calvin cycle as is diagrammed in Fig. 24-31. The existence of many of its postulated reactions was eventually confirmed by *in vitro* studies using purified enzymes.

a. The Calvin Cycle Generates GAP from CO_2 via a Two-Stage Process

The Calvin cycle may be considered to have two stages:

Stage 1 The production phase (top line of Fig. 24-31), in which three molecules of Ru5P react with three molecules of CO_2 to yield six molecules of glyceraldehyde-3-phosphate (GAP) at the expense of nine ATP and six NADPH molecules. *The cyclic nature of the pathway makes this process equivalent to the synthesis of one GAP from three CO_2 molecules.* Indeed, at this point, one GAP can be bled off from the cycle for use in biosynthesis (see Stage 2).

Stage 2 The recovery phase (bottom lines of Fig. 24-31), in which the carbon atoms of the remaining five GAPs are shuffled in a remarkable series of reactions, similar to those of the pentose phosphate pathway (Section 23-4), to reform the three Ru5Ps with which the cycle began. Indeed, the elucidation of the pentose phosphate pathway at about the same time that the Calvin cycle was being worked out provided much of the biochemical evidence in support of the Calvin cycle. This stage can be conceptually decomposed into four sets of reactions (with the numbers keyed to the corresponding reactions in Fig. 24-31):

6. $\text{C}_3 + \text{C}_3 \longrightarrow \text{C}_6$
8. $\text{C}_3 + \text{C}_6 \longrightarrow \text{C}_4 + \text{C}_5$
9. $\text{C}_3 + \text{C}_4 \longrightarrow \text{C}_7$
11. $\text{C}_3 + \text{C}_7 \longrightarrow \text{C}_5 + \text{C}_5$

The overall stoichiometry for this process is therefore



Note that this stage of the Calvin cycle occurs without further input of free energy (ATP) or reducing power (NADPH).

b. Most Calvin Cycle Reactions Also Occur in Other Metabolic Pathways

The types of reactions in the Calvin cycle are all familiar (Section 23-4), with the exception of the carboxylation reaction. This first stage of the Calvin cycle begins with the phosphorylation of Ru5P by **phosphoribulokinase** to form **ribulose-1, 5-bisphosphate (RuBP)**. Following the

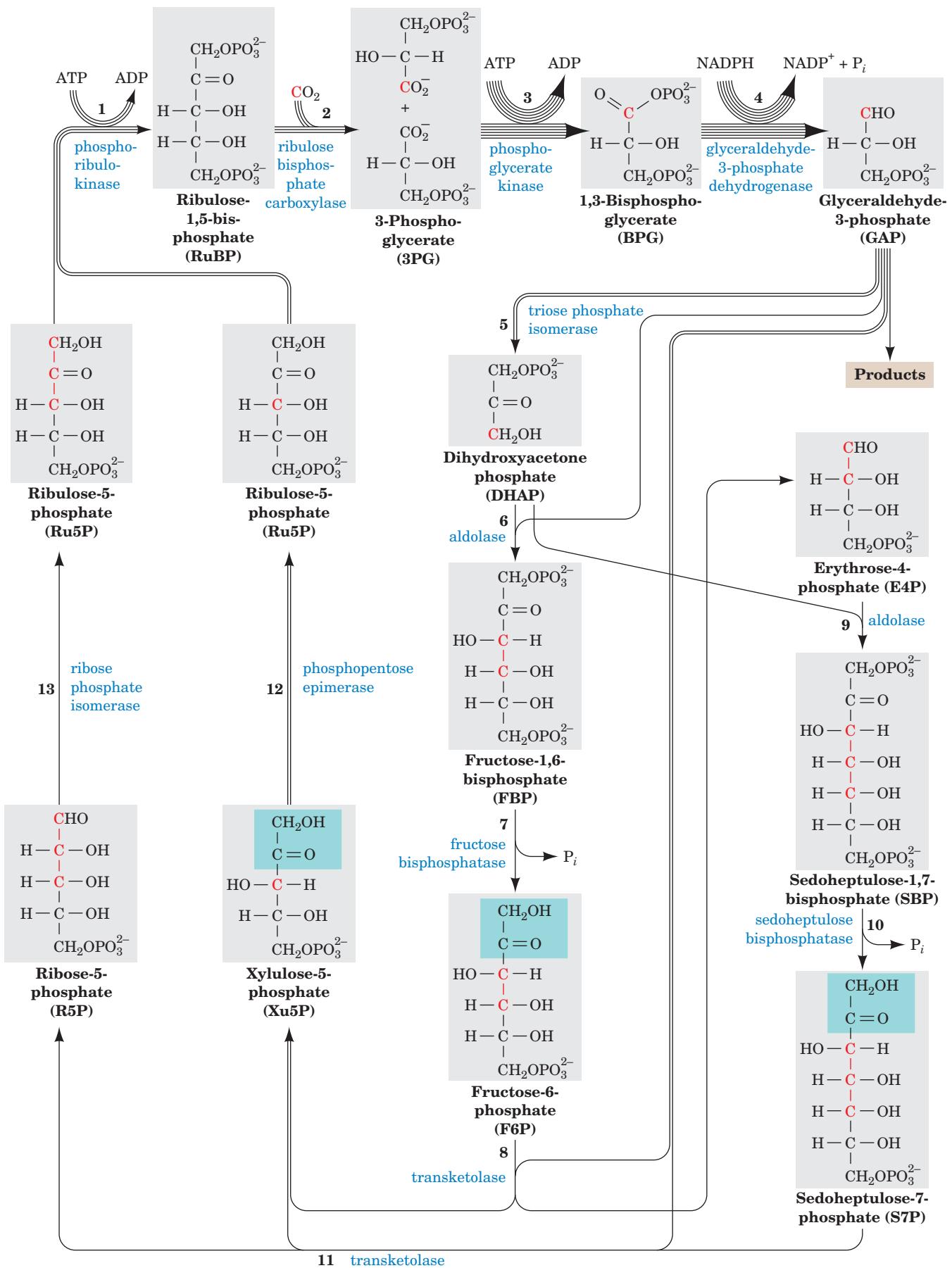


Figure 24-31 The Calvin cycle. (Opposite) The number of lines in an arrow indicates the number of molecules reacting in that step for a single turn of the cycle that converts three CO_2 molecules to one GAP molecule. For the sake of clarity, the sugars are all shown in their linear forms, although the hexoses and heptoses predominantly exist in their cyclic forms (Section 11-1B). The ^{14}C -labeling patterns generated in one turn of the cycle through the use of $^{14}\text{CO}_2$ are indicated in red. Note that two of the Ru5Ps are labeled only at C3, whereas the third Ru5P is equally labeled at C1, C2, and C3. 

carboxylation step, which is discussed below, the resulting 3PG is converted first to 1,3-bisphosphoglycerate (BPG) and then to GAP. This latter sequence is the reverse of two consecutive glycolytic reactions (Sections 17-2G and 17-2F) except that the Calvin cycle reaction involves NADPH rather than NADH.

The second stage of the Calvin cycle begins with the reverse of a familiar glycolytic reaction, the isomerization of GAP to dihydroxyacetone phosphate (DHAP) by triose phosphate isomerase (Section 17-2E). Following this, DHAP is directed along two analogous paths (Fig. 24-31): Reactions 6–8 or Reactions 9–11. Reactions 6 and 9 are

aldolase-catalyzed aldol condensations in which DHAP is linked to an aldehyde (aldolase is specific for DHAP but accepts a variety of aldehydes). Reaction 6 is also the reverse of a glycolytic reaction (Section 17-2D). Reactions 7 and 10 are phosphate hydrolysis reactions that are catalyzed, respectively, by fructose bisphosphatase (FBPase, which we previously encountered in our discussion of glycolytic substrate cycles and gluconeogenesis; Sections 17-4Ff and 23-1Ah) and **sedoheptulose bisphosphatase (SBPase)**. The remaining Calvin cycle reactions are catalyzed by enzymes that also participate in the pentose phosphate pathway. In Reactions 8 and 11, both catalyzed by **transketolase**, a C_2 keto unit (shaded in green in Fig. 24-31) is transferred from a ketose to GAP to form **xylulose-5-phosphate (Xu5P)** and leave the aldoses **erythrose-4-phosphate (E4P)** in Reaction 8 and **ribose-5-phosphate (R5P)** in Reaction 11. The E4P produced by Reaction 8 feeds into Reaction 9. The Xu5Ps produced by Reactions 8 and 11 are converted to Ru5P by **phosphopentose epimerase** in Reaction 12. The R5P from Reaction 11 is also converted to Ru5P by **ribose phosphate isomerase** in Reaction 13, thereby completing a turn of the Calvin cycle. Thus only 3 of the 11 Calvin cycle enzymes, phosphoribulokinase, the carboxylation enzyme **ribulose bisphosphate carboxylase**, and SBPase, have no equivalents in animal tissues.

c. RuBP Carboxylase Catalyzes CO_2 Fixation in an Exergonic Process

The enzyme that catalyzes CO_2 fixation, ribulose bisphosphate carboxylase (**RuBP carboxylase**), is arguably the world's most important enzyme, since nearly all life on Earth ultimately depends on its action. This protein, presumably as a consequence of its particularly low catalytic efficiency ($k_{\text{cat}} = \sim 3 \text{ s}^{-1}$), comprises up to 50% of leaf proteins and is therefore the most abundant protein in the biosphere (it is estimated to be synthesized at the rate of $\sim 4 \times 10^9$ tons/year, which fixes $\sim 10^{11}$ tons of CO_2 /year; in comparison crude oil is consumed at the rate of $\sim 3 \times 10^9$ tons/year). RuBP carboxylase from higher plants and most photosynthetic microorganisms consists of eight large (L) subunits (477 residues in tobacco leaves) encoded by chloroplast DNA and eight small (S) subunits (123 residues) specified by a nuclear gene (the RuBP carboxylase from certain photosynthetic bacteria is an L₂ dimer whose L subunit has 28% sequence identity with and is structurally similar to that of the L₈S₈ enzyme). X-ray studies by Carl-Ivar Brändén and by David Eisenberg demonstrated that the L₈S₈ enzyme has the D₄ symmetry of a square prism (Fig. 24-33a,b). The L subunit contains the enzyme's catalytic site, as is demonstrated by its enzymatic activity in the absence of the S subunit. It consists of two domains (Fig. 24-33c): Residues 1 to 150 form a mixed five-stranded β sheet and residues 151 to 477 fold into an α/β barrel (Fig. 8-19b,c) which, as do nearly all known α/β barrel enzymes (Section 8-3Bh), contains the enzyme's active site at the mouth of the barrel near the C-terminus of its β strands. The function of the S subunit is unknown; attempts to show that it has a regulatory role, in analogy with other enzymes, have been unsuccessful.

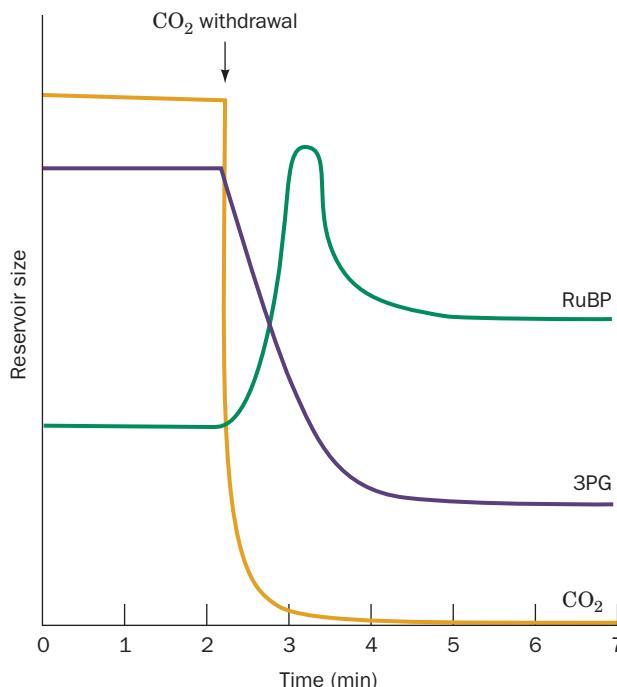


Figure 24-32 Algal 3PG and RuBP levels on removal of CO_2 . The time course of the levels of 3PG (purple curve) and RuBP (green curve) in steady-state $^{14}\text{CO}_2$ -labeled, illuminated algae is shown during a period in which the CO_2 (orange curve) is abruptly withdrawn. In the absence of CO_2 , the 3PG concentration rapidly decreases because it is taken up by the reactions of the Calvin cycle but cannot be replenished by them. Conversely, the RuBP concentration transiently increases as it is synthesized from the residual pool of Calvin cycle intermediates but, in the absence of CO_2 , cannot be used for their regeneration.

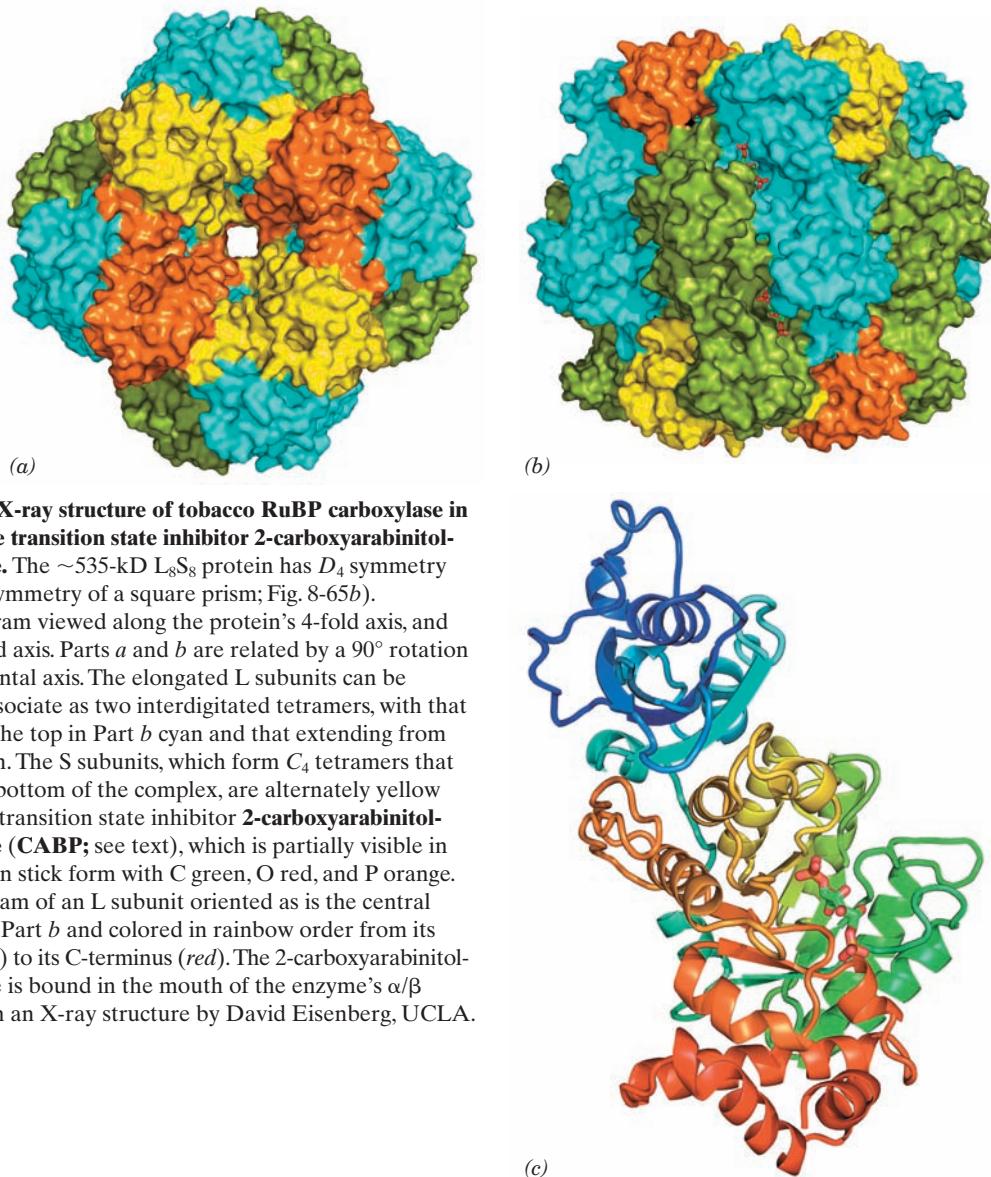
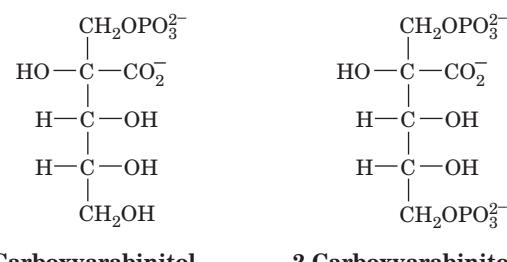


Figure 24-33 X-ray structure of tobacco RuBP carboxylase in complex with the transition state inhibitor 2-carboxyarabinitol-1,5-bisphosphate. The ~ 535 -kD L_8S_8 protein has D_4 symmetry (the rotational symmetry of a square prism; Fig. 8-65b). (a) Surface diagram viewed along the protein's 4-fold axis, and (b) along a 2-fold axis. Parts *a* and *b* are related by a 90° rotation about the horizontal axis. The elongated L subunits can be considered to associate as two interdigitated tetramers, with that extending from the top in Part *b* cyan and that extending from the bottom green. The S subunits, which form C_4 tetramers that cap the top and bottom of the complex, are alternately yellow and orange. The transition state inhibitor **2-carboxyarabinitol-1,5-bisphosphate (CABP)** (see text), which is partially visible in Part *b*, is drawn in stick form with C green, O red, and P orange. (c) Ribbon diagram of an L subunit oriented as is the central green subunit in Part *b* and colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The 2-carboxyarabinitol-1,5-bisphosphate is bound in the mouth of the enzyme's α/β barrel. [Based on an X-ray structure by David Eisenberg, UCLA. PDBid 1RLC.]

The accepted mechanism of RuBP carboxylase, which was largely formulated by Calvin, is indicated in Fig. 24-34. Abstraction of the C3 proton of RuBP, the reaction's rate-determining step, generates an enediolate that nucleophilically attacks CO_2 (not HCO_3^-). The resulting β -keto acid is rapidly attacked at its C3 position by H_2O to yield an adduct that splits, by a reaction similar to aldol cleavage, to yield the two product 3PG molecules. The following evidence favors this mechanism:

1. The C3 proton of enzyme-bound RuBP exchanges with solvent, an observation compatible with the existence of the enediolate intermediate.
2. The C2 and C3 oxygen atoms remain attached to their respective C atoms, which eliminates mechanisms involving a covalent adduct such as a Schiff base between RuBP and the enzyme.
3. The trapping of the proposed β -keto acid intermediate by borohydride reduction and the tight enzymatic binding of its analogs, such as **2-carboxyarabinitol-1-phosphate (CA1P)**

(CA1P) and 2-carboxyarabinitol-1,5-bisphosphate (CABP),



provide strong evidence for the existence of this intermediate.

The driving force for the overall reaction, which is highly exergonic ($\Delta G' = -35.1 \text{ kJ} \cdot \text{mol}^{-1}$), is provided by the

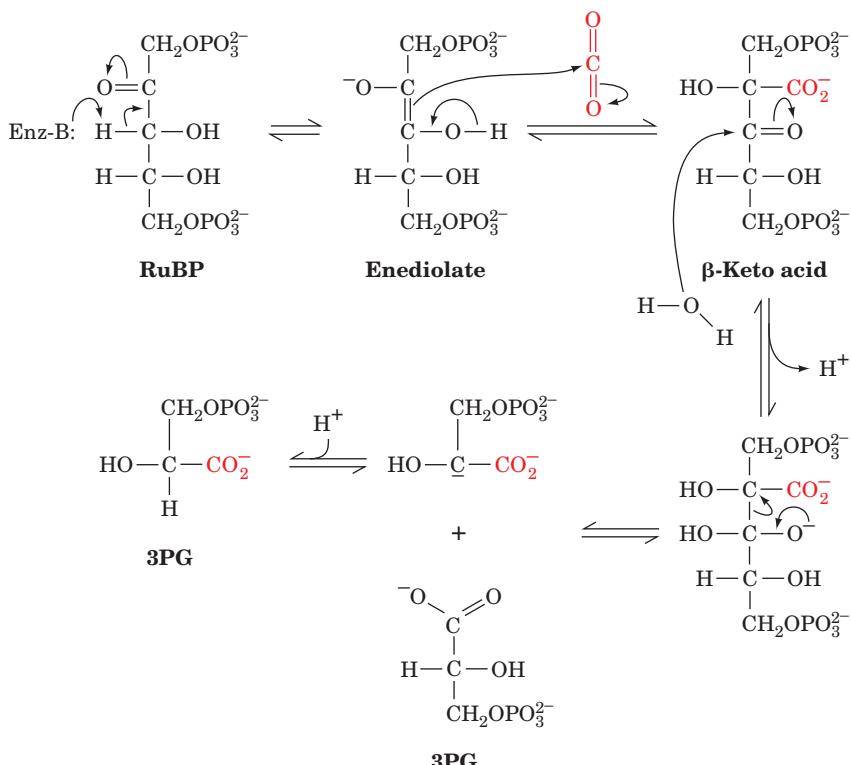


Figure 24-34 Probable mechanism of the carboxylation reaction catalyzed by RuBP carboxylase. The reaction proceeds via an enediolate intermediate that nucleophilically attacks CO₂

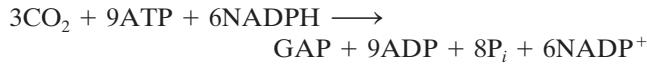
to form a β -keto acid. This intermediate reacts with water to yield two molecules of 3PG.  See the Animated Figures

cleavage of the β -keto acid intermediate to yield an additional resonance-stabilized carboxylate group.

RuBP carboxylase activity requires a bound divalent metal ion, physiologically Mg²⁺, which acts to stabilize the developing negative charges during catalysis. The Mg²⁺ is, in part, bound to the enzyme by a catalytically essential carbamate group that is generated by the reaction of a non-substrate CO₂ with the ϵ -amino group of Lys 201 (R—NH₂ + CO₂ \rightarrow R—NH—COO⁻ + H⁺). Although the *in vitro* activation reaction occurs spontaneously in the presence of Mg²⁺ and HCO₃⁻, it is blocked *in vivo* by the particularly tight binding of RuBP to active sites lacking carbamate. This inhibition is relieved, however, by the release of RuBP in an ATP-driven process catalyzed by **RuBP carboxylase activase**.

d. Calvin Cycle Products Are Converted to Starch, Sucrose, and Cellulose

The overall stoichiometry of the Calvin cycle is

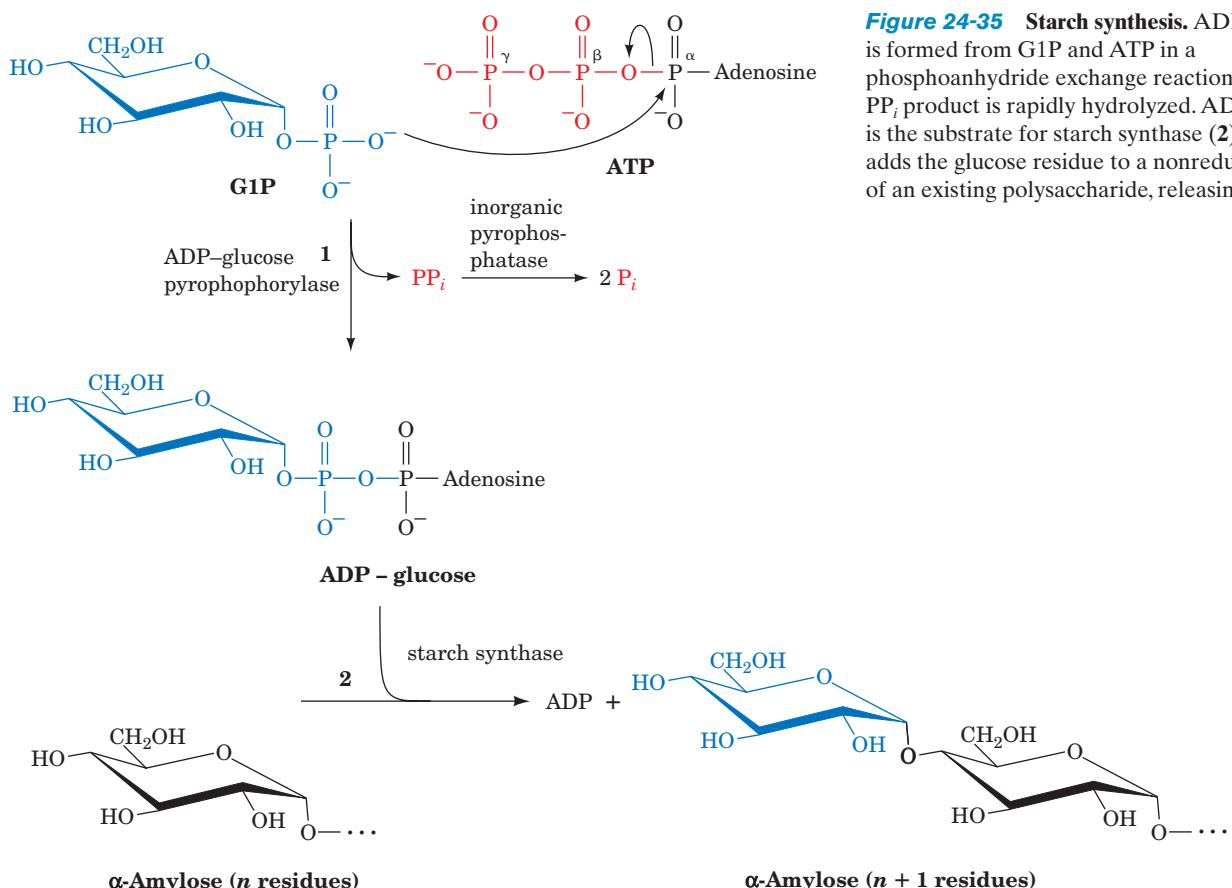


GAP, the primary product of photosynthesis, is used in a variety of biosynthetic pathways, both inside and outside the chloroplast. For example, fatty acids and amino acids are synthesized from GAP as is described, respectively, in Sections 25-4 and 26-5. GAP can also be converted to fructose-

6-phosphate by the further action of Calvin cycle enzymes and then to glucose-1-phosphate (G1P) by phosphoglucomutase (Section 17-2B) and phosphoglucomutase (Section 18-1B). G1P is the precursor of the higher carbohydrates characteristic of plants.

The polysaccharide α -amylose, a major component of starch (Section 11-2D), is synthesized in the chloroplast stroma as a temporary storage depot for glucose units. It is also synthesized as a long-term storage molecule elsewhere in the plant, including leaves, seeds, and roots. G1P is first activated by its reaction with ATP to form ADP-glucose as catalyzed by **ADP-glucose pyrophosphorylase**. **Starch synthase** then transfers the glucose residue to a nonreducing end of an α -amylose or amylopectin molecule, forming a new glycosidic linkage (Fig. 24-35). The overall reaction is driven by the exergonic hydrolysis of the PP_i released in the formation of ADP-glucose. A similar reaction sequence occurs in glycogen synthesis, which uses UDP-glucose (Section 18-2A). The $\alpha(1 \rightarrow 6)$ branches of amylopectin (Section 11-2D) are made by **starch-branched enzyme**, which functions similarly to glycogen branching enzyme (Section 18-2C).

Sucrose, a disaccharide of glucose and fructose (Section 11-2B), is the major transport sugar for delivering carbohydrates to nonphotosynthesizing cells and hence is the major photosynthetic product of green leaves. Since sucrose is synthesized in the cytosol, either glyceraldehyde-3-phosphate



or dihydroxyacetone phosphate is transported out of the chloroplast by an antiporter that exchanges phosphate for a triose phosphate. Two trioses combine to form fructose-6-phosphate (F6P) and subsequently glucose-1-phosphate (G1P), which is then activated by UTP to form UDP-glucose. Next, sucrose-6-phosphate is produced in a reaction catalyzed by **sucrose-phosphate synthase**. Finally, sucrose-6-phosphate is hydrolyzed by **sucrose-phosphate phosphatase** to yield sucrose (Fig. 24-36), which is then exported to other plant tissues.

Cellulose, which consists of long chains of $\beta(1 \rightarrow 4)$ -linked glucose units and is the major polysaccharide of plants, is also synthesized from UDP-glucose. Plant cell walls consist of almost crystalline cables containing 36 par-

allel cellulose chains, each of 500 to 15,000 glucose units, all embedded in an amorphous matrix of other polysaccharides and lignin (Section 11-2C). Unlike starch in plants or glycogen in mammals, cellulose is synthesized by multisubunit enzyme complexes in the plant plasma membrane and extruded into the extracellular space.

B. Control of the Calvin Cycle

During the day, plants satisfy their energy needs via the light and dark reactions of photosynthesis. At night, however, like other organisms, they must use their nutritional reserves to generate their required ATP and NADPH through glycolysis, oxidative phosphorylation, and the pentose phosphate pathway. Since the stroma contains the enzymes of glycolysis and the pentose phosphate pathway as well as those of the Calvin cycle, *plants must have a light-sensitive control mechanism to prevent the Calvin cycle from consuming this catabolically produced ATP and NADPH in a wasteful futile cycle*.

As we saw in Section 17-4F, the control of flux in a metabolic pathway occurs at enzymatic steps that are far from equilibrium; that is, those that have a large negative value of ΔG . Inspection of Table 24-1 indicates that the four best candidates for flux control in the Calvin cycle are the reactions catalyzed by phosphoribulokinase, RuBP carboxylase, FBPase, and SBPase (Reactions 1, 2, 7, and 10, Fig. 24-31). In fact, the catalytic efficiencies of these four enzymes all vary, *in vivo*, with the level of illumination.

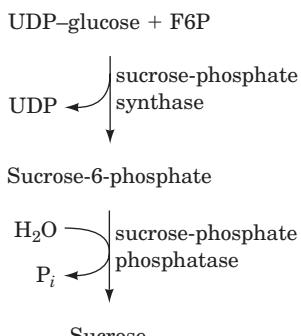


Figure 24-36 The synthesis of sucrose.

Table 24-1 Standard and Physiological Free Energy Changes for the Reactions of the Calvin Cycle

Step ^a	Enzyme	$\Delta G^{\circ'} \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$	$\Delta G \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$
1	Phosphoribulokinase	-21.8	-15.9
2	Ribulose bisphosphate carboxylase	-35.1	-41.0
3 + 4	Phosphoglycerate kinase + glyceraldehyde-3-phosphate dehydrogenase	+18.0	-6.7
5	Triose phosphate isomerase	-7.5	-0.8
6	Aldolase	-21.8	-1.7
7	Fructose bisphosphatase	-14.2	-27.2
8	Transketolase	+6.3	-3.8
9	Aldolase	-23.4	-0.8
10	Sedoheptulose bisphosphatase	-14.2	-29.7
11	Transketolase	+0.4	-5.9
12	Phosphopentose epimerase	+0.8	-0.4
13	Ribose phosphate isomerase	+2.1	-0.4

^aRefer to Fig. 24-31.

Source: Bassham, J.A. and Buchanan, B.B., in Govindjee (Ed.), *Photosynthesis*, Vol. II, p. 155, Academic Press (1982).

The activity of RuBP carboxylase responds to three light-dependent factors:

1. It varies with pH. On illumination, the pH of the stroma increases from around 7.0 to about 8.0 as protons are pumped from the stroma into the thylakoid lumen. RuBP carboxylase has a sharp pH optimum near pH 8.0.

2. It is stimulated by Mg^{2+} . Recall that the light-induced influx of protons to the thylakoid lumen is accompanied by the efflux of Mg^{2+} to the stroma (Section 24-2Db).

3. It is strongly inhibited by its transition state analog 2-carboxyarabinitol-1-phosphate (CA1P; Section 24-3Ac), which many plants synthesize only in the dark. RuBP carboxylase activase (Section 24-3Ac) also facilitates the release of the tight-binding CA1P from RuBP carboxylase.

FBPase and SBPase are also activated by increased pH and Mg^{2+} , and by NADPH as well. The action of these factors is complemented by a second regulatory system that responds to the redox potential of the stroma. **Thioredoxin (Trx)**, an ~105-residue protein that occurs in many types of cells, contains a redox-active disulfide group. Reduced Trx activates five Calvin cycle enzymes by disulfide interchange reactions (Fig. 24-37): phosphoribulokinase, glyceraldehyde-3-phosphate dehydrogenase, FBPase, SBPase, and RuBP carboxylase activase. This explains why these enzymes are activated by reduced disulfide reagents such as dithiothreitol. The redox level of Trx is maintained by **ferredoxin-thioredoxin reductase (FTR)**, which contains a redox-active disul-

fide that is closely associated with a [4Fe-4S] cluster through which the protein directly responds to the redox state of soluble ferredoxin (Fd) in the stroma. This, as we have seen (Section 24-2Ch), varies with the illumination level.

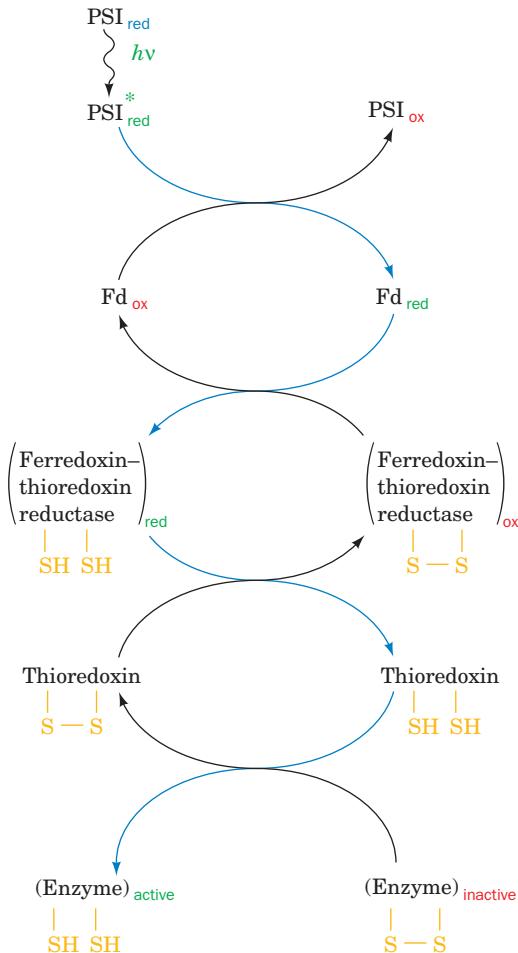


Figure 24-37 Light-activation mechanism of chloroplast enzymes.

Photoactivated PSi reduces soluble ferredoxin (Fd), which reduces ferredoxin-thioredoxin reductase, which, in turn, reduces the disulfide linkage of thioredoxin. Reduced thioredoxin reacts with its target enzymes by disulfide interchange, thereby activating or deactivating these enzymes. In the dark, these processes are quickly reversed by reaction with oxygen.

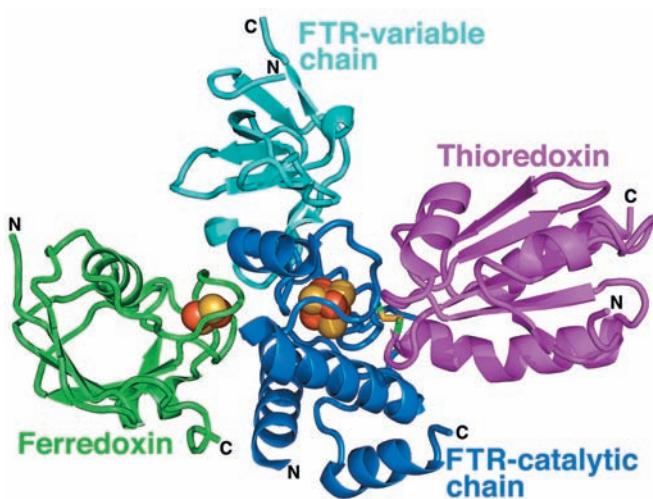


Figure 24-38 X-ray structure of a Fd–FTR–Trx complex. The subunits are drawn in ribbon form with Fd green, the catalytic subunit of FTR blue, its variable subunit cyan, and Trx magenta. The [2Fe–2S] cluster of Fd and the [4Fe–4S] cluster of FTR are drawn in space-filling form with S yellow and Fe red-brown. The two Cys side chains forming the disulfide bond between FTR and Trx are shown in stick form with C green and S yellow. [Based on an X-ray structure by Hans Eklund, Swedish University of Agricultural Sciences, Uppsala, Sweden. PDBid 2PVO.]

The X-ray structure of a Fd–FTR–Trx complex (Fig. 24-38), determined by Hans Eklund, reveals that redox-active Cys residues of FTR and thioredoxin have formed a disulfide bond. Moreover, the distance between this disulfide bond and the [2Fe–2S] cluster of Fd is only ~ 20 Å as a consequence of the remarkably thin disklike catalytic subunit of FTR, and that the [4Fe–4S] cluster of FTR lies on a straight line between these two redox centers. This maximizes the efficiency of electron transfer from the [2Fe–2S] cluster of Fd to the redox active disulfide of Trx.

Reduced Trx also deactivates the chloroplastic enzymes phosphofructokinase (PFK), the main flux-generating enzyme of glycolysis (Section 17-4Fb), and glucose-6-phosphate dehydrogenase, the first enzyme in the pentose phosphate pathway (Section 23-4A), whose products, ATP and NADPH, would otherwise be used by the Calvin cycle in a futile cycle. Thus in plants, *light stimulates the Calvin cycle while deactivating glycolysis and the pentose phosphate pathway, whereas darkness has the opposite effect (that is, the so-called dark reactions do not occur in the dark)*. Moreover, chloroplast ATP synthase is activated by reduced Trx, thus preventing it from uselessly hydrolyzing glycolytically produced ATP in the dark. Indeed, the redox state of Trx regulates a great variety of plant metabolic processes.

We have seen that ferredoxin reduces ferredoxin–NADP⁺ reductase (Section 24-2Ch) and FTR, as well as supplying electrons to the cyclic pathway of chloroplast photosynthesis (Section 24-2Ch). In addition, ferredoxin is the reducing agent for three metabolically essential chloroplast enzymes: **sulfite reductase** (which reduces SO_3^{2-} to S^-), **nitrite reductase** (which reduces NO_2^- to NH_4^+), and **glutamate synthase** (which catalyzes the reaction of α -ketoglutarate and NH_4^+ to form glutamate; Section 26-5Aa). Thus Fd stands at the center of a complex web of enzymatic and regulatory processes.

C. Photorespiration and the C₄ Cycle

It has been known since the 1960s that *illuminated plants consume O_2 and evolve CO_2 in a pathway distinct from oxidative phosphorylation. In fact, at low CO_2 and high O_2 levels, this photorespiration process can outstrip photosynthetic CO_2 fixation*. The basis of photorespiration was unexpected: O_2 competes with CO_2 as a substrate for RuBP carboxylase (RuBP carboxylase is therefore also called **RuBP carboxylase–oxygenase** or **RuBisCO**). In the oxygenase reaction, O_2 reacts with RuBisCO's second substrate, RuBP, to form 3PG and **2-phosphoglycolate** (Fig. 24-39). The 2-phosphoglycolate

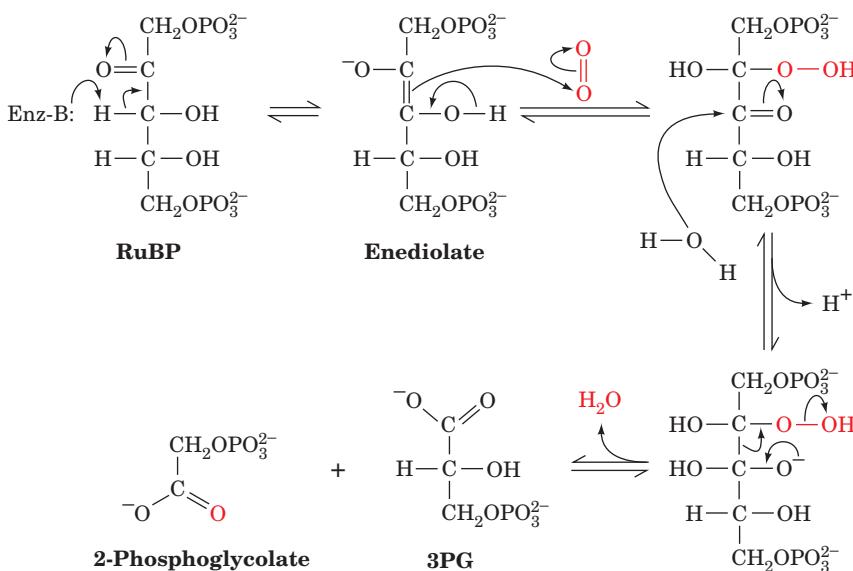


Figure 24-39 Probable mechanism of the oxygenase reaction catalyzed by RuBP carboxylase–oxygenase. Note the similarity of this mechanism to that of the carboxylase reaction catalyzed by the same enzyme (Fig. 24-34).

is hydrolyzed to **glycolate** by **glycolate phosphatase** and, as described below, is partially oxidized to yield CO_2 by a series of enzymatic reactions that occur in the peroxisome and the mitochondrion. Thus photorespiration is a seemingly wasteful process that undoes some of the work of photosynthesis. In the following subsections we discuss the biochemical basis of photorespiration, its significance, and how certain plants manage to evade its deleterious effects.

a. Photorespiration Dissipates ATP and NADPH

The photorespiration pathway is outlined in Fig. 24-40. Glycolate is exported from the chloroplast to the peroxisome (also called the glyoxisome, Sections 1-2Ad and 23-2), where it is oxidized by **glycolate oxidase** to **glyoxylate** and H_2O_2 . The H_2O_2 , a powerful and potentially harmful oxidizing agent, is disproportionated to H_2O and O_2 in the peroxisome by the heme-containing enzyme **catalase**. Some of the glyoxylate is further oxidized by glycolate oxidase to oxalate. The remainder is converted to glycine in a **transamination reaction**, as discussed in Section 26-1A, and exported to the mitochondrion. There, two molecules of glycine are converted to one molecule of serine and one of CO_2 by a reaction described in Section 26-3B. *This is the origin of the CO_2 generated by photorespiration.* The serine is transported back to the peroxisome, where a transamination reaction converts it to **hydroxypyruvate**. This substance is reduced to **glycerate** and phosphorylated in the cytosol to 3PG, which reenters the chloroplast, where it is reconverted to RuBP in the Calvin cycle. *The net result of this complex photorespiration cycle is that some of the ATP and NADPH generated by the light reactions is uselessly dissipated.*

Although photorespiration has no known metabolic function, the RuBisCOs from the great variety of photosynthetic organisms so far tested all exhibit oxygenase activity. Yet, over the eons, the forces of evolution must have optimized the function of this important enzyme. It is thought that photosynthesis evolved at a time when Earth's atmosphere contained large quantities of CO_2 and very little O_2 , so that photorespiration was of no consequence. It has therefore been suggested that the RuBisCO reaction has an obligate intermediate that is inherently autoxidizable. Another possibility is that photorespiration protects the photosynthetic apparatus from photooxidative damage when insufficient CO_2 is available to otherwise dissipate its absorbed light energy. This hypothesis is supported by the observation that when chloroplasts or leaf cells are brightly illuminated in the absence of both CO_2 and O_2 , their photosynthetic capacity is rapidly and irreversibly lost.

b. Photorespiration Limits the Growth Rates of Plants

The steady-state CO_2 concentration attained when a photosynthetic organism is illuminated in a sealed system is named its **CO_2 compensation point**. For healthy plants, this is the CO_2 concentration at which the rates of photosynthesis and photorespiration are equal. For many species it is ~ 40 to 70 ppm (parts per million) CO_2 (the normal

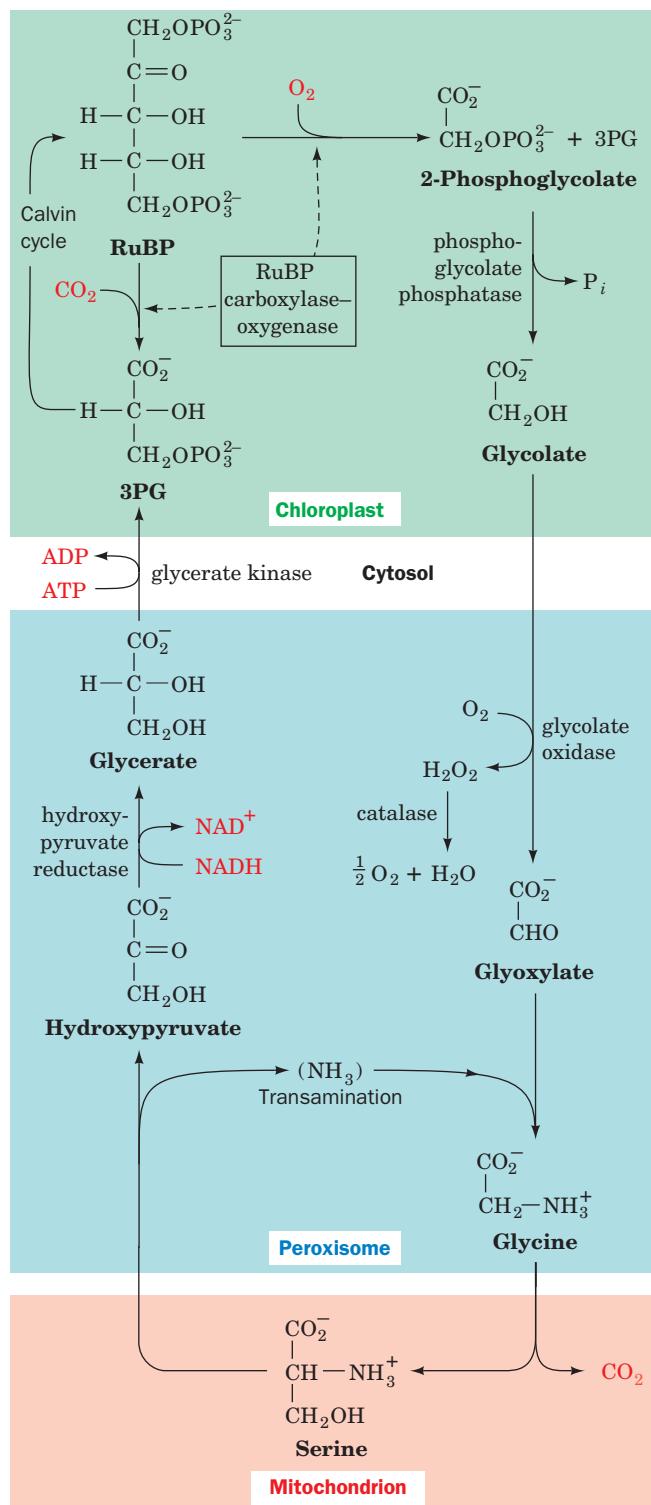


Figure 24-40 Photorespiration. This pathway metabolizes the phosphoglycolate produced by the RuBP carboxylase-catalyzed oxidation of RuBP. The reactions occur, as indicated, in the chloroplast, the peroxisome, the mitochondrion, and the cytosol. Note that two glycines are required to form serine + CO_2 (Section 26-3B).

atmospheric concentration of CO_2 is 330 ppm), so their photosynthetic CO_2 fixation usually dominates their photorespiratory CO_2 release. However, the CO_2 compensation point increases with temperature because the oxygenase activity of RuBisCO increases more rapidly with temperature than does its carboxylase activity. Thus, *on a hot bright day, when photosynthesis has depleted the level of CO_2 at the chloroplast and raised that of O_2 , the rate of photorespiration may approach that of photosynthesis. This phenomenon is, in fact, a major limiting factor in the growth of many plants*. Indeed, plants possessing a RuBisCO with significantly less oxygenase activity would not only have increased photosynthetic efficiency but would need less water because they could spend less time with their **stomata** (the pores leading to their internal leaf spaces) open acquiring CO_2 and would have a reduced need for fertilizer because they would require less RuBisCO. The control of photorespiration is therefore an important unsolved agricultural problem that is presently being attacked through genetic engineering studies (Section 5-5).

c. C_4 Plants Concentrate CO_2

Certain species of plants, such as sugarcane, corn, and most important weeds, have a metabolic cycle that concentrates CO_2 in their photosynthetic cells, thereby almost totally preventing photorespiration (their CO_2 compensation points are in the range 2 to 5 ppm). The leaves of plants that have this so-called **C_4 cycle** have a characteristic anatomy. Their fine veins are concentrically surrounded by a single layer of so-called **bundle-sheath cells**, which in turn are surrounded by **mesophyll cells**.

The C_4 cycle (Fig. 24-41) was elucidated in the 1960s by Marshall Hatch and Roger Slack. It begins with the uptake of atmospheric CO_2 by the mesophyll cells, which, lacking RuBisCO in their chloroplasts, do so by condensing it as HCO_3^- with phosphoenolpyruvate (PEP) to yield oxaloacetate. The oxaloacetate is reduced by NADPH to **malate**, which is exported to the bundle-sheath cells (the name C_4 refers to these four-carbon acids). There the malate is oxidatively decarboxylated by NADP^+ to form CO_2 , pyruvate, and NADPH. The CO_2 , which has been concentrated by this process, enters the Calvin cycle. The pyruvate is returned to the mesophyll cells, where it is phosphorylated to again form PEP. The enzyme that mediates this reaction, **pyruvate-phosphate dikinase**, has the unusual action of activating a phosphate group through the hydrolysis of ATP to AMP + PP_i . This PP_i is further hydrolyzed to two P_i , which is tantamount to the consumption of a second ATP. CO_2 is thereby concentrated in the bundle-sheath cells at the expense of two ATPs per CO_2 . The dark reactions of photosynthesis in C_4 plants therefore consume a total of five ATPs per CO_2 fixed versus the three ATPs required by the Calvin cycle alone. The additional ATP is presumably generated through the cyclic flow of electrons in the light reactions (Section 24-2Ch).

C_4 plants, which comprise ~5% of terrestrial plants, occur largely in unshaded areas of tropical regions because they grow faster under hot and sunny conditions than other, so-called **C_3 plants** (so named because they initially fix CO_2 in the form of three-carbon acids). In cooler climates, where photorespiration is less of a burden, C_3 plants have the advantage because they require less energy to fix CO_2 .

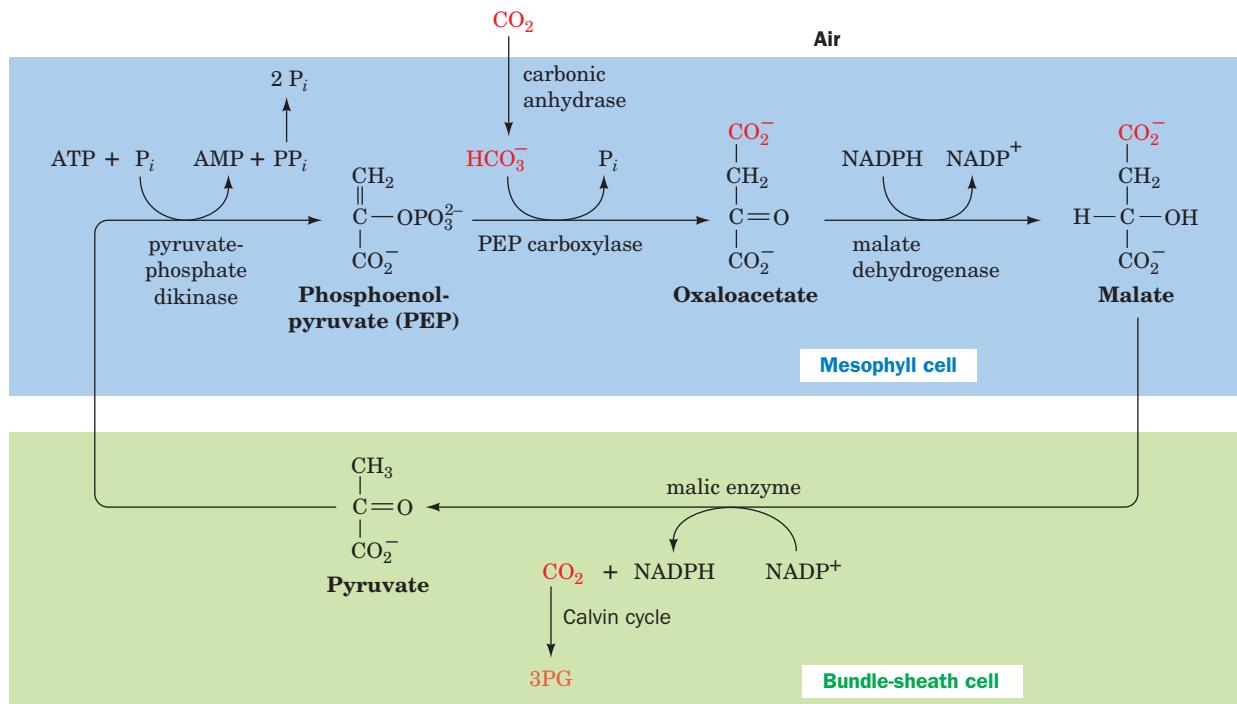


Figure 24-41 The C_4 pathway. CO_2 is concentrated in the mesophyll cells and transported to the bundle-sheath cells for entry into the Calvin cycle.

d. CAM Plants Store CO₂ through a Variant of the C₄ Cycle

Desert-adapted plants known as succulents (e.g., cacti) employ a variant of the C₄ cycle that separates CO₂ acquisition and the Calvin cycle in time rather than in space. If, as most plants, they opened their stomata by day to acquire CO₂, they would simultaneously transpire (lose by evaporation) what for them would be unacceptable amounts of water. To minimize this loss, these succulents only absorb CO₂ at night when the temperature is relatively cool. They store

this CO₂, in a process known as **crassulacean acid metabolism (CAM)**; so named because it was first discovered in plants of the family Crassulaceae), by the synthesis of malate through the reactions of the C₄ pathway (Fig. 24-41). The large amount of PEP necessary to store a day's supply of CO₂ is obtained by the breakdown of starch via glycolysis. During the course of the day, this malate is broken down to CO₂, which enters the Calvin cycle, and pyruvate, which is used to resynthesize starch. CAM plants are able, in this way, to carry out photosynthesis with minimal water loss.

CHAPTER SUMMARY

1 Chloroplasts Photosynthesis is the light-driven fixation of CO₂ to form carbohydrates and other biological molecules. In plants, photosynthesis takes place in the chloroplast, which consists of an inner and an outer membrane surrounding the stroma, a concentrated enzyme solution in which the thylakoid membrane system is immersed. Photosynthesis occurs in two stages, the so-called light reactions in which light energy is harnessed to synthesize ATP and NADPH, and the dark reactions in which these products are used to drive the synthesis of carbohydrates from CO₂ and H₂O. The thylakoid membrane is the site of the photosynthetic light reactions, whereas the dark reactions take place in the stroma. The counterpart of the thylakoid in photosynthetic bacteria is a specialized portion of the plasma membrane named the chromatophore.

2 Light Reactions Chlorophyll is the principal photoreceptor of photosynthesis. Light is absorbed initially by light-harvesting complexes (LHCs) that contain chlorophyll and accessory pigments such as carotenoids. The resulting excitation then migrates via exciton transfer until it reaches the reaction center chlorophyll, where it is trapped. LH2 from purple photosynthetic bacteria is a transmembrane protein that consists of eight or nine rotationally related subunits that each bind three BChl *a* molecules and one carotenoid. LH1, which is similarly arranged but 16-fold symmetric, contains a central hole that binds a photosynthetic reaction center (RC). Light energy absorbed by LH2 is transmitted to LH1, which, in turn, transmits it to the RC.

The purple photosynthetic bacterial RC (PbRC) is a protein that consists of three subunits and several redox-active small molecules that are arranged as two pseudosymmetrically related chains of electron carriers. The primary photon absorbing species of the *Rps. viridis* bacterial reaction center is a special pair of BChl *b* molecules known as P960. By rapid measurement techniques it has been determined that the electron ejected by P960* passes by a third BChl *b* to a BPheo *b* molecule located in only one of the two chains (the other is apparently nonfunctional) and then sequentially to a menaquinone (Q_A) and a ubiquinone (Q_B). The resulting Q_B⁺ is subsequently further reduced in a second one-electron transfer process and then takes up two protons from the cytosol to form Q_BH₂. The electrons taken up by this species are returned to P960 via a cytochrome *bc*₁ complex, cytochrome *c*₂, and, in some purple photosynthetic bacteria, a four-heme *c*-type cytochrome associated with the photosynthetic reaction center. This cyclic electron-transport process functions to

translocate protons, via a Q cycle mediated by the cytochrome *bc*₁, from the cytoplasm to the outside of the cell. The resulting proton gradient, in a process known as photophosphorylation, drives the synthesis of ATP. Since bacterial photosynthesis does not generate the reducing equivalents needed in many biosynthetic processes, photosynthetic bacteria require an outside source of reducing agents such as H₂S.

In plants and cyanobacteria, the light reactions occur in two reaction centers, those of PSI and PSII, which are electrically “connected” in series. This enables the system to generate sufficient electromotive force to form NADPH by oxidizing H₂O in a noncyclic pathway known as the Z-scheme. PSI and PSII both contain core antenna systems and their RCs are evolutionarily related to each other and to the PbRC. PSII contains an Mn₄CaO₄ complex that oxidizes two H₂O molecules to four H⁺ and O₂ in four one-electron steps. The electrons are passed singly, through a Tyr side chain named Z, to photooxidized P680, the reaction center's photon-absorbing species, a special pair that consists of two Chl *a* molecules. The electron previously ejected from P680* passes through a series of carriers resembling those of the PbRC to a pool of plastoquinone molecules. The electrons then enter the cytochrome *b*₆*f* complex, which transports protons, via a Q cycle, from the stroma to the thylakoid space. These electrons are transferred individually, by a plastocyanin carrier, directly to PSI's photooxidized photon-absorbing pigment, P700, a pair of Chl *a*'s that resembles the PbRC's special pair. The electron that had been previously ejected from P700* migrates through both sides of a bifurcated chain of Chl *a* molecules and then through a chain of three [4Fe–4S] clusters to a soluble ferredoxin (Fd) that contains a [2Fe–2S] cluster. The electron then reduces NADP⁺ in a noncyclic process mediated by ferredoxin–NADP⁺ reductase. Alternatively, it may be returned, presumably via ferredoxin–plastoquinone reductase, to the plastoquinone pool in a cyclic process that does not require electron input from PSII and only translocates protons across the thylakoid membrane. ATP is synthesized by the CF₁CF₀-ATP synthase, which closely resembles the analogous mitochondrial complex, in a reaction driven by the dissipation of the proton gradient across the thylakoid membrane.

3 Dark Reactions CO₂ is fixed in the photosynthetic dark reactions of plants and cyanobacteria by the reactions of the Calvin cycle. The first stage of the Calvin cycle, in sum, mediates the reaction 3RuBP + 3CO₂ → 6GAP with the consumption of 9 ATP and 6 NADPH generated by the light reactions. The second stage reshuffles the atoms of five GAPs to

reform the three RuBPs with which the cycle began, a process that requires no further input of free energy or reduction equivalents. The sixth GAP, the product of the Calvin cycle, is used to synthesize carbohydrates, amino acids, and fatty acids. The flux-controlling enzymes of the Calvin cycle are activated in the light through variations in the pH and the Mg^{2+} and NADPH concentrations, and by the redox level of thioredoxin. The central enzyme of the Calvin cycle, RuBP carboxylase, catalyzes both a carboxylase and an oxygenase reaction with RuBP. The latter reaction is the first step in the photorespiration cycle that liberates CO_2 . The rate of photorespiration

increases with temperature and decreases with CO_2 concentration, so photorespiration constitutes a significant energetic drain on most plants on hot bright days. Calvin cycle products are converted to sucrose, starch, and cellulose, as well as fatty acids and amino acids. C_4 plants, which are most common in the tropics, have a system for concentrating CO_2 in their photosynthetic cells so as to minimize the effects of photorespiration but at the cost of 2 ATP per CO_2 fixed. Certain desert plants conserve water by absorbing CO_2 at night and releasing it to the Calvin cycle by day. This crassulacean acid metabolism (CAM) occurs through a process similar to the C_4 cycle.

REFERENCES

General

- Blankenship, R.E., *Molecular Mechanisms of Photosynthesis*, Blackwell Science (2002).
- Buchanan, B.B., Gruissem, W., and Jones, R.L. (Eds.), *Biochemistry and Molecular Biology of Plants*, American Society of Plant Physiologists (2000).
- Hall, D.O. and Rao, K.K., *Photosynthesis* (6th ed.), Cambridge (1999).
- Heldt, H.-W., *Plant Biochemistry*, Elsevier (2005).
- Lawlor, D.W., *Photosynthesis* (3rd ed.), BIOS Scientific Publishers Ltd. (2001).
- Nicholls, D.G. and Ferguson, S.J., *Bioenergetics 3*, Chapter 6, Academic Press (2002).

Chloroplasts

- Bogorad, L. and Vasil, I.K. (Eds.), *The Molecular Biology of Plastids*, Academic Press (1991).
- Hoover, J.K., *Chloroplasts*, Plenum Press (1984).

Light Reactions

- Amunts, A., Drory, O., and Nelson, N., The structure of plant photosystem I at 3.4 Å resolution, *Nature* **447**, 58–63 (2007).
- Barber, J., Photosystem II: a multisubunit membrane protein that oxidises water, *Curr. Opin. Struct. Biol.* **12**, 523–530 (2002).
- Chitnis, P.R., Photosystem I: Function and physiology, *Annu. Rev. Plant Physiol. Plant Biol.* **52**, 593–626 (2001).
- Deisenhofer, J., Epp, O., Sining, I., and Michel, H., Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*, *J. Mol. Biol.* **246**, 429–457 (1995).
- Deisenhofer, J. and Michel, H., High-resolution structures of photosynthetic reaction centers, *Annu. Rev. Biophys. Biophys. Chem.* **20**, 247–266 (1991); and Structures of bacterial photosynthetic reaction centers, *Annu. Rev. Cell Biol.* **7**, 1–23 (1991).
- Deng, Z., Aliverti, A., Zanetti, G., Arakaki, A.K., Ottado, J., Orellano, E.G., Calcaterra, N.B., Ceccarelli, E.A., Carrillo, N., and Karplus, P.A., A productive $NADP^+$ binding mode of ferredoxin– $NADP^+$ reductase revealed by protein engineering and crystallographic studies, *Nature Struct. Biol.* **6**, 847–853 (1999); and Bruns, C.M. and Karplus, P.A., Refined crystal structure of spinach ferredoxin reductase at 1.7 Å resolution: Oxidized, reduced, and 2'-phospho-5'-AMP bound states, *J. Mol. Biol.* **247**, 125–145 (1995).
- Diner, B.A. and Rappaport, F., Structure, dynamics, and energetics of the primary photochemistry of photosystem II of oxygenic photosynthesis, *Annu. Rev. Plant Biol.* **53**, 551–580 (2002).

Eberhard, S., Finazzi, G., and Wollman, F.-A., The dynamics of photosynthesis, *Annu. Rev. Genet.* **42**, 463–515 (2008).

El-Kabbani, O., Chang, C.-H., Tiede, D., Norris, J., and Schiffer, M., Comparison of reaction centers from *Rhodobacter sphaeroides* and *Rhodopseudomonas viridis*: Overall architecture and protein-pigment interactions, *Biochemistry* **30**, 5361–5369 (1991).

Fromme, P. (Ed.), *Photosynthetic Protein Complexes. A Structural Approach*, Wiley–Blackwell (2008).

Guskov, A., Kern, J., Gabdulkhakov, A., Broser, M., Zouni, A., and Saenger, W., Cyanobacterial photosystem II at 2.9-Å resolution and the roles of quinones, lipids, channels, and chloride, *Nature Struct. Mol. Biol.* **16**, 334–342 (2009); Loll, B., Kern, J., Saenger, W., Zouni, A., and Biesiadka, J., Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II, *Nature* **438**, 1040–1044 (2005); and Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S., Architecture of the photosynthetic oxygen-evolving center, *Science* **303**, 1831–1838 (2004). [X-ray structures of PSII.]

Heathcote, P., Fyfe, P.K., and Jones, M.R., Reaction centers: The structure and evolution of biological solar power, *Trends Biochem. Sci.* **27**, 79–87 (2002).

Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N., Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution, *Nature* **411**, 909–917 (2001).

Koepke, J., Hu, X., Muenke, C., Schulen, K., and Michel, H., The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum molischianum*, *Structure* **4**, 581–597 (1996); and McDermott, G., Prince, S.M., Freer, A.A., Horthornthwaite-Lawless, A.M., Papiz, M.Z., Cogdell, R.J., and Isaacs, N.W., Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria, *Nature* **374**, 517–521 (1995). [The X-ray structures of LH2s.]

Kurisu, G., Kusunoki, M., Katoh, E., Yamazaki, T., Teshima, K., Onda, Y., Kimata-Ariga, Y., and Hase, T., Structure of the electron transfer complex between ferredoxin and ferredoxin– $NADP^+$ reductase, *Nature Struct. Biol.* **8**, 117–121 (2001).

Kurisu, G., Zhang, H., Smith, J.L., and Cramer, W.A., Structure of the cytochrome b_6f complex of oxygenic photosynthesis: Tuning the cavity, *Science* **302**, 1009–1014 (2003); Stroebel, D., Choquet, Y., Popot, J.-L., and Picot, D., An atypical haem in the cytochrome b_6f complex, *Nature* **426**, 413–418 (2003); and Cramer, W.A., Zhang, H., Yan, J., Kurisu, G., and Smith, J.L., Transmembrane traffic in the cytochrome b_6f complex, *Annu. Rev. Biochem.* **75**, 769–790 (2006).

Nelson, N. and Yocum, C.F., Structure and function of photosystems I and II, *Annu. Rev. Plant Biol.* **57**, 521–565 (2006).

- Renger, G. and Tenger, T., Photosystem II: The machinery of photosynthetic water splitting, *Photosyn. Res.* **98**, 53–80 (2008).
- Ruban, A.V., et al., Identification of a mechanism of photoprotective energy dissipation in higher plants, *Nature* **450**, 575–578 (2007).
- Standfuss, J., van Scheltinga, A.C.T., Lamborghini, M., and Kühlbrandt, W., Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution, *EMBO J.* **24**, 919–928 (2005). [The X-ray structure of LHC-II.]
- Yano, J., et al., Where water is oxidized to dioxygen: Structure of the photosynthetic Mn₄Ca cluster, *Science* **314**, 821–825 (2006).

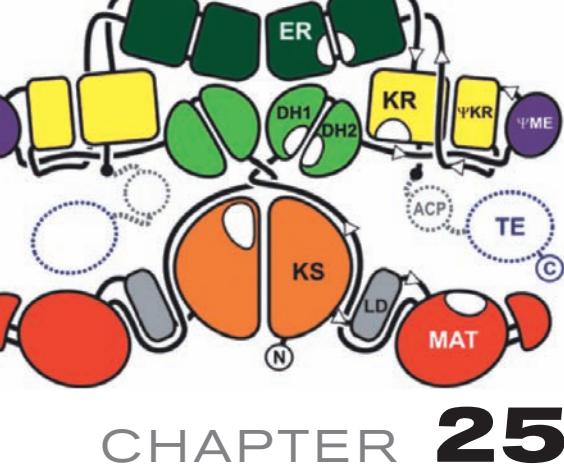
Dark Reactions

- Black, C.C. and Osmond, C.B., Crassulacean acid metabolism photosynthesis: ‘working the night shift,’ *Photosynthesis Res.* **76**, 329–331 (2003).
- Cushman, J.C. and Bohnert, H.J., Crassulacean acid metabolism: Molecular genetics, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 305–332 (1999).
- Dai, S., Friemann, R., Glauser, D.A., Bourquin, F., Manieri, W., Schürmann, P., and Eklund, H., Structural snapshots along the reaction pathway of ferredoxin-thioredoxin reductase, *Nature* **448**, 92–96 (2007).
- Hartman, F.C. and Harpel, M.R., Chemical and genetic probes of the active site of D-ribulose-1,5-bisphosphate carboxylase/oxygenase: A retrospective based on the three-dimensional structure, *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 1–75 (1993).

- Hatch, M.D., C₄ photosynthesis: A unique blend of modified biochemistry, anatomy, and ultrastructure, *Biochim. Biophys. Acta* **895**, 81–106 (1987).
- Portis, A.R., Jr., Regulation of ribulose 1,5-bisphosphate carboxylase/oxygenase activity, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 415–437 (1992); and Rubisco activase, *Biochim. Biophys. Acta* **1015**, 15–28 (1990).
- Saxena, I.M. and Brown, R.M., Jr., Cellulose biosynthesis: current views and evolving concepts, *Ann. Bot.* **96**, 9–21 (2005).
- Schneider, G., Lindqvist, Y., and Branden, C.-I., RUBISCO: Structure and mechanism, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 119–143 (1992).
- Schreuder, H.A., Knight, S., Curmi, P.M.G., Andersson, I., Cascio, D., Sweet, R.M., Brändén, C.-I., and Eisenberg, D., Crystal structure of activated tobacco rubisco complexed with the reaction-intermediate analogue 2-carboxy-arabinitol 1,5-bisphosphate, *Protein Sci.* **2**, 1136–1146 (1993).
- Schürmann, P., Redox signaling in the chloroplast: The ferredoxin/thioredoxin system, *Antioxidants Redox Signaling* **5**, 69–79 (2003).
- Spreitzer, R.J. and Salvucci, M.E., Rubisco: structure, regulatory interactions, and possibilities for a better enzyme, *Annu. Rev. Plant Biol.* **53**, 449–475 (2002).
- Taylor, T.C. and Andersson, I., The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate, *J. Mol. Biol.* **265**, 432–444 (1997).

PROBLEMS

1. Why is chlorophyll green in color when it absorbs in the red and the blue regions of the spectrum (Fig. 24-5)?
2. The “red tide” is a massive proliferation of certain algal species that cause seawater to become visibly red. Describe the spectral characteristics of the dominant photosynthetic pigments in these algae.
3. H₂¹⁸O is added to a suspension of chloroplasts capable of photosynthesis. Where does the label appear when the suspension is exposed to light?
4. Indicate, where appropriate, the analogous components in the photosynthetic electron-transport chains of purple photosynthetic bacteria and chloroplasts.
5. Antimycin inhibits photosynthesis in chloroplasts. Indicate its most likely site of action and explain your reasoning.
6. Calculate the energy efficiency of cyclic and noncyclic photosynthesis in chloroplasts using 680-nm light. What would this efficiency be with 500-nm light? Assume that ATP formation requires 59 kJ · mol⁻¹ under physiological conditions.
- *7. What is the minimum pH gradient required to synthesize ATP from ADP + P_i? Assume [ATP]/([ADP][P_i]) = 10³, T = 25°C, and that three protons must be translocated per ATP generated. (See Table 16-3 for useful thermodynamic information.)
8. Indicate the average Calvin cycle labeling pattern in ribulose-5-phosphate after two rounds of exposure to ¹⁴CO₂.
9. Chloroplasts are illuminated until the levels of their Calvin cycle intermediates reach a steady state. The light is then turned off. How do the levels of RuBP and 3PG vary after this time?
10. What is the energy efficiency of the Calvin cycle combined with glycolysis and oxidative phosphorylation; that is, what percentage of the input energy can be metabolically recovered in synthesizing starch from CO₂ using photosynthetically produced NADPH and ATP rather than somehow directly storing these “high-energy” intermediates? Assume that each NADPH is energetically equivalent to three ATPs and that starch synthesis and breakdown are energetically equivalent to glycogen synthesis and breakdown.
11. Predict the effect of an uncoupler such as dinitrophenol (Fig. 22-47) on production of (a) ATP and (b) NADPH in a chloroplast.
12. Describe the effects of an increase in oxygen pressure on the dark reactions of photosynthesis.
13. If a C₃ plant and a C₄ plant are placed together in a sealed illuminated box with sufficient moisture, the C₄ plant thrives while the C₃ plant sickens and eventually dies. Explain.
14. The leaves of some species of desert plants taste sour in the early morning but, as the day wears on, they become tasteless and then bitter. Explain.



Lipid Metabolism

CHAPTER 25

1 Lipid Digestion, Absorption, and Transport

2 Fatty Acid Oxidation

- A. Fatty Acid Activation
- B. Transport Across the Mitochondrial Membrane
- C. β Oxidation
- D. Oxidation of Unsaturated Fatty Acids
- E. Oxidation of Odd-Chain Fatty Acids
- F. Peroxisomal β Oxidation
- G. Minor Pathways of Fatty Acid Oxidation

3 Ketone Bodies

4 Fatty Acid Biosynthesis

- A. Pathway Overview
- B. Acetyl-CoA Carboxylase
- C. Fatty Acid Synthase
- D. Transport of Mitochondrial Acetyl-CoA Into the Cytosol
- E. Elongases and Desaturases
- F. Synthesis of Triacylglycerols

5 Regulation of Fatty Acid Metabolism

6 Cholesterol Metabolism

- A. Cholesterol Biosynthesis
- B. Control of Cholesterol Biosynthesis and Transport
- C. Cholesterol Utilization

7 Eicosanoid Metabolism: Prostaglandins, Prostacyclins, Thromboxanes, Leukotrienes, and Lipoxins

- A. Background
- B. The Cyclic Pathway of Eicosanoid Metabolism: Prostaglandins, Prostacyclins, and Thromboxanes
- C. The Linear Pathway of Eicosanoid Metabolism: Leukotrienes and Lipoxins

8 Phospholipid and Glycolipid Metabolism

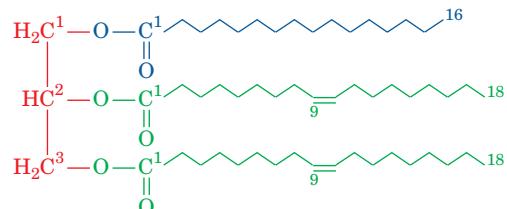
- A. Glycerophospholipids
- B. Sphingophospholipids
- C. Sphingoglycolipids

Lipids play indispensable roles in cell structure and metabolism. For example, triacylglycerols are the major storage form of metabolic energy in animals; cholesterol is a vital component of cell membranes and a precursor of the steroid hormones and bile salts; arachidonate, a C_{20} unsaturated fatty acid, is the precursor of the prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins, potent intercellular mediators that control a variety of complex processes; and complex glycolipids and phospholipids are major components of biological membranes. We dis-

cussed the structures of simple and complex lipids in Section 12-1. In the first half of this chapter, we consider the metabolism of fatty acids and triacylglycerols, including their digestion, oxidation, and biosynthesis. We then consider how cholesterol is synthesized and utilized, and how prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins are synthesized. We end by studying how complex glycolipids and phospholipids are synthesized from their simpler lipid and carbohydrate components.

1 LIPID DIGESTION, ABSORPTION, AND TRANSPORT

Triacylglycerols (also called *fats* or *triglycerides*) constitute ~90% of the dietary lipid and are the major form of metabolic energy storage in humans. Triacylglycerols consist of glycerol triesters of fatty acids such as palmitic and oleic acids



1-Palmitoyl-2,3-dioleoyl-glycerol

(the names and structural formulas of some biologically common fatty acids are listed in Table 12-1). Like glucose, they are metabolically oxidized to CO_2 and H_2O . Yet, since most carbon atoms of triacylglycerols have lower oxidation states than those of glucose, the oxidative metabolism of fats yields over twice the energy of an equal weight of dry carbohydrate or protein (Table 25-1). Moreover, fats, being nonpolar, are stored in an anhydrous state, whereas glycogen, the storage form of glucose, is polar and is consequently stored in a hydrated form that contains about twice its dry weight of water. Fats therefore provide up to six times the metabolic energy of an equal weight of hydrated glycogen.

a. Lipid Digestion Occurs at Lipid-Water Interfaces

Since triacylglycerols are water insoluble, whereas digestive enzymes are water soluble, *triacylglycerol digestion takes place at lipid-water interfaces*. The rate of triacylglyc-

Table 25-1 Energy Content of Food Constituents

Constituent	ΔH (kJ · g ⁻¹ dry weight)
Carbohydrate	16
Fat	37
Protein	17

Source: Newsholme, E.A. and Leech, A.R., *Biochemistry for the Medical Sciences*, p. 16, Wiley (1983).

erol digestion therefore depends on the surface area of the interface, a quantity that is greatly increased by the churning peristaltic movements of the intestine combined with the emulsifying action of **bile salts**. These are powerful digestive detergents that, as we shall see in Section 25-6C, are synthesized by the liver and secreted via the gallbladder into the small intestine where lipid digestion and absorption mainly take place.

b. Pancreatic Lipase Requires Activation and Has a Catalytic Triad

Pancreatic **lipase** (**triacylglycerol lipase**) catalyzes the hydrolysis of triacylglycerols at their 1 and 3 positions to form sequentially **1,2-diacylglycerols** and **2-acylglycerols**, together with the Na⁺ and K⁺ salts of fatty acids (soaps). These soaps, being amphipathic, aid in the lipid emulsification process.

The enzymatic activity of pancreatic lipase greatly increases when it contacts the lipid–water interface, a phenomenon known as **interfacial activation**. Binding to the lipid–water interface requires the presence of mixed micelles of phosphatidylcholine (Fig. 12-4) and bile salts, as well as the pancreatically produced protein named **colipase**, which forms a 1:1 complex with lipase. This complex aids in the adsorption of the enzyme to emulsified oil droplets as well as stabilizes the enzyme in an active conformation. The X-ray structures, determined by Christian Cambillau, of pancreatic lipase–colipase complexes, alone and cocrystallized with mixed micelles of phosphatidylcholine and bile salts, have revealed the structural basis of lipase activation as well as how colipase and micelles aid lipase in binding to the lipid–water interface (Fig. 25-1).

The active site of the 449-residue pancreatic lipase, which is contained in the enzyme's N-terminal domain, has a catalytic triad that closely resembles that in the serine proteases (Section 15-3B; recall that ester hydrolysis is mechanistically similar to peptide hydrolysis). In aqueous solution (Fig. 25-1a), the lipase's active site is covered by a 26-residue helical lid. However, in the presence of the mixed micelles (Fig. 25-1b), the lid undergoes a complex structural reorganization that exposes the active site; causes a contacting 10-residue loop, the β 5 loop, to change conformation in a way that forms the active enzyme's oxyanion hole; and generates a hydrophobic surface about the entrance to the active site. Indeed, the active site of the mixed micelle-containing complex contains a long rod of electron density that contacts the catalytic triad's Ser residue and appears to be a phosphatidylcholine molecule.

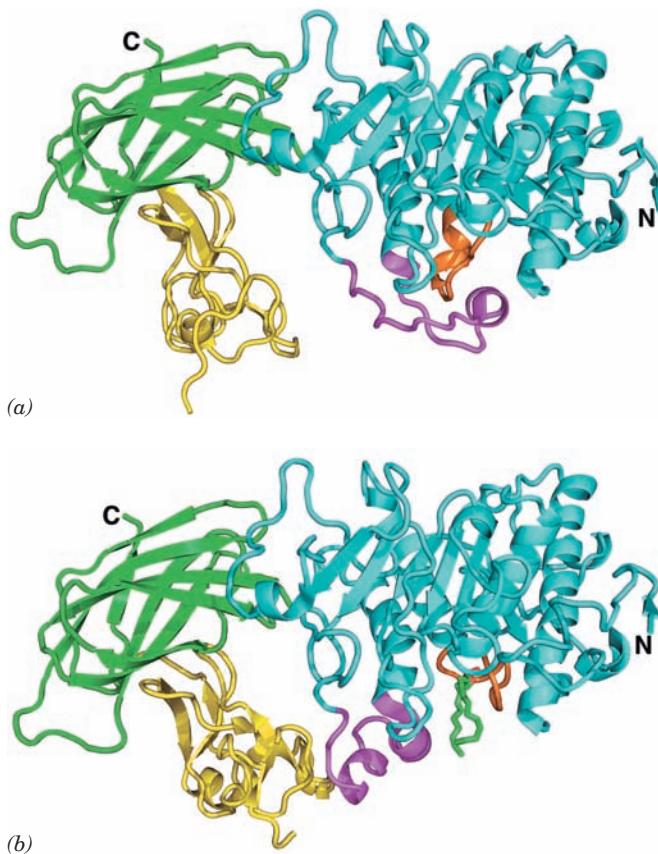


Figure 25-1 X-ray structures of pancreatic lipase in complex with colipase. (a) In aqueous solution, and (b) cocrystallized with mixed micelles of phosphatidylcholine and bile salts. The lipase is drawn in ribbon form with its N-terminal domain (residues 1–336) cyan, its C-terminal domain (residues 337–449) green, the lid (residues 237–262) magenta, and the β 5 loop (residues 76–85) orange. The colipase is yellow. A phosphatidylcholine molecule that is bound in the lipase active site in Part b is shown in stick form with C green, O red, and P orange. The micelles, which have irregular structures (e.g., Fig. 2-9), are not visible. [Based on X-ray structures by Christian Cambillau, LCCMB-CNRS, Marseille, France. PDBIDs 1N8S and 1LPB.]

Colipase binds to the C-terminal domain of lipase such that the hydrophobic tips of the three loops that comprise much of this 90-residue protein extend from the complex on the same face as lipase's active site. A continuous hydrophobic plateau is thereby created that extends over a distance of >50 Å past the active site (bottom of Fig. 25-1b) and that, presumably, helps bind the complex to the lipid surface. In the presence of the mixed micelles, colipase changes conformation so as to form three hydrogen bonds to the opened lid, thereby stabilizing it in this conformation.

The mixed micelles are not visible in the X-ray structure. However, neutron diffraction studies, by Juan Fontecilla-Camps, of crystals of a lipase–colipase–micelle complex in which the lipase is in its active conformation reveal that the activating micelle interacts, not with the substrate site, but with the concave face of colipase and the adjacent tip of the lipase's C-terminal domain (left side of

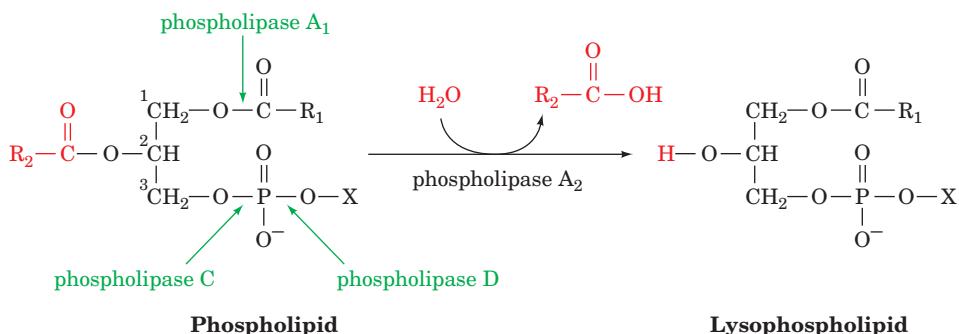


Figure 25-2 Catalytic action of phospholipase A₂. Phospholipase A₂ hydrolytically excises the C2 fatty acid residue from a phospholipid to yield the corresponding lysophospholipid. The

bonds hydrolyzed by other types of phospholipases, which are named according to their specificities, are also indicated.

Fig. 25-1b). Apparently, micelle binding and substrate binding involve different regions of the lipase–colipase complex. Hence, strictly speaking, lipase activation appears not to be interfacial but, instead, occurs in the aqueous phase and requires the binding of colipase and a micelle.

c. Pancreatic Phospholipase A₂ Has a Modified Catalytic Triad

Phospholipids are degraded by pancreatic **phospholipase A₂**, which hydrolytically excises the fatty acid residue at C2 to yield the corresponding **lysophospholipids** (Fig. 25-2), which are also powerful detergents. Indeed, the phospholipid lecithin (phosphatidylcholine) is secreted in the bile, presumably to aid in lipid digestion.

Phospholipase A₂, as does triacylglycerol lipase, preferentially catalyzes reactions at interfaces. However, as Paul Sigler's determinations of the X-ray structures of the phospholipases A₂ from cobra venom and bee venom revealed, its mechanism of interfacial activation differs from that of triacylglycerol lipase in that it does not change its conformation. Instead, phospholipase A₂ contains a hy-

drophobic channel that provides the substrate with direct access from the phospholipid aggregate (micelle or membrane) surface to the bound enzyme's active site. Hence, on leaving its micelle to bind to the enzyme, the substrate need not become solvated and then desolvated (Fig. 25-3). In contrast, soluble and dispersed phospholipids must first surmount these significant kinetic barriers in order to bind to the enzyme.

The catalytic mechanism of phospholipase A₂ also differs substantially from that of triacylglycerol lipase. Although the phospholipase A₂ active site contains the His and Asp components of a catalytic triad, an enzyme-bound water molecule occupies the position expected for an active site Ser. Moreover, the active site contains a bound Ca²⁺ ion and does not form an acyl–enzyme intermediate. Sigler therefore proposed that phospholipase A₂ catalyzes the direct hydrolysis of phospholipid with a His–Asp “catalytic dyad” activating an active site water molecule for nucleophilic attack on the ester, and with the Ca²⁺ ion stabilizing the oxyanion transition state. However, the subsequently determined X-ray structure, by Mahendra Jain and

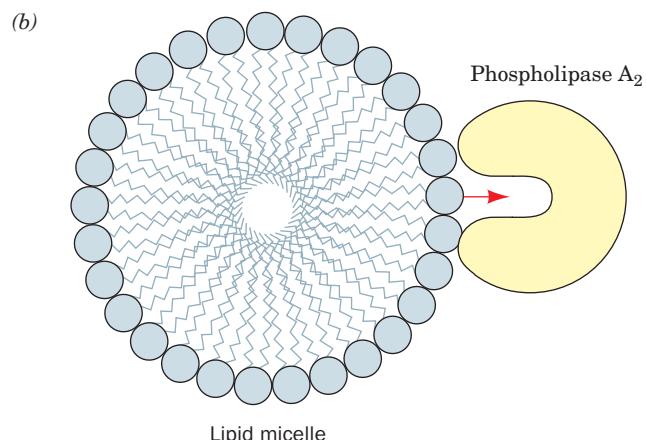
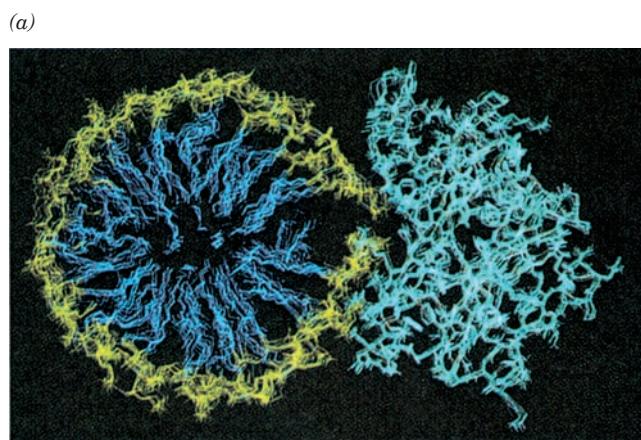
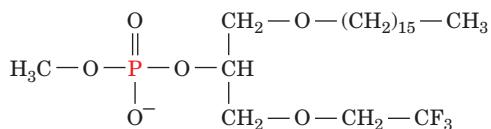


Figure 25-3 Substrate binding to phospholipase A₂. (a) A hypothetical model of phospholipase A₂ in complex with a micelle of lysophosphatidylethanolamine as shown in cross section. The protein is drawn in cyan, the phospholipid head groups are yellow, and their hydrocarbon tails are blue. The calculated atomic

motions of the assembly are indicated through a series of superimposed images taken at 5-ps intervals. [Courtesy of Raymond Salemme, E.I. du Pont de Nemours & Company.] (b) Schematic diagram of a productive interaction between phospholipase A₂ and a phospholipid contained in a micelle.

Brian Bahnson, of phospholipase A₂ in complex with the tetrahedral intermediate mimic MJ33



MJ33 [1-Hexadecyl-3-(trifluoroethyl)-*sn*-glycero-2-phosphomethanol]

suggests that a second, previously unobserved water molecule, which is liganded by the Ca²⁺ ion, is the attacking nucleophile (Fig. 25-4a). This has led to the formulation of a reaction mechanism (Fig. 25-4b) in which the Asp–His–water catalytic triad and the Ca²⁺ ion both activate the second water molecule, with the Ca²⁺ ion also stabilizing the resulting tetrahedral intermediate.

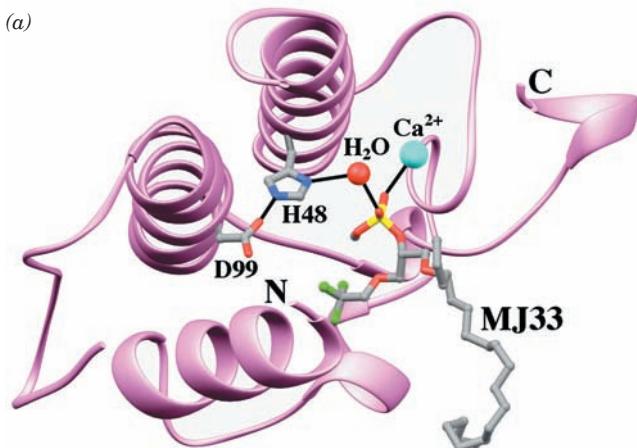
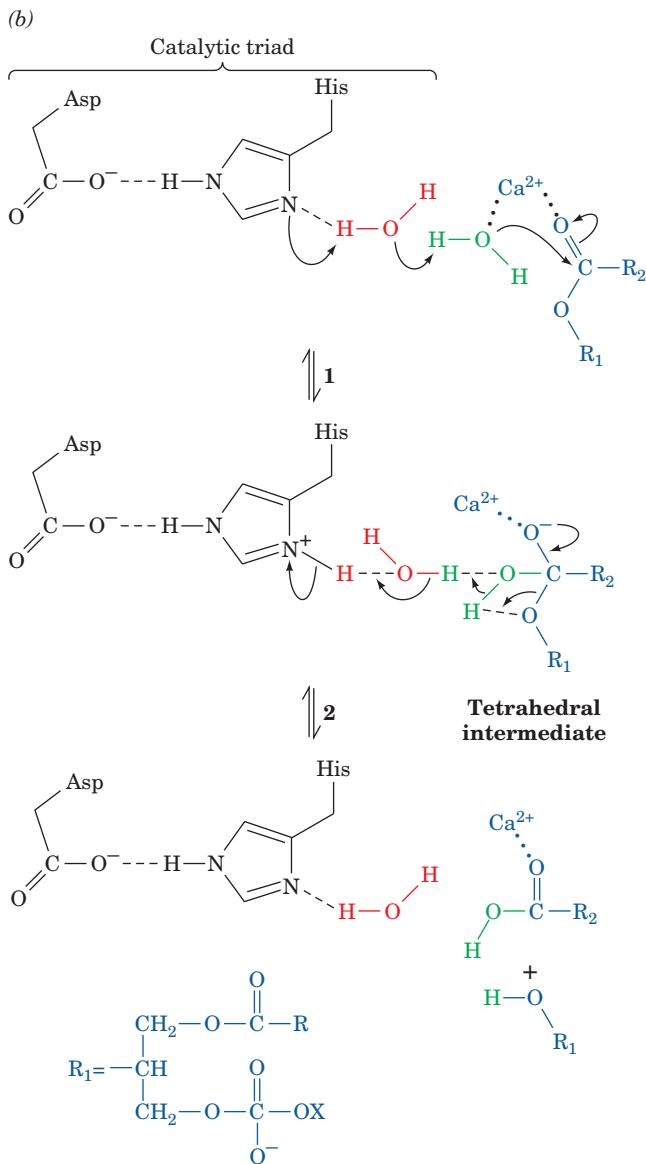


Figure 25-4 Structure and mechanism of phospholipase A₂. (a) The X-ray structure of the 124-residue monomeric porcine phospholipase A₂ (*lavender*) in complex with the tetrahedral intermediate mimic MJ33. The enzyme's active site contains a catalytic triad similar to those of the serine proteases (Fig. 15-20) with a water molecule replacing the catalytic Ser. The His 48 and Asp 99 side chains of the catalytic triad together with MJ33 are drawn in stick form colored according to atom type (C gray, N blue, O red, F green, and P yellow). The water molecule of the catalytic triad and the catalytically important Ca²⁺ ion are represented by red and cyan spheres. Catalytically important hydrogen bonds and Ca²⁺ liganding interactions are represented by thin black lines. The tetrahedral phosphoryl group of MJ33 presumably occupies the site of the unobserved H₂O. Residues 65 to 74 of the protein have been deleted for clarity. [Based on an X-ray structure by Mahendra Jain and Brian Bahnson, University of Delaware. PDBid 1FXF.] (b) The catalytic mechanism of phospholipase A₂. (1) The catalytic triad activates a second water molecule to attack the scissile carbonyl carbon with Ca²⁺ coordinating the activated water molecule as well as electrostatically stabilizing the resulting tetrahedral intermediate (rather than doing so via nucleophilic catalysis as occurs in the serine proteases; Fig. 15-23). (2) The tetrahedral intermediate decomposes to yield products. [After Berg, O.G., Gelb, M.H., Tsai, M.-D., and Jain, M.K., *Chem. Rev.* **101**, 2638 (2001).]

d. Bile Salts and Fatty Acid–Binding Protein Facilitate the Intestinal Absorption of Lipids

The mixture of fatty acids and mono- and diacylglycerols produced by lipid digestion is absorbed by the cells lining the small intestine (the intestinal mucosa) in a process facilitated by bile salts. The micelles formed by the bile salts take up the nonpolar lipid degradation products so as to permit their transport across the unstirred aqueous boundary layer at the intestinal wall. The importance of this process is demonstrated in individuals with obstructed bile ducts: They absorb little of their dietary lipids but, rather, eliminate them in hydrolyzed form in their feces (**steatorrhea**). Evidently, *bile salts are not only an aid to lipid digestion but are essential for the absorption of lipid digestion products*. Bile salts are likewise required for the efficient intestinal absorption of the lipid-soluble vitamins A, D, E, and K.

Inside the intestinal cells, fatty acids form complexes with **intestinal fatty acid–binding protein (I-FABP)**, a cyto-



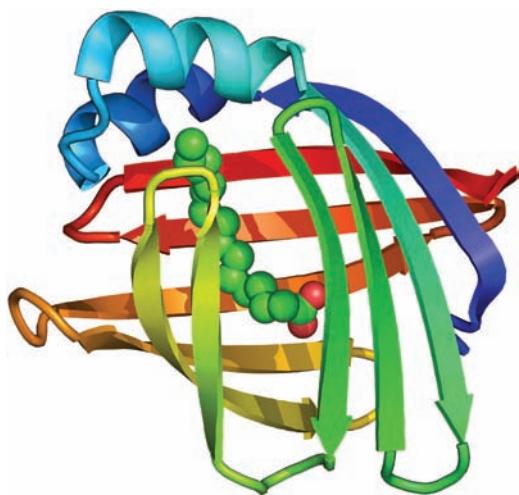


Figure 25-5 X-ray structure of rat intestinal fatty acid-binding protein in complex with palmitate. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The palmitate is shown in space-filling form with C green and O red. [Based on an X-ray structure by James Sacchettini, Albert Einstein College of Medicine. PDBid 2IFB.]

plasmic protein, which serves to increase the effective solubility of these water-insoluble substances and also to protect the cell from their detergent-like effects (recall that soaps are fatty acid salts). The X-ray structures of rat I-FABP, both alone and in complex with a single molecule of palmitate, were determined by James Sacchettini. This monomeric, 131-residue protein consists largely of 10 antiparallel β strands organized into a stack of two approximately orthogonal β sheets (Fig. 25-5). The palmitate occupies a gap between two of the β strands such that it lies between the β sheets with an orientation that, over much of its length, is more or less parallel to the gapped β strands (this structure has therefore been described as forming a “ β -clam”). The palmitate’s carboxyl group interacts with Arg 106, Gln 115, and two bound water molecules, whereas the methylene chain is encased by the side chains of several hydrophobic, mostly aromatic, residues.

e. Lipids Are Transported in Lipoprotein Complexes

The lipid digestion products absorbed by the intestinal mucosa are converted by these tissues to triacylglycerols (Section 25-4F) and then packaged into lipoprotein parti-

cles called **chylomicrons**. These, in turn, are released into the bloodstream via the lymph system for delivery to the tissues. Similarly, triacylglycerols synthesized by the liver are packaged into **very low density lipoproteins (VLDL)** and released directly into the blood. These lipoproteins, whose origins, structures, and functions are discussed in Section 12-5, maintain their otherwise insoluble lipid components in aqueous solution.

The triacylglycerol components of chylomicrons and VLDL are hydrolyzed to free fatty acids and glycerol in the capillaries of adipose tissue and skeletal muscle by **lipoprotein lipase** (Section 12-5Ba). The resulting free fatty acids are taken up by these tissues while the glycerol is transported to the liver or kidneys. There it is converted to the glycolytic intermediate dihydroxyacetone phosphate by the sequential actions of **glycerol kinase** and **glycerol-3-phosphate dehydrogenase** (Fig. 25-6).

Mobilization of triacylglycerols stored in adipose tissue involves their hydrolysis to glycerol and free fatty acids by **hormone-sensitive triacylglycerol lipase** (or just **hormone-sensitive lipase**). The free fatty acids are released into the bloodstream, where they bind to **serum albumin** (or just **albumin**), a soluble 585-residue monomeric protein that comprises about half of the blood serum protein. In the absence of albumin, the maximum solubility of free fatty acids is $\sim 10^{-6} M$. Above this concentration, free fatty acids form micelles that act as detergents to disrupt protein and membrane structure and would therefore be toxic. However, the effective solubility of fatty acids in fatty acid–albumin complexes is as much as 2 mM. Nevertheless, those rare individuals with **analbuminemia** (severely depressed levels of albumin) suffer no apparent adverse symptoms; evidently, their fatty acids are transported in complex with other serum proteins.

The X-ray structure of human serum albumin in its complexes with a variety of common fatty acids, determined by Stephen Curry, reveals that each albumin molecule can bind up to seven fatty acid molecules (Fig. 25-7). However, these binding sites have different fatty acid-binding affinities so that, under normal physiological conditions, albumin carries between 0.1 and 2 fatty acid molecules per protein molecule. Albumin also binds an extraordinarily broad range of drugs and is thereby a major and usually unpredictable influence on their pharmacokinetics (Section 15-4Ba). Indeed, the large amounts of fatty acids in the blood after meals can significantly affect the pharmacokinetics of a drug through competitive and/or cooperative interactions.

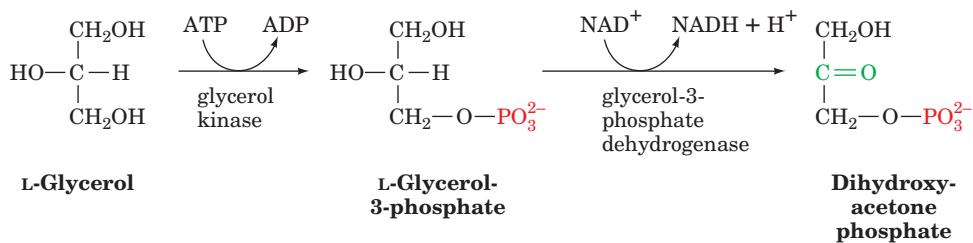


Figure 25-6 Conversion of glycerol to the glycolytic intermediate dihydroxyacetone phosphate.

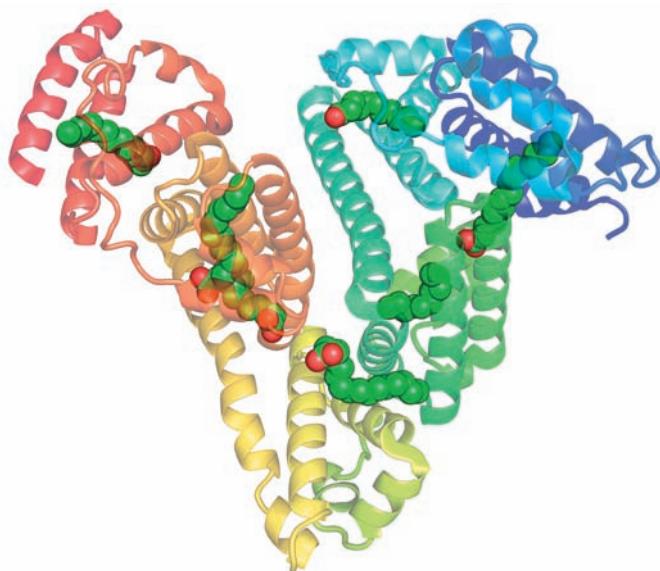


Figure 25-7 X-ray structure of human serum albumin in complex with 7 molecules of palmitic acid. The protein is drawn in semitransparent ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The fatty acids are shown in space-filling form with C green and O red. [Based on an X-ray structure by Stephen Curry, Imperial College of Science, Technology, and Medicine, London, U.K. PDBid 1E7H.]

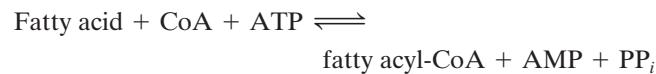
2 FATTY ACID OXIDATION

The biochemical strategy of fatty acid oxidation was understood long before the advent of biochemical techniques involving enzyme purification or the use of radioactive tracers. In 1904, Franz Knoop, in the first use of chemical

labels to trace metabolic pathways, fed dogs fatty acids labeled at their ω (last) carbon atom by a benzene ring and isolated the phenyl-containing metabolic products from their urine. Dogs fed labeled odd-chain fatty acids excreted **hippuric acid**, the glycine amide of **benzoic acid**, whereas those fed labeled even-chain fatty acids excreted **phenylaceturic acid**, the glycine amide of **phenylacetic acid** (Fig. 25-8). Knoop therefore deduced that the oxidation of the carbon atom β to the carboxyl group is involved in fatty acid breakdown. Otherwise, the phenylacetic acid would be further oxidized to benzoic acid. Knoop proposed that this breakdown occurs by a mechanism known as **β oxidation** in which the fatty acid's C_β atom is oxidized. It was not until after 1950, following the discovery of coenzyme A, that the enzymes of fatty acid oxidation were isolated and their reaction mechanisms elucidated. This work confirmed Knoop's hypothesis.

A. Fatty Acid Activation

Before fatty acids can be oxidized, they must be “primed” for reaction in an ATP-dependent acylation reaction to form fatty acyl-CoA. This activation process is catalyzed by a family of at least three **acyl-CoA synthetases** (also called **thiokinases**) that differ according to their chain-length specificities. These enzymes, which are associated with either the endoplasmic reticulum (ER) or the outer mitochondrial membrane, all catalyze the reaction



In the activation of ^{18}O -labeled palmitate by a long-chain acyl-CoA synthetase, both the AMP and the acyl-CoA products become ^{18}O labeled. This observation

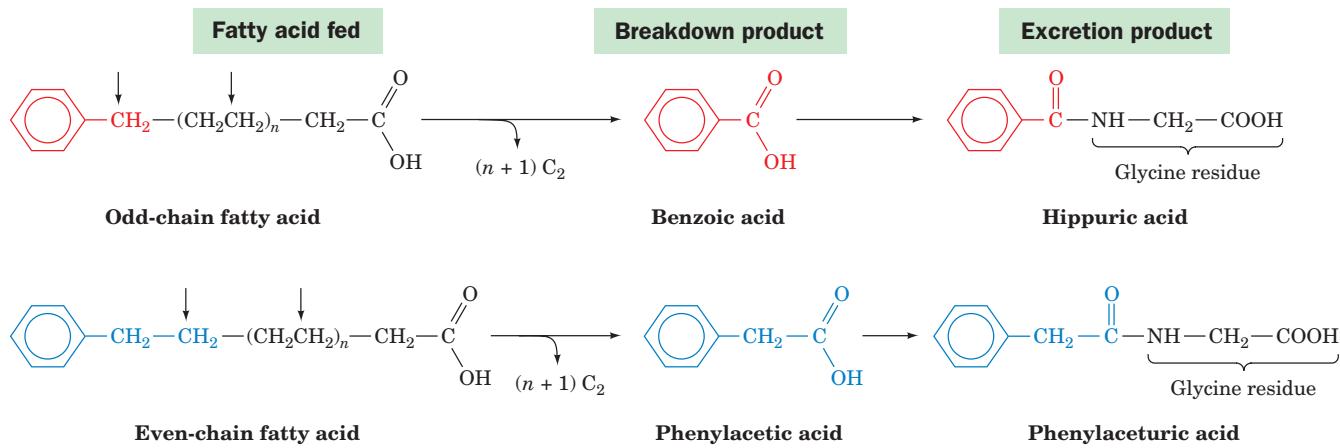


Figure 25-8 Franz Knoop's classic experiment indicating that fatty acids are metabolically oxidized at their β -carbon atom. ω -Phenyl-labeled fatty acids containing an odd number of carbon atoms are oxidized to the phenyl-labeled C_1 product, benzoic acid, whereas those with an even number of carbon atoms are oxidized to the phenyl-labeled C_2 product, phenylacetic acid.

These products are excreted as their respective glycine amides, hippuric and phenylaceturic acids. The vertical arrows indicate the deduced sites of carbon oxidation. The intermediate C_2 products are oxidized to CO_2 and H_2O and were therefore not isolated.

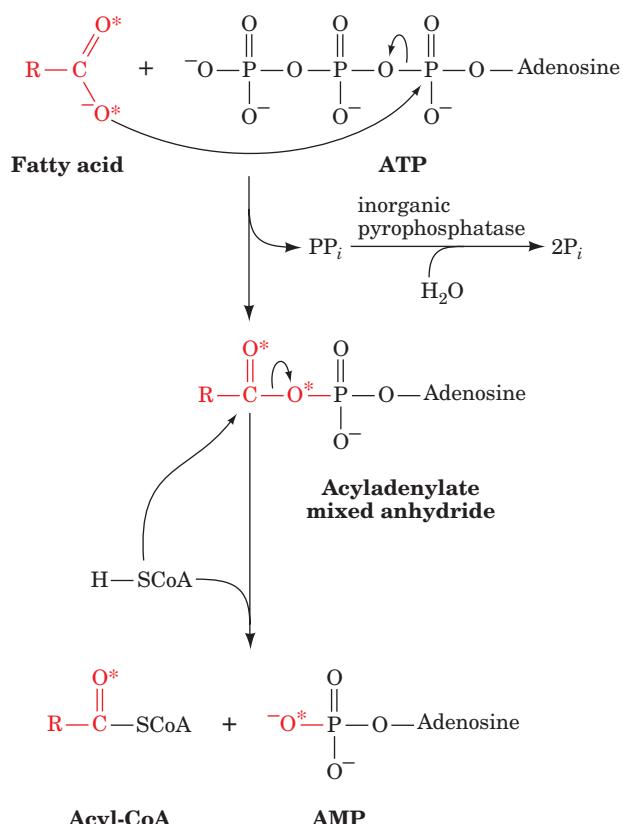


Figure 25-9 Mechanism of fatty acid activation catalyzed by acyl-CoA synthetase. Experiments utilizing ^{18}O -labeled fatty acids (*) demonstrate that the formation of acyl-CoA involves an intermediate acyladenylate mixed anhydride.

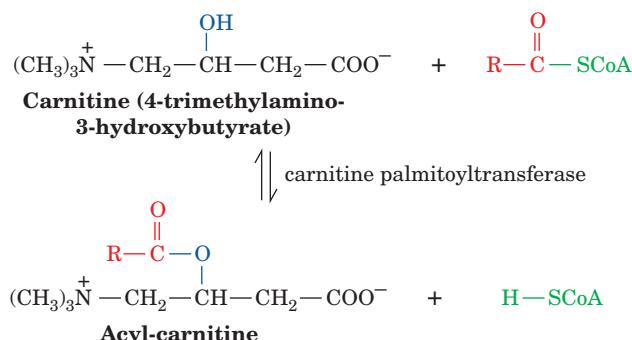


Figure 25-10 Acylation of carnitine catalyzed by carnitine palmitoyltransferase.

indicates that the reaction has an acyladenylate mixed anhydride intermediate that is attacked by the sulphydryl group of CoA to form the thioester product (Fig. 25-9). The reaction involves both the cleavage and the synthesis of bonds with large negative free energies of hydrolysis so that the free energy change associated with the overall reaction is close to zero. The reaction is driven to completion in the cell by the highly exergonic hydrolysis of the product pyrophosphate (PP_i) catalyzed by the ubiquitous **inorganic pyrophosphatase**. Thus, as commonly occurs in metabolic pathways, a reaction forming a “high-energy” bond through the hydrolysis of one of ATP’s phosphoanhydride bonds is driven to completion by the hydrolysis of its second such bond.

B. Transport Across the Mitochondrial Membrane

Although fatty acids are activated for oxidation in the cytosol, they are oxidized in the mitochondrion as Eugene Kennedy and Albert Lehninger established in 1950. We must therefore consider how fatty acyl-CoA is transported across the inner mitochondrial membrane. A long-chain fatty acyl-CoA cannot directly cross the inner mitochondrial membrane. Rather, its acyl portion is first transferred to **carnitine** (Fig. 25-10), a compound that occurs in both plant and animal tissues. This transesterification reaction has an equilibrium constant close to 1, which indicates that the O -acyl bond of **acyl-carnitine** has a free energy of hydrolysis similar to that of the thioester. **Carnitine palmitoyltransferases I** and **II**, which can transfer a variety of acyl groups, are located, respectively, on the external and internal surfaces of the inner mitochondrial membrane. The translocation process itself is mediated by a specific carrier protein that transports acyl-carnitine into the mitochondrion while transporting free carnitine in the opposite direction. Acyl-CoA transport therefore occurs via four reactions (Fig. 25-11):

1. The acyl group of a cytosolic acyl-CoA is transferred to carnitine, thereby releasing the CoA to its cytosolic pool.
2. The resulting acyl-carnitine is transported into the mitochondrial matrix by the transport system.
3. The acyl group is transferred to a CoA molecule from the mitochondrial pool.
4. The product carnitine is returned to the cytosol.

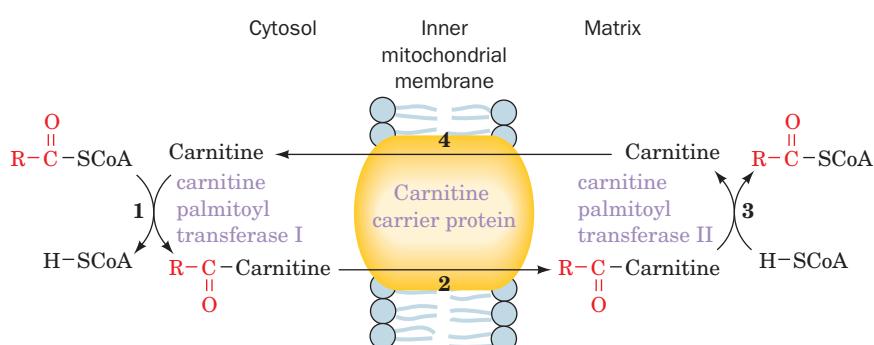


Figure 25-11 Transport of fatty acids into the mitochondrion.

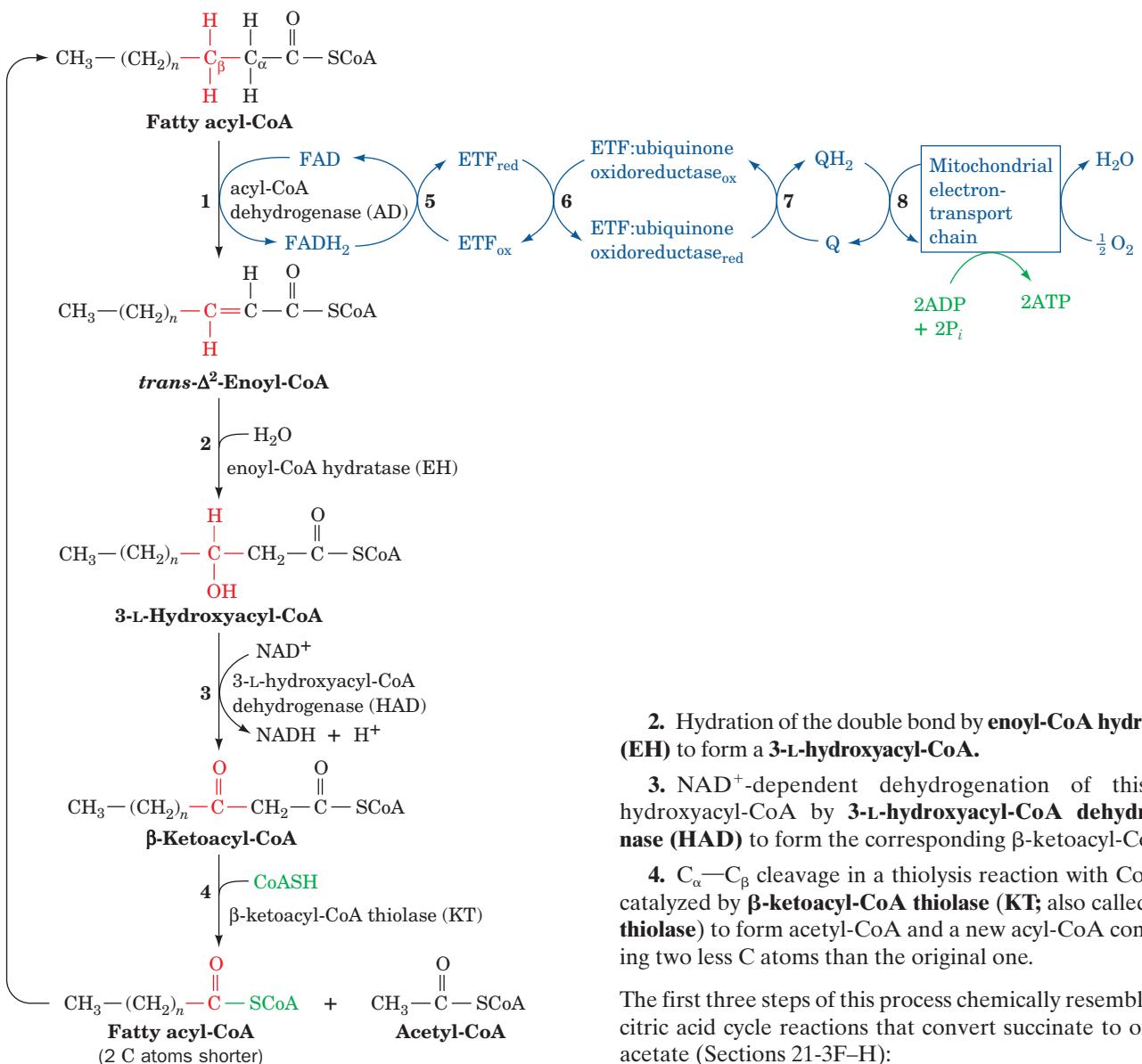


Figure 25-12 The β -oxidation pathway of fatty acyl-CoA.

See the Animated Figures

The cell thereby maintains separate cytosolic and mitochondrial pools of CoA. The mitochondrial pool functions in the oxidative degradation of pyruvate (Section 21-2A) and certain amino acids (Sections 26-3E–G) as well as fatty acids, whereas the cytosolic pool supplies fatty acid biosynthesis (Section 25-4). The cell similarly maintains separate cytosolic and mitochondrial pools of ATP and NAD⁺.

C. *β Oxidation*

Fatty acids are dismembered through the β oxidation of fatty acyl-CoA, a process that occurs in four reactions (Fig. 25-12):

- ### 1. Formation of a trans- α, β double bond through dehydrogenation by the flavoenzyme **acyl-CoA dehydrogenase (AD)**.

2. Hydration of the double bond by **enoyl-CoA hydratase (EH)** to form a **3-L-hydroxyacyl-CoA**.

3. NAD⁺-dependent dehydrogenation of this β -hydroxyacyl-CoA by **3-L-hydroxyacyl-CoA dehydrogenase (HAD)** to form the corresponding β -ketoacyl-CoA.

4. $C_\alpha - C_\beta$ cleavage in a thiolysis reaction with CoA as catalyzed by **β -ketoacyl-CoA thiolase (KT; also called just thiolase)** to form acetyl-CoA and a new acyl-CoA containing two less C atoms than the original one.

The first three steps of this process chemically resemble the citric acid cycle reactions that convert succinate to oxaloacetate (Sections 21-3F–H):

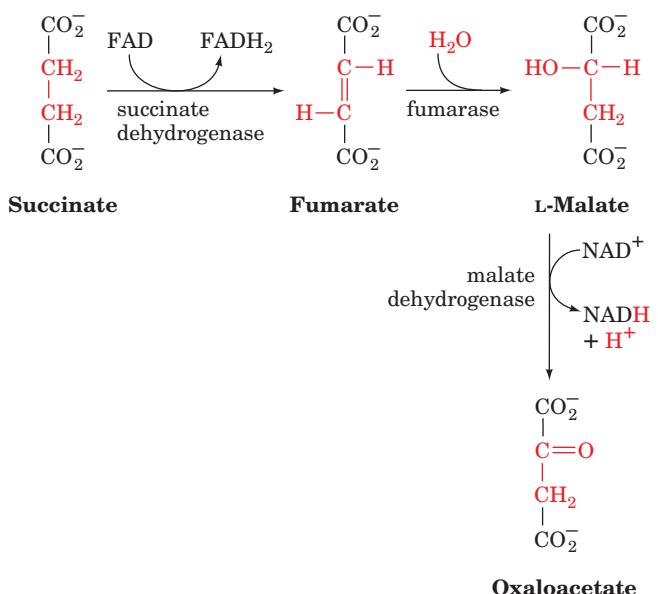




Figure 25-13 Ribbon diagram of the active site region in a subunit of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria in complex with octanoyl-CoA. The enzyme is a tetramer of identical 385-residue subunits, each of which binds an FAD prosthetic group (green) and its octanoyl-CoA substrate (whose octanoyl and CoA moieties are blue and white) in largely extended conformations. The octanoyl-CoA binds such that its C_{α} — C_{β} bond is sandwiched between the carboxylate group of Glu 376 (red) and the flavin ring (green), consistent with the proposal that Glu 376 is the general base that abstracts the α proton in the α,β dehydrogenation reaction catalyzed by the enzyme. [Based on an X-ray structure by Jung-Ja Kim, Medical College of Wisconsin. PDBid 3MDE.] 

Exercise 23

Mitochondria contain four acyl-CoA dehydrogenases, with specificities for short- (C_4 to C_6), medium- (C_6 to C_{10}), long- (between medium and very long), and very long-chain (C_{12} to C_{18}) fatty acyl-CoAs. The reaction catalyzed by these enzymes is thought to involve removal of a proton at C_{α} and transfer of a hydride ion equivalent from C_{β} to FAD (Fig. 25-12, Reaction 1). The X-ray structure of the **medium-chain acyl-CoA dehydrogenase (MCAD)** in complex with **octanoyl-CoA**, determined by Jung-Ja Kim, clearly shows how the enzyme orients the enzyme's base (Glu 376), the substrate C_{α} — C_{β} bond, and the FAD prosthetic group for reaction (Fig. 25-13).

a. Acyl-CoA Dehydrogenase Is Reoxidized via the Electron-Transport Chain

The FADH_2 resulting from the oxidation of the fatty acyl-CoA substrate is reoxidized by the mitochondrial electron-transport chain through the intermediacy of a series of electron-transfer reactions. **Electron-transfer flavoprotein (ETF)** transfers two electrons from FADH_2 to the flavoiron-sulfur protein **ETF:ubiquinone oxidoreductase**, which in turn transfers two electrons to the mitochondrial electron-transport chain by reducing coenzyme Q (CoQ; Fig. 25-12, Reactions 5–8). Reduction of O_2 to H_2O by the electron-transport chain beginning at the CoQ stage re-

sults in the synthesis of 1.5 ATPs per two electrons transferred (Section 22-2Bc).

b. Acyl-CoA Dehydrogenase Deficiency Has Fatal Consequences

The unexpected death of an apparently healthy infant, often overnight, has been, for lack of any real explanation, termed **sudden infant death syndrome (SIDS)**. MCAD has been shown to be deficient in up to 10% of these infants, making this genetic disease more prevalent than **phenylketonuria (PKU)** (Section 26-3Hd), a genetic defect in phenylalanine degradation for which babies born in the United States are routinely tested. Glucose is the principal energy metabolism substrate just after eating, but when the glucose level later decreases, the rate of fatty acid oxidation must correspondingly increase. The sudden death of infants lacking MCAD may be caused by the imbalance between glucose and fatty acid oxidation.

Lys 304, which becomes Glu in the most prevalent mutation among individuals with MCAD deficiency, is ~ 20 Å distant from the enzyme's active site and hence cannot participate in binding substrate or FAD. However, since the side chains of Asp 300 and Asp 346 lie within 6 Å of Glu 304, near a subunit–subunit interface, it seems likely that the high concentration of negative charges resulting from the Lys 304 \rightarrow Glu mutation structurally destabilizes the enzyme.

Deficiency of acyl-CoA dehydrogenase has also been implicated in **Jamaican vomiting sickness**, whose victims suffer violent vomiting followed by convulsions, coma, and death. Severe hypoglycemia is observed in most cases. This condition results from eating unripe **ackee fruit**, which contains **hypoglycin A**, an unusual amino acid, which is metabolized to **methylene cyclopropyl acetyl-CoA (MCPA-CoA)** (Fig. 25-14). MCPA-CoA, a substrate for acyl-CoA dehydrogenase, is thought to undergo the first step of the reaction that this enzyme catalyzes, removal of a proton from C_{α} , to form a reactive intermediate that covalently modifies the enzyme's FAD prosthetic group (Fig. 25-14). Since a normal step in the enzyme's reaction mechanism generates the reactive intermediate, MCPA-CoA is said to be a **mechanism-based inhibitor**.

c. Long-Chain Enoyl-CoAs Are Converted to Acetyl-CoA and a Shorter Acyl-CoA by Mitochondrial Trifunctional Protein

The products of acyl-CoA dehydrogenases are 2-enoyl-CoAs. Depending on their chain lengths their processing is continued by one of three systems (Fig. 25-12): the short-chain, medium-chain, or long-chain 2-enoyl-CoA hydratases (EHs), hydroxyacyl-CoA dehydrogenases (HADs), and β -ketoacyl-CoA thiolases (KTs). The long-chain (LC) versions of these enzymes are contained on one $\alpha_4\beta_4$ octameric protein, **mitochondrial trifunctional protein**, located in the inner mitochondrial membrane. LCEH and LCHAD are contained on the α subunits while LCKT is located on the β subunits. The protein is therefore a combination multifunctional protein (more than one enzyme activity on a sin-

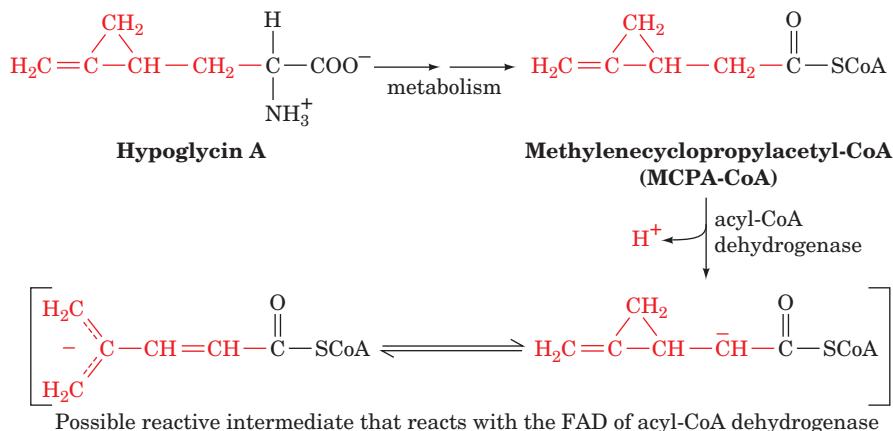


Figure 25-14 Metabolic conversions of hypoglycin A to yield a product that inactivates acyl-CoA dehydrogenase. Spectral

changes suggest that the enzyme's FAD prosthetic group has been modified.

gle polypeptide chain)—multienzyme complex (a complex of polypeptides catalyzing more than one reaction). The advantage of such a trifunctional enzyme is the ability to channel the intermediates toward the final product. Indeed, no long-chain hydroxyacyl-CoA or ketoacyl-CoA intermediates are released into solution by this system.

d. The Thiolase Reaction Occurs via Claisen Ester Cleavage

The final stage of the fatty acid β -oxidation process, the thiolase reaction, forms acetyl-CoA and a new acyl-CoA which is two carbon atoms shorter than the one that began the cycle. This occurs in five reaction steps (Fig. 25-15):

1. An active site thiol is added to the substrate β -keto group.

2. Carbon–carbon bond cleavage forms an acetyl-CoA carbanion intermediate that is stabilized by electron withdrawal into this thioester's carbonyl group. Such a reaction is known as a Claisen ester cleavage (the reverse of a Claisen condensation). The citric acid cycle enzyme citrate synthase also catalyzes a reaction that involves a stabilized acetyl-CoA carbanion intermediate (Section 21-3A).

3. The acetyl-CoA carbanion intermediate is protonated by an enzyme acid group, yielding acetyl-CoA.

4 and 5. Finally, CoA displaces the enzyme thiol group from the enzyme–thioester intermediate, yielding acyl-CoA.

The formation of an enzyme–thioester intermediate involving an active site thiol group is based on the observation that incubation of the enzyme with [14 C]acetyl-CoA

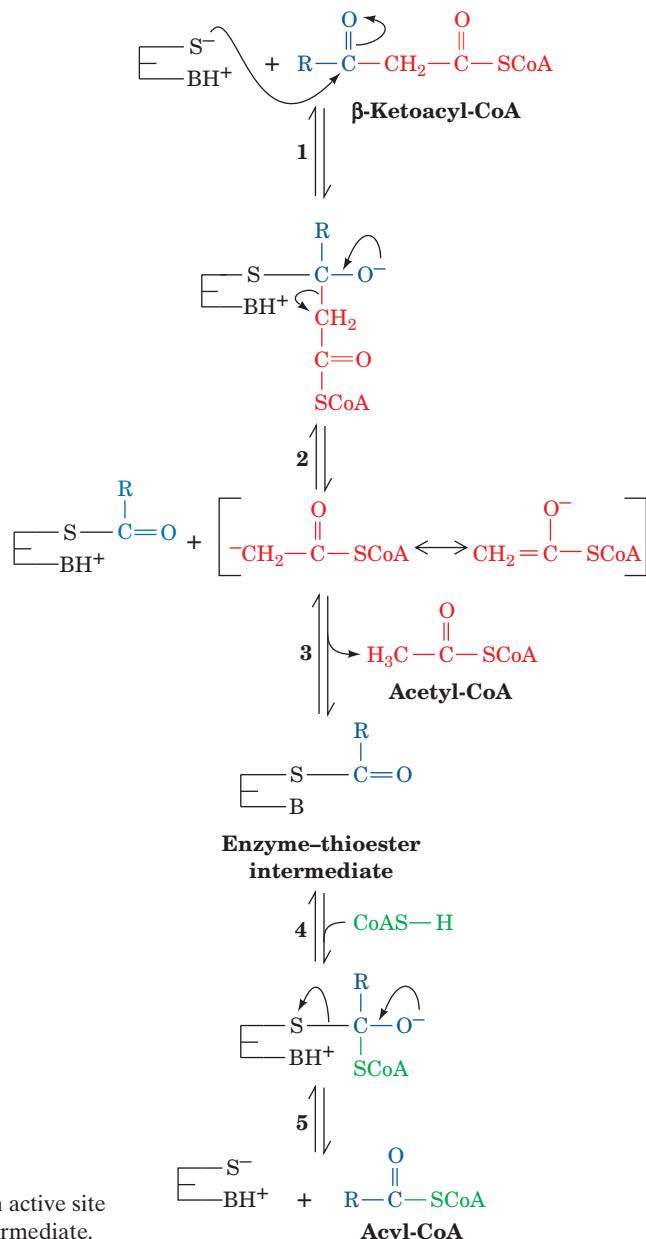
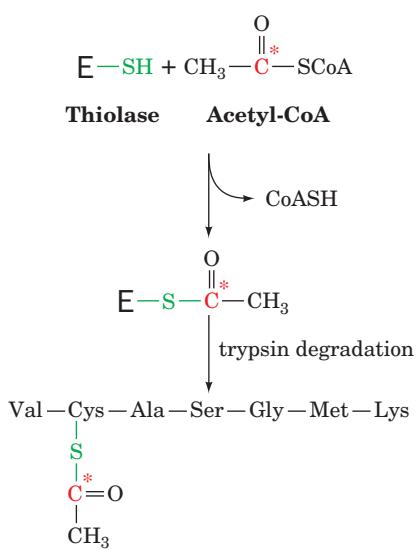


Figure 25-15 Mechanism of action of β -ketoacyl-CoA thiolase. An active site Cys residue participates in the formation of an enzyme–thioester intermediate.

yields a specifically labeled enzyme Cys residue (the reverse of steps 4 and 5):



e. Fatty Acid Oxidation Is Highly Exergonic

The function of fatty acid oxidation is, of course, to generate metabolic energy. Each round of β oxidation produces one NADH, one FADH₂, and one acetyl-CoA. Oxidation of acetyl-CoA via the citric acid cycle generates additional FADH₂ and NADH, which are reoxidized through oxidative phosphorylation to form ATP. Complete oxidation of a fatty acid molecule is therefore a highly exergonic process, which yields numerous ATPs. For example, oxidation of palmitoyl-CoA (which has a C₁₆ fatty acyl group) involves seven rounds of β oxidation, yielding 7 FADH₂, 7 NADH, and 8 acetyl-CoA. Oxidation of the 8 acetyl-CoA, in turn, yields 8 GTP, 24 NADH, and 8 FADH₂. Since oxidative phosphorylation of the 31 NADH molecules yields 77.5 ATP and that of the 15 FADH₂ yields 22.5 ATPs, subtracting the 2 ATP equivalents required for fatty acyl-CoA formation (Section 25-2A), the oxidation of one palmitate molecule has a net yield of 106 ATP.

D. Oxidation of Unsaturated Fatty Acids

Almost all unsaturated fatty acids of biological origin (Section 12-1A) contain only *cis* double bonds, which most often begin between C9 and C10 (referred to as a Δ^9 or 9-double bond; Table 12-1). Additional double bonds, if any, occur at three-carbon intervals and are therefore never conjugated. Two examples of unsaturated fatty acids are oleic acid and linoleic acid (Fig. 25-16). Note that one of the double bonds in linoleic acid is at an odd-numbered carbon atom and the other is at an even-numbered carbon atom. Double bonds at these positions in fatty acids pose three problems for the β -oxidation pathway that are solved through the actions of four additional enzymes (Fig. 25-17):

Problem 1: A β,γ Double Bond

The first enzymatic difficulty occurs on the left-hand pathway in Fig. 25-17 after the third round of β oxidation:

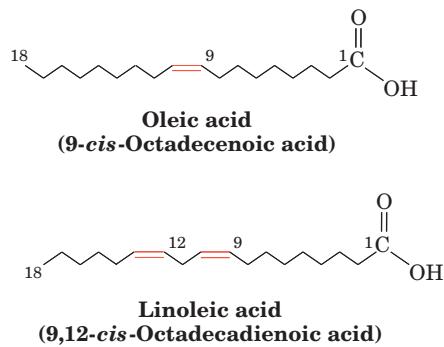
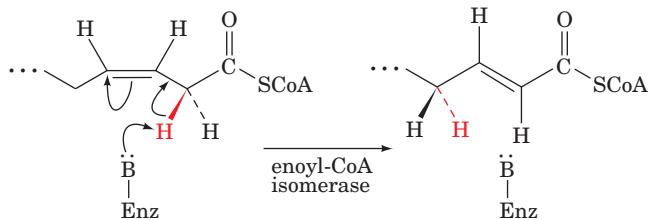


Figure 25-16 Structures of two common unsaturated fatty acids. Most unsaturated fatty acids contain unconjugated *cis* double bonds.

The resulting *cis*- β,γ double bond-containing enoyl-CoA is not a substrate for enoyl-CoA hydratase. **Enoyl-CoA isomerase**, however, mediates conversion of the *cis*- Δ^3 double bond to the more stable, ester-conjugated *trans*- Δ^2 form:

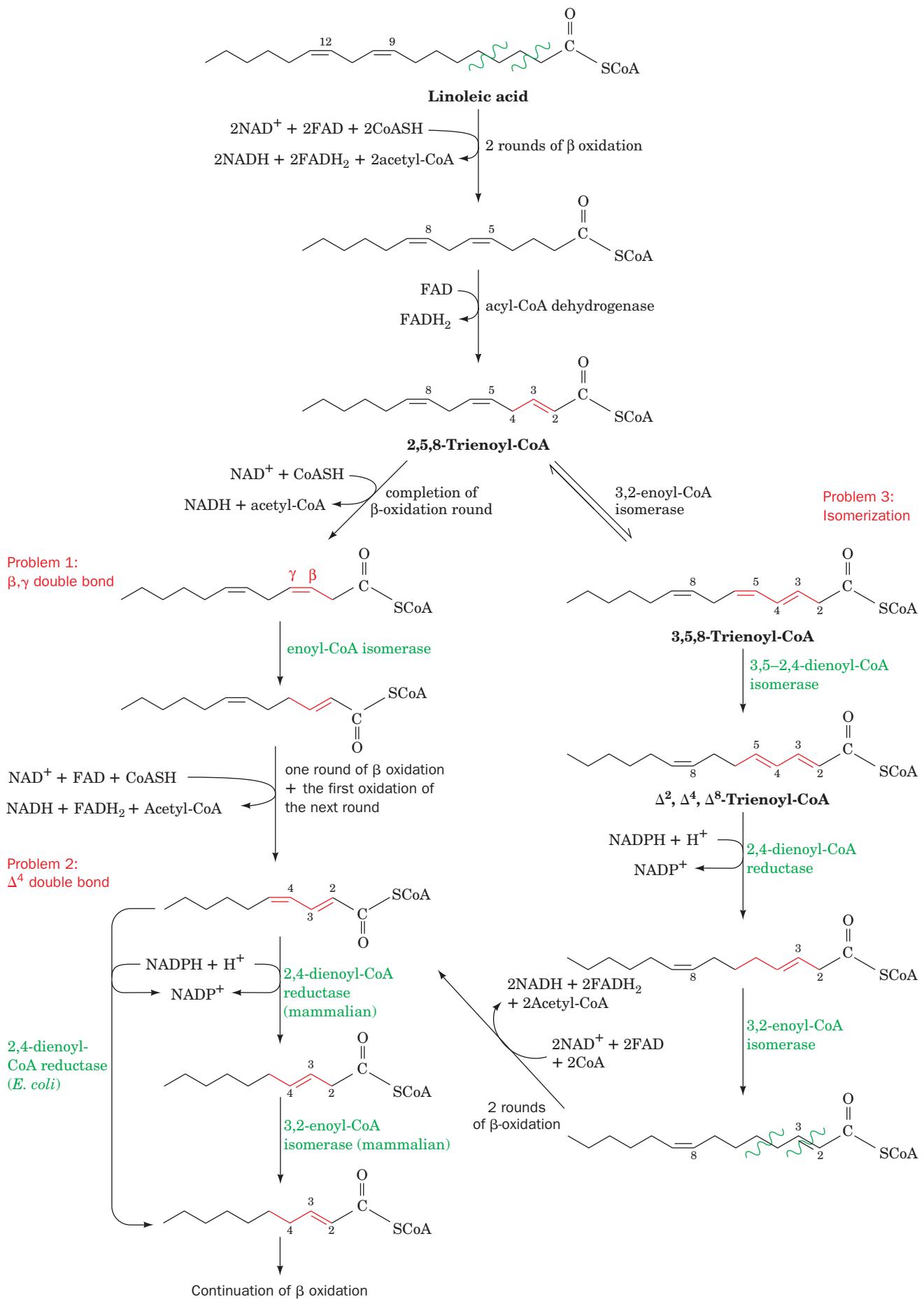


Such compounds are normal substrates of enoyl-CoA hydratase so that β oxidation can then continue.

Problem 2: A Δ^4 Double Bond Inhibits Hydratase Action

The next difficulty arises on the left-hand pathway in Fig. 25-17 in the fifth round of β oxidation. The presence of a double bond at an even-numbered carbon atom results in the formation of 2,4-dienoyl-CoA, which is a poor substrate for enoyl-CoA hydratase. However, NADPH-dependent

Figure 25-17 (Opposite) Problems in the oxidation of unsaturated fatty acids and their solutions. Linoleic acid is used as an example. The first problem, the presence of a β,γ double bond seen in the left-hand pathway, is solved by the bond's enoyl-CoA isomerase-catalyzed conversion to a *trans*- α,β double bond. The second problem in the left-hand pathway, that a 2,4-dienoyl-CoA is not a substrate for enoyl-CoA hydratase, is eliminated by the NADPH-dependent reduction of the Δ^4 bond by 2,4-dienoyl-CoA reductase to yield the β -oxidation substrate *trans*-2-enoyl-CoA in *E. coli* but *trans*-3-enoyl-CoA in mammals. Mammals therefore also have 3,2-enoyl-CoA isomerase, which converts the *trans*-3-enoyl-CoA to *trans*-2-enoyl-CoA. The third problem, the isomerization of 2,5-dienoyl-CoA (originating from the oxidation of unsaturated fatty acids with double bonds at odd-numbered C atoms) to 3,5-dienoyl-CoA by 3,2-enoyl-CoA isomerase, is solved by 3,5-2,4-dienoyl-CoA isomerase, which converts the 3,5-dienoyl-CoA to 2,4-dienoyl-CoA, a substrate for 2,4-dienoyl-CoA reductase.



2,4-dienoyl-CoA reductase reduces the Δ^4 double bond. The *E. coli* reductase produces *trans*-2-enoyl-CoA, a normal substrate of β oxidation. The mammalian reductase, however, yields *trans*-3-enoyl-CoA, which, to proceed along the β -oxidation pathway, must first be isomerized to *trans*-2-enoyl-CoA by **3,2-enoyl-CoA isomerase**.

Problem 3: The Unanticipated Isomerization of 2,5-Enoyl-CoA by 3,2-Enoyl-CoA Isomerase

Mammalian 3,2-enoyl-CoA isomerase catalyzes a reversible reaction that interconverts Δ^2 and Δ^3 double bonds. A carbonyl group is stabilized by being conjugated to a Δ^2 double bond. However, the presence of a Δ^5 double bond (originating from an unsaturated fatty acid with a double bond at an odd-numbered C atom such as the Δ^9 double bond of linoleic acid) is likewise stabilized by being conjugated with a Δ^3 double bond (right-hand pathway of Fig. 25-17). If a 2,5-enoyl-CoA is converted by 3,2-enoyl-CoA isomerase to 3,5-enoyl-CoA, which occurs up to 20% of the time, another enzyme is necessary to continue the oxidation: **3,5-2,4-Dienoyl-CoA isomerase** isomerizes the 3,5-diene to a 2,4-diene, which is then reduced by 2,4-dienoyl-CoA reductase and isomerized by 3,2-enoyl-CoA isomerase as in Problem 2 above. After two more rounds of β oxidation, the *cis*- Δ^4 double bond originating from the *cis*- Δ^{12} double bond of linoleic acid is also dealt with as in Problem 2.

E. Oxidation of Odd-Chain Fatty Acids

Most fatty acids have even numbers of carbon atoms and are therefore completely converted to acetyl-CoA. Some plants and marine organisms, however, synthesize fatty acids with an odd number of carbon atoms. *The final round of β oxidation of these fatty acids forms propionyl-CoA, which, as we shall see, is converted to succinyl-CoA for entry into the citric acid cycle.* Propionate or propionyl-CoA is also produced by oxidation of the amino acids isoleucine, valine, and methionine (Section 26-3E). Furthermore, ruminant animals such as cattle derive most of their caloric intake from the acetate and propionate produced in their rumen (stomach) by bacterial fermentation of carbohydrates. These products are absorbed by the animal and metabolized after conversion to the corresponding acyl-CoA.

a. Propionyl-CoA Carboxylase Has a Biotin Prosthetic Group

The conversion of propionyl-CoA to succinyl-CoA involves three enzymes (Fig. 25-18). The first reaction is that of **propionyl-CoA carboxylase**, a biotin-dependent enzyme (Section 23-1Ab) with subunit composition $\alpha_6\beta_6$. The reaction, which resembles that catalyzed by the homologous biotin-containing enzyme pyruvate decarboxylase (Section 23-1Ac), occurs in two steps (Fig. 25-19):

1. Carboxylation of biotin at N1' by bicarbonate ion as in the pyruvate carboxylase reaction (Fig. 23-4). This step,

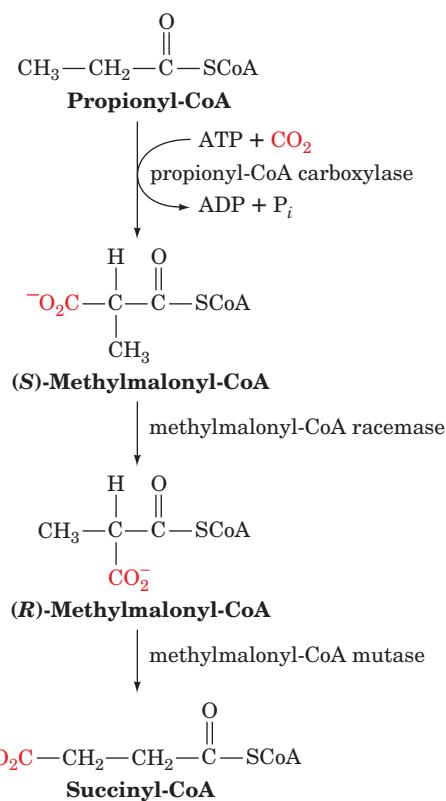


Figure 25-18 Conversion of propionyl-CoA to succinyl-CoA.

which is driven by the concomitant hydrolysis of ATP to ADP and P_i , activates the resulting carboxyl group for transfer without further free energy input.

2. Stereospecific transfer of the activated carboxyl group from carboxybiotin to propionyl-CoA to form **(S)-methylmalonyl-CoA**. This step occurs via nucleophilic attack on carboxybiotin by a carbanion at C2 of propionyl-CoA (see below).

These two reaction steps occur at different catalytic sites on propionyl-CoA carboxylase. It therefore appears that the biotinyllysine linkage attaching the biotin ring to the enzyme forms a flexible tether that permits the efficient transfer of the biotin ring between these two active sites as occurs in pyruvate carboxylase (Section 23-1Ae).

Formation of the C2 carbanion in the second stage of the propionyl-CoA carboxylase reaction involves removal of a proton α to a thioester. This proton is relatively acidic since, as we have seen in Section 25-2Cd, the negative charge on a carbanion α to a thioester can be delocalized into the thioester's carbonyl group. This explains the relatively convoluted path taken in the conversion of propionyl-CoA to succinyl-CoA (Fig. 25-18). It would seem simpler, at least on paper, for this process to occur in one step, with carboxylation occurring on C3 of propionyl-CoA so as to form succinyl-CoA directly. Yet, the C3 carbanion required for

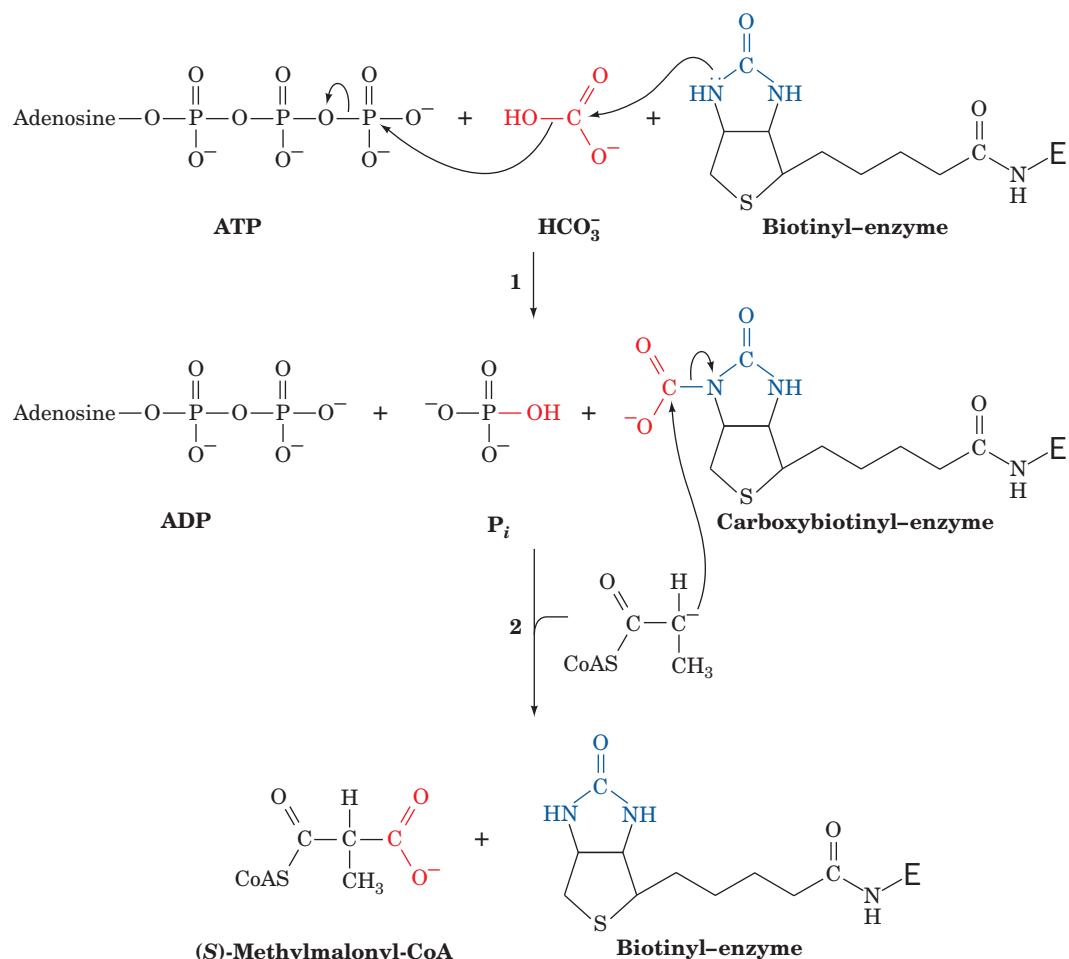


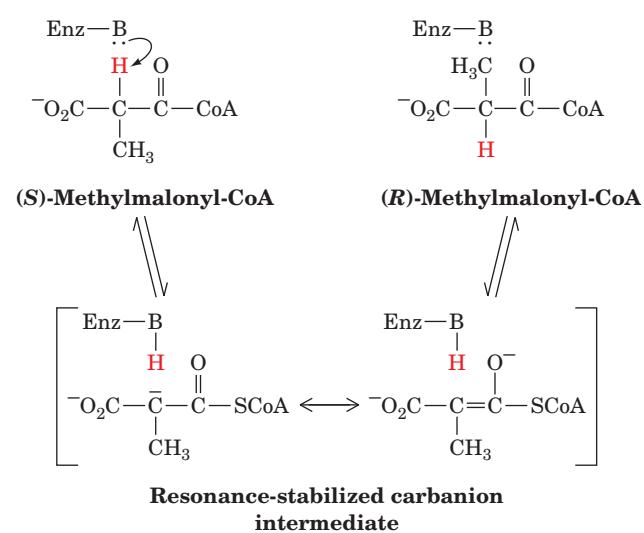
Figure 25-19 The propionyl-CoA carboxylase reaction. (1) The carboxylation of biotin with the concomitant hydrolysis of ATP is followed by (2) the carboxylation of a propionyl-CoA

carbanion by its attack on carboxybiotin. Each reaction step probably involves the intermediate formation of CO_2 as occurs in the pyruvate carboxylase reaction (Fig. 23-4).

such a carboxylation has a high free energy of formation. Nature has instead chosen a more facile, albeit less direct route, which carboxylates propionyl-CoA at a more reactive position and then rearranges the C_4 skeleton to form the desired product.

b. Methylmalonyl-CoA Mutase Contains a Coenzyme B_{12} Prosthetic Group

Methylmalonyl-CoA mutase, which catalyzes the third reaction of the propionyl-CoA to succinyl-CoA conversion (Fig. 25-18), is specific for (*R*)-methylmalonyl-CoA even though propionyl-CoA carboxylase stereospecifically synthesizes (*S*)-methylmalonyl-CoA. This diversion is rectified by **methylmalonyl-CoA racemase**, which interconverts the (*R*) and (*S*) configurations of methylmalonyl-CoA, presumably by promoting the reversible dissociation of its acidic α -H via formation of a resonance-stabilized carbanion intermediate:



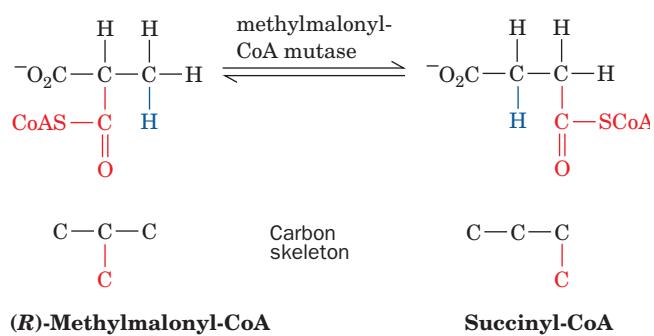


Figure 25-20 The rearrangement catalyzed by methylmalonyl-CoA mutase.

Methylmalonyl-CoA mutase, which catalyzes an unusual carbon skeleton rearrangement (Fig. 25-20), utilizes a **5'-deoxyadenosylcobalamin (AdoCbl)** prosthetic group (also called **coenzyme B₁₂**). Dorothy Hodgkin determined the structure of this complex molecule (Fig. 25-21) in 1956, a landmark achievement, through X-ray crystallographic analysis combined with chemical degradation studies. AdoCbl contains a hemelike **corrin** ring whose four pyrrole N atoms each ligand a 6-coordinate Co ion. The fifth Co ligand in the free coenzyme is an N atom of a **5,6-dimethylbenzimidazole (DMB)** nucleotide that is covalently linked to the corrin D ring. The sixth ligand is a 5'-deoxyadenosyl group in which the deoxyribose C5' atom forms a covalent C—Co bond, *one of only two known carbon–metal bonds in biology* (the other being a C—Ni bond in the bacterial enzyme **carbon monoxide dehydrogenase**). In some enzymes, the sixth ligand instead is a CH₃ group that likewise forms a C—Co bond.

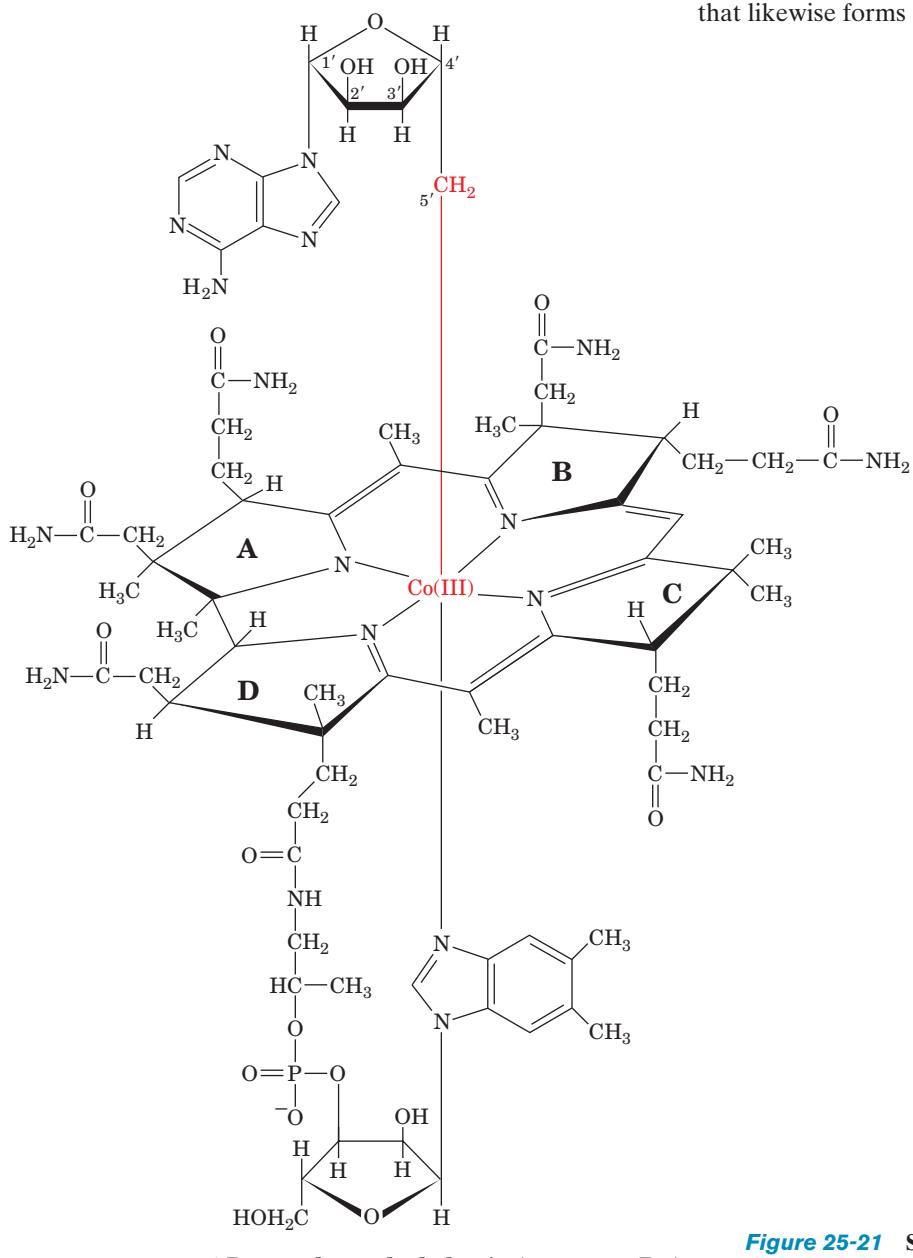
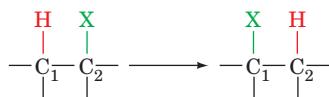


Figure 25-21 Structure of 5'-deoxyadenosylcobalamin (coenzyme B₁₂).

AdoCbl's reactive C—Co bond participates in two types of enzyme-catalyzed reactions:

1. Rearrangements in which a hydrogen atom is directly transferred between two adjacent carbon atoms with concomitant exchange of the second substituent, X:



where X may be a carbon atom with substituents, an oxygen atom of an alcohol, or an amine.

2. Methyl group transfers between two molecules.

There are about a dozen known cobalamin-dependent enzymes. However, only two occur in mammalian systems: (1) methylmalonyl-CoA mutase, which catalyzes a carbon skeleton rearrangement (the X group in the rearrangement is —COSCoA; Fig. 25-20) and is the only B₁₂-containing enzyme that occurs in both eukaryotes and prokaryotes; and (2) **methionine synthase**, a methyl transfer enzyme that participates in methionine biosynthesis (Sections 26-3Ec and 26-5B). Defects in methylmalonyl-CoA mutase result in **methylmalonic aciduria**, a condition that is often fatal in infancy due to **acidosis** (low blood pH) without a diet de-

void of odd-chain fatty acids and low in the amino acid residues that are degraded to propionyl-CoA (Ile, Val, and Met; Section 26-3E).

c. The Methylmalonyl-CoA Mutase Reaction Occurs via a Free Radical Mechanism

Methylmalonyl-CoA mutase from *Propionibacterium shermanii* is an $\alpha\beta$ heterodimer whose catalytically active 728-residue α subunit is 24% identical to its catalytically inactive 638-residue β subunit. In contrast, the human enzyme is a homodimer whose subunits are 60% identical in sequence to *P. shermanii*'s α subunit. Hence *P. shermanii*'s β subunit is thought to be an evolutionary fossil.

The X-ray structure of methylmalonyl-CoA mutase from *P. shermanii* in complex with the substrate analog **2-carboxypropyl-CoA** (which lacks methylmalonyl-CoA's thioester oxygen atom) was determined by Philip Evans. Its AdoCbl cofactor is sandwiched between the α subunit's two domains: a 559-residue N-terminal α/β barrel (TIM barrel, the most common enzymatic motif; Section 8-3B) and a 169-residue C-terminal α/β domain that resembles a Rossmann fold (Section 8-3Bh). The structure of the α/β barrel contains several surprising features (Fig. 25-22):

1. The active sites of nearly all α/β barrel enzymes are located at the C-terminal ends of the barrel's β strands.

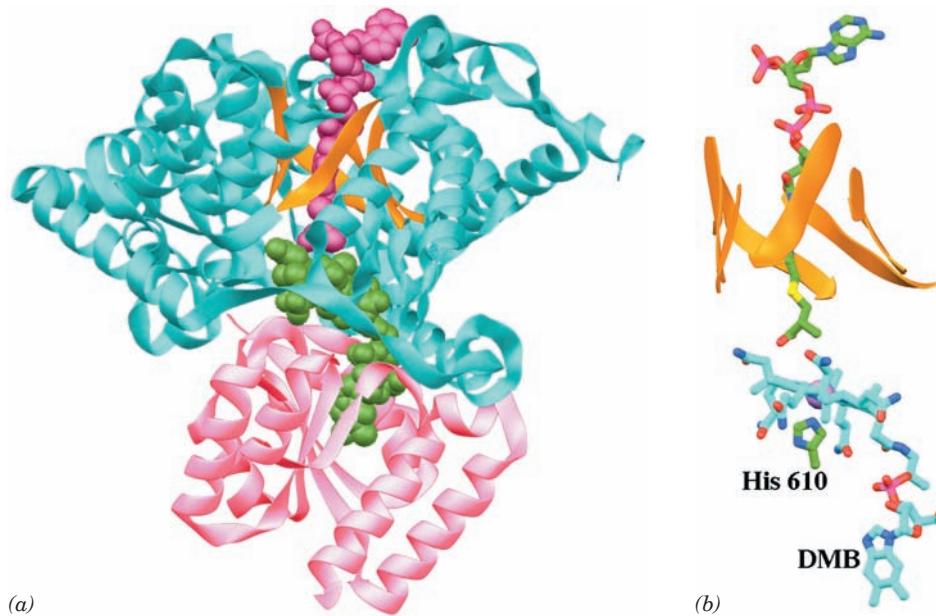


Figure 25-22 X-ray structure of *P. shermanii* methylmalonyl-CoA mutase in complex with 2-carboxypropyl-CoA and AdoCbl. (a) The catalytically active α subunit in which the N-terminal domain is cyan with the β strands of its α/β barrel orange, and the C-terminal domain is pink. The 2-carboxypropyl-CoA (magenta) and AdoCbl (green) are drawn in space-filling form. The 2-carboxypropyl-CoA passes through the center of the α/β barrel and is oriented such that the methylmalonyl group of methylmalonyl-CoA would contact the corrin ring of the AdoCbl, which is sandwiched between the enzyme's N- and C-terminal domains. (b) The arrangement of the AdoCbl and 2-carboxypropyl-CoA molecules which, together with the side

chain of His 610, are represented in stick form colored according to atom type (AdoCbl and His C green, 2-carboxypropyl-CoA C cyan, N blue, O red, P magenta, and S yellow). The corrin ring's Co atom is represented by a lavender sphere and the α/β barrel's β strands are represented by orange ribbons. The view is similar to that in Part a. Note that the DMB group (bottom) has swung away from the corrin ring (seen edgewise) to be replaced by the side chain of His 610 from the C-terminal domain and that the 5'-deoxyadenosyl group is unseen (due to disorder). [Based on an X-ray structure by Philip Evans, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 7REQ.]

See Interactive Exercise 24

However, in methylmalonyl-CoA mutase, the AdoCbl is packed against the N-terminal ends of the barrel's β strands.

2. In free AdoCbl, the Co atom is axially liganded by an N atom of its DMB group and by the adenosyl residue's 5'-CH₂ group (Fig. 25-21). However, in the enzyme, the DMB has swung aside to bind in a separate pocket and has been replaced by the side chain of His 610 from the C-terminal domain. The adenosyl group is not visible in the structure due to disorder and hence has probably also swung aside.

3. In nearly all other α/β barrel-containing enzymes, the center of the barrel is occluded by large, often branched, hydrophobic side chains. However, in methylmalonyl-CoA mutase, the 2-carboxypropyl-CoA's pantetheine group binds in a narrow tunnel through the center of the α/β barrel so as to put the methylmalonyl group of an intact substrate in close proximity to the unliganded face of the cobalamin ring. This tunnel provides the only direct access to the active site cavity, thereby protecting the reactive free radical intermediates that are produced in the catalytic reaction from side reactions (see below). The tunnel is lined by small hydrophilic residues (Ser and Thr).

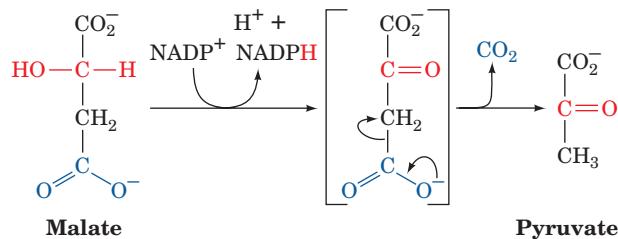
Methylmalonyl-CoA mutase's substrate binding mode resembles those of several other AdoCbl-containing enzymes of known structure, which are collectively unique among α/β barrel-containing proteins.

The proposed methylmalonyl-CoA mutase reaction mechanism (Fig. 25-23) begins with **homolytic cleavage** of the cobalamin C—Co(III) bond (the C and Co atoms each acquire one of the electrons that formed the cleaved electron pair bond). The Co ion therefore fluctuates between its Co(III) and Co(II) oxidation states [the two states are spectroscopically distinguishable: Co(III) is red and diamagnetic (no unpaired electrons), whereas Co(II) is yellow and paramagnetic (unpaired electrons)]. Note that a homolytic cleavage reaction is unusual in biology; most other biological bond cleavage reactions occur via **heterolytic cleavage** (in which the electron pair forming the cleaved bond is fully acquired by one of the separating atoms).

The role of AdoCbl in the catalytic process is that of a reversible free radical generator. The C—Co(III) bond is well suited to this function because it is inherently weak (dissociation energy = 109 kJ · mol⁻¹) and appears to be further weakened through steric interactions with the enzyme. Indeed, as Fig. 25-22 indicates, the Co atom in methylmalonyl-CoA mutase has no sixth ligand and hence, as confirmed by spectroscopic measurements, is in its Co(II) state. The His N—Co bond is extremely long (2.5 Å vs 1.9–2.0 Å in various other B₁₂-containing structures). It is proposed that this strained and hence weakened bond stabilizes the Co(II) state with respect to the Co(III) state, thus favoring the formation of the adenosyl radical and facilitating the homolytic cleavage through which the catalyzed reaction occurs (Fig. 23-23). The adenosyl radical presumably abstracts a hydrogen atom from the substrate, thereby facilitating the rearrangement reaction through the intermediate formation of a cyclopropoxy radical.

d. Succinyl-CoA Cannot Be Directly Consumed by the Citric Acid Cycle

Methylmalonyl-CoA mutase catalyzes the conversion of a metabolite to a C₄ citric acid cycle intermediate, not acetyl-CoA. The route of succinyl-CoA oxidation is therefore not as simple as it may first appear. The citric acid cycle regenerates all of its C₄ intermediates so that these compounds are really catalysts, not substrates. Consequently, succinyl-CoA cannot undergo net degradation by citric acid cycle enzymes alone. Rather, *in order for a metabolite to undergo net oxidation by the citric acid cycle, it must first be converted either to pyruvate or directly to acetyl-CoA*. Net degradation of succinyl-CoA begins with its conversion, via the citric acid cycle, to malate. At high concentrations, malate is transported, by a specific transport protein, to the cytosol, where it may be oxidatively decarboxylated to pyruvate and CO₂ by **malic enzyme (malate dehydrogenase, decarboxylating)**:



(We previously encountered this enzyme in the C₄ cycle of photosynthesis; Fig. 24-41.) Pyruvate is then completely oxidized via pyruvate dehydrogenase and the citric acid cycle.

e. Pernicious Anemia Results from Vitamin B₁₂ Deficiency

The existence of **vitamin B₁₂** came to light in 1926 when George Minot and William Murphy discovered that **pernicious anemia**, a rare but often fatal disease of the elderly characterized by decreased numbers of red blood cells, low hemoglobin levels (for reasons explained in Section 26-4D), and progressive neurological deterioration (caused by the accumulation of odd-chain fatty acid residues in neuronal membranes), can be treated by the daily consumption of large amounts of raw liver (a treatment that some patients considered worse than the disease). It was not until 1948, however, after a bacterial assay for antipernicious anemia factor had been developed, that vitamin B₁₂ was isolated.

Vitamin B₁₂ is synthesized by neither plants nor animals but only by a few species of bacteria. Herbivores obtain their vitamin B₁₂ from the bacteria that inhabit their gut (in fact, some animals, such as rabbits, must periodically eat some of their feces to obtain sufficient amounts of this essential substance). Humans, however, obtain almost all their vitamin B₁₂ directly from their diet, particularly from meat. The vitamin is specifically bound in the intestine by the glycoprotein **intrinsic factor** that is secreted by the stomach. This complex is absorbed by a specific receptor in the intestinal mucosa, where the complex is dissociated and the liberated vitamin B₁₂ transported to the bloodstream.

There it is bound by at least three different plasma globulins, called **transcobalamins**, which facilitate its uptake by the tissues.

Pernicious anemia is not usually a dietary deficiency disease but, rather, results from insufficient secretion of intrinsic factor, often due to an autoimmune attack against the

cells that produce it. The normal human requirement for cobalamin is very small, $\sim 3 \mu\text{g} \cdot \text{day}^{-1}$, and the liver stores a 3- to 5-year supply of this vitamin. This accounts for the insidious onset of pernicious anemia and the fact that true dietary deficiency of vitamin B_{12} , even among strict vegetarians, is extremely rare.

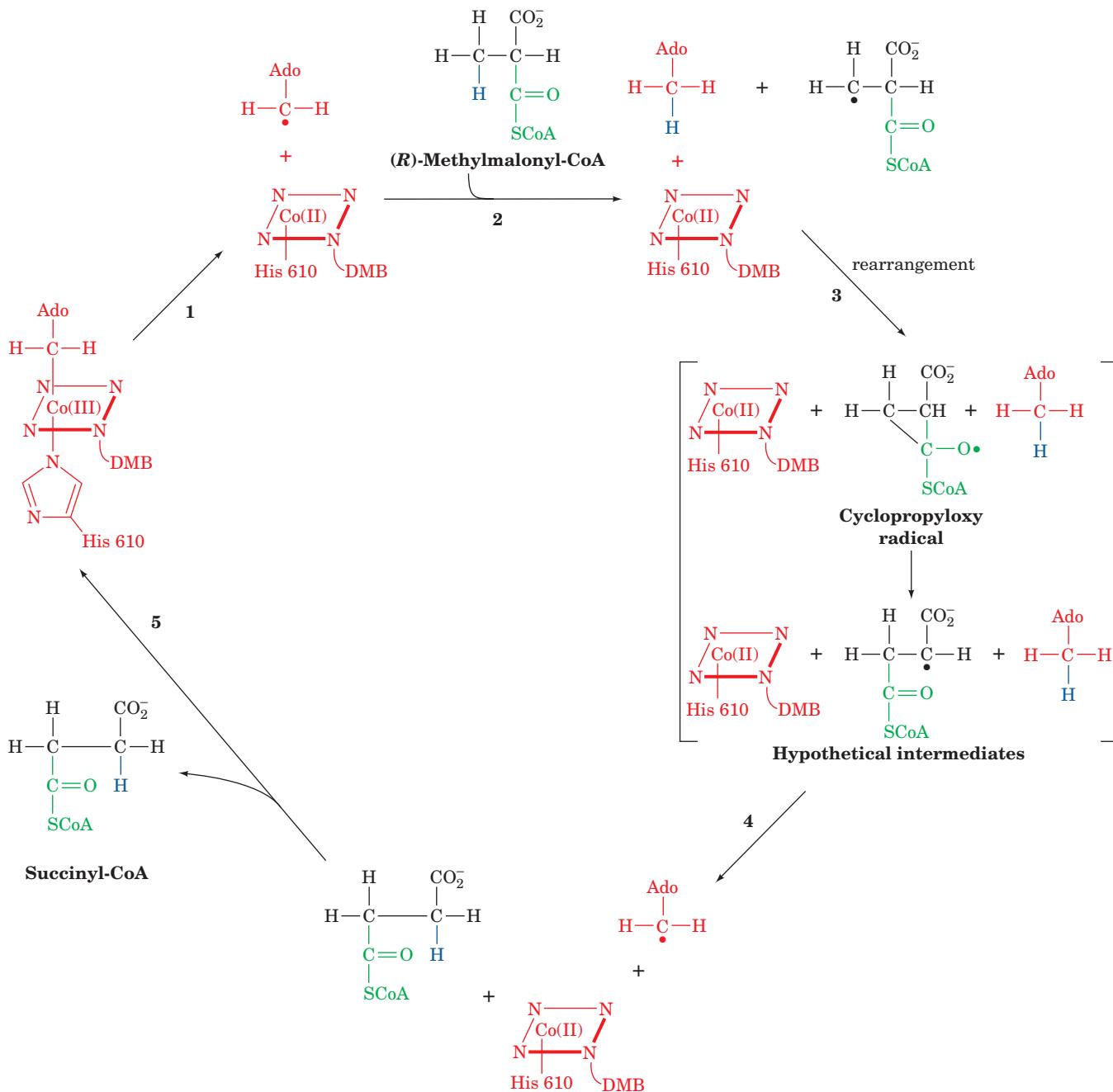


Figure 25-23 Proposed mechanism of methylmalonyl-CoA mutase. (1) The homolytic cleavage of the C—Co(III) bond yielding a 5'-deoxyadenosyl radical and cobalamin in its Co(II) oxidation state. (2) Abstraction of a hydrogen atom from the methylmalonyl-CoA by the 5'-deoxyadenosyl radical, thereby generating a methylmalonyl-CoA radical. (3) Carbon skeleton

rearrangement to form a succinyl-CoA radical via a proposed cyclopropoxy radical intermediate. (4) Abstraction of a hydrogen atom from 5'-deoxyadenosine by the succinyl-CoA radical to form succinyl-CoA and regenerate the 5'-deoxyadenosyl radical. (5) Release of succinyl-CoA and reformation of the coenzyme.

F. Peroxisomal β Oxidation

In mammalian cells, the bulk of β oxidation occurs in the mitochondria, but peroxisomes (Fig. 25-24) also oxidize fatty acids, particularly those with very long chains or branched chains. *Peroxisomal β oxidation in animals functions to shorten very long chain fatty acids (>22 C atoms) so as to facilitate their degradation by the mitochondrial β oxidation system. In yeast and plants, fatty acid oxidation occurs exclusively in the peroxisomes and glyoxysomes (specialized peroxisomes, Sections 23-2 and 1-2Ad).*

The peroxisomal pathway results in the same chemical changes to fatty acids as does the mitochondrial pathway, although the enzymes in these two organelles are different. The protein that transports very long-chain fatty acids into the peroxisome, **ALD protein** (see below), does not have a carnitine requirement. The very long-chain fatty acids that enter this compartment are activated by a peroxisomal very long-chain acyl-CoA synthetase to form their CoA esters, and are oxidized directly. The shorter chain acyl products of this β -oxidation process are then linked to carnitine for transport into mitochondria for further oxidation.

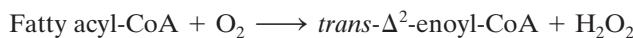
a. Adrenoleukodystrophy Is Caused by a Defect in ALD Protein

Adrenoleukodystrophy (ALD) is a rare X-linked inherited disease that results in progressive brain damage and adrenal gland failure. It causes very long-chain saturated fatty acids to accumulate in the blood and destroy the insulating myelin sheath surrounding the axons of many neurons (Section 20-5Bc). Its varied neurological symptoms present (become evident) between the ages of 4 and 10 years and are usually fatal within 1 to 10 years (except after a successful bone marrow transplant). ALD is caused by a defective ALD protein, an ABC transporter (Section 20-3E). Thus in ALD patients, **lignoceric acid** (24:0; recall that the symbol $n:m$ indicates a C_n fatty acid with m double bonds) is converted to lignoceroyl-CoA at only 13% of the normal rate, although once formed, it undergoes β oxidation at the normal rate.

b. Peroxisomal β Oxidation Differs in Detail from Mitochondrial β Oxidation

The β -oxidation pathway in peroxisomes differs from that in mitochondria as follows:

1. The first enzyme in the peroxisomal pathway, **acyl-CoA oxidase**, catalyzes the reaction



This reaction involves participation of an FAD cofactor but differs from its mitochondrial counterpart in that the abstracted electrons are transferred directly to O_2 rather than passing through the electron-transport chain with its concomitant oxidative phosphorylation (Fig. 25-12). Peroxisomal fatty acid oxidation is therefore less efficient than the mitochondrial process by two ATPs for each C_2 cycle. The H_2O_2 produced is disproportionated to H_2O and O_2

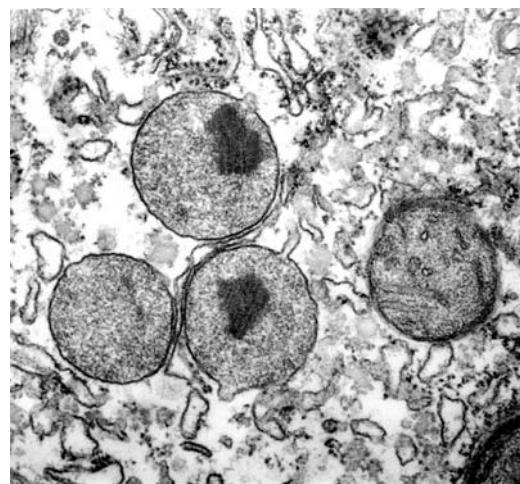


Figure 25-24 **Peroxisomes.** These membrane-bounded organelles perform a variety of metabolic functions, including the oxidation of very long chain fatty acids. [© Donald Fawcett/Visuals Unlimited.]

through the action of peroxisomal catalase (Section 1-2Ad).

2. Peroxisomal enoyl-CoA hydratase and 3-L-hydroxyacyl-CoA dehydrogenase are activities that occur on a single polypeptide and therefore join the growing list of multifunctional enzymes. The reactions catalyzed are identical to those of the mitochondrial system (Fig. 25-12).

3. Peroxisomal thiolase has a different chain-length specificity than its mitochondrial counterpart. It is almost inactive with acyl-CoAs of length C_8 or less so that fatty acids are incompletely oxidized by peroxisomes.

Although peroxisomal β oxidation is not dependent on the transport of acyl groups into the peroxisome as their carnitine esters, the peroxisome contains carnitine acyl-transferases. Acyl-CoAs that have been chain-shortened by peroxisomal β oxidation are thereby converted to their carnitine esters. These substances, for the most part, passively diffuse out of the peroxisome to the mitochondrion, where they are oxidized further.

G. Minor Pathways of Fatty Acid Oxidation

β Oxidation is blocked by an alkyl group at the C_β of a fatty acid, and thus at any odd-numbered carbon atom. One such branched-chain fatty acid, a common dietary component, is **phytanic acid**. This metabolic breakdown product of chlorophyll's phytol side chain (Section 24-2A) is present in dairy products, ruminant fats, and fish although, surprisingly, chlorophyll itself is but a poor dietary source of phytanic acid for humans. The oxidation of branched-chain fatty acids such as phytanic acid is facilitated by α **oxidation** (Fig. 25-25). In this process, the fatty acid is converted to its CoA thioester and its C_α is hydroxylated by the Fe^{2+} -containing **phytanoyl-CoA hydroxylase**. The resulting

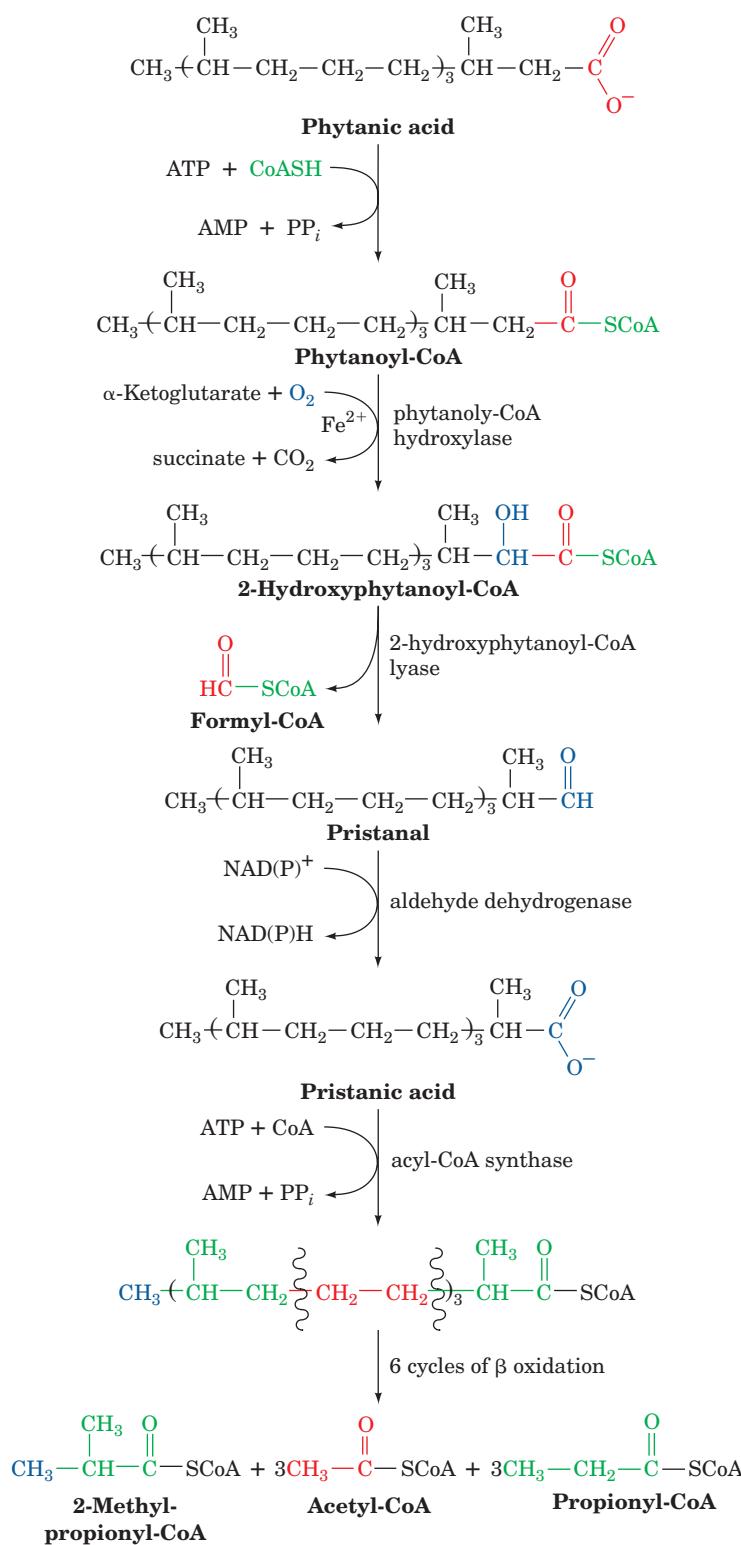


Figure 25-25 Pathway of α oxidation of fatty acids. Phytanic acid, a degradation product of the phytol side chain of chlorophyll, is metabolized through α oxidation to **pristanic acid** followed by β oxidation.

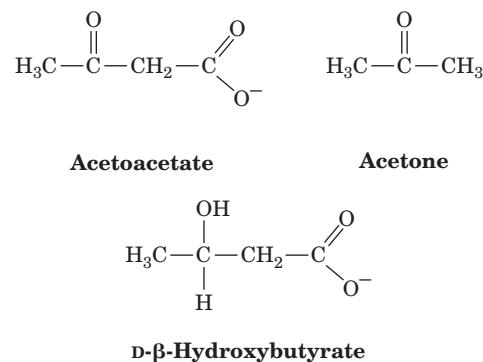
CoA thioester is, in effect, oxidatively decarboxylated to yield a new fatty acid with an unsubstituted C_β. Further degradation of the molecule can then continue via six cycles of normal β oxidation to yield three propionyl-CoAs, three acetyl-CoAs, and one 2-methylpropionyl-CoA (which is converted to succinyl-CoA).

A rare genetic defect, **Refsum's disease** or **phytanic acid storage syndrome**, results from the accumulation of this metabolite throughout the body. The disease, which is characterized by progressive neurological difficulties such as tremors, unsteady gait, and poor night vision, results from a defective phytanoyl-CoA hydroxylase. Its symptoms can therefore be attenuated by a diet that restricts the intake of phytanic acid-containing foods.

Medium- and long-chain fatty acids are converted to dicarboxylic acids through **ω oxidation** (oxidation of the last carbon atom). This process, which is catalyzed by enzymes of the ER, involves hydroxylation of a fatty acid's C_ω atom by a **cytochrome P450**, a monooxygenase that utilizes NADPH and O₂ (Section 15-4Bc). This CH₂—OH group is then oxidized to a carboxyl group, converted to a CoA derivative at either end, and oxidized via the β -oxidation pathway. ω Oxidation is probably of only minor significance in fatty acid oxidation.

3 KETONE BODIES

Acetyl-CoA produced by oxidation of fatty acids in liver mitochondria can be further oxidized via the citric acid cycle as is discussed in Chapter 21. A significant fraction of this acetyl-CoA has another fate, however. By a process known as **ketogenesis**, which occurs primarily in liver mitochondria, acetyl-CoA is converted to **acetoacetate** or **D- β -hydroxybutyrate**. These compounds, which together with **acetone** are somewhat inaccurately referred to as **ketone bodies**,



serve as important metabolic fuels for many peripheral tissues, particularly heart and skeletal muscle. The brain, under normal circumstances, uses only glucose as its energy source (fatty acids are unable to pass the blood-brain barrier), but during starvation, ketone bodies become the brain's major fuel source (Section 27-4A). Ketone bodies are water-soluble equivalents of fatty acids.

Acetoacetate formation occurs in three reactions (Fig. 25-26):

1. Two molecules of acetyl-CoA are condensed to **acetoacetyl-CoA** by thiolase (also called **acetyl-CoA acetyltransferase**) working in the reverse direction from the way it does in the final step of β oxidation (Section 25-2Cd).

2. Condensation of the acetoacetyl-CoA with a third acetyl-CoA by **HMG-CoA synthase** forms **β -hydroxy- β -methylglutaryl-CoA (HMG-CoA)**. The mechanism of this reaction resembles the reverse of the thiolase reaction (Fig. 25-15) in that an active site thiol group forms an acyl-thioester intermediate.

3. Degradation of HMG-CoA to acetoacetate and acetyl-CoA in a mixed aldol-Claisen ester cleavage is catalyzed by **HMG-CoA lyase**. The mechanism of this reaction is analogous to the reverse of the citrate synthase reaction (Section 21-3A). (HMG-CoA is also a precursor in cholesterol biosynthesis and hence may be diverted to this purpose as is discussed in Section 25-6A.)

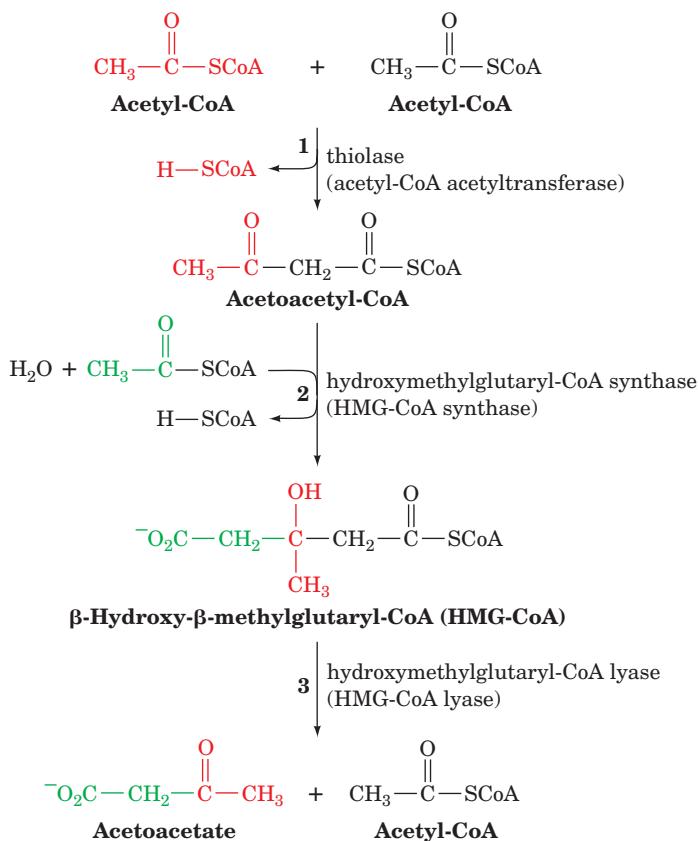


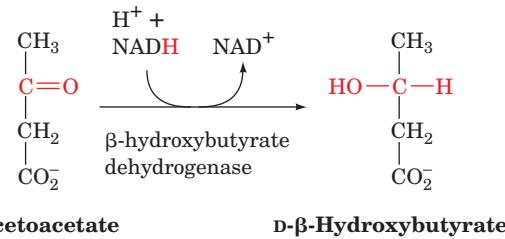
Figure 25-26 Ketogenesis: the enzymatic reactions forming acetoacetate from acetyl-CoA. (1) Two molecules of acetyl-CoA condense to form acetoacetyl-CoA in a thiolase-catalyzed reaction. (2) A Claisen ester condensation of the acetoacetyl-CoA with a third acetyl-CoA to form β -hydroxy- β -methylglutaryl-CoA (HMG-CoA) as catalyzed by HMG-CoA synthase. (3) The degradation of HMG-CoA to acetoacetate and acetyl-CoA in a mixed aldol-Claisen ester cleavage catalyzed by HMG-CoA lyase.

The overall reaction catalyzed by HMG-CoA synthase and HMG-CoA lyase is



One may well ask why this apparently simple hydrolysis reaction occurs in such an indirect manner. The answer is unclear but may lie in the regulation of the process.

Acetoacetate may be reduced to D- β -hydroxybutyrate by **β -hydroxybutyrate dehydrogenase**:



Note that this product is the stereoisomer of the L- β -hydroxyacyl-CoA that occurs in the β -oxidation pathway. Acetoacetate, being a β -keto acid, also undergoes relatively facile nonenzymatic decarboxylation to acetone and

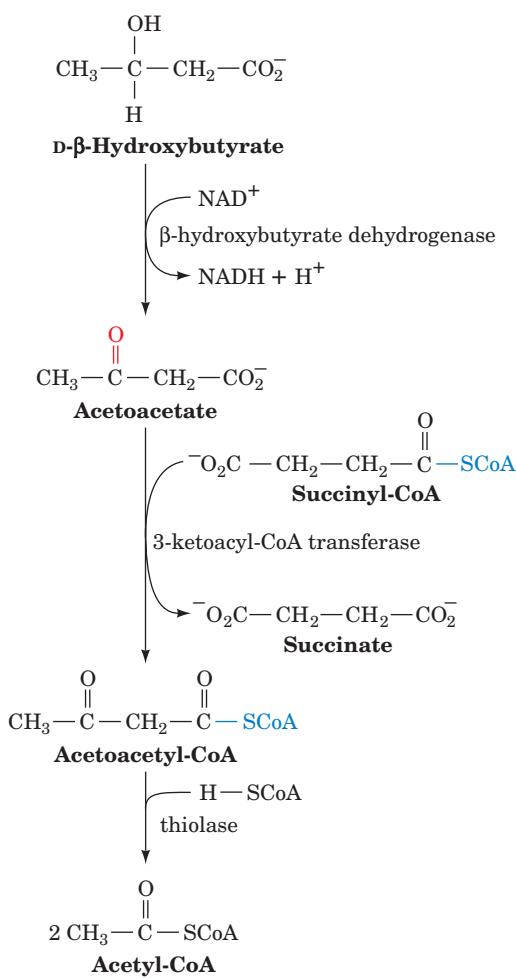


Figure 25-27 The metabolic conversion of ketone bodies to acetyl-CoA.

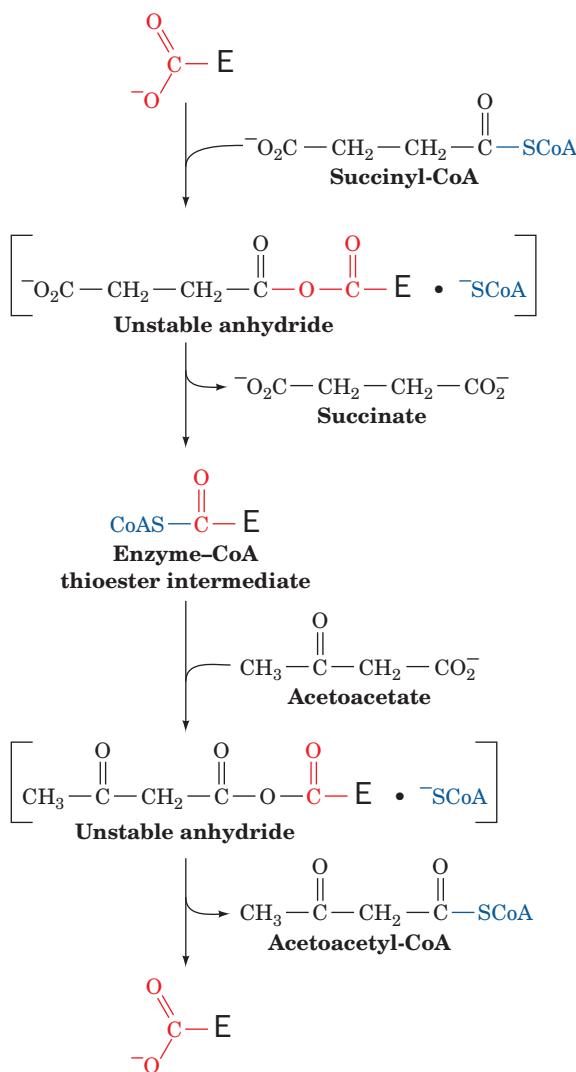


Figure 25-28 Proposed mechanism of 3-ketoacyl-CoA transferase involving an enzyme-CoA thioester intermediate.

CO_2 . Indeed, the breath of individuals with **ketosis** (also called **ketoacidosis**), a potentially pathological condition in which acetoacetate is produced faster than it can be metabolized (a symptom of diabetes; Section 27-4B), has the characteristic sweet smell of acetone.

The liver releases acetoacetate and β -hydroxybutyrate, which are carried by the bloodstream to the peripheral tissues for use as alternative fuels. There, these products are converted to acetyl-CoA as is diagrammed in Fig. 25-27. The proposed reaction mechanism of **3-ketoacyl-CoA transferase** (Fig. 25-28), which catalyzes this pathway's second step, involves the participation of an active site carboxyl group both in an enzyme-CoA thioester intermediate and in an unstable anhydride. Succinyl-CoA, which acts as the CoA donor in this reaction, can also be converted to succinate with the coupled synthesis of GTP in the succinyl-CoA synthetase reaction of the citric acid cycle (Section 21-3Ea). The "activation" of acetoacetate bypasses this

step and therefore "costs" the free energy of GTP hydrolysis. The liver lacks 3-ketoacyl-CoA transferase, which permits it to supply ketone bodies to other tissues.

4 FATTY ACID BIOSYNTHESIS

Fatty acid biosynthesis occurs through condensation of C_2 units, the reverse of the β -oxidation process. Through isotopic labeling techniques, David Rittenberg and Konrad Bloch demonstrated, in 1945, that these condensation units are derived from acetic acid. Acetyl-CoA was soon proven to be a precursor of the condensation reaction, but its mechanism remained obscure until the late 1950s when Salih Wakil discovered a requirement for bicarbonate in fatty acid biosynthesis and malonyl-CoA was shown to be an intermediate. In this section, we discuss the reactions of fatty acid biosynthesis.

A. Pathway Overview

The pathway of fatty acid synthesis differs from that of fatty acid oxidation. This situation, as we saw in Section 18-1D, is typically the case of opposing biosynthetic and degradative pathways because it permits them both to be thermodynamically favorable and independently regulated under similar physiological conditions. Figure 25-29 outlines fatty acid oxidation and synthesis with emphasis on the differences between these pathways. Whereas fatty acid oxidation occurs in the mitochondrion and utilizes fatty acyl-CoA esters, fatty acid biosynthesis occurs in the cytosol with, as Roy Vagelos discovered, the growing fatty acids esterified to **acyl-carrier protein (ACP)** (Fig. 25-30). ACP, like CoA, contains a phosphopantetheine group that forms thioesters with acyl groups. The phosphopantetheine phosphoryl group is esterified to a Ser OH group of ACP, whereas in CoA it is esterified to AMP. In animals, ACP is part of a large multifunctional protein (Type I ACP; see below), whereas in *E. coli* it is a 125-residue polypeptide (Type II ACP). The phosphopantetheine group is transferred from CoA to apo-ACP to form the active holo-ACP by **phosphopantetheine transferase** (alternatively, **ACP synthase**).

The redox coenzymes of the animal fatty acid oxidative and biosynthetic pathways differ (NAD^+ and FAD for oxidation; NADPH for biosynthesis) as does the stereochemistry of their intermediate steps, but their main difference is the manner in which C_2 units are removed from or added to the fatty acyl thioester chain. In the oxidative pathway, β -ketothiolase catalyzes the cleavage of the $\text{C}_\alpha-\text{C}_\beta$ bond of β -ketoacyl-CoA so as to produce acetyl-CoA and a new fatty acyl-CoA, which is shorter by a C_2 unit. The $\Delta G'$ of this reaction is very close to zero so it can also function in the reverse direction (ketone body formation). In the biosynthetic pathway, the condensation reaction is coupled to the hydrolysis of ATP, thereby driving the reaction to completion. This process involves two steps: (1) the ATP-dependent carboxylation of acetyl-CoA by **acetyl-CoA carboxylase (ACC)** to form **malonyl-CoA**, and (2) the

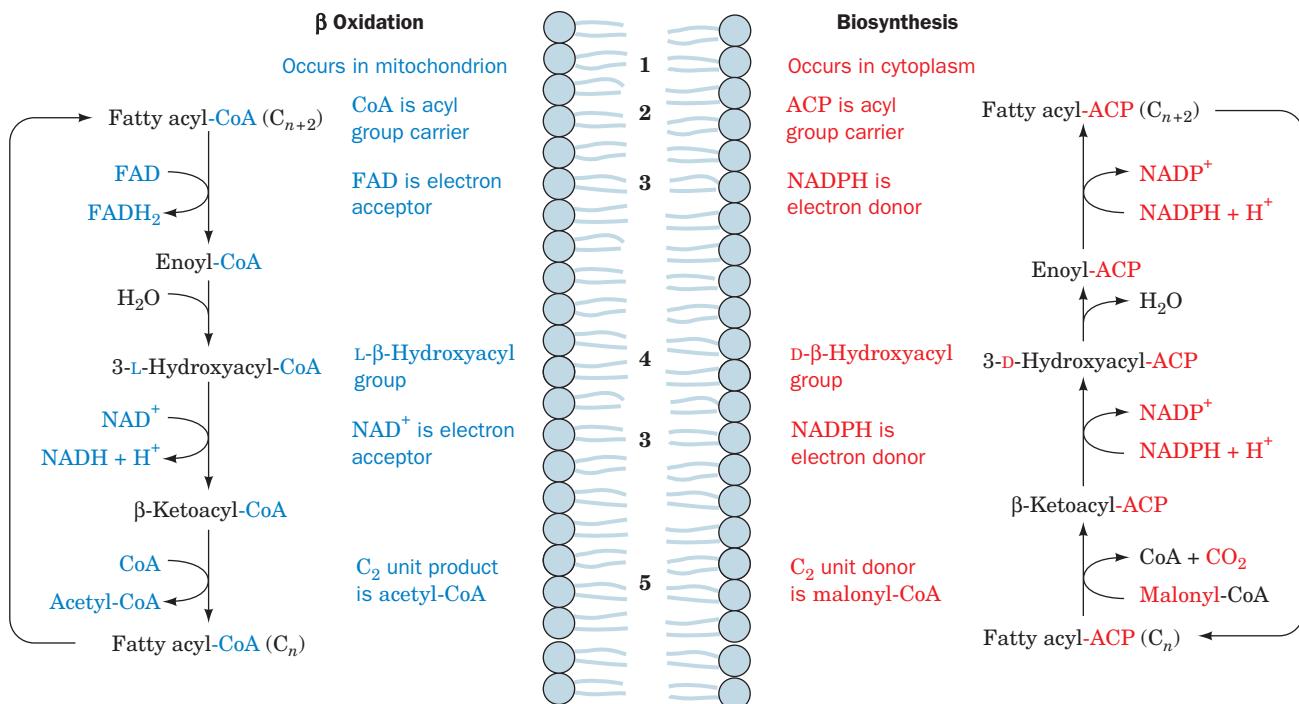


Figure 25-29 Comparison of fatty acid β oxidation and fatty acid biosynthesis. Differences occur in (1) cellular location, (2) acyl group carrier, (3) electron acceptor/donor, (4) stereo-

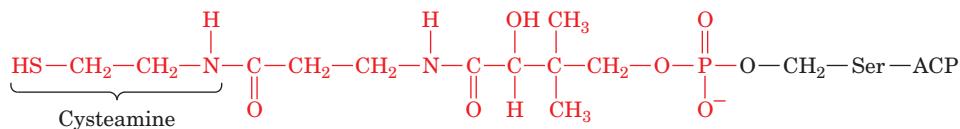
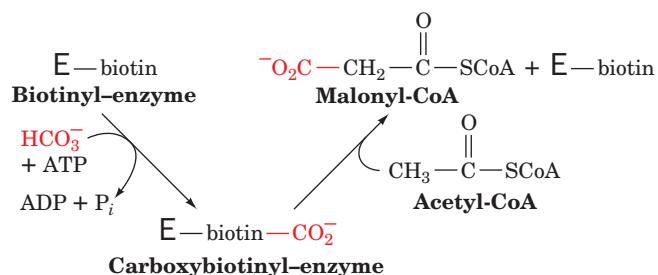
chemistry of the hydration/dehydration reaction, and (5) the form in which C_2 units are produced/donated. 

exergonic decarboxylation of the malonyl group in the condensation reaction catalyzed by **fatty acid synthase**. These enzymes are discussed below.

B. Acetyl-CoA Carboxylase

ACC is a biotin-dependent enzyme that catalyzes the first committed step of fatty acid biosynthesis and one of its rate-controlling steps. It is a member of a family of biotin-dependent carboxylases that, in humans, has only three members besides ACC: propionyl-CoA carboxylase (Section 25-2Ea), pyruvate carboxylase (Fig. 23-4), and **β -methylcrotonyl-CoA carboxylase** (which participates in

the degradation of leucine; Section 26-3F). The ACC reaction, like those of the other biotin-dependent carboxylases, occurs in two steps, a CO_2 activation and a carboxylation:



Phosphopantetheine prosthetic group of ACP

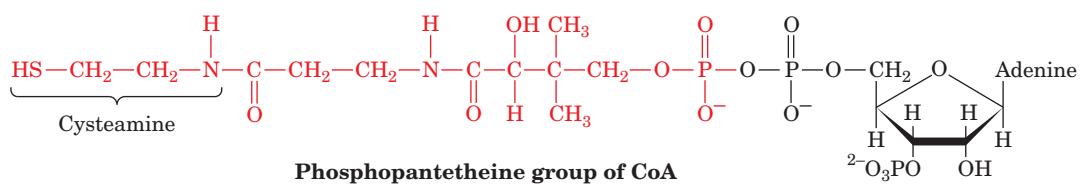


Figure 25-30 The phosphopantetheine group in acyl-carrier protein (ACP) and in CoA.

E. coli ACC is a multienzyme complex in which these steps are catalyzed by separate subunits: the homodimer biotin carboxylase (BC; 456 residues) and the $\alpha_2\beta_2$ heterotetramer carboxyltransferase (CT; 319 and 304 residues). In addition, the biotin is linked as a biocytin residue (Fig. 23-3b) to the biotin carboxyl-carrier protein (BCCP; 156 residues), which forms homodimers. In contrast, mammalian and avian ACCs contain both enzymatic activities as well as the biotin carboxyl carrier on a single 2346-residue polypeptide chain in the order BC-BCCP-CT (which differs from the order in pyruvate carboxylase, which is BC-CT-BCCP; Section 23-1Ae). The structure of an intact ACC has not been determined, although the X-ray structure of the *E. coli* BC subunit closely resembles the BC domain of pyruvate carboxylase (Fig. 23-5a). Interestingly, however, the CT domains of the various biotin-dependent carboxylases differ greatly in sequence and structure.

a. Acetyl-CoA Carboxylase Is Regulated by Hormonally Controlled Reversible Phosphorylation

ACC is subject to hormonal regulation. Glucagon as well as epinephrine and norepinephrine (adrenaline and norepinephrine; Section 18-3Ea) trigger the enzyme's cAMP-dependent increase in phosphorylation, which inactivates the enzyme. Insulin, on the other hand, stimulates enzyme dephosphorylation and thus its activation.

The mechanism by which cAMP causes an increase in the phosphorylation state of ACC is interesting. ACC is phosphorylated, *in vitro*, by two different kinases, the cAMP-dependent protein kinase A (PKA; Section 18-3Cb) at Ser 77 and **AMP-dependent protein kinase (AMPK; Sections 25-5a and 27-1)** (which is cAMP independent) at Ser 79, 1200, and 1215. Yet, when liver cells are incubated with cAMP-elevating hormones in the presence of ^{32}P -ATP, Ser 77 is found not to be labeled. Evidently, an increase in [cAMP] results in a phosphorylation increase at sites modified by AMPK rather than by PKA. How can this be? It appears that, *in vivo*, the cAMP-dependent increase in phosphorylation occurs not through the phosphorylation of new sites but, rather, through the inhibition of dephosphorylation of previously phosphorylated positions. We have already seen such a mechanism in operation in the control of glycogen metabolism, where the cAMP-dependent phosphorylation of phosphoprotein phosphatase inhibitor-1 causes the inhibition of dephosphorylation (Section 18-3C). In the case of ACC, however, dephosphorylation is catalyzed by **phosphoprotein phosphatase-2A**, which is not affected by phosphoprotein phosphatase inhibitor-1. The mechanism by which PKA causes the increase in phosphorylation associated with AMPK activity is, as yet, unknown.

b. Avian and Mammalian Acetyl-CoA Carboxylases Undergo Enzyme Polymerization on Activation

Electron microscopy reveals that both avian and mammalian ACCs form long filaments of 20 to 40 protomers (Fig. 25-31). This polymeric form of the enzyme is catalytically active but the protomer is not. The rate of fatty acid

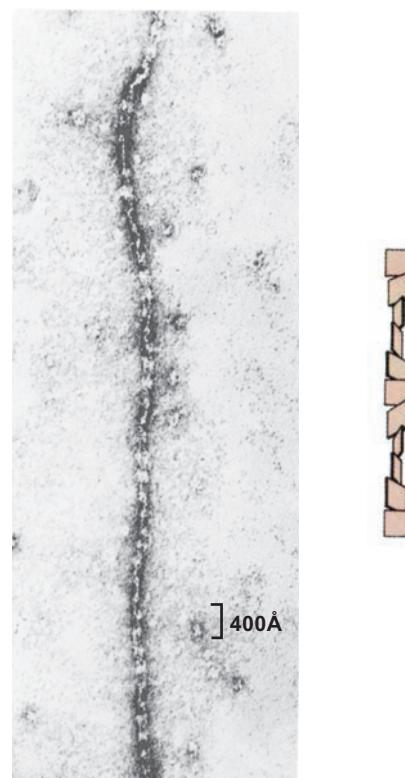
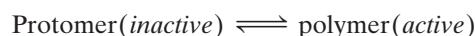


Figure 25-31 Association of acetyl-CoA carboxylase protomers. An electron micrograph with an accompanying interpretive drawing indicates that filaments of avian liver acetyl-CoA carboxylase consist of linear chains of flat rectangular protomers. [Courtesy of Malcolm Lane, The Johns Hopkins University School of Medicine.]

biosynthesis is therefore controlled by the position of the equilibrium between these forms:



Phosphorylation favors the inactive protomer while dephosphorylation favors the active polymer. Several metabolites also affect the activity of acetyl-CoA carboxylase. Citrate promotes the polymerization of ACC, whereas palmitoyl-CoA and other fatty acyl-CoA's promote its depolymerization. Thus, cytosolic citrate, whose concentration increases when the mitochondrial acetyl-CoA concentration builds up (Section 25-4D), activates fatty acid biosynthesis and hence is a feedforward activator, whereas palmitoyl-CoA, the pathway product, is a feedback inhibitor.

c. Mammalian Acetyl-CoA Carboxylase Has Two Major Isoforms

There are two major isoforms of mammalian ACC. **ACC1** occurs in adipose tissue and **ACC2** occurs in tissues that oxidize but do not synthesize fatty acids, such as heart muscle. Tissues that both synthesize and oxidize fatty acids, such as liver, contain both isoforms, which are homologous although the genes encoding them are located on different chromosomes. What is the function of ACC2? The product of the ACC-catalyzed reaction, malonyl-CoA, strongly inhibits

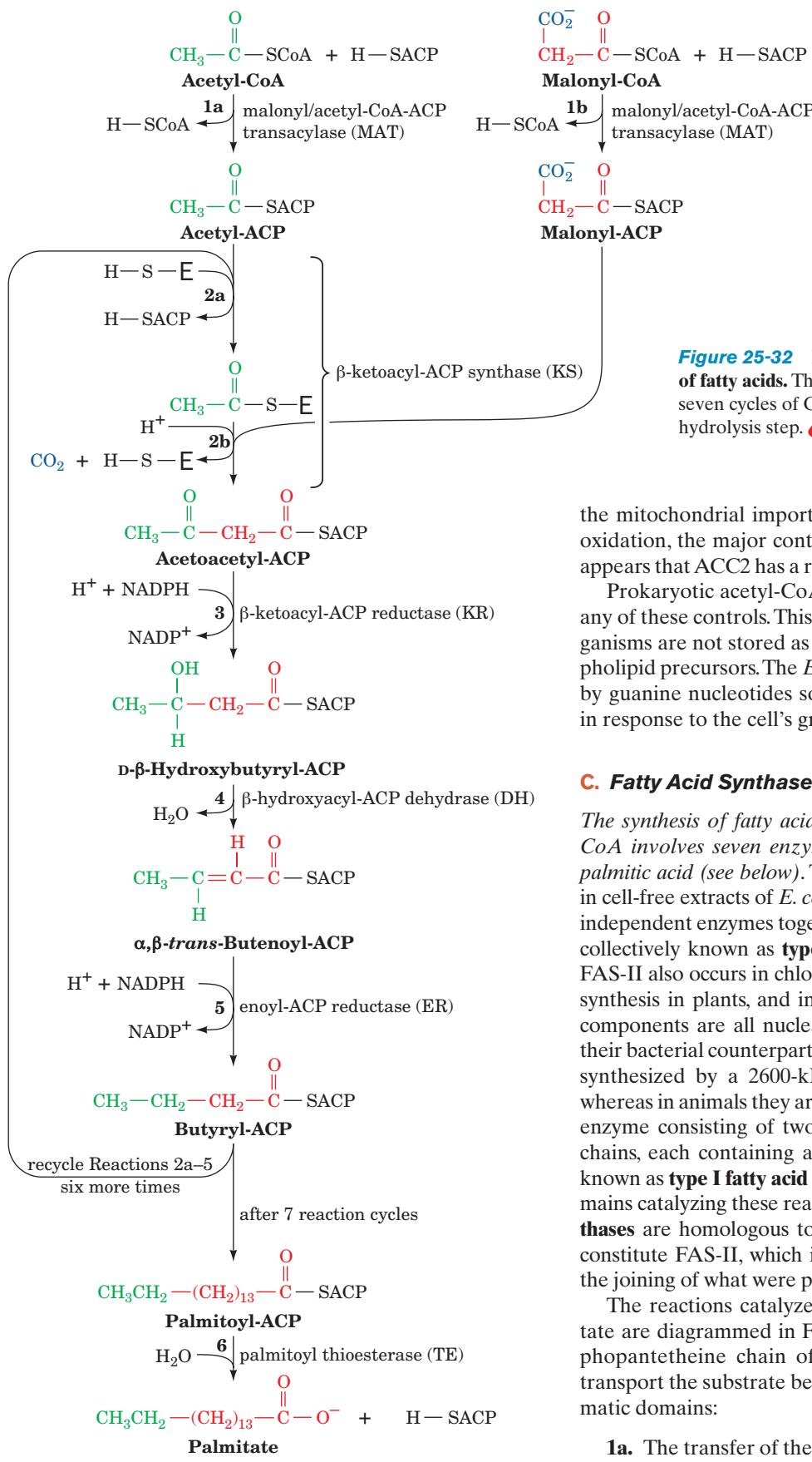


Figure 25-32 Reaction cycle for the biosynthesis of fatty acids. The biosynthesis of palmitate requires seven cycles of C_2 elongation followed by a final hydrolysis step. 

the mitochondrial import of fatty acyl-CoA for fatty acid oxidation, the major control point for this process. Thus it appears that ACC2 has a regulatory function (Section 25-5).

Prokaryotic acetyl-CoA carboxylases are not subject to any of these controls. This is because fatty acids in these organisms are not stored as fats but function largely as phospholipid precursors. The *E. coli* enzyme is instead regulated by guanine nucleotides so that fatty acids are synthesized in response to the cell's growth requirements.

C. Fatty Acid Synthase

The synthesis of fatty acids from acetyl-CoA and malonyl-CoA involves seven enzymatic reactions that yield mainly palmitic acid (see below). These reactions were first studied in cell-free extracts of *E. coli*, in which they are catalyzed by independent enzymes together with ACP. These proteins are collectively known as **type II fatty acid synthase (FAS-II)**. FAS-II also occurs in chloroplasts, the only site of fatty acid synthesis in plants, and in vertebrate mitochondria, whose components are all nuclear encoded and closely resemble their bacterial counterparts. In fungi, however, fatty acids are synthesized by a 2600-kD $\alpha_6\beta_6$ multifunctional enzyme, whereas in animals they are synthesized by a multifunctional enzyme consisting of two identical ~ 275 -kD polypeptide chains, each containing all seven activities plus ACP and known as **type I fatty acid synthase (FAS-I)**. Most of the domains catalyzing these reactions in these so-called **megasynthases** are homologous to the corresponding proteins that constitute FAS-II, which indicates that they arose through the joining of what were previously independent proteins.

The reactions catalyzed by FAS-I to synthesize palmitate are diagrammed in Fig. 25-32. The long flexible phosphopantetheine chain of ACP (Fig. 25-30) functions to transport the substrate between the protein's various enzymatic domains:

- 1a. The transfer of the acetyl group from acetyl-CoA to ACP to yield acetyl-ACP as catalyzed by **malonyl/acetyl-CoA-ACP transacylase (MAT)**.

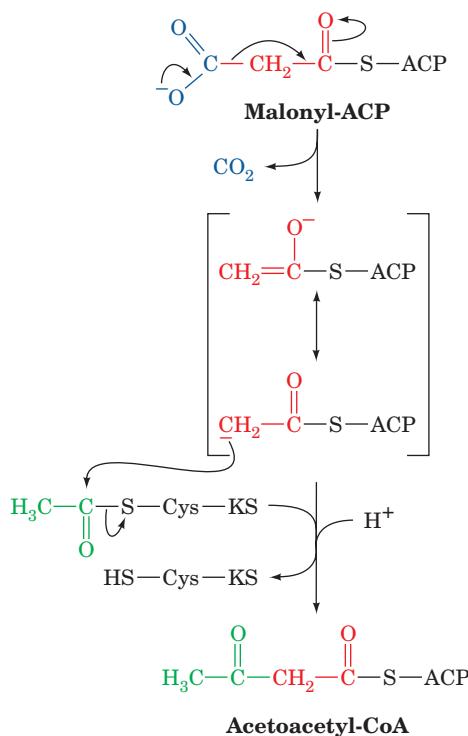


Figure 25-33 The mechanism of carbon–carbon bond formation in fatty acid biosynthesis. The condensation of an acetyl group on the active site Cys of β -ketoacyl-ACP synthase (KS) with a malonyl group on the phosphopantetheine arm of ACP forms a β -ketoacyl-ACP. The reaction is driven by the exergonic elimination of CO_2 from the malonyl group to generate a resonance-stabilized acetyl-ACP carbanion intermediate that functions as a good nucleophile.

2a. The loading of **β -ketoacyl-ACP synthase (KS; also known as condensing enzyme)** by the transfer of the acetyl group from ACP to a KS Cys residue, thus maintaining the acetyl group's thioester linkage.

1b. The formation of malonyl-ACP in a reaction analogous to that of Reaction 1a, which in animals is catalyzed by the same enzyme, MAT.

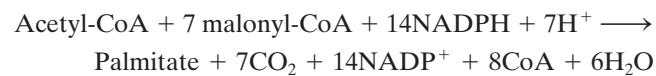
2b. The coupling of the acetyl group to the C_β of the malonyl group on the ACP with the malonyl group's accompanying decarboxylation so as to form acetoacetyl-ACP and free the KS active site Cys-SH group (Fig. 25-33). Consequently, the CO_2 taken up in the acetyl-CoA carboxylase reaction (Section 25-4B) does not appear in the product fatty acid. Rather, the decarboxylation functions to drive carbon–carbon bond formation in the condensation reaction, which through the acetyl-CoA carboxylase reaction, is coupled to ATP hydrolysis.

3–5. The reduction, dehydration, and further reduction of acetoacetyl-ACP so as to form **butyryl-ACP** as sequentially catalyzed by **β -ketoacyl-ACP reductase (KR)**, **β -hydroxyacyl-ACP dehydrase (DH)**, and **enoyl-ACP reductase (ER)**. The coenzyme in both reductive steps is NADPH, whereas in β oxidation, the analogs of Reactions 3 and 5, respectively, use NAD^+ and FAD (Fig. 25-29). Moreover, Reaction 3 produces and Reaction 4 requires a $\text{D-}\beta$ -hydroxyacyl group, whereas the analogous reactions in β oxidation involve the corresponding L isomer.

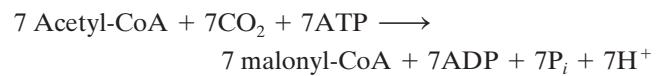
2a to 5 Repeat. The butyryl group from the butyryl-ACP is transferred to the Cys-SH of KS. Thus the acetyl group with which the system was initially loaded has been elongated by a C_2 unit. The ACP is “reloaded” with a malonyl group (Step 1b), and another cycle of C_2 elongation occurs. This process occurs altogether seven times to yield **palmitoyl-ACP**.

6. The palmitoyl-ACP thioester bond is hydrolyzed by **palmitoyl thioesterase (TE)**, yielding palmitate, the normal product of the fatty acid synthase pathway, and regenerating the enzyme for a new round of synthesis.

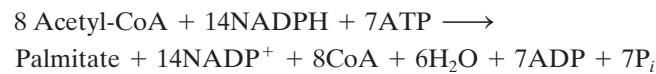
The stoichiometry of palmitate synthesis therefore is



Since the 7 malonyl-CoA are derived from acetyl-CoA as follows:



the overall stoichiometry for palmitate biosynthesis is



a. Animal FAS-I Is a Flexible X-Shaped Dimer

Most but not all of the enzymatic activities of animal FAS-I remain functional when this dimeric enzyme is dissociated into monomers. Moreover, fragments resulting from the limited proteolysis of animal FAS-I exhibit many of the enzymatic activities of the intact protein. Apparently, *contiguous stretches of its polypeptide chain fold to form a series of autonomous domains, each with a specific but different catalytic activity*. Several other enzymes, for example, mammalian acetyl-CoA carboxylase (Section 25-4B), exhibit similar multifunctionality but none has as many separate catalytic activities as does animal FAS-I. The order of the domains along the animal FAS-I polypeptide chain is indicated in Fig. 25-34. Three of these domains besides ACP lack enzymatic activity: a linker domain



Figure 25-34 Domain organization of porcine FAS-I at approximate sequence scale. [Modified from a drawing by Timm Maier and Nenad Ban, ETH Zurich, Switzerland.]

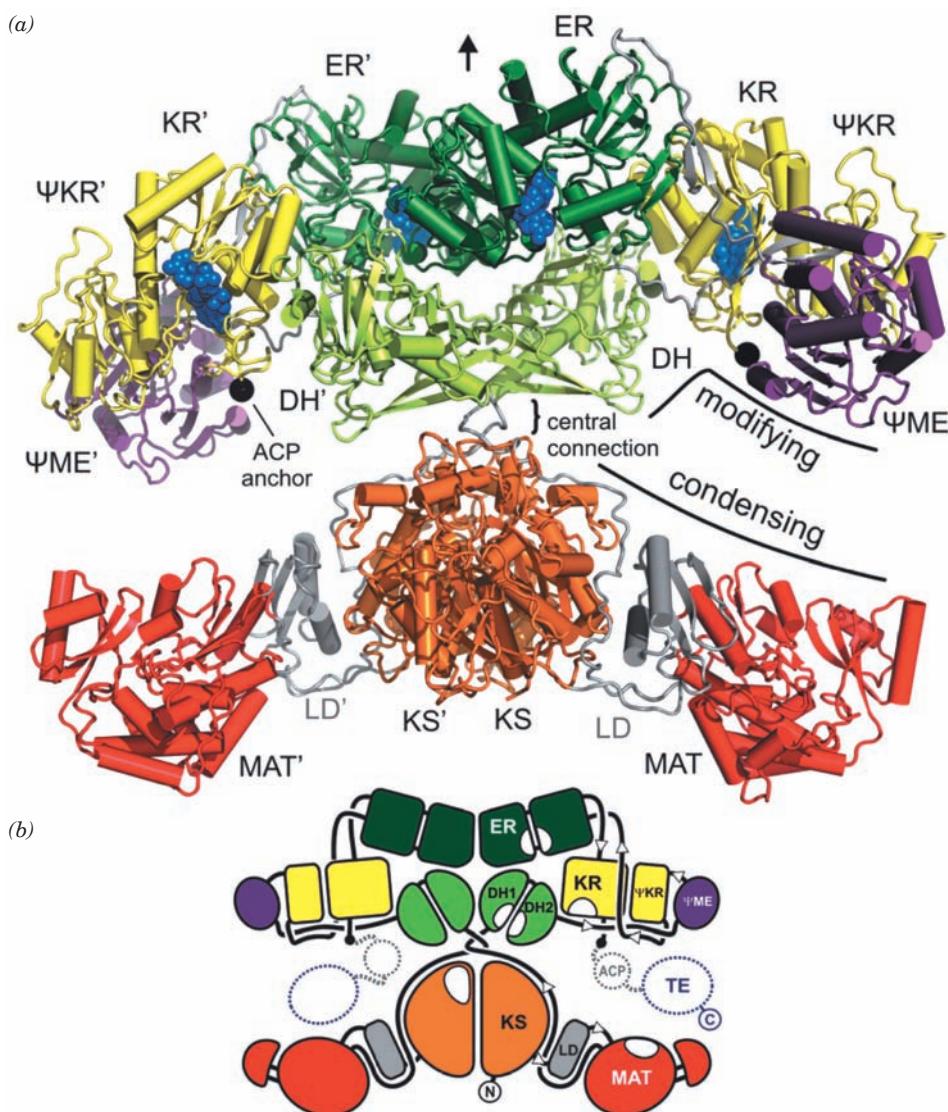


Figure 25-35 X-ray structure of porcine FAS-I in complex with NADP^+ . (a) The $\sim 190\text{-}\text{\AA}$ wide homodimer, as viewed perpendicular to its pseudo-2-fold axis (vertical black arrow), with its various domains colored as in Fig. 25-34 and its linkers gray. The bound NADP^+ cofactors are drawn in space-filling form in blue. The attachment sites of the disordered ACP/TE

domains are indicated by black dots. The domain names of the second subunit have appended primes. (b) A corresponding schematic diagram indicating how the various domains are linked. [Courtesy of Timm Maier and Nenad Ban, ETH Zurich, Switzerland. PDBid 2VZ9.]

(LD) that bridges the KS and MAT domains; a pseudo-methyltransferase (ΨME) domain, so named because it is homologous to the methyltransferase family; and a pseudo-ketoreductase (ΨKR) domain, so named because it is a truncated form of the KR domain.

The X-ray structure of porcine FAS-I in complex with NADP^+ , determined by Nenad Ban, reveals a pseudo-2-fold symmetric, X-shaped dimer (Fig. 25-35). Its two subunits associate via an extensive interface involving over 150 residues per chain from the KS, ER, and DH domains. However the upper and lower portions of the X are only loosely connected. The lower portion of the X contains the enzyme's two condensing activities, MAT and KS (which

catalyze reactions 1 and 2 in Fig. 25-32), whereas the upper portion of the X contains the enzyme's β -carbon modifying activities, DH, KR, and ER (which catalyze reactions 3–5 in Fig. 25-32). The C-terminal ACP and TE domains are flexibly tethered to the enzyme and are therefore unobserved. The KR-ACP linker consists of ~ 13 residues and can therefore span a distance of up to $\sim 40\text{ \AA}$, whereas the ACP-TE linker has a length of ~ 25 residues and can therefore extend $\sim 80\text{ \AA}$.

Each of the two reaction chambers of porcine FAS-I is lined by the full set of catalytic domains from one subunit, all of which must be visited by the phosphopantethiene arm of an ACP to carry out all the FAS reac-

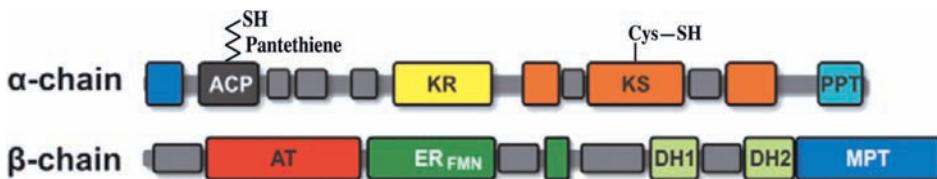


Figure 25-36 Domain organization of fungal FAS-I at approximate sequence scale. [Modified from a drawing by Simon Jenni and Nenad Ban, ETH Zurich, Switzerland.]

tions. The catalytic sites of these domains occur on both the front and back faces of the protein and their distribution has no discernable relationship to the order of the reactions in the catalytic cycle. In the structure shown in Fig. 25-35a, the length of the KR-ACP linker would allow the ACP to reach all the catalytic sites in the same reaction chamber but not those in the opposite reaction chamber (e.g., the distance between the ACP attachment site on one subunit and the MAT domain in the other is ~ 135 Å). Nevertheless, mutational studies by Stuart Smith in which ACP on one subunit and MAT or KS on the other subunit were inactivated yielded functional enzymes albeit with reduced activities, thus indicating that an ACP domain can service both MAT domains and both KS domains. The most plausible explanation for these observations is that the upper portion of the X can rotate 180° with respect to the lower portion. This hypothesis is supported by the observation that the quite tenuous DH/KS contact that joins the top and bottom portions of the X lacks perfect 2-fold symmetry in Fig. 25-35a, thus indicating that this joint is flexible.

b. Fungal FAS-I Has a Barrel-Like Shape

The reactions catalyzed by $\alpha_6\beta_6$ fungal FAS-I differ from those mediated by animal FAS-I in several respects:

1. The bifunctional MAT activity of animal FAS-I transfers both acetyl and malonyl groups from CoA to ACP (Reactions 1a and 1b of Fig. 25-32). However, fungal FAS-I employs a monofunctional **acetyl transferase (AT)** activity to transfer the incoming acetyl group from CoA to ACP and a **malonyl/palmitoyl transferase (MPT)** activity to do so for malonyl groups.

2. Animal FAS-I synthesizes an ACP-linked palmitoyl group, which its TE activity hydrolytically releases as palmitate (Reaction 6 of Fig. 25-32). In contrast, fungal FAS-I synthesizes ACP-linked palmitoyl (C_{16}) and stearoyl (C_{18}) groups in a ratio of roughly 2:3, which its bifunctional MPT activity then transfers to CoA to yield the pathway's final products, palmitoyl-CoA and stearoyl-CoA.

3. The ER activity in animal FAS-I uses NADPH to directly reduce the $C=C$ double bond in Reaction 5 of Fig. 25-32. However, in fungal ER, NADPH reduces FMNH₂, which in turn, reduces the double bond.

4. Fungal FAS-I has a **phosphopantetheine transferase (PPT)** activity that attaches the phosphopantetheine group to ACP. This activity is not part of animal FAS-I.

The distribution of the various enzymatic activities along the α and β chains of fungal FAS-I (Fig. 25-36) bears no resemblance to that of animal FAS-I (Fig. 25-34). Note that the MPT activity is shared by both the α and β chains.

The X-ray structure of FAS-I from the thermophilic fungus *Thermomyces lanuginosus* (Fig. 25-37), determined by Ban, reveals that it forms a hollow barrel-shaped protein with D_3 symmetry (the rotational symmetry of a trigonal prism; Fig. 8-65b). The six α chains form a D_3 -symmetric wheel that is capped on each side by a 3-fold symmetric dome that is predominantly formed by the β chains. The central wheel splits the hollow interior of the barrel into two identical reaction chambers, each of which has several openings in its side walls through which small molecules can enter. The active sites of all the enzymatic activities face the interiors of the reaction chambers. Clearly, animal and fungal FAS-I's have divergently evolved from FAS-II.

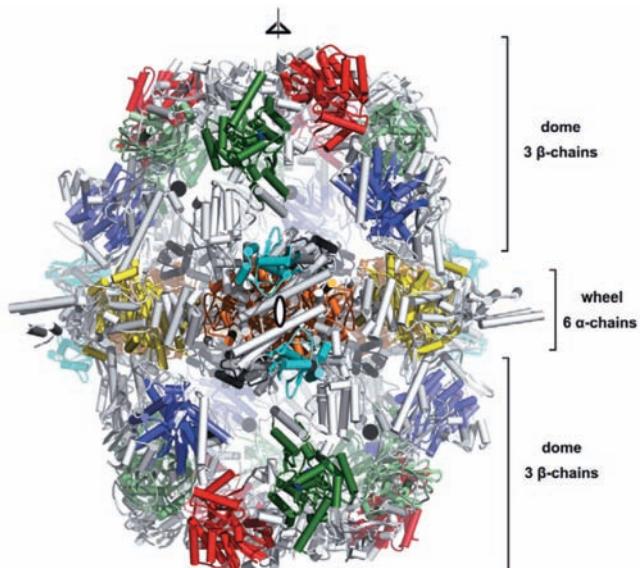


Figure 25-37 X-ray structure of *T. lanuginosus* FAS-I. The ~ 270 -Å-high by ~ 250 -Å-diameter $\alpha_6\beta_6$ heterododecamer is viewed along a 2-fold axis (ellipsoid) and perpendicular to its 3-fold axis (triangle) with its various domains colored as in Fig. 25-36 and its linker domains gray. The attachment sites of the disordered ACP domains are indicated by black dots. [Courtesy of Simon Jenni and Nenad Ban, ETH Zurich, Switzerland. PDBIDs 2UV9 and 2UVA.]

The six ACP domains, which are N-terminally anchored to the chamber walls and C-terminally anchored to the middle of the central wheel, are disordered in the X-ray structure of *T. lanuginosus* FAS-I. However, they are visible in the otherwise closely similar X-ray structure of yeast FAS-I, which was independently determined by Ban and by Thomas Steitz. Structural considerations suggest that each doubly tethered ACP domain can swing to visit the required six catalytic centers, which in the case of the ACP tethered to subchamber 1 are KR from $\alpha 1$, KS from $\alpha 2$, MPT and DH from $\beta 1$, and AT and ER from $\beta 2$, where subchamber 2 is on the clockwise side of subchamber 1 as viewed from the top of the dome.

The PPT domains are located on the outside of the barrel where they cannot interact with the ACP domains. This suggests that they attach the phosphopantetheine groups to the ACPs before the barrel has fully assembled.

c. Fatty Acid Synthase Inhibitors Are Drug Candidates

In well-nourished individuals, fatty acid synthesis proceeds at a low rate. However, certain tissues, particularly cancers, express high levels of FAS-I and produce fatty acids at a high rate. Consequently, inhibitors of animal FAS-I are being investigated as possible anticancer agents. Moreover, the differences between the enzymatic activities of the various types of FAS's, particularly their ER activities, makes FAS-II and fungal FAS-I targets for the development of novel antibiotics.

d. Variations on a Theme: Polyketide Biosynthesis

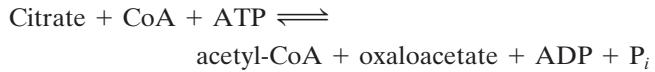
Polyketides are a family of $>10,000$ diverse and structurally complex natural products, many of which have antibacterial, antifungal, antitumor, and immunosuppressive properties, that are synthesized by bacteria, fungi, plants, and certain marine animals. They are made by the modular condensation of acyl-CoA monomers such as acetyl-CoA and propionyl-CoA with malonyl-CoA and methylmalonyl-CoA extender units whose decarboxylation drives the condensation reaction. The name polyketide comes from the fact that the primary condensation products have β -keto functional groups. Palmitate is an example of a polyketide since it is formed by the condensation of one acetyl-CoA primer and seven malonyl-CoA extender units. Following each condensation reaction, the new β -keto group may be reduced, dehydrated, and reduced again as with fatty acids, or may undergo only partial modification.

Polyketides are synthesized by megasynthases. We have already seen that animal FAS-I contains seven enzymatic activities as well as ACP. Another example of a polyketide is **6-deoxyerythronolide B (6dEB)**, the parent **macrolactone** of the antibiotic **erythromycin A** (Section 32-3G), which is synthesized in the soil bacterium *Saccharopolyspora erythraea* from one propionyl-CoA primer and six (S)-methylmalonyl-CoA extenders by **deoxyerythronolide B synthase (DEBS; Fig. 25-38)**. DEBS is a 2000-kD, $\alpha_2\beta_2\gamma_2$ complex of >3000 -residue subunits

whose three homodimeric units each catalyze two elongation/modification cycles. Unlike FAS-I, which catalyzes several cycles of elongation/modification with the same active sites, DEBS catalyzes each elongation/modification cycle on a different module, which permits the differences in the modifications that occur at each cycle. Thus, DEBS, which has 28 different active sites, functions much like an assembly line. Module 4, as Fig. 25-38 indicates, is almost identical in function to FAS-I, containing KS, AT, ACP, KR, DH, and ER, and reducing its primary β -ketone condensation product to a methylene group. However, it does not contain TE because the elongation process is not complete after this phase. Module 3 contains only ACP, KS, and AT, the minimal set of activities for a module, and passes its β -ketone condensation product to module 4 without further modification. Modules 1, 2, 5, and 6 contain only ACP, AT, KS, and KR, the sites necessary for the condensation and ketone reduction steps, thereby generating hydroxy products. The overall organization of the modules therefore creates a polyhydroxy product containing one keto group and one methylene group in the chain. The DEBS final product, 6dEB, is a lactone produced by the reaction of the terminal hydroxyl group with the thioester anchoring the growing chain to the synthase. The various polyketide synthases have different organizations of modules, and consequently synthesize a multitude of different compounds.

D. Transport of Mitochondrial Acetyl-CoA Into the Cytosol

Acetyl-CoA is generated in the mitochondrion by the oxidative decarboxylation of pyruvate as catalyzed by pyruvate dehydrogenase (Section 21-2A) as well as by the oxidation of fatty acids. When the need for ATP synthesis is low, so that the oxidation of acetyl-CoA via the citric acid cycle and oxidative phosphorylation is minimal, this mitochondrial acetyl-CoA may be stored for future use as fat. Fatty acid biosynthesis occurs in the cytosol but the mitochondrial membrane is essentially impermeable to acetyl-CoA. *Acetyl-CoA enters the cytosol in the form of citrate via the tricarboxylate transport system* (Fig. 25-39). Cytosolic **ATP-citrate lyase** then catalyzes the reaction



which resembles the reverse of the citrate synthase reaction (Section 21-3A) except that ATP hydrolysis is required to drive the intermediate synthesis of the “high-energy” citryl-CoA, whose hydrolysis drives the citrate synthase reaction to completion. ATP hydrolysis is therefore required in the ATP-citrate lyase reaction to power the resynthesis of this thioester bond. Oxaloacetate is reduced to malate by malate dehydrogenase. Malate may be oxidatively decarboxylated to pyruvate by malic enzyme (Section 25-2Ed) and be returned in this form to the mitochondrion. The malic enzyme reaction resembles that

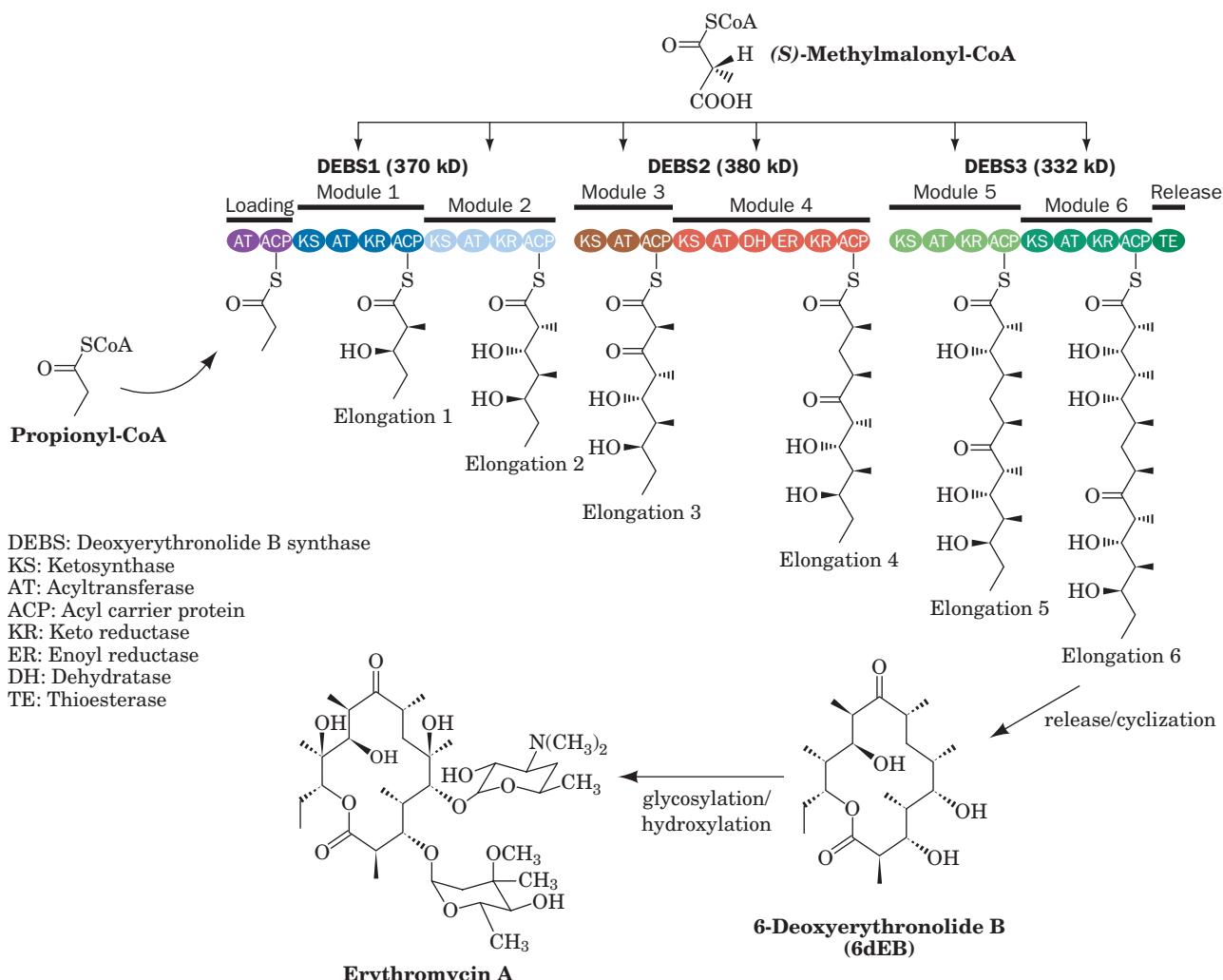


Figure 25-38 An example of polyketide biosynthesis: the synthesis of erythromycin A. [After Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C., *Science* **291**, 1790 (2001).]

of isocitrate dehydrogenase in which a β -hydroxy acid is oxidized to a β -keto acid, whose decarboxylation is strongly favored (Section 21-3C). Malic enzyme's coenzyme is NADP^+ , so when this route is used NADPH is produced for use in the reductive reactions of fatty acid biosynthesis.

Citrate transport out of the mitochondrion must be balanced by anion transport into the mitochondrion. Malate, pyruvate, and P_i can act in this capacity. Malate may therefore also be transported directly back to the mitochondrion without generating NADPH. As we have seen in Section 25-4C, synthesis of each palmitate ion requires 8 molecules of acetyl-CoA and 14 molecules of NADPH. As many as 8 of these NADPH molecules may be supplied with the 8 molecules of acetyl-CoA if all the malate produced in the cytosol is oxidatively decarboxylated. The re-

maining NADPH is provided through the pentose phosphate pathway (Section 23-4).

E. Elongases and Desaturases

Palmitate (16:0), the normal product of the animal fatty acid synthase pathway, is the precursor of longer chain saturated and unsaturated fatty acids through the actions of **elongases** and **desaturases**. Elongases are present in both the mitochondrion and the ER but the mechanisms of elongation at the two sites differ. Mitochondrial elongation (a process independent of the fatty acid synthase pathway) occurs by successive addition and reduction of acetyl units in a reversal of fatty acid oxidation; the only chemical difference between these two pathways occurs in the final reduction step in which NADPH takes the place of FADH_2 as the terminal

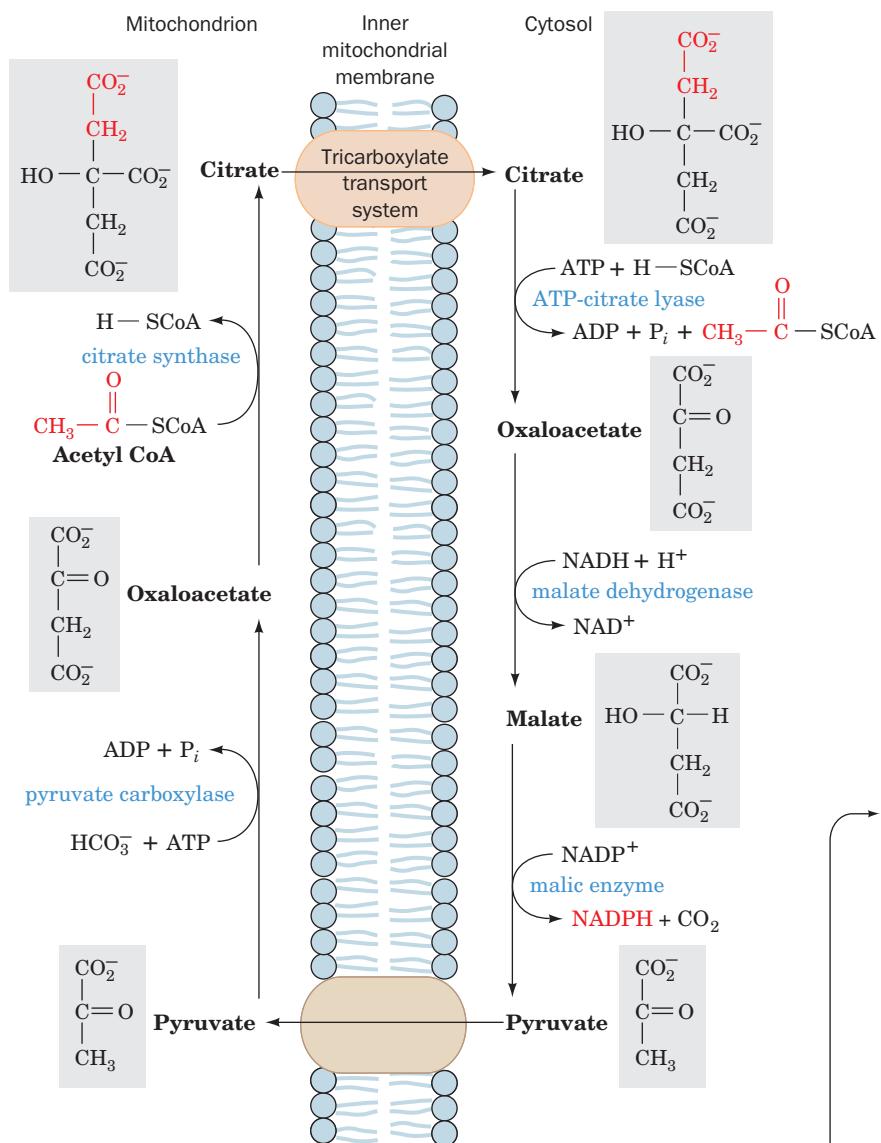


Figure 25-39 Transfer of acetyl-CoA from mitochondrion to cytosol via the tricarboxylate transport system.

redox coenzyme (Fig. 25-40). Elongation in the ER involves the successive condensations of malonyl-CoA with acyl-CoA. These reactions are each followed by NADPH-associated reductions similar to those catalyzed by FAS-I, the only difference being that the fatty acid is elongated as its CoA derivative rather than as its ACP derivative.

Unsaturated fatty acids are produced by **terminal desaturases**. Mammalian systems contain four terminal desaturases of broad chain-length specificities designated Δ^9 -, Δ^6 -, Δ^5 -, and Δ^4 -**fatty acyl-CoA desaturases**. These

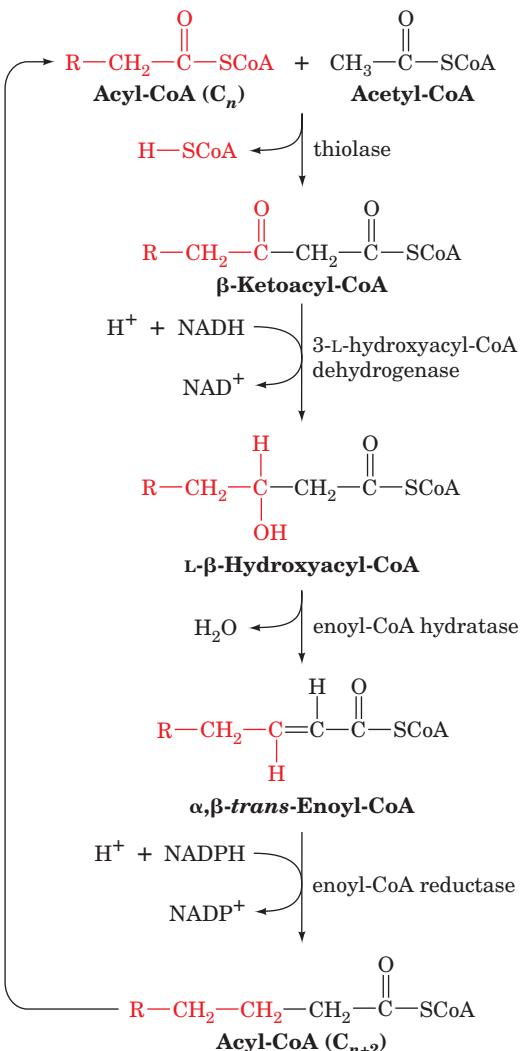
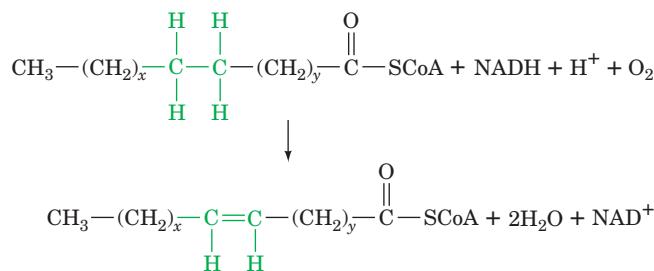


Figure 25-40 Mitochondrial fatty acid elongation. Elongation occurs by the reversal of fatty acid oxidation with the exception that the final reaction employs NADPH rather than FADH₂ as its redox coenzyme.

membrane-bound, nonheme iron-containing enzymes catalyze the general reaction



where x is at least 5 and where (CH₂)_x can contain one or more double bonds. The (CH₂)_y portion of the substrate is always saturated. Double bonds are inserted between existing double bonds in the (CH₂)_x portion of the substrate and the CoA group such that the new double bond is three carbon atoms closer to the CoA group than the next double bond (not conjugated to an existing double bond) and, in animals, never at positions beyond C9. Mammalian terminal desaturases are components of mini-electron-transport systems that contain two other proteins: **cytochrome b_5** and **NADH-cytochrome b_5 reductase**. The electron-transfer reactions mediated by these complexes occur at the inner surface of the ER membrane (Fig. 25-41) and are therefore not associated with oxidative phosphorylation.

a. Some Unsaturated Fatty Acids Must Be Obtained in the Diet

A variety of unsaturated fatty acids may be synthesized by combinations of elongation and desaturation reactions. However, since palmitic acid is the shortest available fatty acid in animals, the above rules preclude the formation of the Δ^{12} double bond of linoleic acid [$\Delta^{9,12}$ -octadecadienoic acid; 18:2n-6 (this nomenclature is explained in Table 12-1)], a required precursor of **prostaglandins**. *Linoleic acid must consequently be obtained in the diet (ultimately from plants that have Δ^{12} - and Δ^{15} -desaturases; it is abundant in most vegetable oils) and is therefore termed an essential fatty acid.* Indeed, animals maintained on a fat-free diet develop an ultimately fatal condition that is initially characterized by poor growth, poor wound healing, and dermatitis. Linoleic acid is also an important constituent of epidermal sphingolipids that function as the skin's water-permeability barrier.

Because of the inability of animal desaturases to add double bonds to positions beyond C9, another essential fatty acid is **α -linolenic acid [ALA; $\Delta^{9,12,15}$ -octadecatrienoic acid (18:3n-3, an ω -3 fatty acid)]. This fatty acid is a precursor to **EPA ($\Delta^{5,8,11,14,17}$ -eicosapentaenoic acid; 20:5n-3)** and **DHA ($\Delta^{4,7,10,13,16,19}$ -docosahexaenoic acid; 22:6n-3)**, polyunsaturated ω -3 fatty acids recently found to be important dietary constituents (present in fish oils) that improve cognitive function and vision, and contribute to protection against inflammation and cardiovascular disease. DHA is, among other things, the predominant fatty acid in the phospholipids of retinal rod outer segments. Substitution of DHA with otherwise identical ω -6 fatty acids in phospholipids results in impaired visual acuity. Deficiency of ω -3 polyunsaturated fatty acids in brain phospholipids is associated with memory loss and diminished cognitive function.**

F. Synthesis of Triacylglycerols

Triacylglycerols are synthesized from fatty acyl-CoA esters and glycerol-3-phosphate or dihydroxyacetone phosphate (Fig. 25-42). The initial step in this process is catalyzed either by **glycerol-3-phosphate acyltransferase** in mitochondria and the ER, or by **dihydroxyacetone phosphate acyltransferase** in the ER or peroxisomes. In the latter case, the product **acyl-dihydroxyacetone phosphate** is reduced to the corresponding **lysophosphatidic acid** by an NADPH-dependent reductase. The lysophosphatidic acid is converted to a triacylglycerol by the successive actions of **1-acylglycerol-3-phosphate acyltransferase**, **phosphatidic acid phosphatase**, and **diacylglycerol acyltransferase**. The intermediate phosphatidic acid and 1,2-diacylglycerol (DAG) can also be converted to phospholipids by the pathways described in Section 25-8. The acyltransferases are not completely specific for particular fatty acyl-CoAs, either in chain length or in degree of unsaturation, but in triacylglycerols of human adipose tissue, palmitate tends to be concentrated at position 1 and oleate at position 2.

a. Glyceroneogenesis Is Important for Triacylglycerol Biosynthesis

The dihydroxyacetone phosphate used to make glycerol-3-phosphate for triacylglycerol synthesis comes either from glucose via the glycolytic pathway (Fig. 17-3) or

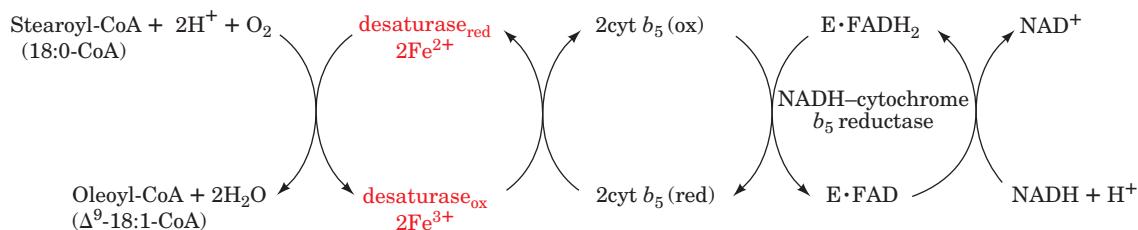


Figure 25-41 The electron-transfer reactions mediated by the Δ^9 -fatty acyl-CoA desaturase complex. Its three proteins, desaturase, cytochrome b_5 , and NADH-cytochrome b_5 reductase, are

situated in the endoplasmic reticulum membrane. [After Jeffcoat, R., *Essays Biochem.* **15**, 19 (1979).]

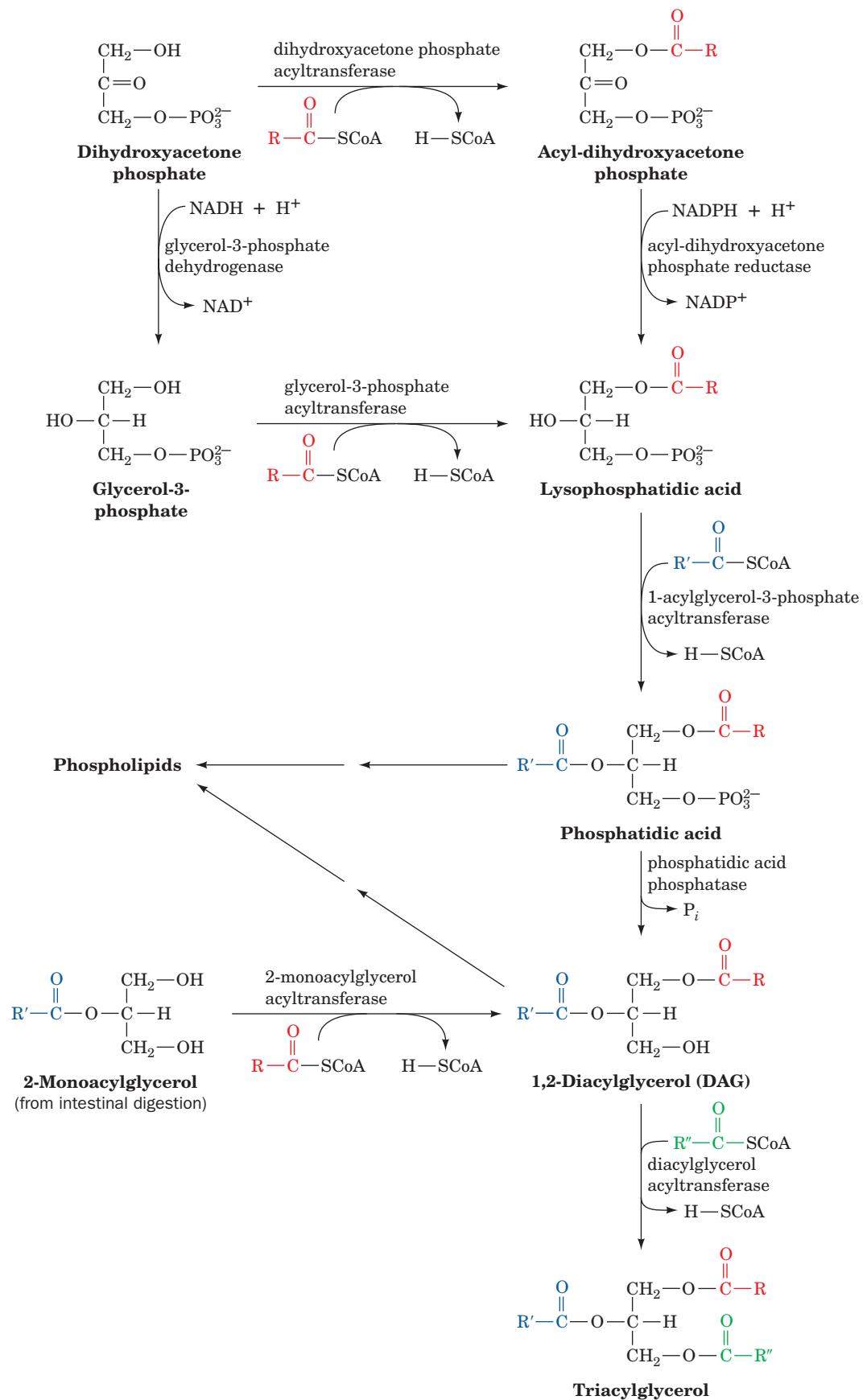


Figure 25-42 The reactions of triacylglycerol biosynthesis.

from oxaloacetate via an abbreviated version of gluconeogenesis (Fig. 23-8) termed **glyceroneogenesis**. Glyceroneogenesis is necessary in times of starvation, since approximately 30% of the fatty acids that enter the liver during a fast are reesterified to triacylglycerol and exported as VLDL (Section 25-1 and 25-6A). Adipocytes also carry out glyceroneogenesis in times of starvation. They do not carry out gluconeogenesis but contain the gluconeogenic enzyme phosphoenolpyruvate carboxykinase (PEPCK), which is upregulated when glucose concentration is low, and participates in the glyceroneogenesis required for triacylglycerol biosynthesis.

5 REGULATION OF FATTY ACID METABOLISM

Discussions of metabolic control are usually concerned with the regulation of metabolite flow through a pathway in response to the differing energy needs and dietary states of an organism. For example, the difference in the energy requirement of muscle between rest and vigorous exertion may be as much as 100-fold. Such varying demands may be placed on the body when it is in either a fed or a fasted state. For instance, Eric Newsholme, an authority on the biochemistry of exercise, enjoys a 2-hour run before breakfast. Others might wish for no greater exertion than the motion of hand to mouth. In both individuals, glycogen and triacylglycerols serve as primary fuels for energy-requiring processes and are synthesized in times of quiet plenty for future use.

a. Hormones Regulate Fatty Acid Metabolism

Synthesis and breakdown of glycogen and triacylglycerols, as detailed in Chapter 18 and above, are processes that concern the whole organism, with its organs and tissues forming an interdependent network connected by the bloodstream. The blood carries the metabolites responsible for energy production: triacylglycerols in the form of chylomicrons and VLDL (Section 12-5A), fatty acids as their albumin complexes (Section 25-1e), ketone bodies, amino acids, lactate, and glucose. The pancreatic α and β cells sense the organism's dietary and energetic state mainly through the glucose concentration in the blood. The α cells respond to the low blood glucose concentration of the fasting and energy-demanding states by secreting glucagon. The β cells respond to the high blood glucose concentration of the fed and resting states by secreting insulin. We have previously discussed (Sections 18-3E and 18-3F) how these hormones are involved in glycogen metabolism. *They also regulate the rates of the opposing pathways of lipid metabolism and therefore control whether fatty acids will be oxidized or synthesized.* Their targets are the regulatory (flux-generating) enzymes of fatty acid synthesis and breakdown in specific tissues (Fig. 25-43).

We are already familiar with most of the mechanisms by which the catalytic activities of regulatory enzymes may be

controlled: substrate availability, allosteric interactions, and covalent modification (phosphorylation). These are examples of **short-term regulation**, regulation that occurs with a response time of minutes or less. *Fatty acid synthesis is controlled, in part, by short-term regulation.* Acetyl-CoA carboxylase, which catalyzes the first committed step of this pathway, is inhibited by palmitoyl-CoA and by the glucagon-stimulated cAMP-dependent increase in phosphorylation, and is activated by citrate and by insulin-stimulated dephosphorylation (Section 25-4B).

Another mechanism exists for controlling a pathway's regulatory enzymes: alteration of the amount of enzyme present by changes in the rates of protein synthesis and/or breakdown. This process requires hours or days and is therefore called **long-term regulation** (the control of protein synthesis and breakdown is discussed in Chapters 31 and 32). *Lipid biosynthesis is also controlled by long-term regulation*, with insulin stimulating and starvation inhibiting the synthesis of acetyl-CoA carboxylase and fatty acid synthase. The presence in the diet of polyunsaturated fatty acids also decreases the concentrations of these enzymes. The amount of adipose tissue lipoprotein lipase, the enzyme that initiates the entry of lipoprotein-packaged fatty acids into adipose tissue for storage (Section 12-5Ba), is also increased by insulin and decreased by starvation. In contrast, the concentration of heart lipoprotein lipase, which controls the entry of fatty acids from lipoproteins into heart tissue for oxidation rather than storage, is decreased by insulin and increased by starvation. *Starvation and/or regular exercise, by decreasing the glucose concentration in the blood, change the body's hormone balance. This situation results in long-term changes in gene expression that increase the levels of fatty acid oxidation enzymes and decrease those of lipid biosynthesis.*

Fatty acid oxidation is regulated largely by the concentration of fatty acids in the blood, which is, in turn, controlled by the hydrolysis rate of triacylglycerols in adipose tissue by hormone-sensitive triacylglycerol lipase. This enzyme is so named because it is susceptible to regulation by phosphorylation and dephosphorylation in response to hormonally controlled cAMP levels. Epinephrine and norepinephrine, as does glucagon, act to increase adipose tissue cAMP concentrations. cAMP allosterically activates protein kinase A (PKA) which, in turn, increases the phosphorylation levels of susceptible enzymes. Phosphorylation activates hormone-sensitive triacylglycerol lipase, thereby stimulating lipolysis in adipose tissue, raising blood fatty acid levels, and ultimately activating the β -oxidation pathway in other tissues such as liver and muscle. In liver, this process leads to the production of ketone bodies that are secreted into the bloodstream for use by peripheral tissues as an alternative fuel to glucose. PKA, acting in concert with AMP-dependent protein kinase (AMPK), also causes the inactivation of acetyl-CoA carboxylase (Section 25-4B), one of the rate-determining enzymes of fatty acid synthesis, so that *cAMP-dependent phosphorylation simultaneously stimulates fatty acid oxidation and inhibits fatty acid synthesis.*

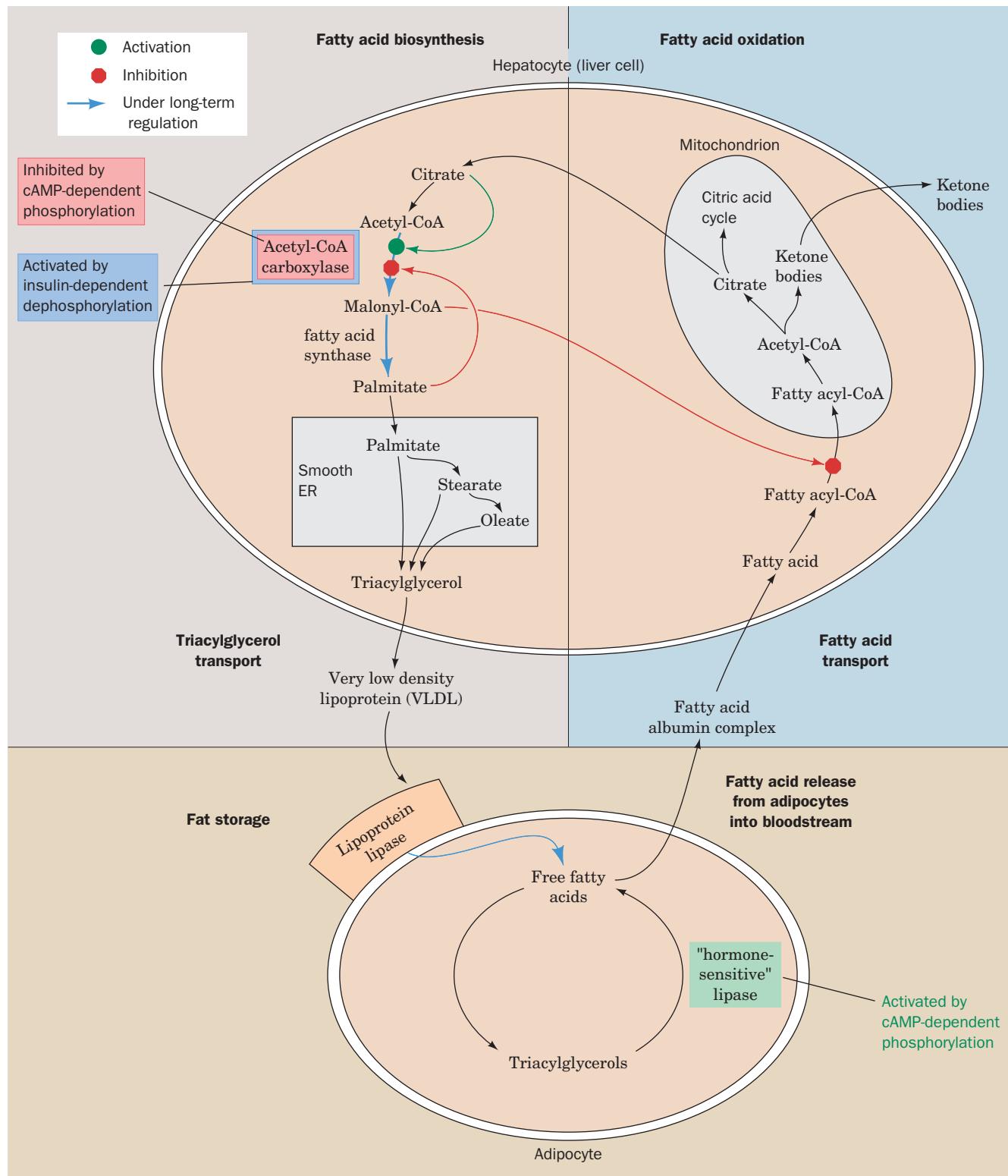


Figure 25-43 Sites of regulation of fatty acid metabolism.

Insulin has the opposite effect of glucagon and epinephrine: It stimulates the formation of glycogen and triacylglycerols. This protein hormone, which is secreted in response to high blood glucose concentrations, triggers a highly complex signal transduction network (Section 19-4F)

that induces the long-term regulation of numerous enzymes as well as decreasing cAMP levels. This latter situation leads to the dephosphorylation and thus the inactivation of hormone-sensitive triacylglycerol lipase, thereby reducing the amount of fatty acid available for oxidation. Insulin

also stimulates the dephosphorylation of acetyl-CoA carboxylase, thereby activating this enzyme (Section 25-4Ba). *The glucagon–insulin ratio is therefore of prime importance in determining the rate and direction of fatty acid metabolism.*

Another control point that inhibits fatty acid oxidation when fatty acid synthesis is stimulated is the inhibition of carnitine palmitoyltransferase I by malonyl-CoA. This inhibition keeps the newly synthesized fatty acids out of the mitochondrion (Section 25-2B) and thus away from the β -oxidation system. As we have seen (Section 25-4Bc), heart muscle, an oxidative tissue that does not carry out fatty acid biosynthesis, contains an isoform of acetyl-CoA carboxylase, ACC2, whose sole function appears to be the synthesis of malonyl-CoA to regulate fatty acid oxidation.

AMPK may itself be an important regulator of fatty acid metabolism. This phosphorylating enzyme is activated by AMP and inhibited by ATP and thus has been proposed to serve as a fuel gauge for the cell. When ATP levels are high, signaling the fed and rested state, this kinase is inhibited, allowing ACC to become dephosphorylated (activated) so as to stimulate malonyl-CoA production for fatty acid synthesis in adipose tissue and for inhibition of fatty acid oxidation in muscle cells. When activity levels increase causing ATP levels to decrease with a concomitant increase in AMP levels, AMPK is activated to phosphorylate (inac-

tivate) ACC. The resulting decrease in malonyl-CoA levels causes fatty acid biosynthesis to decrease in adipose tissue while fatty acid oxidation increases in muscle to provide the ATP for continued activity.

6 CHOLESTEROL METABOLISM

Cholesterol is a vital constituent of cell membranes and the precursor of steroid hormones and bile salts. It is clearly essential to life, yet its deposition in arteries has been associated with cardiovascular disease and stroke, two leading causes of death in humans. In a healthy organism, an intricate balance is maintained between the biosynthesis, utilization, and transport of cholesterol, keeping its harmful deposition to a minimum. In this section, we study the pathways of cholesterol biosynthesis and transport and how they are controlled. We also examine how cholesterol is utilized in the biosynthesis of steroid hormones and bile salts.

A. Cholesterol Biosynthesis

All of the carbon atoms of cholesterol are derived from acetate (Fig. 25-44). Observation of their pattern of incorporation into cholesterol led Konrad Bloch to propose that acetate was first converted to **isoprene units**, C_5 units that have the carbon skeleton of **isoprene**:

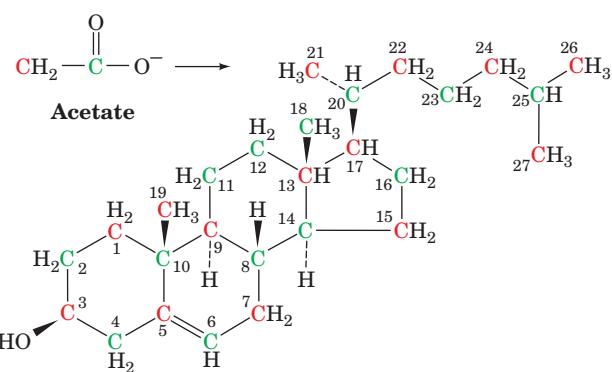
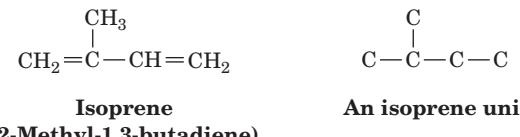


Figure 25-44 All of cholesterol's carbon atoms are derived from acetate.



Isoprene units are condensed to form a linear precursor to cholesterol, and then cyclized.

Squalene, a polyisoprenoid hydrocarbon (Fig. 25-45a), was demonstrated to be the linear intermediate in cholesterol biosynthesis by the observation that feeding isotopically labeled squalene to animals yields labeled cholesterol. Squalene may be folded in several ways that would enable it to cyclize to the four-ring sterol nucleus (Section 12-1E). The folding pattern proposed by Bloch and Robert B. Woodward (Fig. 25-45b) proved to be correct.

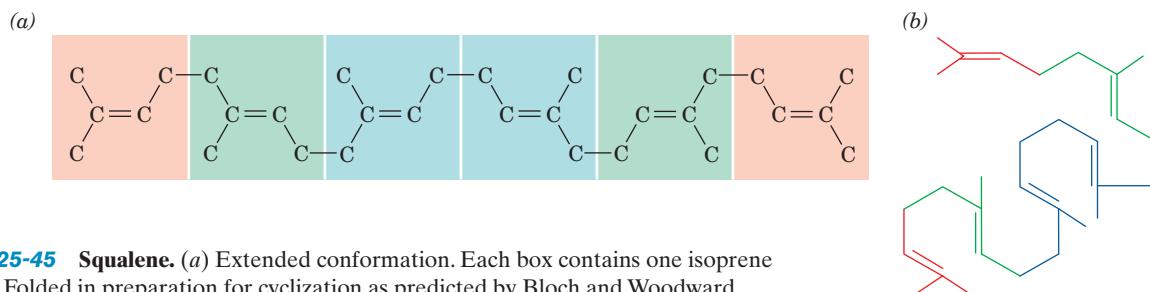


Figure 25-45 Squalene. (a) Extended conformation. Each box contains one isoprene unit. (b) Folded in preparation for cyclization as predicted by Bloch and Woodward.

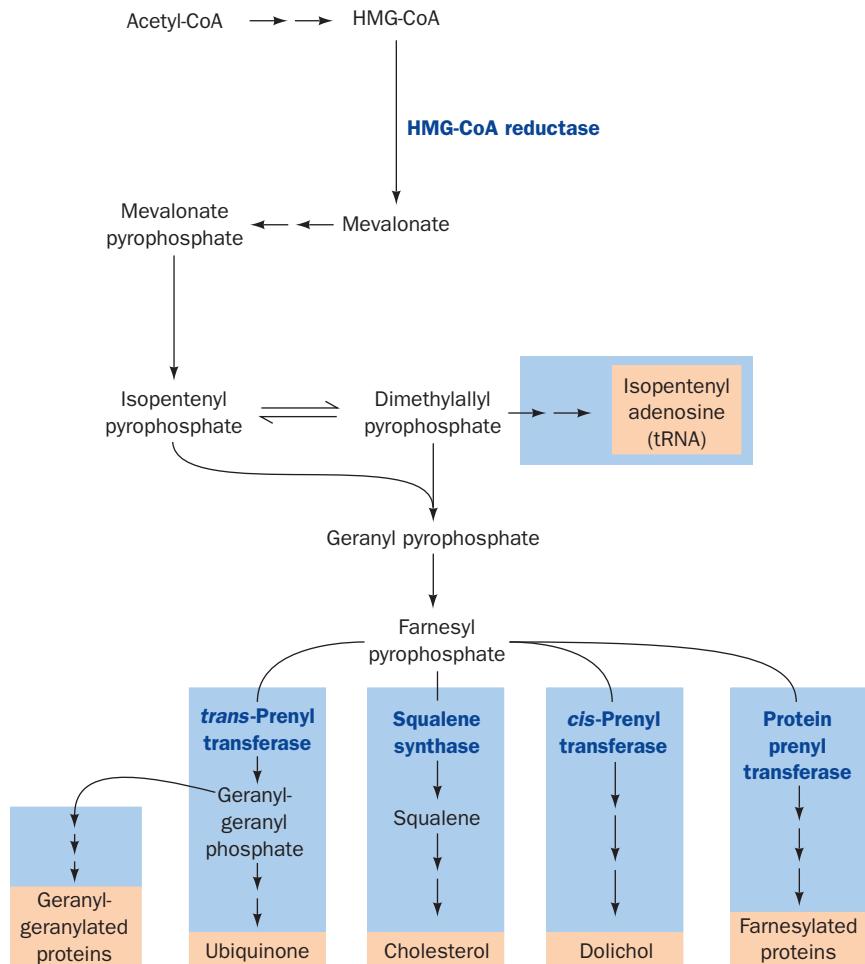


Figure 25-46 The branched pathway of isoprenoid metabolism in mammalian cells. The pathway produces ubiquinone, dolichol, farnesylated and geranylgeranylated

proteins, and isopentenyl adenosine, a modified tRNA base, in addition to cholesterol.

Bloch's outline for the major stages of cholesterol biosynthesis was

Acetate \longrightarrow isoprenoid intermediate \longrightarrow
squalene \longrightarrow cyclization product \longrightarrow cholesterol

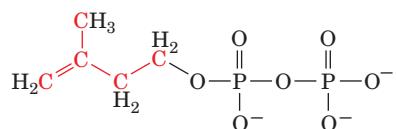
This pathway has been experimentally verified and its details elaborated. It is now known to be part of a branched pathway (Fig. 25-46) that produces several other essential isoprenoids in addition to cholesterol, namely, ubiquinone (CoQ; Fig. 22-17b), dolichol (Fig. 23-15), farnesylated and geranylgeranylated proteins (Fig. 12-29), and **isopentenyl-adenosine** (a modified base of tRNA; Fig. 32-10). We shall examine in detail the portion of this pathway that synthesizes cholesterol. Note, however, that $>25,000$ isoprenoids (also known as **terpenoids**), mostly of plant, fungal, and bacterial origin, have been characterized. These serve as membrane constituents (e.g., cholesterol), hormones

(steroids), pheromones, defensive agents, photoprotective agents (e.g., β -carotene; Section 24-2Ad), and visual pigments (e.g., retinal; Section 12-3Ab), to name only a few of their many biological functions.

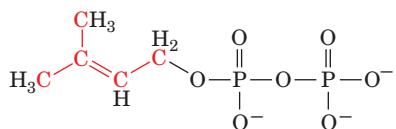
a. HMG-CoA Is a Key Cholesterol Precursor

Acetyl-CoA is converted to isoprene units by a series of reactions that begins with formation of hydroxymethylglutaryl-CoA (HMG-CoA; Fig. 25-26), a compound we previously encountered as an intermediate in ketone body biosynthesis (Section 25-3). HMG-CoA synthesis requires the participation of two enzymes: thiolase and HMG-CoA synthase. The enzymes forming the HMG-CoA leading to ketone bodies occur in the mitochondria, whereas those responsible for the synthesis of the HMG-CoA that is destined for cholesterol biosynthesis are located in the cytosol. Their catalytic mechanisms, however, are identical.

HMG-CoA is the precursor of two isoprenoid intermediates, **isopentenyl pyrophosphate** and **dimethylallyl pyrophosphate**:



Isopentenyl pyrophosphate



Dimethylallyl pyrophosphate

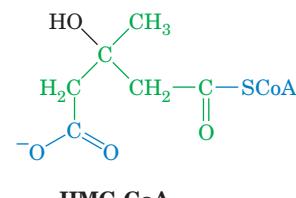
The formation of isopentenyl pyrophosphate involves four reactions (Fig. 25-47):

1. The CoA thioester group of HMG-CoA is reduced to an alcohol in an NADPH-dependent four-electron reduction catalyzed by **HMG-CoA reductase**, yielding **mevalonate**.
2. The new OH group is phosphorylated by **mevalonate-5-phosphotransferase**.
3. The phosphate group is converted to a pyrophosphate by **phosphomevalonate kinase**.
4. The molecule is decarboxylated and the resulting alcohol dehydrated by **pyrophosphomevalonate decarboxylase**.

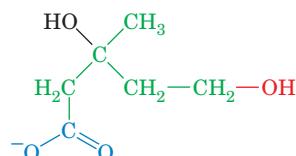
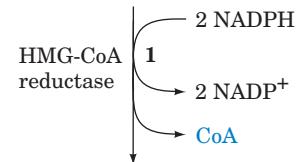
HMG-CoA reductase mediates the rate-determining step of cholesterol biosynthesis and is the most elaborately regulated enzyme of this pathway. This 888-residue ER membrane-bound enzyme is regulated, as we shall see in Section 25-6Bb, by competitive and allosteric mechanisms, phosphorylation/dephosphorylation, and long-term regulation. Cholesterol itself is an important feedback regulator of the enzyme.

b. Pyrophosphomevalonate Decarboxylase Catalyzes an Apparently Concerted Reaction

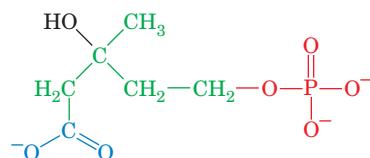
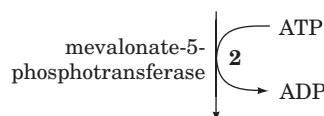
5-Pyrophosphomevalonate is converted to isopentenyl pyrophosphate by an ATP-dependent dehydration-decarboxylation reaction catalyzed by **pyrophosphomevalonate decarboxylase** (Fig. 25-48). When $[3-^{18}\text{O}]5$ -pyrophosphomevalonate (^3O in Fig. 25-48) is used as a substrate, the labeled oxygen appears in P_i . This observation suggests that 3-phospho-5-pyrophosphomevalonate is a reaction intermediate. Since all attempts to isolate this intermediate have failed, however, it has been proposed that phosphorylation, the α,β elimination of CO_2 , and the elimination of P_i occur in a concerted reaction.



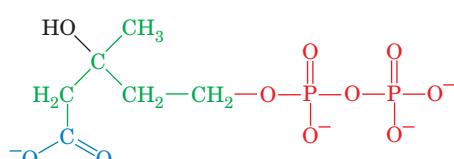
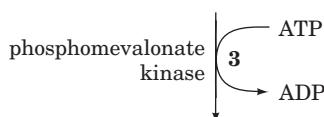
HMG-CoA



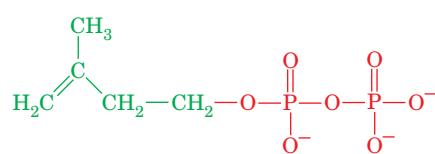
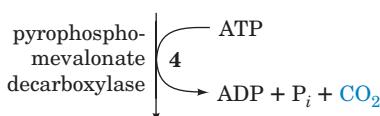
Mevalonate



Phosphomevalonate



5-Pyrophosphomevalonate



Isopentenyl pyrophosphate

Figure 25-47 Formation of isopentenyl pyrophosphate from HMG-CoA.

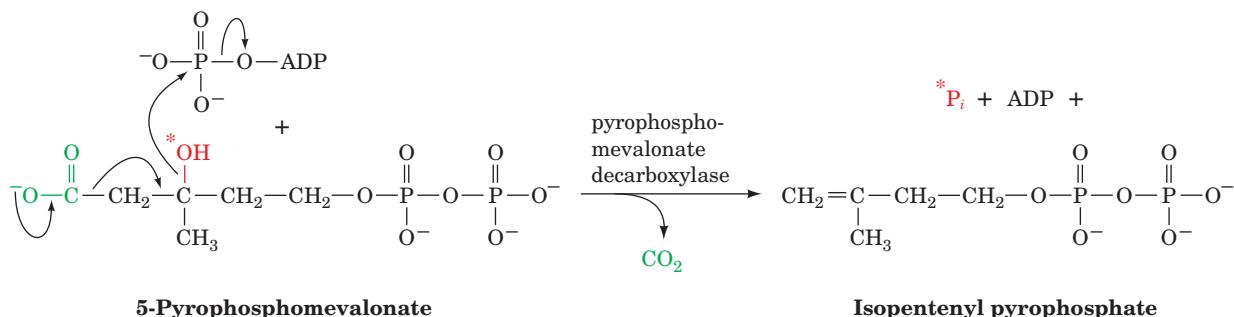


Figure 25-48 Action of pyrophosphomevalonate decarboxylase. The enzyme catalyzes an ATP-dependent concerted dehydration-decarboxylation of pyrophosphomevalonate, yielding isopentenyl pyrophosphate.

The equilibration between isopentenyl pyrophosphate and dimethylallyl pyrophosphate is catalyzed by **isopentenyl pyrophosphate isomerase**. The reaction appears to occur via a protonation/deprotonation reaction involving the intermediacy of a tertiary carbocation intermediate. Cys and Glu residues have been implicated as the general acid and base catalysts, respectively (Fig. 25-49), as supported by site-directed mutagenesis and the X-ray structure of the enzyme. The carbocation is thought to be stabilized through interactions with the aromatic π cloud of an adjacent Trp residue. Aromatic residues provide electron-rich interactions with positively charged groups without forming covalent bonds that would destroy the intermediate.

c. Squalene Is Formed by the Condensation of Six Isoprene Units

Four isopentenyl pyrophosphates and two dimethylallyl pyrophosphates condense to form the C_{30} cholesterol precursor squalene in three reactions catalyzed by two enzymes (Fig. 25-50):

1. Prenyltransferase (farnesyl pyrophosphate synthase) catalyzes the head-to-tail (1'-4) condensation of dimethylallyl pyrophosphate and isopentenyl pyrophosphate to yield **geranyl pyrophosphate**.

2. Prenyltransferase catalyzes a second head-to-tail condensation of geranyl pyrophosphate and isopentenyl pyrophosphate to yield **farnesyl pyrophosphate (FPP)**.

3. Squalene synthase (SQS) then catalyzes the head-to-head (1-1') condensation of two farnesyl pyrophosphate molecules to form squalene. Farnesyl pyrophosphate is also a precursor to dolichol, farnesylated and geranylgeranylated proteins, and ubiquinone (Fig. 25-46).

Prenyltransferase catalyzes the condensation of isopentenyl pyrophosphate with an allylic (conjugated to a C=C double bond) pyrophosphate. It is specific for isopentenyl pyrophosphate but can use either the 5-carbon dimethylallyl pyrophosphate or the 10-carbon **geranyl pyrophosphate** as its allylic substrate. The prenyltransferase-catalyzed condensation mechanism is particularly interesting since it is one of the few known enzyme-catalyzed reactions that pro-

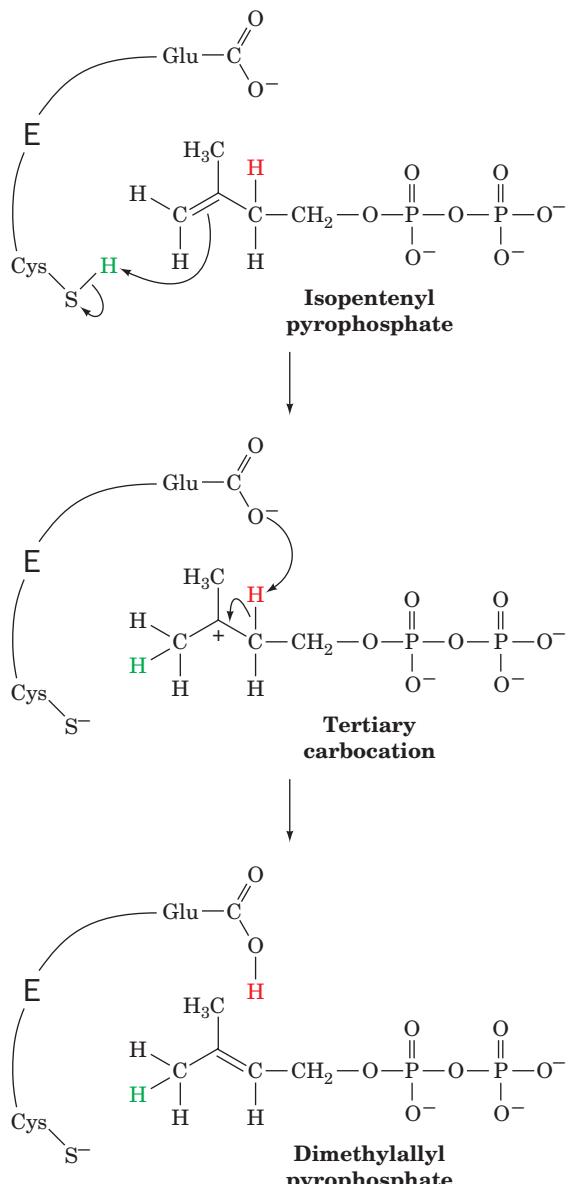
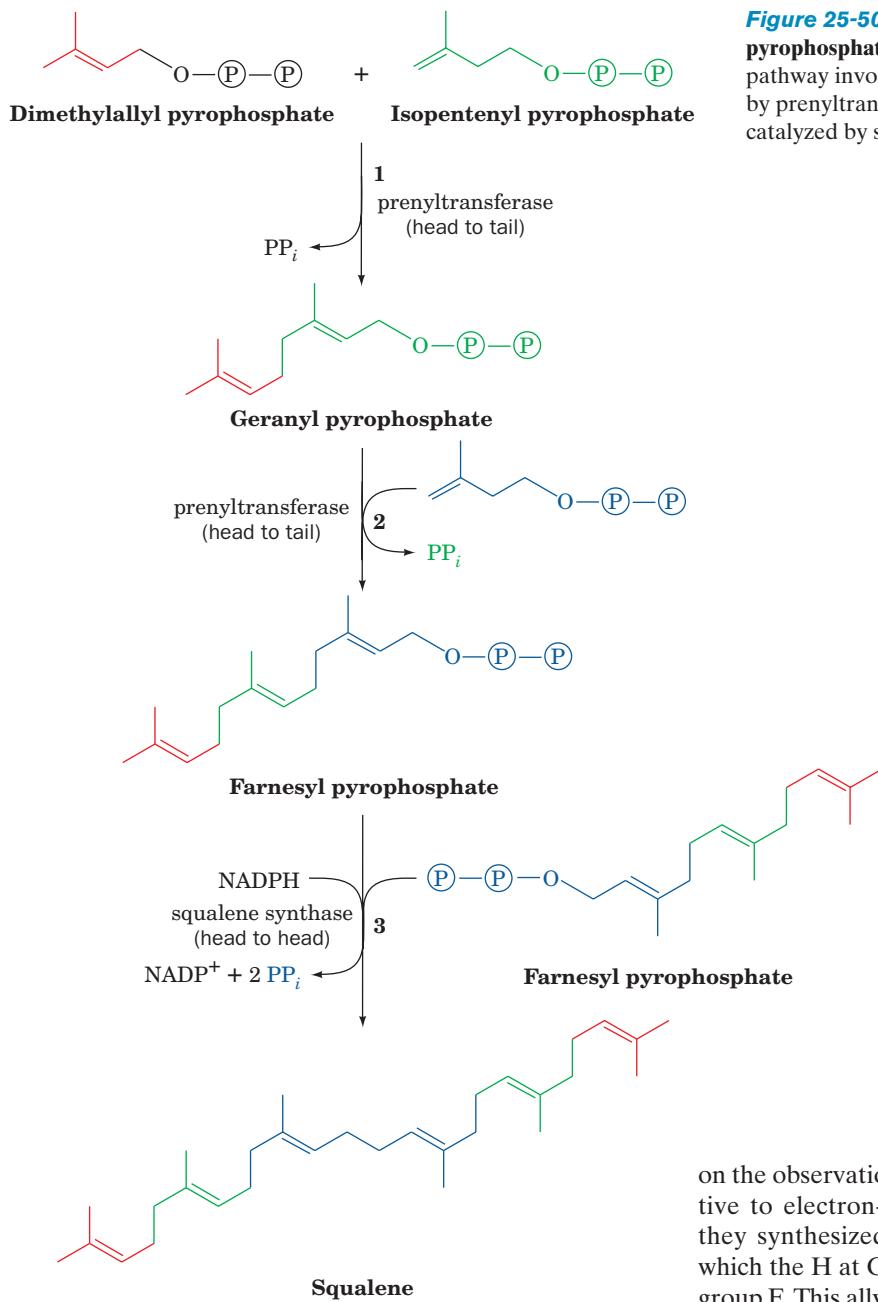


Figure 25-49 Mechanism of isopentenyl pyrophosphate isomerase. The enzyme interconverts isopentenyl pyrophosphate and dimethylallyl pyrophosphate by a protonation/deprotonation reaction involving a carbocation intermediate in which a Cys and a Glu residue act as a proton donor and acceptor. The carbocation intermediate appears to be stabilized by π interactions with a nearby Trp side chain.



ceed via a carbocation intermediate. Two possible condensation mechanisms can be envisioned (Fig. 25-51):

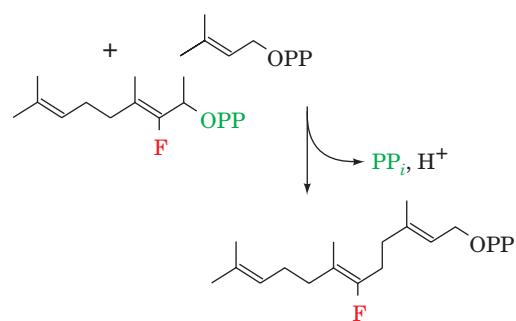
Scheme I An S_N1 mechanism in which an allylic carbocation forms by the elimination of PP_i . Isopentenyl pyrophosphate then condenses with this carbocation, forming a new carbocation that eliminates a proton to form product.

Scheme II An S_N2 reaction in which the allylic PP_i is displaced in a concerted manner. In this case, an enzyme nucleophile, X, assists in the reaction. This group is eliminated in the second step with the loss of a proton to form product.

Dale Poulter and Hans Rilling used chemical logic to differentiate between these two mechanisms. Capitalizing

Figure 25-50 Formation of squalene from isopentenyl pyrophosphate and dimethylallyl pyrophosphate. The pathway involves two head-to-tail condensations catalyzed by prenyltransferase and a head-to-head condensation catalyzed by squalene synthase.

on the observation that S_N1 reactions are much more sensitive to electron-withdrawing groups than S_N2 reactions, they synthesized a geranyl pyrophosphate derivative in which the H at C2 is replaced by the electron-withdrawing group F. This allylic substrate for the second (1'-4) condensation catalyzed by prenyltransferase, not surprisingly, has the same K_M as the natural substrate (F and H have similar atomic radii):



It is, however, the V_{max} of this reaction that tells the story. If the reaction is an S_N2 displacement, the fluoro derivative

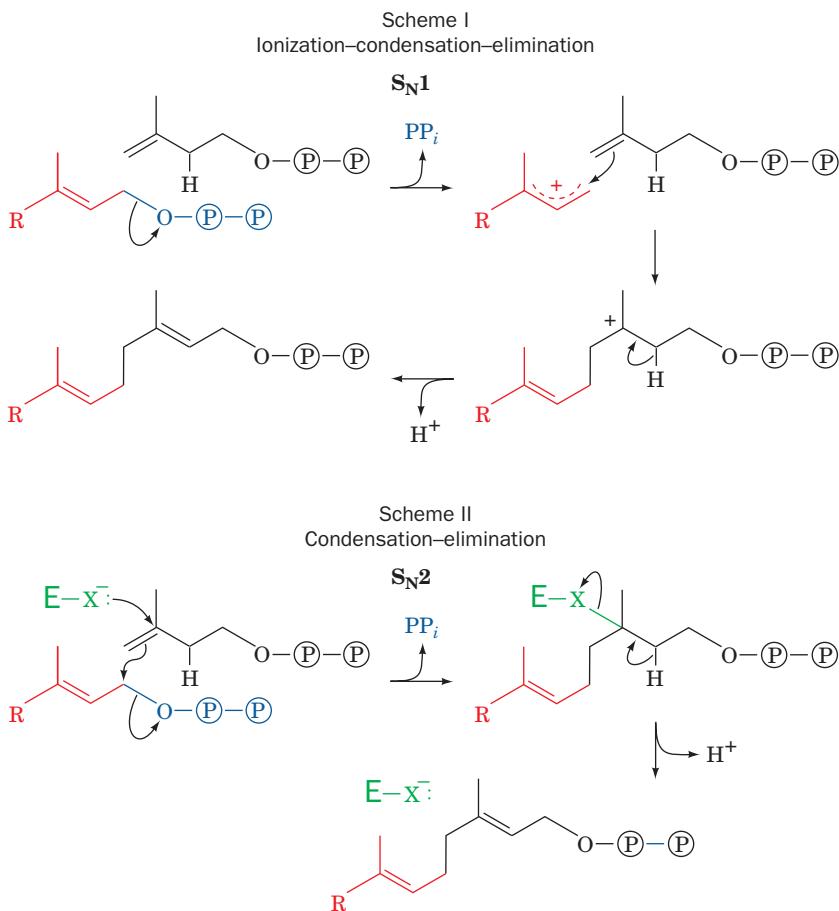


Figure 25-51 Two possible mechanisms for the prenyltransferase reaction. Scheme I involves the formation of a carbocation intermediate, whereas Scheme II involves the participation of an enzyme nucleophile, X.

should react at a rate similar to that of the natural substrate. If, instead, the reaction has an S_N1 mechanism, the fluoro derivative should react orders of magnitude more slowly than the natural substrate. In fact, 3-fluorogeranyl pyrophosphate forms product at <1% of the rate of the natural substrate, strongly supporting an S_N1 mechanism with a carbocation intermediate.

Carbocations are now known to participate in several reactions of isoprenoid biosynthesis. The enzymes are classified according to how they generate these carbocations. Class I enzymes do so via the release of pyrophosphate, as we have seen for prenyltransferase. Class II enzymes do so by protonating a double bond, as does isopentenyl pyrophosphate isomerase (Fig. 25-49), or an epoxide, as we shall see below for oidosqualene cyclase.

Squalene, the immediate sterol precursor, is formed by the head-to-head condensation of two FPP molecules by SQS. Although the enzyme is a Class I enzyme that is structurally related to prenyltransferase and generates carbocations by the release of pyrophosphate, the reaction is not a simple head-to-tail condensation, as might be expected, but, rather, proceeds via a complex two-step mechanism with each step catalyzed by a different active site on the enzyme (Fig. 25-52):

Step I The reaction of two FPP molecules to yield the stable intermediate **presqualene pyrophosphate**. This reac-

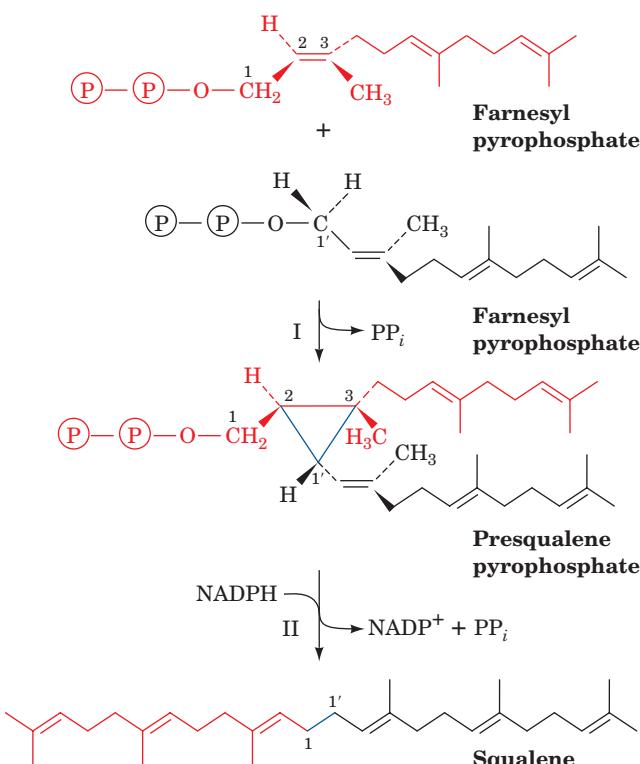


Figure 25-52 Action of squalene synthase. The enzyme catalyzes the head-to-head condensation of two farnesyl pyrophosphate molecules to form squalene.

tion is initiated by the elimination of PP_i from one farnesyl pyrophosphate molecule to form an allylic carbocation at C1 that is stabilized by a π interaction with an essential Tyr residue (Fig. 25-53). The highly reactive electron-deficient carbocation inserts into the electron-rich $\text{C}2=\text{C}3$ double bond of the second molecule, yielding presqualene pyrophosphate, a cyclopropylcarbinyl pyrophosphate.

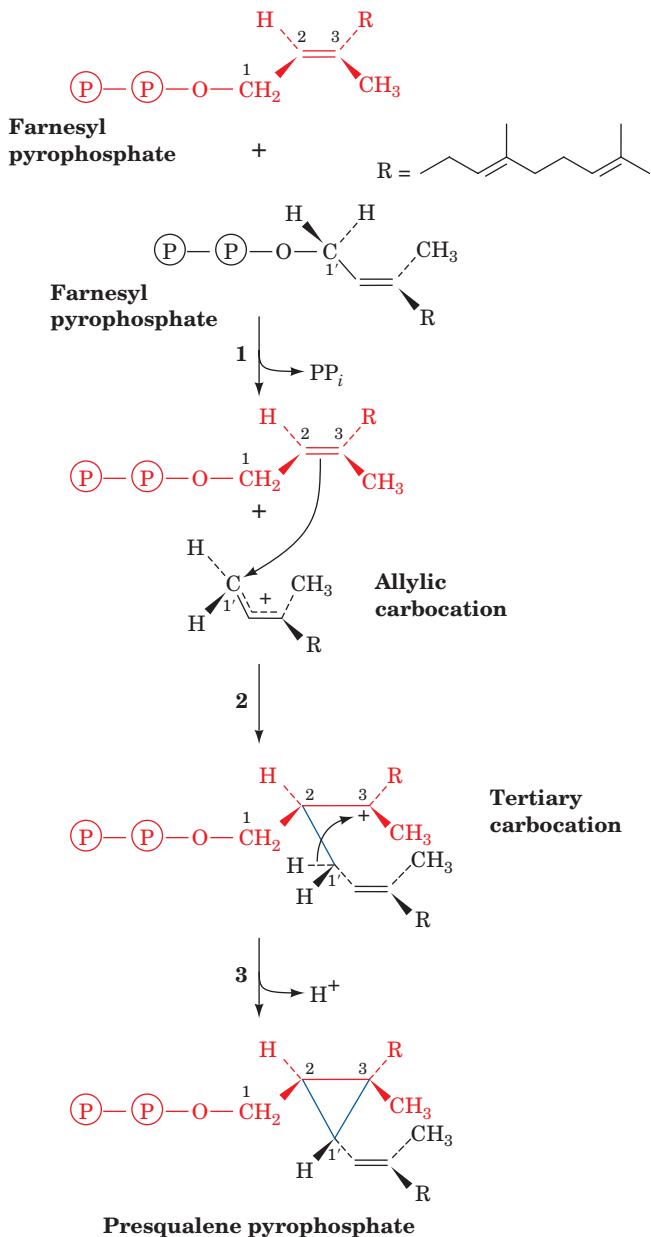


Figure 25-53 Proposed mechanism for the formation of presqualene pyrophosphate from two farnesyl pyrophosphate molecules by squalene synthase (Fig. 25-52, Step I). (1) The pyrophosphate group on one farnesyl pyrophosphate leaves, yielding an allylic carbocation. This reaction step is facilitated by proton donation from the side chain of an essential Tyr residue, which then stabilizes the allylic cation via π -cation interactions. (2) The $\text{C}2=\text{C}3$ double bond of the second farnesyl pyrophosphate nucleophilically attacks the allylic carbocation to form a tertiary carbocation at C3. (3) The abstraction of the pro-*S* proton at C1' by the phenolate group of the essential Tyr residue results in the formation of a C1'-C3 bond yielding presqualene pyrophosphate.

residue (Fig. 25-53). The highly reactive electron-deficient carbocation inserts into the electron-rich $\text{C}2=\text{C}3$ double bond of the second molecule, yielding presqualene pyrophosphate, a cyclopropylcarbinyl pyrophosphate.

Step II The rearrangement and reduction of presqualene pyrophosphate by NADPH to form squalene. This reaction involves the formation and rearrangement of a cyclopropylcarbinyl cation in a complex reaction sequence called a **1'-2-3 process** (Fig. 25-54).

SQS, a monomeric protein, is anchored to the ER membrane via a short C-terminal transmembrane domain, with its active site facing the cytosol. This allows it to accept its water-soluble substrates, farnesyl pyrophosphate and NADPH, from the cytosol and release its hydrophobic product, squalene, in the ER membrane.

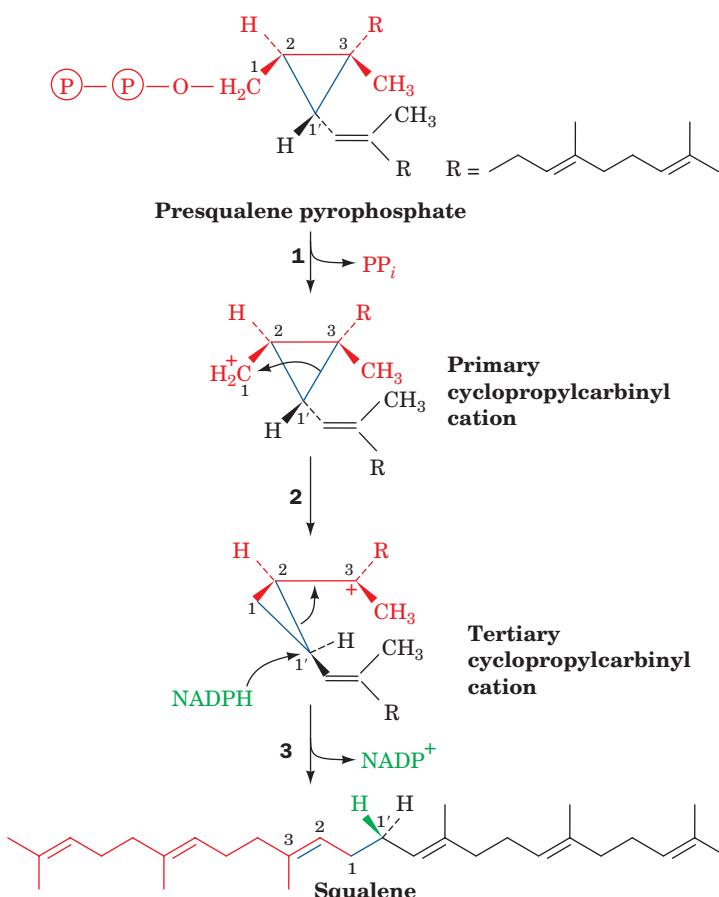
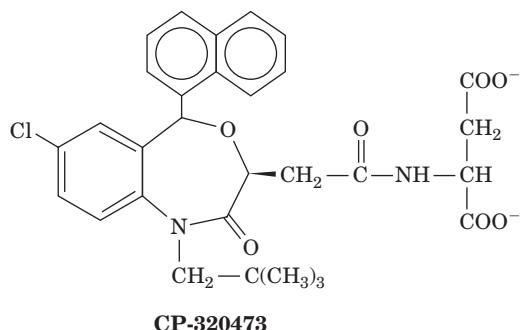


Figure 25-54 Mechanism of rearrangement and reduction of presqualene pyrophosphate to squalene as catalyzed by squalene synthase (Fig. 25-52, Step II). (1) Presqualene's pyrophosphate group leaves, yielding a primary carbocation at C1. (2) The electrons forming the C1'-C3 bond migrate to C1, forming squalene's C1-C1' bond and a tertiary carbocation at C3. (3) The process is completed by the addition of an NADPH-supplied hydride ion to C1' and the formation of the C2=C3 double bond.

The X-ray structure of the 417-residue human SQS, in complex with the inhibitor **CP-320473**,



determined by Jayvardhan Pandit, reveals that the protein folds as a single domain with a large channel across one face into which CP-320473 binds (Fig. 25-55). The channel is lined with Asp and Arg residues that mutagenesis studies indicate are involved in FPP binding. Of these, the conserved Asp 80 and Asp 84 are implicated in binding Mg^{2+} ions that ligand an FPP pyrophosphate group. These Asp residues are adjacent to Tyr 171, which forms the base of the channel and which mutagenesis studies have identified as the essential Tyr that is implicated in stabilizing the allylic carbocation intermediate in Step I of the SQS reaction. Step II of the SQS reaction requires that its highly reactive carbocation intermediates be shielded from contact with the aqueous solvent to prevent it from quenching the reaction. This suggests that for Step II, the presqualene pyrophosphate product of Step I moves deeper into the channel into a pocket that is lined with hydrophobic groups, including Phe 288, whose mutation inactivates the enzyme. This further suggests that Phe 288 functions to stabilize one of the cationic intermediates in Step II (Fig. 25-54) through π -cation interactions.

d. Lanosterol Is Produced by Squalene Cyclization

Squalene, an open-chain C_{30} hydrocarbon, is cyclized to form the tetracyclic steroid skeleton in two steps. **Squalene**

epoxidase catalyzes the oxidation of squalene to form **2,3-oxidosqualene** (Fig. 25-56). **Oxidosqualene cyclase** (alternatively, **lanosterol synthase**) converts this epoxide to **lanosterol**, the sterol precursor of cholesterol. The highly exergonic reaction is a complex process involving cyclization of 2,3-oxidosqualene to a **proto-sterol** cation, via a Class II mechanism involving protonation of the epoxide, and rearrangement of this cation to lanosterol by a series of 1,2 hydride and methyl shifts (Fig. 25-57). Note that this reaction generates six of lanosterol's seven chiral centers.

The X-ray structure of human oxidosqualene cyclase in complex with lanosterol, determined by Armin Ruf, reveals that this 732-residue monomeric and monotopic [integral but not transmembrane] protein contains two structurally similar domains named **α/α barrels** (Fig. 25-58a). An α/α barrel consists of two concentric barrels of 6 helices each, with the helices of the inner barrel largely parallel to each other and antiparallel to those of the outer barrel (much like an α/β barrel with the β strands of the inner barrel replaced by helices but with only 6 α/α units rather than 8 α/β units). The enzyme's active site is located inside an elongated central cavity (Fig. 25-58b) to which the lanosterol is bound and which is accessible from the membrane via a nonpolar channel through the enzyme's membrane-immersed part.

The interactions of 2,3-oxidosqualene with oxidosqualene cyclase cause it to fold and react such that it only forms lanosterol. The active site cavity is lined with several conserved aromatic side chains that are suitably positioned to stabilize the catalyzed reaction's intermediate carbocations, while shielding the cationic intermediates from premature quenching by either enzyme nucleophiles or water. The cationic cascade (Fig. 25-57) is initiated by proton donation by Asp 455 to the epoxide ring of the prefolded 2,3-oxidosqualene, and is quenched by His 232 acting as a general base.

The importance of the proper placement of residues for the formation of the correct product was demonstrated by site-directed mutagenesis: The conversion of Thr 384 in

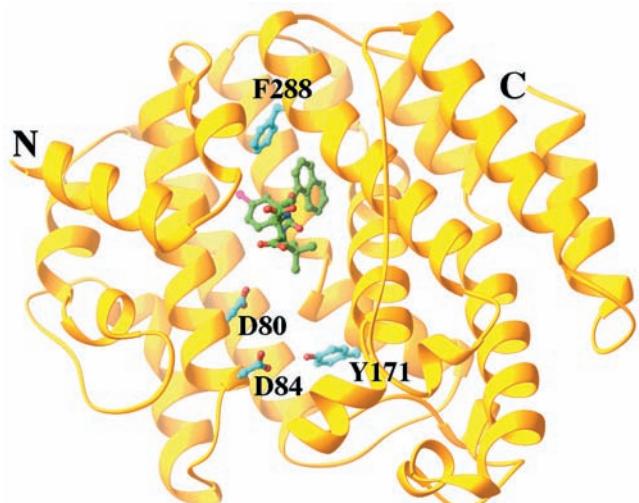


Figure 25-55 X-ray structure of human squalene synthase (SQS) in complex with the inhibitor CP-320473. This inhibitor together with the side chains of D80, D84, Y171, and F288 are drawn in ball-and-stick form colored according to atom type (inhibitor C green, protein C cyan, N blue, O red, and Cl magenta). The protein is viewed looking into its central channel with the putative active sites for Steps I and II of the catalyzed reaction at the bottom and top of the cleft, respectively. The protein's C-terminal transmembrane segment (residues 371–417) together with its N-terminal 30 residues were excised to facilitate its crystallization, which did not affect its *in vitro* catalytic activity. [Based on an X-ray structure by Jayvardhan Pandit, Pfizer Central Research, Groton, Connecticut. PDBid 1EZF.]

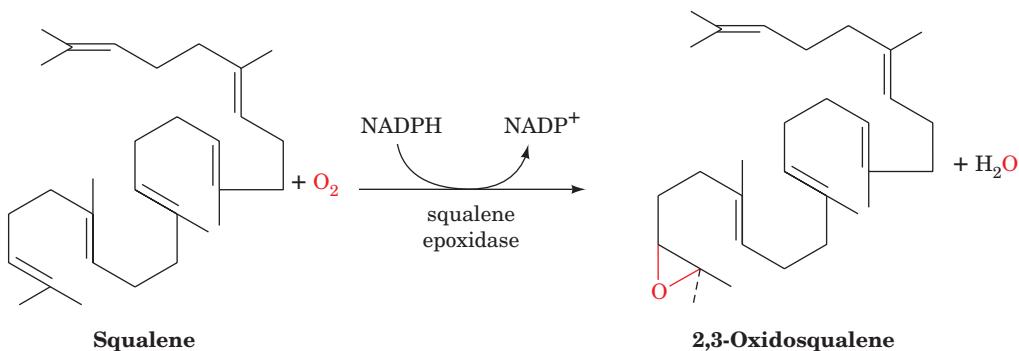
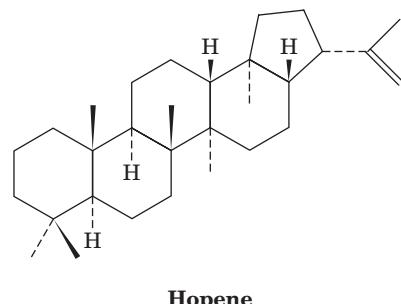
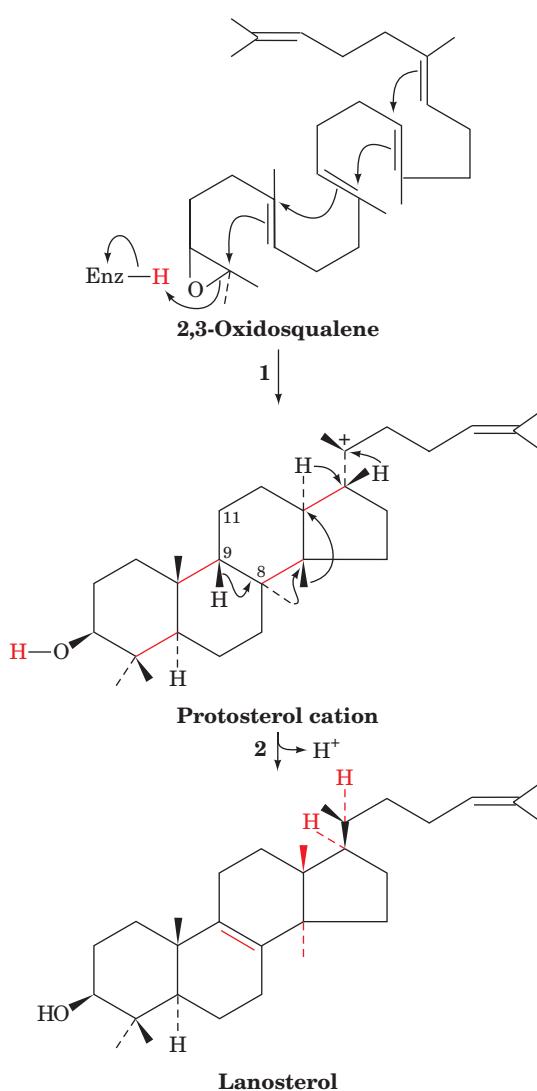


Figure 25-56 The squalene epoxidase reaction.

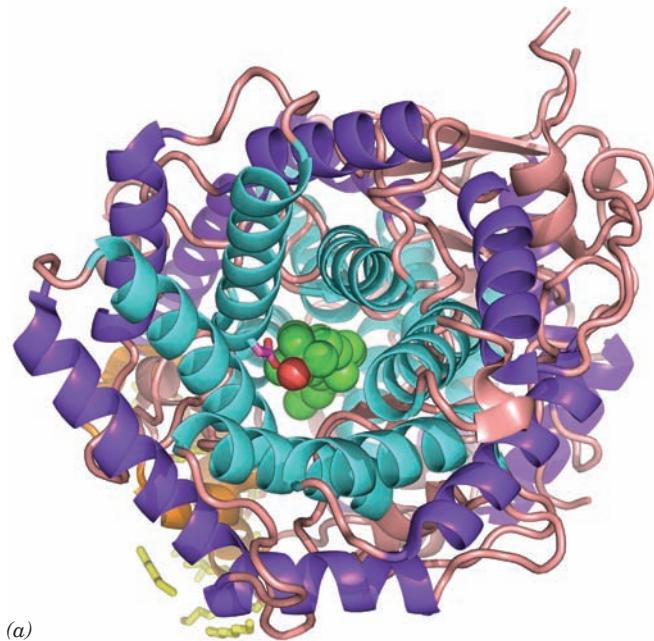
oxidosqualene cyclase to Tyr causes some misplacement of the C8=C9 double bond, with 11% of the product having a C9=C11 double bond and 10% having a 9-hydroxy group; the double mutation, T384Y/V454I, increases the yield of the C9=C11 double bond to 64%. This, together

with different ways of substrate folding, explains how enzymes homologous to oxidosqualene cyclase cause oxidosqualene and squalene to form different products. Indeed, the bacterial enzyme **squalene-hopene synthase**, a homolog of oxidosqualene cyclase, synthesizes the C_{30} compound **hopene**

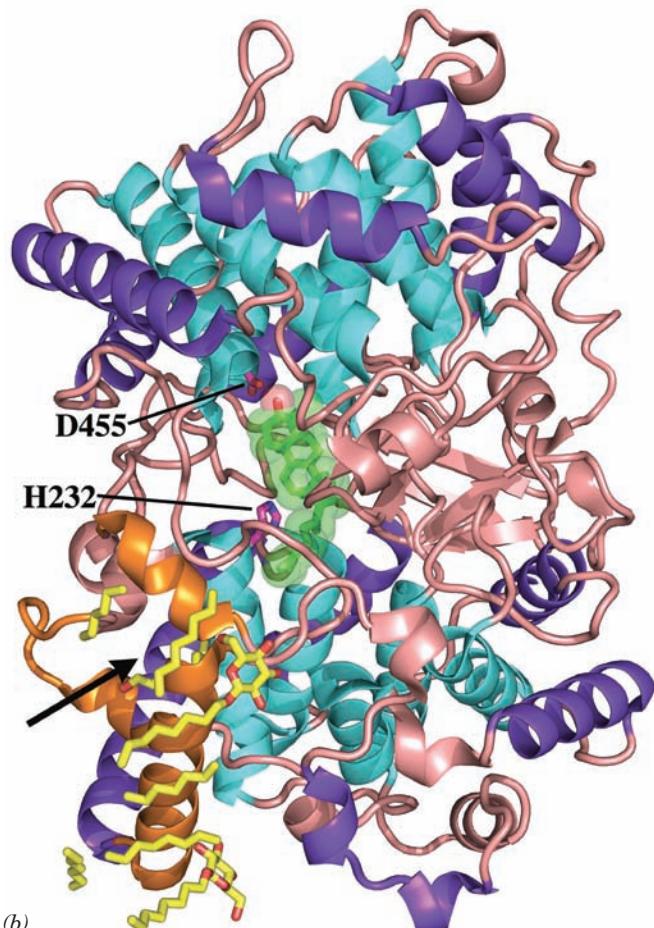


from squalene. Its X-ray structure in complex with 2-azasqualene (squalene with its C2 atom replaced N), determined by Georg Schulz, reveals that the enzyme, which structurally resembles oxidosqualene cyclase, folds squalene in the configuration that is predicted to yield hopene. In this case, the cationic cyclase cascade is initiated by the protonation of a double bond instead of an epoxide.

Figure 25-57 The oxidosqualene cyclase reaction. (1) 2,3-Oxidosqualene is cyclized to the protosterol cation in a process that is initiated by the enzyme-mediated protonation of the squalene epoxide oxygen while this extended molecule is folded in the manner predicted by Bloch and Woodward. The opening of the epoxide leaves an electron-deficient center whose migration drives the series of cyclizations that ultimately form the protosterol cation. (2) A series of methyl and hydride migrations yields a hypothesized intermediate carbocation at C8, which then eliminates a proton from C9 to form the C8=C9 double bond of lanosterol.



(a)



(b)

Figure 25-58 X-ray structure of human oxidosqualene cyclase in complex with its product lanosterol. The monomeric protein is shown in ribbon form (a) viewed along the axis of its α/α barrels, and (b) rotated 90° about the horizontal axis with respect to Part a. The inner helices of its two α/α barrels are cyan, its outer helices are purple, its nonpolar membrane-immersed portion is orange, and the remainder of the protein is pink. The lanosterol, which occupies the enzyme's centrally located active site cavity, is shown in Part a in space-filling form and in Part b in stick form embedded in its semitransparent space-filling form, both with C

green and O red. The catalytically important side chains of His 232 and Asp 455, together with molecules of the detergent β -octylglucoside and fragments of lipids that coat the enzyme's membrane-immersed portion, are drawn in stick form with side chain C magenta, lipid and detergent C yellow, and O red. The arrow in Part b points to the membrane-immersed opening of the enzyme's active site cavity. [Based on an X-ray structure by Armin Ruf, F. Hoffmann-La Roche AG, Basel, Switzerland. PDBid 1W6K.]

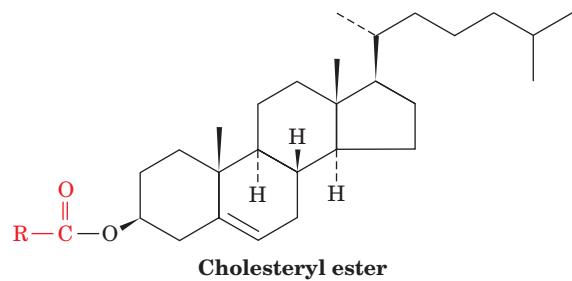
e. Cholesterol Is Synthesized from Lanosterol

Conversion of lanosterol to cholesterol (Fig. 25-59) is a 19-step process that we shall not explore in detail. It involves an oxidation and loss of three methyl groups. The first methyl group is removed as formate and the other two are eliminated as CO_2 in reactions that all require NADPH and O_2 . The enzymes involved in this process are embedded in the ER membrane.

f. Cholesterol Is Transported in the Blood and Taken Up by Cells in Lipoprotein Complexes

Transport and cellular uptake of cholesterol are described in Section 12-5. To recapitulate, cholesterol synthesized by the liver is either converted to bile salts for use in the digestive process (Section 25-1) or esterified by acyl-

CoA:cholesterol acyltransferase (ACAT) to form **cholesteryl esters**



which are secreted into the bloodstream as part of the lipoprotein complexes called **very low density lipoproteins (VLDL)**. As the VLDL circulate, their component triacyl-

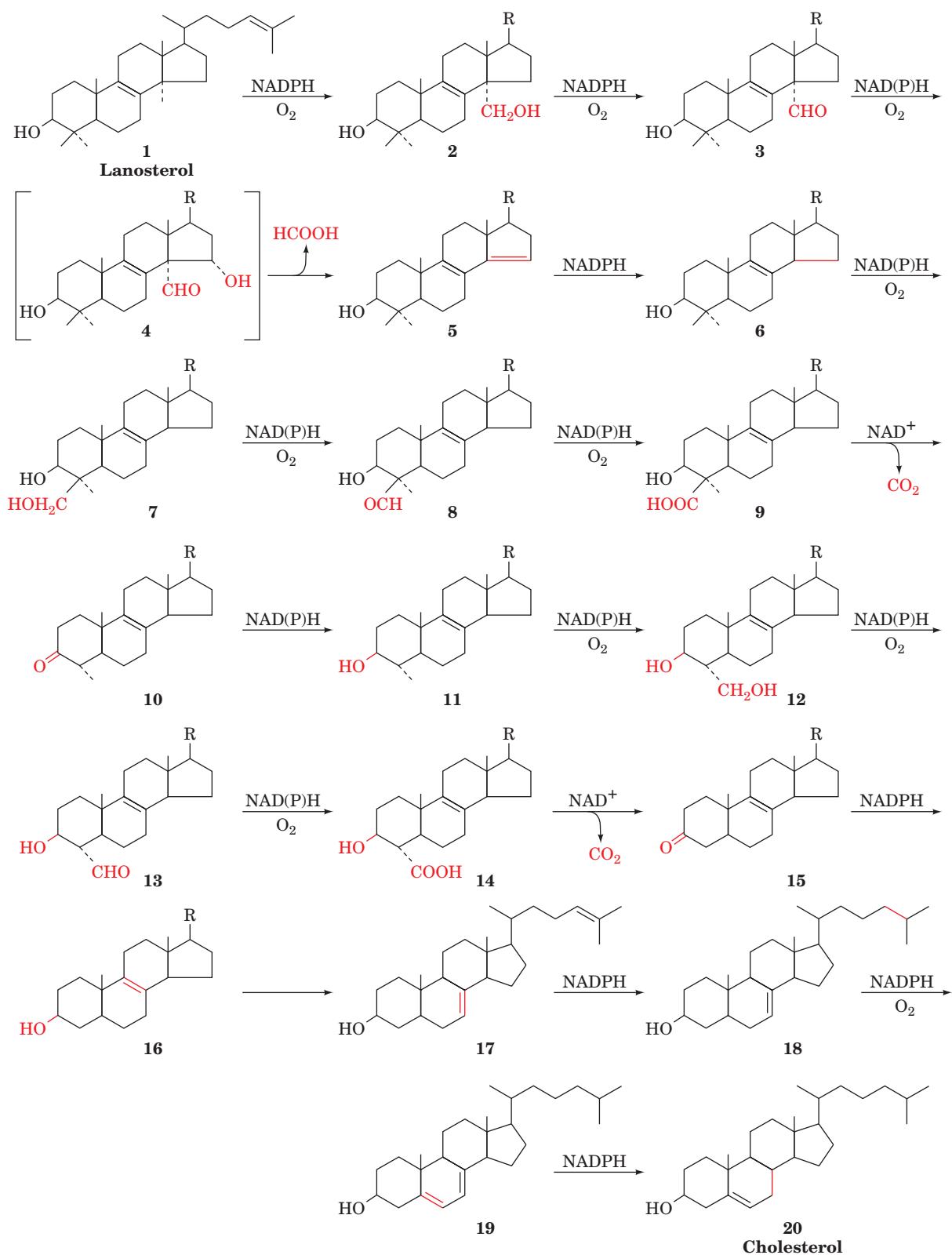


Figure 25-59 The 19-reaction conversion of lanosterol to cholesterol. [After Rilling, H.C. and Chayet, L.T., in Danielsson, H. and Sjövall, J. (Eds.), *Sterols and Bile Acids*, p. 33, Elsevier

(1985), as modified by Bae, S.-H. and Paik, Y.-K., *Biochem. J.* **326**, 609–616 (1997).]

glycerols and most types of their **apolipoproteins** (Table 12-6) are removed in the capillaries of muscle and adipose tissues, sequentially converting the VLDL to **intermediate-density lipoproteins (IDL)** and then to **low-density lipoproteins (LDL)**. Peripheral tissues normally obtain most of their exogenous cholesterol from LDL by receptor-mediated endocytosis (Fig. 25-60; Section 12-5Bc). Inside the cell, cholestryl esters are hydrolyzed by a lysosomal lipase to free cholesterol, which is either incorporated into cell membranes or reesterified by ACAT for storage as cholestryl ester droplets.

Dietary cholesterol, cholestryl esters, and triacylglycerols are transported in the blood by intestinally synthesized lipoprotein complexes called **chylomicrons**. After removal of their triacylglycerols at the peripheral tissues, the resulting **chylomicron remnants** bind to specific liver cell remnant receptors and are taken up by receptor-mediated endocytosis in a manner similar to that of LDL. In the liver, dietary cholesterol is either used in bile salt biosynthesis (Section 25-6C) or packaged into VLDL for export. *Liver and peripheral tissues therefore have two ways of obtaining cholesterol: They may either synthesize it from acetyl-CoA*

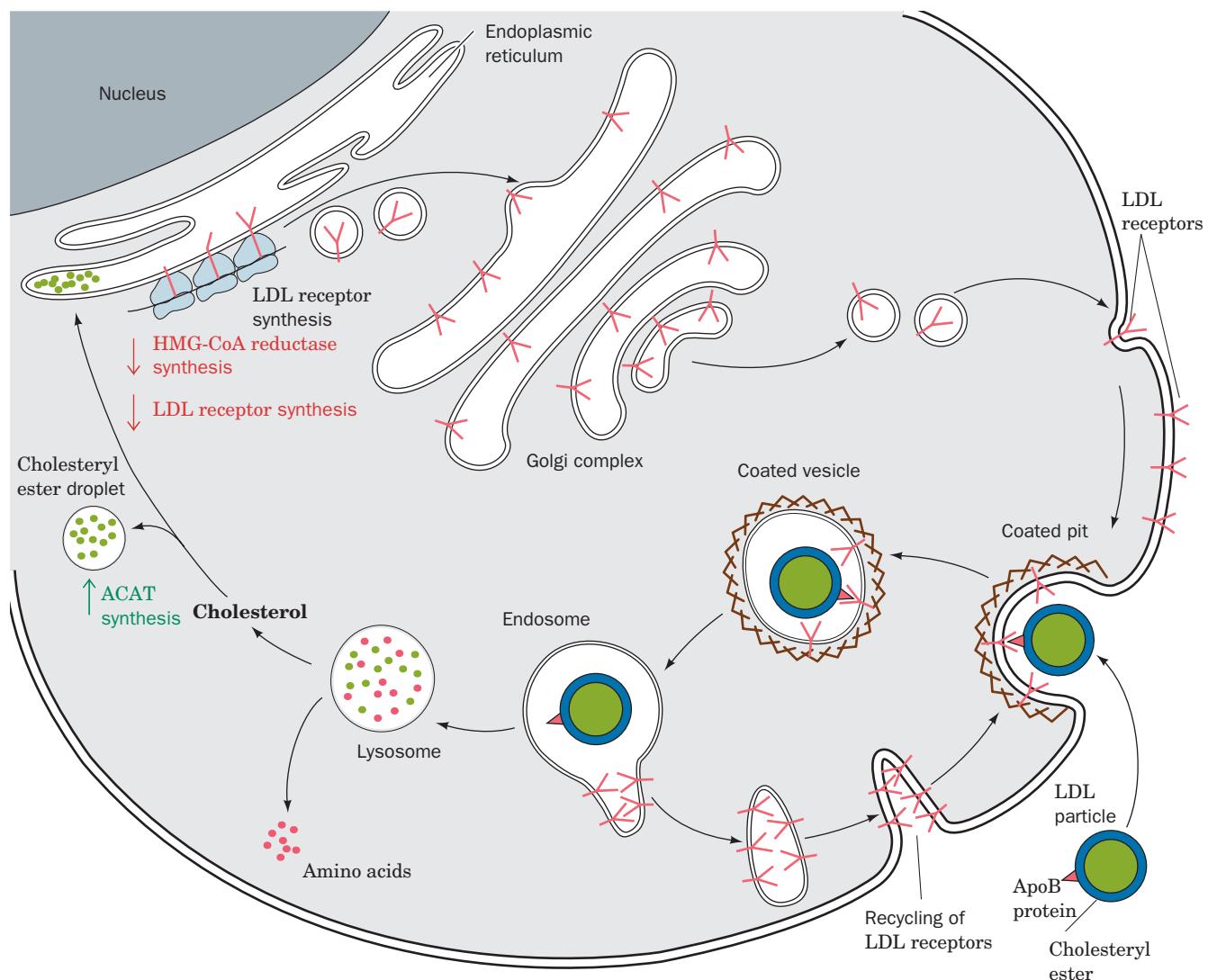


Figure 25-60 LDL receptor-mediated endocytosis in mammalian cells. LDL receptor is synthesized on the endoplasmic reticulum, processed in the Golgi complex, and inserted into the plasma membrane as a component of coated pits. LDL is specifically bound by the LDL receptor on the coated pit and brought into the cell in endosomes that deliver LDL to lysosomes while recycling LDL receptor to the plasma

membrane (Section 12-5Bc). Lysosomal degradation of LDL releases cholesterol, whose presence decreases the rate of synthesis of HMG-CoA reductase and LDL receptors (down arrows) while increasing that of acyl-CoA:cholesterol acyltransferase (ACAT; up arrow). [After Brown, M.S. and Goldstein, J.L., *Curr. Top. Cell. Reg.* **26**, 7 (1985).] 

by the *de novo* pathway we have just discussed, or they may obtain it from the bloodstream by receptor-mediated endocytosis. A small amount of cholesterol also enters cells by a non-receptor-mediated pathway. Note however, that the brain, which comprises ~2% of the human body mass but contains ~30% of its cholesterol, must synthesize all of its cholesterol because cholesterol cannot pass the blood-brain barrier.

Cholesterol actually circulates back and forth between the liver and peripheral tissues. While LDL transports cholesterol from the liver, cholesterol is transported back to the liver by **high-density lipoproteins (HDL)**. Surplus cholesterol is disposed of by the liver as bile salts, thereby protecting the body from an overaccumulation of this water-insoluble substance.

B. Control of Cholesterol Biosynthesis and Transport

Cholesterol biosynthesis and transport must be tightly regulated. There are three ways in which the cellular cholesterol supply is maintained:

1. By regulating the activity of HMG-CoA reductase, the enzyme catalyzing the rate-limiting step in the *de novo* cholesterol biosynthesis pathway. This is accomplished in two ways:
 - (i) Short-term regulation of the enzyme's catalytic activity by (a) competitive inhibition, (b) allosteric effects, and (c) covalent modification involving reversible phosphorylation.
 - (ii) Long-term regulation of the enzyme's concentration by modulating its rates of synthesis and degradation.
2. By regulating the rate of LDL receptor synthesis, and therefore the rate of cholesterol uptake. High intracellular concentrations of cholesterol suppress LDL receptor synthesis, whereas low cholesterol concentrations stimulate it.
3. By regulating the rate of esterification and hence the removal of free cholesterol. ACAT, the enzyme that catalyzes intracellular cholesterol esterification, is regulated by reversible phosphorylation and by long-term control.

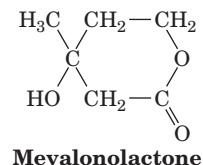
a. HMG-CoA Reductase Is the Primary Control Site for Cholesterol Biosynthesis

HMG-CoA reductase is the rate-limiting enzyme in cholesterol biosynthesis and, as therefore might be expected, constitutes the pathway's main regulatory site. The pathway branches after this reaction, however (Fig. 25-46); ubiquinone, dolichol, farnesylated and geranylgeranylated proteins, and isopentenyl adenosine are also essential, albeit minor, products. HMG-CoA is therefore subject to "multivalent" control, both long-term and short-term, in order to coordinate the synthesis of all of these products.

b. Long-Term Feedback Regulation of HMG-CoA Reductase Is Its Primary Means of Control

The main way in which HMG-CoA reductase is controlled is by long-term feedback control of the amount of enzyme present in the cell. When either LDL-cholesterol or

mevalonate levels fall, the amount of HMG-CoA reductase present in the cell can rise as much as 200-fold, due to an increase in enzyme synthesis combined with a decrease in its degradation. When LDL-cholesterol or **mevalonolactone** (an internal ester of mevalonate that is hydrolyzed to mevalonate and metabolized in the cell)



are added back to a cell, these effects are reversed.

The mechanism by which cholesterol serves to control the expression of the >20 genes involved in its biosynthesis and uptake, such as those encoding HMG-CoA reductase and the LDL receptor, has been elucidated by Michael Brown and Joseph Goldstein. These genes all contain a DNA sequence upstream from the transcription initiation site called the **sterol regulatory element (SRE)**. In order for these genes to be transcribed, a specific transcription factor, the **sterol regulatory element binding protein (SREBP)**, must bind to the SRE (eukaryotic gene expression is discussed in Section 34-3). SREBP is synthesized as an integral membrane protein that, when the cholesterol concentration is sufficiently high (Fig. 25-61a), resides in the ER membrane in complex with **SREBP cleavage-activating protein (SCAP)** and a protein named **Insig**. SREBP (~1160 residues) consists of three domains (Fig. 25-61a): (1) an ~480-residue cytosolic N-terminal domain that is a member of the **basic helix-loop-helix/leucine zipper (bHLH/Z)** family of transcription factors (Section 34-3Br), which specifically binds to SREs; (2) an ~90-residue transmembrane (TM) domain consisting of two TM helices connected by an ~30-residue hydrophilic luminal loop; and (3) an ~590-residue cytosolic C-terminal regulatory domain. SCAP (1276 residues) consists of two domains: (1) a 730-residue N-terminal domain that contains eight TM helices; and (2) a 546-residue cytosolic C-terminal domain that contains five copies of the protein-protein interaction motif known as a **WD repeat** (also called a WD40 sequence motif because it is ~40 residues long; Section 19-2Cc) and which presumably forms a 5-bladed β propeller similar to the 7-bladed β propeller of the G_{β} subunit (Fig. 19-19b). Insig (named to denote insulin-induced gene, although the insulin effect was later found to be indirect) consists mainly of six TM helices. SCAP and SREBP associate through the interaction of the SCAP's regulatory domain with SREBP's WD domain (Fig. 25-61a).

SCAP functions as a sterol sensor. An ~170-residue segment of its TM domain, the **sterol-sensing domain**, interacts with sterols although how it does so is unknown. When the cholesterol in the ER membrane is depleted (Fig. 25-61b), SCAP changes conformation and then

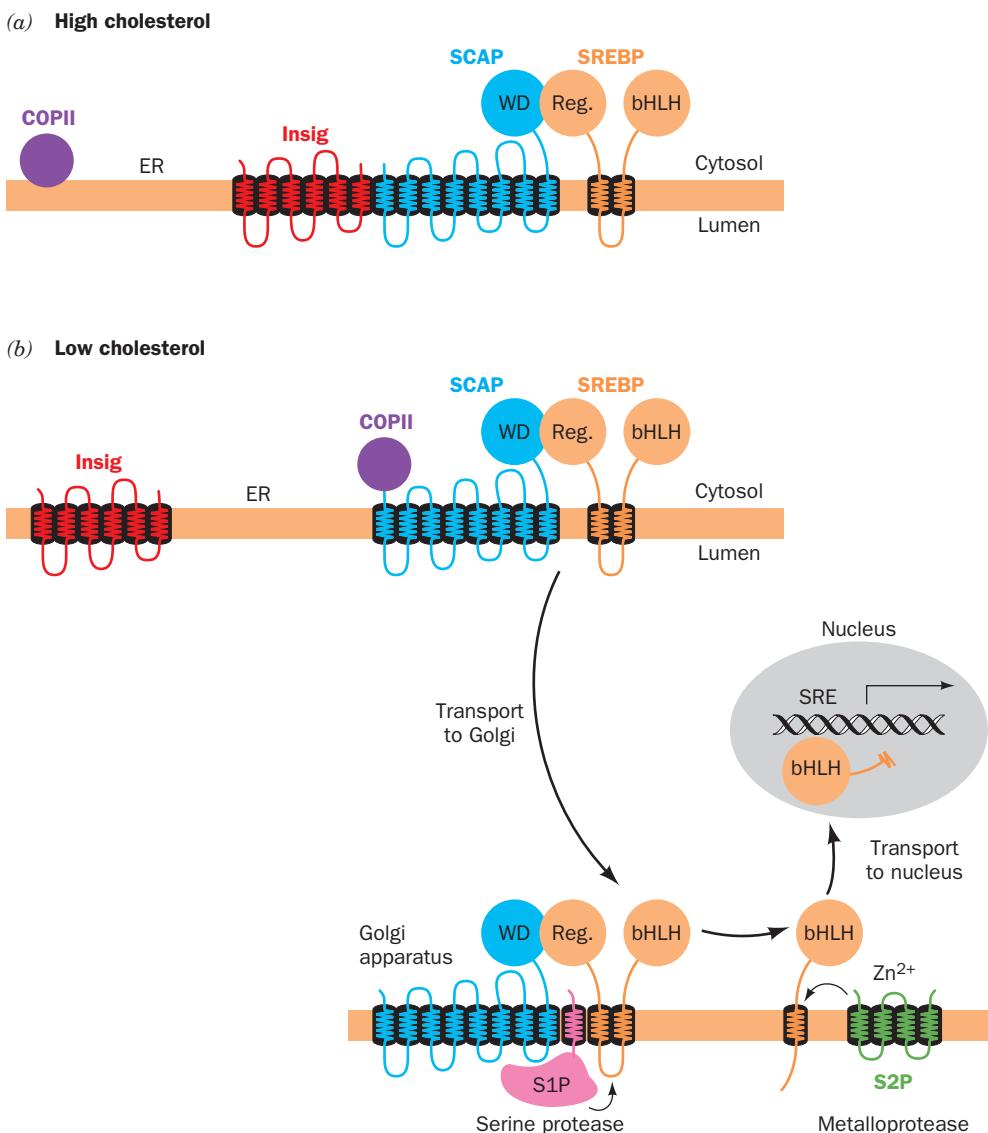


Figure 25-61 The cholesterol-mediated proteolytic activation of SREBP. (a) When cholesterol levels in the cell are high, the Insig–SCAP–SREBP complex resides in the ER membrane. (b) When cholesterol levels are low, SCAP releases Insig and binds COPII protein. The SCAP–SREBP complex is then transported, via COPII-coated vesicles, to the Golgi apparatus, where SREBP

undergoes sequential proteolytic cleavage by the membrane-bound proteases S1P and S2P. This releases SREBP's bHLH/Z-containing N-terminal domain, which enters the nucleus where it binds to the SREs of its target genes, thereby inducing their transcription. [After Goldstein, J., Rawson, R.B., and Brown, M., *Arch. Biochem. Biophys.* **397**, 139 (2002).]

releases Insig, which allows SCAP to bind to the Sec24 cargo-binding subunit of the COPII coated vesicle protein (Section 12-4Ca). The COPII vesicle then escorts its bound SREBP–SCAP complex to the Golgi apparatus.

In the Golgi apparatus, SREBP is sequentially cleaved by two membrane-bound proteases (Fig. 25-61b). **Site-1 protease (S1P)**, a serine protease of the subtilisin family, cleaves SREBP in the luminal loop that connects its two transmembrane helices but only when it is associated with SCAP. This cleavage exposes a peptide bond located 3 residues from the beginning of SREBP's N-terminal

TM helix to cleavage by **site-2 protease (S2P)**, a zinc metalloprotease. This releases the bHLH/Z domain to migrate to the nucleus where it activates the transcription of its target genes. The cholesterol level in the cell thereby rises until SCAP no longer induces the translocation of SREBP to the Golgi, a classic case of feedback inhibition.

This complex regulatory pathway was elucidated, in part, through the generation of several lines of transgenic mice that overexpress one or another of the foregoing proteins, and knockout mice that lack one or another of these

proteins. For example, knockout mice lacking either SCAP or S1P in their livers have decreased expression of both HMG-CoA reductase and LDL receptor, even when fed a cholesterol-deficient diet. In contrast, transgenic mice that overexpress SREBP or SCAP have greatly increased expression of the foregoing proteins. In fact, animals overproducing only the bHLH/Z domain of SREBP have massively enlarged livers (up to 4-fold larger than normal) due to engorgement with triacylglycerols and cholesteryl esters and yet they continue to transcribe SREBP's target genes such that their mRNA levels are up to 75-fold greater than normal. Many individuals suffering from obesity or diabetes caused by insulin resistance (type 2 diabetes; Section 27-4B) have fatty livers, which in some cases leads to liver failure. Fatty livers due to insulin resistance appear to be caused by elevated levels of SREBP in response to elevated insulin levels.

The level of HMG-CoA reductase also responds to the level of the cholesterol precursor lanosterol (Section 25-6Ad). HMG-CoA reductase's ER membrane-bound N-terminal domain contains eight TM helices, whereas its C-terminal domain, which contains its active site and is linked to the N-terminal domain via a flexible Pro-rich sequence, projects into the cytosol. Insig binds to an enzymatic complex that marks proteins for degradation by covalently linking them to the protein **ubiquitin** (Section 32-6B). When lanosterol accumulates in the ER membrane, the N-terminal domain of HMG-CoA reductase also binds to Insig, and is thus marked for destruction. Consequently, the >12-hour half-life of HMG-CoA reductase in sterol-deprived cells decreases to <1 hour when sterols are plentiful.

c. Regulation of HMG-CoA Reductase by Covalent Modification Is a Means of Cellular Energy Conservation

HMG-CoA reductase exists in interconvertible more active and less active forms, as do glycogen phosphorylase (Section 18-3Ca), glycogen synthase (Section 18-3D), pyruvate dehydrogenase (Section 21-2Cb), and acetyl-CoA carboxylase (Section 25-4Ba), among others. The unmodified form of HMG-CoA reductase is more active and the phosphorylated form is less active. HMG-CoA reductase is phosphorylated (inactivated) at its Ser 871 in a bicyclic cascade system by the covalently modifiable enzyme AMP-dependent protein kinase (AMPK), which, as we saw in Section 25-4Ba, also acts on acetyl-CoA carboxylase [in this context, this enzyme was originally named **HMG-CoA reductase kinase (RK)**, until it was found to be identical to AMPK]. It appears that this control is exerted to conserve energy when ATP levels fall and AMP levels rise, by inhibiting biosynthetic pathways. This hypothesis was tested by Brown and Goldstein, who used genetic engineering techniques to produce hamster cells containing a mutant HMG-CoA reductase with Ala replacing Ser 871 and therefore incapable of phosphorylation control. These cells respond normally to feedback regulation of cholesterol biosynthesis by LDL-cholesterol and mevalonate but, un-

like normal cells, do not decrease their synthesis of cholesterol on ATP depletion, supporting the idea that control of HMG-CoA reductase by phosphorylation is involved in energy conservation.

d. LDL Receptor Activity Controls Cholesterol Homeostasis

LDL receptors clearly play an important role in the maintenance of plasma LDL-cholesterol levels. In normal individuals, about half of the IDL formed from the VLDL reenters the liver through LDL receptor-mediated endocytosis (IDL and LDL both contain apolipoproteins that specifically bind to the LDL receptor; Section 12-5Bc). The remaining IDL are converted to LDL (Fig. 25-62a). *The serum concentration of LDL therefore depends on the rate at which liver removes IDL from the circulation, which, in turn, depends on the number of functioning LDL receptors on the liver cell surface.*

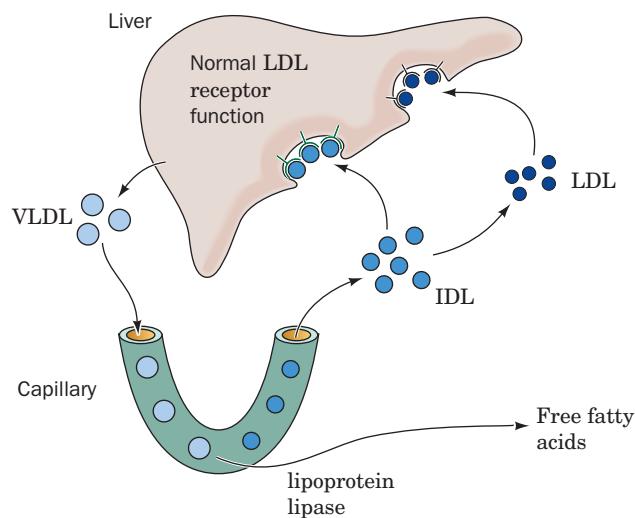
High blood cholesterol (**hypercholesterolemia**), which results from the overproduction and/or underutilization of LDL, is known to be caused by either of two metabolic irregularities: (1) the genetic disease **familial hypercholesterolemia (FH)** or (2) the consumption of a high-cholesterol diet. FH is a dominant genetic defect that results in a deficiency of functional LDL receptors (Section 12-5Ca). Homozygotes for this disorder lack functional LDL receptors, so their cells can absorb neither IDL nor LDL by receptor-mediated endocytosis. The increased concentration of IDL in the bloodstream leads to a corresponding increase in LDL, which is, of course, underutilized since it cannot be taken up by the cells (Fig. 25-62b). FH homozygotes therefore have plasma LDL-cholesterol levels three to five times higher than average. FH heterozygotes, which are far more common, have about half of the normal number of functional LDL receptors and plasma LDL-cholesterol levels about twice the average.

The long-term ingestion of a high-cholesterol diet has an effect similar, although not as extreme, as FH (Fig. 25-62c). Excessive dietary cholesterol enters the liver cells in chylomicron remnants and represses the synthesis of LDL receptor protein. The resulting insufficiency of LDL receptors on the liver cell surface has consequences similar to those of FH.

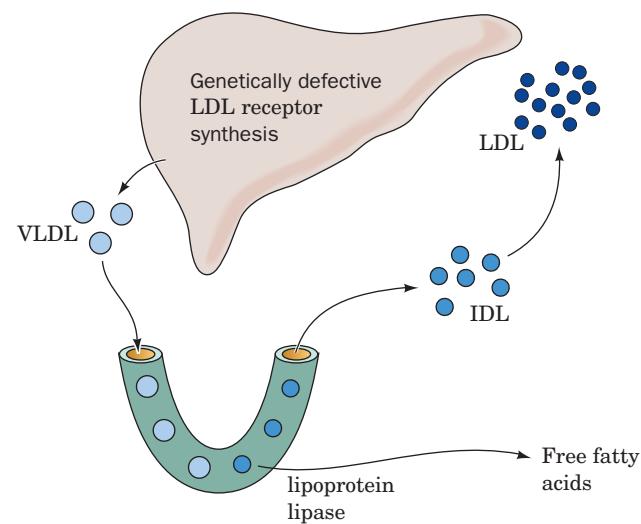
LDL receptor deficiency, whether of genetic or dietary origin, raises the LDL level by two mechanisms: (1) increased LDL production resulting from decreased IDL uptake and (2) decreased LDL uptake. Two strategies for reversing these conditions (besides maintaining a low-cholesterol diet) are being used in humans:

1. *Ingestion of anion exchange resins (Section 6-3A) that bind bile salts, thereby preventing their intestinal absorption* (resins are insoluble in water). Bile salts, which are derived from cholesterol, are normally efficiently recycled by the liver (Section 25-6C). Elimination of resin-bound bile salts in the feces forces the liver to convert more cholesterol to bile salts than otherwise. The consequent decrease in the serum cholesterol concentration induces synthesis of LDL receptors (of course, not in FH homozygotes). Unfortunately, the

(a) Normal



(b) Familial hypercholesterolemia



(c) High cholesterol diet

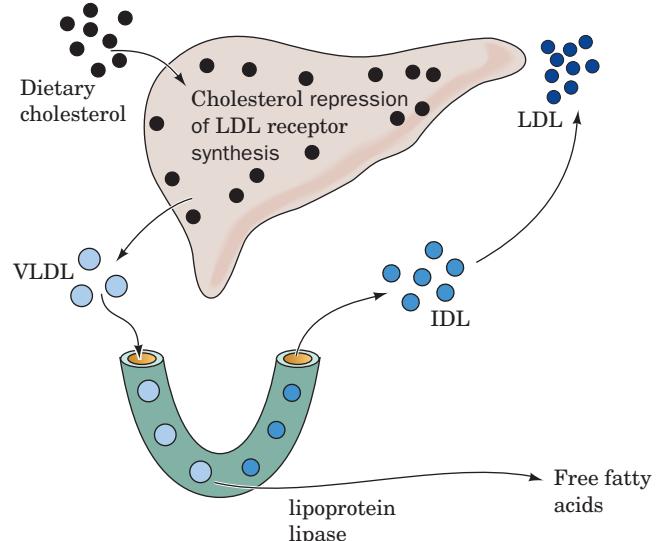


Figure 25-62 Control of plasma LDL production and uptake by liver LDL receptors. (a) In normal human subjects, VLDL is secreted by the liver and converted to IDL in the capillaries of the peripheral tissues. About half of the plasma IDL particles bind to the LDL receptor and are taken up by the liver. The remainder are converted to LDL at the peripheral tissues. (b) In individuals with familial hypercholesterolemia (FH), liver LDL receptors are diminished or eliminated because of a genetic

defect. (c) In normal individuals who have a long-term high-cholesterol diet, the liver is filled with cholesterol, which represses the rate of LDL receptor production. Receptor deficiency, whether of genetic or dietary cause, raises the plasma LDL level by increasing the rate of LDL production and decreasing the rate of LDL uptake. [After Goldstein, J.L. and Brown, M.S., *J. Lipid Res.* **25**, 1457 (1984).]

decreased serum cholesterol level also induces the synthesis of HMG-CoA reductase, which increases the rate of cholesterol biosynthesis. Ingestion of bile salt-binding resins such as **cholestyramine** (sold as **Questran**) therefore provides only a 15 to 20% drop in serum cholesterol levels.

2. Treatment with competitive inhibitors of HMG-CoA reductase. These include (Fig. 25-63) the fungal derivatives **lovastatin** (also called **mevinolin** and sold as **Mevacor**), **pravastatin** (**Pravachol**), and **simvastatin** (**Zocor**) as well as the synthetic inhibitor **atorvastatin** (**Lipitor**), compounds

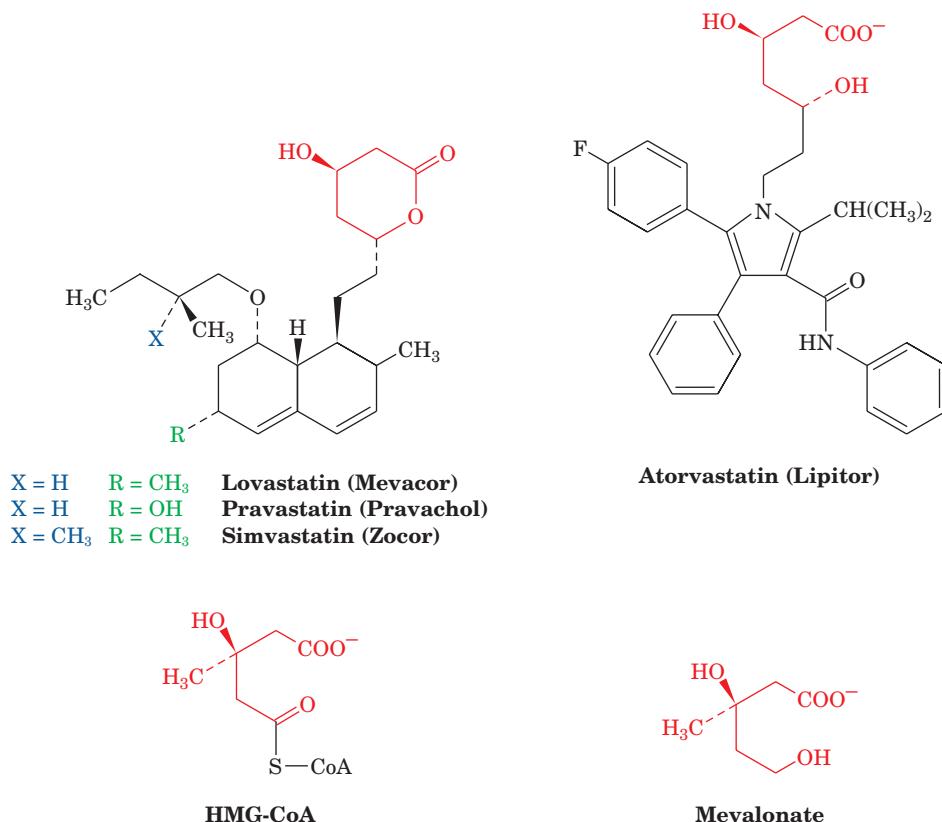


Figure 25-63 Some competitive inhibitors of HMG-CoA reductase used for the treatment of hypercholesterolemia. The molecular formulas of lovastatin (Mevacor), pravastatin (Pravachol), simvastatin (Zocor), and atorvastatin (Lipitor), all of which are potent competitive inhibitors of HMG-CoA

reductase, are given. The structures of HMG-CoA and mevalonate are shown for comparison. Note that lovastatin, pravastatin, and simvastatin are lactones, whereas atorvastatin and mevalonate are hydroxy acids. The lactones are hydrolyzed enzymatically *in vivo* to their active hydroxy-acid forms.

that are collectively known as **statins**. Indeed, Lipitor is presently one of the most widely prescribed drugs in the United States. The initial decreased cholesterol supply in the cell caused by the presence of statins is again met by induction of LDL receptors and HMG-CoA reductase so that, at the new steady state, the HMG-CoA reductase level is almost that of the predrug state. However, the increased number of LDL receptors causes increased removal of both LDL and IDL (the apoB-containing precursor to LDL), decreasing serum LDL levels appreciably. Lipitor-treated FH heterozygotes routinely show a serum cholesterol decrease of 40–50%.

The combined use of these agents, moreover, results in a clinically dramatic 50 to 60% decrease in serum cholesterol levels.

e. Overexpression of LDL Receptor Prevents Diet-Induced Hypercholesterolemia

Experiments are well underway toward the treatment of hypercholesterolemic individuals by **gene therapy**

(Section 5-5Hb). A line of transgenic mice has been developed that overproduce the human LDL receptor. When fed a diet high in cholesterol, fat, and bile salts, these transgenic animals did not develop a detectable increase in plasma LDL. In contrast, normal mice fed the same diet exhibited large increases in plasma LDL levels. Evidently, the unregulated overexpression of LDL receptors can prevent diet-induced hypercholesterolemia, at least in mice.

C. Cholesterol Utilization

Cholesterol is the precursor of steroid hormones and bile salts. Steroid hormones, which are grouped into five categories, **progestins**, **glucocorticoids**, **mineralocorticoids**, **androgens**, and **estrogens**, mediate a wide variety of vital physiological functions (Section 19-1G). All contain the four-ring structure of the sterol nucleus and are remarkably similar in structure, considering the enormous differences in their physiological effects. A simplified

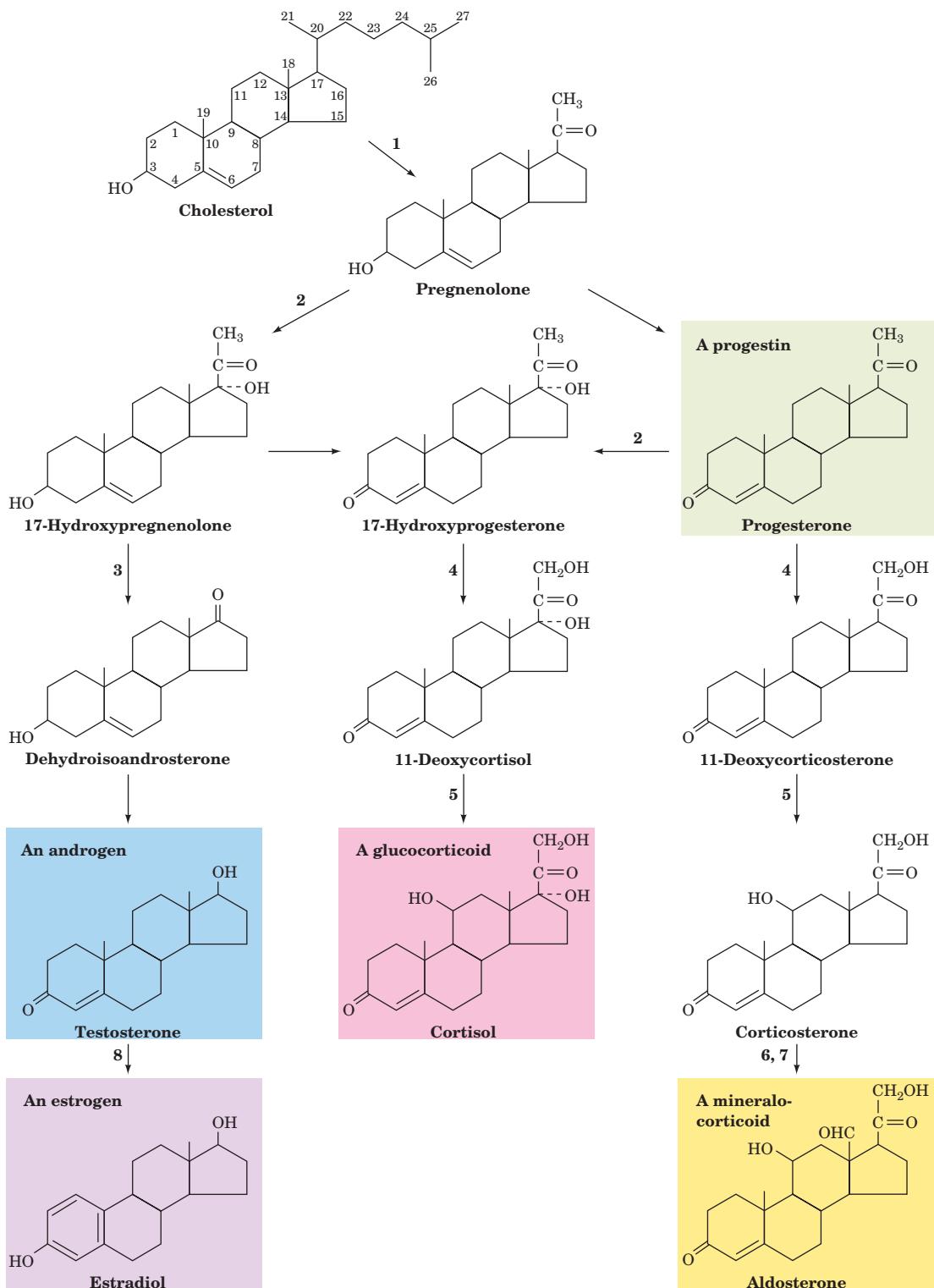


Figure 25-64 Simplified scheme of steroid biosynthesis. The enzymes involved are (1) the cholesterol side chain cleavage enzyme, (2) steroid C17 hydroxylase, (3) steroid C17, C20 lyase,

(4) steroid C21 hydroxylase, (5) steroid 11 β -hydroxylase, (6) steroid C18 hydroxylase, (7) 18-hydroxysteroid oxidase, and (8) aromatase.

biosynthetic scheme (Fig. 25-64) indicates their structural similarities and differences. We shall not discuss the details of these pathways.

The quantitatively most important pathway for the excretion of cholesterol in mammals is the formation of **bile acids**. The major bile acids, **cholate** and **chenodeoxycholate**,

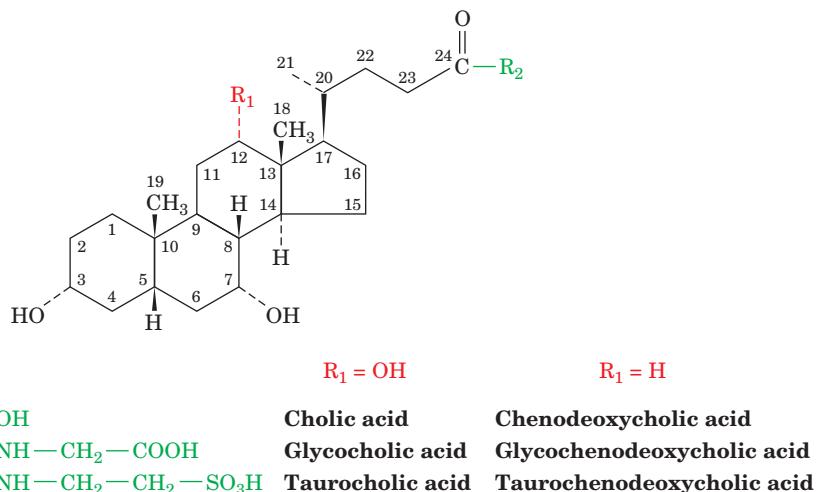


Figure 25-65 Structures of the major bile acids and their glycine and taurine conjugates.

are synthesized in the liver and secreted as their glycine or **taurine** conjugates (Fig. 25-65), which are known as bile salts, into the gallbladder. From there, they are secreted into the small intestine, where they act as emulsifying agents in the digestion and absorption of fats and fat-soluble vitamins (Section 25-1). An efficient recycling system allows the bile salts to reenter the bloodstream and return to the liver for reuse several times each day. The $<1\text{ g} \cdot \text{day}^{-1}$ of bile salts that normally escape this recycling system are further metabolized by microorganisms in the large intestine and excreted. *This is the body's only route for cholesterol excretion.*

Comparison of the structures of cholesterol and the bile acids (Figs. 25-44 and 25-65) indicates that biosynthesis of bile acids from cholesterol involves (1) saturation of the 5,6 double bond, (2) epimerization of the 3β -OH group, (3) introduction of OH groups into the 7α and 12α positions, (4) oxidation of C24 to a carboxylic acid, and (5) conjugation of this side chain carboxylic acid with glycine or taurine. **Cholesterol 7α -hydroxylase** catalyzes the first and rate-limiting step in bile acid synthesis and is closely regulated.

(phosphate buffer-soluble; *fosfat* in Swedish). This began an explosion of research on these potent substances.

*Almost all mammalian cells except red blood cells produce prostaglandins and their related compounds, the prostacyclins, thromboxanes, leukotrienes and lipoxins, known collectively as eicosanoids since they are all C_{20} compounds; Greek: *eikosi*, twenty). The eicosanoids, like endocrine hormones, have profound physiological effects at extremely low concentrations.* For example, they mediate the following: (1) the inflammatory response, notably as it involves the joints (rheumatoid arthritis), skin (psoriasis), and eyes; (2) the production of pain and fever; (3) the regulation of blood pressure; (4) the induction of blood clotting; (5) the control of several reproductive functions such as the induction of labor; and (6) the regulation of the sleep/wake cycle. The enzymes that synthesize these compounds and the receptors to which they bind are therefore the targets of intensive pharmacological research.

The eicosanoids are also hormonelike in that they bind to G-protein-coupled receptors (Section 19-2B), and many of their effects are intracellularly mediated by cAMP. Unlike endocrine hormones, however, they are not transported in the bloodstream to their sites of action. Rather, these chemically and biologically unstable substances (some decompose within minutes or less *in vitro*) are local mediators (paracrine hormones; Section 19-1); that is, *they act in the same environment in which they are synthesized.*

In this section, we discuss the structures of the eicosanoids and outline their biosynthetic pathways and modes of action. As we do so, note the great diversity of their structures and functions, a phenomenon that makes the elucidation of the physiological roles of these potent substances a challenging research area.

A. Background

*Prostaglandins are all derivatives of the hypothetical C_{20} fatty acid **prostanoic acid** in which carbon atoms 8 to 12*

7 EICOSANOID METABOLISM: PROSTAGLANDINS, PROSTACYCLINS, THROMBOXANES, LEUKOTRIENES, AND LIPOXINS

Prostaglandins (PGs) were first identified in human semen by Ulf von Euler in the early 1930s through their ability to stimulate uterine contractions and lower blood pressure. von Euler thought that these compounds originated in the prostate gland (hence their name) but they were later shown to be synthesized in the seminal vesicles. By the time the mistake was realized, the name was firmly entrenched. In the mid-1950s, crystalline materials were isolated from biological fluids and called PGE (ether-soluble) and PGF

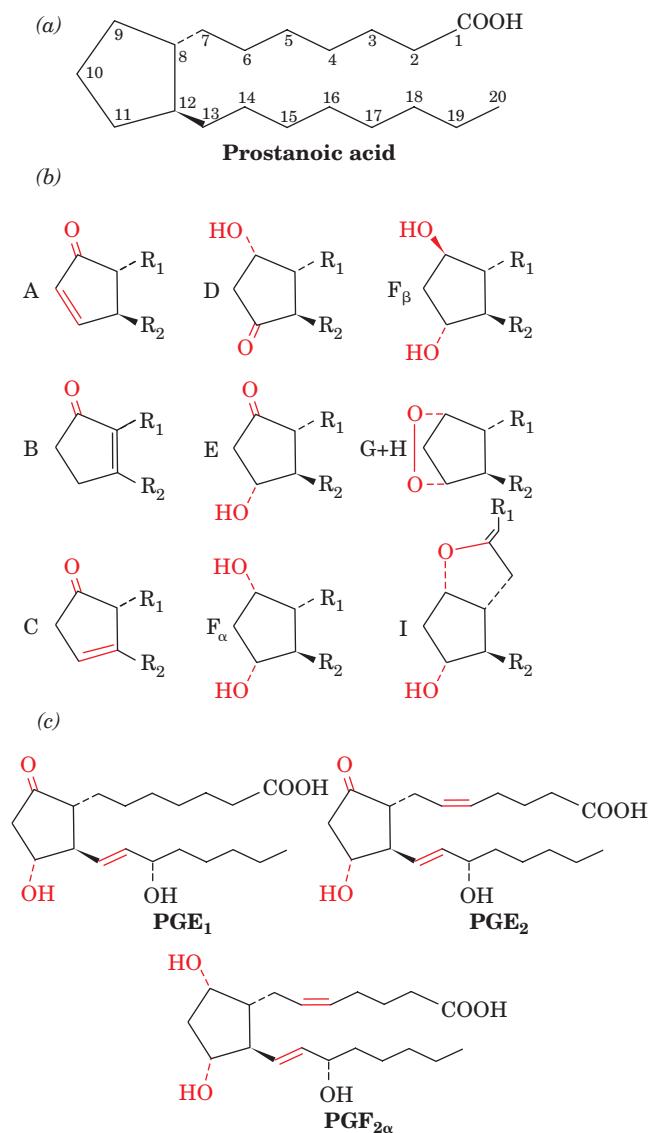


Figure 25-66 Prostaglandin structures. (a) The carbon skeleton of prostanoic acid, the prostaglandin parent compound. (b) Structures of prostaglandins A through I. (c) Structures of prostaglandins E₁, E₂, and F_{2α} (the first prostaglandins to be identified).

form a cyclopentane ring (Fig. 25-66a). Prostaglandins A through I differ in the substituents on the cyclopentane ring (Fig. 25-66b): **PGAs** are α,β -unsaturated ketones, **PGEs** are β -hydroxy ketones, **PGFs** are 1,3-diols, etc. In **PGF_α**, the C9 OH group is on the same side of the ring as R₁; it is on the opposite side in **PGF_β**. The numerical subscript in the name refers to the number of double bonds contained on the side chains of the cyclopentane ring (Fig. 25-66c).

In humans, the most prevalent prostaglandin precursor is **arachidonic acid (5,8,11,14-eicosatetraenoic acid)**, a C₂₀ polyunsaturated fatty acid that has four nonconjugated double bonds. The double bond at C14 is six carbon atoms from the terminal carbon atom (the ω carbon atom), making arachidonic acid an ω -6 fatty acid. As Sune Bergström and Bengt Samuelsson demonstrated, arachidonic acid is syn-

thesized from the essential fatty acid linoleic acid (also an ω -6 fatty acid). This occurs via its desaturation with a Δ^6 -desaturase to yield **γ -linolenic acid (GLA)**, followed by elongation and a second desaturation, this time with a Δ^5 -desaturase (Fig. 25-67; Section 25-4E). Prostaglandins with the subscript 1 (the “series-1” prostaglandins) are synthesized from **dihomo- γ -linolenic acid (DGLA; 8,11,14-eicosatrienoic acid)**, whereas “series-2” prostaglandins are synthesized from arachidonic acid. α -Linolenic acid (ALA), another essential fatty acid since the Δ^{15} -desaturase required for its synthesis occurs only in plants, is a precursor of 5,8,11,14,17-eicosapentaenoic acid (EPA) and the “series-3” prostaglandins. Since arachidonate is the primary prostaglandin precursor in humans, we shall mostly refer to the series-2 prostaglandins in our examples. Note, however, that when dietary linoleic acid and α -linolenic acid are equally available, the relative activities of the Δ^5 - and Δ^6 -desaturases are important in determining the relative amounts of these prostaglandin precursors.

a. Arachidonate Is Generated by Phospholipid Hydrolysis

Arachidonate is stored in cell membranes esterified at glycerol C2 of phosphatidylserine and other phospholipids. The production of arachidonate metabolites is controlled by the rate of arachidonate release from these phospholipids through three alternative pathways (Fig. 25-68):

1. Phospholipase A₂ hydrolyzes acyl groups at C2 of phospholipids (Fig. 25-68b, *left*).

2. Phospholipase C (Section 19-4B) specifically hydrolyzes the phosphatidylserine head group to yield a **1,2-diacylglycerol (DAG)** and **phosphoinositol**. DAG is phosphorylated by **diacylglycerol kinase** to phosphatidic acid, a phospholipase A₂ substrate (Fig. 25-68b, *center*). (Recall that DAG and the various phosphorylated forms of phosphoinositol are also important signaling molecules in that they mediate the phosphoinositide cascade; Section 19-4.)

3. The DAG also may be hydrolyzed directly by diacylglycerol lipase (Fig. 25-68b, *right*). Corticosteroids are used as anti-inflammatory agents because they inhibit phospholipase A₂, reducing the rate of arachidonate production.

b. Aspirin Inhibits Prostaglandin Synthesis

The use of **aspirin** as an analgesic (pain-relieving), antipyretic (fever-reducing), and anti-inflammatory agent has been widespread since the nineteenth century. Yet, it was not until 1971 that John Vane discovered its mechanism of action. Aspirin, as do other **nonsteroidal anti-inflammatory drugs (NSAIDs)**, inhibits the synthesis of prostaglandins from eicosanoid precursors (Section 25-7B). These inhibitors have therefore proved to be valuable tools in the elucidation of prostaglandin biosynthesis pathways and have provided a starting point for the rational synthesis of new anti-inflammatory drugs.

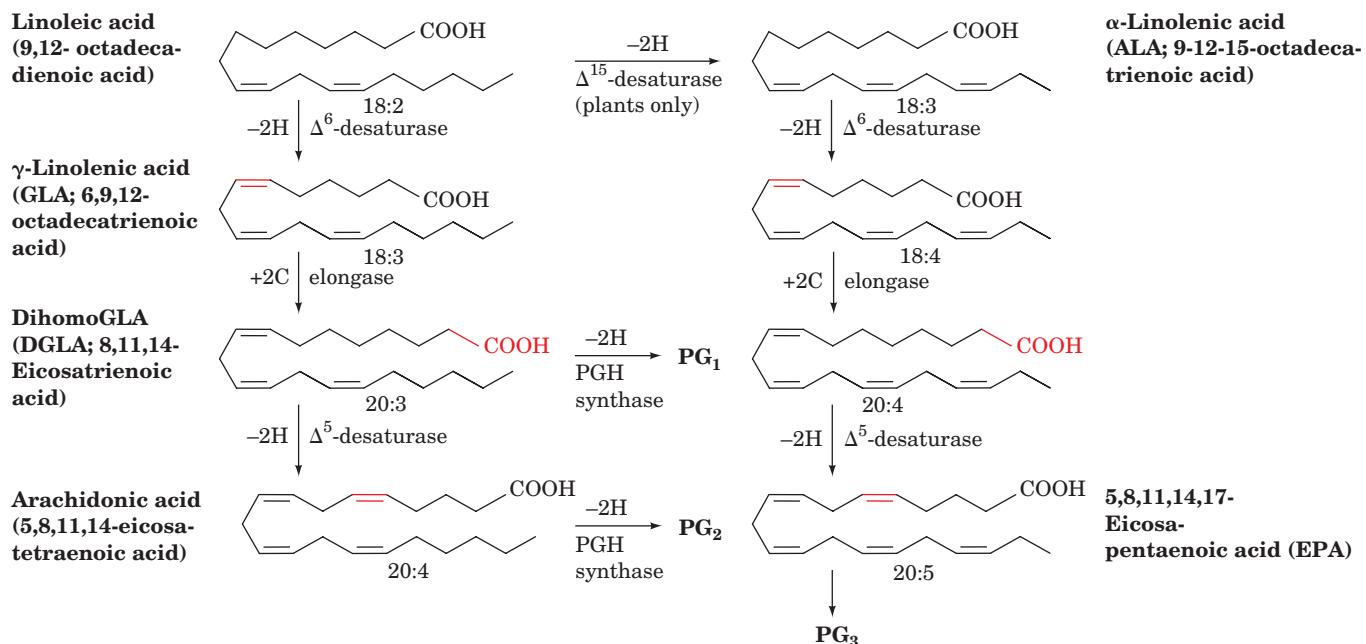


Figure 25-67 Synthesis of prostaglandin precursors. The linoleic acid derivatives dihomogamma-linolenic acid (DGLA), arachidonic acid,

and 5,8,11,14,17-eicosapentaenoic acid (EPA) are the respective precursors of the series-1, series-2, and series-3 prostaglandins.

c. Arachidonic Acid Is a Precursor of Leukotrienes, Thromboxanes, and Prostacyclins

Arachidonic acid also serves as a precursor to compounds whose synthesis is not inhibited by aspirin. In fact,

there are two main pathways of eicosanoid metabolism. The so-called cyclic pathway, which is inhibited by NSAIDs, forms prostaglandin's characteristic cyclopentane ring, whereas the so-called linear pathway, which is

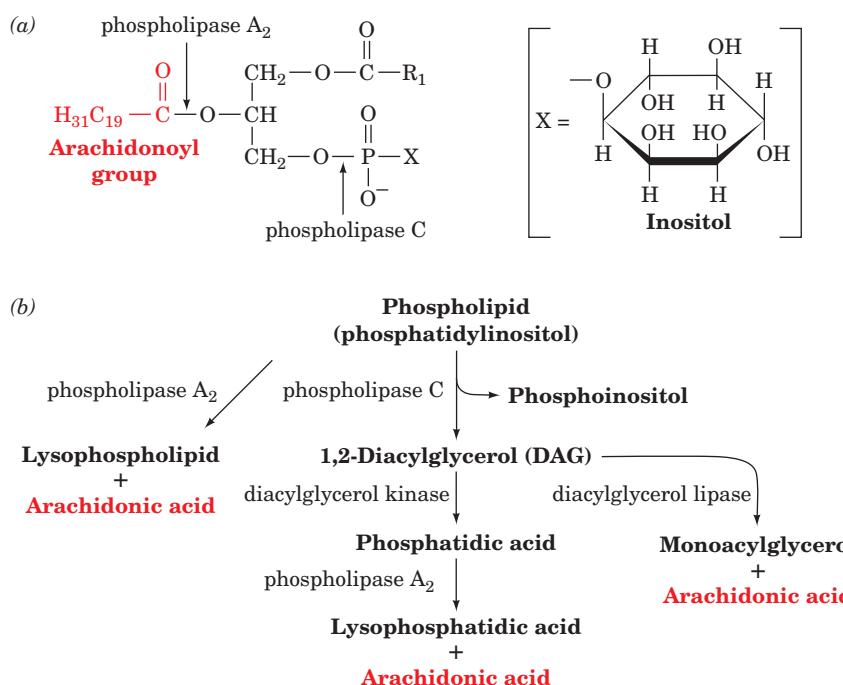


Figure 25-68 Release of arachidonic acid by phospholipid hydrolysis. (a) The sites of hydrolytic cleavage mediated by phospholipases A₂ and C. The polar head group, X, is often

inositol and its various phosphorylated forms (Section 19-4D). (b) Pathways of arachidonic acid liberation from phospholipids.

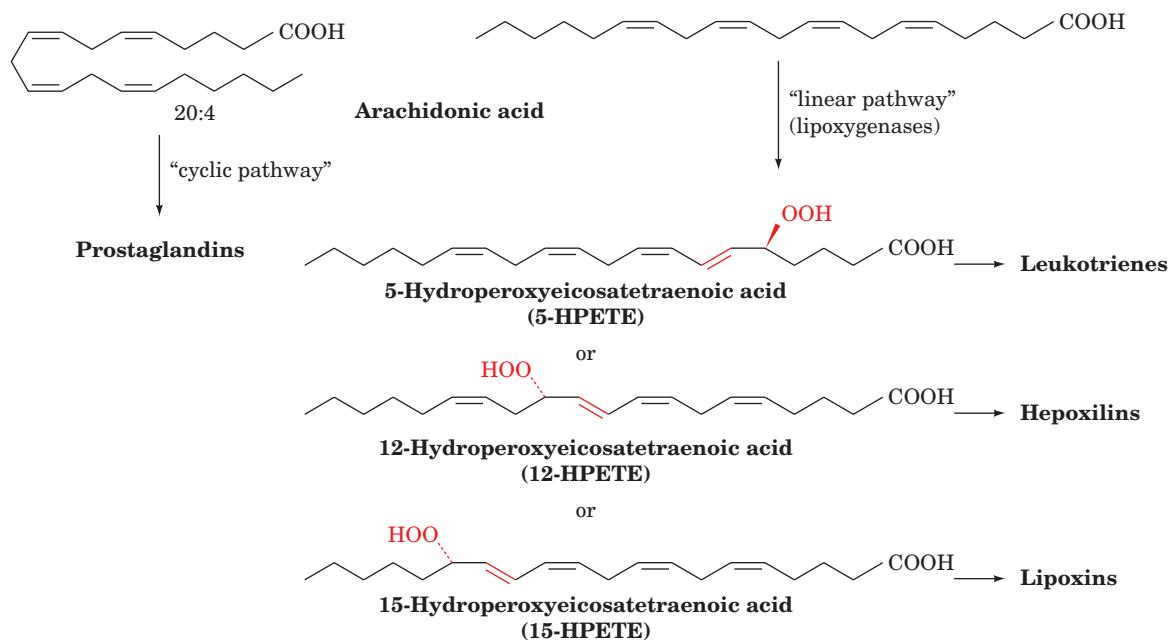


Figure 25-69 The cyclic and linear pathways of arachidonic acid metabolism.

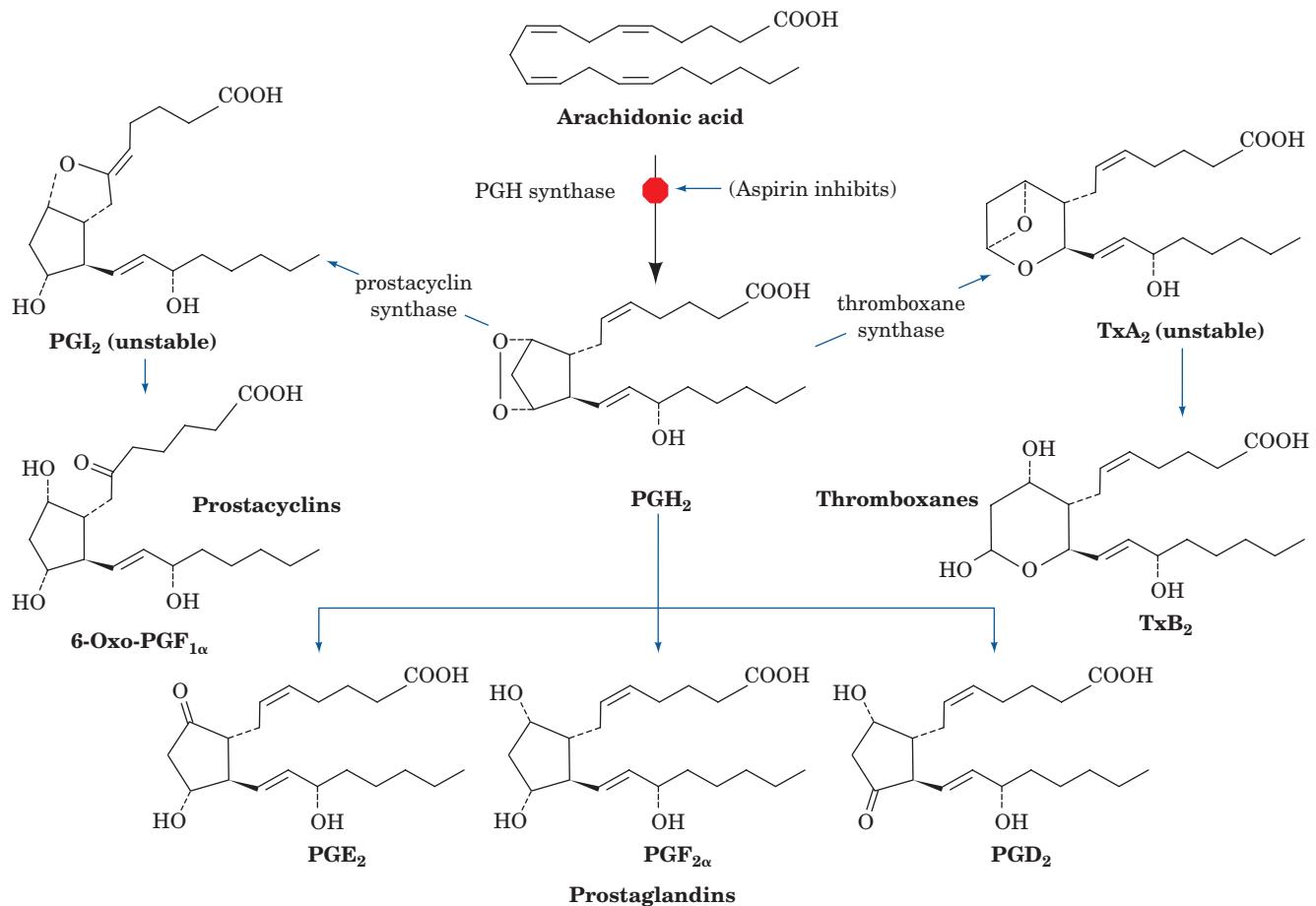


Figure 25-70 The cyclic pathway of arachidonic acid metabolism. This pathway's branches lead to prostaglandins, prostacyclins, and thromboxanes.

not inhibited by these agents, leads to the formation of the **leukotrienes** and **HPETEs** (Fig. 25-69; Section 25-7C).

Studies using NSAIDs helped demonstrate that two structurally related and highly short-lived classes of compounds, the prostacyclins and the thromboxanes (Fig. 25-70), are also products of the cyclic pathway of eicosanoid metabolism. The specific products produced by this branched pathway depend on the tissue involved. For example, blood platelets (thrombocytes) produce thromboxanes almost exclusively; vascular endothelial cells, which make up the walls of veins and arteries, predominantly synthesize the prostacyclins; and heart muscle makes PGI₂, PGE₂, and PGF_{2α} in more or less equal quantities. In the remainder of this section, we study the cyclic and the linear pathways of eicosanoid metabolism.

B. The Cyclic Pathway of Eicosanoid Metabolism: Prostaglandins, Prostacyclins, and Thromboxanes

The first step in the cyclic pathway of eicosanoid metabolism is catalyzed by **PGH synthase** (PGHS; also called **prostaglandin H synthase** and **prostaglandin endoperoxide synthase**; Fig. 25-71). This heme-containing enzyme contains two catalytic activities: a cyclooxygenase activity and a peroxidase activity. The former catalyzes the tyrosyl radical-mediated addition of two molecules of O₂ to arachidonic acid, forming **PGG₂**. The latter converts the hydroperoxy function of PGG₂ to an OH group, yielding **PGH₂**. **PGH₂** is the immediate precursor of all series-2 prostaglandins, prostacyclins, and thromboxanes (Fig. 25-70). The cyclooxygenase activity of the enzyme gives it its common name, **COX** [not to be confused with cytochrome *c* oxidase, which is also called COX (Section 22-2C5)].

PGHS, a homodimeric glycoprotein of 576-residue subunits, is a monotopic membrane protein that extends into the lumen of the endoplasmic reticulum. Its X-ray structure, determined by Michael Garavito, reveals that each of its subunits folds into three domains (Fig. 25-72a): an N-terminal module that structurally resembles **epidermal growth factor (EGF)**; a hormonally active polypeptide that stimulates cell proliferation; Section 12-3Ae); a central membrane-binding motif; and a C-terminal enzymatic domain. The 44-residue membrane-binding motif has a hydrophobic surface that faces away from the body of the protein [as is also true of oxidosqualene cyclase (Fig. 25-58b) and fatty acid amide hydrolase (Fig. 12-28)].

The peroxidase active site of PGHS occurs at the interface between the large and small lobes of the catalytic domain, in a shallow cleft that contains the enzyme's Fe(III)-heme prosthetic group. The cleft exposes a large portion of the heme to solvent and is therefore thought to comprise the substrate binding site.

The cyclooxygenase active site lies on the opposite side of the heme at the end of a long narrow hydrophobic channel (~8 × 25 Å) extending from the outer surface of the membrane-binding motif to the center of each subunit (Fig. 25-72b). This channel allows access of the membrane-associated substrate to the active site. Tyr 385, which lies near the top of the channel, just beneath the heme, has

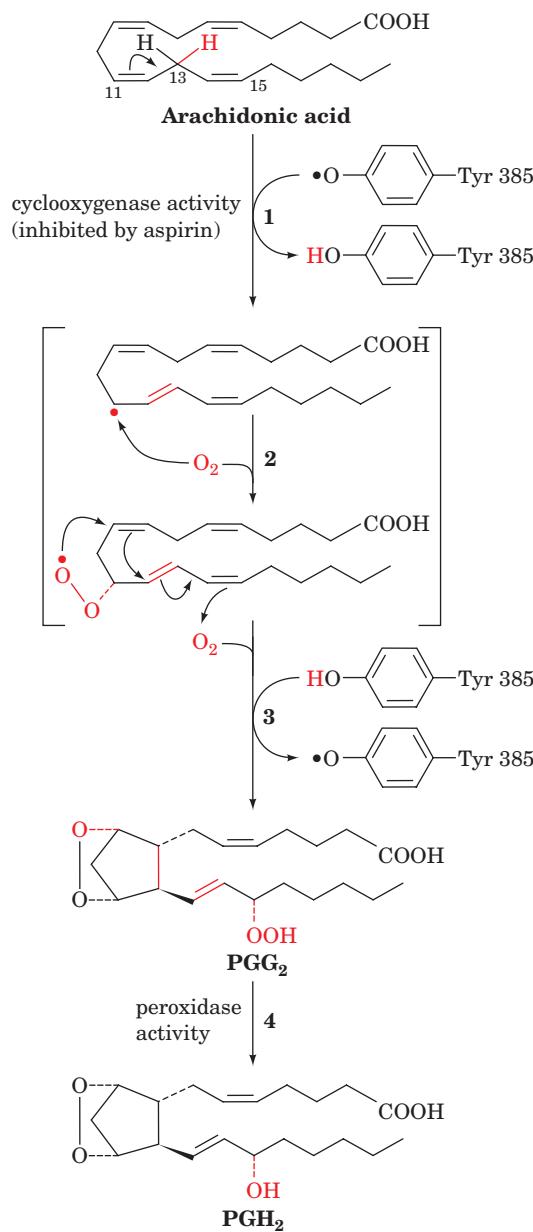


Figure 25-71 Reactions catalyzed by PGH synthase (PGHS). The enzyme contains two activities: a cyclooxygenase, which catalyzes Steps 1 to 3 and is inhibited by aspirin; and a peroxidase, which catalyzes Step 4. (1) A radical at Tyr 385 that is generated by the enzyme's heme cofactor stereospecifically abstracts a hydrogen atom from C13 of arachidonic acid, which then rearranges so that the radical is on C11. (2) The radical reacts with O₂ to yield a hydroperoxide radical. (3) The radical cyclizes and reacts with a second O₂ molecule at C15 to yield a peroxide in a process that regenerates the Tyr radical. (4) The enzyme's peroxidase activity converts the peroxide at C15 to a hydroxyl group.

been shown to form a transient radical during the cyclooxygenase reaction as does, for example, Tyr 244 of cytochrome *c* oxidase (Section 22-2C5c). Indeed, the mutagenic replacement of PGHS's Tyr 385 by Phe abolishes its cyclooxygenase activity. The Tyr 385 radical is generated via an intramolecular oxidation by the heme cofactor.

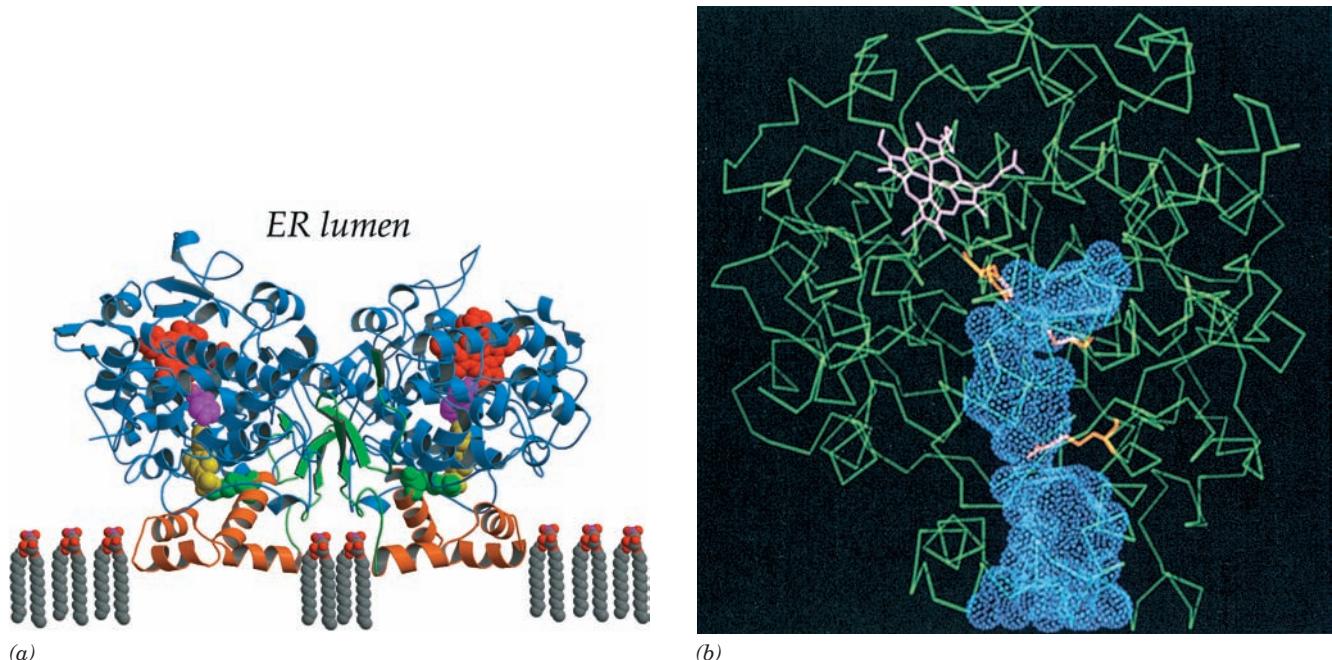


Figure 25-72 X-ray structure of PGH synthase (PGHS) from sheep seminal vesicles in complex with the NSAID flurbiprofen. (a) This homodimeric monotypic membrane protein is viewed parallel to the plane of the ER membrane with its 2-fold axis of symmetry vertical. The EGF-like module is green, the membrane-binding motif is orange, and the catalytic domain is blue. The heme (red); flurbiprofen (yellow); Tyr 385 (magenta), which forms a transient radical during the cyclooxygenase reaction; and Arg 120 (green), which forms an ion pair with flurbiprofen, are drawn in space-filling form. (b) A C_{α} diagram of

a PGHS subunit (green), the left subunit in Part *a* as viewed from 30° to the left. The peroxidase active site is located above the heme (pink). The hydrophobic channel, which penetrates the subunit from the membrane-binding motif at the bottom of the figure to the cyclooxygenase active site below the heme, is represented by its van der Waals surface (blue dots). The three residues in the channel that are shown in stick form in orange are, from top to bottom: Tyr 385, Ser 530, which is acetylated by aspirin, and Arg 120. [Courtesy of Michael Garavito, Michigan State University. PDBid 1CQE.]

The fate of PGH_2 depends on the relative activities of the enzymes catalyzing the specific interconversions (Fig. 25-70). Platelets contain **thromboxane synthase**, which mediates the formation of **thromboxane A₂ (TxA₂)**, a vasoconstrictor and stimulator of platelet aggregation (an initial step in blood clotting; Section 35-1). Vascular endothelial cells contain **prostacyclin synthase**, which catalyzes the synthesis of **prostacyclin I₂ (PGI₂)**, a vasodilator and inhibitor of platelet aggregation. These two substances act in opposition, maintaining a balance in the cardiovascular system.

a. NSAIDs Inhibit PGH Synthase

Nonsteroidal anti-inflammatory drugs (NSAIDs; Fig. 25-73) inhibit the synthesis of the prostaglandins, prostacyclins, and thromboxanes by inhibiting or inactivating the cyclooxygenase activity of PGHS. Aspirin (**acetylsalicylic acid**), for example, acetylates this enzyme: If [^{14}C -acetyl] acetylsalicylic acid is incubated with the enzyme, radioactivity becomes irreversibly associated with the inactive enzyme as Ser 530 becomes acetylated (Fig. 25-74). The X-ray structure of PGHS reveals that Ser 530, which is not implicated in catalysis, extends into the cyclooxygenase channel just below Tyr 385 such that its acetylation would block arachidonic acid's access to the active site (Fig. 25-72b). The structure of PGHS, which was crystallized with the NSAID **flurbiprofen** (Fig. 25-73), indicates that this drug binds in the

cyclooxygenase channel, with its carboxyl group forming an ion pair with Arg 120 (Fig. 25-72a). Evidently, flurbiprofen, and by implication other NSAIDs, inhibits the cyclooxygenase activity of PGHS by blocking its active site channel.

Low doses of aspirin, ~ 80 mg (baby aspirin) every day, significantly reduce the long-term incidence of heart attacks and strokes. Such low doses selectively inhibit platelet aggregation and thus blood clot formation because these enucleated cells, which have a lifetime in the circulation of ~ 10 days, cannot resynthesize their inactivated enzymes. Vascular endothelial cells are not so drastically affected since, for the most part, they are far from the site where aspirin is absorbed, are exposed to lesser concentrations of aspirin and, in any case, can synthesize additional PGHS.

b. COX-2 Inhibitors Lack the Side Effects of Other NSAIDs

PGHS has two isoforms, **COX-1** and **COX-2**, that share a high degree (60%) of sequence identity and structural homology. COX-1 is constitutively (without regulation) expressed in most, if not all, mammalian tissues, thereby supporting levels of prostaglandin synthesis necessary to maintain organ and tissue homeostasis such as that of the gastrointestinal mucosa. In contrast, COX-2 is only expressed in certain tissues in response to inflammatory stim-

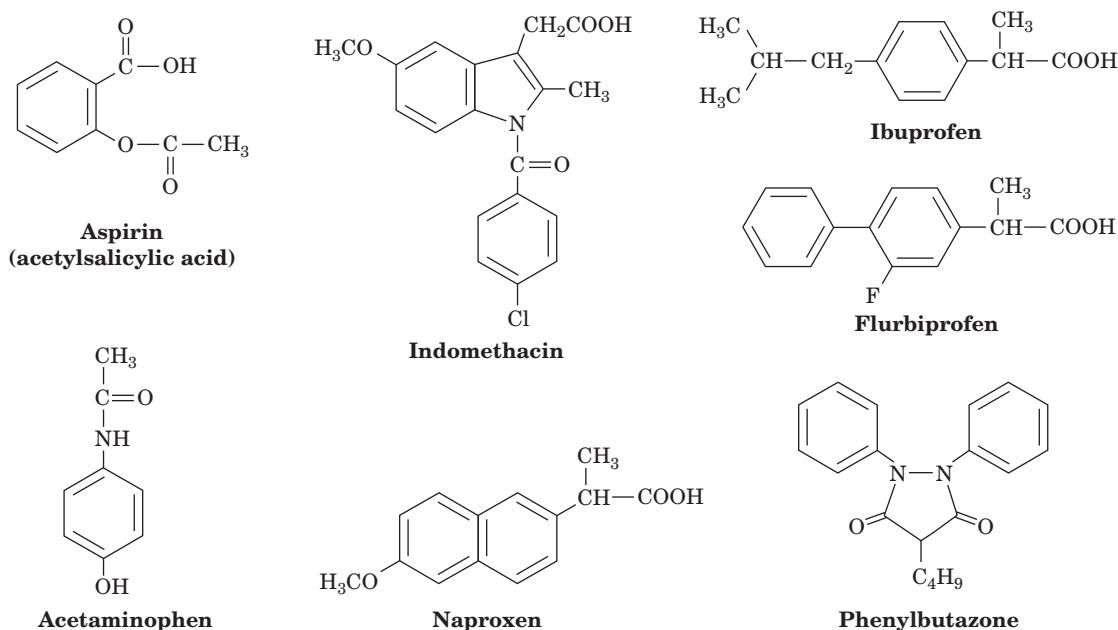


Figure 25-73 Some nonsteroidal anti-inflammatory drugs (NSAIDs).

uli such as cytokines, protein growth factors, and endotoxins, and hence is responsible for the elevated prostaglandin levels that cause inflammation. The NSAIDs in Fig. 25-73 are relatively nonspecific and therefore can have adverse side effects, most notably gastrointestinal ulceration, when used to treat inflammation or fever. A structure-based drug design program (Section 15-4Ad) was therefore instituted to create inhibitors that would target COX-2 but not COX-1. The three-dimensional structures of COX-1 and COX-2 are almost identical. However, their amino acid differences, specifically I523V, I434V, and H513R (COX-1 amino acid on the left and COX-2 amino acid on the right), make COX-2's active site channel ~20% larger in volume

than that of COX-1. In addition, the fourth helix of the membrane-binding domain is oriented slightly differently so as to provide a larger opening to the channel. Medicinal chemists therefore synthesized inhibitors, collectively known as **coxibs**, that could enter the COX-2 channel but are excluded from that of COX-1. Two of these inhibitors, **rofecoxib (Vioxx)** and **celecoxib (Celebrex)** (Fig. 25-75), became major drugs for the treatment of inflammatory diseases such as arthritis because they lack the major side

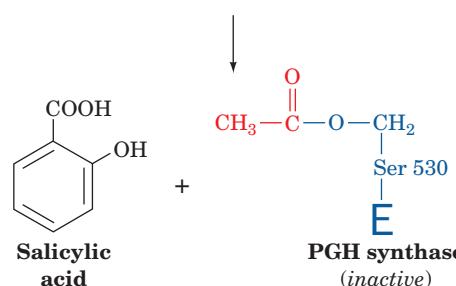
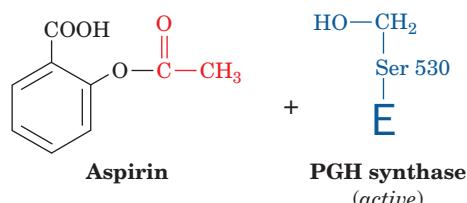
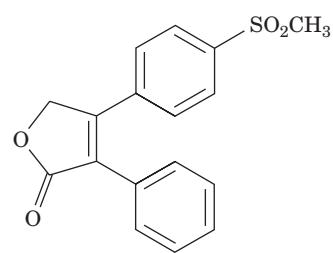
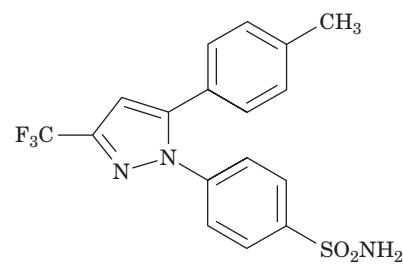


Figure 25-74 Inactivation of PGH synthase by aspirin. Aspirin acetylates Ser 530 of PGH synthase, thereby blocking the enzyme's cyclooxygenase activity.



Rofecoxib (Vioxx)



Celecoxib (Celebrex)

Figure 25-75 COX-2 inhibitors. Rofecoxib and celecoxib are specific inhibitors of COX-2 (PGH synthase-2).

effects of the nonspecific NSAIDs. However, in 2004, Vioxx was withdrawn from the pharmaceutical market because of unanticipated cardiac side effects arising from its attenuation of PGI₂ formation.

c. COX-3 May Be the Target of Acetaminophen

Acetaminophen, which is among the most widely used analgesic/antipyretic drugs (but possesses little anti-inflammatory activity, so that it is not really an NSAID), does not significantly bind to either COX-1 or COX-2. Thus its mechanism of action remained a mystery until the discovery by Daniel Simmons of a third COX isozyme, **COX-3**, a splice variant of COX-1, that is selectively inhibited by acetaminophen as well as by certain NSAIDs. This suggests that COX-3 is the primary target of drugs that decrease pain and fever.

C. The Linear Pathway of Eicosanoid Metabolism: Leukotrienes and Lipoxins

Arachidonic acid can be converted by a linear pathway to several different **hydroperoxyeicosatetraenoic acids (HPETEs)** by the **5-, 12-, and 15-lipoxygenases (5-, 12-, and 15-LOs)** (Fig. 25-69). **Hepoxilins** are hydroxy epoxy derivatives of **12-HPETE** whose functions are not as yet well understood. **Lipoxins**, the products of a second lipoxygenase acting on **15-HPETE**, are anti-inflammatory substances. Leukotrienes, derived from the 5-LO reaction, are synthesized by a variety of white blood cells, mast cells (connective tissue cells derived from the blood-forming tissues that secrete substances which mediate inflammatory and allergic reactions), as well as lung, spleen, brain, and heart. **Peptidoleukotrienes (LTC₄, LTD₄, and LTE₄)** are now recognized to be the components of the **slow reacting substances of anaphylaxis (SRS-A)**; anaphylaxis is a violent and potentially fatal allergic reaction released from sensitized lung after immunological challenge. These substances act at very low concentrations (as little as 10⁻¹⁰ M) to contract vascular, respiratory, and intestinal smooth muscle. Peptidoleukotrienes, for example, are ~10,000-fold more potent than histamine, a well-known stimulant of allergic reactions. In the respiratory system, they constrict bronchi, especially the smaller airways; increase mucus secretion; and are thought to be the mediators in asthma. They are also implicated in immediate hypersensitivity (allergic) reactions, inflammatory reactions, and heart attacks.

a. Leukotriene Synthesis

The first two reactions in the conversion of arachidonic acid to leukotrienes are both catalyzed by 5-LO, which contains a nonheme, non-[Fe–S] cluster iron atom that must be in its Fe(III) state to be active. These reactions occur as follows (Fig. 25-76):

1. The oxidation of arachidonic acid to form 5-HPETE, a substance that, in itself, is not a physiological mediator. This reaction occurs in three steps:

(a) The active site iron atom, in its active Fe(III) state,

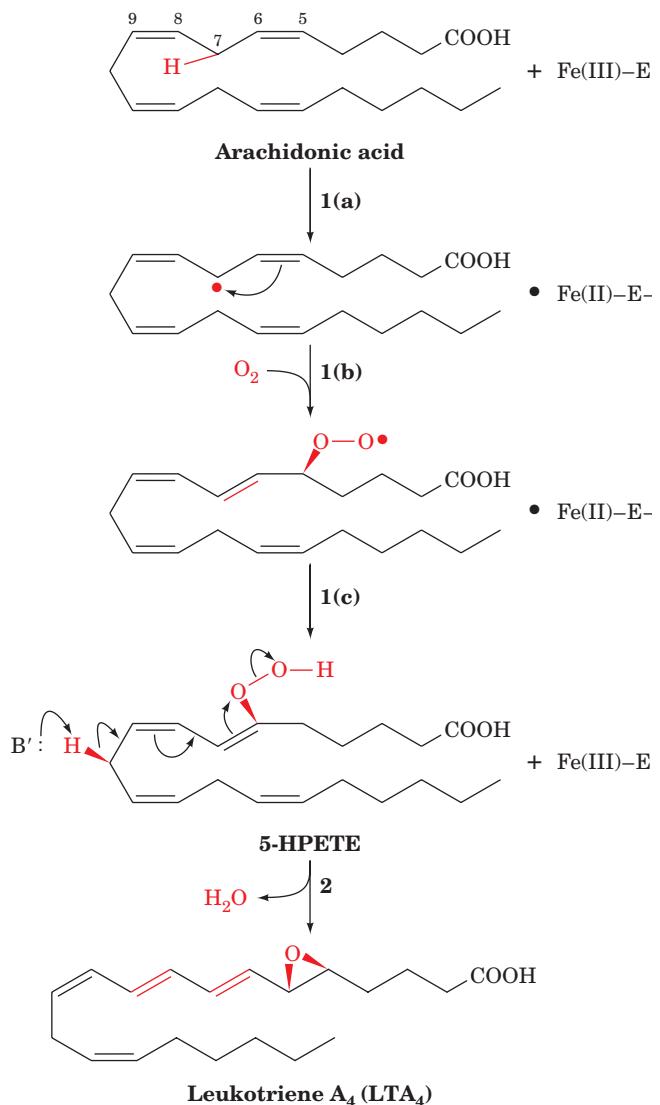


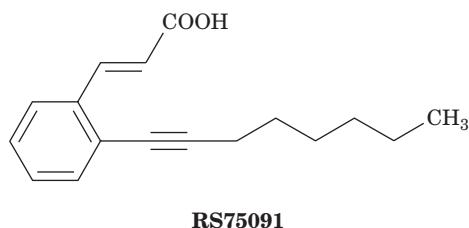
Figure 25-76 The 5-LO-catalyzed oxidation of arachidonic acid to LTA₄ via the intermediate 5-HPETE.

abstracts an electron from the central methylene group of the 5,8-pentadiene moiety of arachidonate and the resulting free radical loses a proton to an enzymatic base.

- (b) The free radical rearranges and adds O₂ to form a hydroperoxide radical.
- (c) The hydroperoxide radical reacts with the active site iron, now in its Fe(II) form, to yield the hydroperoxide in its anionic form, which the enzyme then protonates to yield the hydroperoxide product, regenerating the active Fe(III) enzyme.

- 2. The base-catalyzed elimination of water to form the unstable epoxide **leukotriene A₄ (LTA₄)** (the subscript indicates the number of carbon–carbon double bonds in the molecule, which is also its series number).

The X-ray structure of the rabbit reticulocyte 15-LO, a homolog of 5-LO, in complex with the competitive inhibitor **RS75091**



was determined by Michelle Browner. This 663-residue monomeric protein consists of an N-terminal 8-stranded β barrel domain and a C-terminal catalytic domain (Fig. 25-77). Its active site Fe atom is coordinated by four invariant His residues and by a C-terminal carboxylate oxygen in a liganding arrangement that is best described as a distorted octahedron with one of its six vertices unoccupied. The Fe, which is well below the protein surface, faces an internal cavity occupied by RS75091. This identifies the substrate binding cavity, which is lined with mostly hydrophobic residues and follows an irregular pathway past the Fe atom to the protein surface. Intriguingly, 15-LO (as does soybean **lipoxygenase-1**) contains two rarely observed π helices (Fig. 8-14c), each of which contains two of the Fe-liganding His residues. Each of these π helices is embedded in a longer helix rather than being at the end of an α helix as is the case for all previously observed π helices.

The sizes of the substrate-binding cavities of the 5- and 12-LOs have been predicted through their homology modeling (Section 9-3B) with 15-LO. 5-LO and 12-LO have smaller amino acids substituted for those in 15-LO, such that, for example, 5-LO is predicted to have a cavity with $\sim 20\%$ greater volume than that of 15-LO. The mutagenesis of 5-LO by Harmut Kuhn so as to decrease the size of its cavity yielded an enzyme with the specificity of 15-LO, thus supporting the proposal that it is the size of the cavity that determines lipoxygenase specificity.

b. Peptidoleukotrienes

LTA₄ is converted to peptidoleukotrienes by reaction with **LTC₄ synthase**, a **glutathione-S-transferase** that catalyzes the addition of the glutathione sulfhydryl group to the LTA₄ epoxide, forming the first of the peptidoleukotrienes, **leukotriene C₄ (LTC₄)**, Fig. 25-78). **γ -Glutamyltransferase** removes glutamic acid, converting LTC₄ to **leukotriene D₄ (LTD₄)**. LTD₄ is converted to **leukotriene E₄ (LTE₄)** by a dipeptidase that removes glycine. LTA₄ can also be hydrolyzed to **leukotriene B₄ (LTB₄)**, a potent chemotactic agent (a substance that attracts motile cells) involved in attracting certain types of white blood cells to fight infection.

Various inflammatory and hypersensitivity disorders (such as asthma) are associated with elevated levels of leukotrienes. The development of drugs that inhibit leukotriene synthesis has therefore been an active field of

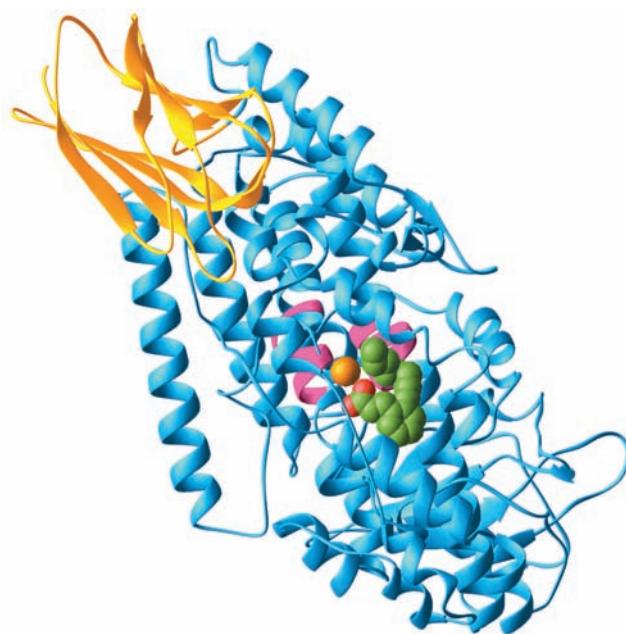
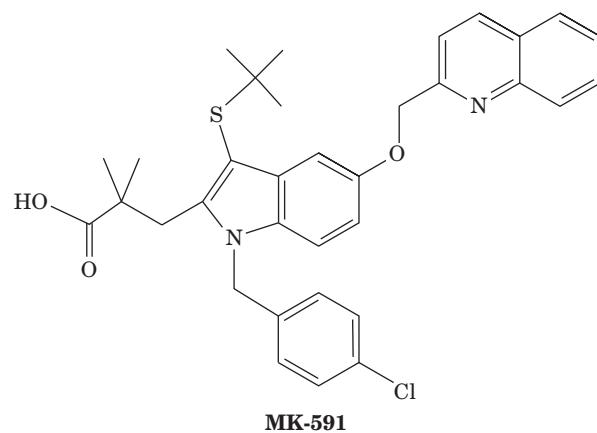
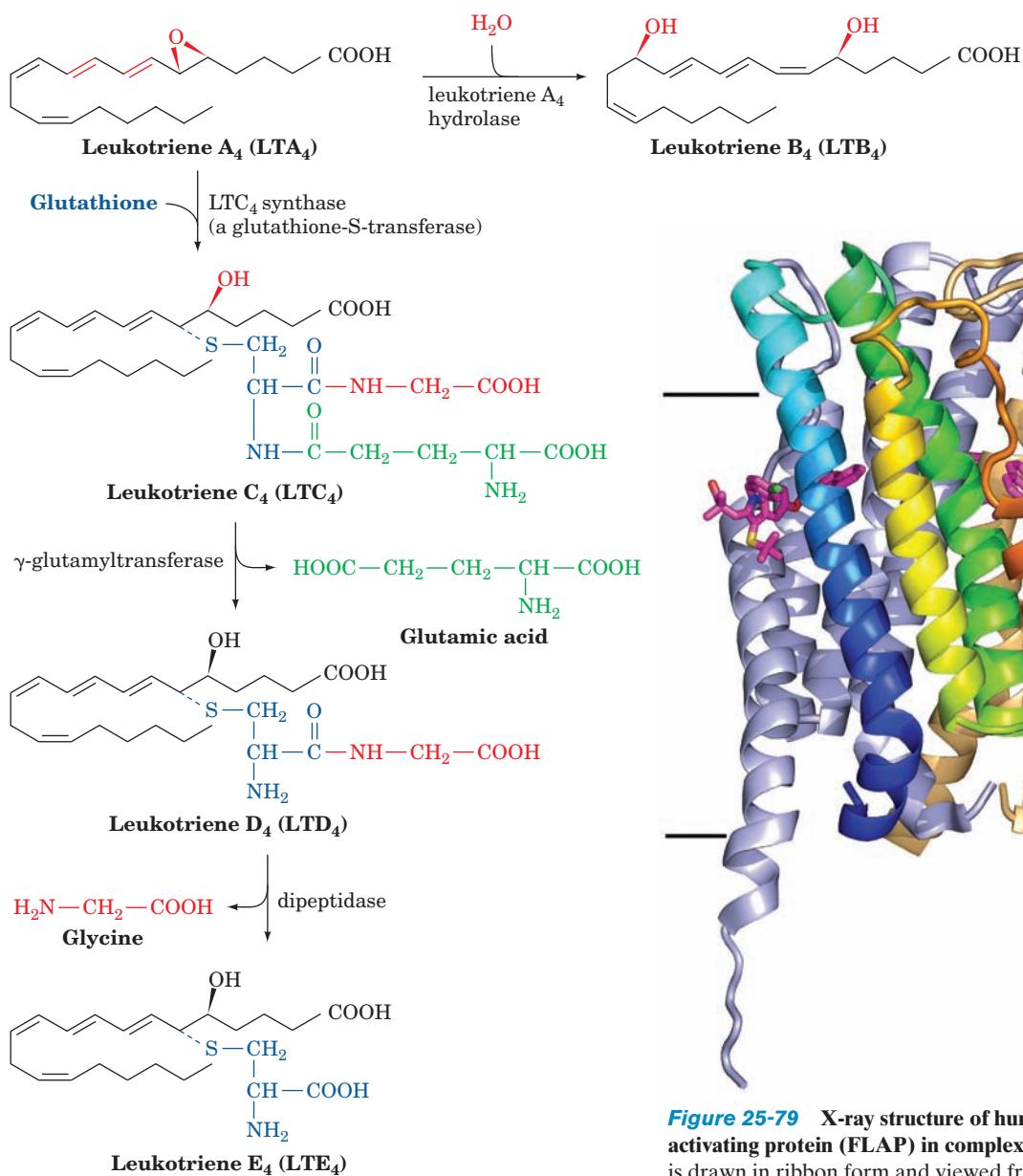


Figure 25-77 X-ray structure of rabbit reticulocyte 15-lipoxygenase (15-LO) in complex with its competitive inhibitor **RS75091**. The N-terminal β barrel domain is gold and the C-terminal catalytic domain is light blue with its two Fe-liganding π helical segments magenta. The Fe is represented by an orange sphere and the RS75091 is drawn in space-filling form with C green and O red. [Based on an X-ray structure by Michelle Browner, Roche Bioscience, Palo Alto, California. PDBid 1LOX.]

research. 5-LO activity requires the presence of **5-lipoxygenase-activating protein (FLAP)**, a homotrimeric integral membrane protein of 161-residue subunits that is located in both the nuclear and ER membranes. FLAP, which has no enzymatic activity, binds the arachidonic acid substrate of 5-LO and facilitates enzyme–substrate binding as well as 5-LO’s interaction with the membrane. Several inhibitors of leukotriene synthesis, such as **MK-591**,



bind to FLAP so as to inhibit both of its functions.

Figure 25-78 Formation of the leukotrienes from LTA₄.

The X-ray structure of human FLAP in complex with MK-591 (Fig. 25-79), determined by Joseph Becker, reveals that each subunit forms an up-down-up-down four-helix bundle whose N- and C-termini are on the luminal side of the membrane, and that its three identical subunits have extensive intersubunit contacts, thereby forming a 12-helix bundle. MK-591 binds in lipid-exposed grooves between pairs of subunits. Moreover, the trimer forms an elongated pocket along its 3-fold axis that is open to the lumen. This, together with biochemical data, suggests that one 5-LO molecule binds to the cytosolic surface of a FLAP trimer where, through conformational changes in FLAP, it accepts arachidonic acid molecules that FLAP has extracted from the membrane. Inhibitors such as MK-591 presumably block this extraction process.

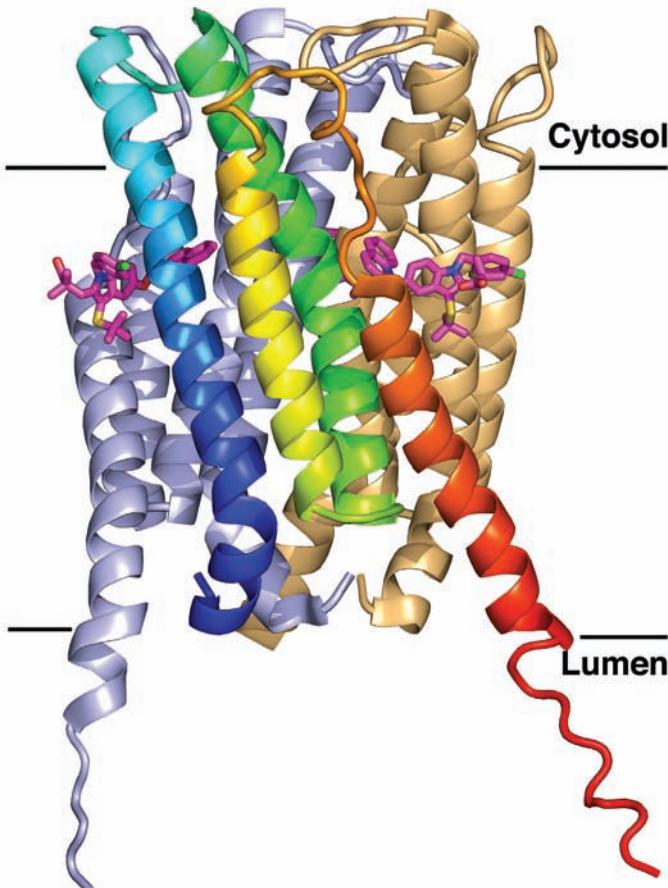


Figure 25-79 X-ray structure of human 5-lipoxygenase-activating protein (FLAP) in complex with MK-591. The protein is drawn in ribbon form and viewed from within the membrane with its approximate molecular 3-fold axis vertical. The subunit closest to the viewer of this homotrimeric protein is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The other subunits are pale blue and pale orange. The bound MK-591 is drawn in stick form with C magenta, N blue, O red, S yellow, and C green. The inferred position of the membrane is indicated by the horizontal black lines. [Based on an X-ray structure by Joseph Becker, Merck Research Laboratories, Rahway, New Jersey. PDBid 2Q7M.]

c. Diets Rich in Marine Lipids May Decrease Cholesterol, Prostaglandin, and Leukotriene Levels

Greenland Eskimos have a very low incidence of coronary heart disease and thrombosis despite their high dietary intake of cholesterol and fat. Their consumption of marine animals provides them with a higher proportion of unsaturated fats than the typical American diet. A major unsaturated component of marine lipids is 5,8,11,14,17-

eicosapentaenoic acid (EPA; Fig. 25-67), an ω -3 fatty acid, rather than the arachidonic acid precursor linoleic acid, an ω -6 fatty acid. EPA inhibits formation of TxA_2 (Fig. 25-70) and is a precursor of the **series-5 leukotrienes**, compounds with substantially lower physiological activities than their arachidonate-derived (series-4) counterparts. This suggests that a diet containing marine lipids should decrease the extent of prostaglandin- and leukotriene-mediated inflammatory responses. Indeed, dietary enrichment with EPA inhibits the *in vitro* chemotactic and aggregating activities of neutrophils (a type of white blood cell). Moreover, an EPA-rich diet decreases the cholesterol and triacylglycerol levels in the plasma of hypertriacylglycerolemic patients.

d. Lipoxins and Aspirin-Induced *epi*-Lipoxins Have Anti-Inflammatory Properties

Eicosanoids are usually associated with the inflammatory response. However, some eicosanoids have anti-inflammatory

properties. The **lipoxins (LXs)**, products of the 12- and 15-LO pathways (sometimes also involving 5-LO), are so named because they are synthesized through *lipoxygenase* interactions. Their activities appear to inhibit those of leukotrienes and hence they are anti-inflammatory. There are many ways for LXs to be synthesized by combinations of the actions of 5-, 12-, and 15-LOs. Here we discuss only one such route (Fig. 25-80, *left*). **Lipoxin A₄ (LXA₄)** synthesis from arachidonic acid begins in endothelial and epithelial cells by the 15-LO-catalyzed synthesis of **(15S)-hydroperoxyeicosatetraenoic acid [(15S)-HPETE]**, which is reduced by **glutathione peroxidase** to **(15S)-hydroxyeicosatetraenoic acid [(15S)-HETE]**. The (15S)-HETE then makes its way to leukocytes where it is converted by 5-LO and a hydrolase.

Charles Serhan discovered an additional pathway for the anti-inflammatory action of aspirin that also involves lipoxin production. As we have previously seen (Fig. 25-74), aspirin covalently inhibits the cyclooxygenase activity of

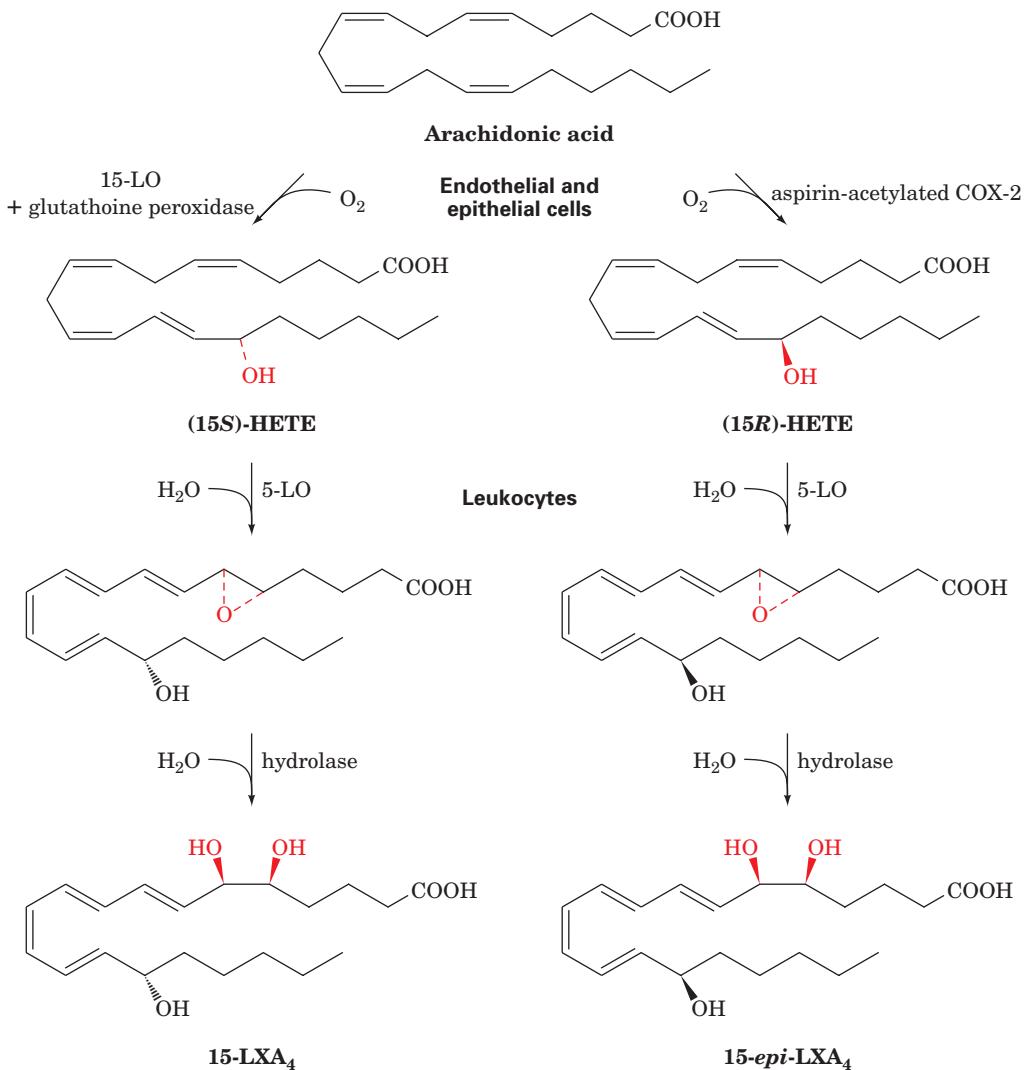


Figure 25-80 Lipoxin biosynthesis. The biosynthesis of the lipoxin LXA₄ (*left*) and the aspirin-triggered *epi*-lipoxin (ATL) 15-*epi*-LXA₄ (*right*). In endothelial and epithelial cells, arachidonic acid is converted by 15-LO and glutathione

peroxidase to (15S)-HETE or by aspirin-acetylated COX-2 to (15R)-HETE. After transfer to leukocytes, 5-LO and a hydrolase convert these intermediate products to LXA₄ and 15-*epi*-LXA₄.

PGHS (COX). However, aspirin-acetylated COX-2 retains a residual 15-LO activity (Steps 3 and 4 in Fig. 25-71) through which it initiates a pathway that converts arachidonic acid to the anti-inflammatory agents called **aspirin-triggered epi-lipoxins (ATLs; Fig. 25-80, right)**. This pathway begins in endothelial and epithelial cells with the aspirin-acetylated COX-2-catalyzed conversion of arachidonic acid to **(15R)-hydroxyeicosatetraenoic acid [(15R)-HETE]**, the epimer of (15S)-HETE. In leukocytes, 5-LO and a hydrolase then convert (15R)-HETE to the anti-inflammatory agent **15-epi-lipoxin A₄ (15-epi-LXA₄)**.

These are indeed exciting times in the study of eicosanoid metabolism and its physiological manifestations. As the mechanisms of action of the prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins are becoming better understood, they are providing the insights required for the development of new and improved therapeutic agents.

8 PHOSPHOLIPID AND GLYCOLIPID METABOLISM

The “complex lipids” are dual-tailed amphipathic molecules composed of either 1,2-diacyl-sn-glycerol or N-acylsphingosine (ceramide) linked to a polar head group that is either a carbohydrate or a phosphate ester (Fig. 25-81; Sections 12-1C and 12-1D; *sn* stands for stereospecific numbering, which assigns the 1 position to the group occupying the *pro-S* position of a prochiral center). Hence, there are two categories of phospholipids, **glycerophospholipids** and **sphingophospholipids**, and two categories of glycolipids, **glyceroglycolipids** and **sphingoglycolipids** (also called **glycosphingolipids; GSLs**). In this section we describe the biosynthesis of the complex lipids from their simpler components. We shall see that the great variety of these substances is matched by the numerous enzymes required for their specific syntheses. Note also that these substances are synthesized in membranes, mostly on the cytosolic face of the endoplasmic reticulum, and from there are transported to their final cellular destinations as indicated in Sections 12-4B–D.

A. Glycerophospholipids

Glycerophospholipids have significant asymmetry in their C1- and C2-linked fatty acyl groups: C1 substituents are

mostly saturated fatty acids, whereas those at C2 are by and large unsaturated fatty acids. We shall examine the major pathways of biosynthesis and metabolism of the glycerophospholipids with an eye toward understanding the origin of this asymmetry.

a. Biosynthesis of Diacylglycerophospholipids

The triacylglycerol precursors 1,2-diacyl-sn-glycerol and phosphatidic acid are also the precursors of certain glycerophospholipids (Figs. 25-42 and 25-81). Activated phosphate esters of the polar head groups (Table 12-2) react with the C3 OH group of 1,2-diacyl-sn-glycerol to form the phospholipid’s phosphodiester bond. In some cases the phosphoryl group of phosphatidic acid is activated and reacts with the unactivated polar head group.

The mechanism of activated phosphate ester formation is the same for both the polar head groups **ethanolamine** and **choline** (Fig. 25-82):

1. ATP first phosphorylates the OH group of choline or ethanolamine.

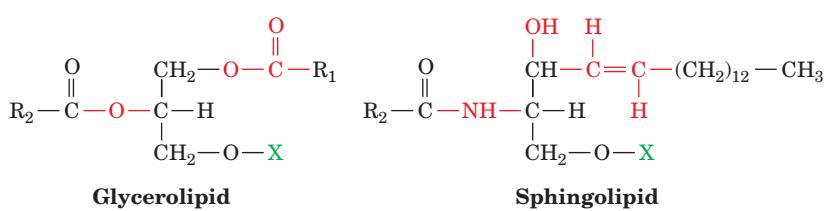
2. The phosphoryl group of the resulting **phosphoethanolamine** or **phosphocholine** then attacks CTP, displacing PP_i, to form the corresponding CDP derivatives, which are activated phosphate esters of the polar head group.

3. The C3 OH group of 1,2-diacyl-sn-glycerol attacks the phosphoryl group of the activated CDP–ethanolamine or CDP–choline, displacing CMP to yield the corresponding glycerophospholipid.

The liver also converts phosphatidylethanolamine to phosphatidylcholine by trimethylating its amino group, using **S-adenosylmethionine** (Section 26-3Ea) as the methyl donor.

Phosphatidylserine is synthesized from phosphatidylethanolamine by a head group exchange reaction catalyzed by **phosphatidylethanolamine:serine transferase** in which serine’s OH group attacks the donor’s phosphoryl group (Fig. 25-83). The original head group is then eliminated, forming phosphatidylserine.

In the synthesis of **phosphatidylinositol** and **phosphatidylglycerol**, the hydrophobic tail is activated rather than the polar head group. Phosphatidic acid, the precursor of 1,2-diacyl-sn-glycerol (Fig. 25-42), attacks the α -phosphoryl group of CTP to form the activated **CDP-**



X = H
X = Carbohydrate
X = Phosphate ester

1,2-Diacylglycerol
Glyceroglycolipid
Glycerophospholipid

N-Acylsphingosine (ceramide)
Sphingoglycolipid (glycosphingolipid)
Sphingophospholipid

Figure 25-81 The glycerolipids and sphingolipids. The structures of the common head groups, X, are presented in Table 12-2.

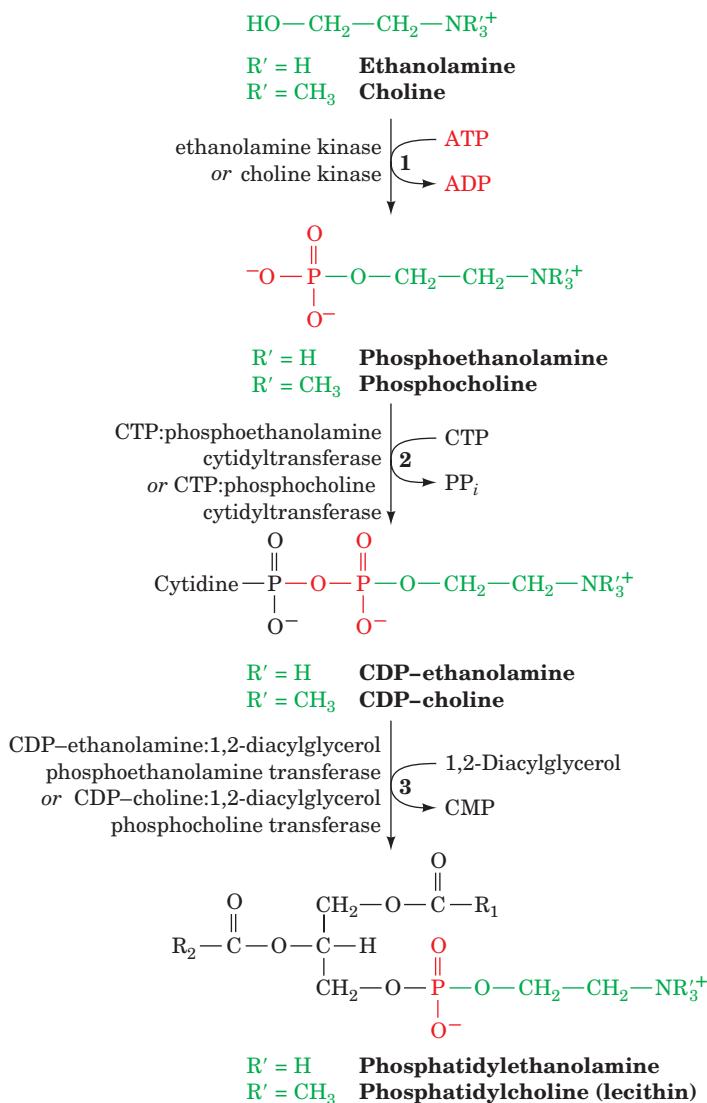


Figure 25-82 The biosynthesis of phosphatidylethanolamine and phosphatidylcholine. In mammals, CDP-ethanolamine and CDP-choline are the precursors of the head groups.

diacylglycerol and PP_i (Fig. 25-84). Phosphatidylinositol results from the attack of inositol on CDP-diacylglycerol. Phosphatidylglycerol is formed in two reactions: (1) attack of the C1 —OH group of *sn*-glycerol-3-phosphate on CDP-diacylglycerol, yielding **phosphatidylglycerol phosphate**; and (2) hydrolysis of the phosphoryl group to form phosphatidylglycerol.

Cardiolipin, an important phospholipid first isolated from heart tissue, is synthesized from two molecules of phosphatidylglycerol (Fig. 25-85). The reaction occurs by the attack of the C1 OH group of one of the phosphatidylglycerol molecules on the phosphoryl group of the other, displacing a molecule of glycerol.

Enzymes that synthesize phosphatidic acid have a general preference for saturated fatty acids at C1 and for unsaturated fatty acids at C2. Yet, this general preference cannot account, for example, for the observations that ~80%

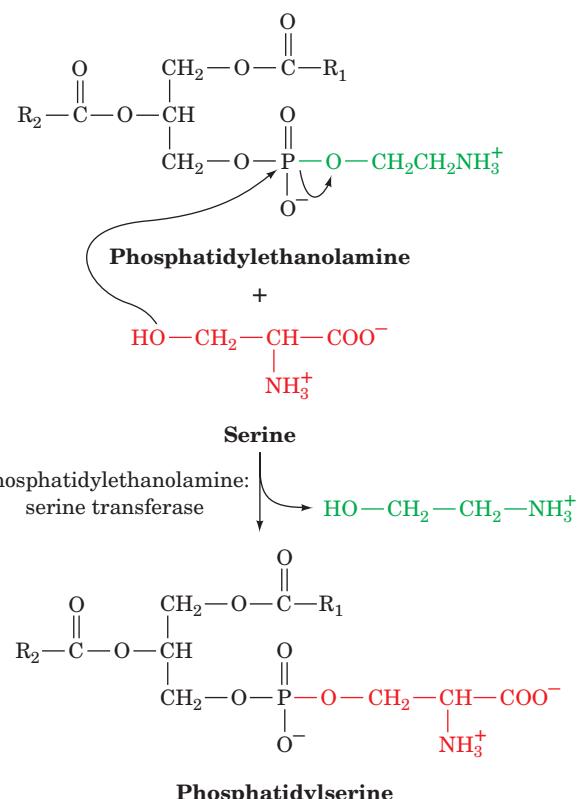


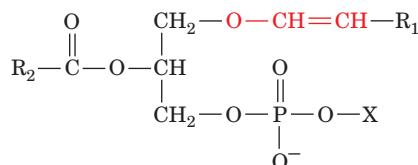
Figure 25-83 Phosphatidylserine synthesis. Serine replaces ethanolamine in phosphatidylethanolamine by a head group exchange reaction.

of brain phosphatidylinositol has a stearoyl group (18:0) at C1 and an arachidonoyl group (20:4) at C2, and that ~40% of lung phosphatidylcholine has palmitoyl groups (16:0) at both positions (this latter substance is the major component of the surfactant that prevents the lung from collapsing when air is expelled; its deficiency is responsible for **respiratory distress syndrome** in premature infants). William Lands showed that such side chain specificity results from “remodeling” reactions in which specific acyl groups of individual glycerophospholipids are exchanged by specific phospholipases and acyltransferases.

b. Biosynthesis of Plasmalogens and Alkylacylglycerophospholipids

Eukaryotic membranes contain significant amounts of two other types of glycerophospholipids:

1. Plasmalogens, which contain a hydrocarbon chain linked to glycerol C1 via a vinyl ether linkage:



A plasmalogen

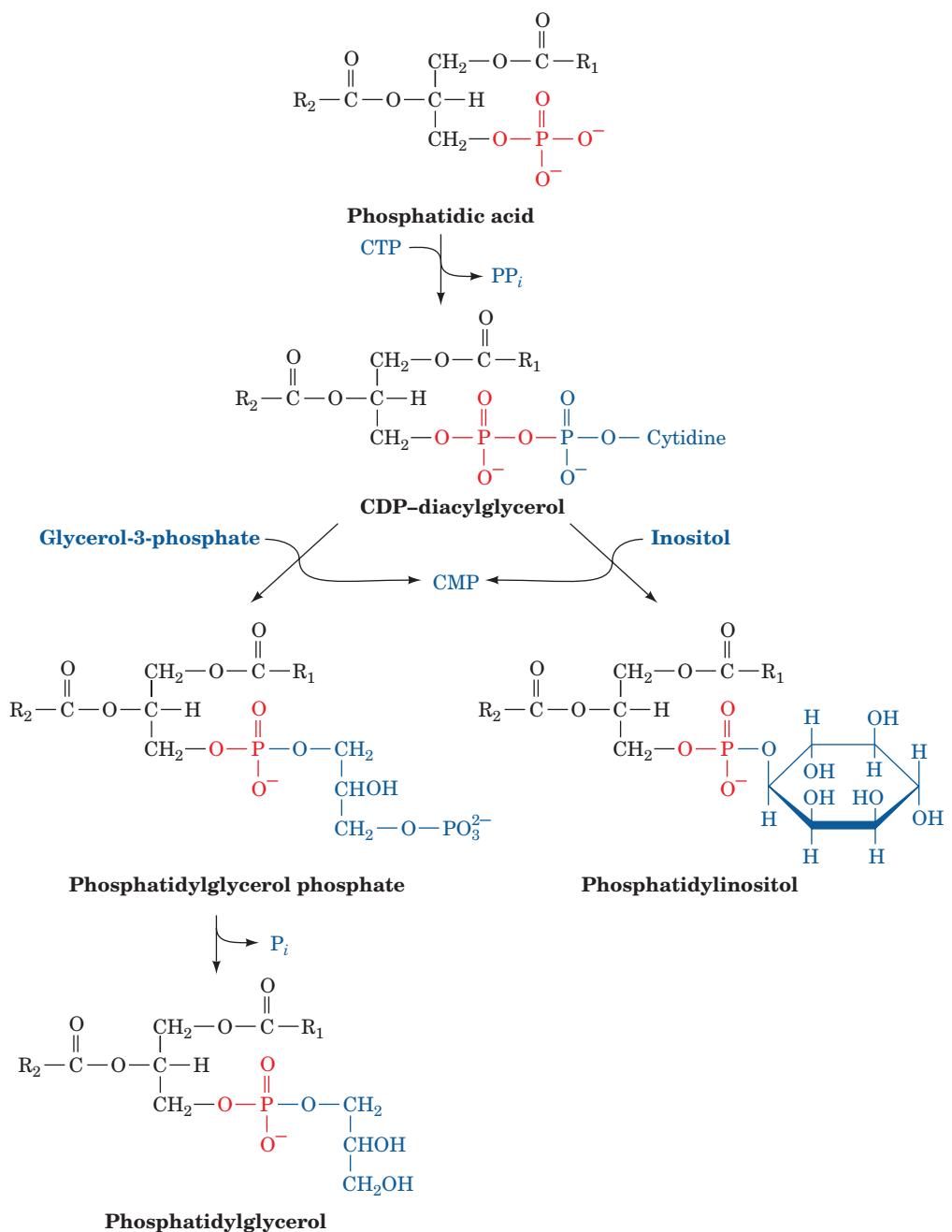


Figure 25-84 The biosynthesis of phosphatidylinositol and phosphatidylglycerol. In mammals, this process involves a CDP-diacylglycerol intermediate.

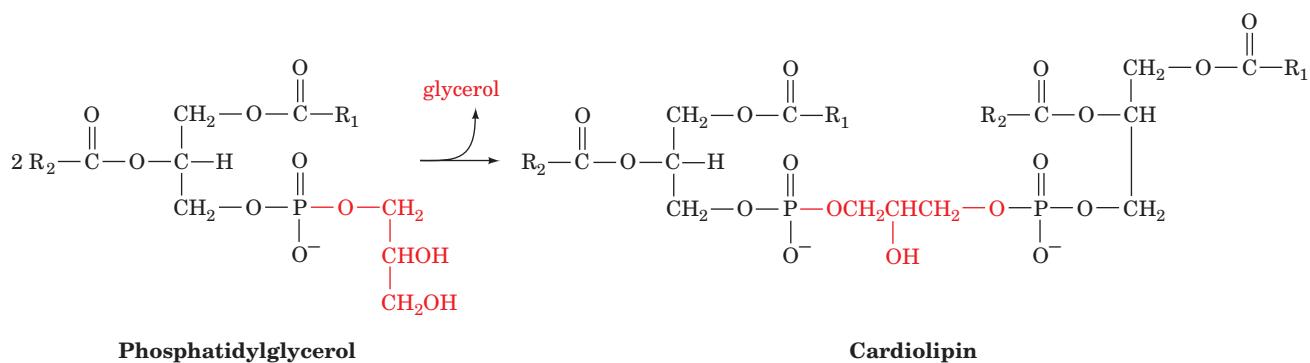


Figure 25-85 The formation of cardiolipin.

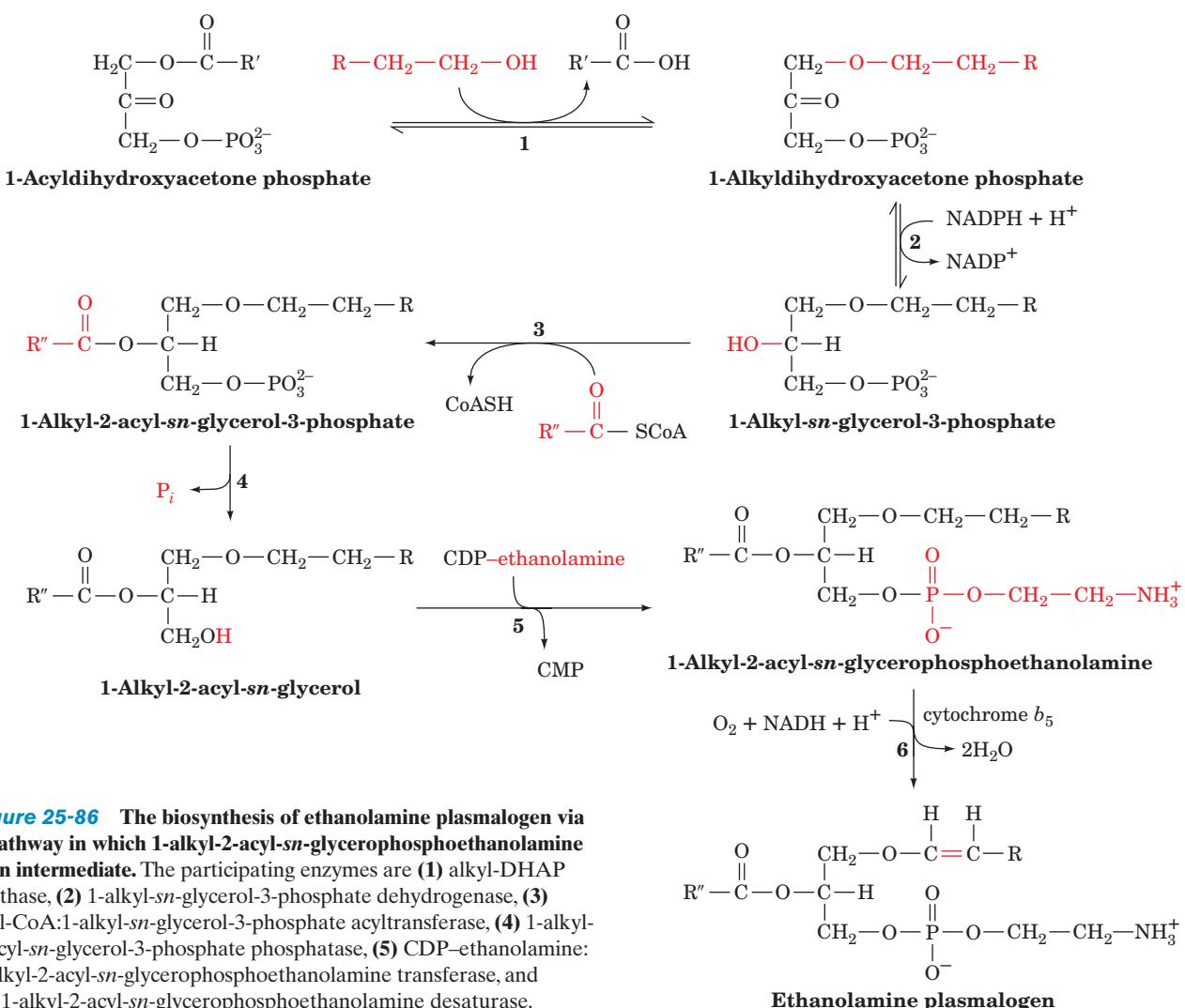
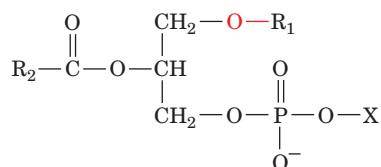


Figure 25-86 The biosynthesis of ethanolamine plasmalogen via a pathway in which 1-alkyl-2-acyl-sn-glycerophosphoethanolamine is an intermediate. The participating enzymes are (1) alkyl-DHAP synthase, (2) 1-alkyl-sn-glycerol-3-phosphate dehydrogenase, (3) acyl-CoA:1-alkyl-sn-glycerol-3-phosphate acyltransferase, (4) 1-alkyl-2-acyl-sn-glycerol-3-phosphate phosphatase, (5) CDP-ethanolamine:1-alkyl-2-acyl-sn-glycerophosphoethanolamine transferase, and (6) 1-alkyl-2-acyl-sn-glycerophosphoethanolamine desaturase.

2. Alkylacylglycerophospholipids, in which the alkyl substituent at glycerol C1 is attached via an ether linkage:



An alkylacylglycerophospholipid

About 20% of mammalian glycerophospholipids are plasmalogens. The exact percentage varies both from species to species and from tissue to tissue within a given organism. While plasmalogens comprise only 0.8% of the phospholipids in human liver, they account for 23% of those in human nervous tissue. The alkylacylglycerophospholipids are less abundant than the plasmalogens; for instance, 59% of the ethanolamine glycerophospholipids of human heart are plasmalogens, whereas only 3.6% are alkylacylglycerophospholipids. However, in bovine erythrocytes, 75% of the ethanolamine glycerophospholipids are of the alkylacyl type.

The pathway forming ethanolamine plasmalogens and alkylacylglycerophospholipids involves several reactions (Fig. 25-86):

1. Exchange of the acyl group of **1-acyldihydroxyacetone phosphate** for an alcohol.
2. Reduction of the ketone to **1-alkyl-sn-glycerol-3-phosphate**.
3. Acylation of the resulting C2 OH group by acyl-CoA.
4. Hydrolysis of the phosphoryl group to yield an alkylacylglycerol.
5. Attack by the new OH group of alkylacylglycerol on CDP-ethanolamine to yield **1-alkyl-2-acyl-sn-glycerophosphoethanolamine**.
6. Introduction of a double bond into the alkyl group to form the plasmalogen by a desaturase having the same cofactor requirements as the fatty acid desaturases (Section 25-4E).

Recall that the precursor–product relationship between the alkylacylglycerophospholipid and the plasmalogen was

established through studies using [¹⁴C]ethanolamine (Section 16-3Be).

The alkylacylglycerophospholipid with an acetyl group at R₂ and a choline polar head group (X), **1-O-hexadecyl-2-acetyl-sn-glycero-3-phosphocholine**, is known as **platelet-activating factor (PAF)**. This molecule has diverse functions and acts at very low concentrations ($10^{-10} M$) to lower blood pressure and to cause blood platelets to aggregate.

B. *Sphingophospholipids*

Only one major phospholipid contains ceramide (*N*-acylsphingosine) as its hydrophobic tail: **sphingomyelin** (***N*-acylsphingosine phosphocholine**; Section 12-1D), an important structural lipid of nerve cell membranes. The molecule was once thought to be synthesized from *N*-acylsphingosine and CDP-choline. However, it is now known that the main route of sphingomyelin synthesis occurs through donation of the phosphocholine group of phosphatidylcholine to *N*-acylsphingosine (Fig. 25-87). These pathways were differentiated by establishing the precursor-product relationships between CDP-choline, phosphatidylcholine, and sphingomyelin (Section 16-3Be). Mouse liver microsomes were isolated and incubated for a short time with [³H]choline. Radioactivity appeared in sphingomyelin only after first appearing in both CDP-choline and phosphatidylcholine, ruling out the direct transfer of phosphocholine from CDP-choline to *N*-acylsphingosine.

The most prevalent acyl groups of sphingomyelin are palmitoyl (16:0) and stearoyl (18:0) groups. Longer chain fatty acids such as nervonic acid (24:1) and behenic acid (22:0) occur with lesser frequency in sphingomyelins.

C. *Sphingoglycolipids*

Most sphingolipids are sphingoglycolipids, that is, their polar head groups consist of carbohydrate units (Section 12-1D). The principal classes of sphingoglycolipids, as indicated in Fig. 25-88, are **cerebrosides** (ceramide monosaccharides), **sulfatides** (ceramide monosaccharide sulfates), **globosides** (neutral ceramide oligosaccharides), and **gangliosides** (acidic, sialic acid-containing ceramide oligosaccharides). The carbohydrate unit is glycosidically attached to the *N*-acylsphingosine at its C1 OH group (Fig. 25-81).

The lipids providing the carbohydrate that covers the external surfaces of eukaryotic cells are sphingoglycolipids. Along with glycoproteins (Section 23-3), they are biosynthesized on the luminal surfaces of the endoplasmic reticulum and the Golgi apparatus and reach the plasma membrane through vesicle flow (Sections 12-4C and 12-4D), where membrane fusion results in their facing the external surface of the lipid bilayer (Fig. 12-60). Degradation of sphingoglycolipids occurs in the lysosomes after endocytosis from the plasma membrane.

In the following subsections, we discuss the biosynthesis and breakdown of *N*-acylsphingosine and sphingoglycolipids and consider the diseases caused by deficiencies in their degradative enzymes.

a. Biosynthesis of Ceramide (*N*-Acylsphingosine)

Biosynthesis of *N*-acylsphingosine occurs in four reactions from the precursors palmitoyl-CoA and serine (Fig. 25-89):

1. 3-Ketosphinganine synthase (serine palmitoyltrans-

ferase), a pyridoxal phosphate-dependent enzyme, catalyzes the condensation of palmitoyl-CoA with serine yielding **3-ketosphinganine** (pyridoxal phosphate-dependent reactions are discussed in Section 26-1A).

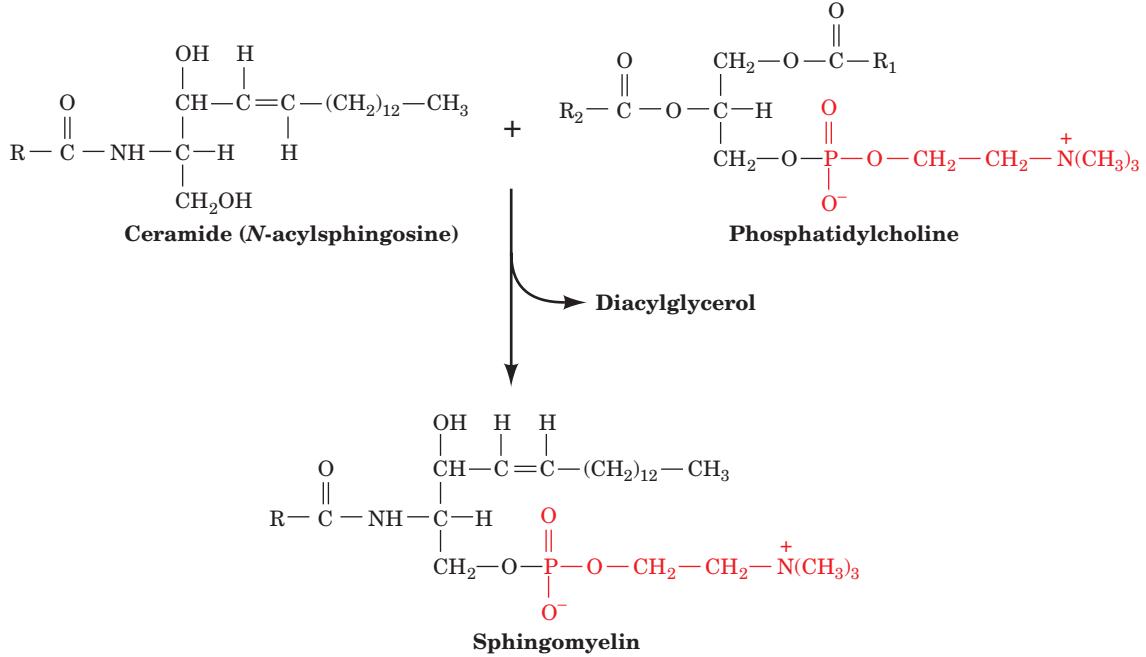


Figure 25-87 The synthesis of sphingomyelin from *N*-acylsphingosine and phosphatidylcholine.

used synonymously with galactocerebroside. Both are synthesized from ceramide by addition of a glycosyl unit from the corresponding UDP-hexose (Fig. 25-90). Galactocerebroside is a common component of brain lipids. Glucocerebroside, although relatively uncommon, is the precursor of globosides and gangliosides.

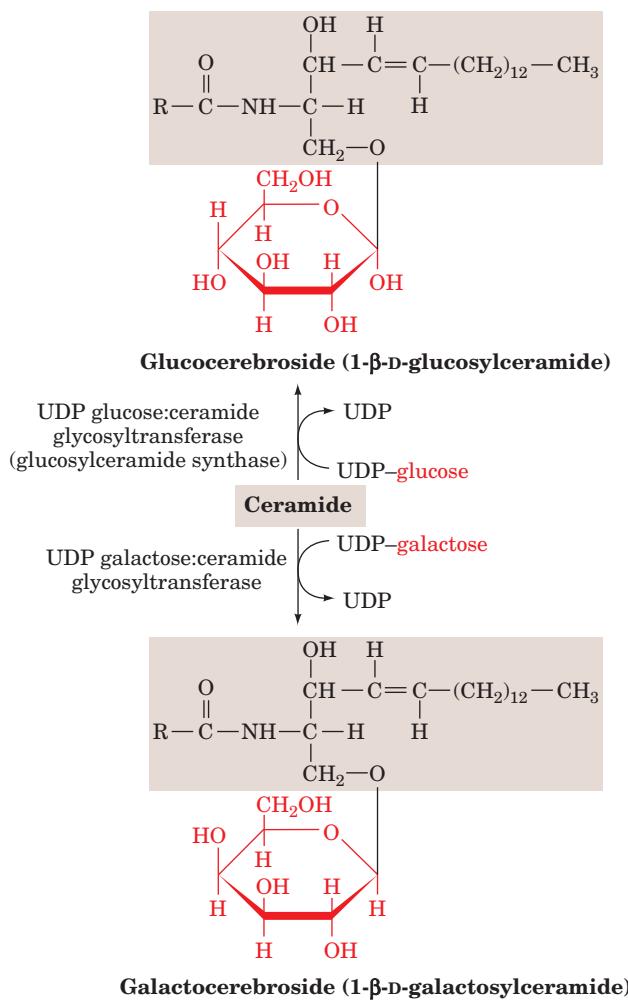


Figure 25-90 The biosynthesis of cerebrosides.

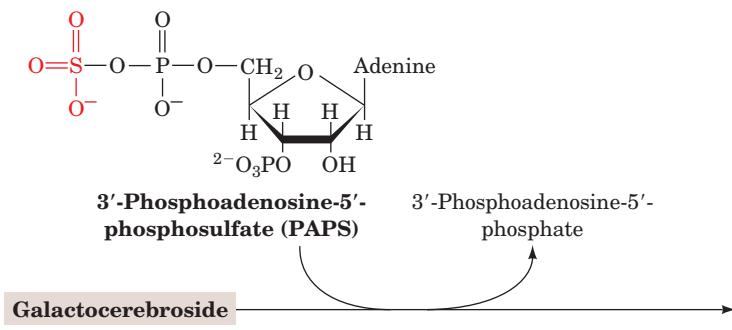


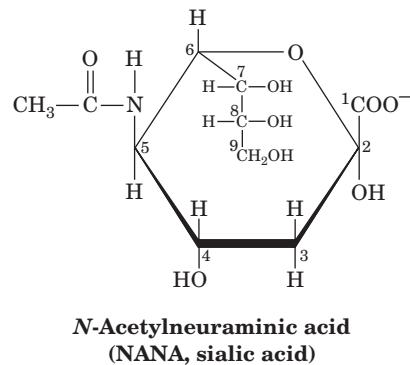
Figure 25-91 The biosynthesis of sulfatides.

c. Biosynthesis of Sulfatides

Sulfatides (galactocerebroside-3-sulfate) account for 15% of the lipids of white matter in the brain. They are formed by transfer of an “activated” sulfate group from **3'-phosphoadenosine-5'-phosphosulfate (PAPS)** to the C3 OH group of galactose in galactocerebroside (Fig. 25-91).

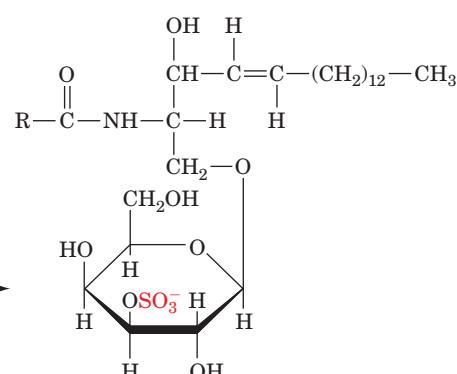
d. Biosynthesis of Globosides and Gangliosides

Biosynthesis of both globosides (neutral ceramide oligosaccharides) and gangliosides (acidic, sialic acid-containing ceramide oligosaccharides) is catalyzed by a series of **glycosyltransferases**. While the reactions are chemically similar, each is catalyzed by a specific enzyme. The pathways begin with transfer of a galactosyl unit from UDP-Gal to glucocerebroside to form a $\beta(1 \rightarrow 4)$ linkage (Fig. 25-92). Since this bond is the same as that linking glucose and galactose in lactose, this glycolipid is often referred to as **lactosyl ceramide**. Lactosyl ceramide is the precursor of both globosides and gangliosides. To form a globoside, one galactosyl and one *N*-acetylgalactosaminyl unit are sequentially added to lactosyl ceramide from UDP-Gal and UDP-GalNAc, respectively. The G_M gangliosides are formed by addition of ***N*-acetylneurameric acid (NANA, sialic acid)**



N-Acetylneurameric acid
(NANA, sialic acid)

from CMP-NANA to lactosyl ceramide in $\alpha(2 \rightarrow 3)$ linkage yielding G_{M3} . The sequential additions to G_{M3} of the *N*-acetylgalactosamine and galactose units from UDP-GalNAc and UDP-Gal yield gangliosides G_{M2} and G_{M1} . Other gangliosides are formed by adding a second



Sulfatide (galactocerebroside-3-sulfate)

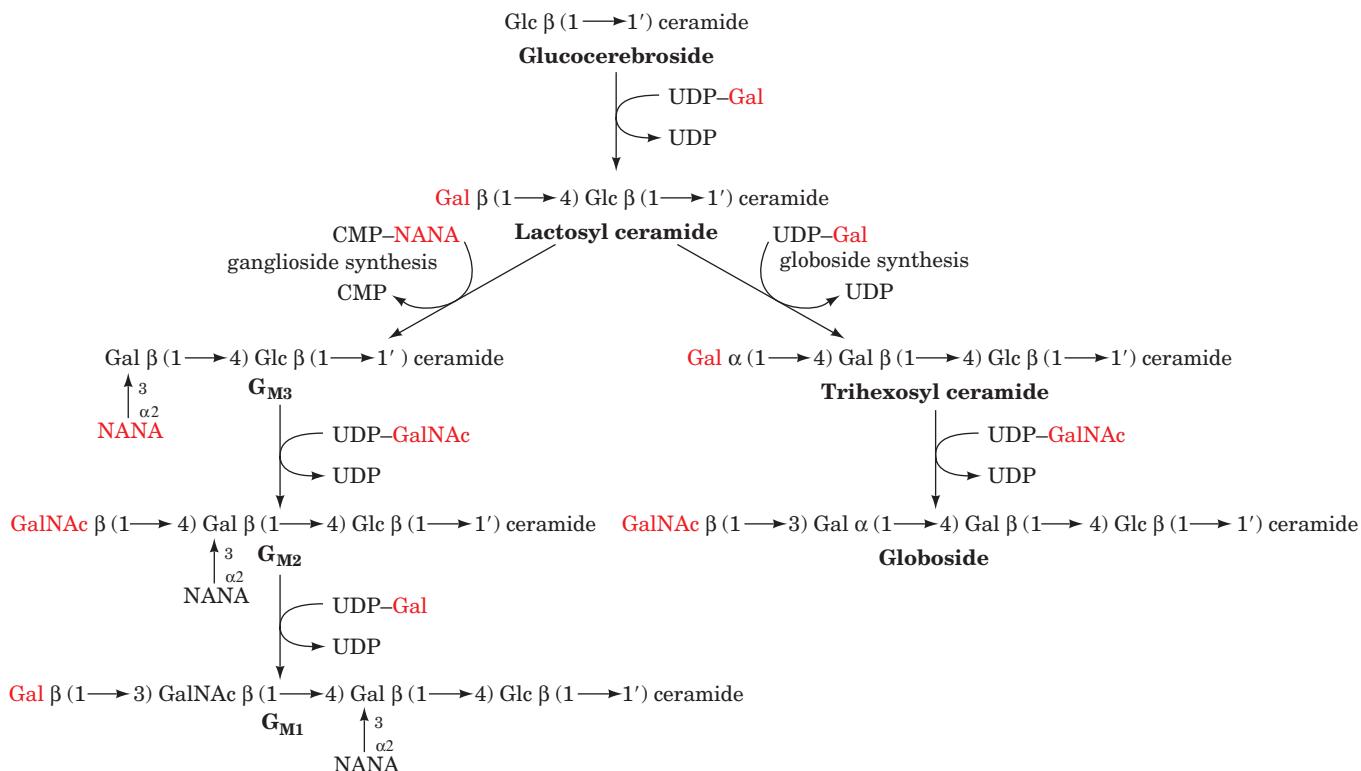


Figure 25-92 The biosynthesis of globosides and G_M gangliosides.

NANA group to G_{M3} , forming G_{D3} , or by adding an *N*-acetylglucosamine unit to lactosyl ceramide before NANA addition, forming G_{A2} . Hundreds of different gangliosides are known.

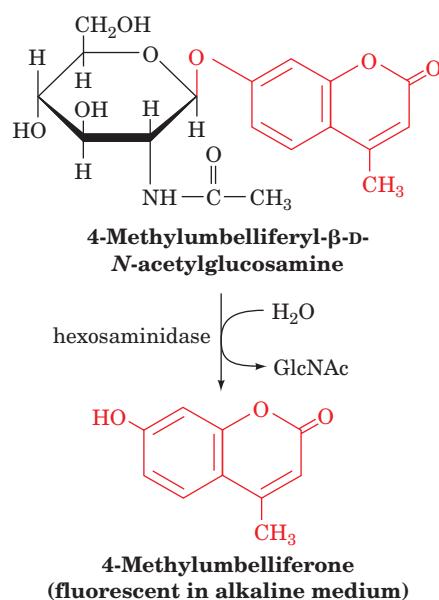
e. Sphingoglycolipid Degradation and Lipid Storage Diseases

Sphingoglycolipids are lysosomally degraded by a series of enzymatically mediated hydrolytic reactions (Fig. 25-93). These reactions are catalyzed at the lipid–water interface by soluble enzymes, often with the aid of **sphingolipid activator proteins (SAPs)**; including **saposins**, G_{M2} -activator protein, and **SAP-A** through **SAP-D**. These nonenzymatic ancillary proteins are thought to increase the accessibility of the carbohydrate moiety of the sphingolipid to the degradation enzyme. For example, G_{M2} -activator binds G_{M2} and helps expose it to the surface of the membrane. The G_{M2} -activator– G_{M2} complex can then bind **hexosaminidase A**, an $\alpha\beta$ dimer that hydrolyzes *N*-acetylgalactosamine from G_{M2} at the lipid–water interface (Fig. 25-94).

The hereditary absence of one of the sphingolipid hydrolases or a SAP results in a **sphingolipid storage disease** (Table 25-2). One of the most common such conditions is **Tay–Sachs disease**, an autosomal recessive deficiency in hexosaminidase A. The absence of hexosaminidase A activity results in the neuronal accumulation of G_{M2} as shell-like inclusions (Fig. 25-95).

Although infants born with Tay–Sachs disease at first appear normal, by ~1 year of age, when sufficient G_{M2} has accumulated to interfere with neuronal function, they be-

come progressively weaker, retarded, and blinded until they die, usually by the age of 3 years. It is possible, however, to screen potential carriers of this disease by a simple serum assay. It is also possible to detect the disease *in utero* by assay of amniotic fluid or amniotic cells obtained by amniocentesis. The assay involves use of an artificial hexosaminidase substrate, **4-methylumbelliferyl- β -D-*N*-acetylglucosamine**, which yields a fluorescent product on hydrolysis:



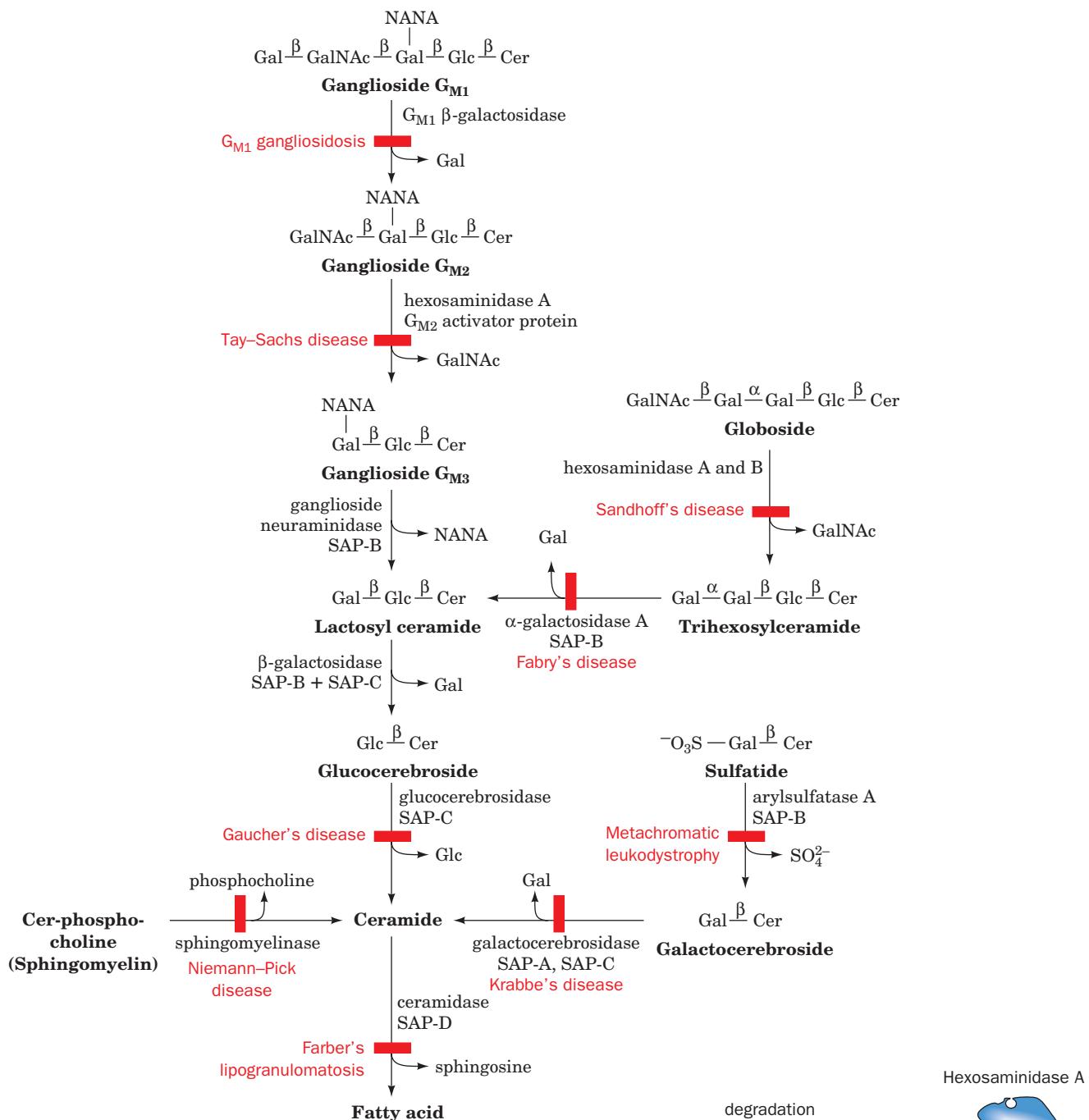


Figure 25-93 The breakdown of sphingolipids by lysosomal enzymes. The genetic diseases caused by the corresponding enzyme deficiencies are noted in red.

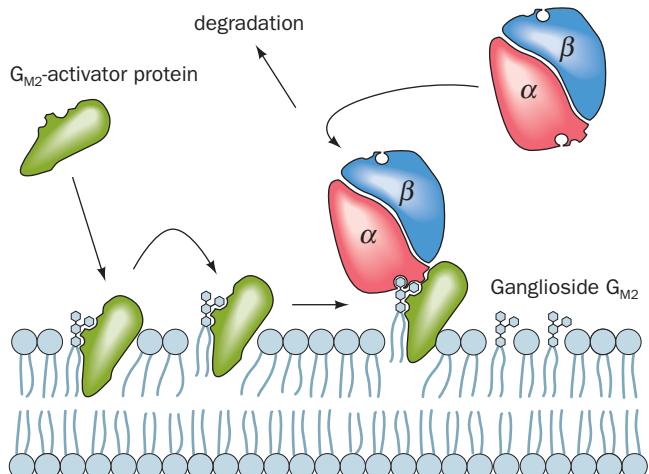


Figure 25-94 Model for G_{M2} -activator protein-stimulated hydrolysis of ganglioside G_{M2} by hexosaminidase A. G_{M2} -activator protein binds and lifts G_{M2} out of the membrane so that it can be recognized and cleaved by the $\alpha\beta$ dimer of hexosaminidase A. [After Kolter, T. and Sandhoff, K., *Angew. Chem. Int. Ed.* **38**, 1532 (1999).]

Table 25-2 Sphingolipid Storage Diseases

Disease	Enzyme Deficiency	Principal Storage Substance	Major Symptoms
G _{M1} Gangliosidosis	G _{M1} β -galactosidase	Ganglioside G _{M1}	Mental retardation, liver enlargement, skeletal involvement, death by age 2
Tay–Sachs disease	Hexosaminidase A	Ganglioside G _{M2}	Mental retardation, blindness, death by age 3
Fabry's disease	α -Galactosidase A	Trihexosylceramide	Skin rash, kidney failure, pain in lower extremities
Sandhoff's disease	Hexosaminidases A and B	Ganglioside G _{M2} and globoside	Similar to Tay–Sachs disease but progressing more rapidly
Gaucher's disease	Glucocerebrosidase	Glucocerebroside	Liver and spleen enlargement, erosion of long bones, mental retardation in infantile form only
Niemann–Pick disease	Sphingomyelinase	Sphingomyelin	Liver and spleen enlargement, mental retardation
Farber's lipogranulomatosis	Ceramidase	Ceramide	Painful and progressively deformed joints, skin nodules, death within a few years
Krabbe's disease	Galactocerebrosidase	Deacylated galactocerebroside	Loss of myelin, mental retardation, death by age 2
Metachromatic leukodystrophy (Sulfatide lipidosis)	Arylsulfatase A	Sulfatide	Mental retardation, death in first decade



Figure 25-95 Cytoplasmic membranous body in a neuron affected by Tay–Sachs disease. [Courtesy of John S. O'Brien, University of California at San Diego Medical School.]

Since this substrate is also recognized by **hexosaminidase B**, which is unaffected in Tay–Sachs disease, the hexosaminidase B is first heat inactivated since it is more heat labile than hexosaminidase A. As a result of mass screening efforts, the tragic consequences of this genetic enzyme deficiency are being averted. The other sphingolipid storage diseases, although less common, have similar consequences (Table 25-2). In recent years, DNA sequencing techniques that screen potential carriers for the most common mutations have become economically feasible.

CHAPTER SUMMARY

1 Lipid Digestion, Absorption, and Transport Triacylglycerols, the storage form of metabolic energy in animals, provide up to six times the metabolic energy of an equal weight of hydrated glycogen. Dietary lipids are digested by pancreatic digestive enzymes such as lipase and phospholipase A₂ that are active at the lipid–water interface of bile salt–

stabilized emulsions. Bile salts are also essential for the intestinal absorption of dietary lipids, as is fatty acid–binding protein. Dietary triacylglycerols and those synthesized by the liver are transported in the blood as chylomicrons and VLDL, respectively. Triacylglycerols present in these lipoproteins are hydrolyzed by lipoprotein lipase outside the cells and enter

them as free fatty acids. Fatty acids resulting from hydrolysis of adipose tissue triacylglycerols by hormone-sensitive triacylglycerol lipase are transported in the bloodstream as fatty acid-albumin complexes.

2 Fatty Acid Oxidation Before fatty acids are oxidized, they are converted to their acyl-CoA derivatives by acyl-CoA synthase in an ATP-requiring process, transported into mitochondria as carnitine esters, and reconverted inside the mitochondrial matrix to acyl-CoA. β Oxidation of fatty acyl-CoA occurs in 2-carbon increments so as to convert even-chain fatty acyl-CoAs completely to acetyl-CoA. The pathway involves FAD-dependent dehydrogenation of an alkyl group, hydration of the resulting double bond, NAD⁺-dependent oxidation of this alcohol to a ketone, and C—C bond cleavage to form acetyl-CoA and a new fatty acyl-CoA with two fewer carbon atoms. The process then repeats itself. Complete oxidation of the acetyl-CoA, NADH, and FADH₂ is achieved by the citric acid cycle and oxidative phosphorylation. Oxidation of unsaturated fatty acids and odd-chain fatty acids also occurs by β oxidation but requires the participation of additional enzymes. Odd-chain fatty acid oxidation generates propionyl-CoA, whose further metabolism requires the participation of (1) propionyl-CoA carboxylase, which has a biotin prosthetic group, (2) methylmalonyl-CoA racemase, and (3) methylmalonyl-CoA mutase, which contains AdoCbl (coenzyme B₁₂). Methylmalonyl-CoA mutase catalyzes a carbon skeleton rearrangement reaction via a free radical mechanism in which the free radical is generated by the homolytic cleavage of AdoCbl's C—Co(III) bond.

β Oxidation of fatty acids takes place in the peroxisomes in addition to the mitochondrion. The peroxisomal pathway differs from the mitochondrial pathway in that the FADH₂ produced in the first step, rather than generating ATP by oxidative phosphorylation, is directly oxidized by O₂ to produce H₂O₂. Peroxisomal enzymes are specific for long-chain fatty acids and function in a chain-shortening process via β oxidation. The resultant intermediate chain-length products are transferred to the mitochondrion for complete oxidation.

3 Ketone Bodies A significant fraction of the acetyl-CoA produced by fatty acid oxidation in the liver is converted to acetoacetate and D- β -hydroxybutyrate, which, together with acetone, are referred to as ketone bodies. The first two compounds serve as important fuels for the peripheral tissues.

4 Fatty Acid Biosynthesis Fatty acid biosynthesis differs from fatty acid oxidation in several respects. Whereas fatty acid oxidation occurs in the mitochondrion utilizing fatty acyl-CoA esters, fatty acid biosynthesis occurs in the cytosol with the growing fatty acids esterified to acyl-carrier protein (ACP). The redox coenzymes differ (FAD and NAD⁺ for oxidation; NADPH for biosynthesis), as does the stereochemistry of the pathway's intermediate steps. Oxidation produces acetyl-CoA, whereas malonyl-CoA is the immediate precursor in biosynthesis. Malonyl-CoA is produced by the ATP-driven reaction of HCO₃⁻ with acetyl-CoA as catalyzed by the biotin-containing enzyme acetyl-CoA carboxylase, although this HCO₃⁻ is not incorporated into the final fatty acid product. In bacteria, fatty acids synthesis is carried out by a series of independent enzymes that are collectively known as FAS-I. In animals, fatty acid synthesis yielding palmitate occurs on a single X-shaped homodimeric protein known as FAS-II, each of whose subunits contain all six activities required to do so on

separate domains that are serviced by ACP. In fungi, FAS-II is an $\alpha_6\beta_6$ heterododecamer that forms a barrel-shaped protein that contains two equivalent reaction chambers. Similar but more extensive systems are used in the synthesis of the various polyketides. Acetyl-CoA is transferred from the mitochondrion to the cytosol as citrate via the tricarboxylate transport system and released by citrate cleavage to yield acetyl-CoA and oxaloacetate. Oxaloacetate is converted to malate and then to pyruvate for transport back to the mitochondrion, a process that also generates some of the NADPH required for biosynthesis. Palmitate is the primary product of fatty acid biosynthesis in animals. Longer chain fatty acids and unsaturated fatty acids are synthesized from palmitate by elongation and desaturation reactions. Certain essential unsaturated fatty acids cannot be synthesized by animals and hence must be obtained from the diet. Triacylglycerols are synthesized from fatty acyl-CoA esters and glycerol-3-phosphate.

5 Regulation of Fatty Acid Metabolism Fatty acid metabolism is regulated through the allosteric control of hormone-sensitive triacylglycerol lipase and acetyl-CoA carboxylase, phosphorylation/dephosphorylation, and/or changes in the rates of protein synthesis and breakdown. This regulation is mediated by the hormones glucagon, epinephrine, and norepinephrine, which activate degradation, and by insulin, which activates biosynthesis. These hormones interact to control the cAMP concentration, which in turn controls phosphorylation/dephosphorylation ratios via PKA. AMPK, which senses the level of ATP, is also an important regulator of fatty acid metabolism.

6 Cholesterol Metabolism Cholesterol is a vital constituent of cell membranes and is the precursor of the steroid hormones and bile salts. Its biosynthesis, transport, and utilization are rigidly controlled. Cholesterol is synthesized in the liver from acetate in a pathway that involves formation of HMG-CoA from three molecules of acetate followed by reduction, phosphorylation, decarboxylation, and dehydration to the isoprene units isopentenyl pyrophosphate and dimethylallyl pyrophosphate. Four of these isoprene units are then condensed via cationic mechanisms to form squalene, which, in turn, undergoes a cyclization reaction, via a cationic cascade, to form lanosterol, the sterol precursor to cholesterol.

The pathway's major control point is at HMG-CoA reductase. This enzyme is regulated by competitive and allosteric mechanisms, by phosphorylation/dephosphorylation, and, most importantly, by long-term control of the rates of enzyme synthesis and degradation. Long-term control is mediated by the integral membrane protein SREBP, which when the cholesterol level is low, is escorted by SCAP to the Golgi apparatus via COPII-coated vesicles. There it is sequentially cleaved by the proteases S1P and S2P, thereby releasing its soluble bHLH/Z domain to travel to the nucleus where it induces the transcription of SRE-containing genes such as those encoding HMG-CoA reductase and the LDL receptor. In addition, when sterol levels are high, the rate at which HMG-CoA reductase is proteolytically degraded is greatly increased.

The liver secretes cholesterol into the bloodstream in esterified form as part of the VLDL. This complex is sequentially converted to IDL and then to LDL. LDL, which is brought into the cells by receptor-mediated endocytosis, carries the major portion of cholesterol to peripheral tissues for utilization. Excess cholesterol is returned to the liver from peripheral tissues

by HDL. The cellular supply of cholesterol is controlled by three mechanisms: (1) long- and short-term regulation of HMG-CoA reductase; (2) control of LDL receptor synthesis by cholesterol concentration; and (3) long- and short-term regulation of acyl-CoA:cholesterol acyltransferase (ACAT), which mediates cholesterol esterification. Cholesterol is the precursor to the steroid hormones, which are classified as progestins, glucocorticoids, mineralocorticoids, androgens, and estrogens. The only pathway for the excretion of cholesterol in mammals is through the formation and elimination of bile salts.

7 Eicosanoid Metabolism: Prostaglandins, Prostacyclins, Thromboxanes, Leukotrienes, and Lipoxins Prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins are eicosanoid products produced largely by the metabolism of arachidonate. These highly unstable compounds have profound physiological effects at extremely low concentrations. They are involved in the inflammatory response, the production of pain and fever, the regulation of blood pressure, and many other important physiological processes. Arachidonate is synthesized from linoleic acid, an essential fatty acid, and stored as phosphatidylinositol and other phospholipids. Prostaglandins, prostacyclins, and thromboxanes are synthesized via the cyclic pathway, whereas leukotrienes and lipoxins are synthesized via the linear pathway. Aspirin and other non-steroidal anti-inflammatory drugs (NSAIDs) inhibit the cyclic pathway but not the linear pathway. COX-2 inhibitors are NSAIDs that bind to COX-2 but not COX-1, thereby eliminating the side effects of other NSAIDs. Peptidoleukotrienes have been identified as the slow reacting substances of anaphylaxis (SRS-A) released from sensitized lung after immuno-

logical challenge. Lipoxins and aspirin-induced *epi*-lipoxins have anti-inflammatory properties.

8 Phospholipid and Glycolipid Metabolism Complex lipids have either a phosphate ester or a carbohydrate as their polar head group and either 1,2-diacyl-*sn*-glycerol or ceramide (*N*-acylsphingosine) as their hydrophobic tail. Phospholipids are either glycerophospholipids or sphingophospholipids, whereas glycolipids are either glyceroglycolipids or sphingoglycolipids. The polar head groups of glycerophospholipids, which are phosphate esters of either ethanolamine, serine, choline, inositol, or glycerol, are attached to 1,2-diacyl-*sn*-glycerol's C3 OH group by means of CTP-linked transferase reactions. The specific long-chain fatty acids found at the C1 and C2 positions are incorporated by "remodeling reactions" after the addition of the polar head group. Plasmalogens and alkylacylglycerophospholipids, respectively, contain a long-chain alkyl group in a vinyl ether linkage or in an ether linkage to glycerol's C1 OH group. Platelet-activating factor (PAF) is an important alkylacylglycerophospholipid. The only major sphingophospholipid is sphingomyelin (*N*-acylsphingosine phosphocholine), an important structural lipid of nerve cell membranes. Most sphingolipids contain polar head groups composed of carbohydrate units and are therefore referred to as sphingoglycolipids. The principal classes of sphingoglycolipids are cerebrosides, sulfatides, globosides, and gangliosides. Their carbohydrate units, which are attached to *N*-acylsphingosine's C1 OH group by glycosidic linkages, are formed by stepwise addition of activated monosaccharide units. Several lysosomal sphingolipid storage diseases, including Tay-Sachs disease, result from deficiencies in the enzymes that degrade sphingoglycolipids.

REFERENCES

General

- Newsholme, E.A. and Leech, A.R., *Biochemistry for the Medical Sciences*, Wiley (1983). [Chapters 6–8 contain a wealth of information on the control of fatty acid metabolism and its integration into the overall scheme of metabolism.]
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Contain numerous chapters on defects in lipid metabolism.]
- Vance, D.E. and Vance J.E. (Eds.), *Biochemistry of Lipids, Lipoproteins, and Membranes* (5th ed.), Elsevier (2008).

Lipid Digestion

- Berg, O.G., Gelb, M.H., Tsai, M.-D., and Jain, M.K., Interfacial enzymology: the secreted phospholipase A₂-paradigm, *Chem. Rev.* **101**, 2613–2653 (2001).
- Bhattacharya, A.A., Grüne, T., and Curry, S., Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin, *J. Mol. Biol.* **303**, 721–732 (2002).
- Borgström, B., Barrowman, J.A., and Lindström, M., Roles of bile acids in intestinal lipid digestion and absorption, in Danielsson, H. and Sjövall, J. (Eds.), *Sterols and Bile Acids*, pp. 405–425, Elsevier (1985).
- Brady, L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L., and Menge, U., A serine pro-

tease triad forms the catalytic centre of a triacylglycerol lipase, *Nature* **343**, 767–770 (1990).

- Hermoso, J., Pignol, D., Penel, S., Roth, M., Chapus, C., and Fontecilla-Camps, J.C., Neutron crystallographic evidence of lipase–colipase complex activation by a micelle, *EMBO J.* **16**, 5531–5536 (1997).
- Scapin, G., Gordon, J.I., and Sacchettini, J.C., Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2-Å resolution, *J. Biol. Chem.* **267**, 4253–4269 (1992).
- van Tilburgh, H., Bezzine, S., Cambillau, C., Verger, R., and Carrière, F., Colipase: structure and interaction with pancreatic lipase, *Biochim. Biophys. Acta* **1441**, 173–184 (1999).

Fatty Acid Oxidation

- Bannerjee, R., Radical peregrinations catalyzed by coenzyme B₁₂-dependent enzymes, *Biochem.* **40**, 6191–6198 (2001).
- Bartlett, K. and Eaton, S., Mitochondrial β-oxidation, *Eur. J. Biochem.* **271**, 462–469 (2004). [Discusses the reactions and enzymes of mitochondrial β oxidation as well as their regulation.]
- Bieber, L.L., Carnitine, *Annu. Rev. Biochem.* **88**, 261–283 (1988).
- Brownsey, R.W., Boone, A.N., Elliott, J.E., Kulpa, J.E., and Lee, W.M., Regulation of acetyl-CoA carboxylase, *Biochem. Soc. Trans.* **34**, 223–227 (2006).
- Chou, C.-Y., Yu, L.P.C., and Tong, L., Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism, *J. Biol. Chem.* **284**, 11690–11697 (2009).

- Kim, J.-J.P. and Battaile, K.P., Burning fat: The structural basis of fatty acid β -oxidation, *Curr. Opin. Struct. Biol.* **12**, 721–728 (2002).
- Kim, J.-J.P., Wang, M., and Pashke, R., Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate, *Proc. Natl. Acad. Sci.* **90**, 7523–7527 (1993).
- Kindl, H., Fatty acid degradation in plant peroxisomes: Function and biosynthesis of the enzymes involved, *Biochimie* **75**, 225–230 (1993).
- Mancia, R., Smith, G.A., and Evans, P.R., Crystal structure of substrate complexes of methylmalonyl-CoA mutase, *Biochem.* **38**, 7999–8005 (1999).
- Marsh, E.N.G. and Drennan, C.L., Adenosylcobalamin-dependent isomerases: new insights into structure and mechanism, *Curr. Opin. Chem. Biol.* **5**, 499–505 (2001).
- Rinaldo, P., Matern, D., and Bennett, M.J., Fatty acid oxidation disorders, *Annu. Rev. Physiol.* **64**, 477–502 (2002).
- Shoukry, K. and Schulz, H., Significance of the reductase-dependent pathway for the β -oxidation of unsaturated fatty acids with odd-numbered double bonds: mitochondrial metabolism of 2-trans-5-cis-octadienoyl-CoA, *J. Biol. Chem.* **273**, 6892–6899 (1998).
- Sudden infant death and inherited disorders of fat oxidation, *Lancet*, 1073–1075, Nov. 8, 1986.
- Tong, L., Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery, *Cell. Mol. Life Sci.* **62**, 1784–1893 (2005).
- van den Bosch, H., Schutgens, R.B.H., Wanders, R.J.A., and Tager, J.M., Biochemistry of peroxisomes, *Annu. Rev. Biochem.* **61**, 157–197 (1992).
- Watkins, P.A., Very-long-chain acyl-CoA synthetases, *J. Biol. Chem.* **283**, 1773–1777 (2008).

Fatty Acid Biosynthesis

- Bartlett, K. and Eaton, S., Mitochondrial β -oxidation, *Eur. J. Biochem.* **271**, 462–469 (2004). [Discusses the reactions and enzymes of mitochondrial β oxidation as well as their regulation.]
- Brownsey, R.W. and Denton, R.M., Acetyl-coenzyme A carboxylase, in Boyer, P.D. and Krebs, E.G. (Eds.), *The Enzymes* (3rd ed.), Vol. 18, pp. 123–146, Academic Press (1987).
- Brownsey, R.W., Zhande, R., and Boone, A.N., Isoforms of acetyl-CoA carboxylase: structures, regulatory properties and metabolic functions, *Biochem. Soc. Trans.* **25**, 1232–1238 (1997).
- Haapalainen, A.M., Meriläinen, G., and Wierenga, R.K., The thiolase superfamily: condensing enzymes with diverse reaction specificities, *Trends Biochem. Sci.* **31**, 64–71 (2006).
- Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolásek, B., and Ban, N., Structure of fungal fatty acid synthase and implications for iterative substrate shuttling, *Science* **316**, 254–261 (2007).
- Jump, D.B., The biochemistry of *n*-3 polyunsaturated fatty acids, *J. Biol. Chem.* **277**, 8755–8758 (2002).
- Khosla, C., Tang, Y., Chen, A.Y., Schnarr, N.A., and Cane, D.E., Structure and mechanism of the 6-deoxyerythronolide B synthase, *Annu. Rev. Biochem.* **76**, 195–221 (2007).
- Leibundgut, M., Maier, T., Jenni, S., and Ban, N., The multienzyme architecture of eukaryotic fatty acid synthases, *Curr. Opin. Struct. Biol.* **18**, 714–725 (2008).
- Lomakin, I.B., Xiong, Y., and Steitz, T.A., The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together, *Cell* **129**, 319–332 (2007); and Leibundgut, M., Jenni, S., Frick, C., and Ban, N., Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase, *Science* **316**, 288–290 (2007).

- Los, D.A. and Murata, N., Structure and expression of fatty acid desaturases, *Biochim. Biophys. Acta* **1394**, 3–15 (1998).
- Maier, T., Leibundgut, M., and Ban, N., The crystal structure of a mammalian fatty acid synthase, *Science* **321**, 1315–1322 (2008).
- Smith, S. and Tsai, S.-C., The type I fatty acid and polyketide synthases: a tale of two megasynthases, *Nat. Prod. Rep.* **24**, 1041–1072 (2007).
- Wakil, S.J., Fatty acid synthase, a proficient multifunctional enzyme, *Biochemistry* **28**, 4523–4530 (1989).
- Walsh, C.T., Polyketide and nonribosomal peptide antibiotics: Modularity and versatility, *Science* **303**, 1805–1810 (2004).
- White, S.W., Zheng, J., Zhang, Y.-M., and Rock, C.O., The structural biology of type II fatty acid biosynthesis, *Annu. Rev. Biochem.* **74**, 791–831 (2005).

Regulation of Fatty Acid Metabolism

- Eaton, S., Control of mitochondrial β -oxidation flux, *Prog. Lipid Res.* **41**, 197–239 (2002).
- Hardie, D.G. and Carling, D., The AMP-activated protein kinase: fuel gauge of the mammalian cell? *Eur. J. Biochem.* **246**, 259–273 (1997).
- Hardie, D.G., Carling, D., and Carlson, M., The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Annu. Rev. Biochem.* **67**, 821–855 (1998).
- Munday M.R. and Hemingway C.J., The regulation of acetyl-CoA carboxylase—A potential target for the action of hypolipidemic agents, *Adv. Enzym. Regul.* **39**, 205–234 (1999).
- Witters, L.A., Watts, T.D., Daniels, D.L., and Evans, J.L., Insulin stimulates the dephosphorylation and activation of acetyl-CoA carboxylase, *Proc. Natl. Acad. Sci.* **85**, 5473–5477 (1988).

Cholesterol Metabolism

- Bloch, K., The biological synthesis of cholesterol, *Science* **150**, 19–28 (1965).
- Chang, T.Y., Chang, C.C.Y., and Cheng, D., Acyl-coenzyme A:cholesterol acyltransferase, *Annu. Rev. Biochem.* **66**, 613–638 (1997).
- Durbecq, V., et al., Crystal structure of isopentenyl diphosphate: dimethylallyl diphosphate isomerase, *EMBO J.* **20**, 1530–1537 (2001).
- Edwards, P.A., Sterols and isoprenoids: signaling molecules derived from the cholesterol biosynthetic pathway, *Annu. Rev. Biochem.* **68**, 157–185 (1999).
- Goldstein, J.L. and Brown, M.S., Regulation of the mevalonate pathway, *Nature* **343**, 425–430 (1990).
- Goldstein, J.L., DeBose-Boyd, R.A., and Brown, M.S., Protein sensors for membrane sterols, *Cell* **124**, 35–46 (2006).
- Goldstein, J.L., Rawson, R.B., and Brown, M.S., Mutant mammalian cells as tools to delineate the sterol regulatory element binding protein pathway for feedback regulation of lipid synthesis, *Arch. Biochem. Biophys.* **397**, 139–148 (2002).
- Ikonen, E., Cellular cholesterol trafficking and compartmentalization, *Nature Rev. Mol. Cell Biol.* **9**, 125–138 (2008).
- Istvan, E.S., Bacterial and mammalian HMG-CoA reductases: related enzymes with distinct architectures, *Curr. Opin. Struct. Biol.* **11**, 746–751 (2001).
- Istvan, E.S. and Deisenhofer, J., Structural mechanism for statin inhibition of HMG-CoA reductase, *Science* **292**, 1160–1164 (2001).
- Knopp, R.H., Drug therapy: drug treatment of lipid disorders, *New Engl. J. Med.* **341**, 498–511 (1999).
- Meyer, M.M., Segura, M.J.R., Wilson, W.K., and Matsuda, S.P.T., Oxidosqualene cyclase residues that promote formation of cycloartenol, lanosterol and parkeol, *Angew. Chem. Int. Ed.* **39**, 4090–4092 (2000).

- Reinert, D.J., Balliano, G., and Schulz, G.E., Conversion of squalene to the pentacyclic hopene, *Chem. Biol.* **11**, 121–126 (2004). [The X-ray structure of squalene-hopene cyclase in complex with 2-azasqualene.]
- Russell, D.W. and Setchell, K.D.R., Bile acid biosynthesis, *Biochemistry* **31**, 4737–4749 (1992).
- Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebl, J., Dehmiow, H., Hennig, M., Stihle, M., and Ruf, A., Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase, *Nature* **432**, 118–122 (2004).
- Wang, K.C. and Ohnuma, S.-I., Isoprenyl diphosphate synthases, *Biochim. Biophys. Acta* **1529**, 33–48 (2000).
- Wendt, K.U., Schulz, G.E., Corey, E.J. and Liu, D.R., Enzyme mechanisms for polycyclic triterpene formation, *Angew. Chem. Int. Ed.* **39**, 2812–2833 (2000).
- Yokode, M., Hammer, R.E., Ishibashi, S., Brown, M.S., and Goldstein, J.L., Diet-induced hypercholesterolemia in mice: Prevention by overexpression of LDL receptors, *Science* **250**, 1273–1275 (1990).

Eicosanoid Metabolism

- Abramovitz, M., Wong, E., Cox, M.E., Richardson, C.D., Li, C., and Vickers, P.J., 5-Lipoxygenase-activating protein stimulates the utilization of arachidonic acid by 5-lipoxygenase, *Eur. J. Biochem.* **215**, 105–111 (1993).
- Chandrasekharan, N.V., Dai, H., Roos, K.L.T., Evanson, N.K., Tomsik, J., Elton, T.S., and Simmons, D.L., COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: Cloning, structure, and expression, *Proc. Natl. Acad. Sci.* **99**, 13926–13931 (2002).
- Ferguson, A.D., et al., Crystal structure of inhibitor-bound human 5-lipoxygenase-activating protein, *Science* **317**, 510–512 (2007).
- Ford-Hutchinson, A.W., Gresser, M., and Young, R.N., 5-Lipoxygenase, *Annu. Rev. Biochem.* **63**, 383–417 (1994).
- Gillmor, S.A., Villaseñor, A., Fletterick, R., Sigal, E., and Browner, M.F., The structure of mammalian 15-lipoxygenase reveals similarity to the lipases and the determinants of substrate specificity, *Nature Struct. Biol.* **4**, 1003–1009 (1997).
- Kurumbail, R.G., Kiefer, J.R., and Marnett, L.J., Cyclooxygenase enzymes: catalysis and inhibition, *Curr. Opin. Struct. Biol.* **11**, 752–760 (2001).
- Phillipson, B.E., Rothrock, D.W., Conner, W.E., Harris, W.S., and Illingworth, D.R., Reduction of plasma lipids, lipoproteins and apoproteins by dietary fish oils in patients with hypertriglyceridemia, *New Engl. J. Med.* **312**, 1210–1216 (1985).
- Picot, D., Loll, P.J., and Garavito, R.M., The X-ray crystal structure

- of the membrane protein prostaglandin H₂ synthase-1, *Nature* **367**, 243–249 (1994).
- Rådmark, O., Werz, O., Steinhilber, D., and Samuelsson, B., 5-Lipoxygenase: regulation of expression and enzyme activity, *Trends Biochem. Sci.* **32**, 332–341 (2007).
- Samuelsson, B. and Funk, C.D., Enzymes involved in the biosynthesis of leukotriene B₄, *J. Biol. Chem.* **264**, 19469–19472 (1989).
- Schwarz, K., Walther, M., Anton, M., Gerth, C., Feussner, I., and Kuhn, H., Structural basis for lipoxygenase specificity: conversion of the human leukocyte 5-lipoxygenase to a 15-lipoxygenating enzyme species by site-directed mutagenesis, *J. Biol. Chem.* **276**, 773–339 (2001).
- Serhan, C.N., Lipoxins and novel aspirin-triggered 15-*epi*-lipoxins (ATL): a jungle of cell-cell interactions or a therapeutic opportunity? *Prostaglandins* **53**, 107–137 (1997).
- Smith, W.L., Nutritionally essential fatty acids and biologically indispensable cyclooxygenases, *Trends Biochem. Sci.* **33**, 27–37 (2003).
- Smith, W.L., DeWitt, D.L., and Garavito, R.M., Cyclooxygenases: structural, cellular and molecular biology, *Annu. Rev. Biochem.* **69**, 145–182 (2000).
- Turini, M.E. and DuBois, R.N., Cyclooxygenase-2: a therapeutic target, *Annu. Rev. Med.* **53**, 35–57 (2002).
- Warner, T.D. and Mitchell, J.A., Cyclooxygenases: new forms, new inhibitors, and lessons from the clinic, *FASEB J.* **18**, 790–804 (2004).

Phospholipid and Glycolipid Metabolism

- Conzelmann, E. and Sandhoff, K., Glycolipid and glycoprotein degradation, *Adv. Enzymol.* **60**, 89–216 (1987).
- Dowhan, W., Molecular basis for membrane diversity: Why are there so many lipids? *Annu. Rev. Biochem.* **66**, 199–232 (1997).
- Kent, C., Eukaryotic phospholipid synthesis, *Annu. Rev. Biochem.* **64**, 315–342 (1995).
- Colter, T. and Sandhoff, K., Sphingolipids—their metabolic pathways and the pathobiochemistry of neurodegenerative diseases, *Angew. Chem. Int. Ed.* **38**, 1532–1568 (1999).
- Neufeld, E.F., Natural history and inherited disorders of a lysosomal enzyme, β -hexosaminidase, *J. Biol. Chem.* **264**, 10927–10930 (1989).
- Prescott, S.M., Zimmerman, G.A., and McIntire, T.M., Platelet-activating factor, *Biol. Chem.* **265**, 17381–17384 (1990).
- Tifft, C.J. and Proila, R.L., Stemming the tide: glycosphingolipid synthesis inhibitors as therapy for storage diseases, *Glycobiology* **10**, 1249–1258 (2000).
- van Echten, G. and Sandhoff, K., Ganglioside metabolism, *J. Biol. Chem.* **268**, 5341–5344 (1993).

PROBLEMS

- The venoms of many poisonous snakes, including rattlesnakes, contain a phospholipase A₂ that causes tissue damage that is seemingly far out of proportion to the small amount of enzyme injected. Explain.
- Explain why individuals with a hereditary deficiency of carnitine palmitoyltransferase II have muscle weakness. Why are these symptoms more severe during fasting?
- Why are the livers of Jamaican vomiting sickness victims usually depleted of glycogen?
- Compare the metabolic efficiencies, in moles of ATP pro-

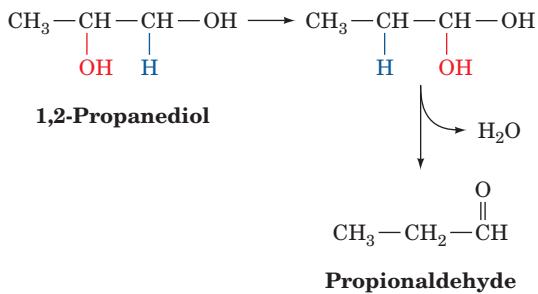
duced per gram, of completely oxidized fat (tripalmitoyl glycerol) versus glucose derived from glycogen. Assume that the fat is anhydrous and the glycogen is stored with twice its weight of water.

- Methylmalonyl-CoA mutase is incubated with deuterated methylmalonyl-CoA. The coenzyme B₁₂ extracted from this mutase is found to contain deuterium at its 5'-methylene group. Account for the transfer of label from substrate to coenzyme.
- What is the energetic price, in ATP units, of converting acetoacetyl-CoA to acetoacetate and then resynthesizing acetoacetyl-CoA?

7. A fasting animal is fed palmitic acid that has a ^{14}C -labeled carboxyl group. (a) After allowing sufficient time for fatty acid breakdown and resynthesis, what would be the ^{14}C -labeling pattern in the animal's palmitic acid residues? (b) The animal's liver glycogen becomes ^{14}C labeled although there is no net increase in the amount of this substance present. Indicate the sequence of reactions whereby the glycogen becomes labeled. Why is there no net glycogen synthesis?

8. What is the ATP yield from the complete oxidation of a molecule of (a) α -linolenic acid (9,12,15-octadecatrienoic acid, 18:3n-3) and (b) **margaric acid** (heptadecanoic acid, 17:0)? Which has the greater amount of available biological energy on a per carbon basis?

***9.** The role of coenzyme B₁₂ in mediating hydrogen transfer was established using the coenzyme B₁₂-dependent bacterial enzyme **dioldehydrase**, which catalyzes the reaction:

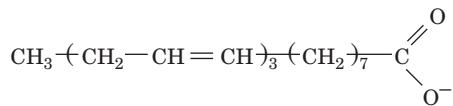


The enzyme converts $[1-^3\text{H}]$ 1,2-propanediol to $[1,2-^3\text{H}]$ propionaldehyde with the incorporation of tritium into both C5' positions of 5'-deoxyadenosylcobalamin's 5'-deoxyadenosyl residue. Suggest the mechanism of this reaction. What would be the products of the dioldehydrase reaction if the enzyme was supplied with $[5'-^3\text{H}]$ deoxyadenosylcobalamin and unlabeled 1,2-propanediol?

10. Why is it important that liver cells lack 3-ketoacyl-CoA transferase (Fig. 25-27)?

11. What is the energetic price, in ATP equivalents, of breaking down palmitic acid to acetyl-CoA and then resynthesizing it?

12. Is the fatty acid shown below likely to be synthesized in animals? Explain.



13. What is the energetic price, in ATP equivalents, of synthesizing cholesterol from acetyl-CoA?

14. What would be the ^{14}C -labeling pattern in cholesterol if it were synthesized from HMG-CoA that was ^{14}C labeled (a) at C5, its carboxyl carbon atom, or (b) C1, its thioester carbon atom?

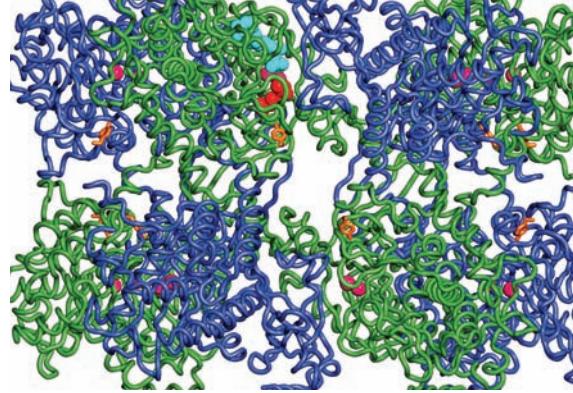
15. Hypercholesterolemic individuals taking statins are sometimes advised to take supplements of coenzyme Q. Explain.

***16.** A child suffering from severe abdominal pain is admitted to the hospital several hours after eating a meal consisting of hamburgers, fried potatoes, and ice cream. Her blood has the appearance of "creamed tomato soup" and on analysis is found to contain massive quantities of chylomicrons. As the attending physician, what is your diagnosis of the patient's difficulty (the cause of the abdominal pain is unclear)? What treatment would you prescribe to alleviate the symptoms of this inherited disease?

17. Although linoleic acid is an essential fatty acid in animals, it is not required by animal cells in tissue culture. Explain.

18. The inactivation of the peroxidase function of prostaglandin H synthase (PGHS) also inactivates its cyclooxygenase function but not vice versa. Explain.

Amino Acid Metabolism



CHAPTER 26

1 Amino Acid Deamination

- A. Transamination
- B. Oxidative Deamination: Glutamate Dehydrogenase
- C. Other Deamination Mechanisms

2 The Urea Cycle

- A. Carbamoyl Phosphate Synthetase: Acquisition of the First Urea Nitrogen Atom
- B. Ornithine Transcarbamoylase
- C. Argininosuccinate Synthetase: Acquisition of the Second Urea Nitrogen Atom
- D. Argininosuccinase
- E. Arginase
- F. Regulation of the Urea Cycle

3 Metabolic Breakdown of Individual Amino Acids

- A. Amino Acids Can Be Glucogenic, Ketogenic, or Both
- B. Alanine, Cysteine, Glycine, Serine, and Threonine Are Degraded to Pyruvate
- C. Asparagine and Aspartate Are Degraded to Oxaloacetate
- D. Arginine, Glutamate, Glutamine, Histidine, and Proline Are Degraded to α -Ketoglutarate
- E. Isoleucine, Methionine, and Valine Are Degraded to Succinyl-CoA
- F. Leucine and Lysine Are Degraded to Acetoacetate and/or Acetyl-CoA
- G. Tryptophan Is Degraded to Alanine and Acetoacetate
- H. Phenylalanine and Tyrosine Are Degraded to Fumarate and Acetoacetate

4 Amino Acids as Biosynthetic Precursors

- A. Heme Biosynthesis and Degradation
- B. Biosynthesis of Physiologically Active Amines
- C. Glutathione
- D. Tetrahydrofolate Cofactors: The Metabolism of C₁ Units

5 Amino Acid Biosynthesis

- A. Biosynthesis of the Nonessential Amino Acids
- B. Biosynthesis of the Essential Amino Acids

6 Nitrogen Fixation

α -Amino acids, in addition to their role as protein monomeric units, are energy metabolites and precursors of many biologically important nitrogen-containing compounds, notably heme, physiologically active amines, glutathione, nucleotides, and nucleotide coenzymes. Amino acids are classified into two groups: **essential** and **nonessential**. Mammals synthesize the nonessential amino acids from

metabolic precursors but must obtain the essential amino acids from their diet. Excess dietary amino acids are neither stored for future use nor excreted. Rather, they are converted to common metabolic intermediates such as pyruvate, oxaloacetate, acetyl-CoA, and α -keto-glutarate. Consequently, *amino acids are also precursors of glucose, fatty acids, and ketone bodies and are therefore metabolic fuels*.

In this chapter, we consider the pathways of amino acid breakdown, synthesis, and utilization. We begin by examining the three common stages of amino acid breakdown:

1. Deamination (amino group removal), whereby amino groups are converted either to ammonia or to the amino group of aspartate.

2. Incorporation of ammonia and aspartate nitrogen atoms into urea for excretion.

3. Conversion of amino acid carbon skeletons (the α -keto acids produced by deamination) to common metabolic intermediates.

Many of these reactions are similar to those we have considered in other pathways. Others employ enzyme cofactors we have not previously encountered. One of our goals in studying amino acid metabolism is to understand the mechanisms of action of these cofactors.

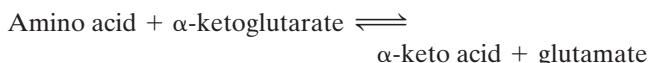
After our discussion of amino acid breakdown, we examine the pathways by which amino acids are utilized in the biosynthesis of heme, physiologically active amines, and glutathione (the synthesis of nucleotides and nucleotide coenzymes is the subject of Chapter 28). Next, we study amino acid biosynthesis pathways. The chapter ends with a discussion of nitrogen fixation, a process that converts atmospheric N₂ to ammonia and is therefore the ultimate biological source of metabolically useful nitrogen.

1 AMINO ACID DEAMINATION

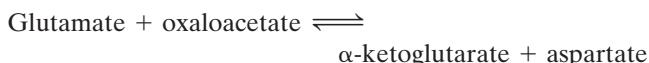
The first reaction in the breakdown of an amino acid is almost always removal of its α -amino group with the object of excreting excess nitrogen and degrading the remaining carbon skeleton or converting it to glucose. Urea, the predominant nitrogen excretion product in terrestrial mammals, is synthesized from ammonia and aspartate. Both of these latter substances are derived mainly from glutamate, a product of most deamination reactions. In this section we

examine the routes by which α -amino groups are incorporated into glutamate and then into aspartate and ammonia. In Section 26-2, we discuss urea biosynthesis from these precursors.

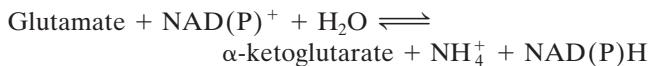
Most amino acids are deaminated by **transamination**, the transfer of their amino group to an α -keto acid to yield the α -keto acid of the original amino acid and a new amino acid, in reactions catalyzed by **aminotransferases** (alternatively, **transaminases**). The predominant amino group acceptor is α -ketoglutarate, producing glutamate as the new amino acid:



Glutamate's amino group, in turn, is transferred to oxaloacetate in a second transamination reaction, yielding aspartate:



Transamination, of course, does not result in any net deamination. Deamination occurs largely through the oxidative deamination of glutamate by **glutamate dehydrogenase (GDH)**, yielding ammonia. The reaction requires NAD^+ or NADP^+ as an oxidizing agent and regenerates α -ketoglutarate for use in additional transamination reactions:

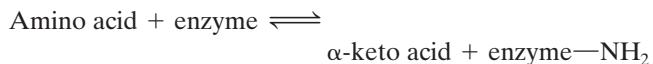


The mechanisms of transamination and oxidative deamination are the subjects of this section. We also consider other means of amino group removal from specific amino acids.

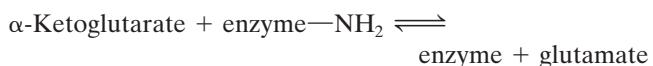
A. Transamination

a. Aminotransferase Reactions Occur in Two Stages

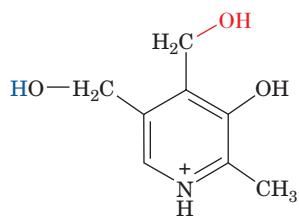
1. The amino group of an amino acid is transferred to the enzyme, producing the corresponding keto acid and the aminated enzyme.



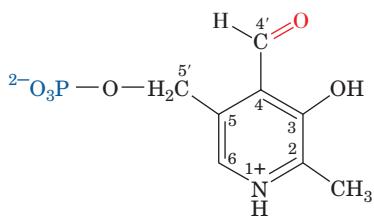
2. The amino group is transferred to the keto acid acceptor (e.g., α -ketoglutarate), forming the amino acid product (e.g., glutamate) and regenerating the enzyme.



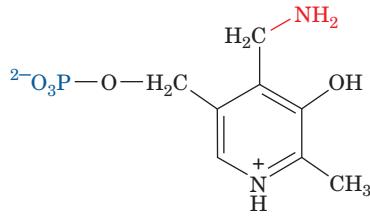
To carry the amino group, aminotransferases require participation of an aldehyde-containing coenzyme, **pyridoxal-5'-phosphate (PLP)**, a derivative of **pyridoxine (vitamin B₆)** (Fig. 26-1a,b). The amino group is accommodated by conversion of this coenzyme to **pyridoxamine-5'-phosphate (PMP)** (Fig. 26-1c). PLP is covalently attached to the enzyme via a Schiff base (imine) linkage formed by the



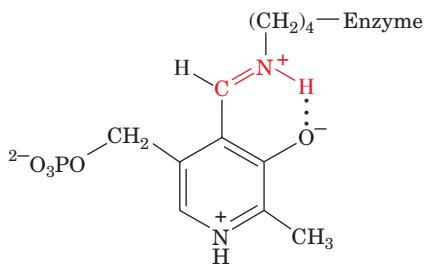
(a) Pyridoxine (vitamin B₆)



(b) Pyridoxal-5'-phosphate (PLP)



(c) Pyridoxamine-5'-phosphate (PMP)



(d) Enzyme-PLP Schiff base

(c) Pyridoxamine-5'-phosphate (PMP). (d) The Schiff base that forms between PLP and an enzyme ϵ -amino group.

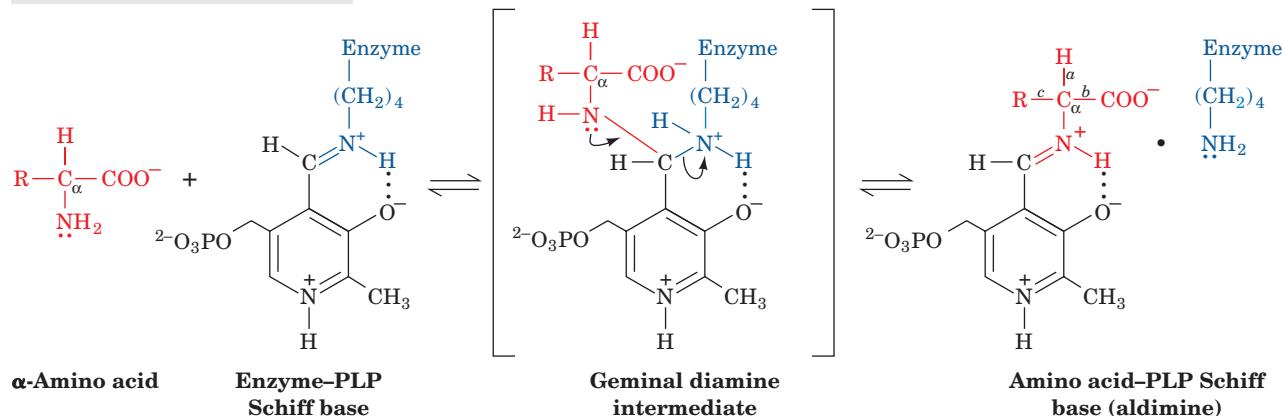
Figure 26-1 Forms of pyridoxal-5'-phosphate.

(a) Pyridoxine (vitamin B₆). (b) Pyridoxal-5'-phosphate (PLP).

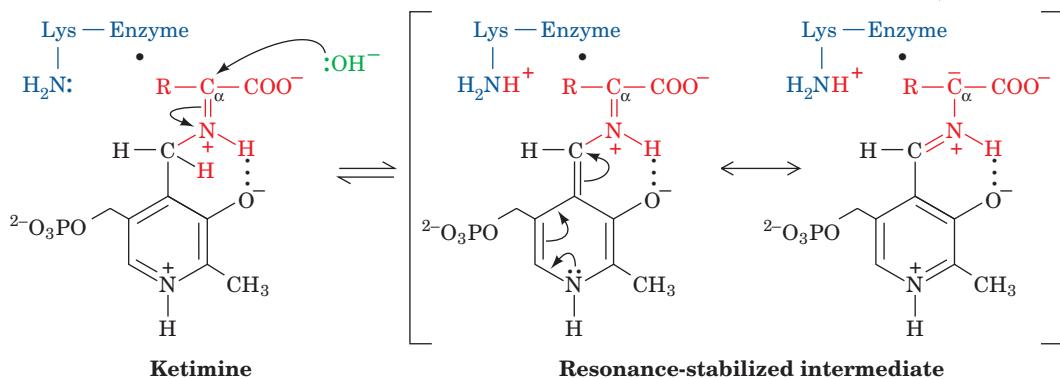
condensation of its aldehyde group with the ϵ -amino group of an enzymatic Lys residue (Fig. 26-1d). This Schiff base, which is conjugated to the coenzyme's pyridinium ring, is the focus of the coenzyme's activity.

Esmond Snell, Alexander Braunstein, and David Metzler demonstrated that the aminotransferase reaction occurs via a Ping Pong Bi Bi mechanism whose two stages consist of three steps each (Fig. 26-2):

Steps 1 & 1': Transimination:



Steps 2 & 2': Tautomerization:



Steps 3 & 3': Hydrolysis:

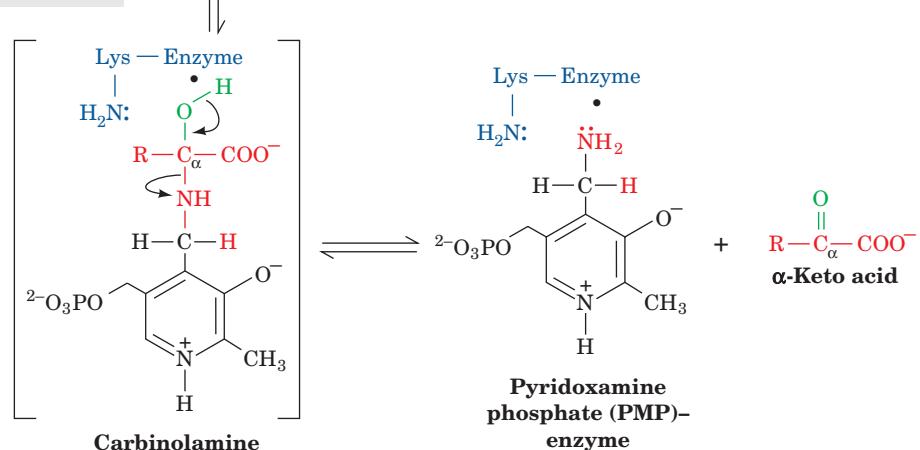


Figure 26-2 The mechanism of PLP-dependent enzyme-catalyzed transamination. The first stage of the reaction, in which the α -amino group of an amino acid is transferred to PLP yielding an α -keto acid and PMP, consists of three steps: (1) transimination; (2) tautomerization, in which the Lys released during the transimination reaction acts as a general acid–base

catalyst; and (3) hydrolysis. The second stage of the reaction, in which the amino group of PMP is transferred to a different α -keto acid to yield a new α -amino acid and PLP, is essentially the reverse of the first stage: Steps 3', 2', and 1' are, respectively, the reverse of Steps 3, 2, and 1. 

b. Stage I: Conversion of an Amino Acid to an α -Keto Acid

Step 1. The amino acid's nucleophilic amino group attacks the enzyme-PLP Schiff base carbon atom in a **trans-similation (trans-Schiffization)** reaction to form an amino acid-PLP Schiff base (aldimine), with concomitant release of the enzyme's Lys amino group. This Lys is then free to act as a general base at the active site.

Step 2. The amino acid-PLP Schiff base tautomerizes to an α -keto acid-PMP Schiff base by the active site Lys-catalyzed removal of the amino acid α hydrogen and protonation of PLP atom C4' via a resonance-stabilized carbanion intermediate. This resonance stabilization facilitates the cleavage of the C_{α} -H bond.

Step 3. The α -keto acid-PMP Schiff base is hydrolyzed to PMP and an α -keto acid.

c. Stage II: Conversion of an α -Keto Acid to an Amino Acid

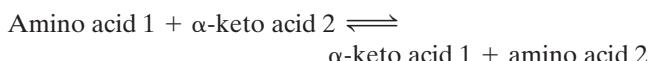
To complete the aminotransferase's catalytic cycle, the coenzyme must be converted from PMP back to the enzyme-PLP Schiff base. This involves the same three steps as above, but in reverse order:

Step 3'. PMP reacts with an α -keto acid to form a Schiff base.

Step 2'. The α -keto acid-PMP Schiff base tautomerizes to form an amino acid-PLP Schiff base.

Step 1'. The ϵ -amino group of the active site Lys residue attacks the amino acid-PLP Schiff base in a transimination reaction to regenerate the active enzyme-PLP Schiff base, with release of the newly formed amino acid.

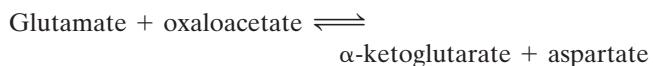
The reaction's overall stoichiometry therefore is



Examination of the amino acid-PLP Schiff base's structure (Fig. 26-2, Step 1) reveals why this system is called "an electron-pusher's delight." *Cleavage of any of the amino acid C_{α} atom's three bonds (labeled a, b, and c) produces a resonance-stabilized C_{α} carbanion whose electrons are*

delocalized all the way to the coenzyme's protonated pyridinium nitrogen atom; that is, PLP functions as an electron sink. For transamination reactions, this electron-withdrawing capacity facilitates removal of the α proton (*a* bond cleavage) in the tautomerization of the Schiff base. PLP-dependent reactions involving *b* bond cleavage (amino acid decarboxylation) and *c* bond labilization are discussed in Section 26-4B and in Sections 26-3Bb and 26-3G, respectively.

Aminotransferases differ in their specificity for amino acid substrates in the first stage of the transamination reaction, thereby producing the correspondingly different α -keto acid products. Most aminotransferases, however, accept only α -ketoglutarate or (to a lesser extent) oxaloacetate as the α -keto acid substrate in the second stage of the reaction, thereby yielding glutamate or aspartate as their only amino acid products. *The amino groups of most amino acids are consequently funneled into the formation of glutamate or aspartate, which are themselves interconverted by glutamate-aspartate aminotransferase:*



Oxidative deamination of glutamate (Section 26-1B) yields ammonia and regenerates α -ketoglutarate for another round of transamination reactions. Ammonia and aspartate are the two amino group donors in the synthesis of urea.

d. The Glucose-Alanine Cycle Transports Nitrogen to the Liver

An important exception to the foregoing is a group of muscle aminotransferases that accept pyruvate as their α -keto acid substrate. The product amino acid, alanine, is released into the bloodstream and transported to the liver, where it undergoes transamination to yield pyruvate for use in gluconeogenesis (Section 23-1A). The resulting glucose is returned to the muscles, where it is glycolytically degraded to pyruvate. This is the **glucose-alanine cycle** (Fig. 26-3). The amino group ends up in either ammonium ion or aspartate for urea biosynthesis. Evidently, the glucose-alanine cycle functions to transport nitrogen from muscle to liver.

During starvation the glucose formed in the liver by this route is also used by the other peripheral tissues, breaking the

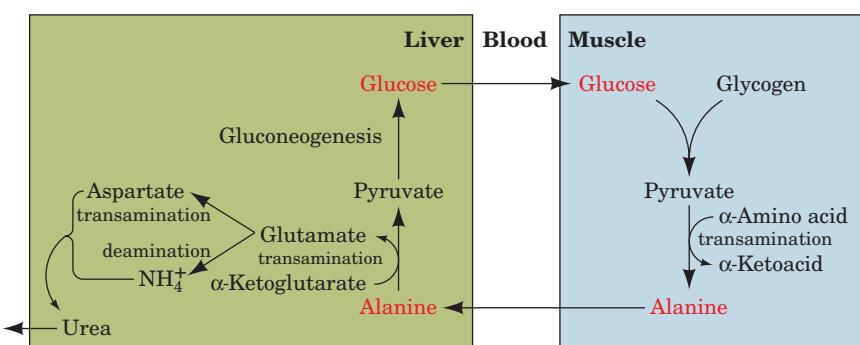


Figure 26-3 The glucose-alanine cycle.  See the Animated Figures

cycle. Under these conditions both the amino group and the pyruvate originate from muscle protein degradation, providing a pathway yielding glucose for other tissue use (recall that muscle is not a gluconeogenic tissue; Section 23-1).

Nitrogen is also transported to the liver in the form of glutamine, synthesized from glutamate and ammonia in a reaction catalyzed by **glutamine synthetase** (Section 26-5Ab). The ammonia is released for urea synthesis in liver mitochondria or for excretion in the kidney through the action of **glutaminase** (Section 26-3D).

B. Oxidative Deamination: Glutamate Dehydrogenase

Glutamate is oxidatively deaminated in the mitochondrial matrix by glutamate dehydrogenase (GDH), the only known enzyme that, in at least some organisms, can accept either NAD^+ or NADP^+ as its redox coenzyme. Oxidation is thought to occur with transfer of a hydride ion from glutamate's C_α to $\text{NAD}(\text{P})^+$, thereby forming α -iminoglutarate, which is hydrolyzed to α -ketoglutarate and ammonia (Fig. 26-4). GDH is allosterically inhibited by GTP, NADH, and nonpolar compounds such as palmitoyl-CoA and steroid hormones. It is activated by ADP, NAD^+ , and leucine (the most abundant amino acid in proteins; Table 4-1).

a. The X-Ray Structures of GDH Reveal its Allosteric Mechanism

The X-ray structures of homohexameric GDH from bovine and human liver mitochondria, determined by Thomas Smith, reveal that each monomer has three domains, a substrate domain, a coenzyme domain, and an antenna domain. The protein, which has D_3 symmetry, can be considered to be a dimer of trimers, with the antenna domains of each trimer wrapping around each other about the 3-fold axis (Fig. 26-5a). Structural comparison of a 501-residue

monomer of the bovine GDH–glutamate– NADH –GTP complex (Fig. 26-5b) with that of the 96% identical human **apoenzyme** (no active site or regulatory ligands bound; Fig. 26-5c) reveals that, on binding ligands, the coenzyme binding domain rotates about the so-called pivot helix so as to close the cleft between the coenzyme and substrate domains. Simultaneously, the antenna domain twists in a way that unwinds one turn of the antenna helix that is connected to the pivot helix. Although the closed form is required for catalysis, the open form favors the association and dissociation of substrates and products. In the open state, Arg 463 (human numbering) in the center of the pivot helix interacts with the activator ADP (whose binding site in the bovine complex is occupied by the ADP moiety of an NADH; Fig. 26-5b), whereas in the closed state, the side chain of His 454 hydrogen-bonds to the γ -phosphate of the inhibitor GTP. The GTP binding site is distorted and blocked in the open state so that GTP binding favors the closed form of the enzyme. This results in tight binding of substrates and products and hence inhibition of the enzyme. ADP binding favors the open form, allowing product dissociation, and therefore activates the enzyme. Allosteric interactions appear to be communicated between subunits through the interactions of the antenna domains. In fact, bacterial GDHs, which lack allosteric regulation, differ from mammalian GDHs mainly by the absence of antenna domains.

b. Hyperinsulinism/Hyperammonemia (HI/HA) Is Caused by Uncontrolled GDH Activity

Charles Stanley has reported a new form of congenital hyperinsulinism that is characterized by hypoglycemia and **hyperammonemia (HI/HA)** (hyperammonemia is elevated levels of ammonia in the blood) and has shown that it is caused by mutations in GDH at the N-terminal end of its pivot helix in the GTP binding site or in the antenna domain near its joint with the pivot helix. The mutant enzymes have reduced sensitivity to GTP inhibition but retain their ability to be activated by ADP. The GDH mutants S448P, H454Y, and R463A, which were respectively designed to affect the antenna region, the GTP binding site, and the ADP binding site (Fig. 26-5b), all have decreased sensitivity to GTP inhibition (Fig. 26-6), with H454Y and S448P, which were previously known to be associated with HI/HA, conferring the most resistance to GTP inhibition. The hypoglycemia and hyperammonemia in HI/HA patients arises from the increased activity of the GDH mutants in the breakdown direction, producing increased amounts of α -ketoglutarate and NH_3^+ . The increased levels of α -ketoglutarate stimulate the citric acid cycle and oxidative phosphorylation, which has been shown to lead to increased insulin secretion and hypoglycemia, thereby producing the symptoms of the disease. The NH_3^+ produced is usually converted to urea (Section 26-2) but can also be exported to the bloodstream.

If this scenario for the cause of HI/HA is correct, it requires a reassessment of the role of GDH in ammonia homeostasis. The equilibrium position of the GDH reaction greatly favors the synthesis of Glu ($\Delta G^\circ \approx 30 \text{ kJ} \cdot \text{mol}^{-1}$ for the reaction as written in Fig. 26-4), but studies of cellular

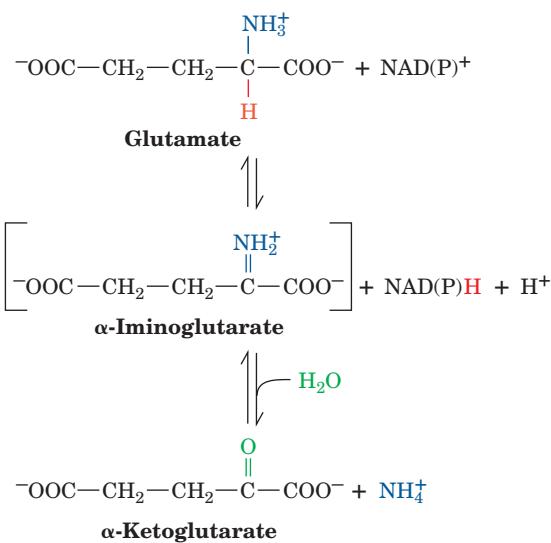


Figure 26-4 The oxidative deamination of glutamate by glutamate dehydrogenase. This reaction involves the intermediate formation of α -iminoglutarate.



(a)

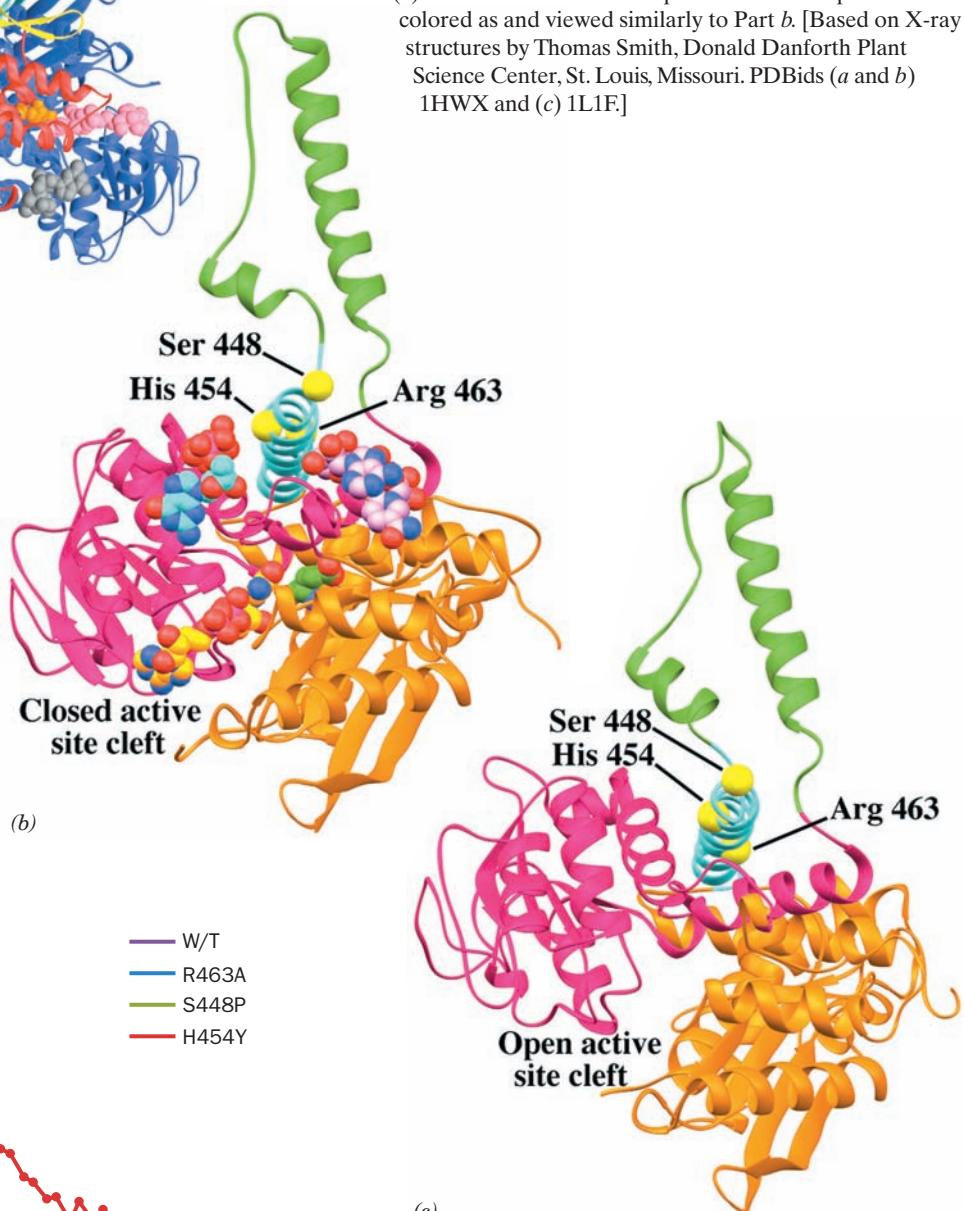
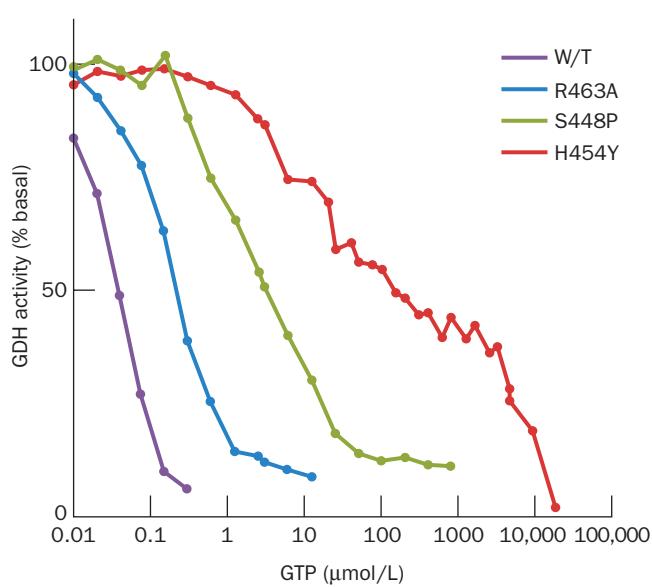


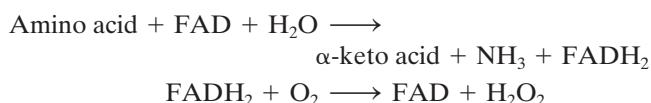
Figure 26-6 Inhibition of human glutamate dehydrogenase (GDH) by GTP. Human wild-type and mutant GDHs were expressed in *E. coli* and assayed for sensitivity to GTP inhibition. The midpoint of each curve corresponds to the concentration of GTP causing 50% inhibition. [After Fang, J., Hsu, B.Y.L., MacMullen, C.M., Poncz, M., Smith, T.J., and Stanley, C.A., *Biochem. J.* **363**, 81 (2002)].



substrate and product concentrations suggested that the enzyme functions close to equilibrium ($\Delta G \approx 0$) *in vivo*. It was therefore widely accepted that increases in $[\text{NH}_3]$, high levels of which are toxic, would cause GDH to act in reverse, removing NH_3 and hence preventing its buildup to toxic levels. However, since HI/HA patients have increased GDH activity yet have higher levels of NH_3 than normal, this accepted role of GDH cannot be correct. Indeed, if GDH functioned close to equilibrium, changes in its activity resulting from allosteric interactions would not result in significant flux changes.

C. Other Deamination Mechanisms

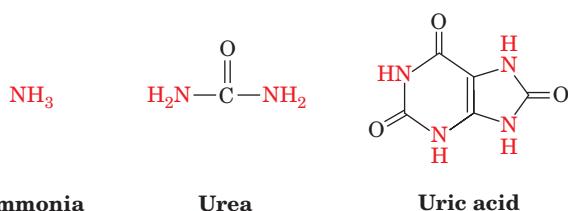
Two nonspecific amino acid oxidases, **L-amino acid oxidase** and **D-amino acid oxidase**, catalyze the oxidation of L- and D-amino acids, utilizing FAD as their redox coenzyme [rather than NAD(P)^+]. The resulting FADH_2 is reoxidized by O_2 .



D-Amino acid oxidase occurs mainly in kidney. Its function is an enigma since D-amino acids are associated mostly with bacterial cell walls (Section 11-3Ba). A few amino acids, such as serine and histidine, are deaminated nonoxidatively (Sections 26-3B and 26-3D).

2 THE UREA CYCLE

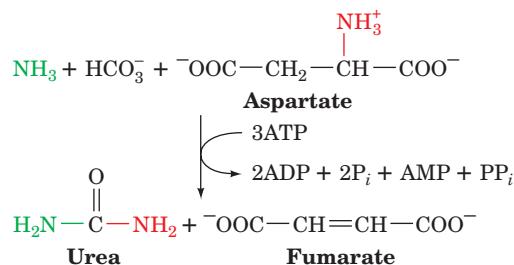
Living organisms excrete the excess nitrogen resulting from the metabolic breakdown of amino acids in one of three ways. Many aquatic animals simply excrete ammonia. Where water is less plentiful, however, processes have evolved that convert ammonia to less toxic waste products that therefore require less water for excretion. One such product is urea, which is excreted by most terrestrial vertebrates; another is **uric acid**, which is excreted by birds and terrestrial reptiles:



Accordingly, living organisms are classified as being either **ammonotelic** (ammonia excreting), **ureotelic** (urea excreting), or **uricotelic** (uric acid excreting). Some animals can shift from ammonotelism to ureotelism or uricotelism if their water supply becomes restricted. Here we focus our attention on urea formation. Uric acid biosynthesis is discussed in Section 28-4A.

Urea is synthesized in the liver by the enzymes of the urea cycle. It is then secreted into the bloodstream and

sequestered by the kidneys for excretion in the urine. The urea cycle was elucidated in outline in 1932 by Hans Krebs and Kurt Henseleit (the first known metabolic cycle; Krebs did not elucidate the citric acid cycle until 1937). Its individual reactions were later described in detail by Sarah Ratner and Philip Cohen. The overall urea cycle reaction is



Thus, the two urea nitrogen atoms are contributed by NH_3 and aspartate, whereas the carbon atom comes from HCO_3^- . Five enzymatic reactions are involved in the urea cycle, two of which are mitochondrial and three cytosolic (Fig. 26-7). In this section, we examine the mechanisms of these reactions and their regulation.

A. Carbamoyl Phosphate Synthetase: Acquisition of the First Urea Nitrogen Atom

Carbamoyl phosphate synthetase (CPS) is technically not a urea cycle enzyme. It catalyzes the condensation and activation of NH_3 and HCO_3^- to form **carbamoyl phosphate**, the first of the cycle's two nitrogen-containing substrates, with the concomitant hydrolysis of two ATPs. Eukaryotes have two forms of CPS:

1. Mitochondrial **CPS I** uses NH_3 as its nitrogen donor and participates in urea biosynthesis.
2. Cytosolic **CPS II** uses glutamine as its nitrogen donor and is involved in pyrimidine biosynthesis (Section 28-2A). The reaction catalyzed by CPS I involves three steps (Fig. 26-8):
 1. Activation of HCO_3^- by ATP to form **carboxyphosphate** and ADP.
 2. Nucleophilic attack of NH_3 on carboxyphosphate, displacing the phosphate to form **carbamate** and P_i .
 3. Phosphorylation of carbamate by the second ATP to form carbamoyl phosphate and ADP.

The reaction is essentially irreversible and is the rate-limiting step of the urea cycle. CPS I is subject to allosteric activation by **N-acetylglutamate** as is discussed in Section 26-2F.

E. coli contains only one type of CPS, which is homologous to both CPS I and CPS II. The enzyme is a heterodimer but when allosterically activated by ornithine (a urea cycle intermediate), it forms a tetramer of heterodimers, $(\alpha\beta)_4$. Its small subunit (382 residues) functions to hydrolyze glutamine and deliver the resulting NH_3 to its large subunit (1073 residues). However, if the enzyme's **glutaminase (glutamine**

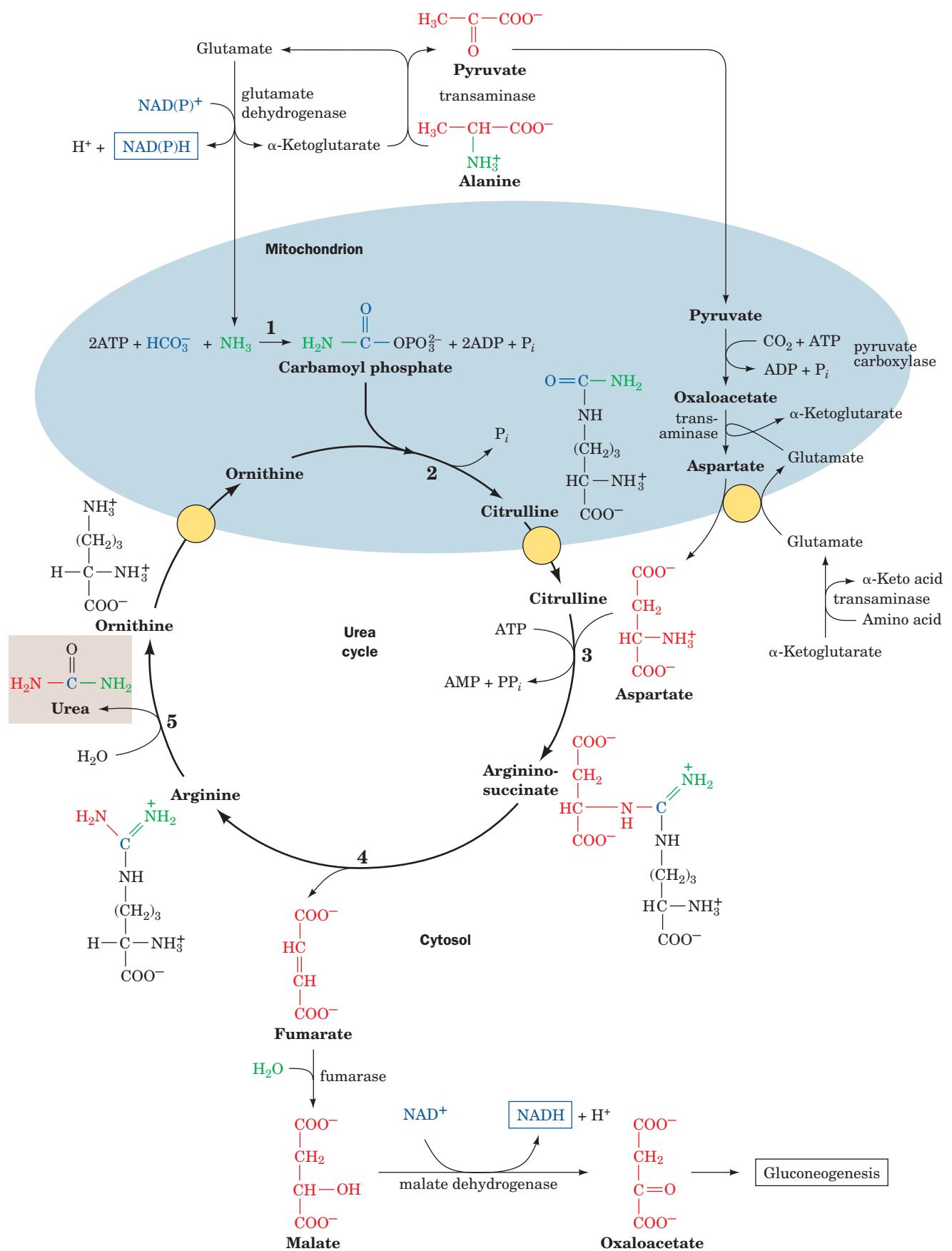


Figure 26-7 (Opposite) **The urea cycle.** Its five enzymes are (1) carbamoyl phosphate synthetase, (2) ornithine transcarbamoylase, (3) argininosuccinate synthetase, (4) argininosuccinase, and (5) arginase. The reactions occur in part in the mitochondrion and in part in the cytosol with ornithine and citrulline being transported across the mitochondrial membrane by specific transport systems (yellow circles). One of the urea amino groups (green) originates as the NH_3 product of the glutamate dehydrogenase reaction (top). The other amino group (red) is obtained from aspartate through the transfer of an amino acid to oxaloacetate via transamination (right). The fumarate product of the argininosuccinate reaction is converted to oxaloacetate for entry into gluconeogenesis via the same reactions that occur in the citric acid cycle but take place in the cytosol (bottom). The ATP utilized in Reactions 1 and 3 of the cycle can be regenerated by oxidative phosphorylation from the NAD(P)H produced in the glutamate dehydrogenase (top) and malate dehydrogenase (bottom) reactions. 

amidotransferase) activity is eliminated (e.g., by site-directed mutagenesis), the large subunit can still produce carbamoyl phosphate if NH_3 is supplied in high enough concentration. The large subunit is composed of two nearly

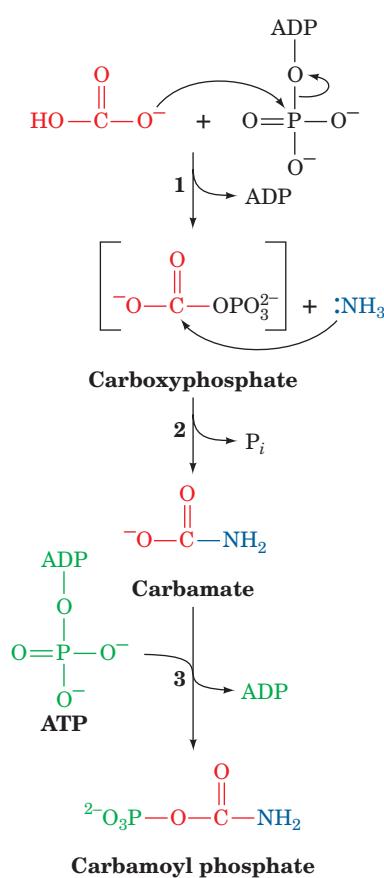


Figure 26-8 The mechanism of action of CPS I.

(1) Activation of HCO_3^- by phosphorylation forms the intermediate, carboxyphosphate; (2) nucleophilic attack on carboxyphosphate by NH_3 forms the reaction's second intermediate, carbamate; and (3) phosphorylation of carbamate by ATP yields the reaction product carbamoyl phosphate.

superimposable halves that have 40% sequence identity. The N-terminal half contains the carboxyphosphate synthetic component and an oligomerization domain while the C-terminal half contains the carbamoyl phosphate synthetic component and an allosteric binding domain.

a. *E. coli* CPS Contains an Extraordinarily Long Tunnel

The X-ray structure of *E. coli* CPS in complex with Mn^{2+} , ADP , P_i , and ornithine, determined by Hazel Holden and Ivan Rayment, reveals that the active site for synthesis of the carboxyphosphate intermediate is $\sim 45 \text{ \AA}$ away from the ammonia synthesis site and also $\sim 35 \text{ \AA}$ away from the carbamoyl phosphate synthesis active site. Astonishingly, the three sites are connected by a narrow 96- \AA -long molecular tunnel that runs nearly the length of the elongated protein molecule (Fig. 26-9). It therefore appears that CPS guides its intermediate products from the active site in which they are formed to that in which they are utilized. This phenomenon, in which the intermediate of two reactions is directly transferred from one enzyme active site to

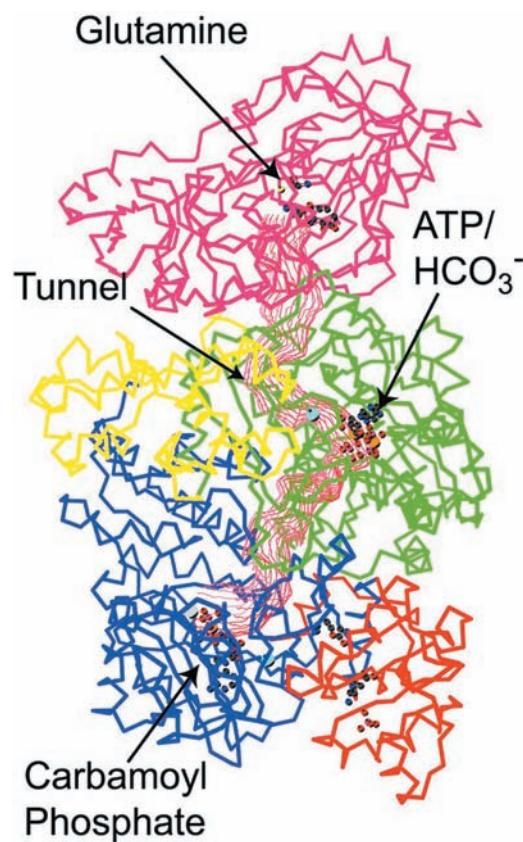


Figure 26-9 X-ray structure of *E. coli* carbamoyl phosphate synthetase (CPS). The protein is represented by its C_α backbone. The small subunit (magenta) contains the glutamine binding site where NH_3 is produced or bound. The large subunit consists of the carboxyphosphate domain (green), the oligomerization domain (yellow), the carbamoyl phosphate domain (blue), and the allosteric binding domain (orange). The 96- \AA -long tunnel connecting the three active sites is outlined in red. [Courtesy of Hazel Holden and Ivan Rayment, University of Wisconsin. PDBid 1JDB.]

another, is called **channeling** (the term “tunneling” is reserved for certain quantum mechanical phenomena).

Channeling increases the rate of a metabolic pathway by preventing the loss of its intermediate products as well as protecting the intermediate from degradation. NH_3 must travel $\sim 45 \text{ \AA}$ down the CPS tunnel to react with carboxyphosphate to form the next intermediate, carbamate. The carbamate, in turn, must travel an additional $\sim 35 \text{ \AA}$ to the site where it is phosphorylated by ATP to form the final product carbamoyl phosphate. The NH_3 transfer tunnel is lined with polar groups capable of forming hydrogen bonds with NH_3 , whereas the tunnel through which carbamate travels is lined with backbone atoms and lacks charged groups that might induce its hydrolysis as it diffuses between active sites. Shielding and channeling are necessary because the intermediates carboxyphosphate and carbamate are extremely reactive, having half-lives of 28 and 70 ms, respectively, at neutral pH. Also, channeling allows the local concentration of NH_3 to reach a higher value than is present in the cellular medium. We shall encounter several other examples of channeling in our studies of metabolic enzymes, but the CPS tunnel is far longer than that in any other known enzyme.

B. Ornithine Transcarbamoylase

Ornithine transcarbamoylase transfers the carbamoyl group of carbamoyl phosphate to **ornithine**, yielding **citrulline** (Fig. 26-7, Reaction 2; note that both of these compounds are “nonstandard” α -amino acids in that they do not occur in proteins). The reaction occurs in the mitochondrion so that ornithine, which is produced in the cytosol, must enter the mitochondrion via a specific transport system. Likewise, since the remaining urea cycle reactions occur in the cytosol, citrulline must be exported from the mitochondrion.

C. Argininosuccinate Synthetase: Acquisition of the Second Urea Nitrogen Atom

Urea’s second nitrogen atom is introduced in the urea cycle’s third reaction by the condensation of citrulline’s ureido group with an aspartate amino group by **argininosuccinate synthetase** (Fig. 26-10). The ureido oxygen atom is activated as a leaving group through formation of a

citrullyl-AMP intermediate, which is subsequently displaced by the aspartate amino group. Support for the existence of the citrullyl-AMP intermediate comes from experiments using ^{18}O -labeled citrulline (* in Fig. 26-10). The label was isolated in the AMP produced by the reaction, demonstrating that at some stage of the reaction, AMP and citrulline are linked covalently through the ureido oxygen atom.

D. Argininosuccinase

With formation of argininosuccinate, all of the urea molecule components have been assembled. However, the amino group donated by aspartate is still attached to the aspartate carbon skeleton. This situation is remedied by the **argininosuccinase**-catalyzed elimination of arginine from the aspartate carbon skeleton forming fumarate (Fig. 26-7, Reaction 4). Arginine is urea’s immediate precursor. The fumarate produced in the argininosuccinase reaction reacts via the fumarase and malate dehydrogenase reactions to form oxaloacetate (Fig. 26-7, bottom), which is then used in gluconeogenesis (Section 23-1).

E. Arginase

The urea cycle’s fifth and final reaction is the **arginase**-catalyzed hydrolysis of arginine to yield urea and regenerate ornithine (Fig. 26-7, Reaction 5). Ornithine is then returned to the mitochondrion for another round of the cycle. The urea cycle thereby converts two amino groups, one from NH_3 and one from aspartate, and a carbon atom from HCO_3^- to the relatively nontoxic excretion product urea at the cost of four “high-energy” phosphate bonds (three ATP hydrolyzed to two ADP, two P_i , AMP, and PP_i , followed by rapid PP_i hydrolysis). This energetic cost, together with that of gluconeogenesis, is supplied by the oxidation of the acetyl-CoA formed by the breakdown of amino acid carbon skeletons (e.g., threonine, Fig. 26-12). Indeed, half the oxygen that the liver consumes is used to provide this energy.

F. Regulation of the Urea Cycle

Carbamoyl phosphate synthetase I, the mitochondrial enzyme that catalyzes the first committed reaction of the urea

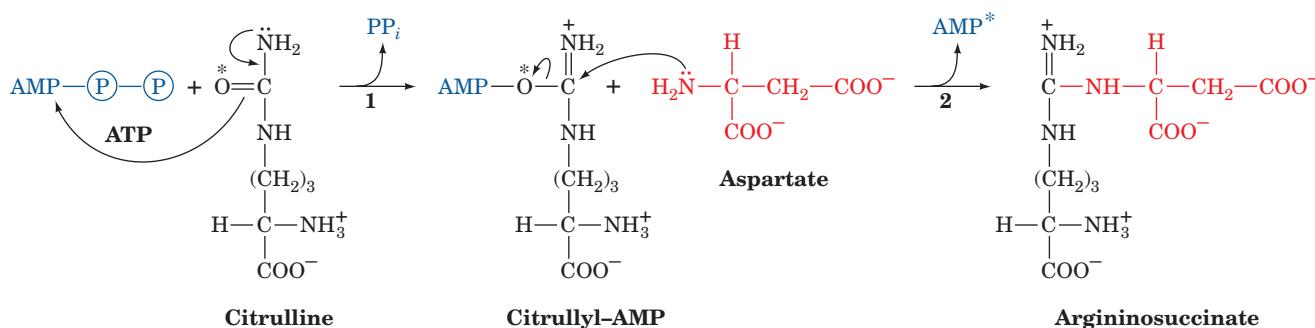
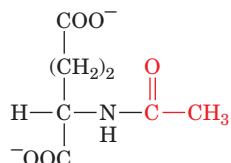


Figure 26-10 The mechanism of action of argininosuccinate synthetase. The steps involved are (1) activation of the ureido oxygen of citrulline through the formation of citrullyl-AMP and

(2) displacement of AMP by the α -amino group of aspartate. The asterisk (*) traces the fate of ^{18}O originating in citrulline’s ureido group.

cycle, is allosterically activated by ***N*-acetylglutamate**:



***N*-Acetylglutamate**

This metabolite is synthesized from glutamate and acetyl-CoA by ***N*-acetylglutamate synthase** and hydrolyzed by a specific hydrolase. The rate of urea production by the liver is, in fact, correlated with the *N*-acetylglutamate concentration. Increased urea synthesis is required when amino acid breakdown rates increase, generating excess nitrogen that must be excreted. Increases in these breakdown rates are signaled by an increase in glutamate concentration through transamination reactions (Section 26-1). This situation, in turn, causes an increase in *N*-acetylglutamate synthesis, stimulating carbamoyl phosphate synthetase and thus the entire urea cycle.

The remaining enzymes of the urea cycle are controlled by the concentrations of their substrates. Thus, inherited deficiencies in urea cycle enzymes other than arginase do not result in significant decreases in urea production (the total lack of any urea cycle enzyme results in death shortly after birth). Rather, the deficient enzyme's substrate builds up, increasing the rate of the deficient reaction to normal. The anomalous substrate buildup is not without cost, however. The substrate concentrations become elevated all the way back up the cycle to NH₃, resulting in hyperammonemia. Although the root cause of NH₃ toxicity is not completely understood, high [NH₃] puts an enormous strain on the NH₃-clearing system, especially in the brain (symptoms of urea cycle enzyme deficiencies include mental retardation and lethargy). This clearing system has been proposed to involve glutamate dehydrogenase (working in reverse) and **glutamine synthetase**, which decrease the α -ketoglutarate and glutamate pools (Sections 26-1 and 26-5Ab). The brain is most sensitive to the depletion of these pools. Depletion of α -ketoglutarate decreases the rate of the energy-generating citric acid cycle, whereas decreasing the glutamate concentration disturbs neuronal function, since it is both a neurotransmitter and a precursor to γ -aminobutyrate (GABA), another neurotransmitter (Section 20-5Cf). Glutamate depletion would also decrease the functioning of the urea cycle, since it is also the precursor to *N*-acetylglutamate, the major regulator of the cycle. The involvement of GDH in NH₃ clearance is a subject of debate in light of the observation that HI/HA involves deinhibition of GDH (Section 26-1Bb), suggesting that increased GDH activity increases the NH₃ concentration rather than decreasing it.

3 METABOLIC BREAKDOWN OF INDIVIDUAL AMINO ACIDS

The degradation of amino acids converts them to citric acid cycle intermediates or their precursors so that they can be metabolized to CO₂ and H₂O or used in gluconeogenesis.

Indeed, oxidative breakdown of amino acids typically accounts for 10 to 15% of the metabolic energy generated by animals. In this section we consider how amino acid carbon skeletons are catabolized. The 20 “standard” amino acids (the amino acids of proteins) have widely differing carbon skeletons, so their conversions to citric acid cycle intermediates follow correspondingly diverse pathways. We shall not describe all of the many reactions involved in detail. Rather, we shall consider how these pathways are organized and focus on a few reactions of chemical and/or medical interest.

A. Amino Acids Can Be Glucogenic, Ketogenic, or Both

“Standard” amino acids are degraded to one of seven metabolic intermediates: pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, oxaloacetate, acetyl-CoA, or acetoacetate (Fig. 26-11). The amino acids may therefore be divided into two groups based on their catabolic pathways (Fig. 26-11):

1. Glucogenic amino acids, whose carbon skeletons are degraded to pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, or oxaloacetate and are therefore glucose precursors (Section 23-1A).

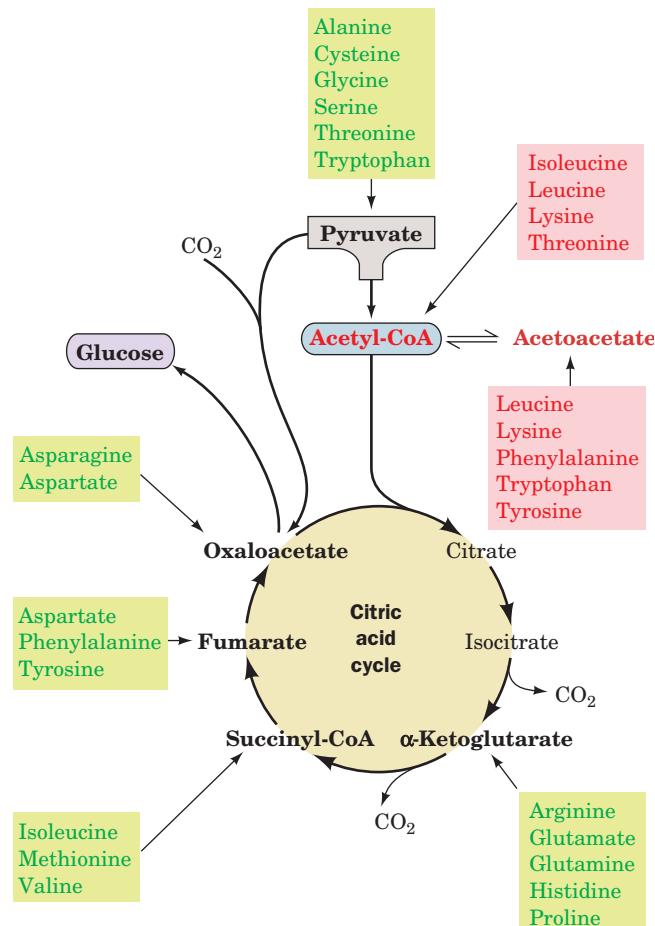


Figure 26-11 Degradation of amino acids to one of seven common metabolic intermediates. Glucogenic and ketogenic degradations are indicated in green and red, respectively.

2. Ketogenic amino acids, whose carbon skeletons are broken down to acetyl-CoA or acetoacetate and can thus be converted to ketone bodies or fatty acids (Sections 25-3 and 25-4).

For example, alanine is glucogenic because its transamination product, pyruvate (Section 26-1A), can be converted to glucose via gluconeogenesis (Section 23-1A). Leucine, on the other hand, is ketogenic; its carbon skeleton is converted to acetyl-CoA and acetoacetate (Section 26-3F). Since animals lack any metabolic pathway for the net conversion of acetyl-CoA or acetoacetate to gluconeogenic precursors, no net synthesis of carbohydrates is possible from leucine, or from lysine, the only other purely ketogenic amino acid. Isoleucine, phenylalanine, threonine, tryptophan, and tyrosine, however, are both glucogenic and ketogenic; isoleucine, for example, is broken down to succinyl-CoA and acetyl-CoA and hence is a precursor of both carbohydrates and ketone bodies (Section 26-3Ed). The remaining 13 amino acids are purely glucogenic.

In studying the specific pathways of amino acid breakdown, we shall organize the amino acids into groups that are degraded into each of the seven metabolic intermediates mentioned above: pyruvate, oxaloacetate, α -ketoglutarate, succinyl-CoA, fumarate, acetyl-CoA, and acetoacetate. When acetoacetyl-CoA is a product in amino acid degradation, it can, of course, be directly converted to acetyl-CoA (Section 25-3). We also discuss the pathway by which, in liver, it is converted instead to acetoacetate for use as an alternative fuel source in peripheral tissues (Section 25-3).

B. Alanine, Cysteine, Glycine, Serine, and Threonine Are Degraded to Pyruvate

Five amino acids, alanine, cysteine, glycine, serine, and threonine, are broken down to yield pyruvate (Fig. 26-12). Tryptophan should also be included in this group since one of its breakdown products is alanine (Section 26-3G), which, as we have seen (Section 26-1Ad), is transaminated to pyruvate.

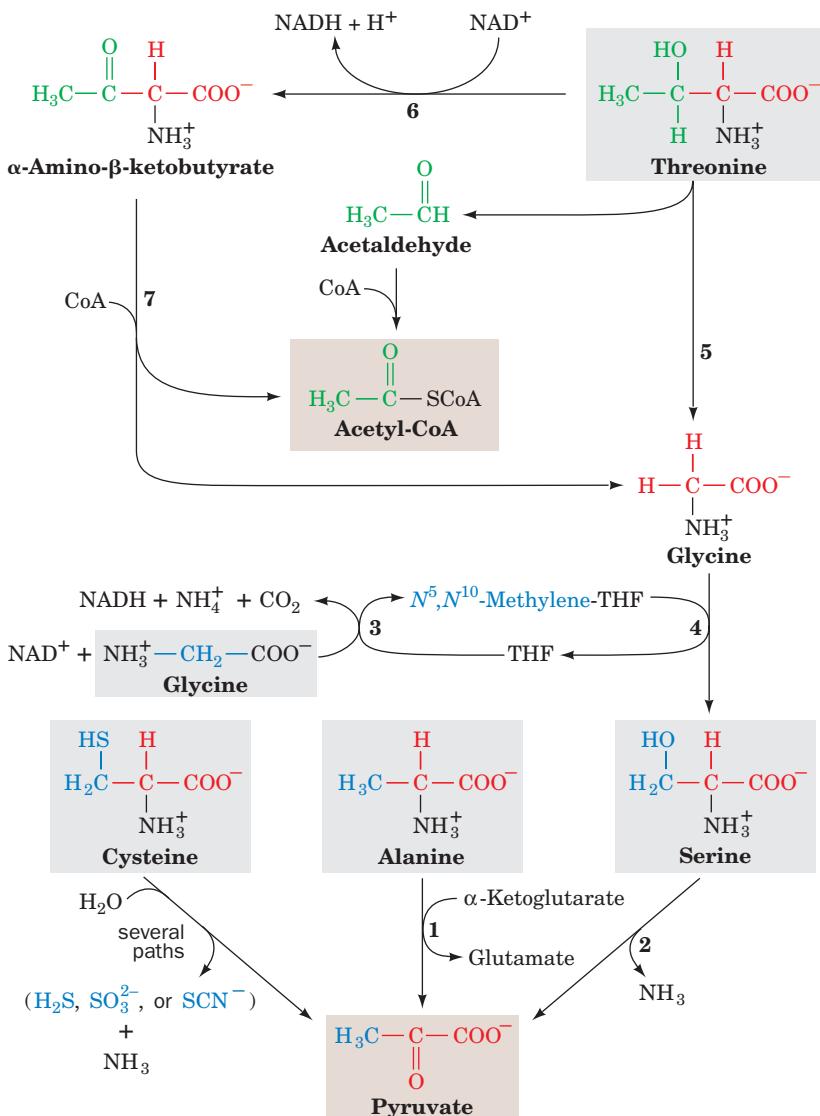


Figure 26-12 The pathways converting alanine, cysteine, glycine, serine, and threonine to pyruvate. The enzymes involved are (1) alanine aminotransferase, (2) serine dehydratase, (3) glycine cleavage system, (4) and (5) serine hydroxymethyltransferase, (6) threonine dehydrogenase, and (7) α -amino- β -ketobutyrate lyase.

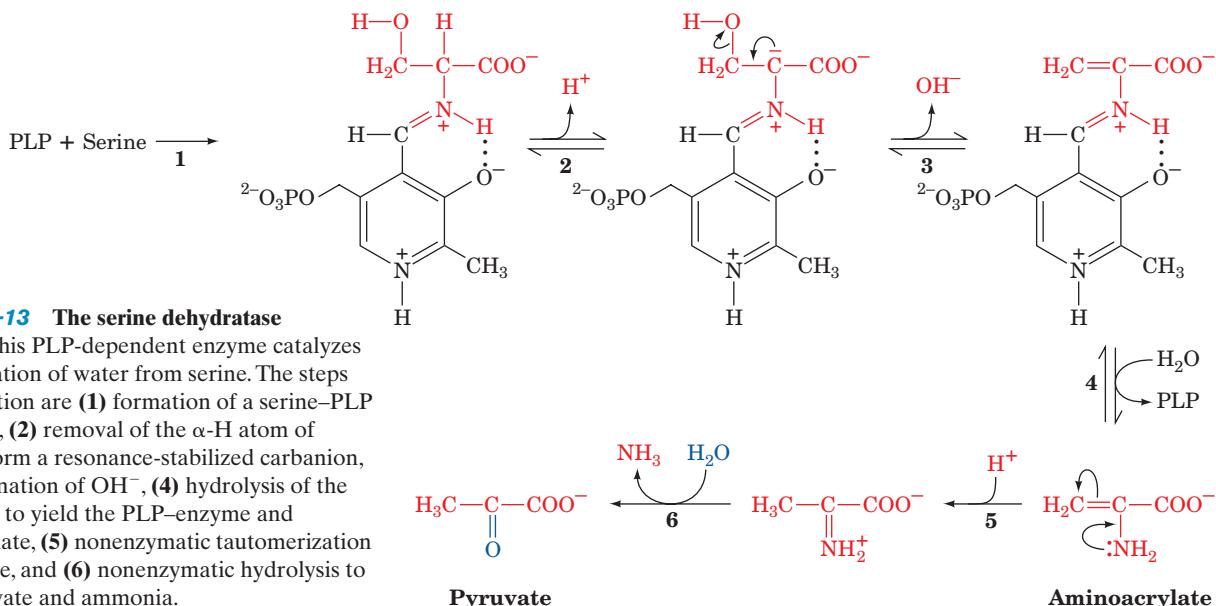


Figure 26-13 The serine dehydratase

reaction. This PLP-dependent enzyme catalyzes the elimination of water from serine. The steps in the reaction are (1) formation of a serine-PLP Schiff base, (2) removal of the α -H atom of serine to form a resonance-stabilized carbanion, (3) β elimination of OH^- , (4) hydrolysis of the Schiff base to yield the PLP-enzyme and aminoacrylate, (5) nonenzymatic tautomerization to the imine, and (6) nonenzymatic hydrolysis to form pyruvate and ammonia.

Serine is converted to pyruvate through dehydration by **serine dehydratase**. This PLP-enzyme, like the aminotransferases (Section 26-1), functions by forming a PLP-amino acid Schiff base, which facilitates the removal of the amino acid's α -hydrogen atom. In the serine dehydratase reaction, however, the C_α carbanion breaks down with the elimination of the amino acid's C_β OH, rather than with tautomerization (Fig. 26-2, Step 2), so that the substrate undergoes α,β elimination of H_2O rather than deamination (Fig. 26-13). The product of the dehydration, the enamine **aminoacrylate**, tautomerizes nonenzymatically to the corresponding imine, which spontaneously hydrolyzes to pyruvate and ammonia.

Cysteine may be converted to pyruvate via several routes in which the sulphydryl group is released as H_2S , SO_3^{2-} or SCN^- .

Glycine is converted to serine by the enzyme **serine hydroxymethyltransferase**, another PLP-containing enzyme (Fig. 26-12, Reaction 4). This enzyme utilizes N^5,N^{10} -methylene-tetrahydrofolate (N^5,N^{10} -methylene-THF) as a cofactor to provide the C_1 unit necessary for this conversion. We shall defer a detailed discussion of THF cofactors until Section 26-4D.

a. The Glycine Cleavage System Is a Multienzyme Complex

The methylene group of the N^5,N^{10} -methylene-THF utilized in the conversion of glycine to serine is obtained from the methylene group of a second glycine via a reaction in which this glycine's remaining atoms are released as CO_2 and NH_4^+ (Fig. 26-12, Reaction 3). This reaction is catalyzed by the **glycine cleavage system** (also called the **glycine decarboxylase multienzyme system** in plants and **glycine synthase** when acting in the reverse direction; Section 26-5Ae), a complex resembling the pyruvate dehydrogenase complex (Section 21-2A) that consists of four proteins (Fig. 26-14):

1. A PLP-dependent glycine decarboxylase (**P-protein**).
2. A lipoamide-containing aminomethyl carrier (**H-protein**), which carries the aminomethyl group remaining after glycine decarboxylation.

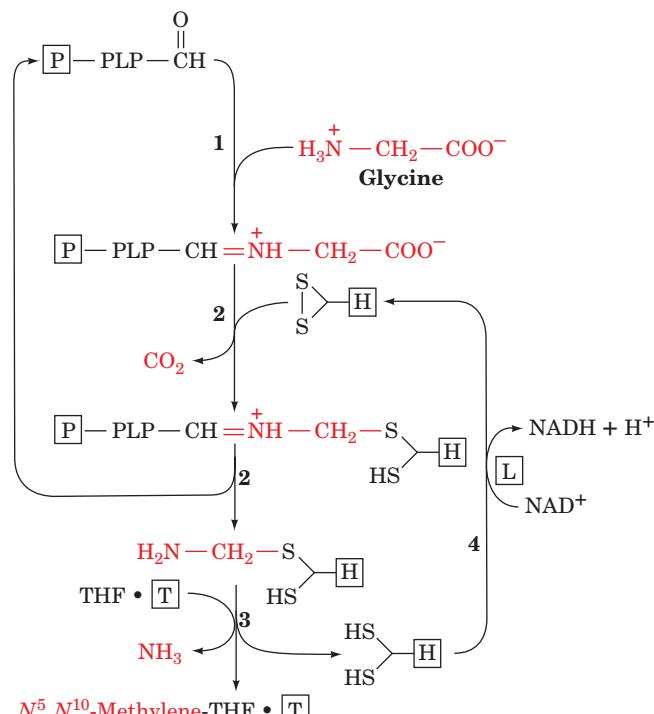


Figure 26-14 The reactions catalyzed by the glycine cleavage system, a multienzyme complex. The enzymes involved are (1) a PLP-dependent glycine decarboxylase (P-protein), (2) a lipoamide-containing protein (H-protein), (3) a THF-requiring enzyme (T-protein), and (4) an NAD^+ -dependent, FAD-requiring dihydrolipoil dehydrogenase (L-protein).

3. An N^5,N^{10} -methylene-THF synthesizing enzyme [T-protein; alternatively **aminomethyltransferase (AMT)**], which accepts a methylene group from the aminomethyl carrier (H-protein; the amino group is released as ammonia).

4. An NAD^+ -dependent, FAD-requiring dihydrolipoyl dehydrogenase (**L-protein**), a protein shared by and known as E3 in the pyruvate dehydrogenase complex (Section 21-2A).

Unlike the pyruvate dehydrogenase complex, the glycine cleavage system components are only loosely associated and hence are isolated as individual proteins. Nevertheless, H-protein has the central role in this multienzyme system: Its oxidized lipoyllysyl arm (Section 21-2Ad) is reduced as it accepts an aminomethyl group from P-protein (Fig. 26-14, Step 2), it donates the methylene group to THF in complex with T-protein as ammonia is released (Fig. 26-14, Step 3), and is then reoxidized by L-protein (Fig. 26-14, Step 4). The X-ray structure of pea leaf H-protein (Fig. 26-15), determined by Roland Douce, reveals that it is largely composed of a sandwich of 3-stranded and 6-stranded antiparallel β sheets that structurally resembles the lipoil domain of E2 in the pyruvate dehydrogenase complex (Fig. 21-10a).

The aminomethylthio group is unstable and ordinarily is rapidly hydrolyzed to formaldehyde and NH_4^+ . However, on its aminomethylation, the previously exposed lipoil group (Fig. 26-15a) inserts into a hydrophobic cleft

in the H-protein where its amino group hydrogen-bonds to Glu 14 (Fig. 26-15b), thereby shielding the aminomethyl group from hydrolysis. Indeed, the replacement of Glu 14 by Ala results in the rapid hydrolysis of the aminomethyl group. It therefore appears that the T-protein in the $\text{THF} \cdot \text{T} \cdot \text{H}$ complex functions to release the lipoil group from the H-protein cleft and to orient the THF for approach to the methylene C of the aminomethyl group for reaction.

Two observations indicate that the above pathway is the major route of glycine degradation in mammalian tissues:

1. The serine isolated from an animal that has been fed $[2-^{14}\text{C}]$ glycine is ^{14}C labeled at both C2 and C3. This observation indicates that the methylene group of the N^5,N^{10} -methylene-THF utilized by serine hydroxymethyltransferase is obtained from glycine C2.

2. The inherited human disease **nonketotic hyperglycinemia**, which is characterized by mental retardation and accumulation of large amounts of glycine in body fluids, results from the absence of one of the components of the glycine cleavage system.

The glycine cleavage system and serine hydroxymethyltransferase occupy a vital role in green leaves, catalyzing the rapid destruction of the huge amounts of glycine produced by photorespiration (Section 24-3Ca). In fact, these

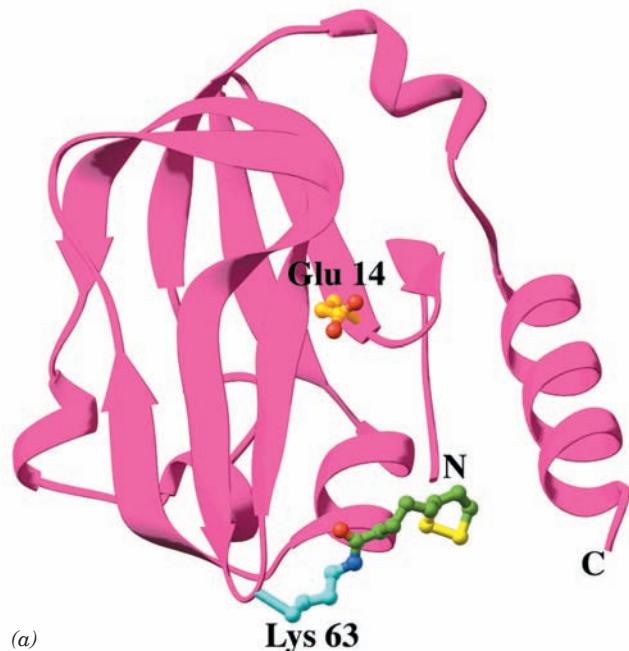
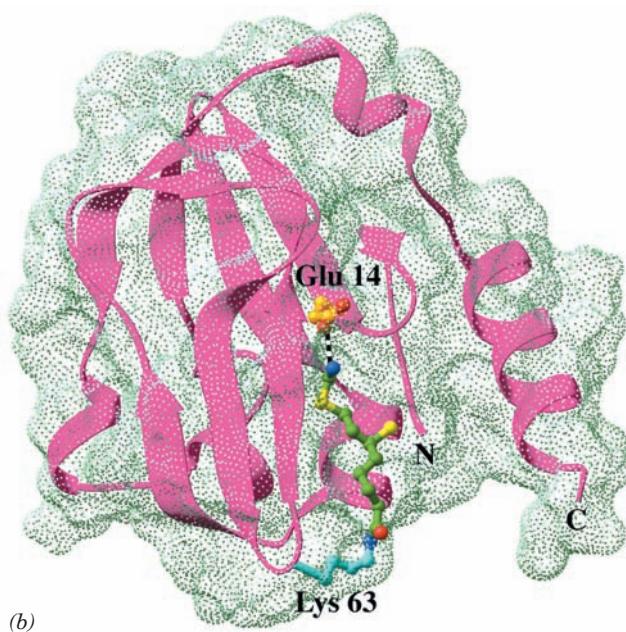


Figure 26-15 X-ray structure of H-protein from the pea leaf glycine cleavage system. (a) The oxidized lipoamide-containing form in which the side chain of Glu 14 together with that of Lys 63 with its covalently linked lipoil group are represented in ball-and-stick form colored according to atom type (Glu 14 C gold, Lys 63 C cyan, lipoil C green, N blue, O red, and S yellow). (b) The reduced aminomethyl-dihydrolipoamide form of H-protein viewed and colored as in Part a. The dot surface represents the protein's solvent-accessible surface. Note how the



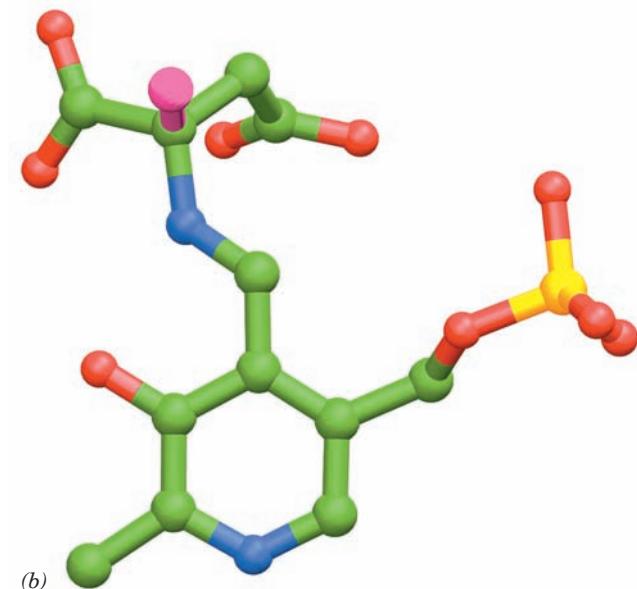
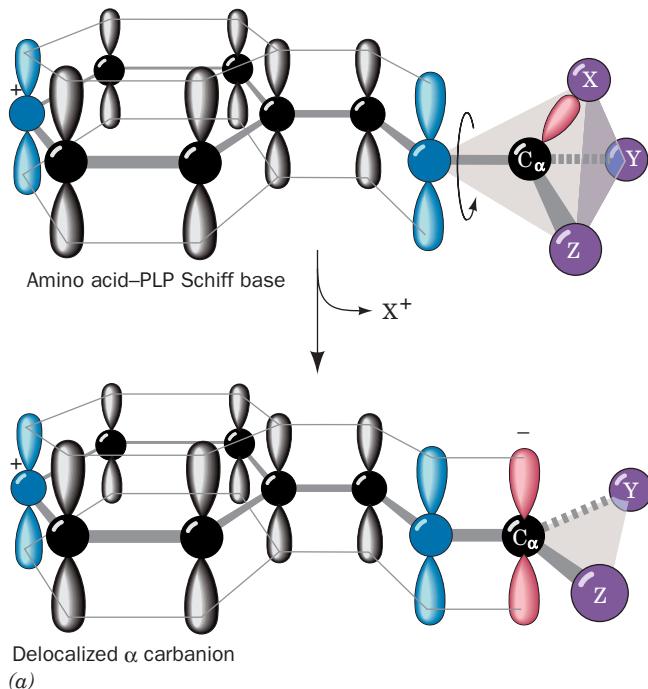
aminomethyl-dihydrolipoamide has changed conformation relative to the lipoamide in Part a so as to bind in a hydrophobic cleft in the protein where its amino group is hydrogen bonded to Glu 14 (dashed black bond). This protects the aminomethyl group from hydrolysis. [Based on X-ray structures by Roland Douce, Centre National de la Recherche Scientifique et Commissariat à l'Énergie Atomique, Grenoble, France. PDBids (a) 1HPC and (b) 1HTP.]

enzymes comprise about half the proteins present in the mitochondria from pea and spinach leaves.

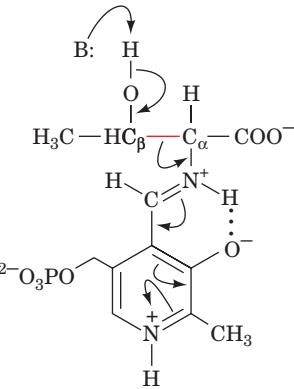
Threonine is both glucogenic and ketogenic, since one of its degradation routes produces both pyruvate and acetyl-CoA (Fig. 26-12, Reactions 6 and 7). Its major route of breakdown is through **threonine dehydrogenase**, producing **α -amino- β -ketobutyrate**, which is converted to acetyl-CoA and glycine by **α -amino- β -ketobutyrate lyase**. The glycine may be converted, through serine, to pyruvate.

b. Serine Hydroxymethyltransferase Catalyzes PLP-Dependent C_α — C_β Bond Cleavage

Threonine may also be converted directly to glycine and acetaldehyde (the latter being subsequently oxidized to



acetyl-CoA), at least *in vitro*, via Reaction 5 of Fig. 26-12. Surprisingly, this reaction is catalyzed by serine hydroxymethyltransferase. We have heretofore considered PLP-catalyzed reactions that begin with the cleavage of an amino acid's C_α —H bond (Fig. 26-2). Degradation of threonine to glycine and acetaldehyde by serine hydroxymethyltransferase demonstrates that PLP also facilitates cleavage of an amino acid's C_α — C_β bond by delocalizing the electrons of the resulting carbanion into the conjugated PLP ring:



c. PLP Facilitates the Cleavage of Different Bonds in Different Enzymes

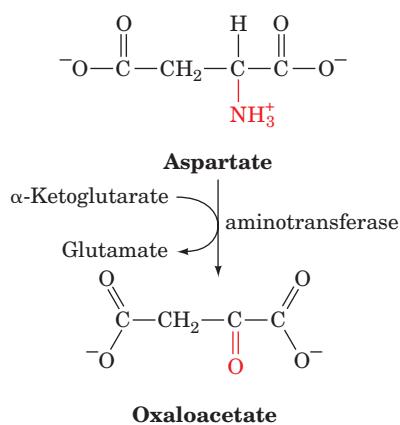
How can the same amino acid-PLP Schiff base be involved in the cleavage of the different bonds to an amino acid C_α in different enzymes? The answer to this conundrum was suggested by Harmon Dunathan. For electrons to be withdrawn into the conjugated ring system of PLP, the π -orbital system of PLP must overlap with the bonding orbital containing the electron pair being delocalized. This is possible only if the bond being broken lies in the plane perpendicular to the plane of the PLP π -orbital system (Fig. 26-16a). Different bonds to C_α can be placed in this plane by rotation about the C_α —N bond. Indeed, the X-ray structure of aspartate aminotransferase reveals that the C_α —H of its aspartate substrate assumes just this

Figure 26-16 Bond orientation in a PLP-amino acid Schiff base. (a) The π -orbital framework of a PLP-amino acid Schiff base. The bond to C_α in the plane perpendicular to the PLP π -orbital system (from X in the illustration) is labile as a consequence of its overlap with the π system, which permits the broken bond's electron pair to be delocalized over the conjugated molecule. (b) The Schiff base complex of the inhibitor **α -methylaspartate** with PLP in the X-ray structure of porcine aspartate aminotransferase as viewed normal to the pyridoxal ring. This inhibitor is drawn in ball-and-stick form with C green, N blue, O red, and P gold, with the exception that the methyl C atom and the bond linking it to the aspartate residue are magenta. Here the methyl C occupies the position of the H atom that the enzyme normally excises from aspartate. Note that the bond linking the methyl C to aspartate is in the plane perpendicular to the pyridoxal ring and is thus ideally oriented for bond cleavage. [Part b based on an X-ray structure by David Metzler and Arthur Arnone, University of Iowa. PDBid 1AJS.]

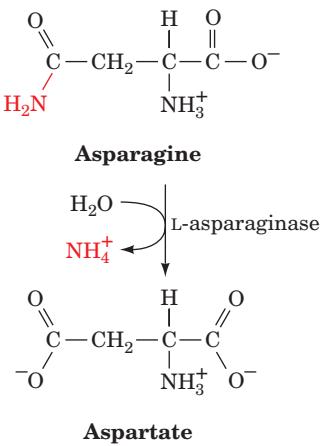
conformation (Fig. 26-16b). Evidently, each enzyme specifically cleaves its corresponding bond because the enzyme binds the amino acid-PLP Schiff base adduct with this bond in the plane perpendicular to that of the PLP ring. This is an example of stereoelectronic assistance (Section 15-1Eb): The enzyme binds substrate in a conformation that minimizes the electronic energy of the transition state.

C. Asparagine and Aspartate Are Degraded to Oxaloacetate

Transamination of aspartate leads directly to oxaloacetate:



Asparagine is also converted to oxaloacetate in this manner after its hydrolysis to aspartate by **L-asparaginase**:



Interestingly, L-asparaginase is an effective chemotherapeutic agent in the treatment of cancers that must obtain asparagine from the blood, particularly **acute lymphoblastic leukemia**. The cancerous cells express particularly low levels of the enzyme asparagine synthetase (Section 26-5Ab) and hence die without an external source of asparagine. However, L-asparaginase treatment may select for cells with increased levels of asparagine synthetase expression, and hence, in these cases, the surviving cancer cells are resistant to this treatment.

D. Arginine, Glutamate, Glutamine, Histidine, and Proline Are Degraded to α -Ketoglutarate

Arginine, glutamine, histidine, and proline are all degraded by conversion to glutamate (Fig. 26-17), which in turn is oxidized to α -ketoglutarate by glutamate dehydrogenase (Section 26-1). Conversion of glutamine to glutamate involves only one reaction: hydrolysis by **glutaminase**. Histidine's conversion to glutamate is more complicated: It is nonoxidatively deaminated, then it is hydrated, and its imidazole ring is cleaved to form **N-formiminoglutamate**. The formimino group is then transferred to tetrahydrofolate forming glutamate and **N⁵-formimino-tetrahydrofolate** (Section 26-4D). Both arginine and proline are converted to glutamate through the intermediate formation of **glutamate-5-semialdehyde**.

E. Isoleucine, Methionine, and Valine Are Degraded to Succinyl-CoA

Isoleucine, methionine, and valine have complex degradative pathways that all yield propionyl-CoA. Propionyl-CoA, which is also a product of odd-chain fatty acid degradation, is converted, as we have seen, to succinyl-CoA by a series of reactions involving the participation of biotin and coenzyme B₁₂ (Section 25-2E).

a. Methionine Breakdown Involves Synthesis of S-Adenosylmethionine and Cysteine

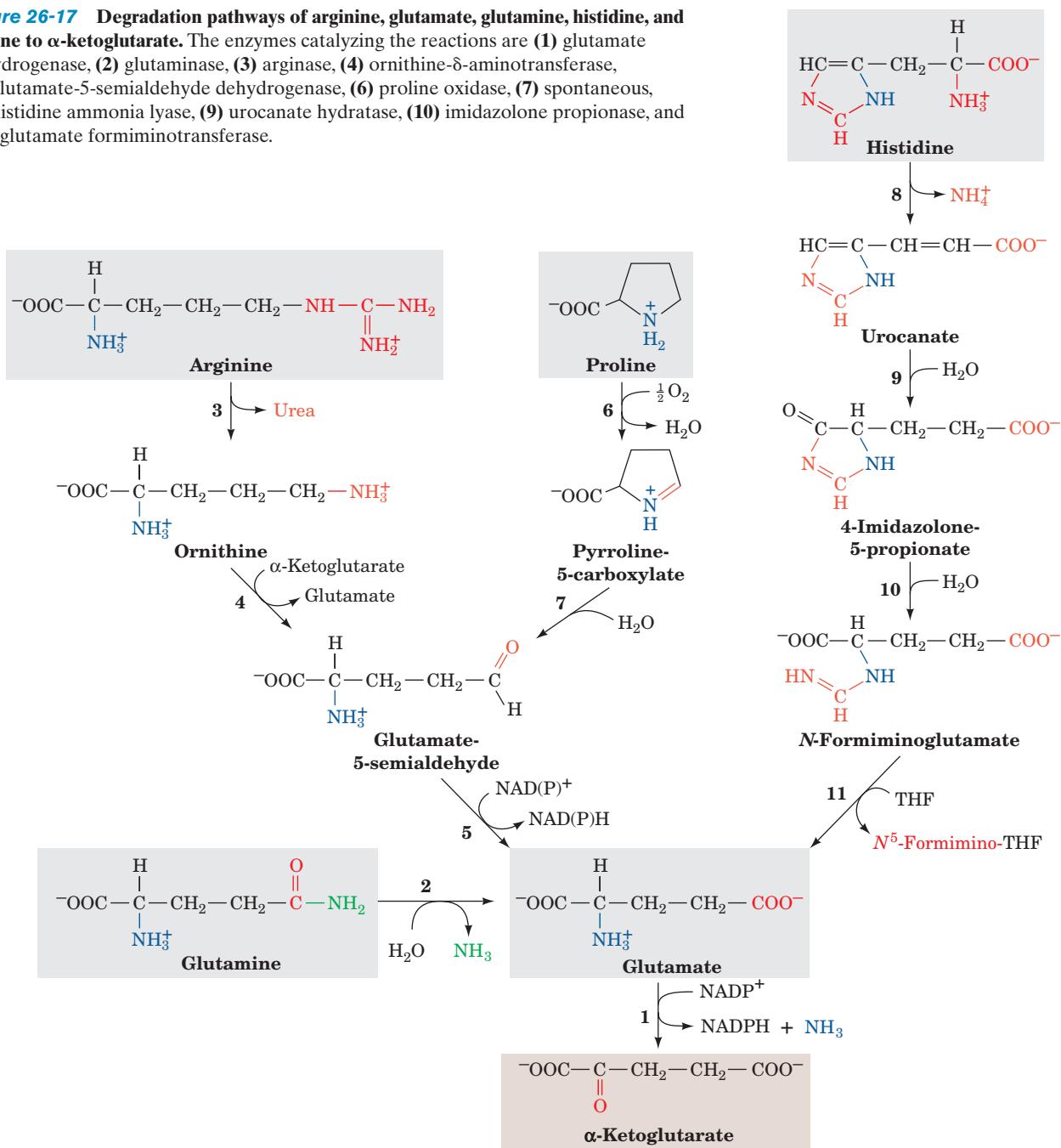
Methionine degradation (Fig. 26-18) begins with its reaction with ATP to form **S-adenosylmethionine (SAM; alternatively AdoMet)**. This sulfonium ion's highly reactive methyl group makes it an important biological methylating agent. For instance, we have already seen that SAM is the methyl donor in the synthesis of phosphatidylcholine from phosphatidylethanolamine (Section 25-8Aa). It is also the methyl donor in the conversion of norepinephrine to epinephrine (Section 26-4B).

Methylation reactions involving SAM yield **S-adenosylhomocysteine** in addition to the methylated acceptor. The former product is hydrolyzed to adenosine and **homocysteine** in the next reaction of the methionine degradation pathway. The homocysteine may be methylated to form methionine via a B₁₂-requiring reaction in which **N⁵-methyl-THF** is the methyl donor. Alternatively, the homocysteine may combine with serine to yield **cystathionine** in a PLP-requiring reaction, which subsequently forms cysteine (cysteine biosynthesis) and **α -ketobutyrate**. The α -ketobutyrate continues along the degradative pathway to propionyl-CoA and then succinyl-CoA.

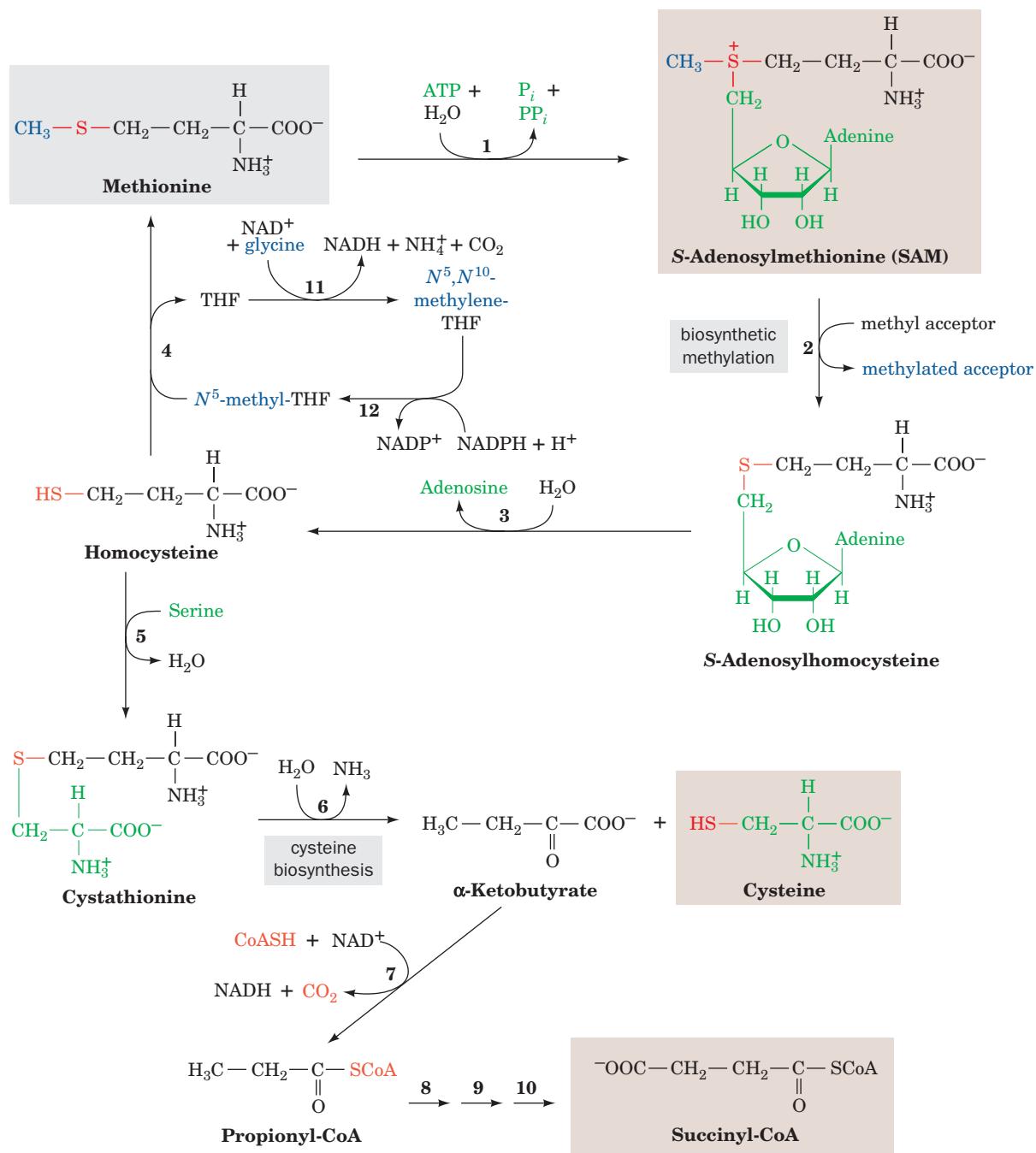
b. Hyperhomocysteinemia Is Associated with Disease

Imbalance between the rate of production of homocysteine through methylation reactions utilizing SAM (Fig. 26-18, Reactions 2 and 3) and its rate of breakdown by either remethylation to form methionine (Fig. 26-18, Reaction 4) or reaction with serine to form cystathionine in the

Figure 26-17 Degradation pathways of arginine, glutamate, glutamine, histidine, and proline to α -ketoglutarate. The enzymes catalyzing the reactions are (1) glutamate dehydrogenase, (2) glutaminase, (3) arginase, (4) ornithine- δ -aminotransferase, (5) glutamate-5-semialdehyde dehydrogenase, (6) proline oxidase, (7) spontaneous, (8) histidine ammonia lyase, (9) urocanate hydratase, (10) imidazolone propionase, and (11) glutamate formiminotransferase.



cysteine biosynthesis pathway (Fig. 26-18, Reaction 5) can result in an increase in the release of homocysteine to the extracellular medium and ultimately the plasma and urine. Moderately elevated concentrations of homocysteine in the plasma, **hyperhomocysteinemia**, for reasons that are poorly understood, are closely associated with cardiovascular disease, cognitive impairment, and **neural tube defects** [the cause of a variety of severe birth defects including **spina bifida** (defects in the spinal column that often result in paralysis) and **anencephaly** (the invariably fatal failure of the brain to develop, which is the leading cause of infant death due to congenital anomalies)]. Hyperhomocysteinemia is readily controlled by ingesting the vitamin precursors of the coenzymes that participate in homocysteine breakdown, namely, B_6 (pyridoxine, the PLP precursor; Fig. 26-1), B_{12} (Fig. 25-21), and folate (Section 26-4D). Folate, especially, appears to alleviate hyperhomocysteinemia; its administration to pregnant women dramatically reduces the incidence of neural tube defects in newborns. This has led to the discovery that 10% of the population is homozygous for the A222V mutation in **N^5,N^{10} -methylene-tetrahydrofolate reductase (MTHFR)** (Fig. 26-18, Reaction 12; Section 26-4D), the enzyme that generates N^5 -methyl-THF for the methionine synthase reaction (Fig. 26-18,



Reaction 4). This mutation does not affect this homotetrameric enzyme's reaction kinetics but instead increases the rate at which it dissociates into dimers that readily lose their essential flavin cofactor. Folate derivatives that bind to the enzyme decrease its rate of dissociation and

acid dehydrogenase, (8) propionyl-CoA carboxylase, (9) methylmalonyl-CoA racemase, (10) methylmalonyl-CoA mutase (a coenzyme B₁₂-dependent enzyme; Reactions 8–10 are discussed in Section 25-2E), (11) glycine cleavage system (Figs. 26-12 and 26-14) or serine hydroxymethyltransferase (Fig. 26-12), (12) *N*⁵,*N*¹⁰-methylene-tetrahydrofolate reductase (a coenzyme B₁₂- and FAD-dependent enzyme; Figs. 26-19 and 26-49).

flavin loss, thus increasing the mutant enzyme's overall activity and decreasing the homocysteine concentration.

The X-ray structure of *E. coli* MTHFR (which is 30% identical with the catalytic domain of human MTHFR), determined by Rowena Matthews and Martha Ludwig,

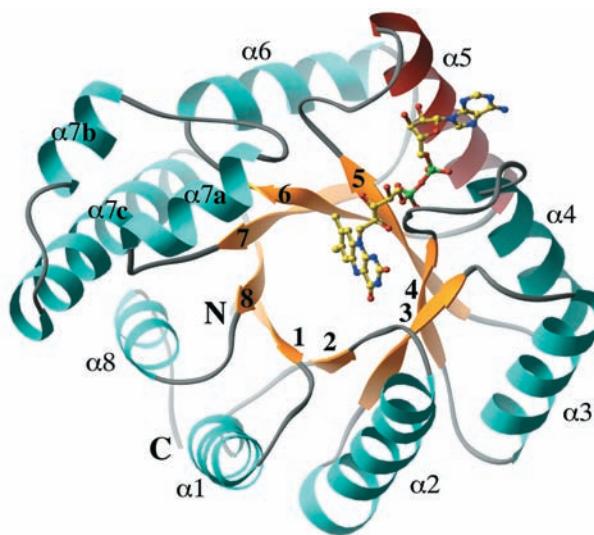


Figure 26-19 X-ray structure of *E. coli* N^5,N^{10} -methylene-tetrahydrofolate reductase (MTHFR). The structure is viewed along the axis of its α/β barrel looking toward the C-terminal ends of its β strands. The protein is colored according to its secondary structure with β strands yellow and α helices cyan except for helix $\alpha 5$, which is red. The enzyme's bound FAD is drawn in ball-and-stick form with C yellow, N blue, O red, and P green. Note that the AMP moiety of the FAD is in contact with helix $\alpha 5$. [Courtesy of Rowena Matthews and Martha Ludwig, University of Michigan. PDBid 1B5T.]

reveals that this 296-residue enzyme forms an α/β barrel. The FAD cofactor binds at the C-terminal ends of barrel strands $\beta 3$, $\beta 4$, and $\beta 5$ and along helix $\alpha 5$ (Fig. 26-19). Ala 177, which corresponds to Ala 222 in the mammalian enzyme, does not interact directly with the active site FAD. Instead it occupies a position flush against helix $\alpha 5$ (which ends with residue 176). It is postulated that the replacement of Ala 177 by a bulkier Val residue would force helix $\alpha 5$ to reorient. Since this helix appears to be involved in the subunit interface as well as in FAD binding, its reorientation is likely to decrease the strength of subunit and FAD interactions.

Why should this mutation be so prevalent in the human population? What selective advantage, if any, might it confer? We have seen that the gene for sickle-cell anemia provides a selective advantage against malaria (Section 7-3Ab). However, the selective advantage of the A222V mutation in human MTHFR is as yet a matter of speculation.

c. Methionine Synthase Is a Coenzyme

B_{12} -Dependent Enzyme

Methionine synthase (alternatively **homocysteine methyltransferase**), the enzyme that catalyzes Reaction 4 in Fig. 26-18, is the only coenzyme B_{12} -associated enzyme in mammals besides methylmalonyl-CoA mutase (Section 25-2Eb). However, in methionine synthase, the cobalamin Co ion is axially liganded by a methyl group forming **methylcobalamin** rather than by a 5'-deoxyadenosyl group as occurs in methylmalonyl-CoA mutase (Fig. 25-21). This

is because the cobalamin functions to accept the methyl group from N^5 -methyl-THF to yield methylcobalamin (and THF), which, in turn, donates the methyl group to homocysteine to yield methionine.

The X-ray structure of the 246-residue methylcobalamin-binding portion of the 1227-residue monomeric *E. coli* methionine synthase, also determined by Matthews and Ludwig, reveals that it consists of two domains, an N-terminal helical domain and a C-terminal Rossmann fold-like α/β domain, with the corrin ring sandwiched between them (Fig. 26-20). The α/β domain resembles the corrin-binding α/β domain in methylmalonyl-CoA mutase (Fig. 25-22) and, in fact, sequence homologies suggest that this domain is a common binding motif in B_{12} -associated enzymes. The Co ion's second axial ligand is a His side chain as is also the case in methylmalonyl-CoA mutase; the coenzyme's 5,6-dimethylbenzimidazole (DMB) moiety, which ligands the Co ion in free methylcobalamin, has swung aside to become anchored to the protein at some distance from the corrin ring.

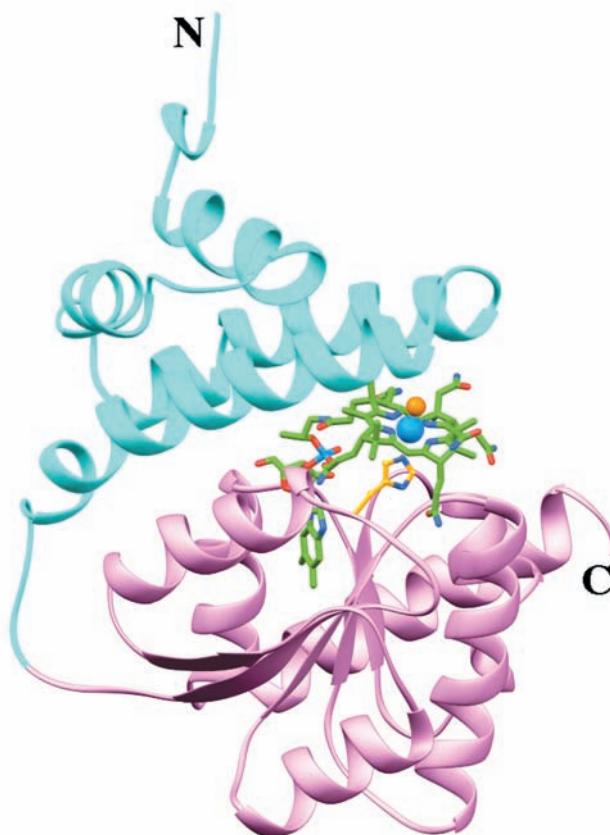


Figure 26-20 X-ray structure of the B_{12} -binding domains of *E. coli* methionine synthase. Its N-terminal helical domain (residues 651–743) is cyan and its C-terminal α/β domain (residues 744–896) is pink. The methylcobalamin cofactor and its axially liganded His 759 side chain are drawn in stick form with cobalamin C green, His C gold, N blue, O red, and the Co ion and its axially liganded methyl group represented by light blue and orange spheres, respectively. [Based on an X-ray structure by Rowena Matthews and Martha Ludwig, University of Michigan. PDBid 1BMT.]

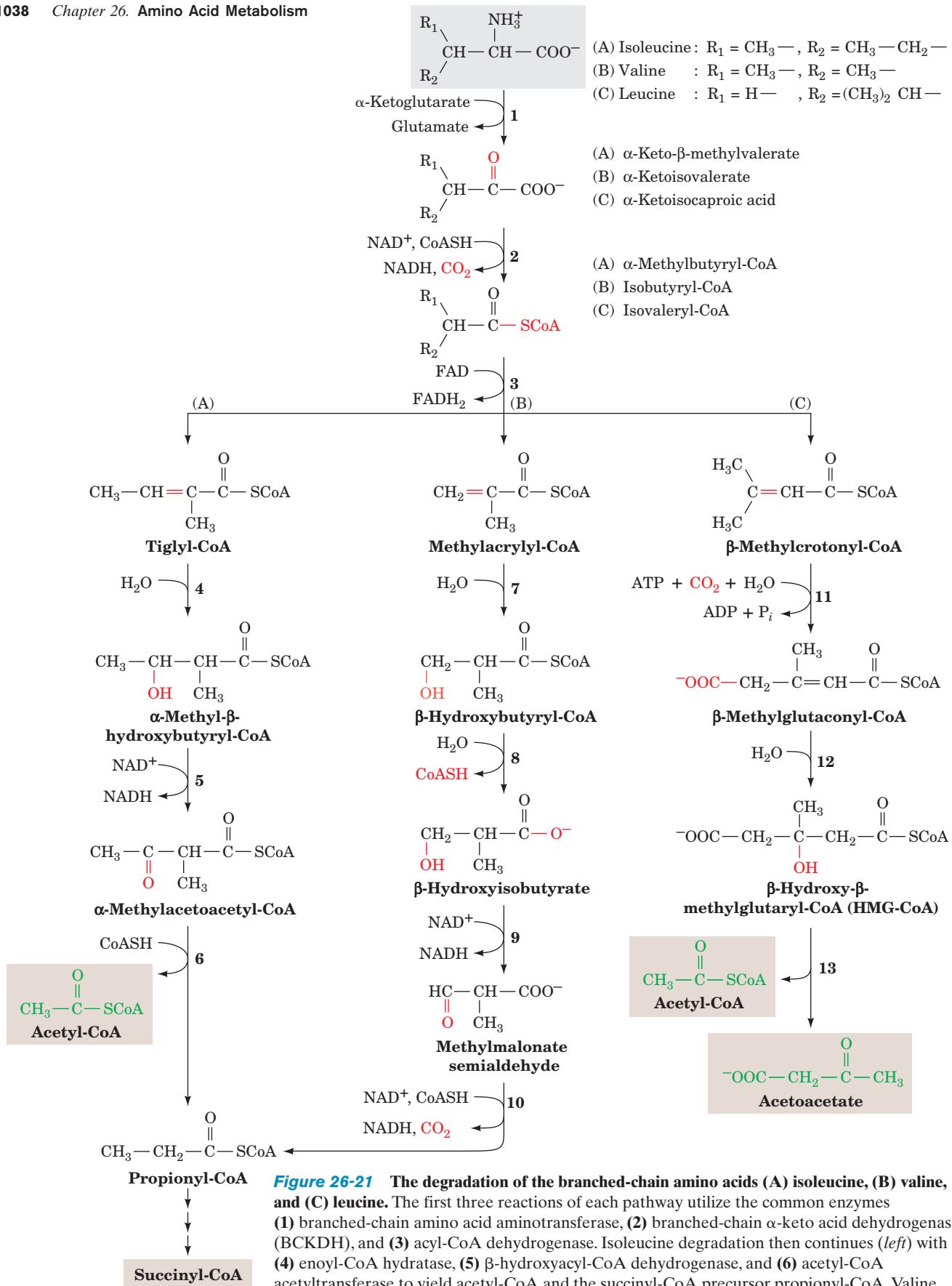


Figure 26-21 The degradation of the branched-chain amino acids (A) isoleucine, (B) valine, and (C) leucine. The first three reactions of each pathway utilize the common enzymes (1) branched-chain amino acid aminotransferase, (2) branched-chain α-keto acid dehydrogenase (BCKDH), and (3) acyl-CoA dehydrogenase. Isoleucine degradation then continues (left) with (4) enoyl-CoA hydratase, (5) β-hydroxyacyl-CoA dehydrogenase, and (6) acetyl-CoA acetyltransferase to yield acetyl-CoA and the succinyl-CoA precursor propionyl-CoA. Valine degradation (center) continues with (7) enoyl-CoA hydratase, (8) β-hydroxyisobutyrate-CoA hydrolase, (9) β-hydroxyisobutyrate dehydrogenase, and (10) methylmalonate semialdehyde dehydrogenase to also yield propionyl-CoA. Leucine degradation (right) continues with (11) β-methylcrotonyl-CoA carboxylase (a biotin-dependent enzyme), (12) β-methylglutaconyl-CoA hydratase, and (13) HMG-CoA lyase to yield acetyl-CoA and acetoacetate.

d. Branched-Chain Amino Acid Degradation Pathways Contain Themes Common to All Acyl-CoA Oxidations

Degradation of the branched-chain amino acids isoleucine, leucine, and valine begins with three reactions that employ common enzymes (Fig. 26-21, *top*): (1) transamination to the corresponding α -keto acid, (2) oxidative decarboxylation to the corresponding acyl-CoA, and (3) dehydrogenation by FAD to form a double bond.

The remainder of the isoleucine degradation pathway (Fig. 26-21, *left*) is identical to that of fatty acid oxidation (Section 25-2C): (4) double-bond hydration, (5) dehydrogenation by NAD^+ , and (6) thiolysis cleavage yielding acetyl-CoA and propionyl-CoA, which is subsequently converted to succinyl-CoA. Valine degradation is a variation on this theme (Fig. 26-21, *center*): Following (7) double-bond hydration, (8) the CoA thioester bond is hydrolyzed before (9) the second dehydrogenation reaction. The thioester bond is then regenerated as propionyl-CoA in the sequence's last reaction (10), an oxidative decarboxylation rather than a thiolysis cleavage.

e. Maple Syrup Urine Disease Results from a Defect in Branched-Chain Amino Acid Degradation

Branched-chain α -keto acid dehydrogenase (BCKDH; also known as α -ketoisovalerate dehydrogenase), which catalyzes Reaction 2 of branched-chain amino acid degradation (Fig. 26-21), is a multienzyme complex containing three enzymatic components, E1, E2, and E3, together with **BCKDH kinase** (phosphorylation inactivates) and **BCKDH phosphatase** (dephosphorylation activates), which impart control by covalent modification. This complex closely resembles the pyruvate dehydrogenase and α -ketoglutarate dehydrogenase multienzyme complexes (Sections 21-2A and 21-3D). Indeed, all three of these multienzyme complexes share a common protein component, E3 (dihydrolipoyl dehydrogenase), and employ the coenzymes thiamine pyrophosphate (TPP), lipoamide, and FAD in addition to their terminal oxidizing agent, NAD^+ .

A genetic deficiency in BCKDH causes **maple syrup urine disease (MSUD)**, so named because the consequent buildup of branched-chain α -keto acids imparts the urine with the characteristic odor of maple syrup. Unless promptly treated by a diet low in branched-chain amino acids (but not too low because they are essential amino acids; Section 26-5), MSUD is rapidly fatal.

MSUD is an autosomal recessive disorder that is caused by defects in any of four of the complex's six subunits, E1 α , E1 β , E2, or E3 (E1 is an $\alpha_2\beta_2$ heterotetramer). The determination of the X-ray structure of human BCKDH E1 by Wim Hol (Fig. 26-22) has enabled the interpretation of several of the mutations causing MSUD. The most common mutation is Y393N- α , the so-called Mennonite mutation, which occurs once in every 176 live births in the Old Order Mennonite population (versus 1 in 185,000 worldwide). This mutation is so common among Old Order Mennonites that it has been attributed to a founder effect, that is, a mutation that originated in one of the handful of founders of this isolated community. The E1 tetramer can be considered to be a dimer of $\alpha\beta$ heterodimers with a TPP cofactor at the interface between an α subunit and a β subunit and with each α subunit

contacting both the β and β' subunits (Fig. 26-22a). The amino acid change in the Mennonite mutation occurs at the α - β' interface: Tyr 393 α is hydrogen bonded to both His 385 α and Asp 328 β' (Fig. 26-22b). Its mutation to Asn disrupts these interactions and thereby impedes tetramerization.

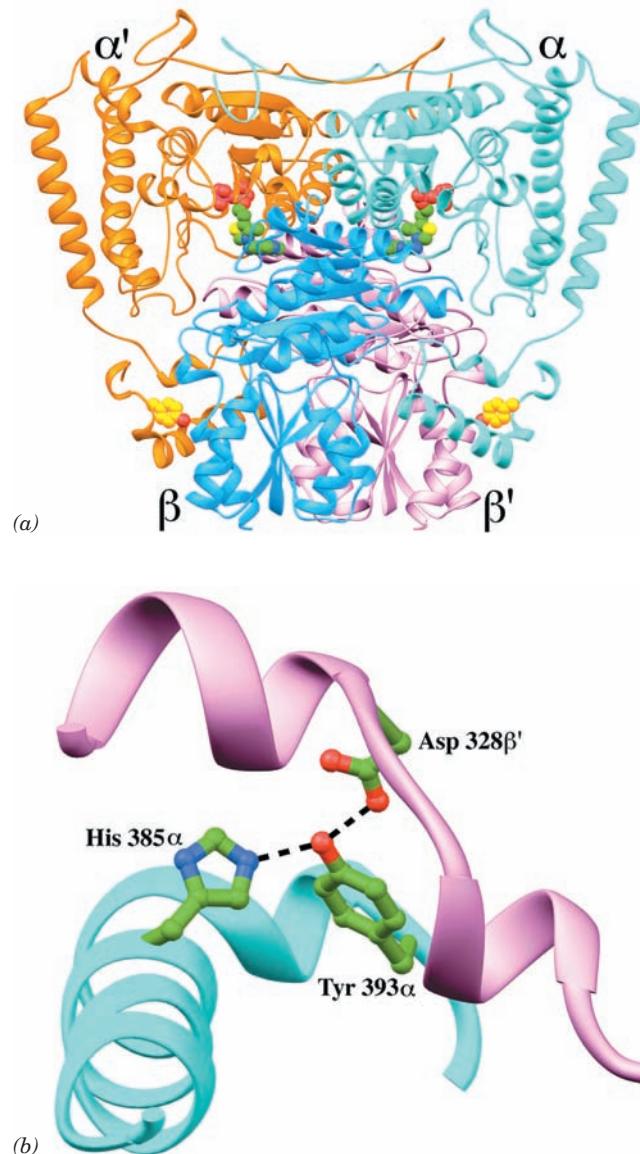


Figure 26-22 X-ray structure of the E1 component of the human branched-chain α -keto acid dehydrogenase multienzyme complex. (a) The $\alpha_2\beta_2$ heterotetramer. The α subunits are colored cyan and orange, and the β subunits are blue and pink. The thiamine pyrophosphate (TPP) cofactor and Tyr 393 α (which is mutated to Asn in the Mennonite mutation, causing maple syrup urine disease) are shown in space-filling form with TPP C green, Tyr 393 α C gold, N blue, O red, S yellow, and P magenta. Note the similarity of this structure to that of the E1 component of the pyruvate dehydrogenase multienzyme complex (Fig. 21-12a). (b) The α - β' interface colored as in Part a and showing the interactions of Tyr 393 α with His 385 α and Asp 328 β' . The side chains of these residues are drawn in ball-and-stick form with C green, N blue, and O red and with the hydrogen bonds between them represented by dashed lines. [Based on an X-ray structure by Wim Hol, University of Washington. PDBid 1DTW.]

F. Leucine and Lysine Are Degraded to Acetoacetate and/or Acetyl-CoA

Leucine is oxidized by a combination of reactions used in β oxidation and ketone body synthesis (Fig. 26-21, right). The first dehydrogenation and the hydration reactions are interspersed by (11) a carboxylation reaction catalyzed by the biotin-containing enzyme **β -methylcrotonyl-CoA carboxylase**. The hydration reaction (12) then produces **β -hydroxy- β -methylglutaryl-CoA (HMG-CoA)**, which is cleaved by HMG-CoA lyase to form acetyl-CoA and the

ketone body acetoacetate (13) (which, in turn, may be converted to 2 acetyl-CoA; Section 25-3).

Although there are several pathways for lysine degradation, the one that proceeds via formation of the α -ketoglutarate–lysine adduct **saccharopine** predominates in mammalian liver (Fig. 26-23). This pathway is of interest because we have encountered 7 of its 11 reactions in other pathways. Reaction 4 is a PLP-dependent transamination. Reaction 5 is the oxidative decarboxylation of an α -keto acid by a multienzyme complex similar to pyruvate dehydrogenase

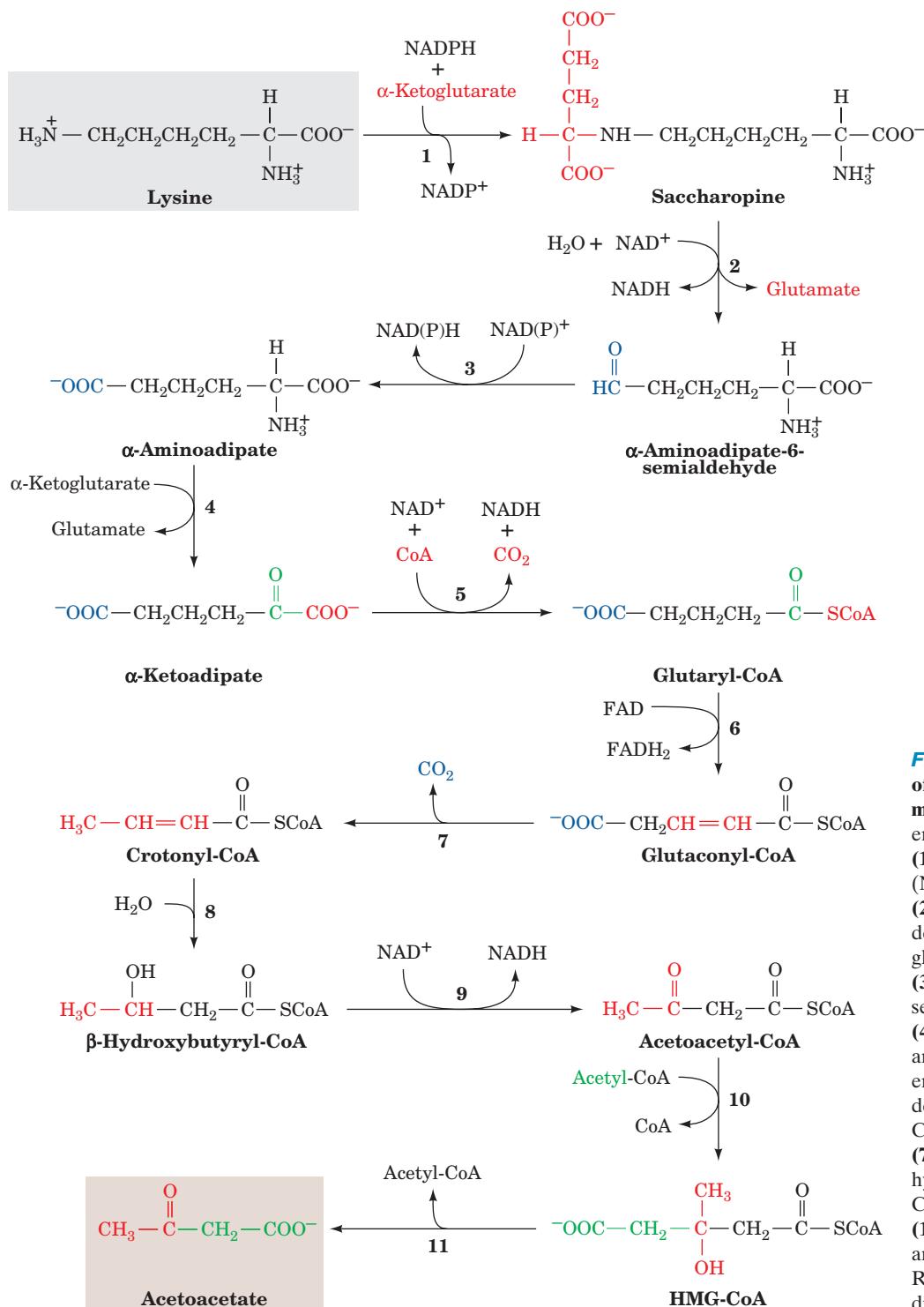


Figure 26-23 The pathway of lysine degradation in mammalian liver. The enzymes involved are (1) saccharopine dehydrogenase (NADP⁺, lysine forming), (2) saccharopine dehydrogenase (NAD⁺, glutamate forming), (3) amino adipate semialdehyde dehydrogenase, (4) amino adipate aminotransferase (a PLP enzyme), (5) α -keto acid dehydrogenase, (6) glutaryl-CoA dehydrogenase, (7) decarboxylase, (8) enoyl-CoA hydratase, (9) β -hydroxyacyl-CoA dehydrogenase, (10) HMG-CoA synthase, and (11) HMG-CoA lyase. Reactions 10 and 11 are discussed in Section 25-3.

and α -ketoglutarate dehydrogenase (Sections 21-2A and 21-3D). Reactions 6, 8, and 9 are standard reactions of fatty acyl-CoA oxidation: dehydrogenation by FAD, hydration, and dehydrogenation by NAD^+ . Reactions 10 and 11 are standard reactions in ketone body formation. Two molecules of CO_2 are produced at Reactions 5 and 7 of the pathway.

The saccharopine pathway is thought to predominate in mammals because a genetic defect in the enzyme that catalyzes Reaction 1 in the sequence results in **hyperlysine** and **hyperlysinuria** (elevated levels of lysine in the blood and urine, respectively) along with mental and physical retardation. This is yet another example of how the study of rare inherited disorders has helped to trace metabolic pathways.

Leucine's carbon skeleton, as we have seen, is converted to one molecule each of acetoacetate and acetyl-CoA, whereas that of lysine is converted to one molecule of acetoacetate and two of CO_2 . Since neither acetoacetate nor acetyl-CoA can be converted to glucose in animals, leucine and lysine are purely ketogenic amino acids.

G. Tryptophan Is Degraded to Alanine and Acetoacetate

The complexity of the major tryptophan degradation pathway (Fig. 26-24) precludes detailed discussion of all of its reactions. However, one reaction in the pathway is of particular interest. Reaction 4, cleavage of **3-hydroxykynurenine**

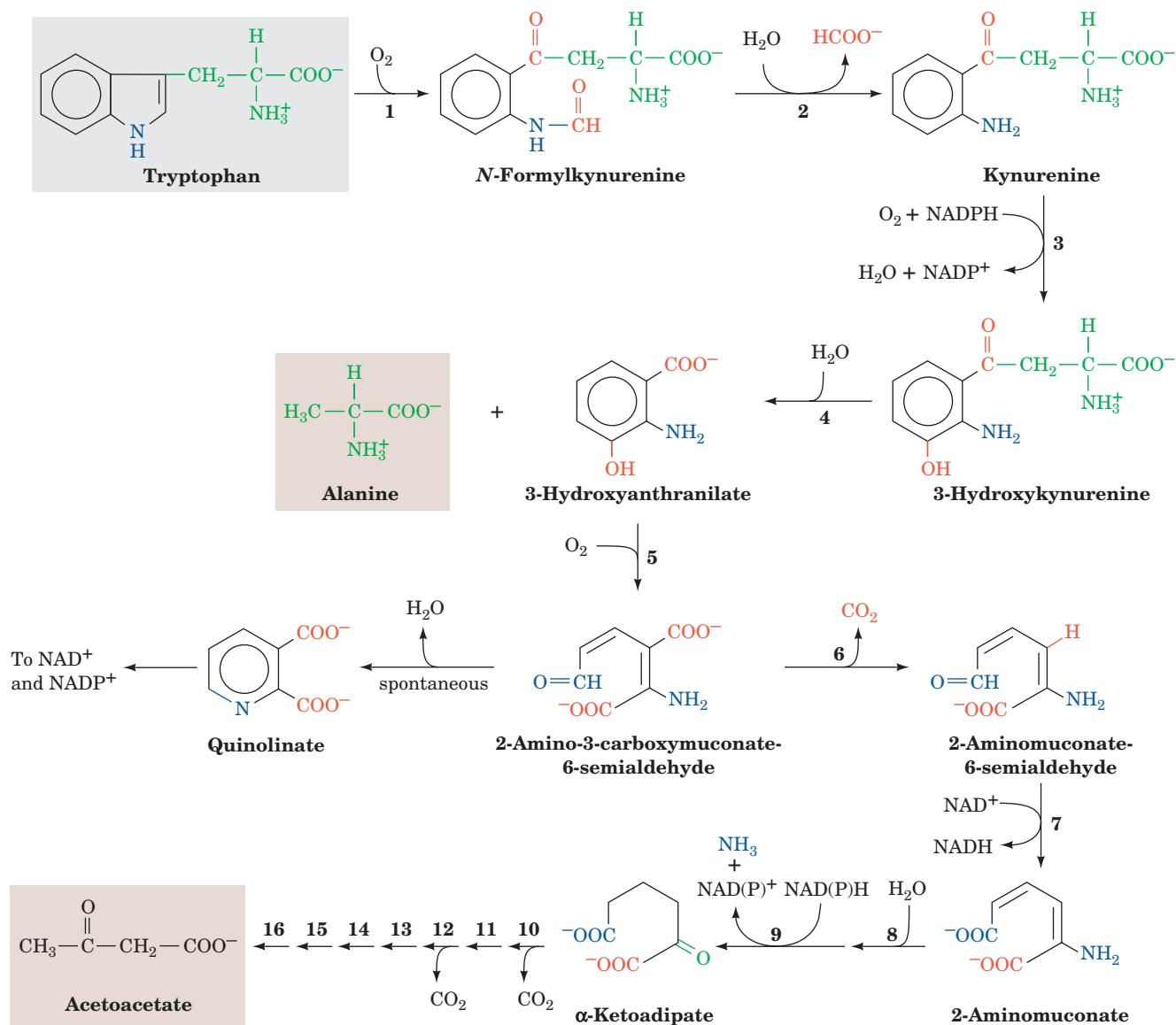


Figure 26-24 The pathway of tryptophan degradation. The enzymes involved are (1) tryptophan-2,3-dioxygenase, (2) formamidase, (3) kynurenine-3-monooxygenase, (4) kynureinase (PLP dependent), (5) 3-hydroxyanthranilate-3,4-dioxygenase, (6) amino carboxymuconate semialdehyde decarboxylase, (7) aminomuconate semialdehyde dehydrogenase,

(8) hydratase, (9) dehydrogenase, and (10-16) enzymes of Reactions 5 through 11 in lysine degradation (Fig. 26-23). 2-Amino-3-carboxymuconate-6-semialdehyde, in addition to undergoing Reaction 6, spontaneously forms **quinolinate**, an NAD^+ and NADP^+ precursor (Section 28-5A).

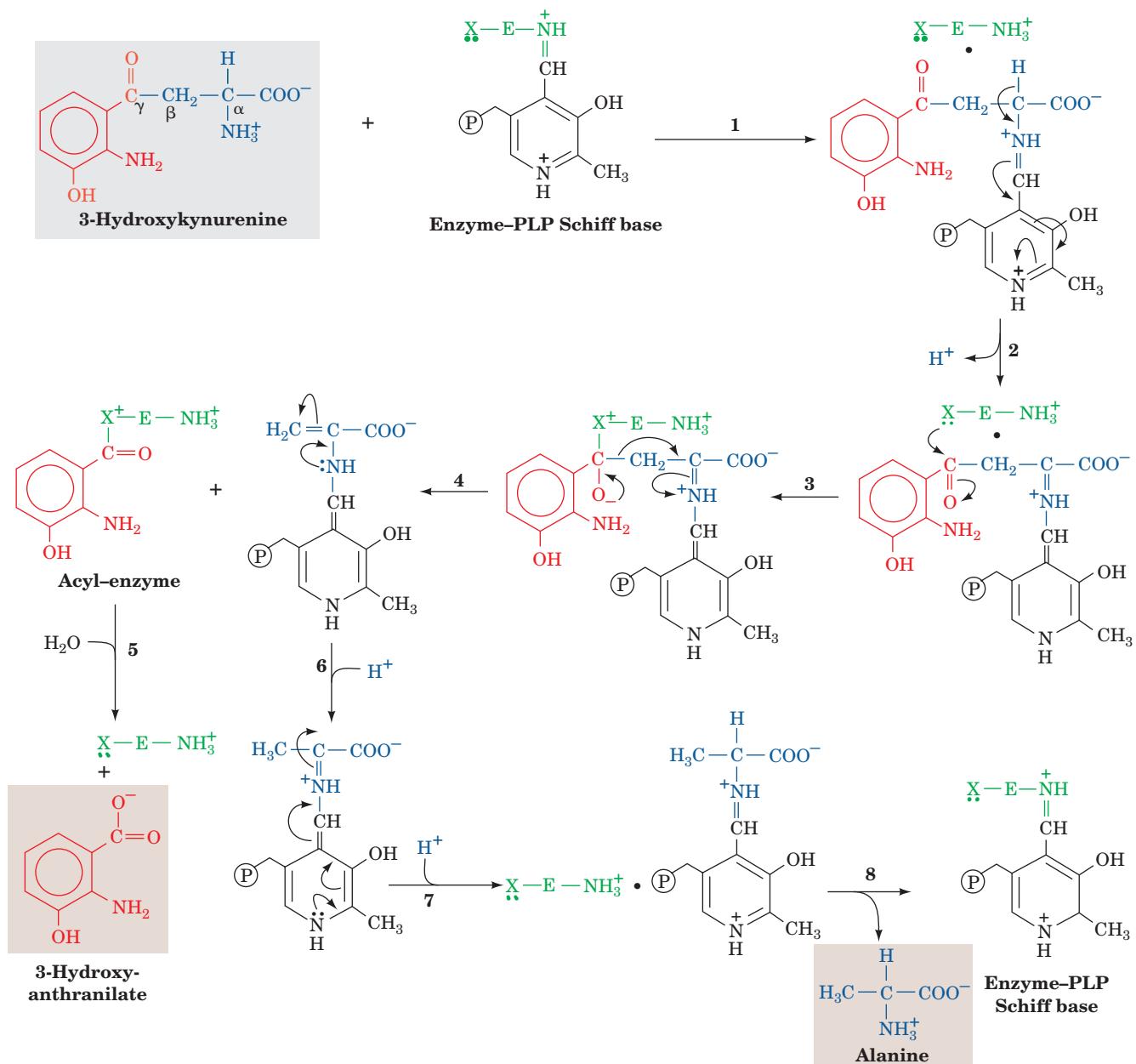


Figure 26-25 Proposed mechanism for the PLP-dependent kynureninase-catalyzed C_β—C_γ bond cleavage of 3-hydroxykynurenine. The reaction occurs in eight steps: (1) transimination; (2) tautomerization; (3) attack of an enzyme nucleophile on the carbonyl carbon (C_γ) of the tautomerized

nucleophile, X; (4) C_β—C_γ bond cleavage with formation of an acyl-enzyme intermediate; (5) acyl-enzyme hydrolysis; (6) and (7) tautomerization; and (8) transimination.

to alanine and **3-hydroxyanthranilate**, is catalyzed by **kynureninase**, a PLP-dependent enzyme. The reaction further demonstrates the enormous versatility of PLP. We have seen how PLP can labilize an α -amino acid's C_α—H and C_α—C_β bonds (Fig. 26-16). Here we see the facilitation of C_β—C_γ bond cleavage. The reaction follows the same steps as transamination reactions but does not hydrolyze the tautomerized Schiff base (Fig. 26-25). The proposed reaction mechanism involves an attack of an enzyme nucleophile on the carbonyl carbon (C_γ) of the tautomerized

3-hydroxykynurenine-PLP Schiff base (Fig. 26-25, Step 3). This is followed by C_β—C_γ bond cleavage to generate an acyl-enzyme intermediate together with a tautomerized alanine-PLP adduct (Fig. 26-25, Step 4). Hydrolysis of the acyl-enzyme then yields 3-hydroxyanthranilate, whose further degradation yields **α -ketoadipate** (Fig. 26-24, Reactions 5–9). α -Ketoadipate is also an intermediate in lysine breakdown (Fig. 26-23, Reaction 4) so that the last seven reactions in the degradation of both these amino acids are identical, forming acetoacetate and two molecules of CO₂.

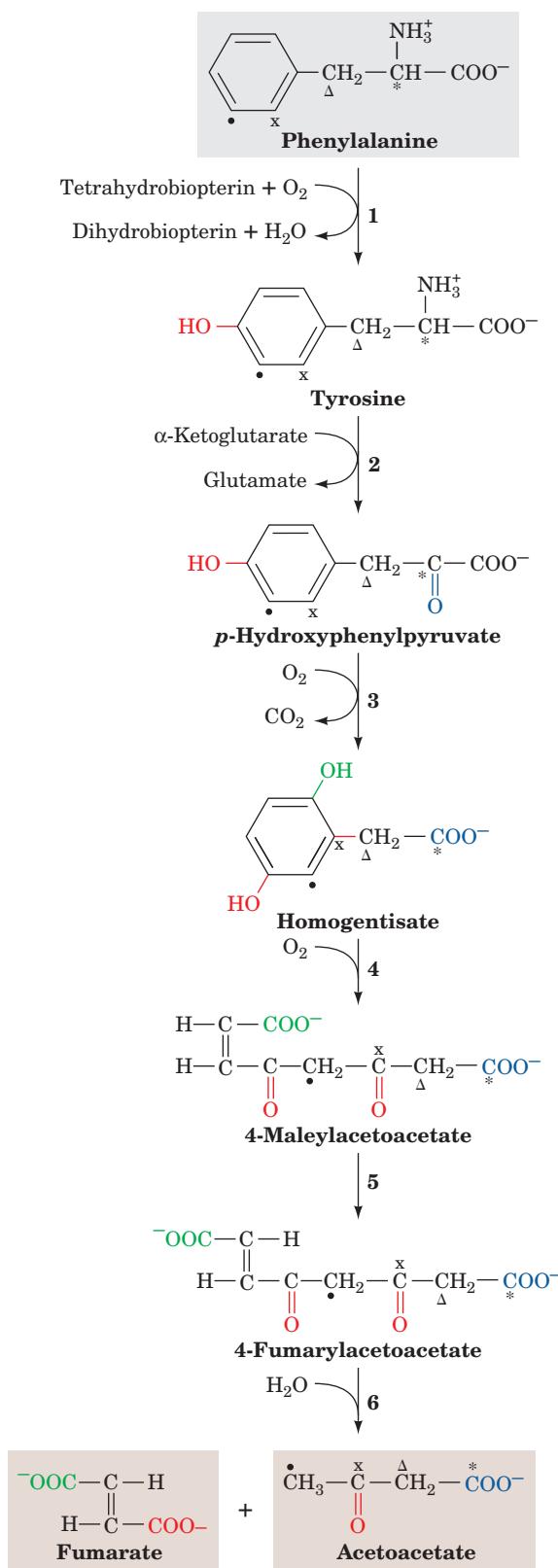


Figure 26-26 The pathway of phenylalanine degradation. The enzymes involved are (1) phenylalanine hydroxylase, (2) aminotransferase, (3) *p*-hydroxyphenylpyruvate dioxygenase, (4) homogentisate dioxygenase, (5) maleylacetoacetate isomerase, and (6) fumarylacetoacetate. The symbols labeling the various carbon atoms serve to indicate the group migration that occurs in Reaction 3 of the pathway (see Fig. 26-31).

H. Phenylalanine and Tyrosine Are Degraded to Fumarate and Acetoacetate

Since the first reaction in phenylalanine degradation is its hydroxylation to tyrosine, a single pathway (Fig. 26-26) is responsible for the breakdown of both of these amino acids. The final products of the six-reaction degradation are fumarate, a citric acid cycle intermediate, and acetoacetate, a ketone body.

a. Pterins Are Redox Cofactors

The hydroxylation of phenylalanine by the nonheme iron-containing homotetrameric enzyme **phenylalanine hydroxylase (PAH)** requires O_2 and that the iron be in the Fe(II) state. The enzyme also requires the participation of **biopterin**, a **pterin** derivative. Pterins are compounds that contain the **pteridine** ring (Fig. 26-27). Note the resemblance between the pteridine ring and the isoalloxazine ring of the flavin coenzymes; the positions of the nitrogen atoms in pteridine are identical with those of the B and C rings of isoalloxazine. Folate derivatives also contain the pterin ring (Section 26-4D). Pterins, like flavins, participate in biological oxidations. The active form of biopterin is the fully reduced form, **5,6,7,8-tetrahydrobiopterin (BH4)**. It is produced from **7,8-dihydrobiopterin** and NADPH, in what

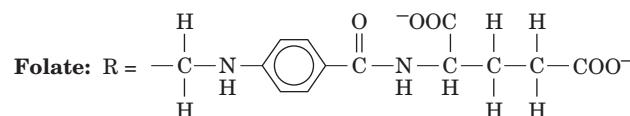
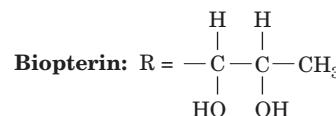
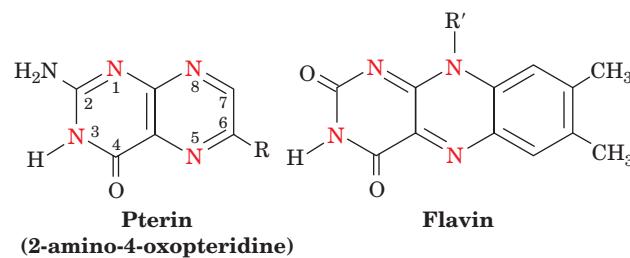
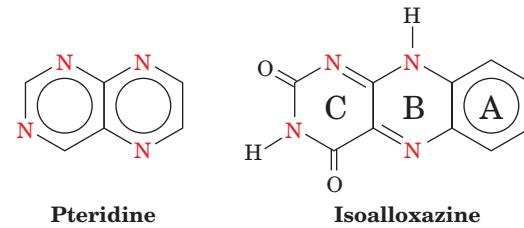


Figure 26-27 The pteridine ring, the nucleus of biopterin and folate. Note the similar structures of pteridine and the isoalloxazine ring of flavin coenzymes.

may be considered a priming reaction, by **dihydrofolate reductase** (Fig. 26-28).

Each 452-residue subunit of the PAH homotetramer contains three domains, an N-terminal regulatory domain, a catalytic domain, and a C-terminal tetramerization domain. However, the 325-residue catalytic domain alone forms catalytically competent dimers. The X-ray structure of the catalytic domain of PAH in its Fe(II) state in complex with BH_4 , determined by Edward Hough, reveals that the Fe(II) is octahedrally coordinated by His 285, His 290, Glu 330, and three water molecules, and that atom O4 of BH_4 is hydrogen bonded to two of these waters (Fig. 26-29).

In the phenylalanine hydroxylase reaction, 5,6,7,8-tetrahydrobiopterin is hydroxylated to **pterin-4a-carbinolamine** (Fig. 26-28), which is converted to **7,8-dihydrobiopterin (quinoid form)** by **pterin-4a-carbinolamine dehydratase**. The quinoid is subsequently reduced by the NAD(P)H-requiring enzyme **dihydropteridine reductase** to regenerate the active cofactor. Note that although dihydrofolate reductase and dihydropteridine reductase produce the same product, they utilize different tautomers of the substrate. Although this suggests that these enzymes may be evolutionarily related, the comparison of their X-ray structures indicates that this is not the case: Dihydropteridine reductase resembles nicotinamide

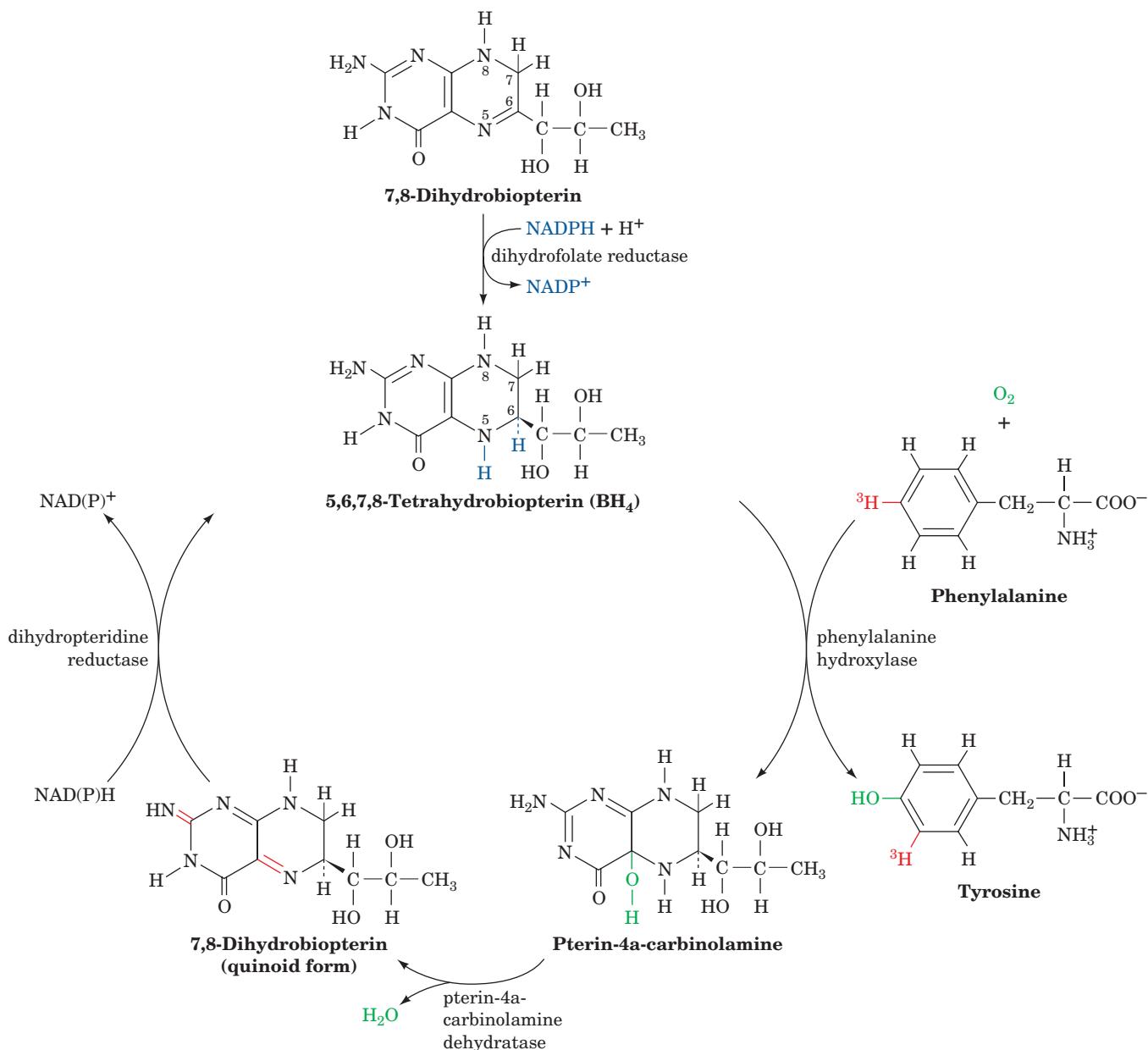


Figure 26-28 The formation, utilization, and regeneration of 5,6,7,8-tetrahydrobiopterin (BH_4) in the phenylalanine hydroxylase reaction.

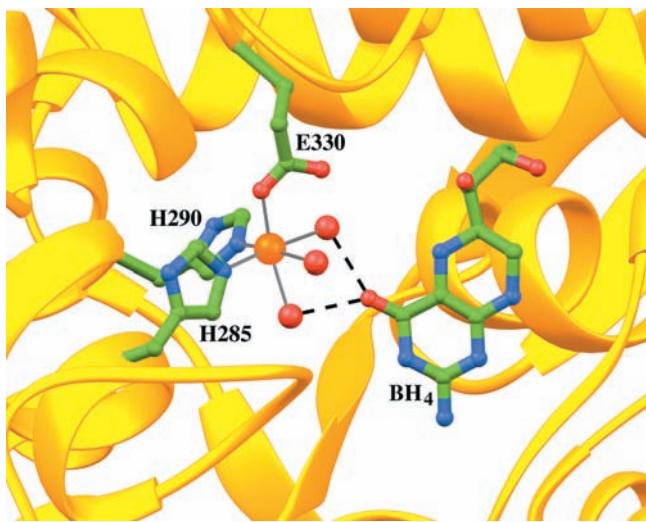


Figure 26-29 The active site of the Fe(II) form of phenylalanine hydroxylase (PAH) in complex with 5,6,7,8-tetrahydrobiopterin (BH₄). The Fe(II) (orange sphere) is octahedrally coordinated (gray lines) by His 285, His 290, and Glu 330 (C green, N blue, and O red) and three water molecules (red spheres). BH₄ atom O4 is hydrogen bonded (black dashed lines) to two of these water molecules. [Based on an X-ray structure by Edward Hough, University of Tromsø, Norway. PDBid 1J8U.]

coenzyme-requiring flavin-dependent enzymes such as glutathione reductase and dihydrolipoyl dehydrogenase (Section 21-2B).

b. Phenylalanine Hydroxylase Is Controlled by Phosphorylation and by Allosteric Interactions

PAH initiates the detoxification of high concentrations of phenylalanine as well as the synthesis of the catecholamine hormones and neurotransmitters (Section 26-4B). It is allosterically activated by its substrate, phenylalanine, and by phosphorylation at its Ser 16 by the cAMP-dependent protein kinase A (PKA; Section 18-3C). Its second substrate, BH₄, allosterically inhibits the enzyme.

c. The NIH Shift

An unexpected aspect of the PAH reaction is that a ³H atom, which begins on C4 of phenylalanine's phenyl ring, ends up on C3 of this ring in tyrosine (Fig. 26-28, right) rather than being lost to the solvent by replacement with the OH group. The mechanism postulated to account for this **NIH shift** (so called because it was first characterized by chemists at the National Institutes of Health) involves the activation of oxygen by the pterin and Fe cofactors to form the pterin-4a-carbinolamine and a reactive oxyferryl group [Fe(IV)=O²⁻; Fig. 26-30, Steps 1 and 2] that reacts with the substrate to form an epoxide across the phenyl ring's 3,4 bond (Fig. 26-30, Step 3). This is followed by epoxide opening to form a carbocation at C3 (Fig. 26-30, Step 4). Migration of a hydride from C4 to C3 forms a more stable carbocation (an oxonium ion; Fig. 26-30, Step 5). This migration is followed by ring aromatization to form tyrosine

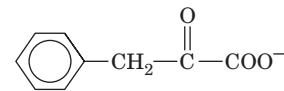
(Fig. 26-30, Step 6). **Tyrosine hydroxylase** and **tryptophan hydroxylase** (Section 26-4B) are both homologous to phenylalanine hydroxylase and utilize this same NIH shift reaction mechanism, although there may not be an epoxide intermediate in these cases.

Reaction 3 in the phenylalanine degradation pathway (Fig. 26-26) provides another example of an NIH shift. This reaction, which is catalyzed by the Fe(II)-containing **p-hydroxyphenylpyruvate dioxygenase**, involves the oxidative decarboxylation of an α -keto acid as well as ring hydroxylation. In this case, the NIH shift involves migration of an alkyl group rather than of a hydride ion to form a more stable carbocation (Fig. 26-31). This shift, which has been demonstrated through isotope-labeling studies (represented by the different symbols in Figs. 26-26 and 26-31), accounts for the observation that C3 is bonded to C4 in **p-hydroxyphenylpyruvate** but to C5 in **homogentisate**.

d. Alkaptonuria and Phenylketonuria Result from Defects in Phenylalanine Degradation

Archibald Garrod realized in the early 1900s that human genetic diseases result from specific enzyme deficiencies. We have repeatedly seen how this realization has contributed to the elucidation of metabolic pathways. The first such disease to be recognized was **alkaptonuria**, which, Garrod observed, resulted in the excretion of large quantities of homogentisic acid. This condition results from deficiency of **homogentisate dioxygenase** (Fig. 26-26, Reaction 4). Alkaptonurics suffer no ill effects other than arthritis later in life (although their urine darkens alarmingly because of the rapid air oxidation of the homogentisate they excrete).

Individuals suffering from **phenylketonuria (PKU)** are not so fortunate. Severe mental retardation occurs within a few months of birth if the disease is not detected and treated immediately (see below). Indeed, ~1% of the patients in mental institutions were, at one time (before routine screening), phenylketonurics. PKU is caused by the inability to hydroxylate phenylalanine (Fig. 26-26, Reaction 1) and therefore results in increased blood levels of phenylalanine (**hyperphenylalaninemia**). The excess phenylalanine is transaminated to **phenylpyruvate**



Phenylpyruvate

by an otherwise minor pathway. The “spillover” of phenylpyruvate (a phenylketone) into the urine was the first observation connected with the disease and gave the disease its name, although it has since been demonstrated that it is the high concentration of phenylalanine itself that gives rise to brain dysfunction. All babies born in the United States are now screened for PKU immediately after birth by testing for elevated levels of phenylalanine in the blood.

Classic PKU results from a deficiency in phenylalanine hydroxylase (PAH). When this was established in 1947, it was the first human inborn error of metabolism whose basic biochemical defect had been identified. Since then, over

Pterin-4a-carbinolamine

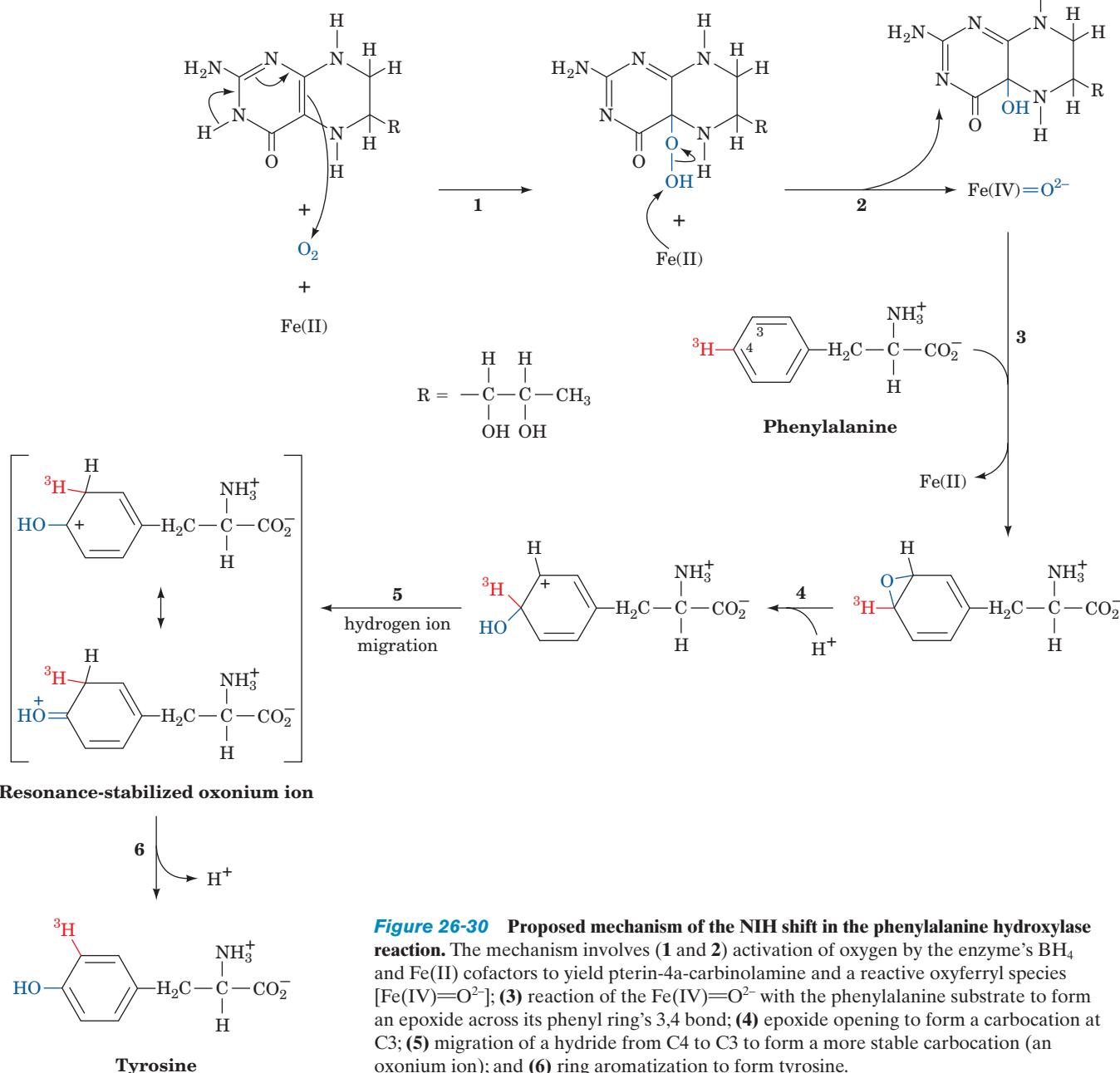
5,6,7,8-Tetrahydrobiopterin (BH_4)

Figure 26-30 Proposed mechanism of the NIH shift in the phenylalanine hydroxylase reaction. The mechanism involves (1 and 2) activation of oxygen by the enzyme's BH_4 and $\text{Fe}(\text{II})$ cofactors to yield pterin-4a-carbinolamine and a reactive oxyferryl species $[\text{Fe}(\text{IV})=\text{O}^{2-}]$; (3) reaction of the $\text{Fe}(\text{IV})=\text{O}^{2-}$ with the phenylalanine substrate to form an epoxide across its phenyl ring's 3,4 bond; (4) epoxide opening to form a carbocation at C3; (5) migration of a hydride from C4 to C3 to form a more stable carbocation (an oxonium ion); and (6) ring aromatization to form tyrosine.

400 mutations have been identified in PAH. Because all of the tyrosine breakdown enzymes are normal, treatment consists in providing the patient with a low-phenylalanine diet and monitoring the blood level of phenylalanine to ensure that it remains within normal limits for the first 5 to 10 years of life (the adverse effects of hyperphenylalaninemia seem to disappear after that age). PAH deficiency also accounts for another common symptom of PKU: Its victims have lighter hair and skin color than their siblings. This is because tyrosine hydroxylation, the first reaction in the formation of the black skin pigment **melanin** (Section 26-4B), is inhibited by elevated phenylalanine levels.

Other causes of hyperphenylalaninemia have been discovered since the introduction of infant screening techniques. These result from deficiencies in the enzymes catalyzing the formation or regeneration of 5,6,7,8-tetrahydrobiopterin (BH_4), the PAH cofactor (Fig. 26-28). In such cases, patients must also be supplied with **L-3,4-dihydroxyphenylalanine (L-DOPA)** and **5-hydroxytryptophan**, metabolic precursors of the neurotransmitters **norepinephrine** and **serotonin**, respectively, since tyrosine hydroxylase and tryptophan hydroxylase, the PAH homologs that produce these physiologically active amines, also require 5,6,7,8-tetrahydrobiopterin (Section 26-4B). Unfortunately, simply adding BH_4 to the diet of

an affected individual is not an effective treatment because BH_4 is unstable and cannot cross the blood–brain barrier.

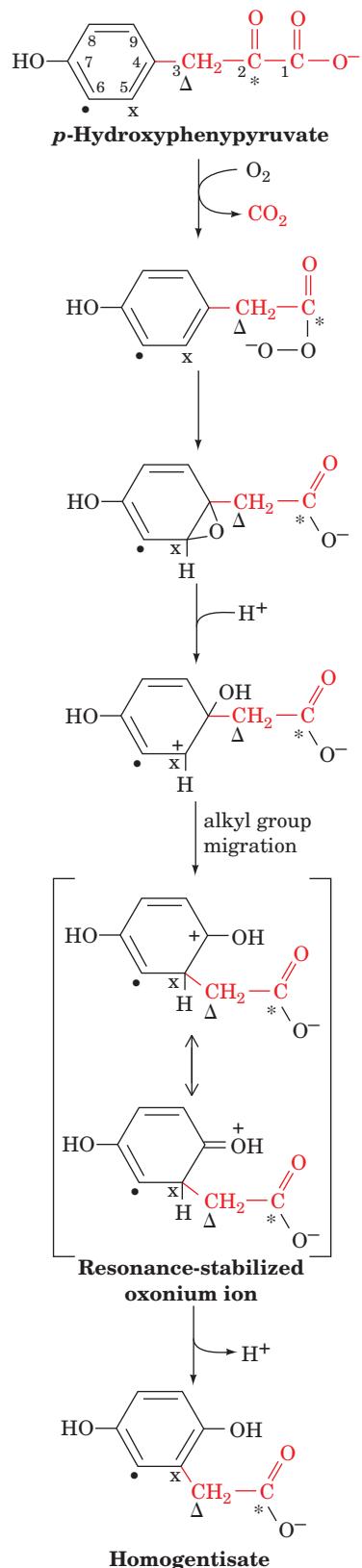


Figure 26-31 The NIH shift in the *p*-hydroxyphenylpyruvate dioxygenase reaction. Carbon atoms are labeled as an aid to following the group migration constituting the shift.

4 AMINO ACIDS AS BIOSYNTHETIC PRECURSORS

Certain amino acids, in addition to their major function as protein building blocks, are essential precursors of a variety of important biomolecules, including nucleotides and nucleotide coenzymes, heme, various hormones and neurotransmitters, and glutathione. In this section, we consider the pathways producing some of these substances. We begin by discussing the biosynthesis of heme from glycine and succinyl-CoA. We then examine the pathways by which tyrosine, tryptophan, glutamate, and histidine are converted to various neurotransmitters and study certain aspects of glutathione biosynthesis and the involvement of this tripeptide in amino acid transport and other processes. Finally, we consider the role of folate derivatives in the biosynthetic transfer of C_1 units. The biosynthesis of nucleotides and nucleotide coenzymes is the subject of Chapter 28.

A. Heme Biosynthesis and Degradation

Heme (Fig. 26-32), as we have seen, is an Fe-containing prosthetic group that is an essential component of many proteins, notably hemoglobin, myoglobin, and the cytochromes. The initial reactions of heme biosynthesis are common to the formation of other tetrapyrroles including

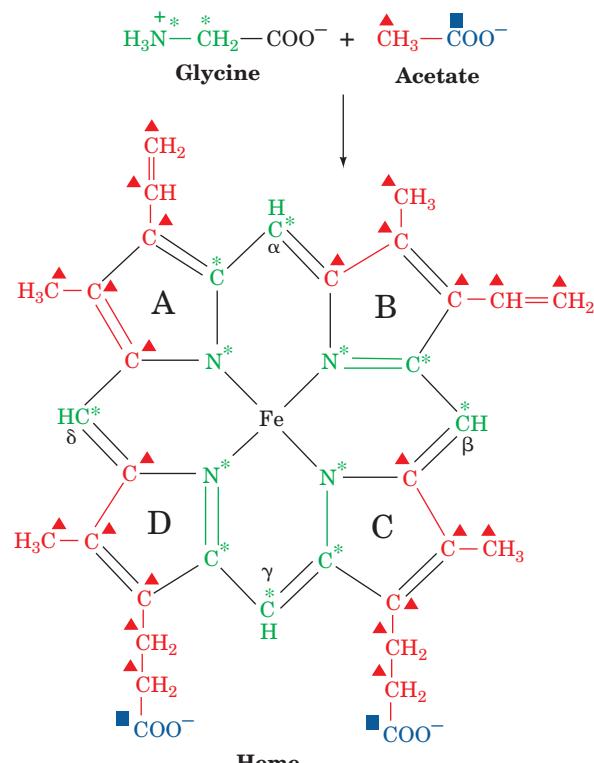


Figure 26-32 Structure of heme. Heme's C and N atoms are derived from those of glycine and acetate.

chlorophyll in plants and bacteria (Section 24-2A) and coenzyme B₁₂ in bacteria (Section 25-2Eb).

a. Porphyrins Are Derived from Succinyl-CoA and Glycine

Elucidation of the heme biosynthesis pathway involved some interesting detective work. David Shemin and David Rittenberg, who were among the first to use isotopic tracers in the elucidation of metabolic pathways, demonstrated, in 1945, that *all of heme's C and N atoms can be derived from acetate and glycine*. Only glycine, out of a variety of ¹⁵N-labeled metabolites they tested (including ammonia, glutamate, leucine, and proline), yielded ¹⁵N-labeled heme in the hemoglobin of experimental subjects to whom these metabolites were administered. Similar experiments, using acetate labeled with ¹⁴C in its methyl or carboxyl groups, or [¹⁴C_α]glycine, demonstrated that 24 of heme's 34 carbon atoms are derived from acetate's methyl carbon, 2 from acetate's carboxyl carbon, and 8 from glycine's C_α atom (Fig. 26-32). None of the heme atoms is derived from glycine's carboxyl carbon atom.

Figure 26-32 indicates that heme C atoms derived from acetate methyl groups occur in groups of three linked atoms. Evidently, acetate is first converted to some other metabolite that has this labeling pattern. Shemin and Rittenberg postulated that this metabolite is succinyl-CoA based on the following reasoning (Fig. 26-33):

1. Acetate is metabolized via the citric acid cycle (Section 21-1B).
2. Labeling studies indicate that atom C3 of the citric acid cycle intermediate succinyl-CoA is derived from acetate's methyl C atom, whereas atom C4 comes from acetate's carboxyl C atom.
3. After many turns of the citric acid cycle, C1 and C2 of succinyl-CoA likewise become fully derived from acetate's methyl C atom.

We shall see that this labeling pattern indeed leads to that of heme.

In the mitochondria of yeast and animals as well as in some bacteria, the first phase of heme biosynthesis is a condensation of succinyl-CoA with glycine followed by decarboxylation to form **δ-aminolevulinic acid (ALA)** as catalyzed by the PLP-dependent enzyme **δ-aminolevulinate synthase (ALA synthase or ALAS)** (Fig. 26-34). The carboxyl group lost in the decarboxylation (Fig. 26-34, Reaction 5) originates in glycine, which is why heme contains no label from this group.

b. The Pyrrole Ring Is the Product of Two ALA Molecules

The pyrrole ring is formed in the next phase of the pathway through linkage of two molecules of ALA to yield **porphobilinogen (PBG)**. The reaction is catalyzed by **porphobilinogen synthase [PBGS; alternatively, δ-aminolevulinic acid dehydratase (ALAD)]**, which in yeast and mammals, is Zn²⁺-dependent and involves Schiff base formation of one of the substrate molecules with an enzyme amine

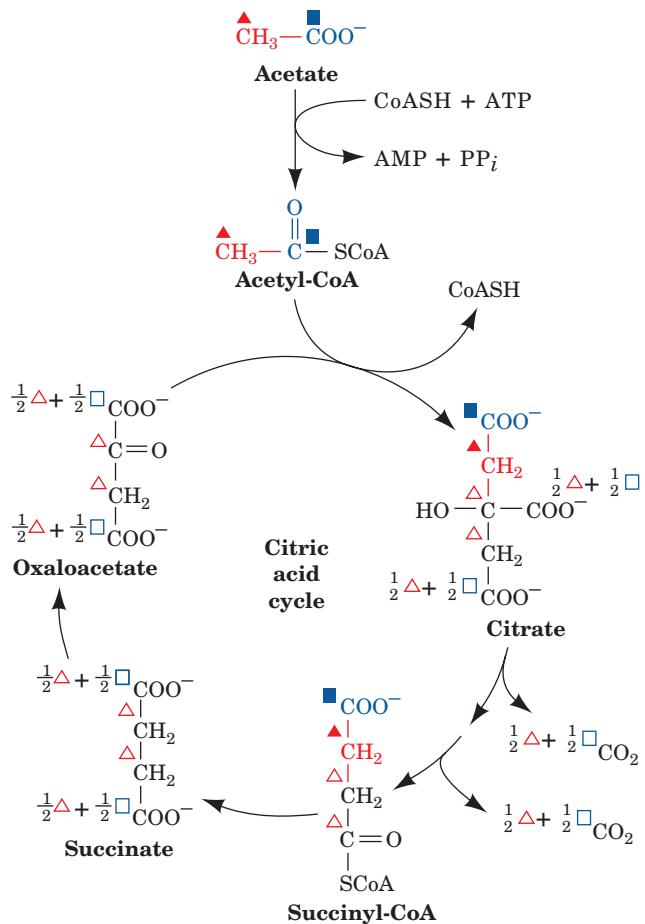


Figure 26-33 The origin of the C atoms of succinyl-CoA as derived from acetate via the citric acid cycle. C atoms labeled with triangles and squares are derived, respectively, from acetate's methyl and carboxyl C atoms. Filled symbols label atoms derived from acetate in the present round of the citric acid cycle, whereas open symbols label atoms derived from acetate in previous rounds of the citric acid cycle. Note that the C1 and C4 atoms of succinyl-CoA are scrambled on forming the 2-fold symmetric succinate.

group (in some bacteria and all plants, Mg²⁺ substitutes for Zn²⁺). One possible mechanism of this condensation-elimination reaction involves formation of a second Schiff base between the ALA-enzyme Schiff base and the second ALA molecule (Fig. 26-35). At this point, if we continue tracing the acetate and glycine labels through the PBG synthase reaction (Fig. 26-35), we can begin to see how heme's labeling pattern arises.

The X-ray structure of human PBGS in covalent complex with its product PBG, determined by Jonathan Cooper, indicates that this enzyme is a homooctamer with D₄ symmetry. Each of its 330-residue subunits consists of an α/β barrel and a 39-residue N-terminal tail that wraps around a neighboring monomer (related to it by 2-fold symmetry) so that the protein is better described as a relatively loosely organized tetramer of compact dimers. As is the case with nearly all α/β barrel enzymes, PBGS's active site (Fig. 26-36a) lies at the mouth of the barrel at the

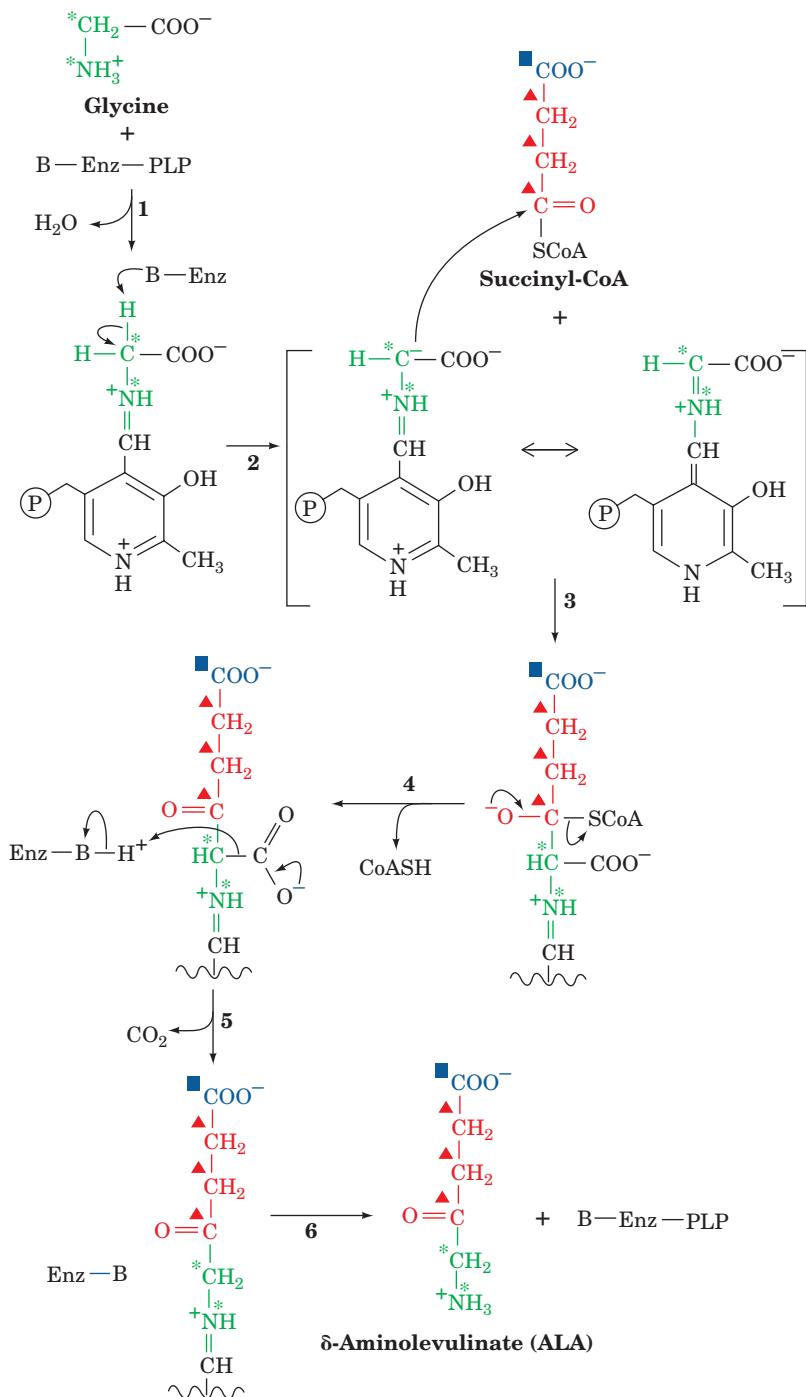
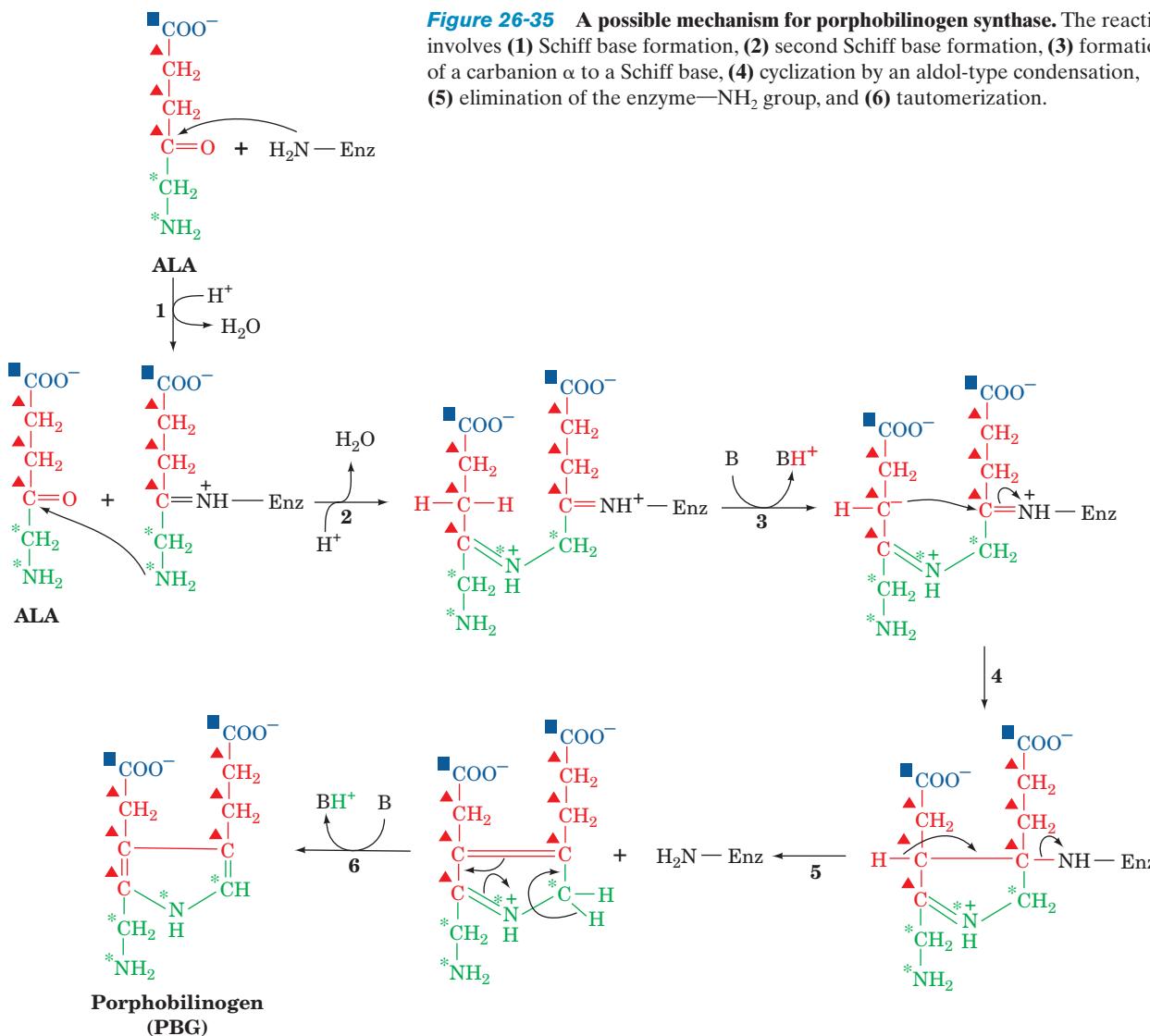


Figure 26-34 The mechanism of action of the PLP-dependent enzyme δ -aminolevulinate synthase (ALAS). The reaction steps are (1) transamination, (2) PLP-stabilized carbanion formation,

(3) C—C bond formation, (4) CoA elimination, (5) decarboxylation facilitated by the PLP-Schiff base, and (6) transimination yielding ALA and regenerating the PLP-enzyme.

C-terminal ends of its β strands. The active site is covered by a loop that comparison with other PBGS structures indicates forms a flexible lid over the substrate, an arrangement that is reminiscent of the glycolytic enzyme triose phosphate isomerase (TIM; Fig. 17-11). PBG is covalently bound to Lys 252 and its free amino group is coordinated to the active site Zn²⁺ ion. Lys 199 appears to be properly positioned to act as a general acid–base catalyst.

Inhibition of PBG synthase by Pb²⁺ (a competitor of its active site Zn²⁺ ion) is one of the major manifestations of lead poisoning, which is among the most common acquired environmental diseases. Indeed, it has been suggested that the accumulation, in the blood, of ALA, which resembles the neurotransmitter γ -aminobutyric acid (Section 26-4B), is responsible for the psychosis that often accompanies lead poisoning.



c. PBGS Has Two Quaternary States with Different Enzymatic Activities

Although the X-ray structure of wild-type human PBGS indicates that it is a D_4 -symmetric homooctamer, the X-ray structure of its rare F12L mutant form, determined by Eileen Jaffe, is a D_3 -symmetric homohexamer (Fig. 26-36b). Moreover, the F12L mutant has far less activity than does wild-type PBGS, even though residue 12 is far from the enzyme's active site in both proteins. This is apparently due to a conformational change in PBGS's N-terminal arm: In the octamer, the N-terminal arms of two adjacent monomers wrap around each other's barrel to form a so-called hugging dimer, whereas in the hexamer, the N-terminal arms extend away from the core of the protein to form a so-called detached dimer (Fig. 26-36b). Nevertheless, the α/β barrels of these two oligomeric forms are closely superimposable. The difference in the activities of these two quaternary forms is caused by the binding of an allosteric Mg²⁺ ion in the octamer's "hugging" interface that is absent in the detached dimer.

The alternate oligomeric forms have been observed in solutions of wild-type PBGS and the equilibrium between these forms can be varied by changing the pH, the enzyme concentration, and the substrate concentration, as well as through mutation. Jaffe has therefore proposed that this quaternary structural change is an allosteric mechanism for controlling the activity of the enzyme. In contrast, in the symmetry and sequential models of allosterism (Sections 10-4B and 10-4C), the quaternary state of the enzyme does not change during an allosteric transition.

Jaffe has termed enzymes that control their activities by changing their oligomeric states **morpheeins** [pronounced morph-ee-in; derived from the verb "to morph" and the classic pronunciation of the word protein (pro-tee-in)]. Although PBGS is the first known morpheein, several others have since been identified (e.g., ribonucleotide reductase; Section 28-3Ad). In fact, it may be that morpheeins are relatively common because when an enzyme is purified, inactive fractions are usually assumed to be denatured and are therefore discarded rather than being characterized.

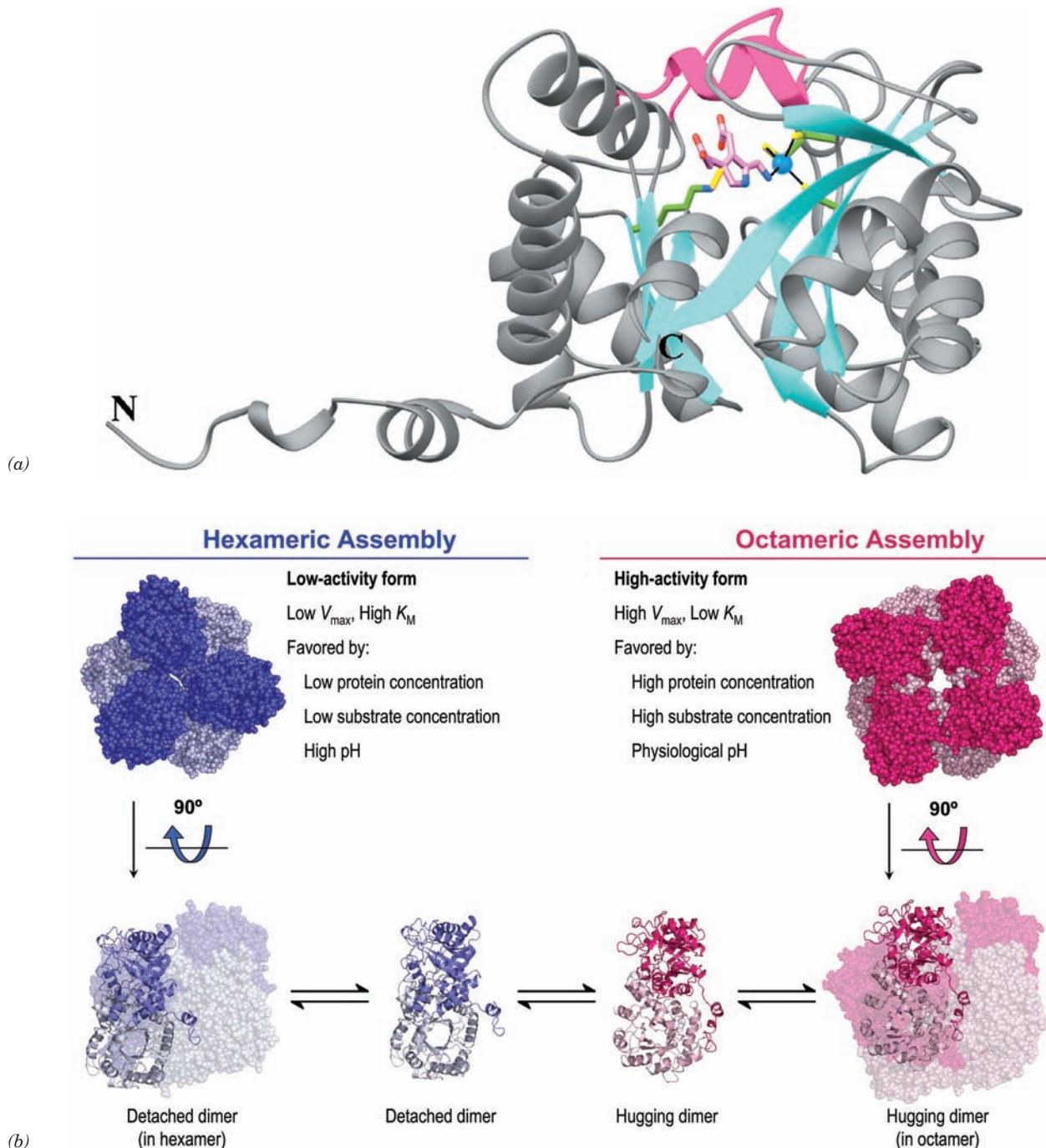
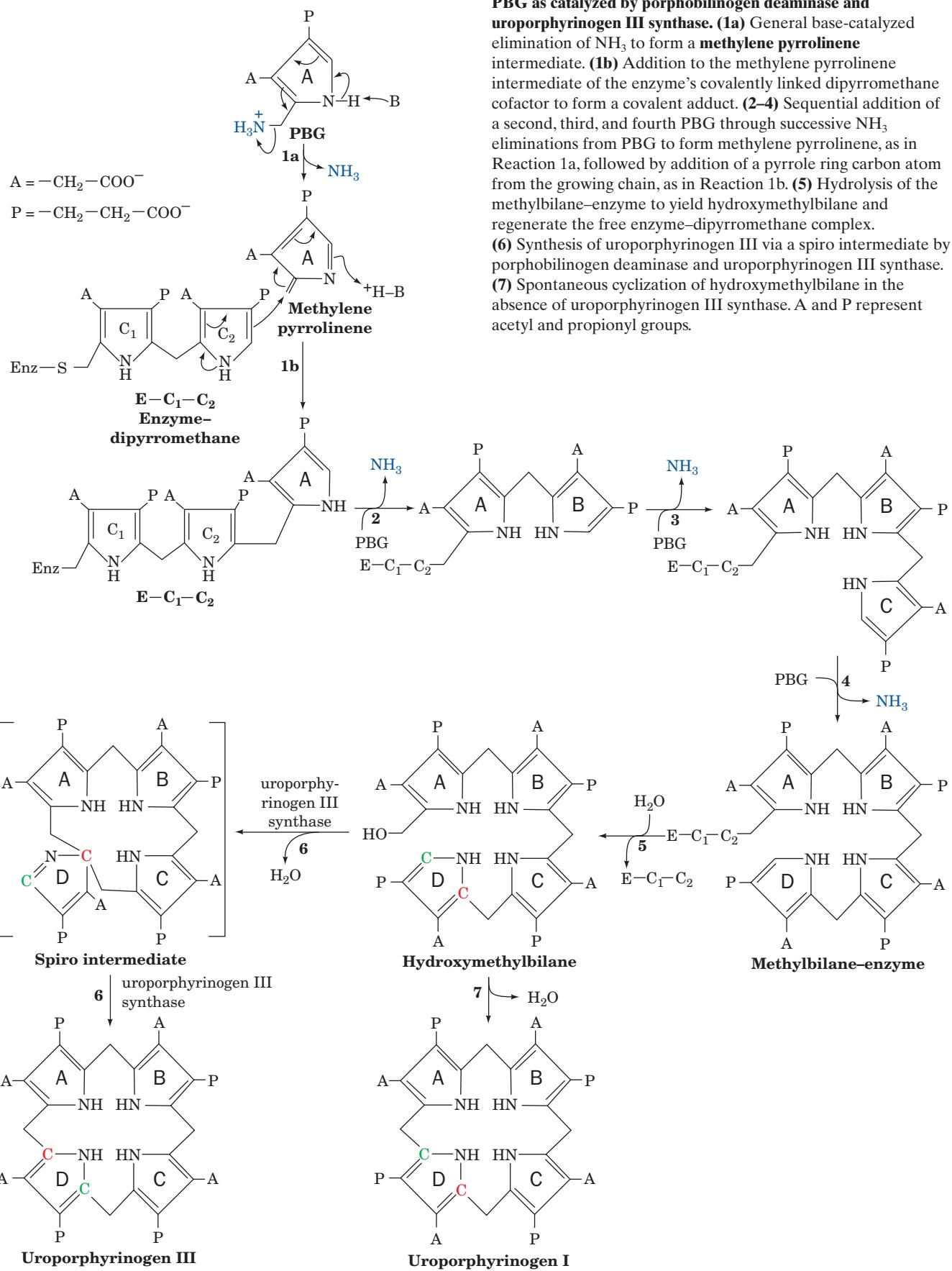


Figure 26-36 X-ray structure of human porphobilinogen synthase (PBGS). (a) PBGS in covalent complex with its porphobilinogen (PBG) product. A monomer of this homo-octameric protein is viewed perpendicular to the axis of its α/β barrel and is drawn in gray with its β strands cyan and the loop forming its flexible lid (residues 201–222) magenta. The PBG product, Lys 252 to which it is covalently linked, and the three Cys side chains that ligand the active site Zn^{2+} ion (blue sphere) are shown in stick form with PBG C pink, side chain C green, N blue, O red, S yellow, and the N—C bond linking Lys 252 to PBG gold. The active site Zn^{2+} ion is liganded (black lines) by the S atoms of Cys 122, Cys 124, Cys 132, and the PBG amino group. Lys 199, which lies directly behind Lys 252 in this view, appears to be properly positioned to act as an acid-base catalyst. [Based on an X-ray structure by Jonathan Cooper,

University of Southampton, U.K. PDBid 1E51.] (b) Quaternary structural changes between the low activity hexameric state of the F12L mutant of human PBGS (blue) and the high activity octameric state of the wild-type enzyme (red). In the upper panel, the proteins are viewed along their 3-fold and 4-fold axes with the subunits closest to the viewer more darkly colored. In the lower panel, the outer drawings are viewed along their 2-fold axes with one dimer drawn in ribbon form and the others in space-filling form. The inner drawings show only the dimers into which the oligomers are assumed to dissociate before reassembling to an alternate quaternary state. [Courtesy of Sarah Lawrence and Eileen Jaffe, The Fox Chase Cancer Center, Philadelphia, Pennsylvania. The X-ray structure of the F12L mutant was determined by Eileen Jaffe. PDBid 1PV8.]



d. The Porphyrin Ring Is Formed from Four PBG Molecules

The next phase of heme biosynthesis is the condensation of four PBG molecules to form **uroporphyrinogen III**, the porphyrin nucleus, in a series of reactions catalyzed by **porphobilinogen deaminase** (alternatively, **hydroxymethylbilane synthase** or **uroporphyrinogen synthase**) and **uroporphyrinogen III synthase**. The reaction (Fig. 26-37) begins with the enzyme's displacement of the amino group in PBG to form a covalent adduct. A second, third, and fourth PBG then sequentially add through the displacement of the primary amino group on one PBG by a carbon atom on the pyrrole ring of the succeeding PBG to yield a linear tetrapyrrole that is hydrolyzed and released from the enzyme as **hydroxymethylbilane** (also called **preuroporphyrinogen**).

e. Porphobilinogen Deaminase Has a Dipyromethane Cofactor

Peter Shoolingin-Jordan and Alan Battersby independently showed that porphobilinogen deaminase contains a unique **dipyromethane** cofactor (two pyrroles linked by a methylene bridge; rings C₁ and C₂ in Fig. 26-37), which is covalently linked to the enzyme via a C—S bond to an enzyme Cys residue. Thus, the methylbilane–enzyme complex really contains a linear hexapyrrole. The subsequent reaction step, also catalyzed by porphobilinogen deaminase (Step 5 in Fig. 26-37), is the hydrolysis of the bond linking the second and third pyrrole units of the hexapyrrole to yield hydroxymethylbilane and the dipyromethane cofactor. This cofactor is still linked to the enzyme, which is therefore ready to catalyze a new round of hydroxymethylbilane synthesis.

How is the dipyromethane cofactor assembled? Shoolingin-Jordan has shown that porphobilinogen deaminase synthesizes its own cofactor from two PBG units using, it appears, the same catalytic machinery with which it synthesizes methylbilane. However, the enzyme Cys reacts much more rapidly with presynthesized hydroxymethylbilane to form a reaction intermediate (the product of Step 2 in Fig. 26-37) that continues to add two more PBG units. When hydroxymethylbilane is released, the enzyme retains its dipyromethane cofactor.

The X-ray structure of human porphobilinogen deaminase (whose sequence is >45% identical to those of the *E. coli* enzyme), in covalent complex with its dipyromethane cofactor, indicates that this monomeric, 364-residue protein folds into three nearly equal sized domains (Fig. 26-38). The dipyromethane cofactor lies deep in a cleft between domains 1 and 2 such that there is still considerable unoccupied space in the cleft. Although the enzyme sequentially appends four PBG residues to the cofactor, it has only one catalytic site.

If the enzyme has only one catalytic site, how does it reposition the polypyrrrole chain after each catalytic cycle so that it can further extend this chain? One possibility is that the polypyrrrole chain fills the cavity next to the cofactor. This model provides a simple steric rationale for why the length of the polypyrrrole chain is limited to six residues (the final four of which are hydrolytically cleaved away by the enzyme to yield the hydroxymethylbilane product and regenerate the dipyromethane cofactor).

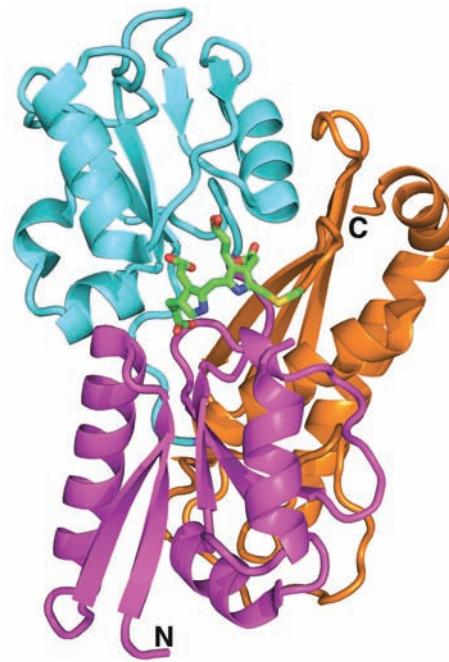


Figure 26-38 X-ray structure of human porphobilinogen deaminase in covalent complex with its dipyromethane cofactor. The protein is shown in ribbon form with domain 1 (residues 1–116 and 216–239) magenta, domain 2 (residues 117–215) cyan, and domain 3 (residues 240–364) orange. The dipyromethane cofactor and the Cys side chain to which it is covalently linked are drawn in stick form with C green, N blue, O red, and C yellow. [Based on an X-ray structure by Zhi-Jie Liu, Institute of Biophysics, Beijing, China. PDBid 3ECR.]

f. Protoporphyrin IX Biosynthesis Requires Four More Reactions

Cyclization of the hydroxymethylbilane product requires **uroporphyrinogen III synthase** (Fig. 26-37). In the absence of this enzyme, hydroxymethylbilane is released from the synthase and rapidly cyclizes nonenzymatically to the symmetric **uroporphyrinogen I**. Heme, however, is an asymmetric molecule; the methyl substituent of pyrrole ring D has an inverted placement compared to those of rings A, B, and C (Fig. 26-32). This ring reversal to yield uroporphyrinogen III has been shown by Battersby to proceed through attachment of the methylenes from rings A and C to the same carbon of ring D so as to form a spiro compound (a bicyclic compound with a carbon atom common to both rings; Fig. 26-37).

Heme biosynthesis takes place partly in the mitochondrion and partly in the cytosol (Fig. 26-39). ALA is mitochondrially synthesized and is transported to the cytosol for conversion to PBG and then to uroporphyrinogen III. **Protoporphyrin IX**, to which Fe is added to form heme, is produced from uroporphyrinogen III in a series of reactions catalyzed by (1) **uroporphyrinogen decarboxylase**, which decarboxylates all four acetate side chains (A) to form methyl groups (M); (2) **coproporphyrinogen oxidase**, which oxidatively decarboxylates two of the propionate side chains (P) to vinyl groups (V); and (3) **protoporphyrinogen**

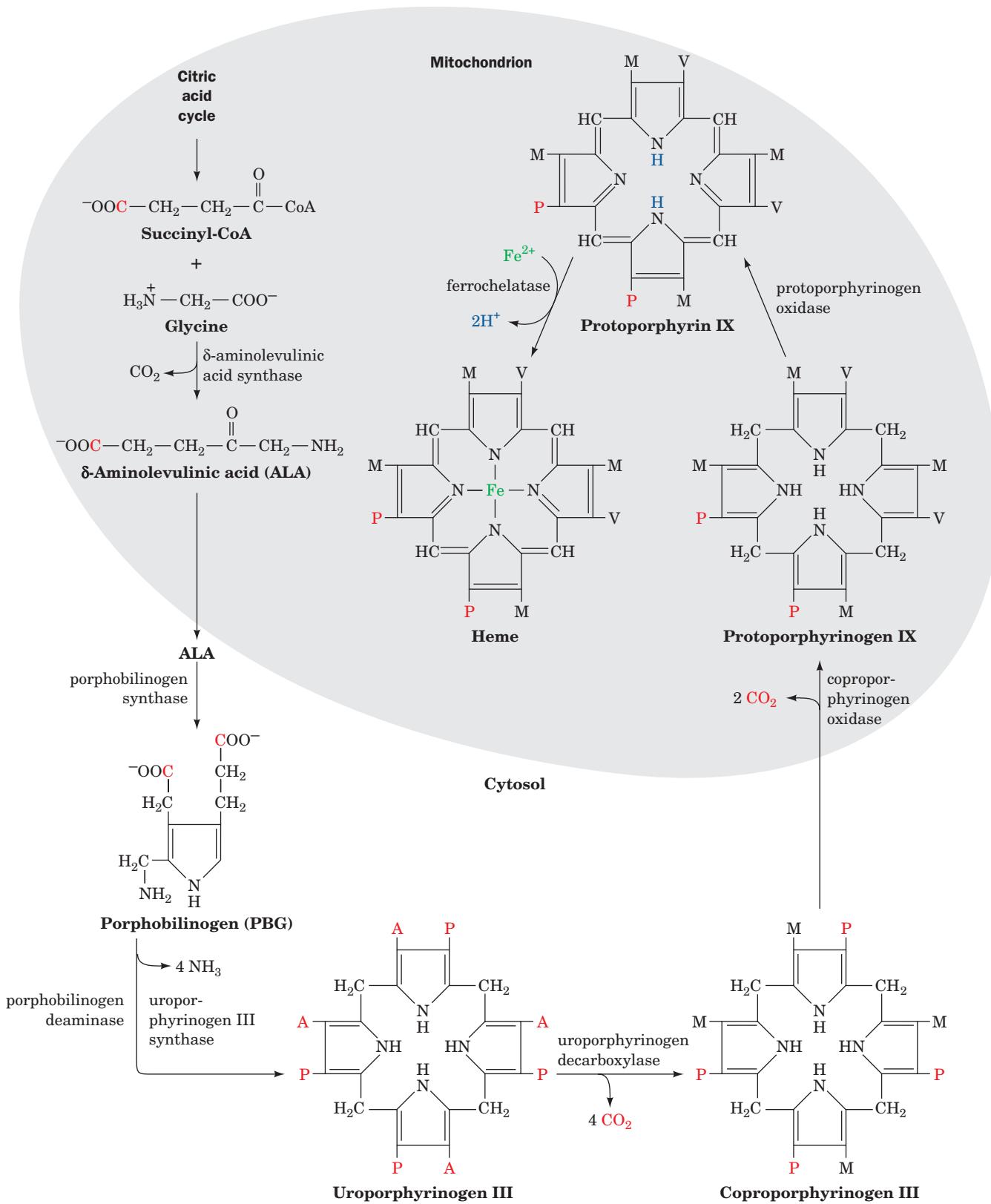


Figure 26-39 The overall pathway of heme biosynthesis. δ -Aminolevulinic acid (ALA) is synthesized in the mitochondrion by ALA synthase. ALA (left) leaves the mitochondrion and is converted to PBG, four molecules of which condense to form a porphyrin ring. The next two reactions involve oxidation of the pyrrole ring substituents yielding protoporphyrinogen IX whose

formation is accompanied by its transport back into the mitochondrion. After oxidation of the methylene groups linking the pyrroles to yield protoporphyrin IX, ferrochelatase catalyzes the insertion of Fe^{2+} to yield heme. A, P, M, and V, respectively, represent acetyl, propionyl, methyl, and vinyl ($-\text{CH}_2=\text{CH}_2$) groups. C atoms originating as the carboxyl group of acetate are red.

oxidase, which oxidizes the methylene groups linking the pyrrole rings to methenyl groups. Altogether, six carboxyl groups originally from carboxyl-labeled acetate are lost as CO_2 . The only remaining C atoms from carboxyl-labeled acetate are the carboxyl groups of heme's two propionate side chains (P). During the coproporphyrinogen oxidase reaction, the macrocycle is transported back into the mitochondrion for the pathway's final reactions.

g. Ferrochelatase Catalyzes the Insertion of Fe(II) Into Protoporphyrin IX to Form Heme

Protoporphyrin IX is converted to heme by the insertion of Fe(II) into the tetrapyrrole nucleus by **ferrochelatase**, a protein that is associated with the inner mitochondrial membrane on the matrix side. The X-ray structure of human ferrochelatase in complex with its heme product (Fig. 26-40), determined by Harry Dailey and William Lanzilotta, reveals that the 369-residue subunits of this homodimeric protein consist of two structurally similar domains and a C-terminal extension that occurs only in animal ferrochelatases (Fig. 26-40). This C-terminal extension participates in hydrogen bonding between the monomers; bacterial ferrochelatases, which lack

this extension, are monomeric although they are otherwise structurally similar to the human enzyme despite their only $\sim 10\%$ sequence identity. In addition, the C-terminal extension is bound to the N-terminal domain by an unusual [2Fe-2S] cluster that is coordinated by C196 of the N-terminal domain and C403, C406, and C411 of the C-terminal extension. The function of this [2Fe-2S] cluster, which is distant from the active site, is unclear although it appears likely that it only has a structural role. Nevertheless, three mutations, C406Y, C406S, and C411G, that inactivate the enzyme and thereby cause the rare inherited disease **erythropoietic protoporphyrinia** (see below) demonstrate the importance of the [2Fe-2S] cluster for activity.

The ferrochelatase active site (Fig. 26-40) consists of two hydrophobic lips that, it has been proposed, participate in the enzyme's association with the inner mitochondrial membrane. The ferrochelatase reaction follows an ordered mechanism in which the Fe(II) binds to the enzyme before the porphyrin. The reaction requires that the two pyrrole NH protons be removed from the porphyrin prior to the binding of the Fe(II) (Fig. 26-39). The invariant H263 appears properly positioned to abstract these protons from the porphyrin (Fig. 26-40), a hypothesis that is supported by mutagenesis studies. Structural and spectroscopic studies indicate that the metalation reaction is accompanied by the folding of the porphyrin by $\sim 12^\circ$ to a nonplanar conformation. Product release is then facilitated by the partial unwinding of a structurally conserved π helix (Fig. 26-40; π helices are discussed in Section 8-1Bb).

h. Heme Biosynthesis Is Regulated Differently in Erythroid and Liver Cells

The two major sites of heme biosynthesis are erythroid cells, which synthesize $\sim 85\%$ of the body's heme groups, and the liver, which synthesizes $\sim 80\%$ of the remainder. An important function of heme in liver is as the prosthetic groups of the **cytochromes P450**, a family of oxidative enzymes involved in detoxification (Section 15-4Bc), whose members are required throughout a liver cell's lifetime in amounts that vary with conditions. In contrast, erythroid cells, in which heme is, of course, a hemoglobin component, engage in heme synthesis only on differentiation, when they synthesize hemoglobin in vast quantities. This is a one-time synthesis; the heme must last the erythrocyte's lifetime (normally 120 days) since heme and hemoglobin synthesis stop on red cell maturation (protein synthesis stops on the loss of nuclei and ribosomes). The different ways in which heme biosynthesis is regulated in liver and in erythroid cells reflect these different demands: In liver, heme biosynthesis must really be "controlled," whereas in erythroid cells, the process is more like breaking a dam.

In liver, the main control target in heme biosynthesis is ALA synthase, the enzyme catalyzing the pathway's first committed step. Heme, or its Fe(III) oxidation product **hemin**, controls this enzyme's activity through three mechanisms: (1) feedback inhibition, (2) inhibition of the transport of ALA synthase (ALAS) from its site of synthesis in the cytosol to its reaction site in the mitochondrion (Fig. 26-39), and (3) repression of ALAS synthesis.

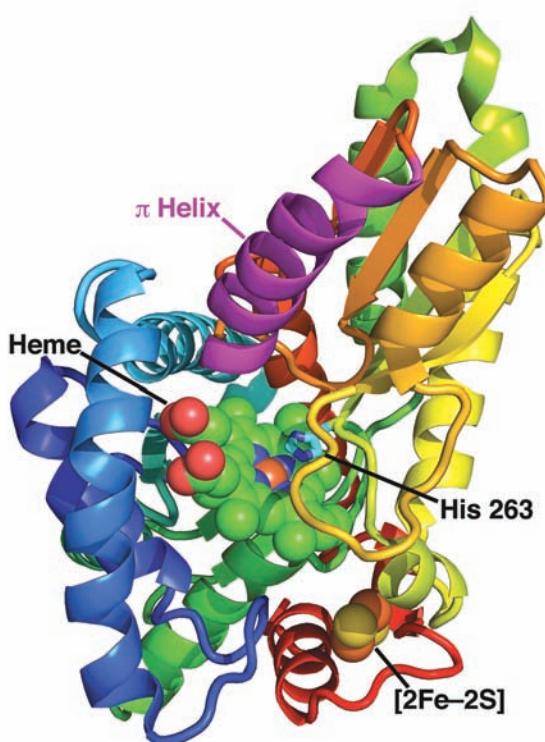


Figure 26-40 X-ray structure of human ferrochelatase in complex with its heme product. A subunit of this homodimeric protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red), but with its conserved π helix (residues 340–360) magenta. The bound heme and [2Fe-2S] cluster are shown in space-filling form with C green, N blue, O red, S yellow, and Fe orange. The side chain of the catalytically essential His 263 is drawn in stick form with C cyan and N blue. [Based on an X-ray structure by Harry Dailey and William Lanzilotta, University of Georgia. PDBid 2QD3.]

In erythroid cells, heme exerts quite a different effect on its biosynthesis. Heme induces, rather than represses, protein synthesis in reticulocytes (immature erythrocytes). Although the vast majority of the protein synthesized by reticulocytes is globin, heme may also induce these cells to synthesize the enzymes of the heme biosynthesis pathway. Moreover, the rate-determining step of heme biosynthesis in erythroid cells may not be the ALA synthase reaction, which in mammalian reticulocytes is catalyzed by a different isozyme (**ALAS-2**) than the ALA synthase that is expressed in other cells (**ALAS-1**). Experiments on various systems of differentiating erythroid cells implicate ferrochelatase and porphobilinogen deaminase in the control of heme biosynthesis in these cells. There are also indications that cellular uptake of iron may be rate limiting. Iron is transported in the plasma complexed with the iron transport protein **transferrin**. The rate at which the iron–transferrin complex enters most cells, including those of liver, is controlled by receptor-mediated endocytosis (Section 12-5Bc). However, lipid-soluble iron complexes that diffuse directly into reticulocytes stimulate *in vitro* heme biosynthesis. The existence of several control points supports the supposition that when erythroid heme biosynthesis is “switched on,” all of its steps function at their maximal rates rather than any one step limiting the flow through the pathway. Heme-stimulated synthesis of globin also ensures that heme and globin are synthesized in the correct ratio for assembly into hemoglobin (Section 32-4Aa).

i. Porphyrias Have Bizarre Symptoms

Seven sets of genetic defects in heme biosynthesis, in liver or erythroid cells, are recognized. All involve the accumulation of porphyrin and/or its precursors and are therefore known as **porphyrias** (Greek: *porphyra*, purple). Two such defects are known to affect erythroid cells: uroporphyrinogen III synthase deficiency (**congenital erythropoietic porphyria**) and ferrochelatase deficiency (**erythropoietic protoporphyrinia**). The former results in accumulation of uroporphyrinogen I and its decarboxylation product **coproporphyrinogen I**. Excretion of these compounds colors the urine red, their deposition in the teeth turns them a fluorescent reddish brown, and their accumulation in the skin renders it extremely photosensitive such that it ulcerates and forms disfiguring scars. Increased hair growth is also observed in afflicted individuals such that fine hair may cover much of the face and extremities. These symptoms have prompted speculation that the werewolf legend has a biochemical basis.

The most common porphyria that primarily affects liver is porphobilinogen deaminase deficiency (**acute intermittent porphyria**). This disease is marked by intermittent attacks of abdominal pain and neurological dysfunction, often brought about by infection, fasting, certain drugs, alcohol, steroids, and other chemicals, all of which induce the expression of ALAS-1. Excessive amounts of ALA and PBG are excreted in the urine during and after such attacks. The urine may become red resulting from the excretion of excess porphyrins synthesized from PBG in nonhepatic cells although the skin does not become unusually photosensitive. King George III,

who ruled England during the American Revolution, and who has been widely portrayed as being mad, in fact had attacks characteristic of acute intermittent porphyria, was reported to have urine the color of port wine, and had several descendants who were diagnosed as having this disease. American history might have been quite different had George III not inherited this metabolic defect.

j. Heme Is Degraded to Bile Pigments

At the end of their lifetime, red cells are removed from the circulation and their components degraded. Heme catabolism (Fig. 26-41) begins with oxidative cleavage, by **heme oxygenase**, of the porphyrin between rings A and B to form **biliverdin**, a green linear tetrapyrrole. Biliverdin’s central methenyl bridge (between rings C and D) is then reduced to form the red-orange **bilirubin**. The changing colors of a healing bruise are a visible manifestation of heme degradation.

The highly lipophilic bilirubin is insoluble in aqueous solutions. Like other lipophilic metabolites, such as free fatty acids, it is transported in the blood in complex with serum albumin. In the liver, its aqueous solubility is increased by esterification of its two propionate side groups with glucuronic acid, yielding **bilirubin diglucuronide**, which is secreted into the bile. Bacterial enzymes in the large intestine hydrolyze the glucuronic acid groups and, in a multistep process, convert bilirubin to several products, most notably **urobilinogen**. Some urobilinogen is reabsorbed and transported via the bloodstream to the kidney, where it is converted to the yellow **urobilin** and excreted, thus giving urine its characteristic color. Most of the urobilinogen, however, is microbially converted to the deeply red-brown **stercobilin**, the major pigment of feces.

When the blood contains excessive amounts of bilirubin, the deposition of this highly insoluble substance colors the skin and the whites of the eyes yellow. This condition, called **jaundice** (French: *jaune*, yellow), signals either an abnormally high rate of red cell destruction, liver dysfunction, or bile duct obstruction. Newborn infants, particularly when premature, often become jaundiced because their livers do not yet make sufficient **bilirubin UDP-glucuronosyltransferase** to glucuronidate the incoming bilirubin. Jaundiced infants are treated by bathing them with light from a fluorescent lamp; this photochemically converts bilirubin to more soluble isomers that the infant can degrade and excrete.

k. Hemoglobin’s Reduced Affinity for CO Prevents Asphyxiation

In the reaction forming biliverdin, the methenyl bridge carbon between porphyrin rings A and B is released as CO (Fig. 26-41, *top*), which, we have seen, is a tenacious heme ligand (with 200-fold greater affinity for hemoglobin and myoglobin than O₂; Section 10-1A). Consequently, ~1% of hemoglobin’s O₂-binding sites are blocked by CO, even in the absence of air pollution. However, free heme in solution binds CO with 20,000-fold greater affinity than it binds O₂. Thus, the globin (protein) portion of hemoglobin (and likewise myoglobin) somehow lowers the affinity of its bound heme for CO, thereby making O₂ transport possible. How does the globin do so?

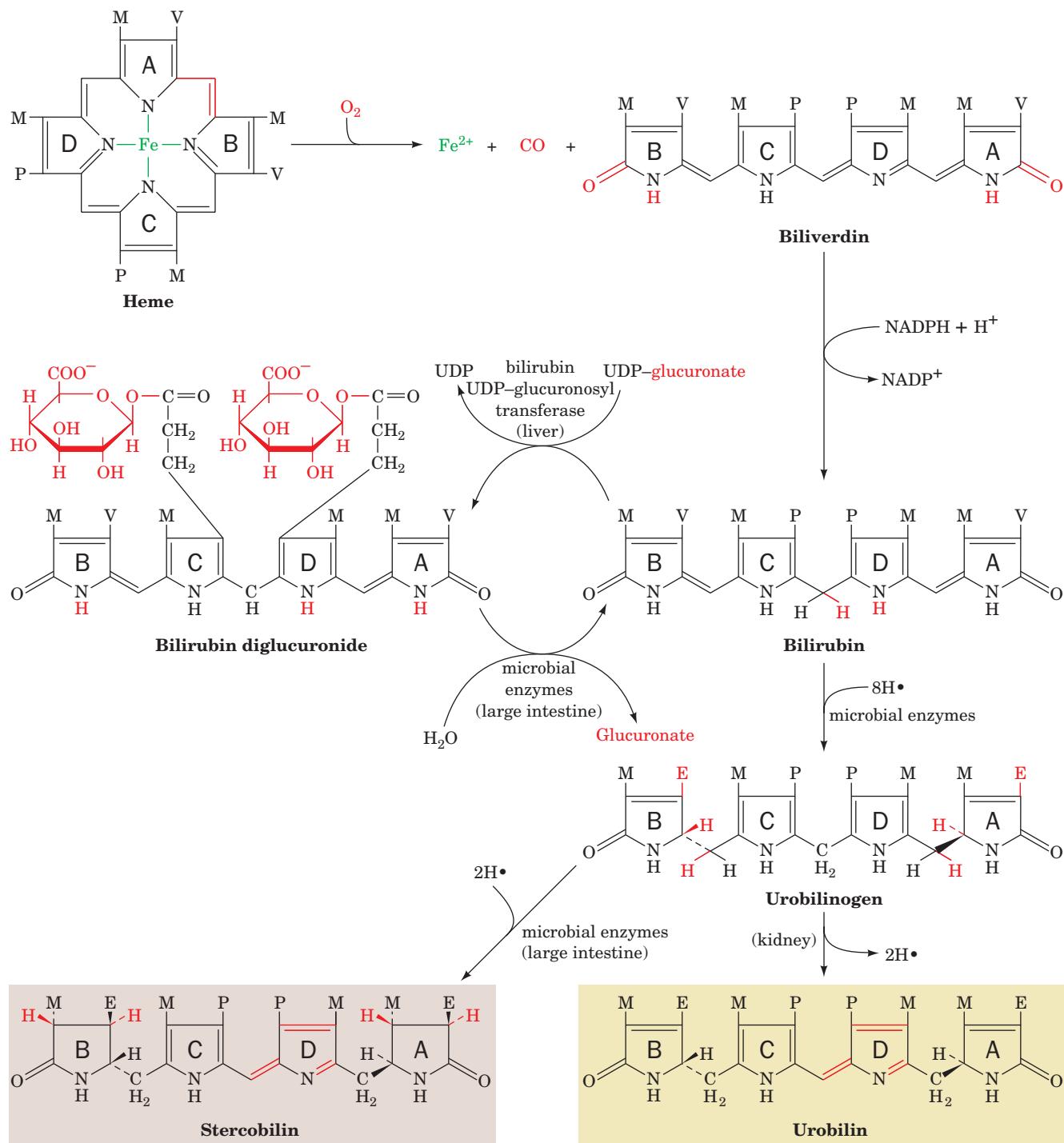


Figure 26-41 The heme degradation pathway. M, V, P, and E, respectively, represent methyl, vinyl, propionyl, and ethyl groups.

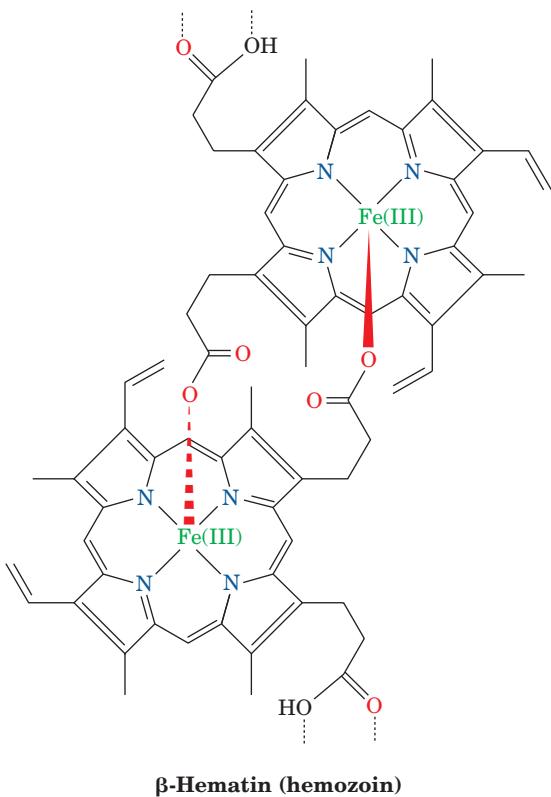
Early X-ray structures of **carboxymyoglobin** (myoglobin with a CO ligand) indicated that the bound CO was inclined from the normal to the heme plane by 40° to 60° (the Fe—C—O bond angle appeared to be 120° to 140°), approximately the same angle with which O₂ binds to heme (Fig. 10-12). Yet, in complexes of CO with porphyrins in the absence of protein, the CO is normal to the heme plane. This suggested that the globin (in both myoglobin and

hemoglobin) sterically bends the bound CO away from its preferred linear geometry, thereby reducing its affinity for CO and hence permitting the CO to be slowly exhaled. However, a variety of spectroscopic investigations together with highly accurate X-ray structures of carboxymyoglobin revealed that the bound CO is, in fact, inclined from the normal to the heme plane by ~7°, a distortion that is too small to explain the reduced affinity of myoglobin for CO.

Of course, this reduced affinity might instead be explained by the distortions that the upright CO ligand imposes on the globin, presumably via the distal His (E7, the His residue that hydrogen bonds to the bound O₂; Section 10-2A). However, studies of the energetics of binding of CO and O₂ to myoglobins in which His E7 has been mutated to nonpolar residues of comparable bulk (e.g., Leu) indicate that this is not the main determinant of the ligand affinity changes. Rather, the reduction in affinity of myoglobin for CO relative to that for O₂ has been shown to arise from the greater hydrogen bonding affinity that His E7 has for O₂ relative to CO together with electrostatic effects due to the differing charge distributions in the O₂ and CO ligands.

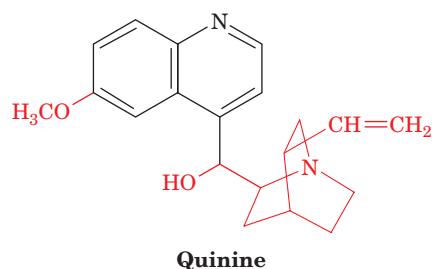
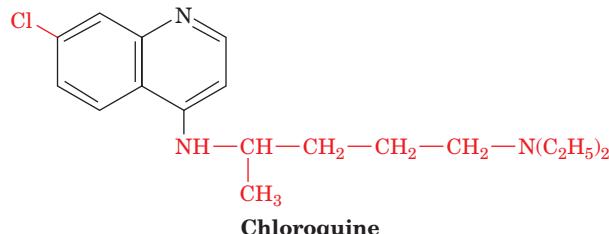
I. Chloroquine Prevents Malaria by Inhibiting Plasmodial Heme Sequestration

Malaria is caused by the mosquito-borne parasite *Plasmodium falciparum* (Section 7-3Ab), which multiplies within and destroys red blood cells in a 2-day cycle. During the intraerythrocytic stages of its life cycle, the parasite partially meets its nutritional needs by proteolyzing up to ~80% of the host cell's hemoglobin in its so-called acid food vacuole, whose pH is 4.7. This process releases heme, which in its soluble form, is toxic to the parasite because it damages cell membranes and inhibits a variety of enzymes. Since, unlike their human hosts, plasmodia cannot degrade heme, they sequester it within their food vacuoles in the form of harmless dark brown granules known as **hemozoin**, which consist of crystals of dimerized hemes linked together by reciprocal iron–carboxylate bonds between the ferric ions and the propionate side chains of adjacent molecules. Hemozoin has been found to be identical to **β-hematin**,



whose X-ray structure has been determined. Dimers interact in the crystals through hydrogen bonds between the remaining carboxyl groups.

Chloroquine,

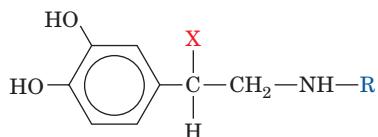


a member of the quinoline ring-containing family of antimalarials, which includes **quinine**, is one of the most successful antimicrobial agents that has been produced. It is effective against plasmodia only during their intraerythrocytic stages. This drug, being a weak base that can readily pass through membranes in its uncharged form, accumulates in the plasmodial acid food vacuole in its acidic (charged) form in millimolar concentrations. Chloroquine and several other quinoline-containing antimalarials inhibit the crystallization of hemes to form hemozoin. This inhibition *in vivo* is almost certainly responsible for the antimalarial properties of these drugs. The mechanism of inhibition is as yet unclear although a plausible hypothesis is that the drug adsorbs onto crystallized hemozoin, inhibiting further crystallization.

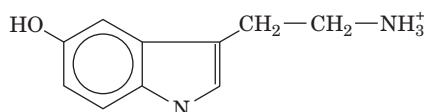
The massive use of chloroquine has, unfortunately, led to the appearance of chloroquine-resistant plasmodia in nearly every malarial region of the world. Resistant plasmodia do not concentrate chloroquine in their food vacuoles to the high levels found in sensitive parasites. Rather, they export this drug out of their food vacuoles at an ~50-fold higher rate than do sensitive organisms. Since chloroquine activity and chloroquine resistance have different mechanisms, it has been possible to modify existing quinoline-containing structures and to develop new hemozoin crystallization inhibitors that are effective antimalarial agents but to which plasmodia are not (yet) resistant.

B. Biosynthesis of Physiologically Active Amines

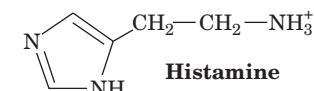
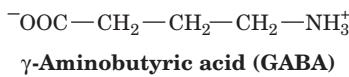
Epinephrine, norepinephrine, dopamine, serotonin (5-hydroxytryptamine), γ-aminobutyric acid (GABA), and

histamine

$X = \text{OH}$, $R = \text{CH}_3$ **Epinephrine (Adrenalin)**
 $X = \text{OH}$, $R = \text{H}$ **Norepinephrine**
 $X = \text{H}$, $R = \text{H}$ **Dopamine**



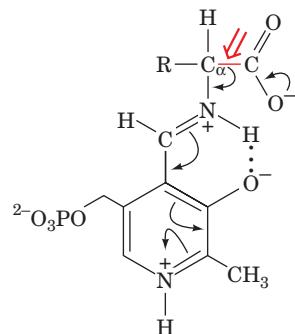
Serotonin
(5-hydroxytryptamine)



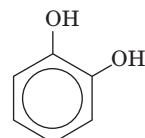
are hormones and/or neurotransmitters derived from amino acids. For instance, epinephrine, as we have seen, activates muscle adenylate cyclase, thereby stimulating glycogen breakdown (Section 18-3E); deficiency in dopamine production is associated with **Parkinson's disease**, a degenerative condition causing "shaking palsy"; serotonin causes smooth muscle contraction; GABA is one of the brain's major inhibitory neurotransmitters (Section 20-5Cf), being released at 30% of its synapses; and histamine is involved in allergic responses (as allergy sufferers who take antihistamines will realize), as well as in the control of acid secretion by the stomach (Section 20-3C).

The biosynthesis of each of these physiologically active amines involves decarboxylation of the corresponding precursor amino acid. Amino acid decarboxylases are PLP-dependent enzymes that form a PLP-Schiff base with the

substrate so as to stabilize the C_α carbanion formed on $C_\alpha-\text{COO}^-$ bond cleavage (Section 26-1Aa):



Formation of histamine and GABA are one-step processes (Fig. 26-42). In the synthesis of serotonin from tryptophan, the decarboxylation is preceded by a hydroxylation (Fig. 26-43) by **tryptophan hydroxylase**, one of three mammalian enzymes that has a 5,6,7,8-tetrahydrobiopterin cofactor (Section 26-3Ha). This hydroxylation involves an NIH shift similar to that occurring in phenylalanine hydroxylase (Fig. 26-30), although no epoxide intermediate has been observed in this case. Dopamine, norepinephrine, and epinephrine are all termed **catecholamines** because they are amine derivatives of **catechol**:



Catechol

The conversion of tyrosine to these various catecholamines occurs as follows (Fig. 26-44):

1. Tyrosine is hydroxylated to **3,4-dihydroxyphenylalanine (L-DOPA)** by **tyrosine hydroxylase**, another 5,6,7,8-tetrahydrobiopterin-requiring enzyme.
2. L-DOPA is decarboxylated to dopamine.
3. A second hydroxylation yields norepinephrine.

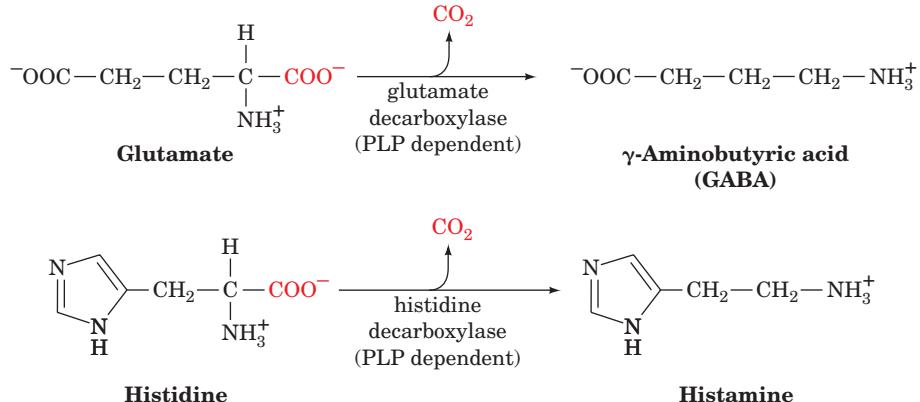
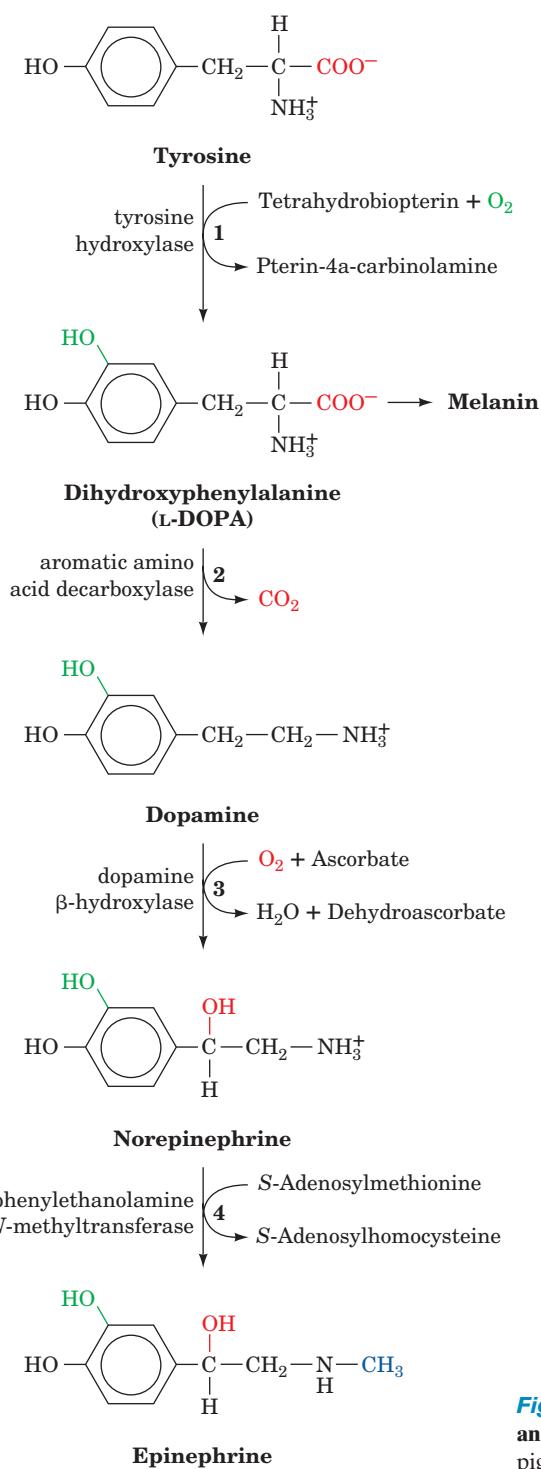
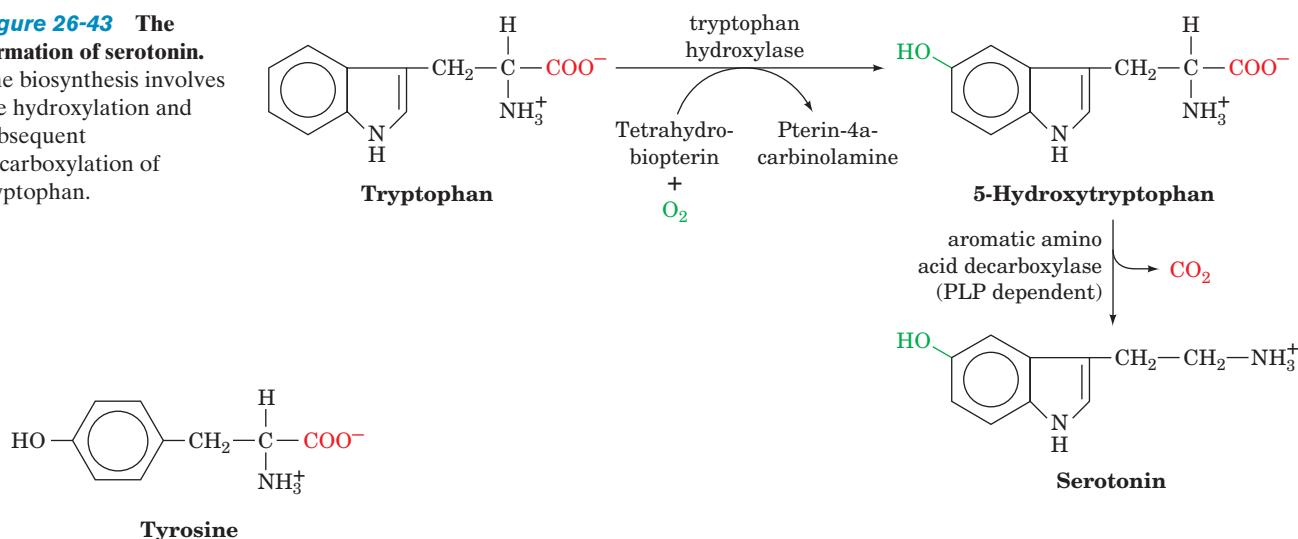


Figure 26-42 The formation of **γ -aminobutyric acid (GABA)** and **histamine**. The reactions involve the decarboxylations of glutamate to form GABA and of histidine to form histamine.

Figure 26-43 The formation of serotonin.

The biosynthesis involves the hydroxylation and subsequent decarboxylation of tryptophan.

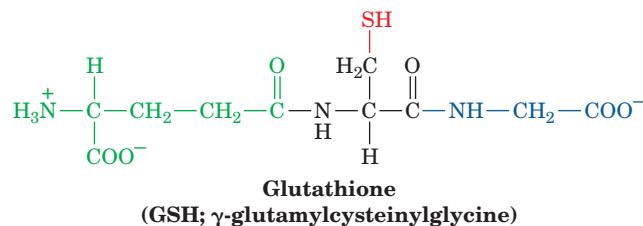


4. Methylation of norepinephrine's amino group by *S*-adenosylmethionine (SAM; Section 26-3Ea) produces epinephrine.

The specific catecholamine that a cell produces depends on which enzymes of the pathway are present. In adrenal medulla, which functions to produce hormones (Section 19-1F), epinephrine is the predominant product. In some areas of the brain, norepinephrine is more common. In other areas, most prominently the **substantia nigra**, the pathway stops at dopamine synthesis. Indeed, Parkinson's disease, which is caused by degeneration of the substantia nigra, has been treated with some success by the administration of L-DOPA, dopamine's immediate precursor. Dopamine itself is ineffective because it cannot cross the blood-brain barrier. L-DOPA, however, is able to get to its sites of action where it is decarboxylated to dopamine. The enzyme catalyzing this reaction, **aromatic amino acid decarboxylase**, decarboxylates all aromatic amino acids and is therefore also responsible for serotonin formation. L-DOPA is also a precursor of the black skin pigment melanin.

C. Glutathione

Glutathione (GSH; γ -glutamylcysteinylglycine),



a tripeptide that contains an unusual γ -amide bond, participates in a variety of detoxification, transport, and metabolic processes (Fig. 26-45). For instance, it is a substrate for

Figure 26-44 The sequential synthesis of L-DOPA, dopamine, norepinephrine, and epinephrine from tyrosine. L-DOPA is also the precursor of the black skin pigment melanin, an oxidized polymeric material.

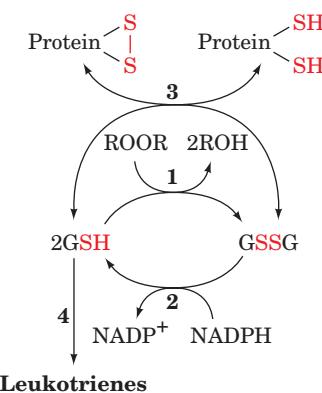
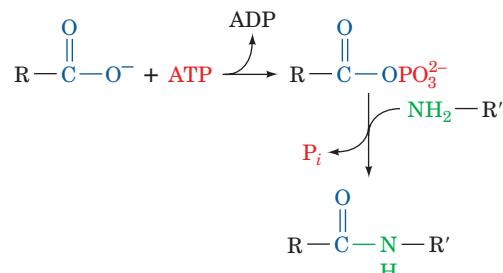


Figure 26-45 Some reactions involving glutathione. The reactions and enzymes are (1) peroxide detoxification by **glutathione peroxidase**, (2) regeneration of GSH from GSSG by **glutathione reductase** (Section 21-2Ba), (3) thiol transferase modulation of protein thiol-disulfide balance, and (4) leukotriene biosynthesis by a glutathione-S-transferase.

peroxidase reactions, helping to destroy peroxides generated by oxidases; it is involved in leukotriene biosynthesis (Section 25-7Cb); and the balance between its reduced (GSH) and oxidized (GSSG) forms maintains the sulfhydryl groups of intracellular proteins in their correct oxidation states.

The **γ -glutamyl cycle**, which was elucidated by Alton Meister, provides a vehicle for the energy-driven transport

of amino acids into cells through the synthesis and breakdown of GSH (Fig. 26-46). GSH is synthesized from glutamate, cysteine, and glycine by the consecutive action of **γ -glutamylcysteine synthetase** and **GSH synthetase** (Fig. 26-46, Reactions 1 and 2). ATP hydrolysis provides the free energy for each reaction. The carboxyl group is activated for peptide bond synthesis by formation of an acyl phosphate intermediate:



The breakdown of GSH is catalyzed by **γ -glutamyl transpeptidase**, **γ -glutamyl cyclotransferase**, **5-oxoprolinase**, and an intracellular protease (Fig. 26-46, Reactions 3–6).

Amino acid transport occurs because, whereas GSH is synthesized intracellularly and is located largely within the cell, γ -glutamyl transpeptidase, which catalyzes GSH breakdown (Fig. 26-46, Reaction 3), is situated on the cell membrane's external surface and accepts amino acids, notably cysteine and methionine. GSH is first transported to the

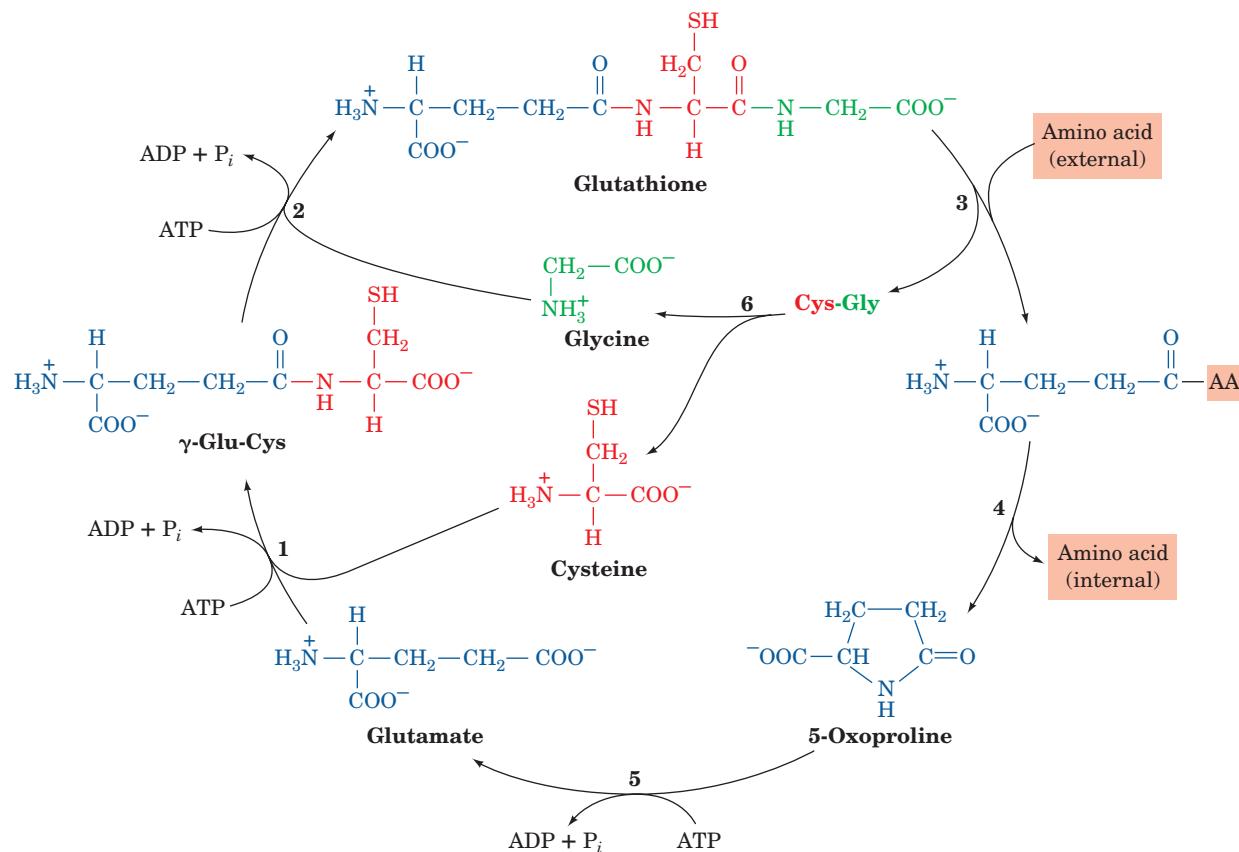


Figure 26-46 Glutathione synthesis as part of the γ -glutamyl cycle of glutathione metabolism. The cycle's reactions are catalyzed by (1) γ -glutamylcysteine synthetase, (2) glutathione

synthetase, (3) γ -glutamyl transpeptidase, (4) γ -glutamyl cyclotransferase, (5) 5-oxoprolinase, and (6) an intracellular protease.

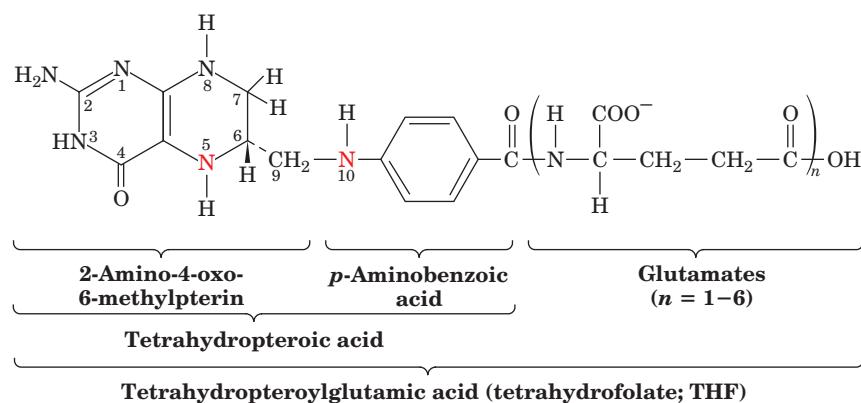


Figure 26-47 Tetrahydrofolate (THF).

external surface of the cell membrane, where the transfer of the γ -glutamyl group from GSH to an external amino acid occurs. The γ -glutamyl amino acid is then transported back into the cell and converted to glutamate by a two-step process in which the transported amino acid is released and **5-oxoproline** is formed as an intermediate. The last step in the cycle, the hydrolysis of 5-oxoproline, requires ATP hydrolysis. This surprising observation (amide bond hydrolysis is almost always an exergonic process) is a consequence of 5-oxoproline's unusually stable internal amide bond.

D. Tetrahydrofolate Cofactors: The Metabolism of C_1 Units

Many biosynthetic processes involve the addition of a C_1 unit to a metabolic precursor. A familiar example is carboxylation. For instance, gluconeogenesis from pyruvate begins with the addition of a carboxyl group to form oxaloacetate (Section 23-1Aa). The coenzyme involved in this and most other carboxylation reactions is biotin (Section 23-1Ab). In contrast, *S*-adenosylmethionine functions as a methylating agent (Section 26-3Ea).

Tetrahydrofolate (THF) is more versatile than the above cofactors in that it functions to transfer C_1 units in several oxidation states. THF is a 6-methylpterin derivative linked in sequence to *p*-aminobenzoic acid and Glu residues (Fig. 26-47). Up to five additional Glu residues may be linked to the first glutamate via isopeptide bonds to form a polyglutamyl tail.

THF is derived from **folic acid** (Latin: *folium*, leaf), a doubly oxidized form of THF that must be enzymatically reduced before it becomes an active coenzyme (Fig. 26-48). Both reductions are catalyzed by **dihydrofolate reductase (DHFR)**. Mammals cannot synthesize folic acid and so must obtain it from their diets or from intestinal microorganisms.

C_1 units are covalently attached to THF at its positions N5, N10, or both N5 and N10. These C_1 units, which may be at the oxidation levels of formate, formaldehyde, or methanol (Table 26-1), are all interconvertible by enzymatic redox reactions (Fig. 26-49).

The main entry of C_1 units into the THF pool is as **N^5,N^{10} -methylene-THF** through the conversion of serine to glycine by serine hydroxymethyltransferase (Sections 26-3Bb and 26-5Ae) and the cleavage of glycine by glycine synthase (the glycine cleavage system; Section 26-3Ba, Fig. 26-14). Histidine also contributes C_1 units through its degradation with the formation of **N^5 -formimino-THF** (Fig. 26-17, Reaction 11).

A C_1 unit in the THF pool can have several fates (Fig. 26-50):

1. It may be used directly as N^5,N^{10} -methylene-THF in the conversion of the deoxynucleotide dUMP to dTMP by **thymidylate synthase** (Section 28-3Bb).
2. It may be reduced to **N^5 -methyl-THF** for the synthesis of methionine from homocysteine (Section 26-3Ea).
3. It may be oxidized through N^5,N^{10} -methenyl-THF to **N^{10} -formyl-THF** for use in the synthesis of purines (Section

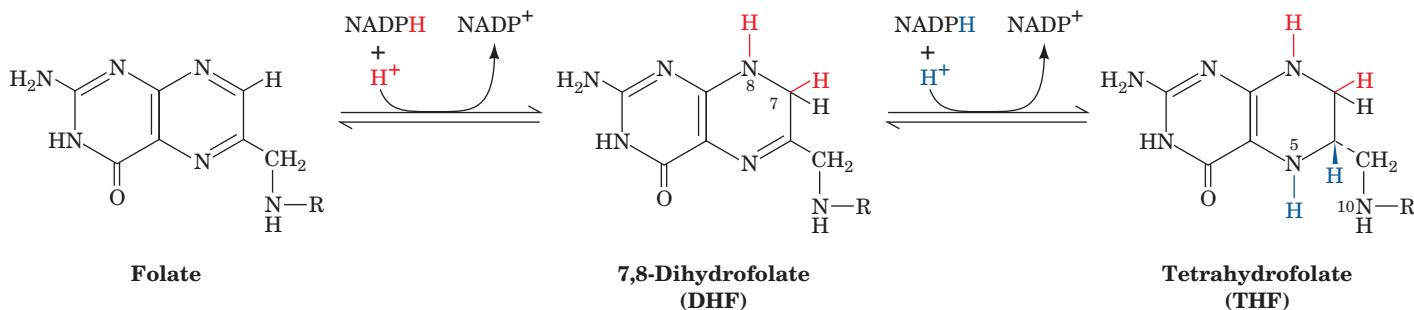


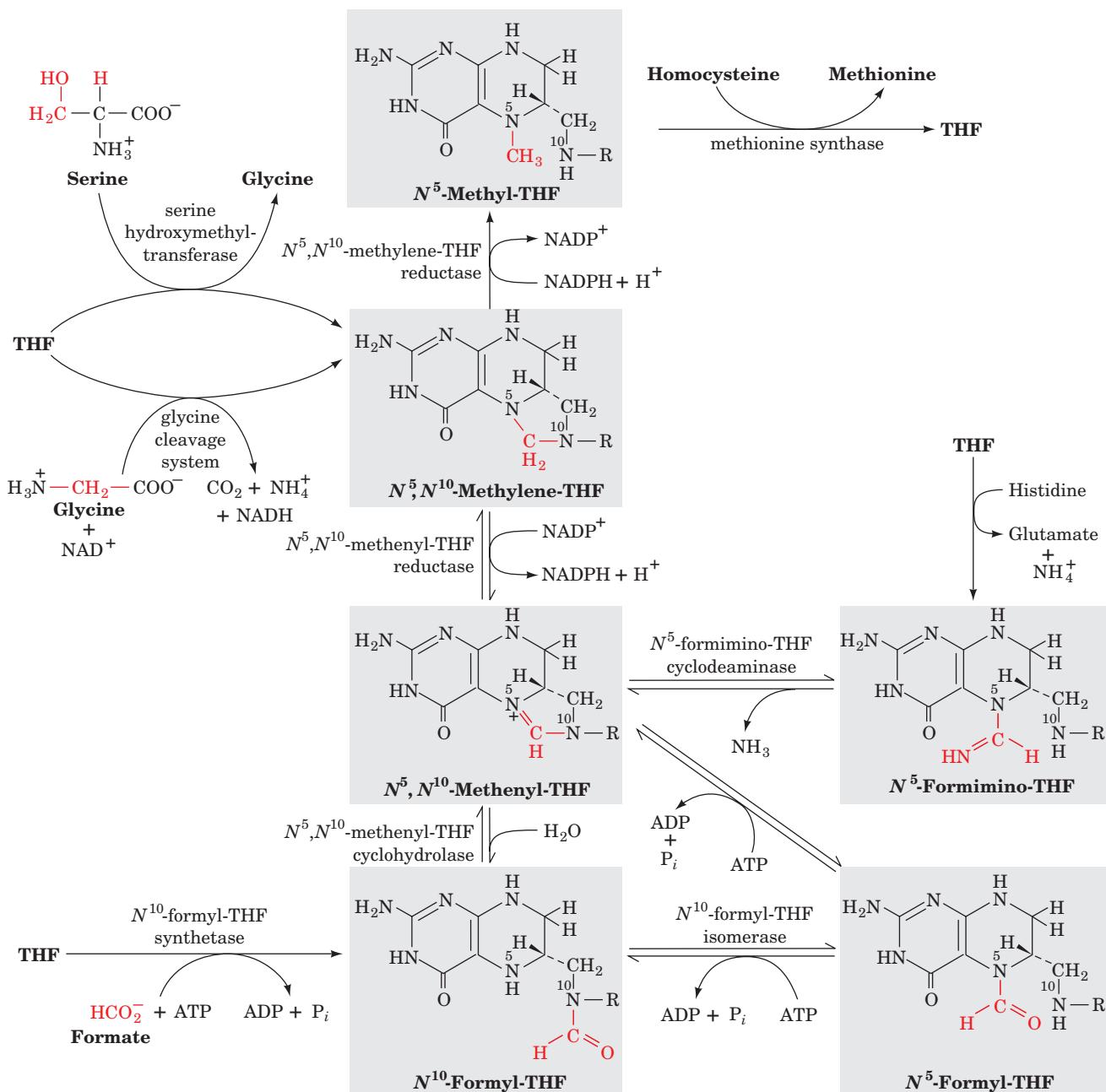
Figure 26-48 The two-stage reduction of folate to THF. Both reactions are catalyzed by dihydrofolate reductase (DHFR).

Table 26-1 Oxidation Levels of C₁ Groups Carried by THF

Oxidation Level	Group Carried	THF Derivative(s)
Methanol	Methyl (—CH ₃)	N ⁵ -Methyl-THF
Formaldehyde	Methylene (—CH ₂ —)	N ⁵ ,N ¹⁰ -Methylene-THF
Formate	Formyl (—CH=O)	N ⁵ -Formyl-THF, N ¹⁰ -formyl-THF
	Formimino (—CH=NH)	N ⁵ -Formimino-THF
	Methenyl (—CH=)	N ⁵ ,N ¹⁰ -Methenyl-THF

28-1A). Since the purine ring of ATP is involved in histidine biosynthesis in microorganisms and plants (Section 26-5Be), N¹⁰-formyl-THF is indirectly involved in this pathway as

well. Prokaryotes use N¹⁰-formyl-THF in a formylation reaction yielding **formylmethionyl-tRNA**, which they require for the initiation of protein synthesis (Section 32-3Ca).

**Figure 26-49** Interconversion of the C₁ units carried by THF.

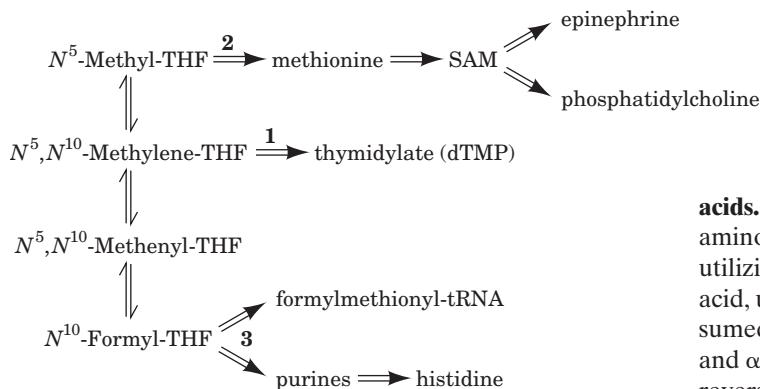
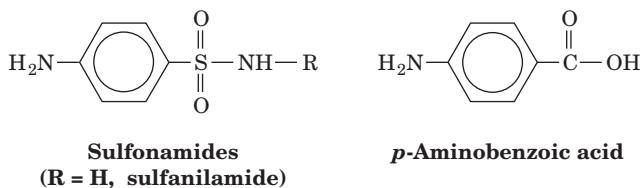


Figure 26-50 The biosynthetic fates of the C₁ units in the THF pool.

A deficiency in folic acid results in **megaloblastic anemia**, a condition in which erythrocytes in the blood are replaced by fewer abnormally large erythrocytes known as **macrococytes**. This occurs as a consequence of the lack of N⁵,N¹⁰-methylene-THF, which is required for the synthesis of the DNA precursor deoxythymidine monophosphate from deoxyuridine monophosphate (Section 28-3Bb). Pernicious anemia (vitamin B₁₂ deficiency; Section 25-2Ee) has similar anemic symptoms. This is because the irreversibility of the reaction forming N⁵-methyl-THF and the now inactive coenzyme B₁₂-dependent methionine synthase (top of Fig. 26-49), the only mammalian enzyme that utilizes N⁵-methyl-THF, result in folates becoming trapped in their N⁵-methyl-THF form. Consequently, the anemic symptoms of pernicious anemia may be alleviated by the administration of folic acid but this does not relieve its neurological symptoms.

Sulfonamides (sulfa drugs) such as **sulfanilamide** are antibiotics that are structural analogs of the *p*-aminobenzoic acid constituent of THF:



They competitively inhibit bacterial synthesis of THF at the *p*-aminobenzoic acid incorporation step, thereby blocking the above THF-requiring reactions. The inability of mammals to synthesize folic acid leaves them unaffected by sulfonamides, which accounts for the medical utility of these widely used antibacterial agents.

5 AMINO ACID BIOSYNTHESIS

Many amino acids are synthesized by pathways that are present only in plants and microorganisms. Since mammals must obtain these amino acids in their diets, these substances are known as **essential amino acids**. The other amino acids, which can be synthesized by mammals from common intermediates, are termed **nonessential amino**

acids. Their α -keto acid carbon skeletons are converted to amino acids by transamination reactions (Section 26-1A) utilizing the preformed α -amino nitrogen of another amino acid, usually glutamate. Yet, although it was originally presumed that glutamate can be synthesized from ammonia and α -ketoglutarate by glutamate dehydrogenase acting in reverse, it now appears that the predominant physiological direction of this enzyme is glutamate breakdown (Section 26-1B). Consequently, *preformed α -amino nitrogen should also be considered to be an essential nutrient*. In this context, it is interesting to note that, in addition to the four well-known taste receptors, those for sweet, sour, salty, and bitter tastes, a fifth taste receptor has been characterized, that for the meaty taste of **monosodium glutamate (MSG)**, which is known as **umami** (Japanese: flavor).

The essential and nonessential amino acids for humans are listed in Table 26-2. Arginine is classified as essential, even though it is synthesized by the urea cycle (Section 26-2D), because it is required in greater amounts than can be produced by this route during the normal growth and development of children (but not adults). The essential amino acids occur in animal and vegetable proteins. Different proteins, however, contain different proportions of the essential amino acids. Milk proteins, for example, contain them all in the proportions required for proper human nutrition. Bean protein, on the other hand, contains an abundance of lysine but is deficient in methionine, whereas wheat is deficient in lysine but contains ample methionine. A balanced protein diet therefore must contain a variety of different protein sources that complement each other to supply the proper proportions of all the essential amino acids.

In this section we study the pathways involved in the formation of the nonessential amino acids. We also briefly consider such pathways for the essential amino acids as they occur in plants and microorganisms. You should note, however, that although we discuss some of the most common pathways for amino acid biosynthesis, there is considerable variation in these pathways among different species. In contrast, as we have seen, the basic pathways of carbohydrate and lipid metabolism are all but universal.

A. Biosynthesis of the Nonessential Amino Acids

All the nonessential amino acids except tyrosine are synthesized by simple pathways leading from one of four common metabolic intermediates: pyruvate, oxaloacetate, α -ketoglutarate, and 3-phosphoglycerate. Tyrosine, which is really misclassified as nonessential, is synthesized by the one-step hydroxylation of the essential amino acid phenylalanine (Section 26-3H). Indeed, the dietary requirement for phenylalanine reflects the need for tyrosine as well. The presence of dietary tyrosine therefore decreases the need for phenylalanine. Since preformed α -amino nitrogen in the form of glutamate is an essential nutrient for nonessential

Table 26-2 Essential and Nonessential Amino Acids in Humans

Essential	Nonessential
Arginine ^a	Alanine
Histidine	Asparagine
Isoleucine	Aspartate
Leucine	Cysteine
Lysine	Glutamate
Methionine	Glutamine
Phenylalanine	Glycine
Threonine	Proline
Tryptophan	Serine
Valine	Tyrosine

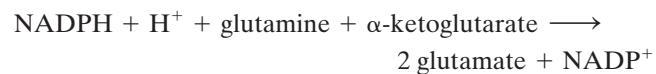
^aAlthough mammals synthesize arginine, they cleave most of it to form urea (Sections 26-2D and 26-2E).

amino acid biosynthesis, we first discuss its production by plants and microorganisms.

a. Glutamate Is Synthesized by Glutamate Synthase

Glutamate synthase, an enzyme that occurs only in microorganisms, plants, and lower animals, converts

α -ketoglutarate and ammonia originating in glutamine to glutamate. The electrons required for this reductive amination come from NADPH or ferredoxin, depending on the organism. The NADPH-dependent glutamate synthase from the nitrogen-fixing bacterium *Azospirillum brasilense*, the best characterized such enzyme, is an $\alpha_2\beta_2$ heterotetramer that binds an FAD and two [4Fe-4S] clusters on each β subunit, and an FMN and a [3Fe-4S] cluster on each α subunit. The overall reaction is



and involves five steps that occur at three distinct active sites (Fig. 26-51):

1. Electrons are transferred from NADPH to FAD at active site 1 on the β subunit to yield FADH_2 .
2. The electrons are transferred from the FADH_2 to FMN at site 2 on a specific α subunit through the iron-sulfur clusters to yield FMNH_2 .
3. Glutamine is hydrolyzed to α -glutamate and ammonia at site 3 on the α subunit.
4. The ammonia produced is transferred to site 2 where it reacts with α -ketoglutarate to form α -iminoglutarate.

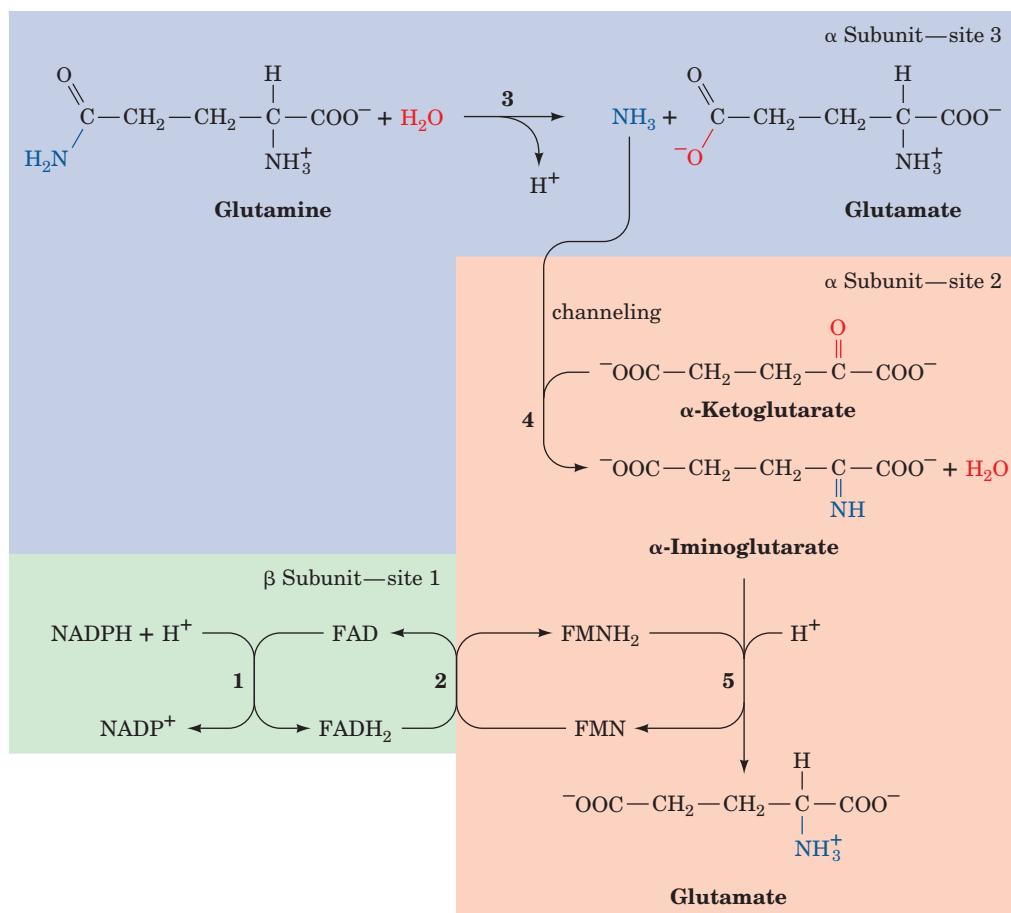


Figure 26-51 The sequence of reactions catalyzed by glutamate synthase.

5. The α -imino glutarate is reduced by FMNH_2 to form glutamate.

In the absence of the β subunit, the α subunit can synthesize glutamate from glutamine and α -ketoglutarate using an artificial electron donor; moreover, it is homologous and functionally similar to ferredoxin-dependent glutamine synthases. The *A. brasiliense* α subunit is therefore considered to be the enzyme's catalytic core.

The X-ray structure of the 1479-residue α subunit of *A. brasiliense* glutamate synthase in complex with a [3Fe-4S] cluster, an FMN, an α -ketoglutarate substrate, and a **methionine sulfone** inhibitor (a tetrahedral transition state analog)

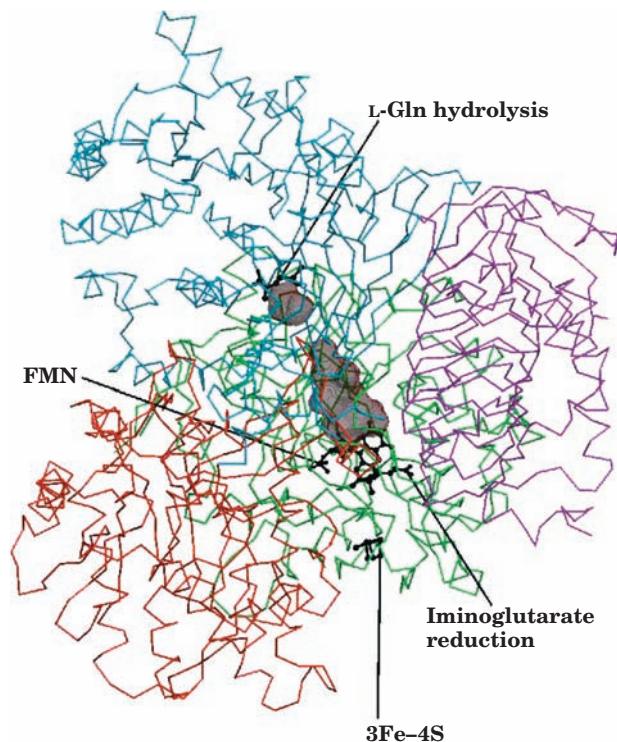
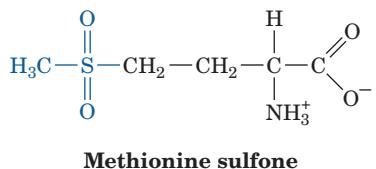
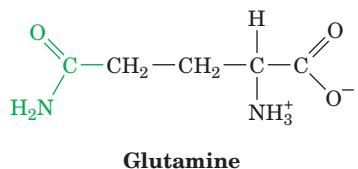


Figure 26-52 X-ray structure of the α subunit of *A. brasiliense* glutamate synthase as represented by its C_α backbone. The subunit has four domains, an N-terminal glutamine amidotransferase domain (blue) to which methionine sulfone (a glutamine tetrahedral transition state analog) is bound; a central domain (red); an FMN-binding domain (green) to which an FMN, a [3Fe-4S] cluster, and an α -ketoglutarate are bound; and a β helix domain (purple). The foregoing ligands are drawn in black in ball-and-stick form. The 31-Å-long tunnel from the methyl group on the methionine sulfone (analogous to the amido group of glutamine) to the α -keto group of the α -ketoglutarate is outlined by a gray surface. The tunnel is blocked in this structure (it is divided into two cavities) by the main chain atoms of four residues that protrude into the tunnel. [Courtesy of Andrea Mattevi, Università degli Studi di Pavia, Italy. PDBid 1EA0.]

was determined by Andrea Mattevi (Fig. 26-52). The subunit has four domains, an N-terminal glutamine amidotransferase domain, a central domain, an FMN-binding domain, and a C-terminal **β -helix** domain (see below). The methionine sulfone binds in the N-terminal amidotransferase domain (site 3), where glutamine is normally hydrolyzed via the nucleophilic attack of the Cys 1 sulphydryl group on the glutamine C_γ atom to transiently form a tetrahedral intermediate that the tetrahedral sulfone group mimics. The FMN-binding domain consists in large part of an α/β barrel, at the mouth of which the α -ketoglutarate and the [3Fe-4S] cluster bind. The distance between the methyl group on methionine sulfone (the analog of the glutamine amido group) and the α -keto group of the α -ketoglutarate is 31 Å. These two sites are connected by a tunnel through which ammonia must diffuse in order to react with α -ketoglutarate (Reaction 4, Fig. 26-51). However, the tunnel is blocked by the main chain atoms of four residues that protrude into the tunnel. Consequently, there must be at least a 2- to 3-Å shift in the structure during the reaction to permit ammonia to diffuse between these sites. Such gating of the tunnel may well be crucial for the control of enzyme function so as to avoid the wasteful hydrolysis of glutamine. In fact, glutamine is hydrolyzed only when the enzyme has bound α -ketoglutarate and reducing equivalents are available for iminoglutamate reduction.

The C-terminal domain of glutamate synthase consists largely of a 7-turn, right-handed β helix (Fig. 26-53). In this unusual motif, the polypeptide chain is wrapped in a wide helix such that neighboring turns of chain interact as do the strands of a parallel β sheet. The 43-Å-long glutamate synthase β helix has an elliptical cross section of 16 by 23 Å. β Helices have been observed in only a handful of mainly bacterial enzymes. The β helix of glutamate synthase does

not contain a residue that is involved in catalysis or electron transfer. However, it does appear to have an important structural role because some of its residues line the tunnel through which ammonia passes.

Glutamine amidotransferase domains or subunits are part of several protein structures in which glutamine is the ammonia donor for a further reaction. In glutamate synthase, it is a domain within the α subunit (Fig. 26-52). In *E. coli* carbamoyl phosphate synthetase (CPS; Section 26-2Aa), it is a complete subunit within a heterodimer. These domains or subunits occur in one of two families that are differentiated by their active site structures. In CPS, it belongs to the **triad family**, so called because it contains an active site Cys in a catalytic triad reminiscent of the catalytic triad in serine proteases (Section 15-3). The glutamine amidotransferase domain of glutamate synthase belongs to the **N-terminal nucleophile (Ntn) family**, which, as we have seen, has an N-terminal Cys that acts as the active site nucleophile. Other enzymes involved in

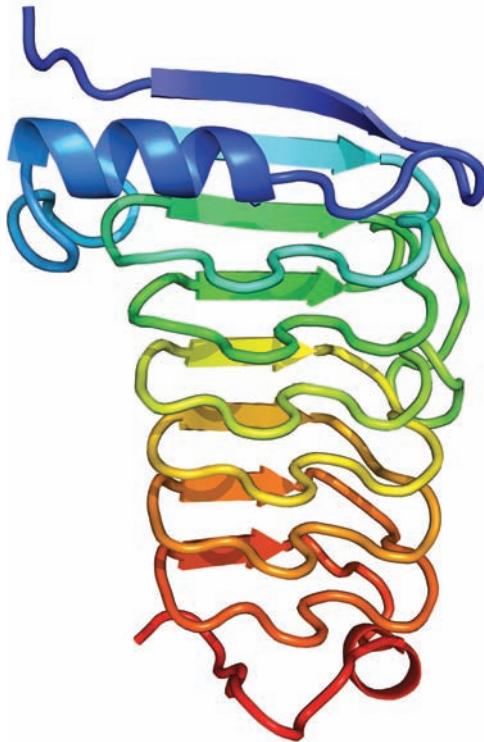


Figure 26-53 The β helix of *A. brasiliense* glutamate synthase. The polypeptide backbone (residues 1225–1416) is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). Neighboring turns of polypeptide chain within the β helix interact as do the strands of parallel β sheets. Nevertheless, the conformations of many of these segments lie outside the normal range for β strands, and hence they are drawn in coil form. [Based on an X-ray structure by Andrea Mattevi, Università degli Studi di Pavia, Italy. PDBid 1EA0.]

amino acid biosynthesis and having a glutamine amidotransferase domain include asparagine synthetase (Fig. 26-54, Reaction 4; see below), a member of the Ntn family, and imidazole glycerol phosphate synthase (Fig. 26-65, Reaction 5), which belongs to the triad family. All of these enzymes have an ammonia-channeling tunnel that connects the amidotransferase site with the ammonia-utilizing site.

b. Alanine, Asparagine, Aspartate, Glutamate, and Glutamine Are Synthesized from Pyruvate, Oxaloacetate, and α -Ketoglutarate

Pyruvate, oxaloacetate, and α -ketoglutarate are the keto acids that correspond to alanine, aspartate, and glutamate, respectively. Indeed, as we have seen (Section 26-1), the synthesis of each of these amino acids is a one-step

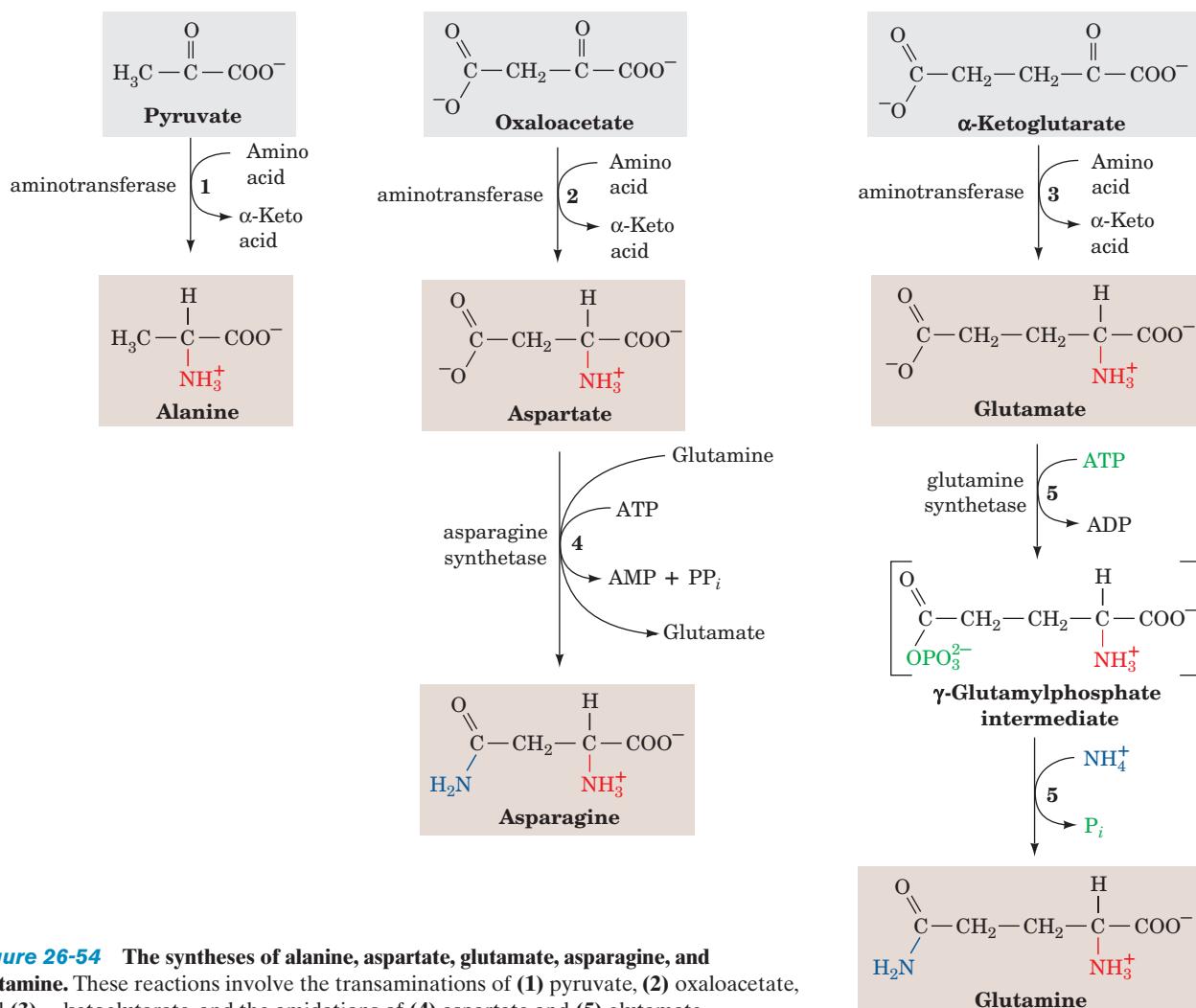


Figure 26-54 The syntheses of alanine, aspartate, glutamate, asparagine, and glutamine. These reactions involve the transaminations of (1) pyruvate, (2) oxaloacetate, and (3) α -ketoglutarate, and the amidations of (4) aspartate and (5) glutamate.

transamination reaction (Fig. 26-54, Reactions 1-3). Asparagine and glutamine are, respectively, synthesized from aspartate and glutamate by amidation (Fig. 26-54, Reactions 4 and 5). **Glutamine synthetase** catalyzes the formation of glutamine in a reaction in which ATP is hydrolyzed to ADP

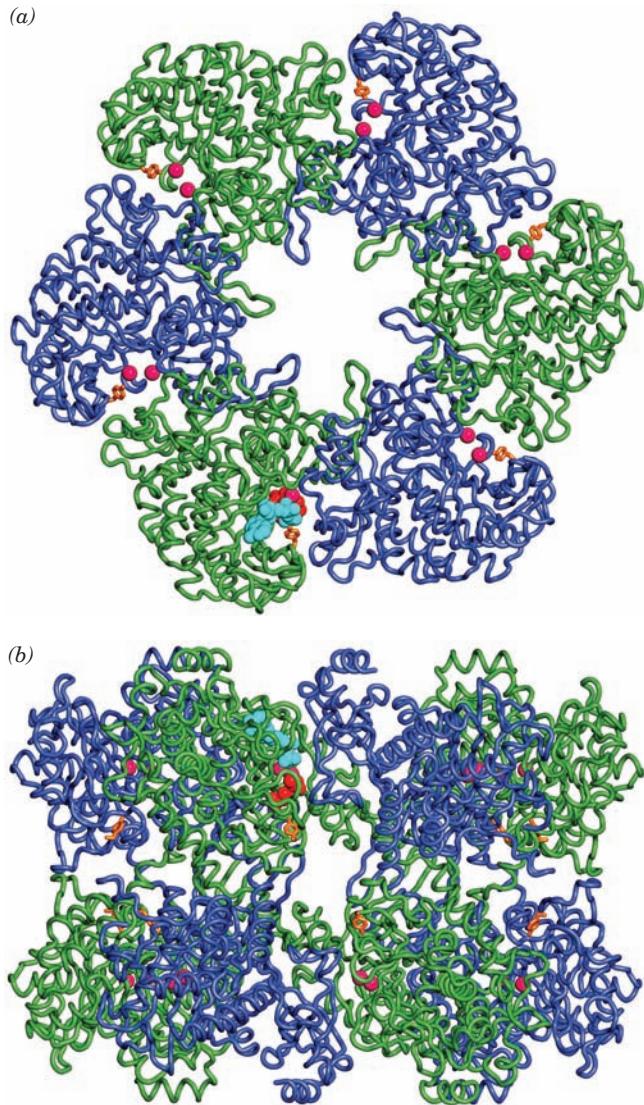
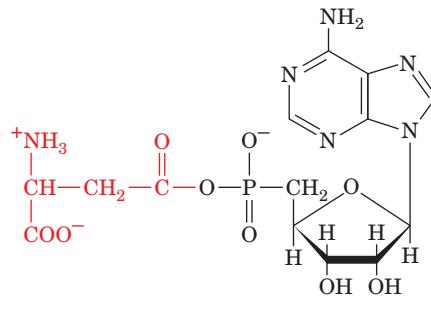


Figure 26-55 X-ray structure of *S. typhimurium* glutamine synthetase. The enzyme consists of 12 identical subunits, here represented by their C_α backbones, arranged with D₆ symmetry (the symmetry of a hexagonal prism). (a) View down the 6-fold axis of symmetry showing only the six subunits of the upper ring in alternating blue and green. The subunits of the lower ring are roughly directly below those of the upper ring. The protein, including its side chains (not shown), has a diameter of 143 Å. The six active sites shown are marked by the pairs of Mn²⁺ ions (*magenta spheres*; divalent metal ions, physiologically Mg²⁺, are required for enzymatic activity). Each adenylylation site, Tyr 397 (*yellow*), lies between two subunits at a higher radius than the corresponding active site. Also drawn in one active site are ADP (*cyan*) and phosphinotricin (*orange*), a competitive inhibitor of glutamate. (b) Side view along one of the enzyme's 2-fold axes showing only the eight nearest subunits. The molecule extends 103 Å along the 6-fold axis, which is vertical in this view. [Based on an X-ray structure by David Eisenberg, UCLA. PDBid 1FPY.]

and P_i via the intermediacy of **γ -glutamylphosphate** and NH_4^+ is the amino group donor (Fig. 26-54, Reaction 5). Curiously, aspartate amidation by **asparagine synthetase** to form asparagine follows a different route; it utilizes glutamine as its amino group donor and hydrolyzes ATP to $AMP + PP_i$ (Fig. 26-54, Reaction 4). This enzyme is composed of a glutamine amidotransferase domain of the Ntn family (see above) and a second domain in which β -aspartyl-AMP



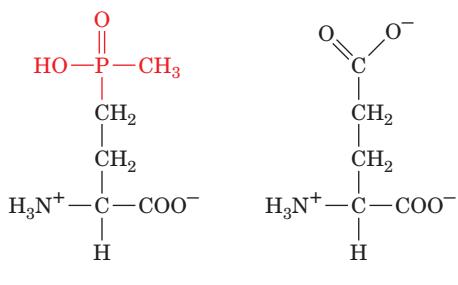
is synthesized from Asp and ATP and then reacts with ammonia to form Asn. As with other glutamine amidotransferase-containing enzymes, the two domains are connected by a tunnel that channels ammonia between their two active sites.

c. Glutamine Synthetase Is a Central Control Point in Nitrogen Metabolism

Glutamine, as we have seen, is the amino group donor in the formation of many biosynthetic products, as well as being a storage form of ammonia. Glutamine synthetase's consequent pivotal position in nitrogen metabolism makes its control a vital aspect of this process. In fact, mammalian glutamine synthetases are activated by α -ketoglutarate, the product of glutamate's oxidative deamination. This control presumably prevents the accumulation of the ammonia produced by that reaction.

Bacterial glutamine synthetase, as Earl Stadtman showed, has a much more elaborate control system. This enzyme, which consists of 12 identical 469-residue subunits arranged with D_6 symmetry (Fig. 26-55), is regulated by several effectors as well as by covalent modification. Although a complete description of this complex enzyme is not given here, several aspects of its catalytic and control systems bear note.

The X-ray structure of *Salmonella typhimurium* glutamine synthetase in complex with the glutamate structural analog **phosphinothricin**,



Phosphinothricin

Glutamate

determined by David Eisenberg, reveals that its catalytic sites occur at the interface between the C-terminal domain

of one subunit and the N-terminal domain of an adjacent subunit. These catalytic sites have a shape described as a “bifunnel” that opens at both the exposed top (ATP binding) and bottom (Glu and NH_4^+ binding) of the molecule (between the two hexameric rings) and is narrow in the plane of its essential metal ions (two per subunit). Nucleotide binding induces conformational changes that increase the enzyme’s affinity for glutamate and ammonium ion, leading to an ordered sequential mechanism.

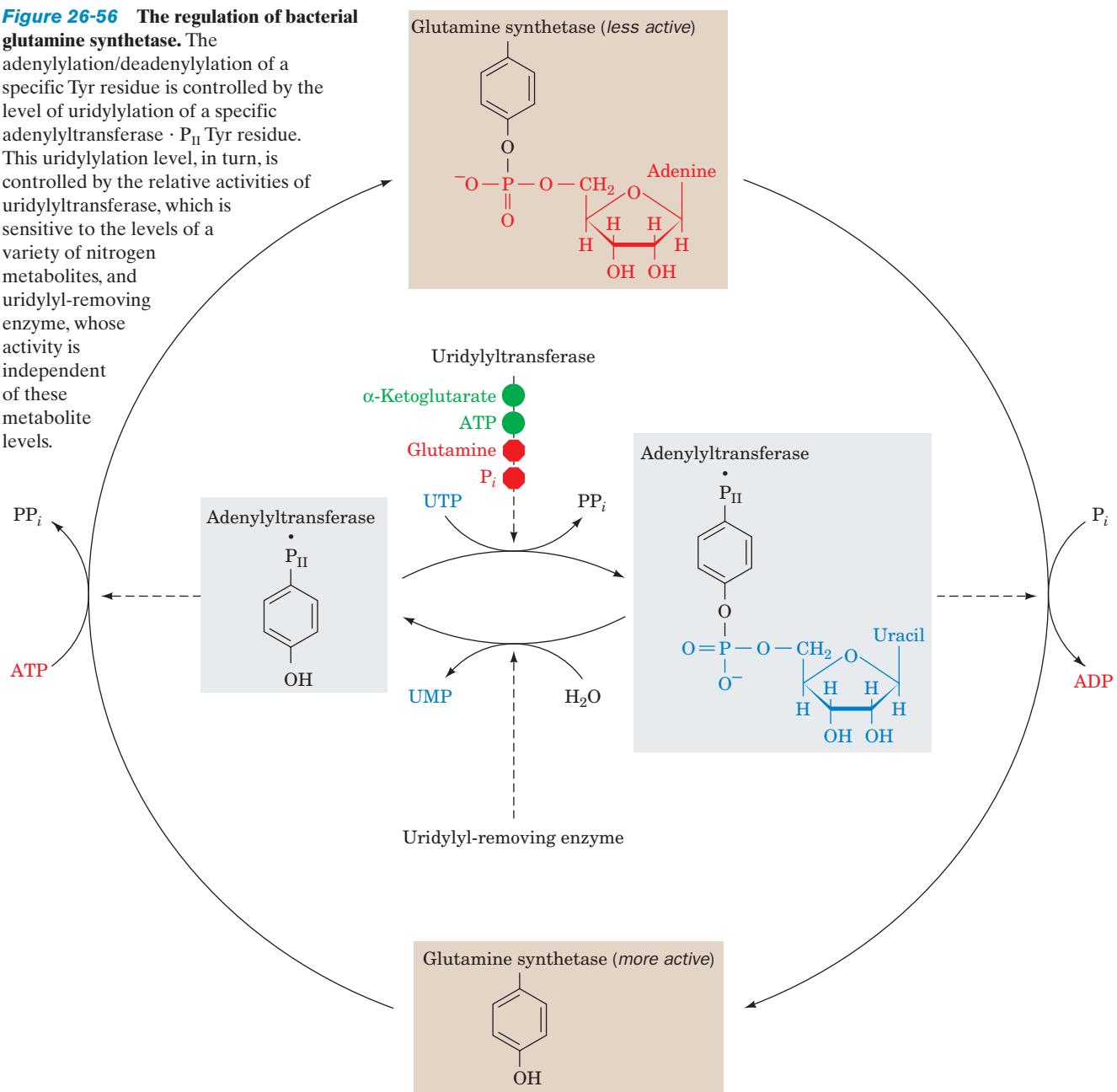
Nine feedback inhibitors cumulatively control the activity of bacterial glutamine synthetase: histidine, tryptophan, carbamoyl phosphate (as synthesized by carbamoyl phosphate synthetase II), glucosamine-6-phosphate, AMP, and CTP are all end products of pathways leading from glutamine, whereas alanine, serine, and glycine reflect the cell’s nitrogen

level. Several of these inhibitors act in a competitive manner, binding either to the glutamate binding site (serine, glycine, and alanine) or to the ATP binding site (AMP and CTP).

E. coli glutamine synthetase is covalently modified by adenyllylation of a specific Tyr residue (Fig. 26-56). The enzyme’s susceptibility to cumulative feedback inhibition increases, and its activity therefore decreases, with its degree of adenyllylation. The level of adenyllylation is controlled by a complex metabolic cascade that is conceptually similar to the one controlling glycogen phosphorylase (although the type of covalent modification differs in that glycogen phosphorylase is phosphorylated at a specific Ser residue; Section 18-3Ca). Both adenyllylation and deadenyllylation of glutamine synthetase are catalyzed by **adenyllyltransferase** in complex with a tetrameric regulatory protein, P_{II} . This

Figure 26-56 The regulation of bacterial glutamine synthetase.

The adenyllylation/deadenyllylation of a specific Tyr residue is controlled by the level of uridylylation of a specific adenyllyltransferase $\cdot \text{P}_{\text{II}}$ Tyr residue. This uridylylation level, in turn, is controlled by the relative activities of uridylyltransferase, which is sensitive to the levels of a variety of nitrogen metabolites, and uridylyl-removing enzyme, whose activity is independent of these metabolite levels.



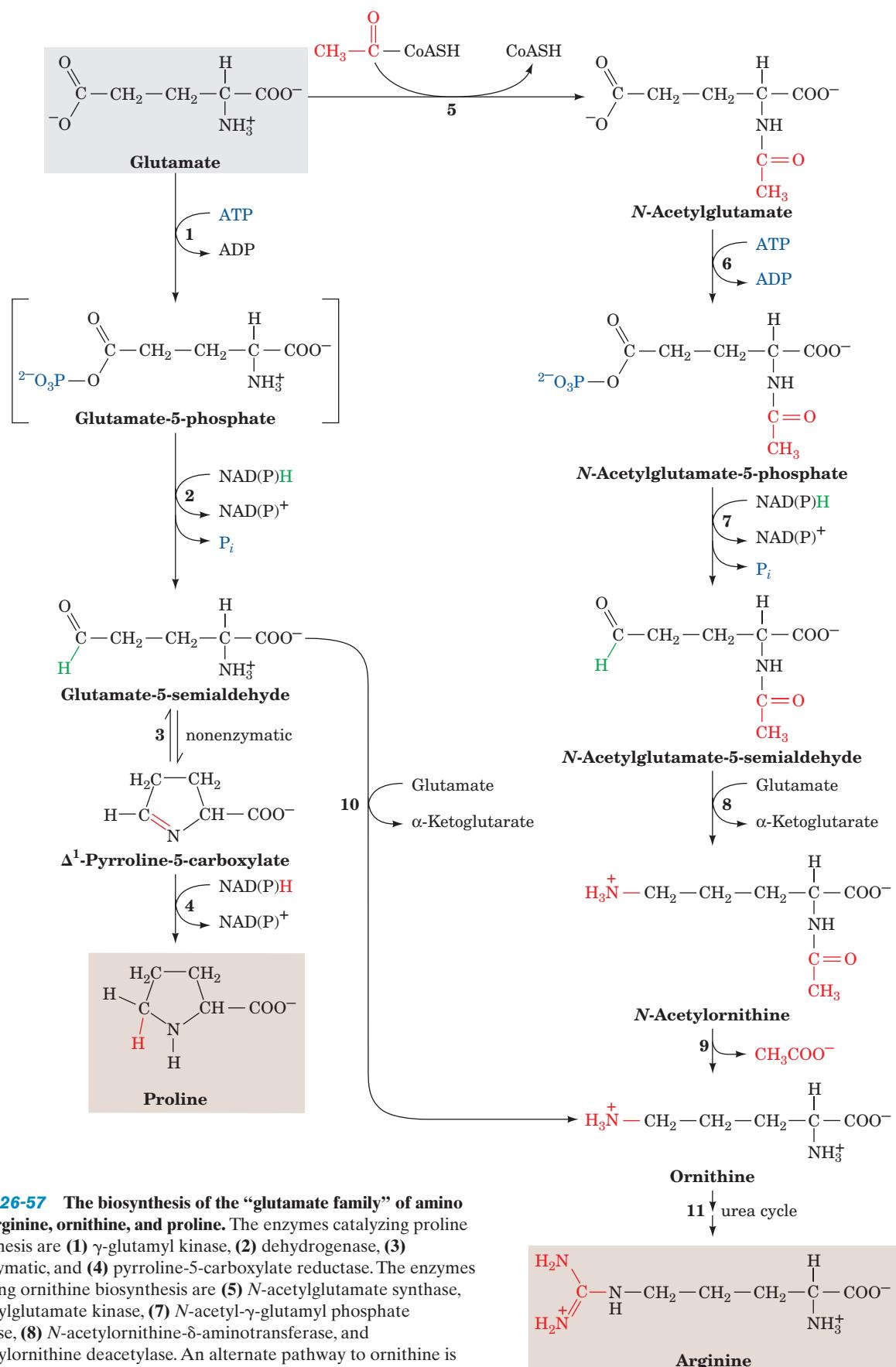


Figure 26-57 The biosynthesis of the “glutamate family” of amino acids: arginine, ornithine, and proline. The enzymes catalyzing proline biosynthesis are (1) γ -glutamyl kinase, (2) dehydrogenase, (3) nonenzymatic, and (4) pyrroline-5-carboxylate reductase. The enzymes catalyzing ornithine biosynthesis are (5) N -acetylglutamate synthase, (6) acetylglutamate kinase, (7) N -acetyl- γ -glutamyl phosphate reductase, (8) N -acetylornithine- δ -aminotransferase, and (9) acetylornithine deacetylase. An alternate pathway to ornithine is through Reaction 10, catalyzed by ornithine- δ -aminotransferase. Ornithine is converted to arginine (11) via the urea cycle (Fig. 26-7, Reactions 2-4).

complex deadenylylates glutamine synthetase when P_{II} is uridylylated (also at a Tyr residue) and adenylates glutamine synthetase when P_{II} lacks UMP residues. The level of P_{II} uridylylation, in turn, depends on the relative activities of two enzymatic activities located on the same protein: a **uridylyltransferase** that uridylylates P_{II} and a **uridylyl-removing enzyme** that hydrolytically excises the attached UMP groups of P_{II} (Fig. 26-56). The uridylyltransferase is activated by α -ketoglutarate and ATP and inhibited by glutamine and P_i , whereas uridylyl-removing enzyme is insensitive to these metabolites. This complex metabolic cascade therefore renders the activity of *E. coli* glutamine synthetase extremely responsive to the cell's nitrogen requirements.

d. Glutamate Is the Precursor of Proline, Ornithine, and Arginine

Conversion of glutamate to proline (Fig. 26-57, Reactions 1–4) involves the reduction of the γ -carboxyl group to an aldehyde followed by the formation of an internal Schiff base whose further reduction yields proline. Reduction of the glutamate γ -carboxyl group to an aldehyde is an endergonic process that is facilitated by the carboxyl group's prior phosphorylation by **γ -glutamyl kinase**. The unstable product, **glutamate-5-phosphate**, has not been isolated from reaction mixtures but is presumed to be the substrate for the reduction that follows. The resulting **glutamate-5-semialdehyde** cyclizes spontaneously to form the internal Schiff base **Δ^1 -pyrroline-5-carboxylate**. The final reduction to proline is catalyzed by **pyrroline-5-carboxylate reductase**. Whether the enzyme requires NADH or NADPH is unclear.

The *E. coli* pathway from glutamate to ornithine and hence to arginine likewise involves the ATP-driven reduction of the glutamate γ -carboxyl group to an aldehyde (Fig. 26-57, Reactions 6 and 7). Spontaneous cyclization of this intermediate, ***N*-acetylglutamate-5-semialdehyde**, is prevented by prior acetylation of its amino group by ***N*-acetylglutamate synthase** to form ***N*-acetylglutamate** (Fig. 26-57, Reaction 5). *N*-Acetylglutamate-5-semialdehyde, in turn, is converted to the corresponding amine by transamination (Fig. 26-57, Reaction 8). Hydrolysis of the acetyl protecting group finally yields ornithine, which, as we have seen (Section 26-2), is converted to arginine via the urea cycle. In humans, however, the pathway to ornithine is more direct. The *N*-acetylation of glutamate that protects it from cyclization does not occur. Rather, glutamate-5-semialdehyde, which is in equilibrium with Δ^1 -pyrroline-5-carboxylate, is directly transaminated to yield ornithine in a reaction catalyzed by **ornithine- δ -aminotransferase** (Fig. 26-57, Reaction 10).

e. Serine, Cysteine, and Glycine Are Derived from 3-Phosphoglycerate

Serine is formed from the glycolytic intermediate 3-phosphoglycerate in a three-reaction pathway (Fig. 26-58):

1. Conversion of 3-phosphoglycerate's 2-OH group to a ketone yielding **3-phosphohydroxypyruvate**, serine's phosphorylated keto acid analog.
2. Transamination of 3-phosphohydroxypyruvate to phosphoserine.
3. Hydrolysis of phosphoserine to yield serine.

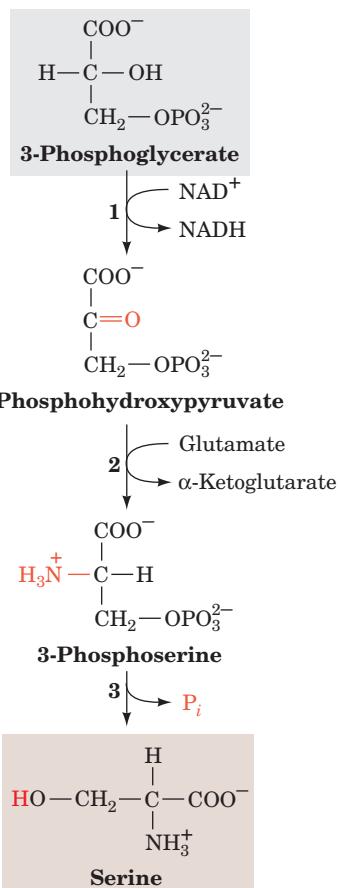


Figure 26-58 The conversion of 3-phosphoglycerate to serine. The pathway enzymes are (1) 3-phosphoglycerate dehydrogenase, (2) a PLP-dependent aminotransferase, and (3) phosphoserine phosphatase.

Serine participates in glycine synthesis in two ways (Section 26-3B):

1. Direct conversion of serine to glycine by serine hydroxymethyl transferase in a reaction that also yields N^5,N^{10} -methylene-THF (Fig. 26-12, Reaction 4 in reverse).
2. Condensation of the N^5,N^{10} -methylene-THF with CO_2 and NH_4^+ by the glycine cleavage system (Fig. 26-12, Reaction 3 in reverse).

We have already discussed the synthesis, in animals, of cysteine from serine and homocysteine, a breakdown product of methionine (Section 26-3Ea). Homocysteine combines with serine to yield cystathionine, which subsequently forms cysteine and α -ketobutyrate (Fig. 26-18, Reactions 5 and 6). Since cysteine's sulphydryl group is derived from the essential amino acid methionine, cysteine is really an essential amino acid. In plants and microorganisms, however, cysteine is synthesized from serine in a two-step reaction involving the activation of the serine —OH by converting it to ***O*-acetylserine** followed by the displacement of acetate by sulfide (Fig. 26-59a). The sulfide required is produced from sulfate in an 8-electron reduction that occurs in *E. coli* as shown in Fig. 26-59b. Sulfate is first activated by the

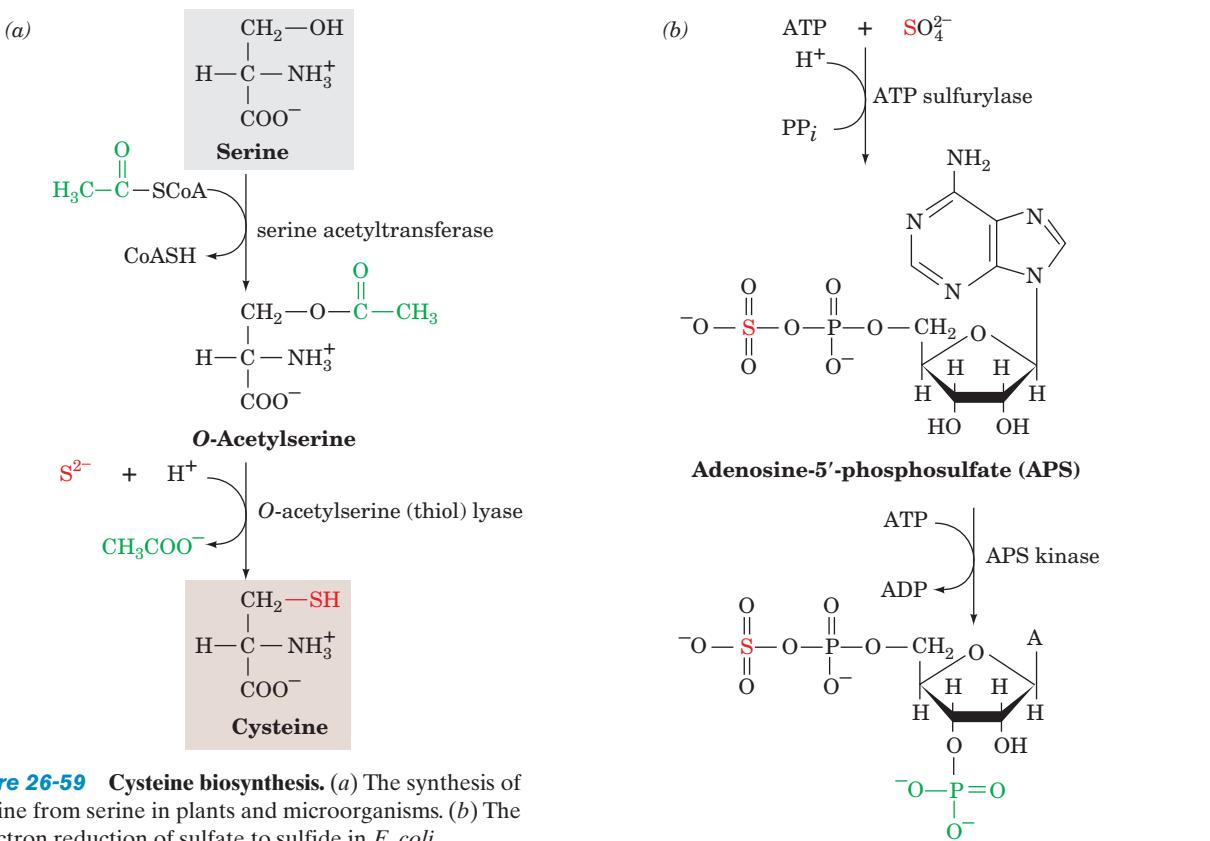


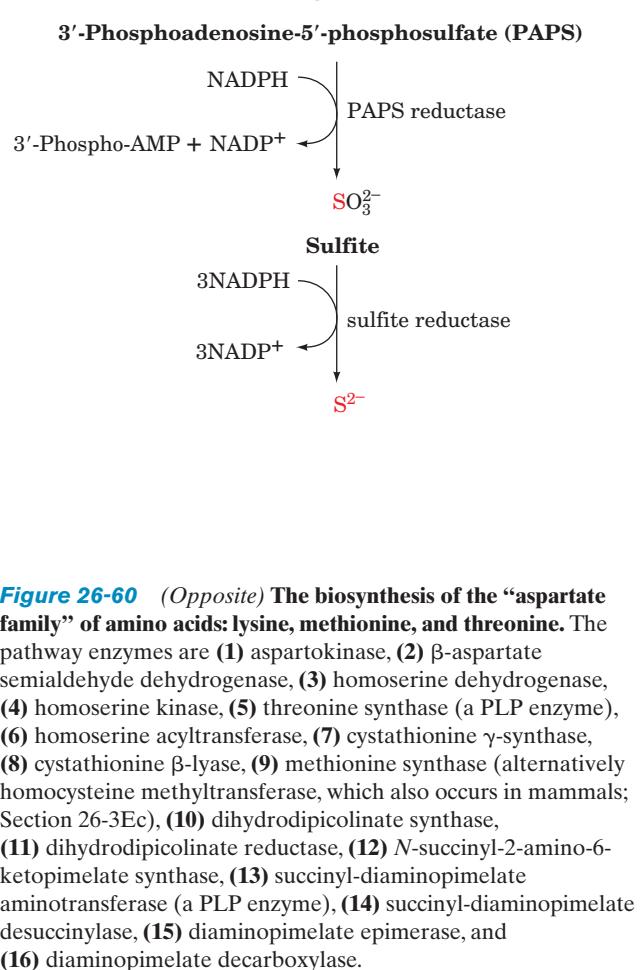
Figure 26-59 Cysteine biosynthesis. (a) The synthesis of cysteine from serine in plants and microorganisms. (b) The 8-electron reduction of sulfate to sulfide in *E. coli*.

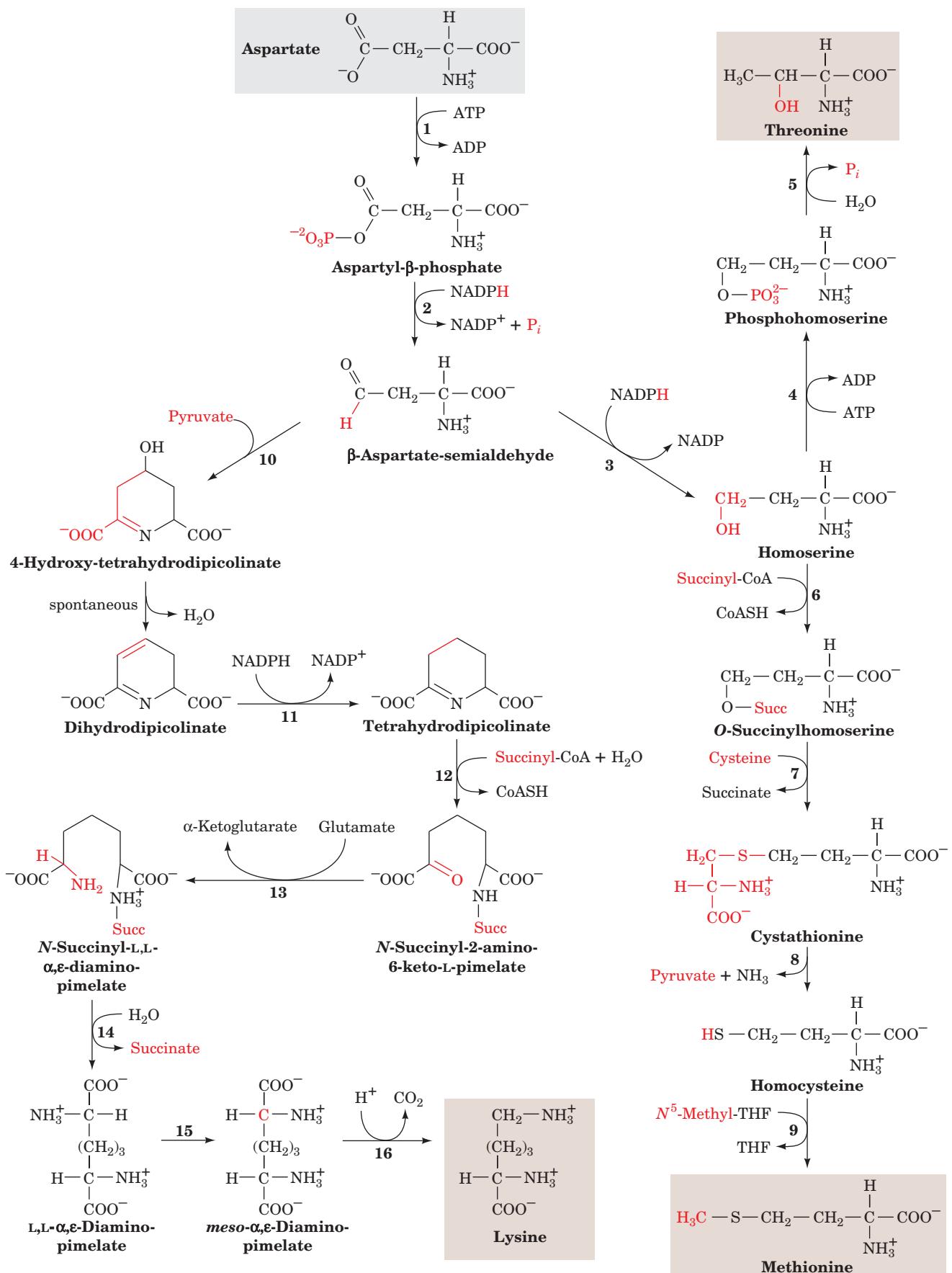
enzymes **ATP sulfurylase** (which is used in the pyrosequencing of DNA; Section 7-2Ca) and **adenosine-5'-phosphosulfate (APS) kinase**. The activated sulfate is then reduced to sulfite by **3'-phosphoadenosine-5'-phosphosulfate (PAPS) reductase** and to sulfide by **sulfite reductase**.

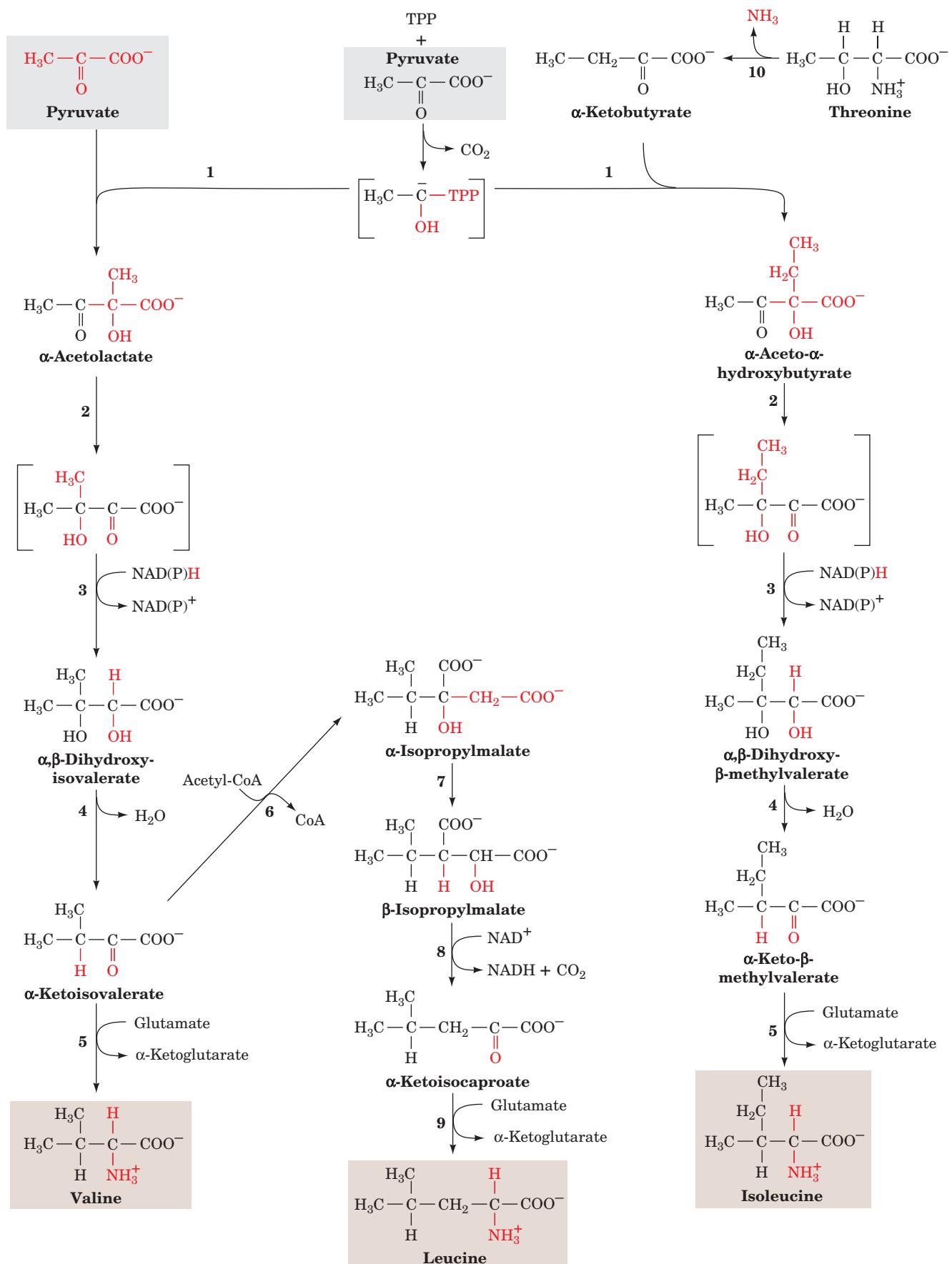
B. Biosynthesis of the Essential Amino Acids

Essential amino acids, like nonessential amino acids, are synthesized from familiar metabolic precursors. Their synthetic pathways are present only in microorganisms and plants, however, and usually involve more steps than those of the nonessential amino acids. For example, lysine, methionine, and threonine are all synthesized from aspartate in pathways whose common first reaction is catalyzed by **aspartokinase**, an enzyme that is present only in plants and microorganisms. Similarly, valine and leucine are formed from pyruvate; isoleucine is formed from pyruvate and α -ketobutyrate; and tryptophan, phenylalanine, and tyrosine are formed from phosphoenolpyruvate and erythrose-4-phosphate. The enzymes that synthesize essential amino acids were apparently lost early in animal evolution, possibly because of the ready availability of these amino acids in the diet.

Time and space prevent a detailed discussion of the many interesting reactions that occur in these pathways. The biosynthetic pathways of the aspartate family of amino acids, the pyruvate family, the aromatic family, and histidine are presented in Figs. 26-60 through 26-63 and 26-65 together with lists of the enzymes involved. Several agriculturally useful







herbicides are specific inhibitors of some of these enzymes. Such herbicides have little toxicity toward animals and hence pose minimal risk to human health and the environment.

a. The Aspartate Family: Lysine, Methionine, and Threonine

In bacteria, aspartate is the common precursor of lysine, methionine, and threonine (Fig. 26-60). The biosyntheses of these essential amino acids all begin with the aspartokinase-catalyzed phosphorylation of aspartate to yield **aspartyl-β-phosphate**. We have seen that the control of metabolic pathways commonly occurs at the first committed step of the pathway. One might therefore expect lysine, methionine, and threonine biosynthesis to be controlled as a group. Each of these pathways is, in fact, independently controlled. *E. coli* does so via three isozymes of aspartokinase that respond differently to the three amino acids in terms both of feedback inhibition of enzyme activity and repression of enzyme synthesis. Table 26-3 summarizes this differential control. In addition, the pathway direction is controlled by feedback inhibition at the branch points by the individual amino acids. Thus methionine inhibits the *O*-acylation of homoserine (Fig. 26-60, Reaction 6), and lysine inhibits dihydrolipicollate synthase (Fig. 26-60, Reaction 10).

b. The Pyruvate Family: Leucine, Isoleucine, and Valine

Valine and isoleucine are both synthesized via the same five-step pathway (Fig. 26-61), the only difference being in the first step of the series. In this TPP-dependent reaction, which resembles those catalyzed by pyruvate decarboxylase (Section 17-3Ba) and transketolase (Section 23-4Ca), pyruvate forms an adduct with TPP, which is decarboxylated to hydroxyethyl-TPP. This resonance-stabilized carbanion adds either to the keto group of a second pyruvate to form **acetolactate** on the way to valine, or to the keto group of threonine-derived **α-ketobutyrate** to form **α-aceto-α-hydroxybutyrate** on the way to isoleucine. The leucine biosynthetic pathway branches off from the valine pathway at **α-ketoisovalerate** (Fig. 26-61, Reaction 6). Reactions 6 to 8 in Fig. 26-61 are reminiscent of the first three reactions of the citric acid cycle (Sections 21-3A–C). Here, acetyl-CoA condenses with **α-ketoisovalerate** to form **α-isopropylmalate**, which then undergoes a dehydration/

Table 26-3 Differential Control of Aspartokinase Isoenzymes in *E. coli*

Enzyme	Feedback Inhibitor	Corepressor(s) ^a
Aspartokinase I	Threonine	Threonine and isoleucine
Aspartokinase II	None	Methionine
Aspartokinase III	Lysine	Lysine

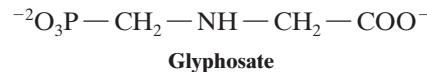
^aCompounds whose presence results in the transcriptional repression of enzyme synthesis (Section 31-3G).

hydration reaction, followed by oxidative decarboxylation and transamination, to yield leucine.

c. The Aromatic Amino Acids: Phenylalanine, Tyrosine, and Tryptophan

The precursors to the aromatic amino acids are the glycolytic intermediate phosphoenolpyruvate (PEP) and erythrose-4-phosphate (an intermediate in the pentose phosphate pathway; Section 23-4Cb). Their condensation forms **2-keto-3-deoxy-D-arabinoheptulosonate-7-phosphate**, a C₇ compound that cyclizes and is ultimately converted to **chorismate** (Fig. 26-62), the branch point for tryptophan biosynthesis. Chorismate is converted either to **anthranilate** and then on to tryptophan, or to **prephenate** and on to either tyrosine or phenylalanine (Fig. 26-63). Although mammals synthesize tyrosine by the hydroxylation of phenylalanine (Section 26-3Ha), many microorganisms synthesize it directly from prephenate.

Since the synthesis of aromatic amino acids only occurs in plants and microorganisms, this pathway is a natural target for herbicides that will not be toxic to animals. For example, **glyphosate**,



the active ingredient in one of the most widely used weed killers, Roundup, is a competitive inhibitor with respect to PEP in the **5-enolpyruvylshikimate-3-phosphate (EPSP) synthase** reaction (Reaction 6 of Fig. 26-62).

d. A Protein Tunnel Channels the Intermediate Product of Tryptophan Synthase between Two Active Sites

The final two reactions in tryptophan biosynthesis, Reactions 5 and 6 in Fig. 26-63, are both catalyzed by **tryptophan synthase**:

1. The α subunit (268 residues) of this $\alpha_2\beta_2$ bifunctional enzyme cleaves **indole-3-glycerol phosphate**, yielding **indole** and glyceraldehyde-3-phosphate (Reaction 5).
2. The β subunit (396 residues) joins indole with L-serine in a PLP-dependent reaction to form L-tryptophan (Reaction 6).

Either subunit alone is enzymatically active, but when they are joined in the $\alpha_2\beta_2$ tetramer, the rates of both reactions

Figure 26-61 (Opposite) The biosynthesis of the “pyruvate family” of amino acids: isoleucine, leucine, and valine. The pathway enzymes are (1) acetolactate synthase (a TPP enzyme), (2) acetolactate mutase, (3) reductase, (4) dihydroxy acid dehydratase, (5) valine aminotransferase (a PLP enzyme), (6) α -isopropylmalate synthase, (7) α -isopropylmalate dehydratase, (8) isopropylmalate dehydrogenase, (9) leucine aminotransferase (a PLP enzyme), and (10) threonine deaminase (serine dehydratase, a PLP enzyme).

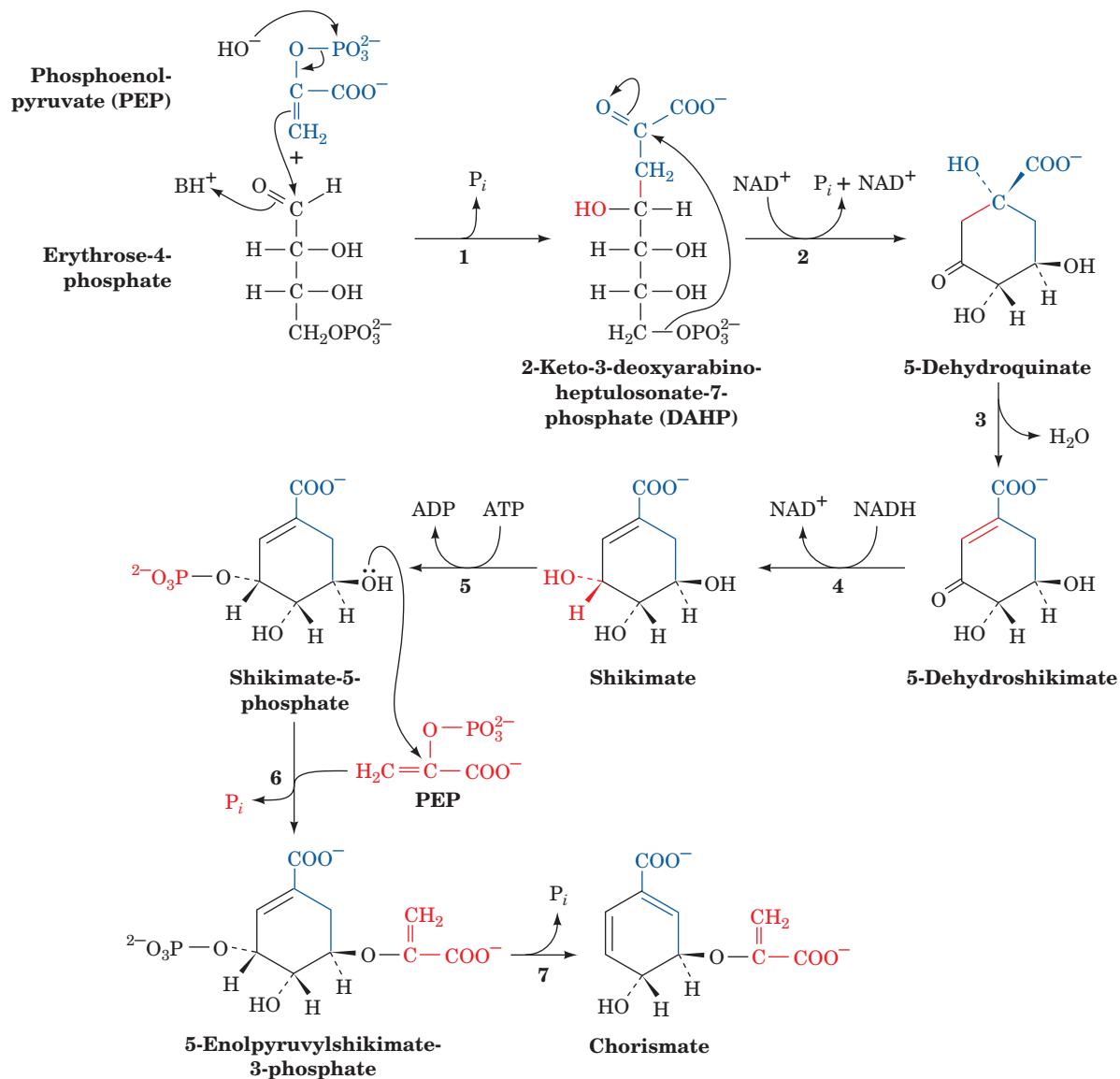


Figure 26-62 The biosynthesis of chorismate, the aromatic amino acid precursor. The pathway enzymes are (1) 2-keto-3-deoxy-D-arabinoheptulonate-7-phosphate synthase, (2) dehydroquinate synthase (an NAD^+ -requiring reaction that yields an unchanged NAD^+ product and is thereby indicative of

an oxidized intermediate as similarly occurs in the UDP-galactose-4-epimerase reaction; Section 17-5B), (3) 5-dehydroquinate dehydratase, (4) shikimate dehydrogenase, (5) shikimate kinase, (6) 5-enolpyruvylshikimate-3-phosphate synthase, and (7) chorismate synthase.

and their substrate affinities are increased by 1 to 2 orders of magnitude. Indole, the intermediate product, does not appear free in solution; the enzyme apparently sequesters it.

The X-ray structure of tryptophan synthase from *Salmonella typhimurium*, determined by Craig Hyde, Edith Miles, and David Davies, explains the latter observation. The protein forms a 150-Å-long, 2-fold symmetric α - β - β - α complex (Fig. 26-64) in which the active sites of neighboring α and β subunits are separated by ~ 25 Å. These active sites are joined by a solvent-filled tunnel that is

wide enough to permit the passage of the intermediate substrate, indole. This structure, the first in which the presence of a tunnel between active sites was observed, suggests the following series of events. The indole-3-glycerol phosphate substrate binds to the α subunit through an opening into its active site, its “front door,” and the glyceraldehyde-3-phosphate product leaves via the same route. Similarly, the β subunit active site has a “front door” opening to the solvent through which serine enters and tryptophan leaves. Both active sites also have “back doors” that are connected

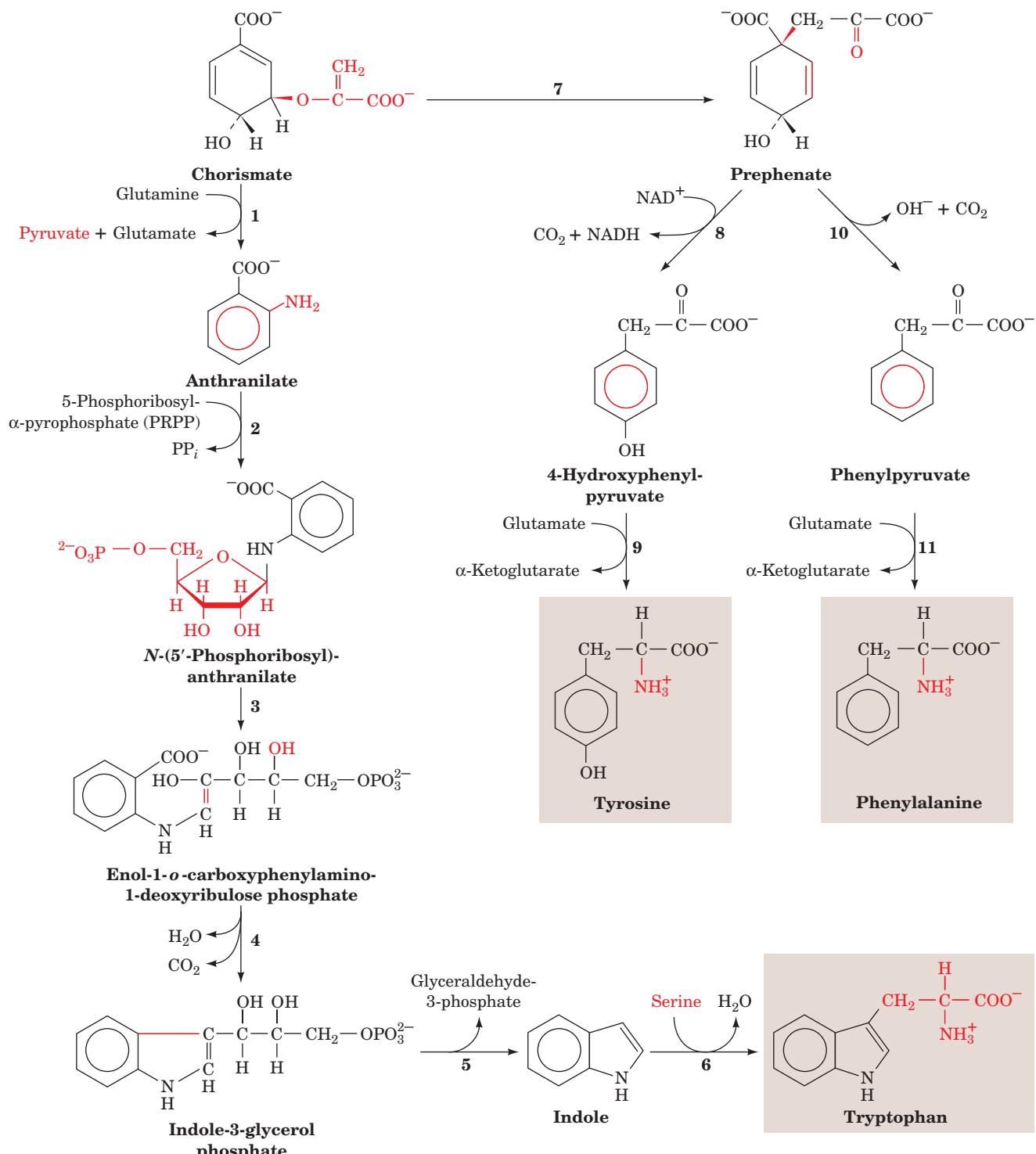


Figure 26-63 The biosynthesis of phenylalanine, tryptophan, and tyrosine from chorismate. The pathway enzymes are (1) anthranilate synthase, (2) anthranilate phosphoribosyltransferase, (3) *N*-(5'-phosphoribosyl)-anthranilate isomerase, (4) indole-3-glycerol phosphate synthase,

(5) tryptophan synthase, α subunit, (6) tryptophan synthase, β subunit, (7) chorismate mutase, (8) prephenate dehydrogenase, (9) aminotransferase, (10) prephenate dehydratase, and (11) aminotransferase.

by the tunnel. The indole intermediate presumably diffuses between the two active sites via the tunnel and hence does not escape to the solvent.

Allosteric interactions between the subunits to control the activity of the α subunit also serve to ensure that indole is only released when the β subunit is ready to accept it.

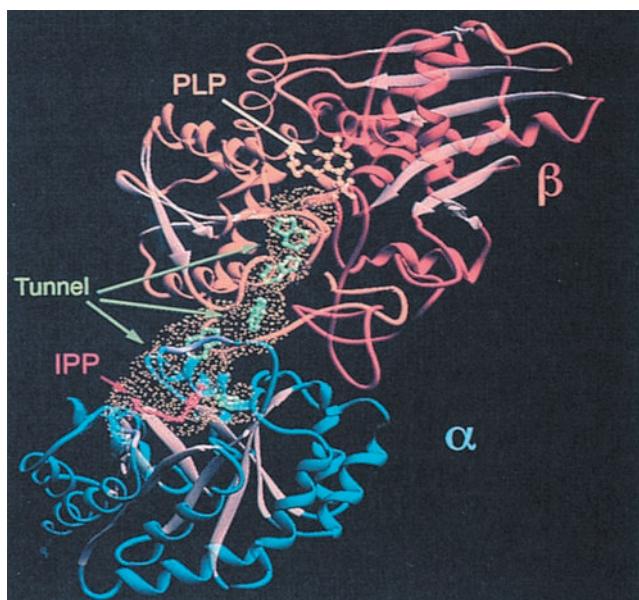
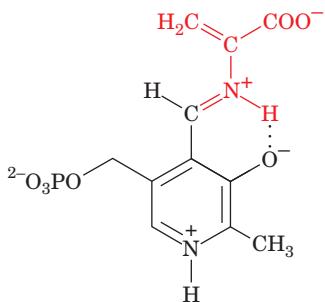


Figure 26-64 A ribbon diagram of the bifunctional enzyme tryptophan synthase from *S. typhimurium*. Only one $\alpha\beta$ protomer of this 2-fold symmetric $\alpha\beta\alpha\beta$ heterotetramer is shown. The α subunit is blue, the β subunit's N-terminal domain is orange, its C-terminal domain is red-orange, and all β sheets are tan. The active site of the α subunit is located by its bound competitive inhibitor, **indolepropanol phosphate (IPP; red ball-and-stick model)**, whereas that of the β subunit is marked by its PLP coenzyme (**yellow ball-and-stick model**). The solvent-accessible surface of the $\sim 25\text{-}\text{\AA}$ -long “tunnel” connecting the active sites of the α and β subunits is outlined by a yellow dot surface. Several indole molecules (**green ball-and-stick models**) have been modeled into the tunnel in head to tail fashion, thereby demonstrating that the tunnel has sufficient width to permit the indole product of the α subunit to pass through the tunnel to the β subunit's active site. [Courtesy of Craig Hyde, Edith Miles, and David Davies, National Institutes of Health.]

See Interactive Exercise 25

Michael Dunn has shown that the elimination of water from the serine-PLP Schiff base on the β subunit to form an aminoacrylate-PLP Schiff base intermediate



Aminoacrylate-PLP Schiff base

triggers a conformational change that activates the α subunit to produce indole. The diffusion of the indole to the β

Figure 26-65 (Opposite) The biosynthesis of histidine. The pathway enzymes are (1) ATP phosphoribosyltransferase, (2) pyrophosphohydrolase, (3) phosphoribosyl-AMP cyclohydrolase, (4) phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide isomerase, (5) imidazole glycerol phosphate synthase (a glutamine amidotransferase), (6) imidazole glycerol phosphate dehydratase, (7) L-histidinol phosphate aminotransferase, (8) histidinol phosphate phosphatase, and (9) histidinol dehydrogenase.

subunit to react with this intermediate then results in the formation of tryptophan.

Channeling may be particularly important for indole since this nonpolar molecule otherwise can escape the bacterial cell by diffusing through its plasma and outer membranes. We have seen similar phenomena in reactions involving glutamine amidotransferases (Sections 26-2Aa and 26-5Aa), as well as in the series of reactions catalyzed by fatty acid synthase, in which the growing product is kept in the vicinity of the multifunctional enzyme's active site by covalent attachment to the enzyme's flexible phosphopantetheine arm (Section 25-4Ca). Channeling is also implicated in the multistep biosyntheses of purines and pyrimidines (Sections 28-1A and 28-2A).

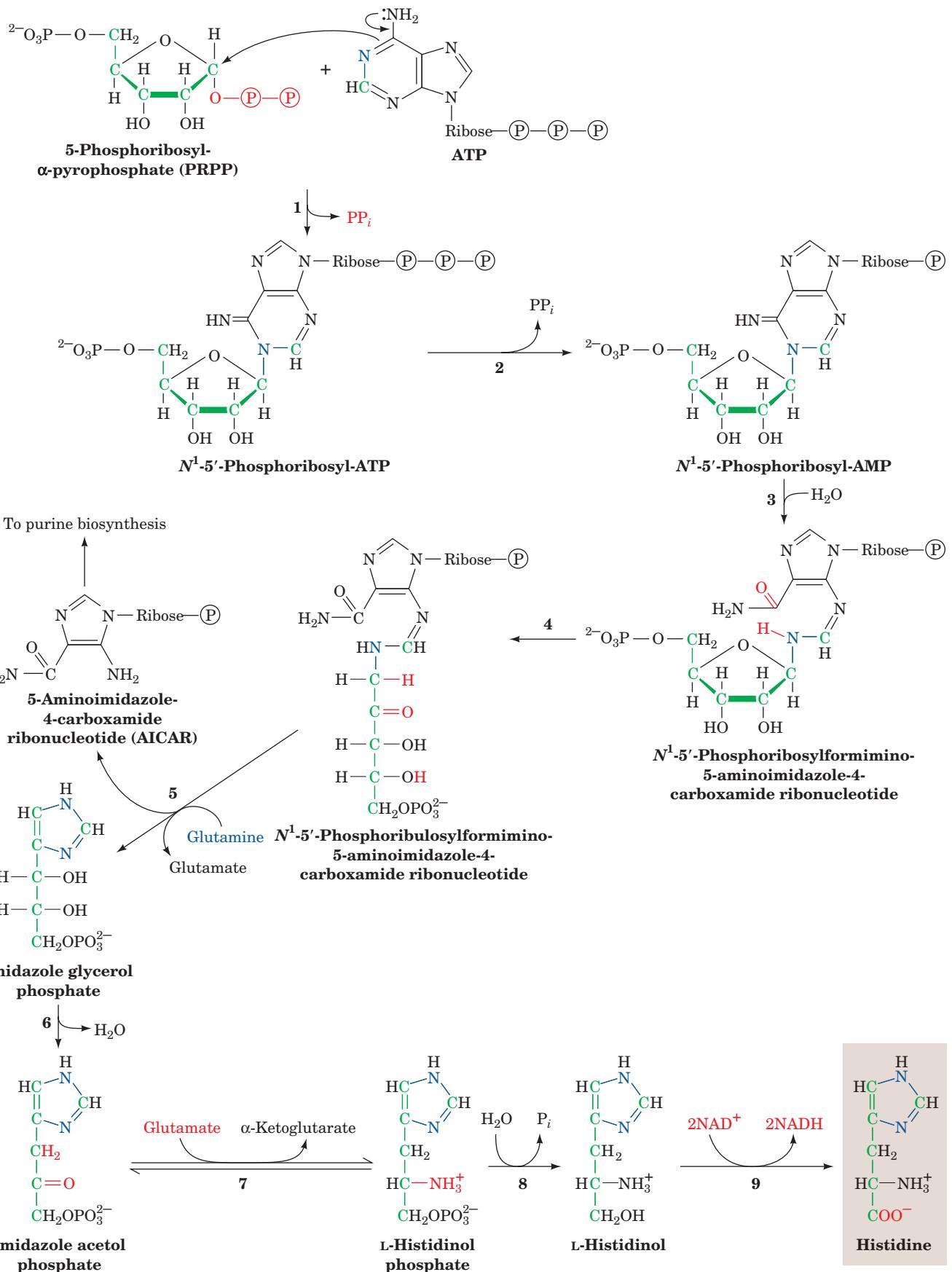
e. Histidine Biosynthesis

Five of histidine's six C atoms are derived from **5-phosphoribosyl-α-pyrophosphate (PRPP; Fig. 26-65)**, an intermediate also involved in the biosynthesis of tryptophan (Fig. 26-63, Reaction 2), purine nucleotides (Section 28-1A), and pyrimidine nucleotides (Section 28-2A). The histidine's sixth carbon originates from ATP. The ATP atoms that are not incorporated into histidine are eliminated as **5-aminoimidazole-4-carboxamide ribonucleotide (AICAR; Fig. 26-65, Reaction 5)**, which is also an intermediate in purine biosynthesis (Section 28-1A).

The unusual biosynthesis of histidine from a purine has been cited as evidence supporting the hypothesis that life was originally RNA based (Section 1-5Ca). His residues, as we have seen, are often components of enzyme active sites, where they act as nucleophiles and/or general acid-base catalysts. The discovery that RNA can have catalytic properties (Section 31-4Ae) therefore suggests that the imidazole moiety of purines plays a similar role in these RNA enzymes (**ribozymes**). This further suggests that the histidine biosynthesis pathway is a “fossil” of the transition to more efficient protein-based life-forms.

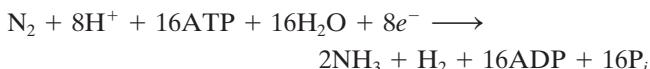
6 NITROGEN FIXATION

The most prominent chemical elements in living systems are O, H, C, N, and P. The elements O, H, and P occur widely in metabolically available forms (e.g., H_2O , O_2 , and P_i). However, the major available forms of C and N, CO_2 and N_2 , are extremely stable (unreactive); for example, the $\text{N}\equiv\text{N}$ triple bond has a bond energy of $945\text{ kJ}\cdot\text{mol}^{-1}$



(versus $351 \text{ kJ} \cdot \text{mol}^{-1}$ for a C—O single bond). CO_2 , with only minor exceptions, is metabolized (fixed) only by photosynthetic organisms (Chapter 24). *N₂* fixation is even less common; this element is converted to metabolically useful forms by only a few strains of bacteria, named **diazatrophs**.

Diazatrophs of the genus *Rhizobium* live in symbiotic relationship with root nodule cells of legumes (plants belonging to the pea family, including beans, clover, and alfalfa; Fig. 26-66) where they convert N_2 to NH_3 :



The NH_3 thus formed can be incorporated either into glutamate by glutamate dehydrogenase (Section 26-1B) or into glutamine by glutamine synthetase (Section 26-5Ab). This nitrogen-fixing system produces more metabolically useful nitrogen than the legume needs; the excess is excreted into the soil, enriching it. It is therefore common agricultural practice to plant a field with alfalfa every few years to build up the supply of usable nitrogen in the soil for later use in growing other crops.

a. Nitrogenase Contains Novel Redox Centers

Nitrogenase, which catalyzes the reduction of N_2 to NH_3 , is a complex of two proteins:

1. The **Fe-protein**, a homodimer that contains one [4Fe–4S] cluster and two ATP binding sites.



Figure 26-66 Photograph showing the root nodules of the legume bird's-foot trefoil. [Vu/Cabisco/Visuals Unlimited.]

2. The **MoFe-protein**, an $\alpha_2\beta_2$ heterotetramer that contains Fe and Mo.

The X-ray structure of *Azotobacter vinelandii* nitrogenase in complex with the inhibitor ADP · AlF_4^- (which mimics the transition state in ATP hydrolysis), determined by Douglas Rees, reveals that each MoFe-protein associates with two molecules of Fe-protein (Fig. 26-67).

Each Fe-protein dimer's single [4Fe–4S] cluster is located in a solvent-exposed cleft between the two subunits

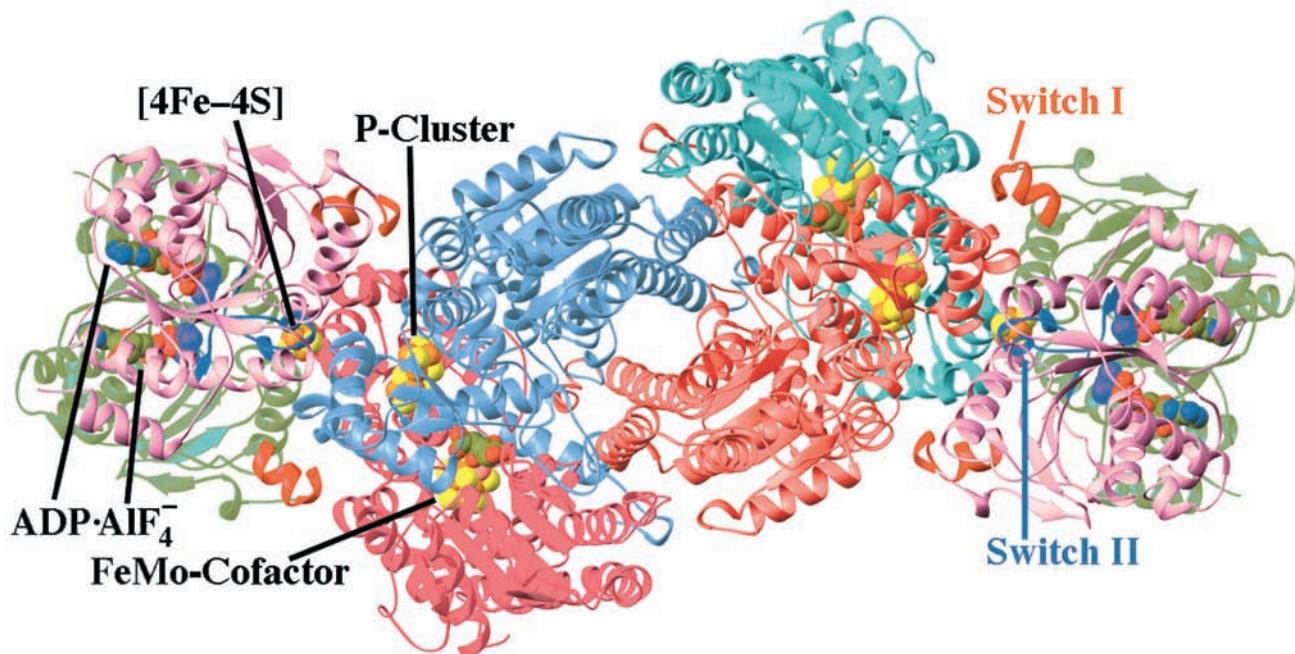


Figure 26-67 X-ray structure of the *A. vinelandii* nitrogenase in complex with ADP · AlF_4^- . The enzyme, which is viewed along its molecular 2-fold axis, is an $(\alpha\beta\gamma_2)_2$ heterooctamer in which the $\beta-\alpha-\alpha-\beta$ assembly, the MoFe-protein, is flanked by two γ_2 Fe-proteins whose 289-residue subunits are related by local 2-fold symmetry. The homologous α subunits (cyan and magenta; 491 residues) and β subunits (light red and light blue; 522 residues) are related by pseudo-2-fold symmetry. The two γ subunits forming each Fe-protein (pink and green with their

Switch I and Switch II segments red and blue) bind to the MoFe-protein with the 2-fold axis relating them coincident with the pseudo-2-fold axis relating the MoFe-protein's α and β subunits. The ADP · AlF_4^- , [4Fe–4S] cluster, FeMo-cofactor, and P-cluster are drawn in space-filling form with C green, N blue, O red, S yellow, Fe orange, Mo pink, and the AlF_4^- ion purple. [Based on an X-ray structure by Douglas Rees, California Institute of Technology. PDBid 1N2C.]

See Interactive Exercise 26

and is symmetrically linked to Cys 97 and Cys 132 from both subunits such that an Fe-protein resembles an “iron butterfly” with the [4Fe–4S] cluster at its head. Its two identical nucleotide binding sites are located at the interface between its two subunits. The [4Fe–4S] cluster cycles between its +1 and +2 oxidation states.

The MoFe-protein’s α and β subunits assume similar folds and extensively associate to form a pseudo-2-fold symmetric $\alpha\beta$ dimer, two of which more loosely associate to form the 2-fold symmetric $\alpha_2\beta_2$ tetramer (Fig. 26-67). Each $\alpha\beta$ dimer has two bound redox centers:

1. The P-cluster (Fig. 26-68a,b), which consists of two [4Fe–3S] clusters linked through an additional sulfide ion forming the eighth corner of each of the clusters to make cubane-like structures, and bridged by two Cys thiol ligands, each coordinating one Fe from each cluster. Four additional Cys thiols coordinate the remaining four Fe atoms. The positions of two of the Fe atoms in one of the [4Fe–3S] clusters change on oxidation, rupturing the bonds from these Fe atoms to the linking sulfide ion. These bonds are replaced in the oxidized state by a Ser oxygen ligand to one of the Fe atoms, and by a bond to the amide N of a Cys from the other Fe atom.

2. The FeMo-cofactor (Fig. 26-68c), which consists of a [4Fe–3S] cluster and a [1Mo–3Fe–3S] cluster bridged by

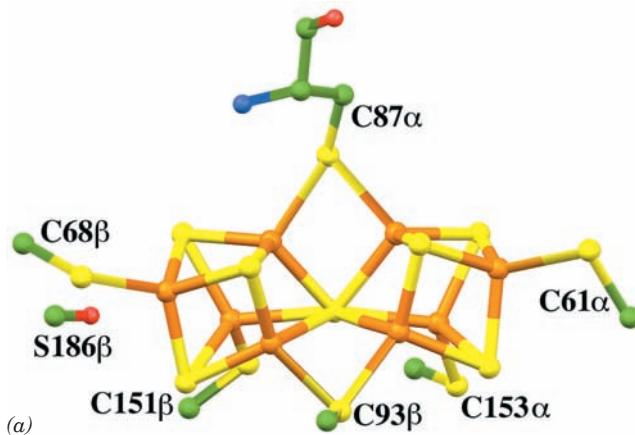
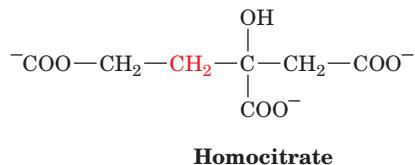


Figure 26-68 The prosthetic groups of the nitrogenase MoFe-protein. The molecules are drawn in ball-and-stick form with C green, N blue, O red, S yellow, Fe orange, and Mo pink.

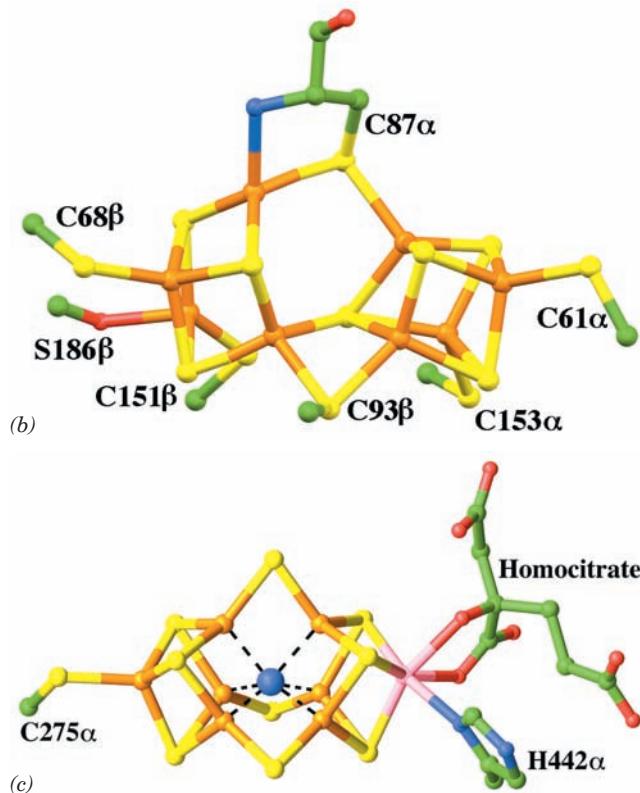
(a) The reduced *Klebsiella pneumoniae* P-cluster. It consists of two [4Fe–3S] complexes linked by an additional sulfide ion forming the eighth corner of each cubane-like structure, and bridged by two Cys thiol ligands, each coordinating one Fe from each cluster. Four additional Cys thiols coordinate the remaining 4 Fe atoms. (b) The 2-electron-oxidized *K. pneumoniae* P-cluster. In comparison with the reduced complex in Part a, two of the Fe–S bonds from the centrally located sulfide ion that bridges the two [4Fe–3S] clusters have been replaced by ligands from the Cys 87 α amide N and the Ser 186 β side chain O yielding a [4Fe–3S] cluster (left) and a [4Fe–4S] cluster (right) that remain linked by a direct Fe–S bond and two bridging Cys thiols. (c) The *A. vinelandii* FeMo-cofactor. It consists of a [4Fe–3S] cluster and a [1Mo–3Fe–3S] cluster that are bridged by three sulfide ions. The FeMo-cofactor is linked to the protein by only two ligands at its opposite ends, one from His 442 α to the Mo atom and the other from Cys 275 α to an Fe atom. The Mo atom is

three sulfide ions. The FeMo-cofactor’s Mo atom is approximately octahedrally coordinated by three cofactor sulfide ions, a His imidazole nitrogen, and two oxygens from a bound **homocitrate** ion:



(an essential component of the FeMo-cofactor). The FeMo-cofactor contains a central cavity that a high resolution (1.16 Å) X-ray structure of *A. vinelandii* MoFe-protein, also determined by Rees, reveals contains what most probably is a nitrogen atom (although a C or an O atom cannot be ruled out). This putative N atom is liganded to the FeMo-cofactor’s central six Fe atoms such that it completes the approximate tetrahedral coordination environment of each of these Fe atoms.

The FeMo-cofactor is located ~10 Å below the α subunit surface, and hence the N_2 is thought to gain access to its binding site through conformational fluctuations of the protein (recall that myoglobin and hemoglobin likewise have no clear path for O_2 to approach its heme binding



additionally doubly liganded by homocitrate. What is most likely an N atom (blue sphere) is liganded to the FeMo-cluster’s six central Fe atoms (dashed black lines). [Parts a and b based on X-ray structures by David Lawson, John Innes Centre, Norwich, U.K. Part c based on an X-ray structure by Douglas Rees, California Institute of Technology. PDBIDs (a) 1QGU, (b) 1QH1, and (c) 1M1N.]

sites in these proteins; Section 10-2). The P-cluster, which is also ~ 10 Å below the protein surface, is at the interface between the α and β subunits on the pseudo-2-fold axis that roughly relates these two subunits. The 2-fold axis of the Fe-protein and the pseudo-2-fold axis of the MoFe-proteins coincide in their complex.

The Fe-protein hydrolyzes two ATP molecules for each electron it transfers from its $[4\text{Fe}-4\text{S}]^{+1}$ cluster to the P-cluster. Since the nucleotide binding sites and the $[4\text{Fe}-4\text{S}]$ cluster on the Fe-protein are separated by ~ 20 Å, a distance too large for direct coupling between electron transfer and ATP hydrolysis, it appears that these processes are allosterically coupled through conformational changes at the subunit interface. Indeed, portions of the Fe-protein resemble those of G-proteins, in which nucleotide hydrolysis is coupled to conformational changes controlling the protein's actions (Sections 19-2Cb and 19-3Cf). Specifically, two regions of the Fe-protein, designated Switch I and Switch II (Fig. 26-67), are homologous with those of Ras (Section 19-3Cf). The binding of $\text{ADP} \cdot \text{AlF}_4^-$ to Fe-protein induces conformational changes in Switch I that affect the interactions between the Fe-protein and the MoFe-protein, and in Switch II that affect the environment of the $[4\text{Fe}-4\text{S}]$ cluster.

In nitrogenase, the $[4\text{Fe}-4\text{S}]$ cluster of the Fe-protein approaches within ~ 14 Å of the P-cluster in the MoFe-protein, whereas the P-cluster and the FeMo-cofactor are ~ 13 Å apart. Hence, the sequence of the electron-transfer steps in the nitrogenase reaction appears to be



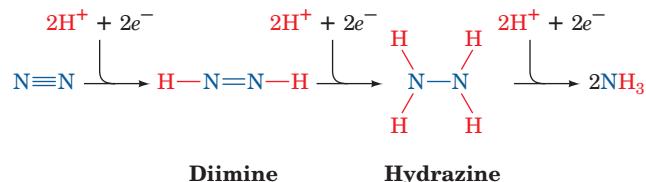
It therefore seems that the role of ATP hydrolysis is to stabilize a conformation in the Fe-protein that it cannot achieve on its own and which facilitates electron transfer from the $[4\text{Fe}-4\text{S}]$ cluster on the Fe-protein to the P-cluster on the MoFe-protein.

b. N_2 Reduction Is Energetically Costly

Nitrogen fixation requires two participants in addition to N_2 and nitrogenase: (1) a source of electrons and (2) ATP. Electrons are generated either oxidatively or photosynthetically, depending on the organism. These electrons are transferred to ferredoxin (Section 22-2C1a), a $[4\text{Fe}-4\text{S}]$ -containing electron carrier that transfers an electron to the Fe-protein of nitrogenase, beginning the nitrogen fixation process (Fig. 26-69). Two molecules of ATP bind to the

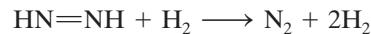
reduced Fe-protein and are hydrolyzed as each electron is passed from the Fe-protein to the MoFe-protein. The ATP hydrolysis-induced conformational change in the Fe-protein alters its redox potential from -0.29 to -0.40 V, making the electron capable of N_2 reduction ($\mathcal{E}^\circ' = -0.34$ V for the half-cell $\text{N}_2 + 6\text{H}^+ + 6\text{e}^- \rightleftharpoons 2\text{NH}_3$).

The actual reduction of N_2 occurs on the MoFe-protein in three discrete steps, each involving an electron pair:



An electron transfer must occur six times per N_2 molecule fixed so that a total of 12 ATPs are required to fix one N_2 molecule. Although the N_2 binding site is almost certainly the FeMo-cofactor, exactly how the N_2 is bound and reduced are largely a matter of speculation. Theoretical studies suggest that the FeMo-cofactor's prismatically arranged Fe atoms provide favorable interaction sites for N_2 and its reduction products. Indeed, it seems highly likely that the putative N atom that is liganded to the FeMo-cofactor (Fig. 26-68c) participates in N_2 reduction.

Nitrogenase also reduces H_2O to H_2 , which in turn reacts with **diimine** to reform N_2 :



The resulting futile cycle is favored when the ATP level is low and/or the reduction of the Fe-protein is sluggish. Even when ATP is plentiful, however, the cycle cannot be suppressed beyond about one H_2 molecule produced per N_2 reduced and hence appears to be a requirement of the nitrogenase reaction. The total cost of N_2 reduction is therefore 8 electrons transferred and 16 ATPs hydrolyzed (physiologically, 20–30 ATPs). Hence nitrogen fixation is an energetically expensive process; indeed, the nitrogen-fixing bacteria in the root nodules of pea plants consume nearly 20% of the ATP that the plant produces.

c. Leghemoglobin Protects Nitrogenase from Oxygen Inactivation

Nitrogenase is rapidly inactivated by O_2 , so the enzyme must be protected from this reactive substance. Cyanobacteria

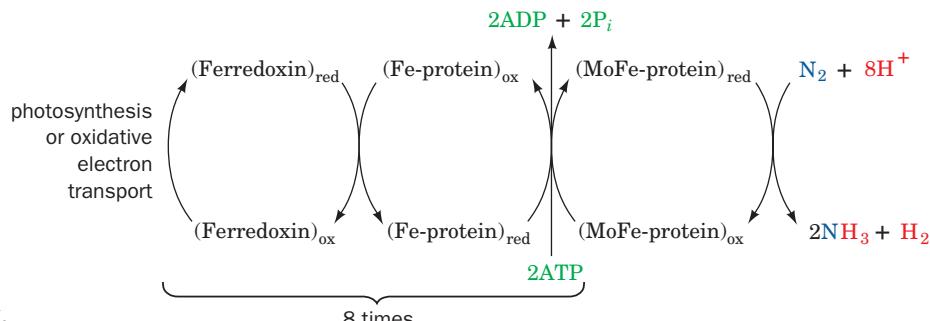


Figure 26-69 The flow of electrons in the nitrogenase-catalyzed reduction of N_2 .

(photosynthetic oxygen-evolving bacteria; Section 1-1Ab) provide protection by carrying out nitrogen fixation in specialized nonphotosynthetic cells called **heterocysts**, which have Photosystem I but lack Photosystem II (Section 24-2Ca). In the root nodules of legumes (Fig. 26-66), however, protection is afforded by the symbiotic synthesis of **leghemoglobin**. The globin portion of this ~145-residue monomeric oxygen-binding protein is synthesized by the plant (an evolutionary curiosity since globins are otherwise known to occur only in animals), whereas the heme is synthesized by the *Rhizobium*. Leghemoglobin has a very high O₂ affinity, thus keeping the pO₂ low enough to protect the nitrogenase while providing passive O₂ transport for the aerobic bacterium.

d. Installing the Nitrogen Fixation Machinery in Nonleguminous Plants Would Revolutionize Agriculture

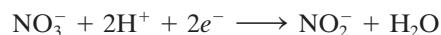
Although atmospheric N₂ is the ultimate nitrogen source for all living things, most plants do not support the symbiotic growth of nitrogen-fixing bacteria. They must therefore depend on a source of “prefixed” nitrogen such as nitrate or ammonia. These nutrients come from lightning discharges (the source of ~10% of naturally fixed N₂), decaying organic matter in the soil, or from fertilizer applied to it. The Haber process, which was invented by Fritz Haber in 1910, is a chemical process for N₂ fixation that is still widely used in fertilizer manufacture. This direct reduction of N₂ by H₂ to form NH₃ requires temperatures of 300 to 500°C, pressures of >300 atm, and an Fe catalyst. Intriguingly, the spacing of the Fe atoms on the surface of this catalyst resembles that of the FeMo-cofactor’s central Fe atoms (Fig. 26-68c).

One of the major long-term goals of genetic engineering is to induce agriculturally useful nonleguminous plants to

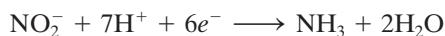
fix their own nitrogen, a complex undertaking in which the plant must be made to provide a hospitable environment for nitrogen fixation as well as to acquire the enzymatic machinery to do so. This would free farmers, particularly those in developing countries, from the need for either purchasing fertilizers, periodically letting their fields lie fallow (giving legumes the opportunity to grow), or following the slash-and-burn techniques that are rapidly destroying the world’s tropical forests and contributing significantly to the greenhouse effect (atmospheric CO₂ pollution causing long-term global warming).

e. The Nitrogen Cycle Describes the Interconversion of Nitrogen in the Biosphere

The ammonia produced by the nitrogenase reaction and incorporated into amino acids is eventually recycled in the biosphere as described by the **nitrogen cycle** (Fig. 26-70). Nitrate is produced by certain bacteria that oxidize NH₃ to NO₂⁻ and then NO₃⁻, a process called **nitrification**. Still other organisms convert nitrate back to N₂, which is known as **denitrification**. In addition, nitrate is reduced to NH₃ by plants, fungi, and many bacteria, a process called **ammonification** in which **nitrate reductase** catalyzes the two-electron reduction of nitrate to nitrite (NO₂⁻):



and then **nitrite reductase** converts nitrite to ammonia,



The direct anaerobic oxidation of NH₃ back to N₂ without the intermediacy of nitrate, the reverse of nitrogen fixation, has recently been discovered in certain bacteria.

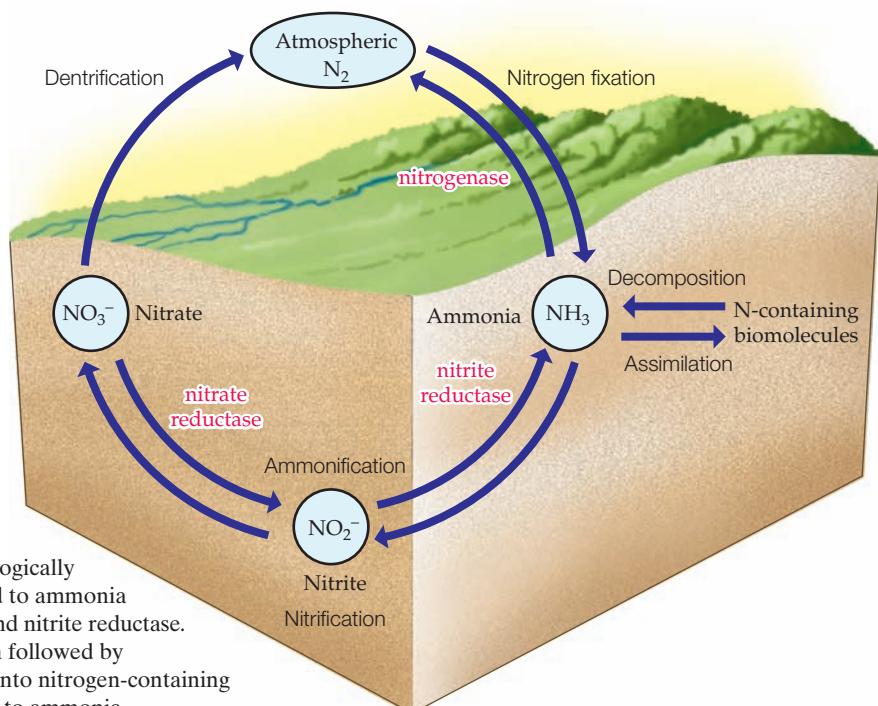


Figure 26-70 The nitrogen cycle. Nitrogen fixation by nitrogenase converts N₂ to the biologically useful ammonia. Nitrate can also be converted to ammonia by the sequential actions of nitrate reductase and nitrite reductase. Ammonia is transformed to N₂ by nitrification followed by denitrification. Ammonia may be assimilated into nitrogen-containing biomolecules, which may be decomposed back to ammonia.

CHAPTER SUMMARY

1 Amino Acid Deamination Amino acids are the precursors for numerous nitrogen-containing compounds such as heme, physiologically active amines, and glutathione. Excess amino acids are converted to common metabolic intermediates for use as fuels. The first step in amino acid breakdown is removal of the α -amino group by transamination. Transaminases require pyridoxal phosphate (PLP) and convert amino acids to their corresponding α -keto acids. The amino group is transferred to α -ketoglutarate to form glutamate, oxaloacetate to form aspartate, or pyruvate to form alanine. Glutamate is subsequently oxidatively deaminated by glutamate dehydrogenase (GDH) to form ammonia and regenerate α -ketoglutarate. Hyperinsulinism/hyperammonemia (HI/HA), a genetic disease, is caused by a mutation of the GDH gene that decreases GTP's ability to inhibit GDH.

2 The Urea Cycle In the urea cycle, amino groups from NH_3 and aspartate combine with HCO_3^- to form urea. This pathway takes place in the liver, partially in the mitochondrion and partially in the cytosol. It begins with the ATP-dependent condensation of NH_3 and HCO_3^- by carbamoyl phosphate synthetase, an enzyme with a 96-Å-long tunnel connecting its three active sites through which its highly reactive intermediate products are channeled. The resulting carbamoyl phosphate then combines with ornithine to yield citrulline, which combines with aspartate to form arginosuccinate, which in turn is cleaved to fumarate and arginine. The arginine is then hydrolyzed to urea, which is excreted, and ornithine, which reenters the urea cycle. *N*-Acetylglutamate regulates the urea cycle by activating carbamoyl phosphate synthetase allosterically.

3 Metabolic Breakdown of Individual Amino Acids The α -keto acid products of transamination reactions are degraded to citric acid cycle intermediates or their precursors. The amino acids leucine and lysine are ketogenic in that they are converted only to the ketone body precursors acetyl-CoA and acetoacetate. The remaining amino acids are, at least in part, glucogenic in that they are converted to the glucose precursors pyruvate, oxaloacetate, α -ketoglutarate, succinyl-CoA, or fumarate. Alanine, cysteine, glycine, serine, and threonine are converted to pyruvate. Serine hydroxymethyltransferase catalyzes the PLP-dependent C_α — C_β bond cleavage of serine to form glycine. This reaction requires the transfer of a methylene group from N^5,N^{10} -methylene-tetrahydrofolate, which the tetrahydrofolate (THF) obtains from the glycine cleavage system, a multienzyme system. Asparagine and aspartate are converted to oxaloacetate. α -Ketoglutarate is a product of arginine, glutamate, glutamine, histidine, and proline degradation. Methionine, isoleucine, and valine are degraded to succinyl-CoA. Methionine breakdown involves the synthesis of *S*-adenosylmethionine (SAM), a sulfonium ion that acts as a methyl donor in many biosynthetic reactions. Hyperhomocysteinemia, a risk factor for cardiovascular disease, cognitive im-

pairment, and neural tube defects, is caused by a deficiency in its folate-dependent degradation. Maple syrup urine disease (MSUD) is caused by an inherited defect in branched-chain amino acid degradation. Branched-chain amino acid degradation pathways contain reactions common to all acyl-CoA oxidations. Tryptophan is degraded to alanine and acetoacetate. Phenylalanine and tyrosine are degraded to fumarate and acetoacetate. Most individuals with the hereditary disease phenylketonuria lack phenylalanine hydroxylase (PAH), which converts phenylalanine to tyrosine.

4 Amino Acids as Biosynthetic Precursors Heme is synthesized from glycine and succinyl-CoA. These precursors condense to form δ -aminolevulinic acid (ALA), which cyclizes to form the pyrrole porphobilinogen (PBG). Four molecules of PBG condense to form uroporphyrinogen III, which then goes on to form heme, with the final reaction, the insertion of Fe(II) into protoporphyrin IX, catalyzed by ferrochelatase. Defects in heme biosynthesis, which are known as porphyrias, have a variety of bizarre symptoms. Heme is degraded to form linear tetrapyrroles, which are subsequently excreted as bile pigments. The hormones and neurotransmitters L-DOPA, epinephrine, norepinephrine, serotonin, γ -aminobutyric acid (GABA), and histamine are all synthesized from amino acid precursors. Glutathione, a tripeptide that is synthesized from glutamate, cysteine, and glycine, is involved in a variety of protective, transport, and metabolic processes. Tetrahydrofolate is a coenzyme that participates in the transfer of C_1 units.

5 Amino Acid Biosynthesis Amino acids are required for many vital functions of an organism. Those amino acids that mammals can synthesize from common α -keto acid carbon skeletons and preformed α -amino nitrogen such as that of glutamate are known as nonessential amino acids; those that mammals must obtain from their diets are called essential amino acids. The biosynthesis of nonessential amino acids involves relatively simple pathways, whereas those forming the essential amino acids are generally more complex.

6 Nitrogen Fixation Although the ultimate source of nitrogen for amino acid biosynthesis is atmospheric N_2 , this nearly inert gas must first be reduced to a metabolically useful form, NH_3 , by nitrogen fixation. This process occurs only in certain types of bacteria, one genus of which occurs in symbiotic relationship with legumes. N_2 is fixed in these organisms by an oxygen-sensitive enzyme, nitrogenase, that consists of two proteins: the Fe-protein dimer, which contains one [4Fe-4S] cluster and two ATP binding sites, and the MoFe-protein $\alpha_2\beta_2$ tetramer, which contains one P-cluster (consisting of two [4Fe-3S] clusters linked by a sulfide ion) and one FeMo-cofactor (a [4Fe-3S] cluster and a [1Mo-3Fe-3S] cluster bridged by three sulfide ions and coordinated with homocitrate) in each $\alpha\beta$ dimer. These cofactors each function as two-electron carriers for the ATP-driven reduction of N_2 to NH_3 .

REFERENCES

General

- Bender, D.A., *Amino Acid Metabolism*, Wiley (1985).
- Frey, P.A. and Hegeman, A.D., *Enzymatic Reaction Mechanisms*, Oxford University Press (2007). [Describes the mechanisms of many of the enzymes discussed in this chapter.]
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Part 8 contains numerous chapters on defects in amino acid metabolism.]

Amino Acid Deamination and the Urea Cycle

- Eliot, A.C. and Kirsch, J.F., Pyridoxal phosphate enzymes, *Annu. Rev. Biochem.* **73**, 383–415 (2004).
- Holden, H.M., Thoden, J.B., and Raushel, F.M., Carbamoyl phosphate synthetase: a tunnel runs through it, *Curr. Opin. Struct. Biol.* **8**, 679–685 (1998); and Thoden, J.B., Holden, H.M., Wessenberg, G., Raushel, F.M., and Rayment, I., Structure of carbamoyl phosphate synthetase: A journey of 96 Å from substrate to product, *Biochemistry* **36**, 6305–6316 (1997).
- Jansonius, J.N., Structure, evolution and action of vitamin B₆-dependent enzymes, *Curr. Opin. Struct. Biol.* **8**, 759–769 (1998).
- Jungas, R.L., Halperin, M.L., and Brosnan, J.T., Quantitative analysis of amino acid oxidation and related gluconeogenesis in humans, *Physiol. Rev.* **72**, 419–448 (1992).
- Miles, E.W., Rhee, S., and Davies, D.R., The molecular basis of substrate channeling, *Biochemistry* **274**, 12193–12196 (1999).
- Saeed-Kothe, A. and Powers-Lee, S.G., Specificity determining residues in ammonia- and glutamine-dependent carbamoyl phosphate synthetases, *J. Biol. Chem.* **277**, 7231–7238 (2002).
- Smith, T.J., Schmidt, T., Fang, J., Wu, J., Siuzdak, G., and Stanley, C.A., The structure of apo human glutamate dehydrogenase details subunit communication and allostery, *J. Mol. Biol.* **318**, 765–777 (2002).
- Smith, T.J. and Stanley, C.A., Untangling the glutamate dehydrogenase allosteric nightmare, *Trends Biochem. Sci.* **33**, 557–564 (2008).
- Stipanuk, M.H. and Watford, M., Amino acid metabolism, Chap. 11, in Stipanuk, M.H. (Ed.), *Biochemical and Physiological Basis of Nutrition*, Saunders (2000).
- Torchinsky, Yu.M., Transamination: its discovery, biological and chemical aspects (1937–1987), *Trends Biochem. Sci.* **12**, 115–117 (1987).

Metabolic Breakdown of Individual Amino Acids

- Anderson, O.A., Flatmark, T., and Hough, E., High resolution crystal structures of the catalytic domain of human phenylalanine hydroxylase in its catalytically active Fe(II) form and binary complex with tetrahydrobiopterin, *J. Mol. Biol.* **314**, 279–291 (2001).
- Evarsson, A., Chuang, J.L., Wynn, R.M., Turley, S., Chuang, D.T., and Hol, W.G.J., Crystal structure of human branched-chain α-ketoacid dehydrogenase and the molecular basis of multienzyme complex deficiency in maple syrup urine disease, *Structure* **8**, 277–291 (2000); and Wynn, R.M., Davie, J.R., Chuang, J.L., Cote, C.D., and Chuang, D.T., Impaired assembly of E1 de-carboxylase of the branched-chain α-ketoacid dehydrogenase complex in type IA maple syrup urine disease, *J. Biol. Chem.* **273**, 13110–13118 (1998).
- Binda, C., Bossi, R.T., Wakatsuki, S., Arzt, S., Coda, A., Curti, B., Vanoni, M.A., and Mattevi, A., Cross-talk and ammonia channeling between active centers in the unexpected domain arrangement of glutamate synthase, *Structure* **8**, 1299–1308 (2000).

- Douce, R., Bourguignon, J., Neuburger, M., and Rébeillé, F., The glycine decarboxylase system: a fascinating complex, *Trends Plant Sci.* **6**, 167–176 (2001).
- Drennen, C.L., Huang, S., Drumond, J.T., Matthews, R., and Ludwig, M.L., How a protein binds B₁₂: A 3.0 Å X-ray structure of B₁₂-binding domains of methionine synthase, *Science* **266**, 1669–1674 (1994).
- Faure, M., Rourguignon, J., Neuburger, M., Macherel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C., and Douce, R., Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system, 2. Crystal structures of H- and L-proteins, *Eur. J. Biochem.* **267**, 2890–2898 (2000).
- Guenther, B.D., Sheppard, C.A., Tran, P., Rozen, R., Matthews, R.G., and Ludwig, M.L., The structure and properties of methylenetetrahydrofolate reductase from *Escherichia coli* suggest how folate ameliorates human hyperhomocysteine, *Nature Struct. Biol.* **6**, 359–365 (1999).
- Guilhaudis, L., Simorre, J.-P., Blackledge, M., Marion, D., Gans, P., Neuburger, M., and Douce, R., Combined structural and biochemical analysis of the H-T complex in the glycine decarboxylase cycle: evidence for a destabilization mechanism of the H-protein, *Biochemistry* **39**, 4259–4266 (2000).
- Huang, X., Holden, H.M., and Raushel, F.M., Channeling of substrates and intermediates in enzyme-catalyzed reactions, *Annu. Rev. Biochem.* **70**, 149–180 (2001).
- Kelly, A. and Stanley, C.A., Disorders of glutamate metabolism, *Mental Retard. Devel. Dis. Res. Rev.* **7**, 287–295 (2001).
- Ludwig, M.L. and Matthews, R.G., Structure-based perspectives on B₁₂-dependent enzymes, *Annu. Rev. Biochem.* **66**, 269–313 (1997).
- Matthews, R.G., Koutmos, M., and Datta, S., Cobalamin-dependent and cobamide-dependent methyltransferases, *Curr. Opin. Struct. Biol.* **18**, 658–666 (2008).
- Medina, M.Á., Urdiales, J.L., and Amores-Sánchez, M.I., Roles of homocysteine in cell metabolism: Old and new functions, *Eur. J. Biochem.* **268**, 3871–3882 (2001).
- Spiro, T.G. and Kozlowski, P.M., Is the CO adduct of myoglobin bent, and does it matter? *Acc. Chem. Res.* **34**, 137–144 (2001).
- Swain, A.L., Jaskólski, M., Housset, D., Rao, J.K.M., and Wladower, A., Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy, *Proc. Natl. Acad. Sci.* **90**, 1474–1478 (1993).
- Varughese, K.I., Skinner, M.M., Whiteley, J.M., Matthews, D.A., and Xuong, N.H., Crystal structure of rat liver dihydropteridine reductase, *Proc. Natl. Acad. Sci.* **89**, 6080–6084 (1992).
- Zalkin, H. and Smith, J.L., Enzymes utilizing glutamine as an amide donor, *Adv. Enzymol.* **72**, 87–144 (1998).
- Amino Acids as Biosynthetic Precursors
- Ajioka, R.S., Phillips, J.D., and Kushner, J.P., Biosynthesis of heme in mammals, *Biochim. Biophys. Acta* **1763**, 723–736 (2006).
- Al-Karadaghi, S., Franco, R., Hansson, M., Shelnutt, J.A., Isaya, G., and Ferreira, G.C., Chelatases: distort to select, *Trends Biochem. Sci.* **31**, 135–142 (2006).
- Battersby, A.R., Tetrapterroroles: the pigments of life, *Nat. Prod. Rep.* **17**, 507–526 (2000).
- Erskine, P.T., Newbold, R., Brindley, A.A., Wood, S.P., Shoolingin-Jordan, P.M., Warren, M.J., and Cooper, J.B., The X-ray structure of yeast 5-aminolaevulinic acid dehydratase complexed with substrate and three inhibitors, *J. Mol. Biol.* **312**, 133–141 (2001).

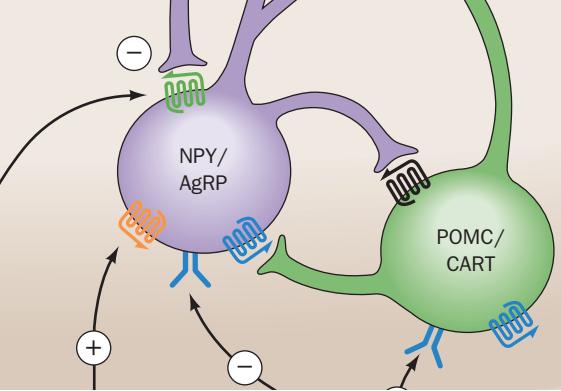
- Fitzpatrick, P.F., Tetrahydropterin-dependent amino acid hydroxylases, *Annu. Rev. Biochem.* **68**, 355–381 (1999).
- Jaffe, E.K., Martins, J., Li, J., Kervinen, J., and Dunbrack, R.L., Jr., The molecular mechanism of lead inhibition of human porphobilinogen synthase, *J. Biol. Chem.* **276**, 1531–1537 (2001).
- Jaffe, E.K., Morpheins—a new structural paradigm for allosteric regulation, *Trends Biochem. Sci.* **39**, 490–497 (2005); and Lawrence, S.H. and Jaffe, E.K., Expanding the concepts in protein structure-function relationships and enzyme kinetics: Teaching using morpheins, *Biochem. Mol. Biol. Educ.* **36**, 274–283 (2008).
- Kauppinen, R., Porphyrias, *Lancet* **365**, 241–252 (2005).
- Louie, G.V., Brownlie, P.D., Lambert, R., Cooper, J.B., Blundell, T.L., Wood, S.P., Warren, M.J., Woodcock, S.C., and Jordan, P.M., Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site, *Nature* **359**, 33–39 (1992).
- Medlock, A.E., Dailey, T.A., Ross, T.A., Dailey, H.A., and Lanzilotta, W.N., A π -helix switch selective for porphyrin deprotonation and product release in human ferrochelatase, *J. Mol. Biol.* **373**, 1006–1016 (2007).
- Pagola, S., Stephens, P.W., Bohle, D.S., Kosar, A.D., and Madsen, S.K., The structure of malaria pigment β -haematin, *Nature* **404**, 307–310 (2000).
- Schneider-Yin, X., Gouya, L., Dorsey, M., Rüfenacht, U., Deybach, J.-C., and Ferreira, G.C., Mutations in the iron-sulfur cluster ligands of the human ferrochelatase lead to erythropoietic protoporphyrinia, *Blood* **96**, 1545–1549 (2000).
- Sellers, V.M., Wu, C.-K., Dailey, T.A., and Dailey, H.A., Human ferrochelatase: Characterization of substrate-iron binding and proton-abstracting residues, *Biochemistry* **40**, 9821–9827 (2001).
- Shoolingin-Jordan, P.M., Warren, M.J., and Awan, S.J., Discovery that the assembly of the dipyromethane cofactor of porphobilinogen deaminase holoenzyme proceeds initially by the reaction of preuroporphyrinogen with the apoenzyme, *Biochem. J.* **316**, 373–376 (1996).
- Song, G., Li, Y., Cheng, C., Zhao, Y., Gao, A., Zhang, R., Joachimiak, A., Shaw, N., and Li, Z.-J., Structural insight into acute intermittent porphyria, *FASEB J.* **23**, 396–404 (2009). [Reports the X-ray structure of human porphobilinogen deaminase.]
- Thunell, S., Porphyrins, porphyrin metabolism and porphyrias. I. Update, *Scand. J. Clin. Lab. Invest.* **60**, 509–540 (2000).
- Wellems, T.E., How chloroquine works, *Nature* **355**, 108–109 (1992).
- Wu, C.-K., Dailey, H.A., Rose, J.P., Burden, A., Sellers, V.M., and Wang, B.-C., The 2.0 Å structure of human ferrochelatase, the terminal enzyme of heme biosynthesis, *Nature Struct. Biol.* **8**, 156–160 (2001).
- Amino Acids Biosynthesis**
- Chaudhuri, B.N., Lange, S.C., Myers, R.S., Chittur, S.V., Davisson, V.J., and Smith, J.L., Crystal structure of imidazole glycerol phosphate synthase: a tunnel through a $(\beta/\alpha)_8$ barrel joins two active sites, *Structure* **9**, 987–997 (2001).
- Dunn, M.F., Niks, D., Ngo, H., Barends, T.R.M., and Schlichting, I., Tryptophan synthase: the workings of a channeling nanomachine, *Trends Biochem. Sci.* **33**, 254–264 (2008).
- Eisenberg, D., Gill, H.S., Pfluegl, M.U., and Rotstein, S.H., Structure-function relationships of glutamine synthetases, *Biochim. Biophys. Acta* **1477**, 122–145 (2000); and Gill, H.S. and Eisenberg, D., The crystal structure of phosphothridin in the active site of glutamine synthetase illuminates the mechanism of enzymatic inhibition, *Biochemistry* **40**, 1903–1912 (2001).
- Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., and Davies, D.R., Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*, *J. Biol. Chem.* **263**, 17857–17871 (1988).
- Katagiri, M. and Nakamura, M., Animals are dependent on preformed α -amino nitrogen as an essential nutrient, *Life* **53**, 125–129 (2002).
- Kishore, G.M. and Shah, D.M., Amino acid biosynthesis inhibitors as herbicides, *Annu. Rev. Biochem.* **57**, 627–663 (1988). [Discusses the biosynthesis of the essential amino acids.]
- Larsen, T.M., Boehlein, S.K., Schuster, S.M., Richards, N.G.J., Thoden, J.B., Holden, H.M., and Rayment, I., Three-dimensional structure of *Escherichia coli* asparagine synthetase B: a short journey from substrate to product, *Biochemistry* **38**, 16146–16167 (1999).
- Stadtman, E.R., The story of glutamine synthetase regulation, *J. Biol. Chem.* **276**, 44357–44364 (2001).
- Stallings, W.C., Abdel-Meguid, S.S., Lim, L.W., Shieh, H.-S., Dayringer, H.E., Leimgruber, N.K., Stegeman, R.A., Anderson, K.S., Sikorski, J.A., Padgett, S.R., and Kishore, G.M., Structure and topological symmetry of the glyphosate target 5-enol-pyruvylshikimate-3-phosphate synthase: A distinctive protein fold, *Proc. Natl. Acad. Sci.* **88**, 5046–5050 (1991). [The enzyme that catalyzes Reaction 6 of Fig. 26-62 in complex with glyphosate, an inhibitor that is a broad-spectrum herbicide.]
- Weeks, A., Lund, L., and Raushel, F.M., Tunneling of intermediates in enzyme-catalyzed reactions, *Curr. Opin. Chem. Biol.* **10**, 465–472 (2006).

Nitrogen Fixation

- Einsle, O., Tezcan, F.A., Andrade, A.L.A., Schmidt, B., Yoshida, M., Howard, J.B., and Rees, D.C., Nitrogenase MoFe-protein at 1.16 Å resolution: A central ligand in the FeMo-cofactor, *Science* **297**, 1696–1700 (2002).
- Fisher, R.F. and Long, S.R., *Rhizobium*–plant signal exchange, *Nature* **357**, 655–660 (1992). [Discusses the signals through which Rhizobiaceae and legumes communicate to symbiotically generate the root nodules in which nitrogen fixation occurs.]
- Jang, S.B., Seefeldt, L.C., and Peters, J.W., Insights into nucleotide signal transduction in nitrogenase: Structure of an iron protein with MgADP bound, *Biochemistry* **39**, 14745–14752 (2000).
- Lawson, D.M. and Smith, B.E., Molybdenum nitrogenases: a crystallographic and mechanistic view, *Metal Ions Biol. Sys.* **39**, 75–120 (2002).
- Mayer, S.M., Lawson, D.M., Gormal, C.A., Roe, S.M., and Smith, B.E., New insights into structure-function relationships in nitrogenase: A 1.6 Å resolution X-ray crystallographic study of *Klebsiella pneumoniae* MoFe-protein, *J. Mol. Biol.* **292**, 871–891 (1999).
- Peters, J.W., Stowell, M.H.B., Soltis, S.M., Finnegan, M.G., Johnson, M.K., and Rees, D.C., Redox-dependent structural changes in the nitrogenase P-cluster, *Biochemistry* **36**, 1181–1187 (1997).
- Peters, J.W. and Szilagyi, R.K., Exploring new frontiers of nitrogenase structure and mechanism, *Curr. Opin. Chem. Biol.* **10**, 101–108 (2006).
- Rees, D.C., Tezcan, F.A., Haynes, C.A., Walton, M.Y., Andrade, S., Einsle, O., and Howard, J.B., Structural basis of biological nitrogen fixation, *Philos. Trans. Roy. Soc. A* **363**, 971–984 (2005); and Howard, J.B. and Rees, D.C., How many metals does it take to fix N_2 ? A mechanistic overview of biological nitrogen fixation, *Proc. Natl. Acad. Sci.* **103**, 17088–17093 (2006).
- Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B., and Rees, D.C., Structure of ADP·AlF₄⁻ stabilized nitrogenase complex and its implications for signal transduction, *Nature* **387**, 370–376 (1997).
- Seefeldt, L.C., Hoffman, B.M., and Dean, D.R., Mechanism of Mo-dependent nitrogenase, *Annu. Rev. Biochem.* **78**, 701–722 (2009).

PROBLEMS

- 1.** Write the reaction for the transamination of an amino acid in terms of Cleland notation (Section 14-5A).
- 2.** Explain why the symptoms of the partial deficiency of a urea cycle enzyme may be attenuated by a low-protein diet.
- 3.** Why are people on a high-protein diet instructed to drink lots of water?
- 4.** A student on a particular diet expends $10,000 \text{ kJ} \cdot \text{day}^{-1}$ while excreting 40 g of urea. Assuming that protein is 16% N by weight and that its metabolism yields $18 \text{ kJ} \cdot \text{g}^{-1}$, what percentage of the student's energy requirement is met by protein?
- 5.** Production of the enzymes that catalyze the reactions of the urea cycle can increase or decrease according to the metabolic needs of the organism. High levels of these enzymes are associated with high-protein diets as well as starvation. Explain this apparent paradox.
- 6.** *Helicobacter pylori*, the bacterium responsible for gastric ulcers, can survive in the stomach (where the pH is as low as 1.5) in part because it synthesizes large amounts of the enzyme urease. (a) Write the reaction for urea hydrolysis by urease. (b) Explain why this reaction could help establish a more hospitable environment for *H. pylori*, which tolerates acid but prefers to grow at near-neutral pH.
- 7.** Why are phenylketonurics warned against eating products containing the artificial sweetener **aspartame** (**NutraSweet**; chemical name L-aspartyl-L-phenylalanine methyl ester)?
- 8.** Demonstrate that the synthesis of heme from PBG as labeled in Fig. 26-35 results in the heme-labeling pattern given in Fig. 26-32.
- 9.** Explain why certain drugs and other chemicals can precipitate an attack of acute intermittent porphyria.
- 10.** Heterozygotes for erythropoietic protoporphyrin show only 20 to 30% residual ferrochelatase activity rather than the 50% that is normally expected for an autosomal dominant inherited disease. Provide a plausible explanation for this observation.
- 11.** One of the symptoms of **kwashiorkor**, the dietary protein deficiency disease in children, is the depigmentation of the skin and hair. Explain the biochemical basis of this symptom.
- 12.** What are the metabolic consequences of a defective uridylyl-removing enzyme in *E. coli*?
- 13.** Figure 26-60, Reaction 9, indicates that methionine is synthesized in microorganisms by the methylation of homocysteine in a reaction in which N^5 -methyl-THF is the methyl donor. Yet, in the breakdown of methionine (Fig. 26-18), its demethylation occurs in three steps in which SAM is an intermediate. Discuss why this reaction does not occur via the simpler one-step reversal of the methylation reaction.
- *14.** In the glucose-alanine cycle (Fig. 26-3), glycolytically derived pyruvate is transaminated to alanine and exported to the liver for conversion to glucose and return to the cell. Explain how a muscle cell is able to participate in this cycle under anaerobic (vigorously contracting) conditions. (Hint: The breakdown of many amino acids yields NH_3 .)
- 15.** Draw the activated intermediates involved in (a) glutamine and (b) asparagine biosynthesis from glutamate and aspartate, respectively. (c) Provide an example of another metabolic activation of a carboxylic acid group analogous to each of these reactions.
- 16.** The $\alpha_2\beta_2$ tetramer of tryptophan synthase catalyzes the PLP-dependent reaction of indole-3-glycerol phosphate and serine to form tryptophan (Fig. 26-63, Reactions 5 and 6). Draw the chemical reactions involved in this synthesis, including the participation of PLP, and use curved arrows to show the flow of electrons. What role does PLP play in the reaction?
- 17.** Suggest a reason why the nitrogen-fixing heterocysts of cyanobacteria have lost Photosystem II but retain Photosystem I.



Energy Metabolism: Integration and Organ Specialization

CHAPTER 27

1 Major Pathways and Strategies of Energy Metabolism: A Summary

2 Organ Specialization

- A. Brain
- B. Muscle
- C. Adipose Tissue
- D. Liver
- E. Kidney

3 Metabolic Homeostasis: Regulation of Appetite, Energy Expenditure, and Body Weight

- A. AMP-Dependent Protein Kinase Is the Cell's Fuel Gauge
- B. Adiponectin Regulates AMPK Activity
- C. Leptin
- D. Insulin
- E. Ghrelin and PYY₃₋₃₆
- F. Hypothalamic Integration of Hormonal Signals
- G. Control of Energy Expenditure by Adaptive Thermogenesis
- H. Did Leptin Evolve as a Thrifty Gene?

4 Metabolic Adaptation

- A. Starvation
- B. Diabetes Mellitus

At this point in our narrative we have studied all of the major pathways of energy metabolism. Consequently, we are now in a position to consider how organisms, mammals in particular, orchestrate the metabolic symphony to meet their energy needs. This chapter therefore begins with a recapitulation of the major metabolic pathways and their control systems, then considers how these processes are apportioned among the various organs of the body, and ends with a discussion of metabolic adaptation, including how the body maintains energy balance (homeostasis), how it deals with the metabolic challenges of starvation and obesity, and how it responds to the loss of control resulting from diabetes mellitus.

1 MAJOR PATHWAYS AND STRATEGIES OF ENERGY METABOLISM: A SUMMARY

Figure 27-1 indicates the interrelationships among the major pathways involved in energy metabolism. Let us review these pathways and their control mechanisms.

1. Glycolysis (Chapter 17) The metabolic degradation of glucose begins with its conversion to two molecules of pyruvate with the net generation of two molecules each of ATP and NADH. Under anaerobic conditions, pyruvate is converted to lactate (or, in yeast, to ethanol) so as to recycle the NADH. Under aerobic conditions, however, when glycolysis serves to prepare glucose for further oxidation, the NAD⁺ is regenerated through oxidative phosphorylation (see below). The flow of metabolites through the glycolytic pathway is largely controlled by the activity of phosphofructokinase (PFK). This enzyme is activated by AMP and ADP, whose concentrations rise as the need for energy metabolism increases, and is inhibited by ATP and citrate, whose concentrations increase when the demand for energy metabolism has slackened. Citrate, a citric acid cycle intermediate, also inhibits PFK and glycolysis when aerobic metabolism takes over from anaerobic metabolism, making glucose oxidation more efficient (the Pasteur effect; Section 22-4C), and when fatty acid and/or ketone body oxidation (which are also aerobic pathways) are providing for energy needs (the glucose–fatty acid or Randle cycle; Section 22-4Bb). PFK is also activated by fructose-2,6-bisphosphate, whose concentration is regulated by the levels of glucagon, epinephrine, and norepinephrine through the intermediacy of cAMP (Section 18-3Fc). Liver and heart muscle F2,6P levels are regulated oppositely: A [cAMP] increase causes an [F2,6P] decrease in liver and an [F2,6P] increase in heart muscle. However, skeletal muscle [F2,6P] does not respond to changes in [cAMP].

2. Gluconeogenesis (Section 23-1) Mammals can synthesize glucose from a variety of precursors, including pyruvate, lactate, glycerol, and glucogenic amino acids (but not fatty acids), through pathways that occur mainly in liver and kidney. Many of these precursors are converted to oxaloacetate which, in turn, is converted to phosphoenolpyruvate and then, through a series of reactions that largely reverse the path of glycolysis, to glucose. The irreversible steps of glycolysis, those catalyzed by PFK and hexokinase, are bypassed in gluconeogenesis by hydrolytic reactions catalyzed, respectively, by fructose-1,6-bisphosphatase (FBPase) and glucose-6-phosphatase. FBPase and PFK may both be at least partially active simultaneously, creating a substrate cycle. This cycle, and the reciprocal regulation of PFK and FBPase, are important in regulating both the rate and direction of flux through glycolysis and

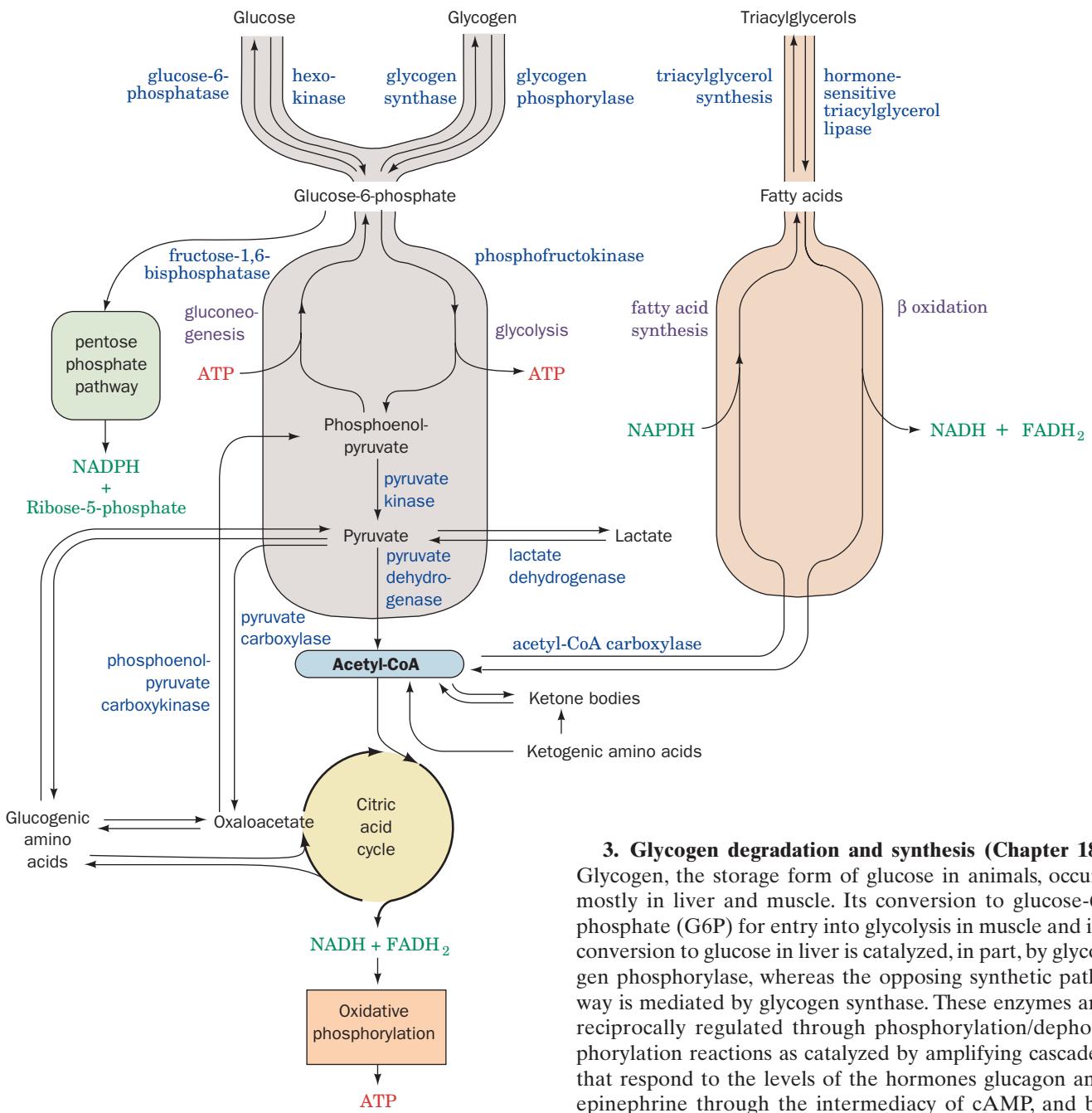


Figure 27-1 The major energy metabolism pathways.

gluconeogenesis (Sections 17-4F and 23-1B). Fatty acid and ketone body oxidation can increase the rate of gluconeogenesis in liver by decreasing the concentration of F2,6P (Section 18-3Fc). This occurs because the increased citrate concentration accompanying activation of the citric acid cycle during fatty acid oxidation inhibits PFK-2 as well as PFK (Table 23-1). Phosphoenolpyruvate carboxykinase (PEPCK) bypasses the third irreversible reaction of glycolysis, that catalyzed by pyruvate kinase (PK), and is controlled exclusively by long-term transcriptional regulation.

3. Glycogen degradation and synthesis (Chapter 18)

Glycogen, the storage form of glucose in animals, occurs mostly in liver and muscle. Its conversion to glucose-6-phosphate (G6P) for entry into glycolysis in muscle and its conversion to glucose in liver is catalyzed, in part, by glycogen phosphorylase, whereas the opposing synthetic pathway is mediated by glycogen synthase. These enzymes are reciprocally regulated through phosphorylation/dephosphorylation reactions as catalyzed by amplifying cascades that respond to the levels of the hormones glucagon and epinephrine through the intermediacy of cAMP, and by insulin (Sections 18-3E and 19-4F). The glucagon–insulin ratio is therefore a crucial factor in determining the rate and direction of glycogen metabolism.

4. Fatty acid degradation and synthesis (Sections 25-1 through 25-5) Fatty acids are broken down in increments of C₂ units through β oxidation to form acetyl-CoA. They are synthesized from this compound via a separate pathway. The activity of the β-oxidation pathway varies with the fatty acid concentration. This, in turn, depends on the activity of “hormone-sensitive” triacylglycerol lipase in adipose tissue that is stimulated, through cAMP-regulated phosphorylation/dephosphorylation reactions, by glucagon and epinephrine but inhibited by insulin. The fatty acid synthesis rate varies with the activity of acetyl-CoA carboxylase,

which is activated by citrate and insulin-dependent dephosphorylation, and inhibited by the pathway product palmitoyl-CoA and by cAMP- and AMP-dependent phosphorylation. Fatty acid synthesis is also subject to long-term regulation through alterations in the rates of synthesis of the enzymes mediating this process as stimulated by insulin and inhibited by fasting. *The glucagon–insulin ratio is therefore of prime importance in determining the rate and direction of fatty acid metabolism.*

5. Citric acid cycle (Chapter 21) The citric acid cycle oxidizes acetyl-CoA, the common degradation product of glucose, fatty acids, ketone bodies, and ketogenic amino acids, to CO_2 and H_2O with the concomitant production of NADH and FADH_2 . Many glucogenic amino acids can also be oxidized via the citric acid cycle through their breakdown, ultimately to pyruvate and then to acetyl-CoA, sometimes via the **cataplerosis** (using up) of a citric acid cycle intermediate (Section 21-5). The activities of the citric acid cycle regulatory enzymes citrate synthase, isocitrate dehydrogenase, and α -ketoglutarate dehydrogenase are controlled by substrate availability and feedback inhibition by cycle intermediates, NADH, and ATP.

6. Oxidative phosphorylation (Chapter 22) This mitochondrial pathway oxidizes NADH and FADH_2 to NAD^+ and FAD with the coupled synthesis of ATP. The rate of oxidative phosphorylation, which is tightly coordinated with the metabolic fluxes through glycolysis and the citric acid cycle, is largely dependent on the concentrations of ATP, ADP, and P_i , as well as O_2 .

7. Pentose phosphate pathway (Section 23-4) This pathway functions to generate NADPH for use in reductive biosynthesis, as well as the nucleotide precursor ribose-5-phosphate, through the oxidation of G6P. Its flux-generating step is catalyzed by glucose-6-phosphate dehydrogenase, which is controlled by the level of NADP^+ . *The ability of enzymes to distinguish between NADH, which is mainly utilized in energy metabolism, and NADPH permits energy metabolism and biosynthesis to be regulated independently.*

8. Amino acid degradation and synthesis (Sections 26-1 through 26-5) Excess amino acids may be degraded to common metabolic intermediates. Most of these pathways begin with an amino acid's transamination to its corresponding α -keto acid with the eventual transfer of the amino group to urea via the urea cycle. Leucine and lysine are ketogenic amino acids in that they can be converted only to acetyl-CoA or acetoacetate and hence cannot be glucose precursors. The other amino acids are glucogenic in that they may be, at least in part, converted to one of the glucose precursors pyruvate, oxaloacetate, α -ketoglutarate, succinyl-CoA, or fumarate. Five amino acids are both ketogenic and glucogenic. Essential amino acids are those that an animal cannot synthesize itself; they must be obtained from plant and microbial sources. Nonessential amino acids can be synthesized by animals utilizing preformed amino groups via pathways that are generally simpler than those synthesizing essential amino acids.

Two compounds lie at the crossroads of the foregoing metabolic pathways: acetyl-CoA and pyruvate (Fig. 27-1).

Acetyl-CoA is the common degradation product of most metabolic fuels, including polysaccharides, lipids, and proteins. Its acetyl group may be oxidized to CO_2 and H_2O via the citric acid cycle and oxidative phosphorylation or used to synthesize fatty acids. Pyruvate is the product of glycolysis, the dehydrogenation of lactate, and the breakdown of certain glucogenic amino acids. It may be oxidatively decarboxylated to yield acetyl-CoA, thereby committing its atoms either to oxidation or to the biosynthesis of fatty acids. Alternatively, it may be carboxylated via the pyruvate carboxylase reaction to form oxaloacetate, which, in turn, either replenishes citric acid cycle intermediates or enters gluconeogenesis via phosphoenolpyruvate, thereby bypassing an irreversible step in glycolysis. Pyruvate is therefore a precursor of several amino acids as well as of glucose.

The foregoing pathways occur in specific cellular compartments. Glycolysis, glycogen synthesis and degradation, fatty acid synthesis, and the pentose phosphate pathway are largely or entirely cytosolically based, whereas fatty acid degradation, the citric acid cycle, and oxidative phosphorylation occur in the mitochondrion. Different phases of gluconeogenesis and amino acid degradation occur in each of these compartments. *The flow of metabolites across compartment membranes is mediated, in most cases, by specific carriers that are also subject to regulation.*

The enormous number of enzymatic reactions that simultaneously occur in every cell (Fig. 16-1) must be coordinated and strictly controlled to meet the cell's needs. Such regulation occurs on many levels. Intercellular communications regulating metabolism occur via certain hormones, including epinephrine, norepinephrine, glucagon, and insulin, as well as through a series of steroid hormones known as **glucocorticoids** (whose effects are discussed in Section 19-1Ga). These hormonal signals trigger a variety of cellular responses, including the synthesis of second messengers such as cAMP in the short term and the modulation of protein synthesis rates in the long term. On the molecular level, enzymatic reaction rates are controlled by phosphorylation/dephosphorylation via amplifying reaction cascades, by allosteric responses to the presence of effectors, which are usually precursors or products of the reaction pathway being controlled, and by substrate availability. The regulatory machinery of opposing catabolic and anabolic pathways is generally arranged such that these pathways are reciprocally regulated.

2 ORGAN SPECIALIZATION

Different organs have different metabolic functions and capabilities. In this section we consider how the special needs of the mammalian body organs are met and how their metabolic capabilities are coordinated to meet these needs. In particular, we discuss brain, muscle, adipose tissue, liver, and kidney (Fig. 27-2).

A. Brain

Brain tissue has a remarkably high respiration rate. For instance, the human brain constitutes only $\sim 2\%$ of the adult

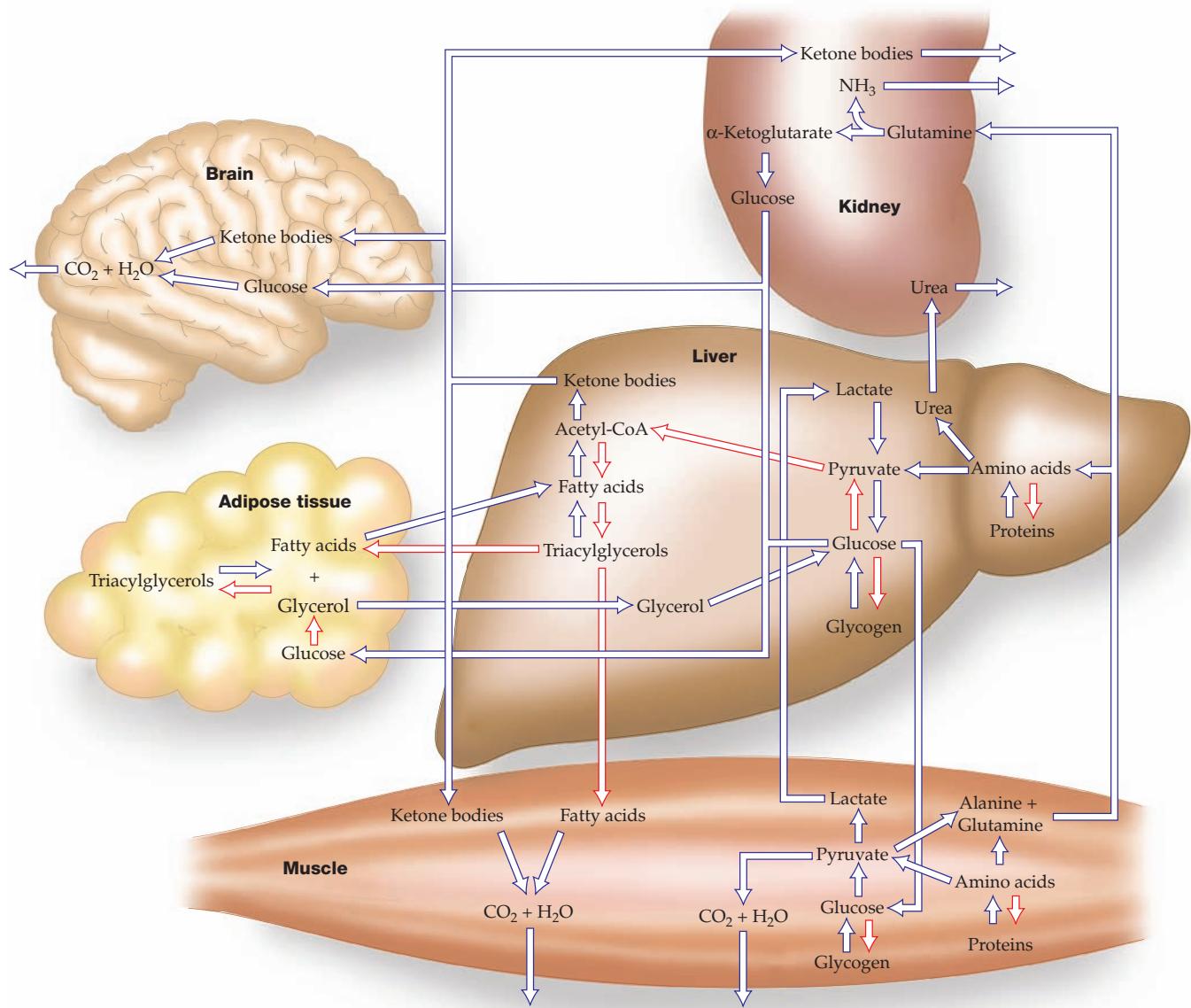


Figure 27-2 The metabolic interrelationships among brain, adipose tissue, muscle, liver, and kidney. The red-outlined arrows indicate pathways that predominate in the well-fed state when

body mass but is responsible for ~20% of its resting O₂ consumption. This consumption, moreover, is independent of the state of mental activity; it varies little between sleep and the intense concentration required of, say, the study of biochemistry. Most of the brain's energy production serves to power the plasma membrane (Na⁺-K⁺)-ATPase (Section 20-3A), which maintains the membrane potential required for nerve impulse transmission (Section 20-5). In fact, the respiration of brain slices is over 50% reduced by the (Na⁺-K⁺)-ATPase inhibitor ouabain (Section 20-3Af).

Under usual conditions, glucose serves as the brain's only fuel (although, with extended fasting, the brain gradually switches to ketone bodies; Section 27-4A). Indeed, since brain cells store very little glycogen, they require a

steady supply of glucose from the blood. A blood glucose concentration of less than half of the normal value of ~5 mM results in brain dysfunction. Levels much below this, for example, caused by severe insulin overdose, result in coma, irreversible damage, and ultimately death. One of the liver's major functions, therefore, is to maintain the blood glucose level (Sections 18-3F and 27-2D).

B. Muscle

Muscle's major fuels are glucose from glycogen, fatty acids, and ketone bodies. Rested, well-fed muscle, in contrast to brain, synthesizes a glycogen store comprising 1 to 2% of its mass. The glycogen serves muscle as a readily available

fuel depot since it can be rapidly converted to G6P for entry into glycolysis (Section 18-1).

Muscle cannot export glucose because it lacks glucose-6-phosphatase. Nevertheless, muscle serves the body as an energy reservoir because, during the fasting state, its proteins are degraded to amino acids, many of which are converted to pyruvate, which in turn, is transaminated to alanine. The alanine is then exported via the bloodstream to the liver, which transaminates it back to pyruvate, a glucose precursor. This process is known as the glucose-alanine cycle (Section 26-1Ad).

Since muscle does not participate in gluconeogenesis, it lacks the machinery that regulates this process in such gluconeogenic organs as liver and kidney. Muscle does not have receptors for glucagon, which, it will be recalled, stimulates an increase in blood glucose levels (Section 18-3F). However, muscle possesses epinephrine receptors (β -adrenergic receptors; Section 19-1F), which through the intermediacy of cAMP control the phosphorylation/dephosphorylation cascade system that regulates glycogen breakdown and synthesis (Section 18-3). This is the same cascade system that controls the competition between glycolysis and gluconeogenesis in liver in response to glucagon.

Heart muscle and skeletal muscle contain different isozymes of PFK-2/FBPase-2. The heart muscle isozyme is controlled by phosphorylation oppositely to that in liver, whereas skeletal muscle PFK-2/FBPase-2 is not controlled by phosphorylation at all (Section 18-3Fc). Thus the concentration of F2,6P rises in heart muscle but falls in liver in response to an increase in [cAMP]. Moreover the muscle isozyme of pyruvate kinase, which, it will be recalled, catalyzes the final step of glycolysis, is not subject to phosphorylation/dephosphorylation as is the liver isozyme (Section 23-1Ba). Thus, whereas an increase in liver cAMP stimulates glycogen breakdown and gluconeogenesis, resulting in glucose export, an increase in heart muscle cAMP activates glycogen breakdown and glycolysis, resulting in glucose consumption. Consequently, epinephrine, which prepares the organism for action (fight or flight), acts independently of glucagon which, acting reciprocally with insulin, regulates the general level of blood glucose.

a. Muscle Contraction Is Anaerobic Under Conditions of High Exertion

Muscle contraction is driven by ATP hydrolysis (Section 35-3Bb) and is therefore ultimately dependent on respiration. Skeletal muscle at rest utilizes $\sim 30\%$ of the O_2 consumed by the human body. A muscle's respiration rate may increase in response to a heavy workload by as much as 25-fold. Yet, its rate of ATP hydrolysis can increase by a much greater amount. The ATP is initially regenerated by the reaction of ADP with phosphocreatine as catalyzed by creatine kinase (Section 16-4Cd):



(phosphocreatine is resynthesized in resting muscle by the reversal of this reaction). Under conditions of maximum exertion, however, such as occurs in a sprint, a muscle has

only an ~ 5 -s supply of phosphocreatine. It must then shift to ATP production via glycolysis of G6P resulting from glycogen breakdown, a process whose maximum flux greatly exceeds those of the citric acid cycle and oxidative phosphorylation. Much of this G6P is therefore degraded anaerobically to lactate (Section 17-3A) which, in the Cori cycle (Section 23-1C), is exported via the bloodstream to the liver, where it is reconverted to glucose through gluconeogenesis. Gluconeogenesis requires ATP generated by oxidative phosphorylation. Muscles thereby shift much of their respiratory burden to the liver and consequently also delay the O_2 -consumption process, a phenomenon known as oxygen debt. The source of ATP during exercise of varying duration is summarized in Fig. 27-3.

b. Muscle Fatigue Has a Protective Function

Muscle fatigue, defined as the inability of a muscle to maintain a given power output, occurs in ~ 20 s under conditions of maximum exertion. Such fatigue is not caused by the exhaustion of the muscle's glycogen supply. Rather, it may result from glycolytic proton generation that can drop the intramuscular pH from its resting value of 7.0 to as low as 6.4 (fatigue does not, as is widely believed, result from the buildup of lactate itself, as is demonstrated by the observation that muscles can sustain a large power output under high lactate concentrations if the pH is maintained near 7.0). Nevertheless, how acidification might cause muscle fatigue is unclear. Two other proposed causes for muscle fatigue are (1) the increased $[P_i]$ arising largely from the

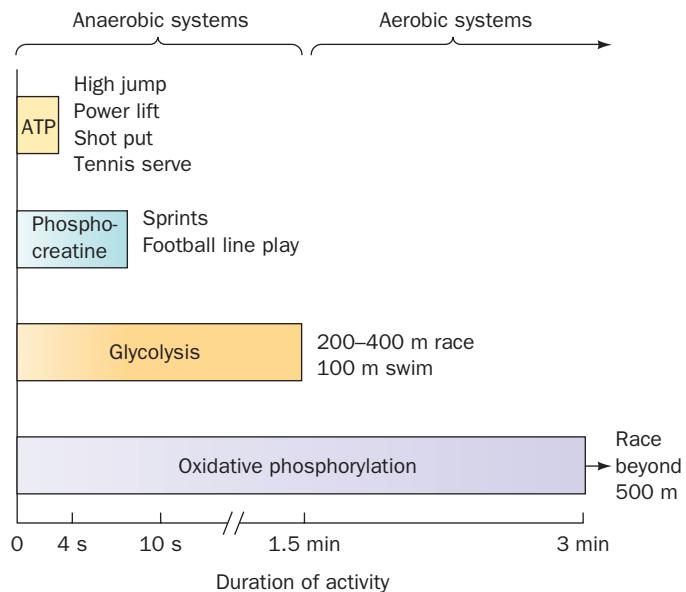


Figure 27-3 Source of ATP during exercise in humans. The supply of endogenous ATP is extended for a few seconds by phosphocreatine, after which anaerobic glycolysis generates ATP. The shift from anaerobic to aerobic metabolism (oxidative phosphorylation) occurs after about 90 s, or slightly later in trained athletes. [Adapted from McArdle, W.D., Katch, F.I., and Katch, V.L., *Exercise Physiology*, 2nd ed., Lea & Febiger (1986), p. 348.]

utilization of ATP may precipitate Ca^{2+} as calcium phosphate (which is highly insoluble), thereby decreasing contractile force (muscle contraction is triggered by the release of Ca^{2+} ion; Section 35-3Cb); and (2) the K^+ ion known to be released from contracting muscle cells may result in their depolarization (Section 20-5Ba) and hence a reduction in their contraction. Whatever its cause(s), it seems likely that muscle fatigue is an adaptation that prevents muscle cells from committing suicide by exhausting their ATP supply (recall that glycolysis and other ATP-generating pathways must be primed by ATP).

C. The Heart Is a Largely Aerobic Organ

The heart is a muscular organ but one that must maintain continuous rather than intermittent activity. Thus heart muscle, except for short periods of extreme exertion, relies entirely on aerobic metabolism. It is therefore richly endowed with mitochondria; they comprise up to 40% of its cytoplasmic space, whereas some types of skeletal muscle are nearly devoid of mitochondria. The heart can metabolize fatty acids, ketone bodies, glucose, pyruvate, and lactate. Fatty acids are the resting heart's fuel of choice but, on the imposition of a heavy workload, the heart greatly increases its rate of consumption of glucose, which is derived mostly from its relatively limited glycogen store.

C. Adipose Tissue

Adipose tissue, which consists of cells known as adipocytes (Fig. 12-2), is widely distributed about the body but occurs most prominently under the skin, in the abdominal cavity, in skeletal muscle, around blood vessels, and in mammary gland. The adipose tissue of a normal 70-kg man contains ~ 15 kg of fat. This amount represents some 590,000 kJ of energy (141,000 dieter's Calories), which is sufficient to maintain life for ~ 3 months. Yet, adipose tissue is by no means just a passive storage depot. In fact, it is second in importance only to liver in the maintenance of metabolic homeostasis (Section 27-3).

Adipose tissue obtains most of its fatty acids from the liver or from the diet as described in Section 25-1. Fatty acids are activated by the formation of the corresponding fatty acyl-CoA and then esterified with glycerol-3-phosphate to form the stored triacylglycerols (Section 25-4F). The glycerol-3-phosphate arises from the reduction of dihydroxyacetone phosphate, which must be glycolytically generated from glucose or gluconeogenically generated from pyruvate or oxaloacetate (a process called **glycroneogenesis**; Section 25-4Fa) because adipocytes lack a kinase that phosphorylates endogenous glycerol.

Adipocytes hydrolyze triacylglycerols to fatty acids and glycerol in response to the levels of glucagon, epinephrine, and insulin through a reaction catalyzed by hormone-sensitive triacylglycerol lipase (Section 25-5). If glycerol-3-phosphate is abundant, many of the fatty acids so formed are reesterified to triacylglycerols. Indeed, the average turnover time for triacylglycerols in adipocytes is only a few days. If, however, glycerol-3-phosphate is in short supply, the fatty acids are released into the bloodstream. *The rate of*

glucose uptake by adipocytes, which is regulated by insulin as well as by glucose availability, is therefore also an important factor in triacylglycerol formation and mobilization. However, glycerol-3-phosphate is also produced via glycconeogenesis under the control of PEPCK, allowing triacylglycerol turnover even when glucose concentration is low.

a. Obesity Results from Aberrant Metabolic Control

The human body regulates glycogen and protein levels within relatively narrow limits, but fat reserves, which are much larger, can become enormous. The accumulation of fatty acids as triacylglycerols in adipose tissue is largely a result of excess fat or carbohydrate intake compared to energy expenditure. Fat synthesis from carbohydrates occurs when the carbohydrate intake is high enough that glycogen stores, to which excess carbohydrate is normally directed, approach their maximum capacity.

Obesity is one of the major health-related problems in industrial countries. An estimated 30% of adults in the United States are obese (are at least 20% above their desirable weights) and another 35% are overweight. Most obese people find it inordinately difficult to lose weight or, having done so, to keep it off. Yet most animals, including humans, tend to have stable weights; that is, if they are given free access to food, they eat just enough to maintain this so-called set point weight. The nature of the regulatory machinery that controls the set point, which in obese individuals seems to be aberrantly high, is just beginning to come to light (see Section 27-3).

Formerly grossly obese individuals who have lost at least 100 kg to reach their normal weights exhibit some of the metabolic symptoms of starvation: they are obsessed with food, have low heart rates, are cold intolerant, and require 25% less caloric intake than normal individuals of similar heights and weights. In both normal and obese individuals, some 50% of the fatty acids liberated by the hydrolysis of triacylglycerols are reesterified before they can leave the adipocytes. In formerly obese subjects, this reesterification rate is only 35 to 40%, a level similar to that observed in normal individuals after a several day fast. The fat cells in normal and obese individuals, moreover, are of roughly the same size; obese people just have more of them. In fact, adipocyte precursor cells from massively obese individuals proliferate excessively in tissue culture compared to those from normal or even moderately obese subjects (adipocytes themselves do not replicate). Since fat cells, once gained, are never lost, this suggests that adipocytes, although highly elastic in size, tend to maintain a certain fixed volume and in doing so influence the metabolism and thus the appetite. This insight, unfortunately, has not yet led to a method for lowering the set points of individuals with a tendency toward obesity.

D. Liver

The liver is the body's central metabolic clearinghouse. It functions to maintain the proper levels of nutrients in the blood for use by the brain, muscles, and other tissues. The liver is uniquely situated to carry out this task because all the

nutrients absorbed by the intestines except fatty acids are released into the portal vein, which drains directly into the liver.

One of the liver's major functions is to act as a blood glucose "buffer." It does so by taking up or releasing glucose in response to the levels of glucagon, epinephrine, and insulin as well as to the concentration of glucose itself. After a carbohydrate-containing meal, when the blood glucose concentration reaches ~ 6 mM, the liver takes up glucose by converting it to G6P. The process is catalyzed by glucokinase (Section 18-3Fa), which differs from hexokinase, the analogous glycolytic enzyme in other cells, in that glucokinase has a much lower affinity for glucose (glucokinase reaches half-maximal velocity at ~ 5 mM glucose vs <0.1 mM glucose for hexokinase) and is not inhibited by G6P. Liver cells, in contrast to muscle and adipose cells, are permeable to glucose, and thus insulin has no direct effect on their glucose uptake. Since the blood glucose concentration is normally less than glucokinase's K_M , the rate of glucose phosphorylation in the liver is more or less proportional to the blood glucose concentration. The other intestinally absorbed sugars, mostly fructose, galactose, and mannose, are also converted to G6P in the liver (Section 17-5). After an overnight fast, the blood glucose level drops to ~ 4 mM. The liver keeps it from dropping below this level by releasing glucose into the blood as is described below. In addition, lactate, the product of anaerobic glucose metabolism in the muscle, is taken up by the liver for use in gluconeogenesis and lipogenesis as well as in oxidative phosphorylation (the Cori cycle; Section 23-1C). Alanine produced in the muscle is taken up by the liver and converted to pyruvate for gluconeogenesis as well (the glucose-alanine cycle; Section 26-1Ad).

a. The Fate of Glucose-6-Phosphate Varies with Metabolic Requirements

G6P is at the crossroads of carbohydrate metabolism; it can have several alternative fates depending on the glucose demand (Fig. 27-1):

1. G6P can be converted to glucose by the action of glucose-6-phosphatase for transport via the bloodstream to the peripheral organs.

2. G6P can be converted to glycogen (Section 18-2) when the body's demand for glucose is low. Yet, increased glucose demand, as signaled by higher levels of glucagon and/or epinephrine, reverses this process (Section 18-1).

3. G6P can be converted to acetyl-CoA via glycolysis and the action of pyruvate dehydrogenase (Chapter 17 and Section 21-2). Most of this glucose-derived acetyl-CoA is used in the synthesis of fatty acids (Section 25-4), whose fate is described below, and in the synthesis of phospholipids (Section 25-8) and cholesterol (Section 25-6A). Cholesterol, in turn, is a precursor of bile salts, which are produced by the liver (Section 25-6C) for use as emulsifying agents in the intestinal digestion and absorption of fats (Section 25-1).

4. G6P can be degraded via the pentose phosphate pathway (Section 23-4) to generate the NADPH required

for fatty acid biosynthesis and the liver's many other biosynthetic functions, as well as ribose-5-phosphate (R5P) for nucleotide biosynthesis (Sections 28-1A and 28-2A).

b. The Liver Can Synthesize or Degrade Triacylglycerols

Fatty acids are also subject to alternative metabolic fates in the liver (Fig. 27-1):

1. When the demand for metabolic fuels is high, fatty acids are degraded to acetyl-CoA and then to ketone bodies (Section 25-3) for export via the bloodstream to the peripheral tissues.

2. When the demand for metabolic fuels is low, fatty acids are used to synthesize triacylglycerols that are secreted into the bloodstream as VLDL for uptake by adipose tissue. Fatty acids may also be incorporated into phospholipids (Section 25-8).

Since the rate of fatty acid oxidation varies only with fatty acid concentration (Section 25-5), fatty acids produced by the liver might be expected to be subject to reoxidation before they can be exported. Such a futile cycle is prevented by the compartmentalization of fatty acid oxidation in the mitochondrion and fatty acid synthesis in the cytosol. Carnitine palmitoyltransferase I, a component of the system that transports fatty acids into the mitochondrion (Section 25-2B), is inhibited by malonyl-CoA, the key intermediate in fatty acid biosynthesis (Section 25-4A). Hence, when the demand for metabolic fuels is low so that fatty acids are being synthesized, they cannot enter the mitochondrion for conversion to acetyl-CoA. Rather, the liver's biosynthetic demand for acetyl-CoA is met through the degradation of glucose.

When the demand for metabolic fuel rises so as to inhibit fatty acid biosynthesis, however, fatty acids are transported into the liver mitochondria for conversion to ketone bodies. Under such conditions of low blood glucose concentrations, glucokinase has reduced activity so that there is net glucose export (there is, however, always a futile cycle between the reactions catalyzed by glucokinase and glucose-6-phosphatase; Section 18-3Fb). The liver cannot use ketone bodies for its own metabolic purposes because liver cells lack 3-ketoacyl-CoA transferase (Section 25-3). Fatty acids rather than glucose or ketone bodies are therefore the liver's major acetyl-CoA source under conditions of high metabolic demand. The liver generates its ATP from this acetyl-CoA through the citric acid cycle and oxidative phosphorylation. The aerobic oxidation of fatty acids inhibits glucose utilization since activation of the citric acid cycle and oxidative phosphorylation increases the concentration of citrate, which inhibits glycolysis (the glucose-fatty acid or Randle cycle; Section 22-4Bb).

c. Amino Acids Are Important Metabolic Fuels

The liver degrades amino acids to a variety of metabolic intermediates (Section 26-3). These pathways mostly begin with amino acid transamination to yield the corresponding α -keto acid (Section 26-1A) with the amino group being

ultimately converted, via the urea cycle (Section 26-2), to the subsequently excreted urea. Glucogenic amino acids can be converted in this manner to pyruvate or citric acid cycle intermediates such as oxaloacetate and are thereby gluconeogenic precursors (Section 23-1). Ketogenic amino acids, many of which are also glucogenic, may be converted to ketone bodies.

The liver's glycogen store is insufficient to supply the body's glucose needs for more than ~ 6 h after a meal. After that, glucose is supplied through gluconeogenesis from amino acids arising mostly from muscle protein degradation to alanine (the glucose–alanine cycle; Section 26-1Ad) and glutamine (the transport form of ammonia; Section 26-1B). Thus proteins, in addition to their structural and functional roles, are important fuel resources. (Animals cannot convert fat to glucose because they lack a pathway for the net conversion of acetyl-CoA to oxaloacetate; Section 23-2).

d. The Liver Is the Body's Major Metabolic Processing Unit

The liver has numerous specialized biochemical functions in addition to those already mentioned. Prominent among them are the synthesis of blood plasma proteins, the degradation of porphyrins (Section 26-4A) and nucleic acid bases (Section 28-4), the storage of iron, and the detoxification of biologically active substances such as drugs, poisons, and hormones by a variety of oxidation (e.g., by cytochromes P450; Section 15-4Bc), reduction, hydrolysis, conjugation, and methylation reactions.

E. Kidney

The kidney functions to filter out the waste product urea from the blood and concentrate it for excretion, to recover important metabolites such as glucose, and to maintain the blood's pH. Blood pH is maintained by regenerating depleted blood buffers such as bicarbonate (lost by the exhalation of CO_2) and by removing for excretion excess H^+ together with the conjugate bases of excess metabolic acids such as the ketone bodies acetoacetate and β -hydroxybutyrate. Phosphate, the major buffer in urine for moderate acid excretion, is accompanied by equivalent quantities of cations such as Na^+ and K^+ . However, large losses of Na^+ and K^+ would upset the body's electrolyte balance, so on the production of large amounts of acids such as lactic acid or ketone bodies, the kidney produces NH_4^+ to aid in the excretion of the excess H^+ (utilizing Cl^- or the conjugate base of a metabolic acid as the counterion). This NH_4^+ is generated from glutamine, which is converted first to glutamate and then to α -ketoglutarate by glutaminase and glutamate dehydrogenase. The overall reaction is



The α -ketoglutarate is converted to malate by the citric acid cycle and then is exported from the mitochondrion and converted either to pyruvate, which is oxidized completely to CO_2 , or via oxaloacetate to PEP and then to glucose via gluconeogenesis. High fat diets, which produce

high blood concentrations of free fatty acids and ketone bodies and hence high acidic loads, cause α -ketoglutarate to be converted completely to CO_2 , and then to bicarbonate, thereby increasing the blood's buffering capacity. During starvation, the α -ketoglutarate enters gluconeogenesis, to the extent that the kidneys generate as much as 50% of the body's glucose supply.

3 METABOLIC HOMEOSTASIS: REGULATION OF APPETITE, ENERGY EXPENDITURE, AND BODY WEIGHT

When a normal animal overeats, the resulting additional fat somehow signals the brain to induce the animal to eat less and to expend more energy. Conversely, the loss of fat stimulates increased eating until the lost fat is replaced. Evidently, animals have a "lipostat" that can keep the amount of body fat constant to within 1% over many years. At least a portion of the lipostat resides in the hypothalamus (a part of the brain that hormonally controls numerous physiological functions; Section 19-1H), since damaging it can yield a grossly obese animal.

Despite this obvious set of controls in animals, there has been an explosion of obesity in many industrial nations. It has, in fact, become a world health problem, leading to diabetes and heart disease. As a result of numerous studies in recent years, researchers have been able to outline the mechanisms involved in **metabolic homeostasis**, the balance between energy influx and energy expenditure, and to identify some of the irregularities that lead to obesity. A variety of mutant strains of rodents have been generated that cause obesity. The study of these mutants has resulted in the identification of several hormones that act in a coordinated manner to regulate appetite.

A. AMP-Dependent Protein Kinase Is the Cell's Fuel Gauge

All of the metabolic pathways discussed in Section 27-1 are affected in one way or another by the need for ATP, as is indicated by the cell's AMP-to-ATP ratio (Section 17-4Fd). Several enzymes are either activated or inhibited allosterically by AMP, and several others are phosphorylated by **AMP-dependent protein kinase (AMPK)**, a major regulator of metabolic homeostasis. *AMPK activates metabolic breakdown pathways that generate ATP while inhibiting biosynthetic pathways so as to conserve ATP for more vital processes.* AMPK is an $\alpha\beta\gamma$ heterotrimer found in all eukaryotic organisms from yeast to humans. The α subunit contains a Ser/Thr protein kinase domain, and the γ subunit contains sites for allosteric activation by AMP and inhibition by ATP. Like other protein kinases, AMPK's kinase domain must be phosphorylated for activity. Binding of AMP to the γ subunit causes a conformational change that exposes Thr 172 in the activation loop of the α subunit, promoting its phosphorylation and increasing its activity at least 100-fold. AMP can activate the phosphorylated enzyme up to 5-fold more. There are two isoforms of the α

subunit, two of the β subunit, and three of the γ subunit, giving rise to 12 possible heterotrimeric combinations, with splice variants yielding further diversity. The major kinase that phosphorylates AMPK is named **LKB1**. The knockout of LKB1 in mouse liver results in the loss of the phosphorylated form of AMPK.

a. AMPK Activates Glycolysis in Ischemic Cardiac Muscle

AMPK's targets include the heart isozyme of the bi-functional enzyme PFK-2/FBPAse-2, which controls the fructose-2,6-bisphosphate (F2,6P) concentration (Section 18-3Fc). The phosphorylation of this isozyme activates the PFK-2 activity, increasing [F2,6P], which in turn activates PFK-1, the rate-determining enzyme of glycolysis (Section 17-4Fb). Consequently, in ischemic (blood-starved) heart muscle cells, which receive insufficient oxygen for oxidative phosphorylation to maintain adequate concentrations of ATP, the resulting AMP buildup causes the cells to switch to anaerobic glycolysis for ATP production.

b. AMPK Inhibits Lipogenesis, Cholesterol Synthesis, and Gluconeogenesis in Liver

AMPK-mediated phosphorylation also inhibits acetyl-CoA carboxylase (ACC), which catalyzes the first committed step of fatty acid synthesis (Section 25-4B), and hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase), which catalyzes the rate-determining step in cholesterol biosynthesis (Section 25-6Aa). Activated AMPK inhibits gluconeogenesis in a more complicated way: It phosphorylates and thereby inactivates the transcriptional coactivator **TORC2** (for *transducer of regulated CREB activity-2*), which in concert with the transcriptional activator CREB, would otherwise induce the transcription of the gene encoding PEP carboxykinase (PEPCK), the enzyme that catalyzes the rate-determining step of gluconeogenesis (Sections 23-1Af and 23-1Bb). Consequently, when the rate of ATP production is inadequate, these biosynthetic pathways are turned off, thereby conserving ATP for more vital cellular functions.

c. AMPK Promotes Fatty Acid Oxidation and Glucose Uptake but Inhibits Glycogen Synthesis in Skeletal Muscle

The inhibition of ACC results in a decrease in the concentration of malonyl-CoA, the starting material for fatty acid biosynthesis. Malonyl-CoA has an additional role, however. It is an inhibitor of carnitine palmitoyltransferase I (Section 25-2B), which is required to transfer cytosolic palmitoyl-CoA into mitochondria for oxidation. The decrease in malonyl-CoA concentration therefore allows more palmitoyl-CoA to be oxidized. AMPK also increases the recruitment of GLUT4 to muscle cell plasma membranes (Section 20-2Ec), as well as stimulating its expression, thus facilitating the insulin-independent entry of glucose into these cells. In addition, AMPK inhibits glycogen synthase (which catalyzes the rate-limiting reaction in glycogen synthesis; Section 18-3B). In fact, the β subunit of AMPK has a glycogen-binding domain that presumably recruits AMPK to the vicinity of glycogen synthase.

d. AMPK Inhibits Fatty Acid Synthesis and Lipolysis in Adipocytes

AMPK inhibits fatty acid synthesis in adipocytes by phosphorylating ACC as described above. Moreover, AMPK phosphorylates hormone-sensitive triacylglycerol lipase in adipose tissue (Section 25-5). This phosphorylation inhibits rather than activates the enzyme, in part by preventing the relocation of the enzyme to the lipid droplet, the cellular location of lipolysis. As a result, fewer triacylglycerol molecules are broken down so that fewer fatty acids are exported to the bloodstream. This latter process seems paradoxical (fatty acid oxidation would help relieve an ATP deficit), although it has been speculated that it prevents the cellular buildup of fatty acids to toxic levels. The major effects of AMPK activation on glucose and lipid metabolism in liver, skeletal muscle, heart muscle, and adipose tissue are diagrammed in Fig. 27-4.

B. Adiponectin Regulates AMPK Activity

Adiponectin is a 247-residue protein hormone, secreted exclusively by adipocytes, that helps regulate energy homeostasis and glucose and lipid metabolism by controlling AMPK activity. Its monomers consist of an N-terminal collagenlike domain and a C-terminal globular domain. Adiponectin occurs in the bloodstream in several forms: a low molecular weight (LMW) trimer formed by the coiling of its collagenlike domains into a triple helix (Section 8-2Ba) as well as hexamers (MMW) and multimers (HMW) that form disulfide cross-linked bouquets (Fig. 27-5). In addition, globular adiponectin, formed by the cleavage of the collagenlike domain to release globular monomers, occurs in lower concentrations.

The binding of adiponectin to **adiponectin receptors**, which occur on the surfaces of both liver and muscle cells, acts to increase the phosphorylation and activity of AMPK. This, as we have seen (Section 27-3A), inhibits gluconeogenesis and stimulates fatty acid oxidation in liver and stimulates glucose uptake and glucose and fatty acid oxidation in muscle. All of these effects act to increase insulin sensitivity, in part because adiponectin and insulin elicit similar responses in tissues such as liver. Decreased adiponectin is associated with insulin resistance (Section 27-4B). Paradoxically, the blood concentration of adiponectin, which is secreted by adipocytes, decreases with increased amounts of adipose tissue. This may be because increased adipose tissue is also associated with increased production of **tumor necrosis factor- α (TNF- α)**, a cytokine that decreases both the expression and secretion of adiponectin from adipose tissue (Section 19-3Db).

C. Leptin

Two of the genes whose mutations cause obesity in mice are known as *obese* (*ob*) and *diabetes* (*db*; the wild-type genes are designated *OB* and *DB*). Homozygotes for defects in either of these recessive genes, *ob/ob* or *db/db*, are grossly obese and have nearly identical phenotypes (Fig. 27-6). Indeed, the way in which these phenotypes

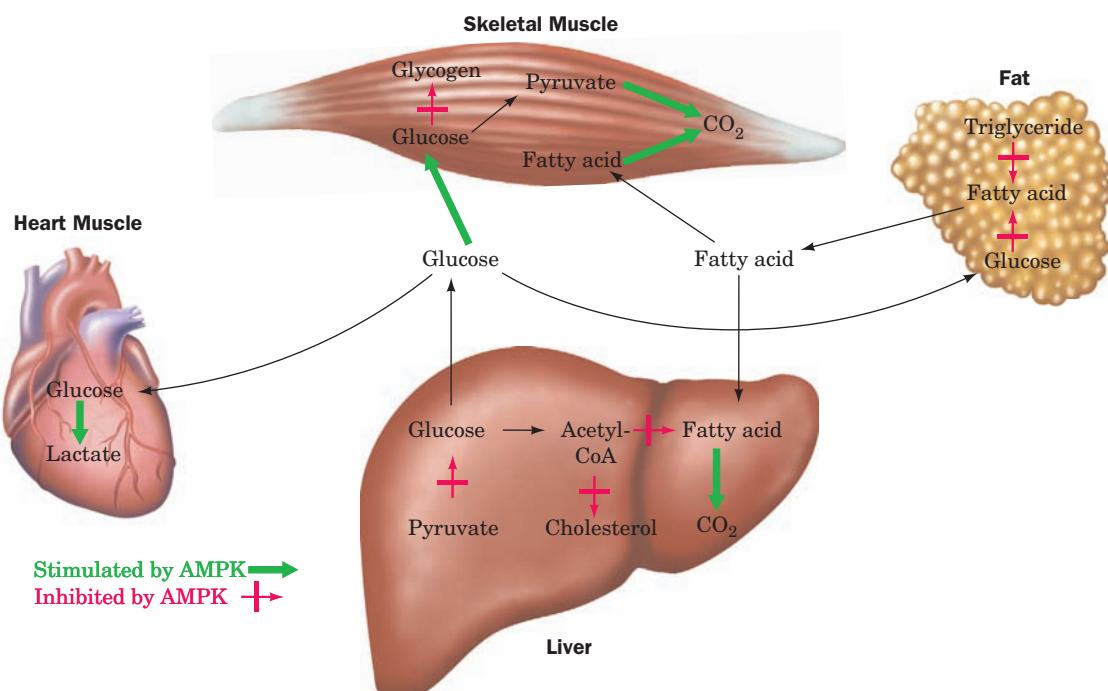


Figure 27-4 Major effects of AMP-activated protein kinase (AMPK) on glucose and lipid metabolism in liver, muscle, and adipose tissue. In skeletal muscle, AMPK stimulates glucose and fatty acid oxidation while inhibiting glycogen synthesis. In heart muscle, AMPK stimulates glycolysis. In liver, AMPK inhibits

were distinguished was by surgically linking the circulation of a mutant mouse to that of a normal (*OB/OB*) mouse, a phenomenon named **parabiosis**. *ob/ob* mice so linked

lipid biosynthesis and gluconeogenesis while activating fatty acid oxidation. In adipose tissue, AMPK inhibits fatty acid biosynthesis, lipolysis, and fatty acid export. [After Towler, M.C. and Hardie, D.G., *Circ. Res.* **100**, 328 (2007).]

exhibit normalization of body weight and reduced food intake, whereas *db/db* mice do not do so. This suggests that *ob/ob* mice are deficient in a circulating factor that regulates appetite and metabolism, whereas *db/db* mice are defective in the receptor for this circulating factor.

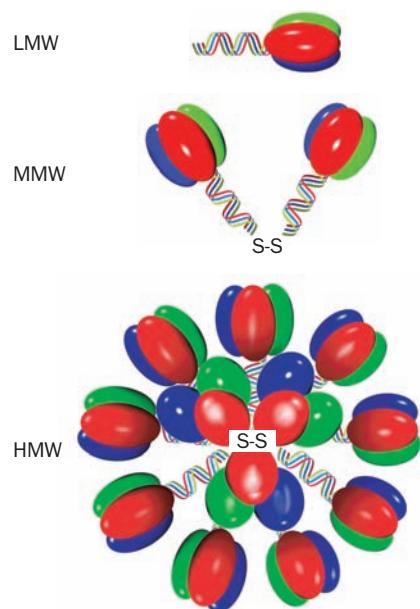


Figure 27-5 Adiponectin trimers, hexamers, and multimers. These complexes are referred to as low molecular weight (LMW), medium molecular weight (MMW), and high molecular weight (HMW) forms. [After Kadowaki, T., and Yamauchi, T., *Endocr. Rev.* **26**, 439 (2005).]



Figure 27-6 Normal (*OB/OB*, left) and obese (*ob/ob*, right) mice. [Courtesy of Richard D. Palmiter, University of Washington.]

The mouse *OB* gene encodes a 146-residue monomeric protein named **leptin** (Greek: *leptos*, thin; Fig. 27-7) that has no apparent homology with proteins of known sequence. Leptin, which was discovered by Jeffrey Friedman, is expressed only by adipocytes, which in doing so appear to inform the brain of how much fat the body carries. Thus, injecting leptin into *ob/ob* mice causes them to eat less and to lose weight. In fact, leptin-treated *ob/ob* mice on a restricted diet lost 50% more weight than untreated *ob/ob* mice on the same diet, which suggests that reduced food intake alone is insufficient to account for leptin-induced weight loss. Leptin appears to control energy expenditure as well.

Leptin injection has no effect on *db/db* mice. The leptin receptor gene was identified by making a cDNA library from mouse brain tissue that specifically bound leptin and then identifying a receptor-expressing clone by its ability to bind leptin (gene cloning techniques are discussed in Section 5-5). This gene, which has been shown to be the *DB* gene, encodes a protein named **OB-R** (for *OB* receptor) that appears to have a single transmembrane segment and an extracellular domain that resembles the receptors for certain cytokines (proteins that regulate the differentiation, proliferation, and activities of various blood cells; Section 19-3Eb).

OB-R protein, which was discovered by Louis Tartaglia, has at least six alternatively spliced forms that appear to be expressed in a tissue-specific manner (alternative gene splicing is discussed in Section 31-4Am). In normal mice, the hypothalamus expresses high levels of a splice variant of OB-R that has a 302-residue cytoplasmic segment. However, in *db/db* mice, this segment has an abnormal splice site that truncates it to only 34 residues, which almost certainly renders this OB-R variant unable to transmit leptin signals. Thus, it appears that leptin's weight-controlling effects are mediated by signal transduction resulting from its binding to the OB-R protein in the hypothalamus.

Human leptin is 84% identical in sequence to that of mice. The use of a radioimmunoassay (Section 19-1A) to measure the serum levels of leptin in normal-weight and obese humans established that in both groups serum leptin concentrations increase with their percentage of body fat as does the *ob* mRNA content of their adipocytes. Moreover, after obese individuals had lost weight, their serum leptin concentrations and adipocyte *ob* mRNA content declined. This suggests that most obese persons produce sufficient amounts of leptin but have developed "leptin resistance." Since leptin must cross the blood-brain barrier in order to exert its effects on the hypothalamus, it has been suggested that this crossing is somehow saturable, thus limiting the concentration of leptin in the brain. The high concentration of leptin in obese individuals is not without affect, however. OB-R is also expressed in peripheral tissues where leptin has been shown to function as well. While not preventing obesity, the hormone has been shown to directly stimulate the oxidation of fatty acids as well as to inhibit the accumulation of lipids in non-adipose tissue. It does so by activating AMP-dependent protein kinase (AMPK), which in turn phosphorylates and thereby inactivates

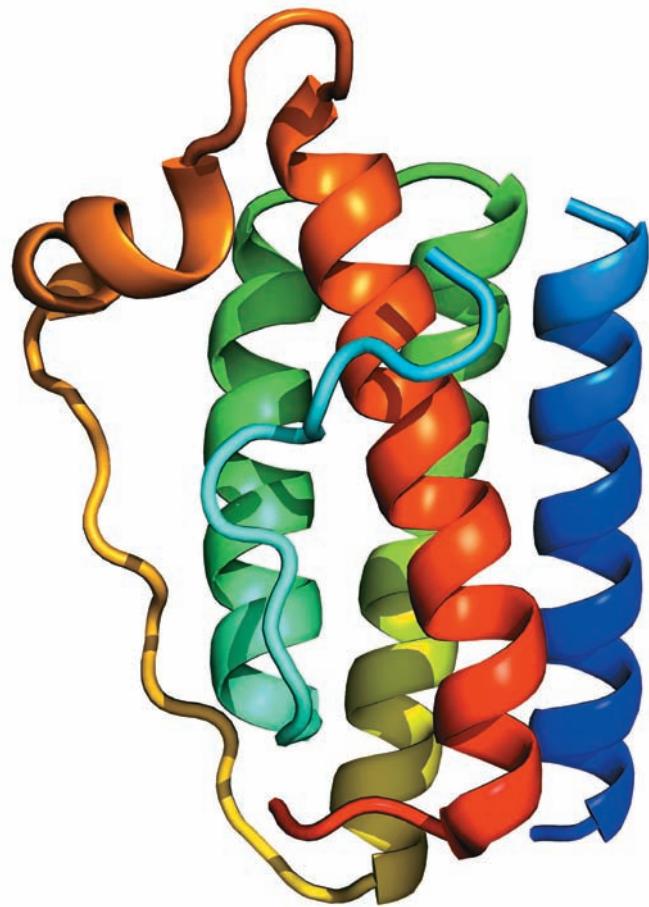


Figure 27-7 X-ray structure of human leptin-E100. This mutant form of leptin (W100E) has comparable biological activity to the wild-type protein but crystallizes more readily. The protein, which is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), forms a four-helix bundle, as do many protein growth factors (e.g., human growth hormone; Fig. 19-10). Residues 25 to 38 are not visible in the X-ray structure. [Based on an X-ray structure by Faming Zhang, Eli Lilly & Co., Indianapolis, Indiana. PDBid 1AX8.]

See Interactive Exercise 27.

vates acetyl-CoA carboxylase (ACC). This reduces the malonyl-CoA concentration, thereby decreasing its inhibition of carnitine palmitoyltransferase I, which then transports fatty acyl-CoA into the mitochondrion for oxidation (Section 25-5). We discuss the function of leptin in peripheral tissues in Section 27-3H.

A small minority of obese individuals have been found to be leptin deficient in a manner similar to *ob/ob* mice. Two grossly obese children who are members of the same highly consanguineous (descended from the same ancestors) family (they are cousins and both sets of parents are cousins) have been shown to be homozygous for a defective OB gene. The children, at the ages of 8 and 2 years old, respectively, weighed 86 and 29 kg and were noted to have remarkably large appetites. Their OB genes have a deletion of a single guanine nucleotide in codon 133, thereby causing a frameshift mutation that, it is likely, renders the

mutant leptin biologically inactive. Moreover, their leptin serum levels were only ~10% of normal. Leptin injections have relieved their symptoms.

D. Insulin

We have discussed the insulin signaling cascade (Section 19-4F) and the role of insulin in peripheral tissues such as muscle and adipose tissue in stimulating the uptake of glucose (Fig. 20-11) and its storage as glycogen (Section 18-3) or fat (Section 25-5). Insulin receptors also occur in the hypothalamus. Consequently, the infusion of insulin into rats with insulin-deficient diabetes inhibits food intake, reversing the overeating behavior characteristic of the disease. Knock-out mice have been developed with a central nervous system-specific disruption of the insulin receptor gene. These mice have no alteration in brain development or survival but become obese, with increased body fat, increased leptin levels, increased serum triacylglycerol, and the elevated plasma insulin levels characteristic of insulin resistance (Section 27-4B). Evidently, insulin also plays a role in the neuronal regulation of food intake and body weight. As we discuss in Section 27-3F, insulin and leptin both act through receptors in the hypothalamus to decrease food intake.

E. Ghrelin and PYY₃₋₃₆

a. Ghrelin and PYY₃₋₃₆ Act as Short-Term Regulators of Appetite

Ghrelin, which was discovered by Masayasu Kojima and Kenji Kanagawa, is an appetite-stimulating gastric peptide that is secreted by the empty stomach. This 28-residue peptide was first discovered and named for its function as a growth hormone-releasing peptide (ghrelin is an abbreviation for growth-hormone-release). Octanoylation of its Ser 3 is required for activity.



Human ghrelin

X = Ser modified with *n*-octanoic acid

Injection of ghrelin has been shown to induce adiposity (increased adipose tissue) in rodents by stimulating an increase in food intake while reducing fat utilization. In humans in states of positive energy balance such as obesity or high caloric intake, circulating ghrelin levels are decreased, whereas during fasting, circulating ghrelin levels increase.

PYY₃₋₃₆



Human PYY₃₋₃₆

is a peptide secreted by the gastrointestinal tract in proportion to the caloric intake of a meal, which acts to inhibit further food intake. Both rodents and humans have been shown to respond to the presence of this peptide by decreasing their food intake for up to 12 hours. Human subjects receiving a 90-minute infusion of PYY₃₋₃₆ ate only 1500 kcal of

food during the next 24-hour period, whereas those receiving saline controls ate 2200 kcal during the same period.

F. Hypothalamic Integration of Hormonal Signals

a. Neurons of the Arcuate Nucleus Region of the Hypothalamus Integrate and Transmit Hunger Signals

About half of the length of the hypothalamus is taken up by the **arcuate nucleus**, a collection of neuronal cell bodies consisting of two cell types: the **NPY/AgRP** cell type and the **POMC/CART** cell type. These cell types are named after the neuropeptides they secrete. **Neuropeptide Y (NPY)**



Neuropeptide Y

The C-terminal carboxyl is amidated

is a potent stimulator of food intake and an inhibitor of energy expenditure, as is **Agouti related peptide (AgRP)**. **Pro-opiomelanocortin (POMC)** is post-translationally processed in the hypothalamus to release **α -melanocyte stimulating hormone (α -MSH**; Section 34-3C). **Cocaine and amphetamine-regulated transcript (CART)** and α -MSH are both inhibitors of food intake and stimulators of energy expenditure.

The balance of the secretions from these two cell types is controlled by leptin, insulin, ghrelin, and PYY₃₋₃₆ (Fig. 27-8). Leptin and insulin signal satiety and therefore decrease appetite by diffusing across the blood-brain barrier to the arcuate nucleus, where they stimulate POMC/CART neurons to produce CART and α -MSH, while inhibiting the production of NPY from NPY/AgRP neurons. Leptin receptors act through the JAK-STAT signal transduction pathway (Section 19-3E). Ghrelin has receptors on NPY/AgRP neurons that stimulate the secretion of NPY and AgRP to increase appetite. Interestingly, PYY₃₋₃₆, a peptide that is homologous to NPY, binds specifically to NPY receptor subtype Y2R on NPY/AgRP neurons. This subtype is an inhibitory receptor, however, so binding of PYY₃₋₃₆ causes a decrease in secretion from NPY/AgRP neurons. The integrated stimuli of all these secretions from the arcuate nucleus control appetite.

G. Control of Energy Expenditure by Adaptive Thermogenesis

The energy content of food is utilized by an organism either in the performance of work or the generation of heat. Excess energy is stored as glycogen or fat for future use. In well-balanced individuals, the storage of excess fuel remains constant over many years. However, when energy consumed is consistently greater than energy expended, obesity results. The body has several mechanisms for preventing obesity. One of them, as discussed above, is appetite control. The other is **diet-induced thermogenesis**, a form of **adaptive thermogenesis** (heat production in

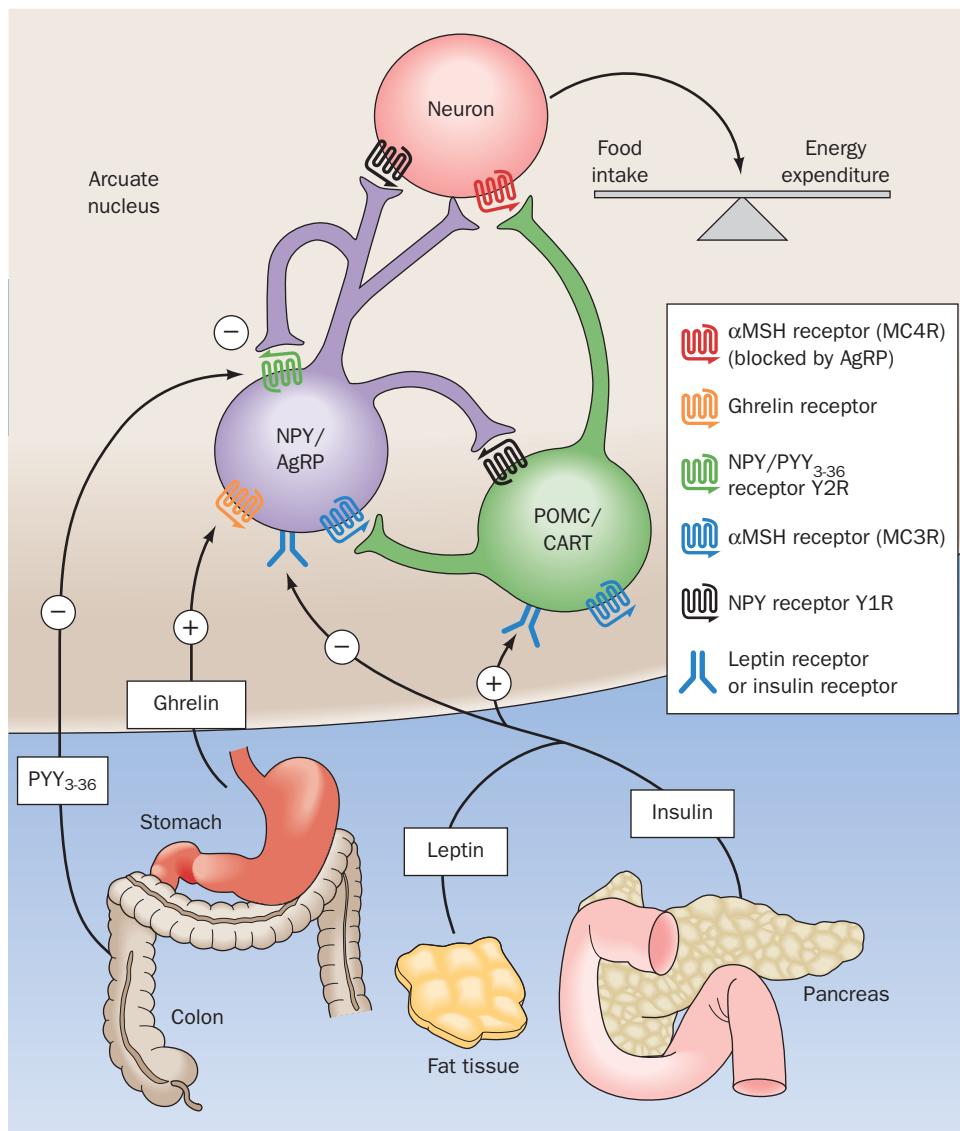


Figure 27-8 Hormones that control the appetite. Leptin and insulin (bottom) circulate in the blood at concentrations proportional to body-fat mass. They decrease appetite by inhibiting NPY/AgRP neurons (center) while stimulating melanocortin-producing neurons in the arcuate nucleus region of the hypothalamus. NPY and AgRP increase the appetite, and melanocortins decrease the appetite, via other neurons (top). Activation of NPY/AgRP-expressing neurons inhibits

melanocortin-producing neurons. The gastric hormone ghrelin stimulates appetite by activating the NPY/AgRP-expressing neurons. PYY₃₋₃₆, released from the gastrointestinal tract, inhibits NPY/AgRP-expressing neurons and thereby decreases the appetite. PYY₃₋₃₆ works in part through the autoinhibitory NPY receptor subtype Y2R. [After Schwartz, M.W. and Morton, G.J., *Nature* **418**, 596 (2002).]

response to environmental stress). We have previously discussed adaptive thermogenesis in response to cold, which occurs in rodents and newborn humans through the uncoupling of oxidative phosphorylation in brown adipose tissue (Section 22-3Da). The mechanism of this thermogenesis involves the release of norepinephrine from the brain in response to cold, its binding to β -adrenergic receptors on brown adipose tissue inducing an increase in [cAMP], which in turn initiates an enzymatic phosphorylation cascade that activates hormone-sensitive triacylglycerol lipase. The resulting increase in the concentration of free fatty acids provides fuel for oxidation as well as inducing

the opening of a proton channel, called uncoupling protein-1 (UCP1) or thermogenin, in the inner mitochondrial membrane. The opening of UCP1 discharges the proton gradient across the inner mitochondrial membrane, thus uncoupling electron transport from ATP production. The energy that would otherwise have been used to drive ATP synthesis is thereby released as heat.

Although metabolic measurements in adult humans clearly demonstrate that an increase in energy intake causes an increase in metabolic rate and thermogenesis, the cause of this increase is unclear. Adult humans have little brown adipose tissue. However, skeletal muscle represents

up to 40% of their total body weight and has high mitochondrial capacity. Homologs of UCP1 have been identified: **UCP2** occurs in many tissues including white adipose tissue, whereas **UCP3** occurs in brown adipose tissue, white adipose tissue, and muscle. Leptin has been shown to up-regulate UCP2. However, it has yet to be demonstrated that UCP3 in muscle participates in diet-induced thermogenesis. ATP-hydrolyzing substrate cycles such as that between fatty acids and triacylglycerol in adipose tissue (Section 27-2C) may also be involved.

H. Did Leptin Evolve as a Thrifty Gene?

The unusual behavior of leptin, which serves to control weight in normal-weight individuals while its concentration continues to climb without apparent effect in obese individuals, has led to the proposal that leptin evolved as a “thrifty gene.” In hunter-gatherer societies, it was a distinct advantage to be able to survive intermittent famines. In order to do this, fat must be stored in adipose tissue in times of plenty, making short-term obesity advantageous. However, the accumulation of fatty acids and lipids in non-adipose tissue results in coronary artery disease, insulin resistance, and diabetes (Section 27-4B). Leptin, by directly stimulating the oxidation of fatty acids as well as inhibiting the accumulation of lipids in non-adipose tissue, is thought to protect against these diseases during short-term obesity, thereby providing an evolutionary advantage. However, in recent times in industrialized nations, the unprecedented availability of food and lack of famine has made obesity a long-term rather than a short-term condition, which is a liability rather than a benefit.

4 METABOLIC ADAPTATION

In this section we consider the body’s responses to two metabolically abnormal situations: (1) starvation and (2) the disease diabetes mellitus.

A. Starvation

Glucose is the metabolite of choice of both brain and working muscle. Yet, the body stores less than a day’s supply of carbohydrate (Table 27-1). Thus, the low blood sugar caused by even an overnight fast results, through an increase in glucagon secretion and a decrease in insulin secretion, in the mobilization of fatty acids from adipose tissue (Section 25-5). The diminished insulin level also inhibits glucose uptake by muscle tissue. Muscles therefore switch from glucose to fatty acid metabolism for energy production. The brain, however, still remains heavily dependent on glucose.

In animals, glucose cannot be synthesized from fatty acids. This is because neither pyruvate nor oxaloacetate, the precursors of glucose in gluconeogenesis (Section 23-1), can be synthesized from acetyl-CoA (the oxaloacetate in the citric acid cycle is derived from acetyl-CoA but the cyclic nature of this process requires that the oxaloacetate be consumed as fast as it is synthesized; Section 21-1A).

Table 27-1 Fuel Reserves for a Normal 70-kg Man

Fuel	Mass (kg)	Calories ^a
<i>Tissues</i>		
Fat (adipose triacylglycerols)	15	141,000
Protein (mainly muscle)	6	24,000
Glycogen (muscle)	0.150	600
Glycogen (liver)	0.075	300
<i>Circulating fuels</i>		
Glucose (extracellular fluid)	0.020	80
Free fatty acids (plasma)	0.0003	3
Triacylglycerols (plasma)	0.003	30
<i>Total</i>		166,000

^aOne (dieter’s) Calorie = 1 kcal = 4.184 kJ.

Source: Cahill, G.F., Jr., *New Engl. J. Med.* **282**, 669 (1970).

During starvation, glucose must therefore be synthesized from the glycerol product of triacylglycerol breakdown and, more importantly, from the amino acids derived from the proteolytic degradation of proteins, the major source of which is muscle. Thus, after a 40-hour fast, gluconeogenesis supplies ~96% of the glucose produced by the liver. However, the continued breakdown of muscle during prolonged starvation would ensure that this process became irreversible since a large muscle mass is essential for an animal to move about in search of food. The organism must therefore make alternate metabolic arrangements.

After several days of starvation, gluconeogenesis has so depleted the liver’s oxaloacetate supply that this organ’s ability to metabolize acetyl-CoA via the citric acid cycle is greatly diminished. Rather, the liver converts the acetyl-CoA to ketone bodies (Section 25-3), which it releases into the blood. The brain gradually adapts to using ketone bodies as fuel through the synthesis of the appropriate enzymes: After a 3-day fast, only about one-third of the brain’s energy requirements are satisfied by ketone bodies but after 40 days of starvation, ~70% of its energy needs are so met. The rate of muscle breakdown during prolonged starvation consequently decreases to ~25% of its rate after a several-day fast. The survival time of a starving individual is therefore much more dependent on the size of his or her fat reserves than it is on his or her muscle mass. Indeed, highly obese individuals can survive for over a year without eating (and have occasionally done so in clinically supervised weight reduction programs).

a. Caloric Restriction May Increase Longevity

Caloric restriction is a modified form of starvation whereby energy intake is reduced 30–40%, while micronutrient (vitamin and mineral) levels are maintained. Rodents kept on such a diet live up to 50% longer than rodents on normal diets and exhibit fewer of the debilitating symptoms of old age. The life spans of a large range of organisms from yeast to primates are similarly extended. Considerable research effort is being expended to determine the biochemical basis of these observations.

B. Diabetes Mellitus

The polypeptide hormone insulin acts mainly on muscle, liver, and adipose tissue cells to stimulate the synthesis of glycogen, fats, and proteins while inhibiting the breakdown of these metabolic fuels. In addition, insulin stimulates the uptake of glucose by most cells, with the notable exception of brain and liver cells. Together with glucagon, which has largely opposite effects, insulin acts to maintain the proper level of blood glucose.

In the disease **diabetes mellitus**, which is the third leading cause of death in the United States after heart disease and cancer, insulin either is not secreted in sufficient amounts or does not efficiently stimulate its target cells. As a consequence, blood glucose levels become so elevated that the glucose “spills over” into the urine, providing a convenient diagnostic test for the disease. Yet, despite these high blood glucose levels, cells “starve” since insulin-stimulated glucose entry into cells is impaired. Triacylglycerol hydrolysis, fatty acid oxidation, gluconeogenesis, and ketone body formation are accelerated and, in a condition termed **ketoacidosis**, ketone body levels in the blood become abnormally high. Since ketone bodies are acids, their high concentration puts a strain on the buffering capacity of the blood and on the kidney, which controls blood pH by excreting the excess H^+ into the urine (Section 27-2E). This unusually high excess H^+ excretion is accompanied by NH_4^+ , Na^+ , K^+ , P_i , and H_2O excretion, causing severe dehydration (which compounds the dehydration resulting from the osmotic effect of the high glucose concentration in the blood; excessive thirst is a classic symptom of diabetes) and a decrease in blood volume—ultimately life-threatening situations.

There are two major forms of diabetes mellitus:

1. Insulin-dependent, type 1, or juvenile-onset diabetes mellitus, which most often strikes suddenly in childhood.

2. Noninsulin-dependent, type 2, or maturity-onset diabetes mellitus, which usually develops rather gradually after the age of 40.

a. Insulin-Dependent Diabetes Is Caused by a Deficiency of Pancreatic β Cells

In insulin-dependent (type 1) diabetes mellitus, insulin is absent or nearly so because the pancreas lacks or has defective β cells. This condition results, in genetically susceptible individuals (see below), from an autoimmune response that selectively destroys their β cells. Individuals with insulin-dependent diabetes, as Frederick Banting and George Best first demonstrated in 1921, require daily insulin injections to survive and must follow carefully balanced diet and exercise regimens. Their life spans are, nevertheless, reduced by up to one-third as a result of degenerative complications such as kidney malfunction, nerve impairment, and cardiovascular disease, which apparently arise from the imprecise metabolic control provided by periodic insulin injections. The **hyperglycemia** (high blood [glucose]) of diabetes mellitus also leads to blindness through retinal degeneration and the glucosylation of lens proteins, which causes cataracts (Fig. 27-9).

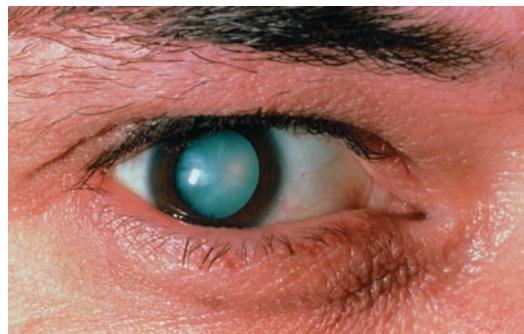


Figure 27-9 Photo of a diabetic cataract. The accumulation of glucose in the lens leads to swelling and precipitation of lens proteins. The resulting opacification causes blurred vision and ultimately complete loss of sight. [© Sue Ford/Photo Researchers.]

Perhaps newly developed systems that monitor blood glucose levels and continuously deliver insulin in the required amounts will rectify this situation.

The usually rapid onset of the symptoms of insulin-dependent diabetes had suggested that the autoimmune attack on the pancreatic β cells responsible for this disease is one of short duration. Typically, however, the disease “brews” for several years as the aberrantly aroused immune system slowly destroys the β cells. Only when >80% of these cells have been eliminated do the classic symptoms of diabetes suddenly emerge. Consequently, one of the most successful treatments for insulin-dependent diabetes is a β -cell transplant, a procedure that became possible with the development of relatively benign immunosuppressive drugs.

Why does the immune system attack the pancreatic β cells? It has long been known that certain alleles (genetic variants) of the **Class II major histocompatibility complex (MHC) proteins** are particularly common in insulin-dependent diabetics [MHC proteins are highly polymorphic (variable within a species) immune system components to which cell-generated antigens such as viral proteins must bind in order to be recognized as foreign; Sections 35-2Aa and 35-2E]. It is thought that autoimmunity against β cells is induced in a susceptible individual by a foreign antigen, perhaps a virus, which immunologically resembles some β cell component. The Class II MHC protein that binds this antigen does so with such tenacity that it stimulates the immune system to launch an unusually vigorous and prolonged attack on the antigen. Some of the activated immune system cells eventually make their way to the pancreas, where they initiate an attack on the β cells due to the close resemblance of the β cell component to the foreign antigen.

b. Noninsulin-Dependent Diabetes Is Characterized by Insulin Resistance as Well as Impaired Insulin Secretion

Noninsulin-dependent (type 2) diabetes mellitus (**NIDDM**), which accounts for over 90% of the diagnosed cases of diabetes and affects 18% of the population over 65 years of age, usually occurs in obese individuals with a genetic predisposition for this condition (although one that differs from that associated with insulin-dependent

diabetes). These individuals may have normal or even greatly elevated insulin levels. Their symptoms arise from **insulin resistance**, an apparent lack of sensitivity to insulin in normally insulin-responsive cells.

The hyperglycemia that accompanies insulin resistance induces the pancreatic β cells to increase their production of insulin. Yet the high basal level of insulin secretion diminishes the ability of the β cells to respond to further increases in blood glucose. Consequently, the hyperglycemia and its attendant complications tend to worsen over time.

A small percentage of cases of type II diabetes result from mutations in the insulin receptor that affect its insulin-binding ability or tyrosine kinase activity. However, a clear genetic cause has not been identified in the vast majority of cases. It is therefore likely that many factors play a role in the development of this disease. For example, the increased insulin production resulting from overeating may eventually suppress the synthesis of insulin receptors. This hypothesis accounts for the observation that diet alone is often sufficient to control this type of diabetes.

Insulin resistance, which may precede NIDDM by as much as 10 to 20 years, appears to be caused by an interruption in the insulin signaling pathway (Section 19-4F). Gerald Shulman has proposed that this interruption is caused by a Ser/Thr kinase cascade that phosphorylates

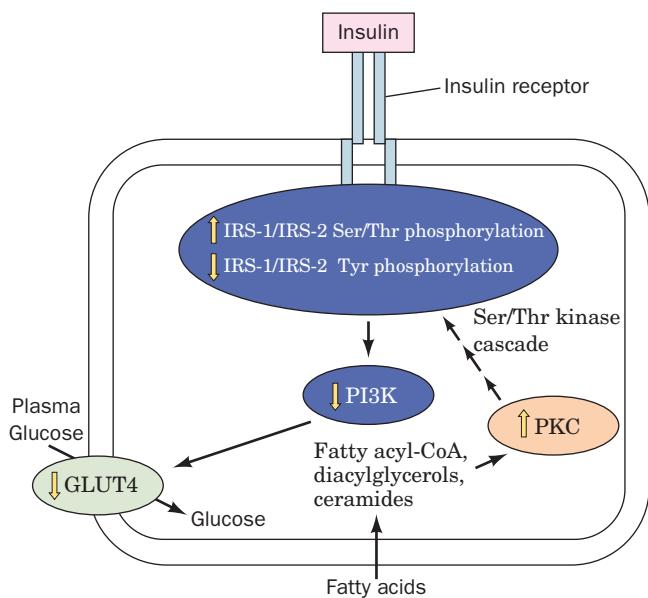


Figure 27-10 The mechanism through which high concentrations of free fatty acids cause insulin resistance. Elevated concentrations of free fatty acids in the blood diffuse into muscle cells where they are converted to fatty acyl-CoA, diacylglycerols, and ceramides. These lipotoxic substances activate an isoform of protein kinase C (PKC), triggering a Ser/Thr kinase cascade that results in the phosphorylation of IRS-1 and IRS-2. This phosphorylation inhibits the Tyr phosphorylation required for transmission of the insulin signal, thereby decreasing the activation of PI3K, which decreases the rate of fusion of GLUT4-containing vesicles with the plasma membrane and hence the amount of glucose entering the cell. [Modified from Shulman, G.I., *J. Clin. Invest.* **106**, 173 (2000).]

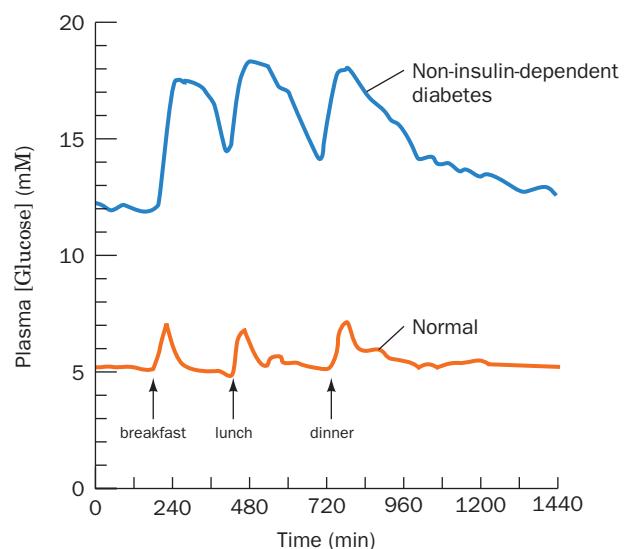
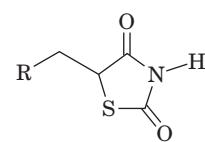
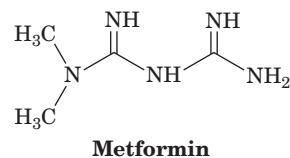


Figure 27-11 Twenty-four-hour plasma glucose profiles in normal and noninsulin-dependent diabetic subjects. The basal level of glucose and the peaks following meals are higher in the diabetic individuals. [After Bell, G.I., Pilkis, S.J., Weber, I.T., and Polonsky, K.S., *Annu. Rev. Physiol.* **58**, 178 (1996).]

proteins known as **insulin receptor substrates (IRSs; Section 19-3Cg)** so as to decrease their ability to be phosphorylated on their Tyr residues by activated insulin receptor. Tyrosine phosphorylation is required for IRS activation and communication with phosphoinositide 3-kinase (PI3K; Section 19-4D), which subsequently activates the translocation of GLUT4-containing vesicles to the cell surface for increased glucose transport into cells (Section 20-2Ec). The original Ser/Thr kinase cascade is triggered by the activation of an isoform of protein kinase C (PKC; Section 19-4C) caused by an increase in fatty acyl-CoA, diacylglycerol, and ceramides (Section 12-1D) resulting from elevated free fatty acids (Fig. 27-10). The failure to activate IRSs decreases the cell's response to insulin (Fig. 27-11).

c. Substances That Activate AMPK Attenuate the Symptoms of Noninsulin-Dependent Diabetes

Other treatments for noninsulin-dependent diabetes are drugs such as **metformin** and the **thiazolidinediones (TZDs)**,



A thiazolidinedione (TZD)

which decrease insulin resistance by either suppressing glucose release by the liver (metformin) or promoting

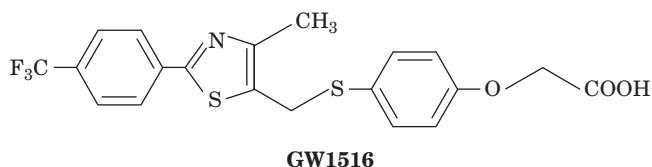
insulin-stimulated glucose disposal in muscle (TZDs). These drugs act by increasing AMPK activity but by different mechanisms. TZDs cause a large increase in the AMP to ATP ratio in muscle cells, with the expected concomitant increase in AMPK phosphorylation and activity. Metformin, however, stimulates LKB1 to phosphorylate and hence activate AMPK (LKB1 knockout mice are insensitive to metformin). In both cases, the increase in AMPK activity decreases gluconeogenesis in liver and increases glucose utilization in muscle (Fig. 27-4). In addition, the TZDs decrease insulin resistance by binding to and activating a transcription factor known as a **peroxisome proliferator-activated receptor γ (PPAR γ)**, primarily in adipose tissue. Among other things, PPAR γ activation induces the synthesis of adiponectin (Section 27-3B), which leads to an increase in AMPK activity. In adipose tissue, AMPK action leads to a decrease in lipolysis and fatty acid export, decreasing the concentration of free fatty acids in the blood and therefore decreasing insulin resistance (see above).

Intriguingly, Ronald Evans has shown that transgenic mice expressing an activated form of PPAR γ in their skeletal muscles can run around twice the distance of wild-type mice and are resistant to weight gain, even on a high fat diet. This activated PPAR γ induces an increase in the number of the aerobic and hence fatty acid-oxidizing slow-twitch (Type I) muscle fibers (Section 17-3Ca) relative to the largely anaerobic and hence less energy-efficient fast-twitch (Type II) muscle fibers.

Rodent adipocytes secrete a 108-residue polypeptide hormone called **resistin**. The hormone is named for its ability to block the action of insulin on adipocytes. In fact, resistin production is decreased by TZDs, a phenomenon that led to the discovery of resistin. Overproduction of resistin was proposed to contribute to the development of noninsulin-dependent diabetes. An interesting difference between rodents and humans is that in humans, resistin is produced by macrophages, a divergence whose evolutionary and functional implications are unclear.

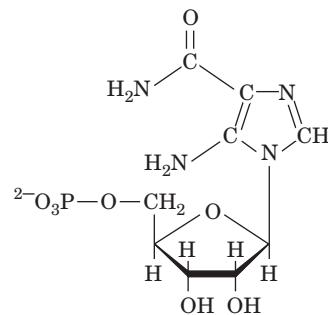
d. Obesity Is a Contributing Factor in Metabolic Syndrome

Metabolic syndrome is a disturbance in metabolism characterized by insulin resistance, inflammation, and a predisposition to several disorders including type 2 diabetes, hypertension, and atherosclerosis. These disorders are accompanied by an increase in coronary heart disease. Obesity, physical inactivity, and possibly genetic determinants have been implicated in its occurrence, which affects as many as 65 million people in the United States alone. Exercise, calorie/weight reduction, adiponectin, leptin, metformin, and TZDs have all been successfully used to treat metabolic syndrome. Similarly, the PPAR γ agonist known as **GW1516**



alleviates the symptoms of metabolic syndrome in obese men, probably by stimulating fatty acid oxidation.

Evans has shown that GW1516 greatly increases exercise endurance in mice, particularly when it is administered together with the AMPK agonist **5-aminoimidazole-4-carboxamide ribotide [AICAR]**:



5-Aminoimidazole-4-carboxamide ribotide (AICAR)

which is also a product of histidine biosynthesis (Section 26-5Be) and an intermediate in purine ribonucleotide biosynthesis (Section 28-1A)]. This treatment mimics the effects of the expression of activated PPAR γ , which suggests that the administration of GW1516 and AICAR can confer some of the benefits of exercise without actually exercising. Indeed, the World Anti-Doping Agency has placed these compounds on the list of performance-enhancing drugs that athletes are forbidden from taking.

e. DNA Chip Technology Permits the Integrated Study of Metabolic Regulation

Our ability to understand the integrated nature of metabolism and its genetic regulation in health and disease has taken a giant step forward with the advent of DNA chips (microarrays; Section 7-6B). For example, Ronald Kahn has used this technology to study the genetic basis of the metabolic abnormalities underlying both obesity and diabetes. To do so, he isolated the mRNA from the skeletal muscle of normal, diabetic, and insulin-treated diabetic mice and reverse-transcribed it to cDNA (Section 5-5Fa), which was then hybridized to oligonucleotide microarrays that represented 14,288 mouse genes. Thereby, 129 up-regulated and 106 down-regulated genes were identified in diabetic mice. Not surprisingly, the expression of the mRNAs encoding enzymes of the fatty acid β -oxidation pathway were increased, whereas those for GLUT4, glucokinase, the E1 component of the pyruvate dehydrogenase multienzyme complex, and the subunits of all four mitochondrial electron-transport chain complexes were coordinately decreased. Intriguingly, only about half of these changes in gene expression could be reversed by insulin treatment. Thus, the post-genomic era will almost certainly witness an explosion in our knowledge of metabolic regulation that should yield major health benefits. Nevertheless, our ability to sensibly interpret this huge influx of information may prove to be the greatest challenge.

CHAPTER SUMMARY

1 Major Pathways and Strategies of Energy Metabolism:

A Summary The complex network of processes involved in energy metabolism are distributed among different compartments within cells and in different organs of the body. These processes function to generate ATP “on demand,” to generate and store glucose, triacylglycerols, and proteins in times of plenty for use when needed, and to keep the concentration of glucose in the blood at the proper level for use by organs such as the brain, whose sole fuel source, under normal conditions, is glucose. The major energy metabolism pathways include glycolysis, glycogen degradation and synthesis, gluconeogenesis, the pentose phosphate pathway, and triacylglycerol and fatty acid synthesis, which are cytosolically based, and fatty acid oxidation, the citric acid cycle, and oxidative phosphorylation, which are confined to the mitochondrion. Amino acid degradation occurs, in part, in both compartments. The mediated membrane transport of metabolites therefore also plays an essential metabolic role.

2 Organ Specialization The brain normally consumes large amounts of glucose. Muscle, under intense ATP demand such as in sprinting, degrades glucose and glycogen anaerobically, thereby producing lactate, which is exported via the blood to the liver for reconversion to glucose through gluconeogenesis. During moderate activity, muscle generates ATP by oxidizing glucose from glycogen, fatty acids, and ketone bodies completely to CO_2 and H_2O via the citric acid cycle and oxidative phosphorylation. Adipose tissue stores triacylglycerols and releases fatty acids into the bloodstream in response to the organism’s metabolic needs. These metabolic needs are communicated to adipose tissue by means of the hormones insulin, which indicates a fed state in which storage is appropriate, and glucagon, epinephrine, and norepinephrine, which signal a need for fatty acid release to provide fuel for other tissues. The liver, the body’s central metabolic clearinghouse, maintains blood glucose concentrations by storing glucose as glycogen in times of plenty and releasing glucose in times of need both by glycogen breakdown and by gluconeogenesis. It also converts fatty acids to ketone bodies for use by peripheral tissues. During a fast, it breaks down amino acids resulting from protein degradation to metabolic intermediates that can

be used to generate glucose. The kidney filters out urea from the blood, recovers important metabolites, and maintains pH balance. To do so, glutamine is broken down to produce NH_4^+ for H^+ excretion. The resulting α -ketoglutarate product is converted to CO_2 to resupply HCO_3^- to the blood to maintain its buffering capacity. During starvation, the kidney uses the α -ketoglutarate from glutamine breakdown for gluconeogenesis.

3 Metabolic Homeostasis: Regulation of Appetite, Energy Expenditure, and Body Weight AMP-dependent protein kinase (AMPK), the cell’s fuel gauge, senses the cell’s need for ATP and activates metabolic breakdown pathways while inhibiting biosynthetic pathways. Adiponectin, an adipocyte hormone that increases insulin sensitivity, acts by activating AMPK. Appetite is suppressed by the actions of leptin, a hormone produced by adipose tissue, insulin, produced by the β cells of the pancreas, and PYY₃₋₃₆, produced by the gastrointestinal tract, which act in the hypothalamus to inhibit the secretion of neuropeptide Y (NPY) and stimulate the secretion of α -MSH and CART. This decreases the appetite and hence food intake. Ghrelin, a hormone secreted by the empty stomach, opposes the actions of leptin, insulin, and PYY₃₋₃₆, stimulating appetite and food intake. Leptin also acts in peripheral tissues to stimulate energy expenditure by fatty acid oxidation and thermogenesis.

4 Metabolic Adaptation During prolonged starvation, the brain slowly adapts from the use of glucose as its sole fuel source to the use of ketone bodies, thereby shifting the metabolic burden from protein breakdown to fat breakdown. Diabetes mellitus is a disease in which insulin either is not secreted or does not efficiently stimulate its target tissues, leading to high concentrations of glucose in the blood and urine. Cells “starve” in the midst of plenty since they cannot absorb blood glucose and their hormonal signals remain those of starvation. Abnormally high production of ketone bodies is one of the most dangerous effects of uncontrolled diabetes. Metabolic syndrome is caused by obesity, physical inactivity, and possibly genetic determinants. Its symptoms can be relieved by substances that activate AMPK.

REFERENCES

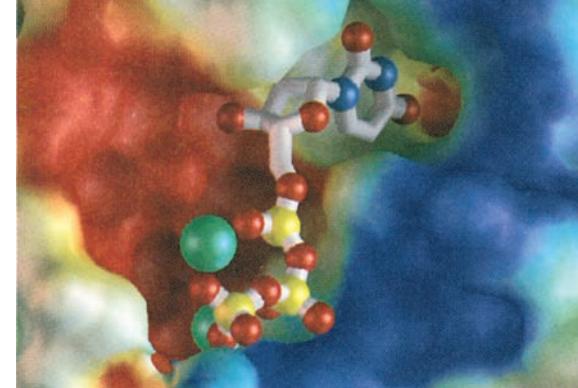
- Chapters 17 to 26 of this text.
- Batterham, R.L., et al., Gut hormone PYY₃₋₃₆ physiologically inhibits food intake, *Nature* **418**, 650–654 (2002).
- Brüning, J.C., Gautam, D., Burks, D.J., Gillette, J., Schubert, M., Orban, P.C., Klein, R., Krone, W., Müller-Weiland, D., and Kahn, C.R., Role of brain insulin receptor in control of body weight and reproduction, *Science* **289**, 2122–2125 (2000).
- Carling, D., The AMP-activated protein kinase cascade—a unifying system for energy control, *Trends Biochem. Sci.* **29**, 18–24 (2004).
- Coll, A.P., Farooqi, I.S., and O’Rahilly, S., The hormonal control of food intake, *Cell* **129**, 251–262 (2007).
- Evans, J.L., Goldfine, I.D., Maddux, B.A., and Grodsky, G.M., Oxidative stress and stress-activated signaling pathways: a unifying hypothesis of type 2 diabetes, *Endocrine Rev.* **23**, 599–622 (2002).
- Flier, J.S., Obesity wars: Molecular progress confronts an expanding epidemic, *Cell* **116**, 337–350 (2004).
- Kadowaki, T. and Yamauchi, T., Adiponectin and adiponectin receptors, *Endocrine Rev.* **26**, 439–451 (2005).
- Lowell, B.B. and Spiegelman, B.M., Towards a molecular understanding of adaptive thermogenesis, *Nature* **404**, 652–660 (2000).
- Montague, C.T., et al., Congenital leptin deficiency is associated with severe early-onset obesity in humans, *Nature* **387**, 903–908 (1997).
- Moreno-Aliaga, M.J., Martí, A., García-Foncillas, J. and Martínez, J.A., DNA hybridization arrays: a powerful technology for nutritional and obesity research, *Br. J. Nutr.* **86**, 119–122 (2001).

- Nakar, V.A., et al., AMPK and PPAR γ agonists are exercise mimetics, *Cell* **134**, 405–415 (2008).
- Nakazato, M., Murakami, N., Date, Y., Kojima, M., Matsuo, H., Kangawa, K., and Matsukura, S., A role for ghrelin in the central regulation of feeding, *Nature* **409**, 194–198 (2001).
- Schwartz, M.W. and Morton, G.J., Keeping hunger at bay, *Nature* **418**, 595–597 (2002).
- Shaw, R.J., Lamia, K.A., Vasquez, D., Koo, S.-H., Bardeesy, N., DePinho, R.A., Montminy, M., and Cantley, L.C., The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin, *Science* **310**, 1642–1646 (2005).
- Shulman, G.I., Cellular mechanisms of insulin resistance, *J. Clin. Invest.* **106**, 171–176 (2000).
- Towler, M.C. and Hardie, D.G., AMP-Activated protein kinase in metabolic control and insulin signaling, *Circ. Res.* **100**, 328–341 (2007).
- Tshöp, M., Smiley, D.L., and Heiman, M.L., Ghrelin induces adiposity in rodents, *Nature* **407**, 908–913 (2000).
- Type 2 Diabetes, *Science* **307**, 369–387 (2005). [A series of informative articles on the origins of type 2 diabetes and its relationship to obesity.]
- Unger, R.H., Leptin physiology: a second look, *Regul. Pept.* **92**, 87–95 (2000).
- Wang, Y.-X., Zhang, C.L., Yu, R.T., Cho, H.K., Nelson, M.C., Bayuga-Ocampo, C.R., Ham, J., Kang, H., and Evans, R.M., Regulation of muscle fiber type and running endurance by PPAR γ , *PLoS Biol.* **2**, e294 (2004).
- Yechoor, V.K., Patti, M.-E., Saccone, R., and Kahn, C.R., Coordinated patterns of gene expression for substrate and energy metabolism in skeletal muscle of diabetic mice, *Proc. Natl. Acad. Sci.* **99**, 10587–10592 (2002).
- Zhang, F., et al., Crystal structure of the *obese* protein leptin-E100, *Nature* **387**, 206–209 (1997).
- Zick, Y., Insulin resistance: a phosphorylation-based uncoupling of insulin signaling, *Trends Cell Biol.* **11**, 437–441 (2001).

PROBLEMS

1. Describe the metabolic effects of liver failure.
2. What is the basis of the hypothesis that athletes' muscles are more heavily buffered than those of normal individuals?
3. Experienced runners know that it is poor practice to ingest large amounts of glucose prior to running a long-distance race such as a marathon. What is the metabolic basis of this apparent paradox?
4. Explain why urea output is vastly decreased during starvation.
5. Explain why people survive longer by total fasting than on a diet consisting only of carbohydrates.
6. Explain why the breath of an untreated diabetic smells of acetone.
7. Among the many eat-all-you-want-and-lose-weight diets that have been popular for a time is one that eliminates all carbohydrates but permits the consumption of all the protein and fat desired. Would such a diet be effective? (Hint: Individuals on such a diet often complain that they have bad breath.)
8. Pancreatic β cells express a receptor for fatty acids. Fatty acid binding to this protein appears to stimulate insulin secretion. (a) Does this phenomenon make metabolic sense? (b) Fatty acids appear to stimulate insulin secretion to a much greater extent when glucose is also present. Why is this significant?
9. High concentrations of free fatty acids in the blood are known to cause insulin resistance in muscle, but only after 5 hours. This suggests that a metabolite of these fatty acids may be responsible for this phenomenon. It is also known that an isoform of protein kinase C is activated during the process and that high concentrations of free fatty acids result in intramuscular accumulation of triacylglycerols. With this information, review the mechanism of activation of PKC and the pathway of triacylglycerol biosynthesis and suggest a metabolite that may be responsible for PKC activation.
10. Discuss, in molecular terms, how physical inactivity might lead to insulin resistance.

Nucleotide Metabolism



CHAPTER 28

1 Synthesis of Purine Ribonucleotides

- A. Synthesis of Inosine Monophosphate
- B. Synthesis of Adenine and Guanine Ribonucleotides
- C. Regulation of Purine Nucleotide Biosynthesis
- D. Salvage of Purines

2 Synthesis of Pyrimidine Ribonucleotides

- A. Synthesis of UMP
- B. Synthesis of UTP and CTP
- C. Regulation of Pyrimidine Nucleotide Biosynthesis

3 Formation of Deoxyribonucleotides

- A. Production of Deoxyribose Residues
- B. Origin of Thymine

4 Nucleotide Degradation

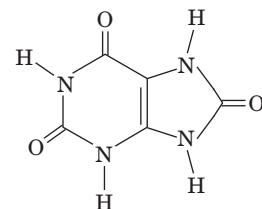
- A. Catabolism of Purines
- B. Fate of Uric Acid
- C. Catabolism of Pyrimidines

5 Biosynthesis of Nucleotide Coenzymes

- A. Nicotinamide Coenzymes
- B. Flavin Coenzymes
- C. Coenzyme A

1 SYNTHESIS OF PURINE RIBONUCLEOTIDES

In this section we commence our considerations of how nucleic acids and their components are synthesized by describing the synthesis of purine ribonucleotides. In 1948, John Buchanan obtained the first clues as to how this process occurs *de novo* by feeding a variety of isotopically labeled compounds to pigeons and chemically determining the positions of the labeled atoms in their excreted **uric acid** (a purine).



Uric acid

He used birds in these experiments because they excrete waste nitrogen almost entirely as uric acid, a water-insoluble and therefore easily isolated substance. The results of his studies, which are summarized in Fig. 28-1, demonstrated that N1 of purines arises from the amine group of aspartate; C2 and C8 originate from formate; N3 and N9 are contributed by the amide group of glutamine; C4, C5, and N7 are derived from glycine (strongly suggesting that this molecule is wholly incorporated into the purine ring); and C6 comes from HCO₃⁻ (CO₂).

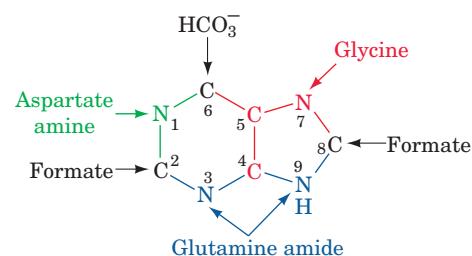
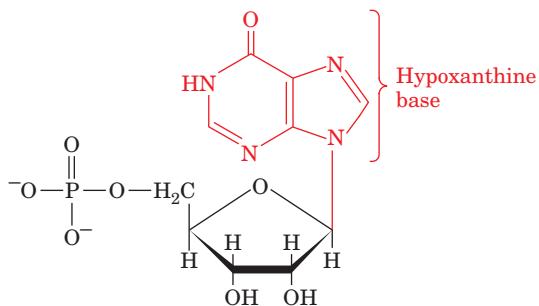


Figure 28-1 The biosynthetic origins of purine ring atoms.

Note that C4, C5, and N7 come from a single glycine molecule but each of the other atoms is derived from an independent precursor.

Nucleotides, as we have seen, are biologically ubiquitous substances that participate in nearly all biochemical processes: They are the monomeric units of DNA and RNA; the hydrolysis of ATP and GTP drives many free energy-requiring processes; the levels of ATP, ADP, and AMP regulate numerous metabolic pathways; cAMP and cGMP mediate hormonal signals; and NAD⁺, NADP⁺, FMN, FAD, and coenzyme A are essential coenzymes in a great variety of enzymatic reactions. The importance of nucleotides in cellular metabolism is indicated by the observation that nearly all cells can synthesize them both *de novo* (anew) and from the degradation products of nucleic acids. In this chapter, we consider the nature of these biosynthetic pathways. In doing so, we shall examine how they are regulated and the consequences of their blockade, both by genetic defects and through the administration of chemotherapeutic agents. We then discuss how nucleotides are degraded. Finally, we outline the biosynthesis of the nucleotide coenzymes.

The actual pathway by which these precursors are incorporated into the purine ring, the subject of Section 28-1A, was elucidated in subsequent investigations performed largely by Buchanan and by G. Robert Greenberg. These investigations showed that the initially synthesized purine derivative is **inosine monophosphate (IMP)**,



Inosine monophosphate (IMP)

the nucleotide of the base **hypoxanthine**. AMP and GMP are subsequently synthesized from this intermediate via separate pathways (Section 28-1B). Thus, contrary to naive expectation, purines are initially formed as ribonucleotides rather than as free bases. Additional studies have demonstrated that such widely divergent organisms as *E. coli*, yeast, pigeons, and humans have virtually identical pathways for the biosynthesis of purine nucleotides, thereby further demonstrating the biochemical unity of life.

A. Synthesis of Inosine Monophosphate

IMP is synthesized in a pathway comprising 11 reactions (Fig. 28-2):

1. Activation of ribose-5-phosphate. The starting material for purine biosynthesis is α -D-ribose-5-phosphate (R5P), a product of the pentose phosphate pathway (Section 23-4). In the first step of *de novo* purine biosynthesis, **ribose phosphate pyrophosphokinase** (also known as **phosphoribosylpyrophosphate synthetase**) activates R5P by reacting it with ATP to form **5-phosphoribosyl- α -pyrophosphate (PRPP)**. This reaction, which occurs via the nucleophilic attack of the R5P's C1—H group on the P_β of ATP, is unusual in that a pyrophosphoryl group is directly transferred from ATP to C1 of R5P and that the product has the α anomeric configuration. PRPP is also a precursor in the biosynthesis of pyrimidines (Section 28-2A) and the amino acids tryptophan and histidine (Section 26-5Bd,e). Thus, as is expected for an enzyme at such an important biosynthetic crossroads, the activity of ribose phosphate pyrophosphokinase varies with the concentrations of numerous metabolites, including PP_i and 2,3-bisphosphoglycerate, which are activators, and ADP and GDP, which are mixed inhibitors (Section 14-3C). The regulation of purine nucleotide biosynthesis is further discussed in Section 28-1C.

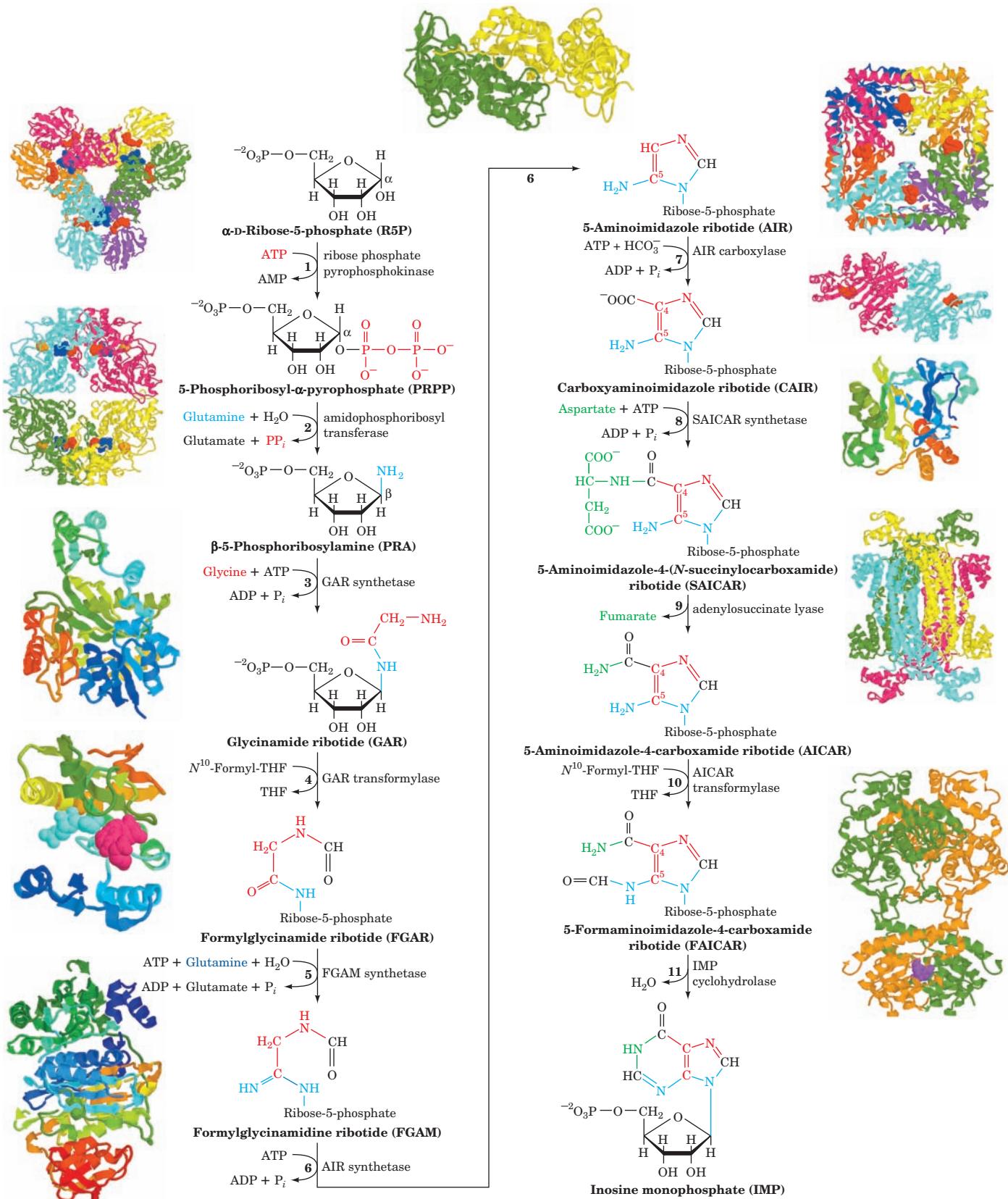
2. Acquisition of purine atom N9. Amidophosphoribosyltransferase (alternatively, **glutamine PRPP aminotransferase** or **PurF**; the latter being named for the *E. coli* gene encoding it, *purF*) catalyzes the displacement of PRPP's

pyrophosphate group by glutamine's amide nitrogen to yield **β -5-phosphoribosylamine (PRA)**. This is the first reaction in the pathway that is unique to *de novo* purine biosynthesis (and hence some sources refer to it as the first reaction of the pathway, which is then said to consist of 10 reactions). This process occurs in two consecutive reactions that take place on separate active sites on the enzyme:

1. Glutamine + H₂O \rightarrow glutamic acid + NH₃
2. NH₃ + PRPP \rightarrow PRA + PP_i

Step 1 is catalyzed by a member of the N-terminal nucleophile (Ntn) amidotransferase family (Section 26-5Aa). Step 2 occurs with inversion of configuration about ribose C1 and hence establishes the anomeric form of the future nucleotide. The NH₃ passes between the two active sites through a 20-Å-long tunnel that is lined with conserved

Figure 28-2 (Opposite) **The metabolic pathway for the *de novo* biosynthesis of IMP.** The purine residue is built up on a ribose ring in 11 enzymatically catalyzed reactions. The X-ray structures for all enzymes are shown to the outside of the corresponding reaction arrow. The peptide chains of monomeric enzymes are colored in rainbow order from their N-termini (blue) to their C-termini (red). The oligomeric enzymes, all of which consist of identical polypeptide chains, are viewed along a rotation axis with their various chains differently colored. Bound ligands are shown in space-filling form. Enzyme 1, determined by Sine Larsen, University of Copenhagen, Denmark, is a D_3 hexamer from *B. subtilis* that binds α, β -methylene-ADP at its catalytic (red) and allosteric (blue) sites; PDBid 1DKU. Enzyme 2, determined by Janet Smith, Purdue University, is a D_2 tetramer from *B. subtilis* that binds GMP (blue), ADP (red), and a [4Fe–4S] cluster (orange, which appears to have a regulatory rather than a redox function); PDBid 1AO0. Enzymes 3 and 6, both from *E. coli*, were determined by JoAnne Stubbe, MIT, and Steven Ealick, Cornell University; PDBids 1GSO and 1CLI. Enzyme 4, from *E. coli*, determined by Robert Almasy, Agouron Pharmaceuticals, San Diego, California, binds GAR (cyan) and 5-deazatetrahydrofolate (red); PDBid 1CDE. Enzyme 5, from *Thermatoga maritima*, was determined by Ian Wilson, Scripps Research Institute, La Jolla, California; PDBid 1VK3. Reaction 7, in *E. coli*, is catalyzed by two sequentially acting enzymes, Class I PurE (above) and PurK (below). Class I PurE, determined by JoAnne Stubbe, MIT, and Steven Ealick, Cornell University, is a D_4 octamer that binds AIR (red); PDBid 1D7A. PurK, determined by JoAnne Stubbe, MIT, and Hazel Holden, University of Wisconsin, is a C_2 dimer that binds ADP (red); PDBid 1B6S. Enzyme 8, from yeast, was determined by Victor Lamzin, Academy of Sciences, Moscow, Russia, and Keith Wilson, EMBL, Hamburg, Germany; PDBid 1A48. Enzyme 9, from *Thermatoga maritima*, determined by Todd Yeates, UCLA, is a D_2 tetramer; PDBid 1C3U. Reactions 10 and 11 in chicken are catalyzed by a bifunctional enzyme that was determined by Stephen Benkovic, Pennsylvania State University, and Ian Wilson, The Scripps Research Institute, La Jolla, California. It forms a C_2 dimer shown with its AICAR transformylase function above and its IMP cyclohydrolase function, which binds GMP (purple), below; PDBid 1G8M.  See the Animated Figures



nonpolar residues that lack hydrogen bonding groups and hence do not impede the diffusion of the NH_3 [we have seen that NH_3 generated by glutamine hydrolysis is similarly channeled to the active site that uses it in carbamoyl phosphate synthetase (Section 26-2Aa) and glutamate synthetase (Section 26-5Aa)]. These reactions, which are driven to completion by the subsequent hydrolysis of the released PP_i , constitute the pathway's flux-generating step. Not surprisingly, therefore, amidophosphoribosyltransferase is subject to feedback inhibition by purine nucleotides (Section 28-1C).

3. Acquisition of purine atoms C4, C5, and N7. Glycine's carboxyl group forms an amide with the amino group of PRA, yielding **glycinamide ribotide (GAR)** in a reaction, catalyzed by **GAR synthetase (PurD)**, that occurs via the intermediate phosphorylation of glycine's carboxyl group. The reaction, which is reversible despite its concomitant hydrolysis of ATP to $\text{ADP} + \text{P}_i$, is the only step of the purine biosynthesis pathway in which more than one purine ring atom is acquired. The observation that PRA is chemically unstable (it is hydrolyzed to R5P and NH_3 with a half-life of 5 s at 37°C) suggests that GAR synthetase and amidophosphoribosyltransferase associate in a way that channels PRA between them. Indeed, a sterically and electrostatically plausible model of such a complex has been built based on the X-ray structures of these two enzymes.

4. Acquisition of purine atom C8. GAR's primary α -amino group is formylated by **GAR transformylase (PurN)** to yield **formylglycinamide ribotide (FGAR)**. The formyl donor in this reaction is N^{10} -formyltetrahydrofolate (N^{10} -formyl-THF), a cofactor that transfers C_1 units from such donors as serine, glycine, and formate to various acceptors in biosynthetic reactions (Section 26-4D). Structural and enzymological studies indicate that the reaction proceeds via the nucleophilic attack of the GAR primary amine group on the formyl carbon of N^{10} -formyl-THF to yield a tetrahedral intermediate.

5. Acquisition of purine atom N3. The amide amino group of a second glutamine is transferred to the growing purine ring to form **formylglycinamidine ribotide (FGAM)**. This reaction, which is catalyzed by **FGAM synthetase (PurL)**, is driven by the coupled hydrolysis of ATP to $\text{ADP} + \text{P}_i$. It is thought to proceed by the mechanism diagrammed in Fig. 28-3. Here the oxygen of the FGAR isoamide form reacts with ATP to yield a phosphoryl ester intermediate. This intermediate then reacts with NH_3 (the glutamine amide nitrogen as stabilized through the transient formation of an enzyme thioester) to form a tetrahedral adduct. The adduct then eliminates P_i to yield the imine product, FGAM. Such reactions, in which a carboxamide oxygen is replaced by an imino group, are common in the biosynthesis of nucleotides. For example, Reaction 6 of this pathway and the reactions converting IMP to AMP (Section 28-1B) and UTP to CTP (Section 28-2B) follow similar mechanisms, that is, conversion of a carboxamide oxygen to a phosphoryl ester that is nucleophilically

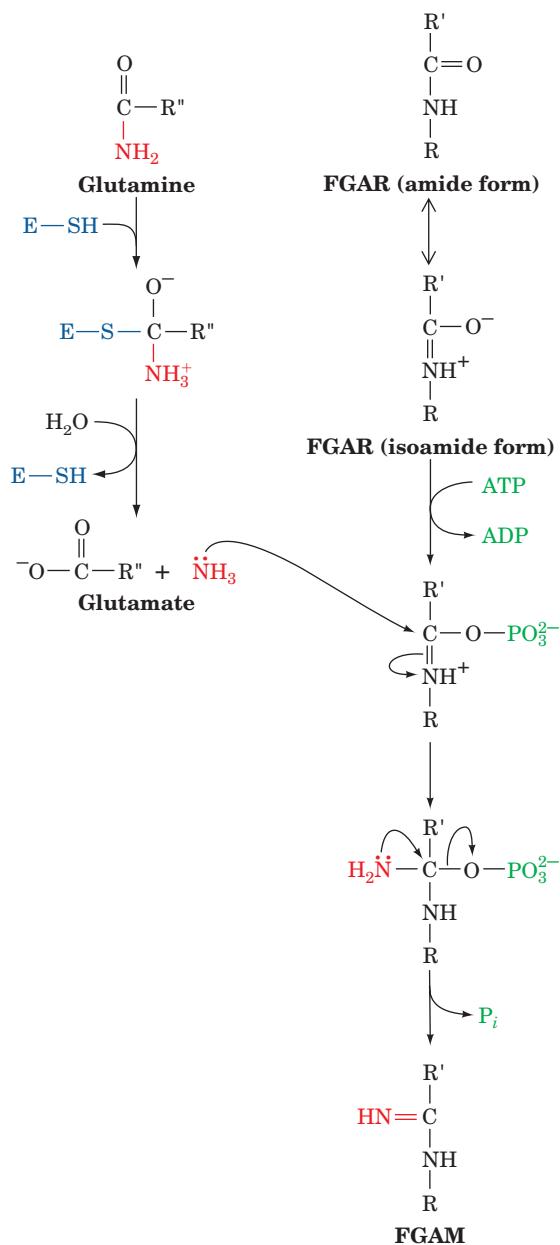


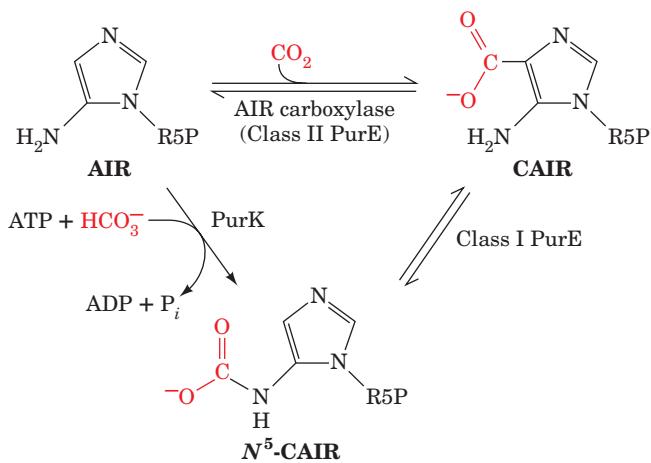
Figure 28-3 The proposed mechanism of formylglycinamide ribotide (FGAM) synthetase. The glutaminase domain of the enzyme contains an active site Cys residue that catalyzes the release of NH_3 with the transient formation of an enzyme thioester (not shown) whose hydrolysis produces glutamate. The isoamide form of FGAR is phosphorylated by ATP and then reacts with " NH_3 " to form a tetrahedral intermediate whose collapse yields FGAM + P_i .

attacked by an amine nitrogen atom to yield a tetrahedral adduct that, in turn, expels P_i to form product.

6. Formation of the purine imidazole ring. The purine imidazole ring is closed in an ATP-requiring intramolecular condensation that yields **5-aminoimidazole ribotide (AIR)** in a reaction catalyzed by **AIR synthetase (PurM)**.

The aromatization of the imidazole ring is facilitated by the tautomeric shift of the reactant from its imine to its enamine form.

7. Acquisition of C6. In higher eukaryotes, purine C6 is introduced as HCO_3^- (CO_2) in an ATP-dependent reaction catalyzed by **AIR carboxylase** that yields **carboxyaminoimidazole ribotide (CAIR)** and $\text{ADP} + \text{P}_i$. However, in yeast, plants, and most prokaryotes (including *E. coli*), this overall reaction occurs in two steps that are mediated by separate enzymatic activities: **PurK** and **Class I PurE**.



PurK catalyzes the ATP-dependent carboxylation of AIR to yield **$N^5\text{-CAIR}$** , which Class I PurE rearranges to yield CAIR. Class I PurE is homologous to AIR carboxylase, which is therefore also called **Class II PurE**. Class I PurE alone can catalyze the AIR carboxylase reaction but since its K_M for HCO_3^- is 110 mM, it requires an unphysiologically high (~100 mM) HCO_3^- concentration to do so at a significant rate. However, the action of PurK decreases the HCO_3^- concentration required for the PurE-catalyzed reaction by >1000-fold, presumably through the ATP-driven formation of carbonyl phosphate, as is also postulated to occur in the carbamoyl phosphate synthetase reaction (Section 26-2A). The observation that $N^5\text{-CAIR}$ is chemically unstable (it decomposes to AIR with a half-life of 15 s at pH 7.5 and 25°C) suggests that $N^5\text{-CAIR}$ is channeled between PurK and Class I PurE. In fact, in yeast and plants, the N-terminus of Class I PurE is fused to the C-terminus of PurK. However, in *E. coli*, these two enzymatic activities occur on separate proteins for which there is no evidence of association.

8. Acquisition of N1. Purine atom N1 is contributed by aspartate in an amide-forming condensation reaction yielding **5-aminoimidazole-4-(N-succinylcarboxamide) ribotide (SAICAR)** that is catalyzed by **SAICAR synthetase (PurC)**. The reaction, which is driven by the hydrolysis of ATP to $\text{ADP} + \text{P}_i$, chemically resembles Reaction 3.

9. Elimination of fumarate. SAICAR is cleaved with the release of fumarate, yielding **5-aminoimidazole-4-carboxamide ribotide (AICAR)** in a reaction catalyzed by **adenylosuccinate lyase (PurB)**. Reactions 8 and 9 chemically

resemble the reactions in the urea cycle in which citrulline is aminated to form arginine (Sections 26-2C and 26-2D). In both pathways, aspartate's amino group is transferred to an acceptor through an ATP-driven coupling reaction followed by the elimination of the aspartate carbon skeleton as fumarate. In plants and microorganisms, AICAR is also formed in the biosynthesis of histidine (Section 26-5Be) but since in that process the AICAR is derived from ATP, it provides for no net purine biosynthesis.

10. Acquisition of C2. The final purine ring atom is acquired through formylation by N^{10} -formyltetrahydrofolate, yielding **5-formaminoimidazole-4-carboxamide ribotide (FAICAR)** in a reaction catalyzed by **AICAR transformylase (PurH)**. In bacteria, this reaction and that of Reaction 4 are indirectly inhibited by sulfonamides, which, it will be recalled, prevent the synthesis of folate by competing with its *p*-aminobenzoate component (Section 26-4D). Animals, including humans, must acquire folate through the diet, since they are incapable of synthesizing it. They are therefore unaffected by sulfonamides. The antibiotic properties of sulfonamides are therefore largely a result of their inhibition of nucleic acid biosynthesis in susceptible bacteria.

11. Cyclization to form IMP. The final reaction in the pathway, ring closure to form IMP, occurs through the elimination of water as catalyzed by **IMP cyclohydrolase (PurJ)**. In contrast to Reaction 6, the cyclization that forms the imidazole ring, this reaction does not entail ATP hydrolysis.

In animals, the activities catalyzing Reactions 3, 4, and 6, Reactions 7 and 8, and Reactions 10 and 11 occur on single polypeptides. The intermediate products of these multifunctional enzymes are not readily released to the medium but are channeled to the succeeding enzymatic activities of the pathway, thereby increasing the overall rates of these multistep processes and protecting the intermediates from degradation by other cellular enzymes. We have previously seen, for example, that the formation of acetyl-CoA from pyruvate takes place on the pyruvate dehydrogenase multienzyme complex, which contains three enzymes catalyzing five consecutive reactions (Section 21-2A); that all seven enzymatic activities catalyzing fatty acid synthesis in animals occur on a single protein molecule (Section 25-4Ca,b); and that the multifunctional enzymes carbamoyl phosphate synthase I (Section 26-2Aa), glutamate synthase (Section 26-5Aa), tryptophan synthase (Section 26-5Bd), and amidophosphoribosyltransferase (see above) pass reactive intermediate products between their active sites via protein tunnels. It is becoming increasingly apparent that the association of functionally related enzymes is a widespread phenomenon.

B. Synthesis of Adenine and Guanine Ribonucleotides

IMP does not accumulate in the cell but is rapidly converted to AMP and GMP. AMP, which differs from IMP only in the replacement of its 6-keto group by an amino

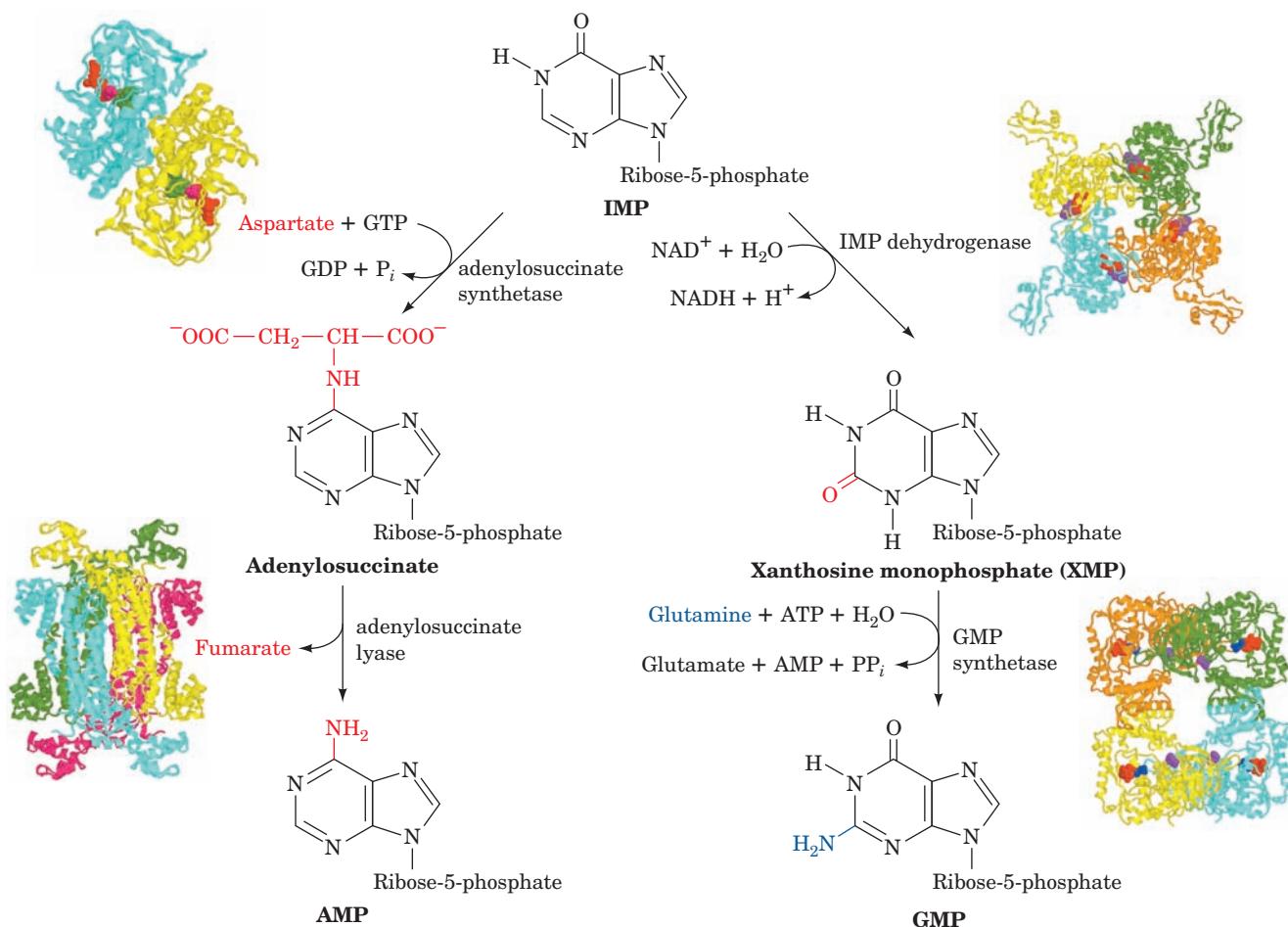


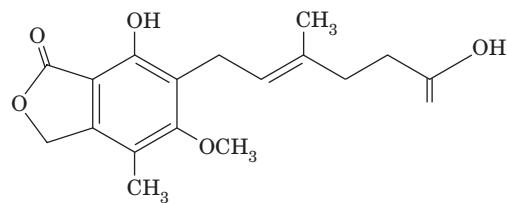
Figure 28-4 IMP is converted to AMP or GMP in separate two-reaction pathways. The X-ray structures for all of the enzymes catalyzing these reactions are shown to the outside of the corresponding reaction arrow. The X-ray structures for these homooligomers are shown as described in the legend to Fig. 28-2. Adenylosuccinate synthetase from *E. coli*, determined by Herbert Fromm and Richard Honzatko, Iowa State University, is a C₂ dimer in complex with IMP (green), GDP (red), and **hadacidin** (magenta; a competitive inhibitor of aspartate); PDBid 1GIM.

Adenylosuccinate lyase, from *Thermatoga maritima*, determined by Todd Yeates, UCLA, is a D₂ tetramer; PDBid 1C3U. IMP dehydrogenase from Chinese hamsters, determined by Keith Wilson, Vertex Pharmaceuticals, Cambridge, Massachusetts, is a C₄ tetramer in complex with oxidized IMP (red) and MPA (purple); PDBid 1JR1. GMP synthetase from *E. coli*, determined by Janet Smith, Purdue University, is a D₂ tetramer in complex with AMP (red), pyrophosphate (blue), and citrate (purple); PDBid 1GPM.

group, is synthesized in a two-reaction pathway (Fig. 28-4, left). In the first reaction, aspartate's amino group is linked to IMP in a reaction driven by the hydrolysis of GTP to GDP + P_i to yield **adenylosuccinate**. In the second reaction, **adenylosuccinate lyase** eliminates fumarate from adenylosuccinate to form AMP. This enzyme also catalyzes Reaction 9 of the IMP pathway (Fig. 28-2).

GMP is also synthesized from IMP in a two-reaction pathway (Fig. 28-4, right). In the first reaction, **IMP dehydrogenase** catalyzes the NAD⁺-dependent oxidation of IMP to form **xanthosine monophosphate (XMP)**; the ribonucleotide of the base **xanthine**). XMP is then converted to GMP by the replacement of its 2-keto group with glutamine's amide nitrogen in a reaction driven by the hydrolysis of ATP to AMP + PP_i (and subsequently to 2 P_i).

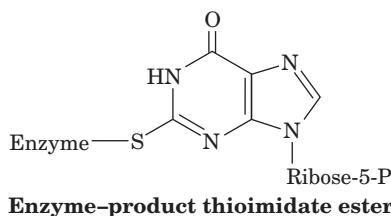
IMP dehydrogenase, a homotetramer of 514-residue subunits, was incubated with IMP, NAD⁺, and the fungally produced inhibitor **mycophenolic acid (MPA)**.



Mycophenolic acid (MPA)

The X-ray structure of the resulting complex, determined by Keith Wilson, reveals that the enzyme had bound MPA together with a reaction intermediate in which IMP atom

C2 had become covalently linked to the Cys 331 S atom and then dehydrogenated by NAD^+ to yield a thioimidate ester:



The mutagenic replacement of Cys 331 by Ala inactivates the enzyme. These observations strongly support a catalytic mechanism in which the Cys 331 thiol group nucleophilically attacks IMP's C2 atom, followed by hydride transfer to NAD^+ to yield the above covalently bound intermediate, which is subsequently hydrolyzed to yield XMP. The MPA binds to the enzyme with its bicyclic ring stacked on the purine ring (as would be expected for NAD^+ 's nicotinamide ring) and with its phenolic hydroxyl group in the proposed hydrolytic water site. This blocks the hydrolysis of the thioimidate ester, thereby inactivating the enzyme.

IMP dehydrogenase activity is essential to the immune response (Section 35-2) because it is required by the immune system cells known as B and T lymphocytes to generate the guanosine nucleotides they need to proliferate. Moreover, certain cancer cells have increased IMP dehydrogenase activity. Hence, IMP dehydrogenase is a target for both immunosuppressive therapy and cancer chemotherapy. Indeed, MPA is in clinical use to prevent the rejection of transplanted kidneys.

a. Nucleoside Diphosphates and Triphosphates Are Synthesized by the Phosphorylation of Nucleoside Monophosphates

In order to participate in nucleic acid synthesis, nucleoside monophosphates must first be converted to the corresponding nucleoside triphosphates. In the first of the two sequential phosphorylation reactions that do so, nucleoside diphosphates are synthesized from the corresponding nucleoside monophosphates by base-specific **nucleoside monophosphate kinases**. For example, adenylate kinase (Section 17-4Fd) catalyzes the phosphorylation of AMP to ADP:



Similarly, GDP is produced by a guanine-specific enzyme:



These nucleoside monophosphate kinases do not discriminate between ribose and deoxyribose in the substrate.

Nucleoside diphosphates are converted to the corresponding triphosphates by **nucleoside diphosphate kinase**; for instance,



Although this reaction is written with ATP as the phosphoryl donor and GDP as the acceptor, nucleoside diphosphate kinase is nonspecific as to the bases on either of its

substrates and as to whether their sugar residues are ribose or deoxyribose. The reaction occurs via a Ping Pong mechanism in which the substrate NTP phosphorylates an enzyme His residue, which in turn, phosphorylates the substrate NDP. The phosphoglycerate mutase reaction of glycolysis also has a phospho-His intermediate (Section 17-2H). The nucleoside diphosphate kinase reaction, as might be expected from the nearly identical structures of its substrates and products, normally operates close to equilibrium ($\Delta G \approx 0$). ADP is, of course, also converted to ATP by a variety of energy-releasing reactions such as those of glycolysis and oxidative phosphorylation. Indeed, it is these reactions that ultimately drive the foregoing kinase reactions.

C. Regulation of Purine Nucleotide Biosynthesis

The pathways involved in nucleic acid metabolism are tightly regulated, as is evidenced, for example, by the increased rates of nucleotide synthesis during cell proliferation. In fact, the pathways synthesizing IMP, ATP, and GTP are individually regulated in most cells so as not only to control the total amounts of purine nucleotides produced but also to coordinate the relative amounts of ATP and GTP. This control network is diagrammed in Fig. 28-5.

The IMP pathway is regulated at its first two reactions: those catalyzing the synthesis of PRPP and 5-phosphoribo-

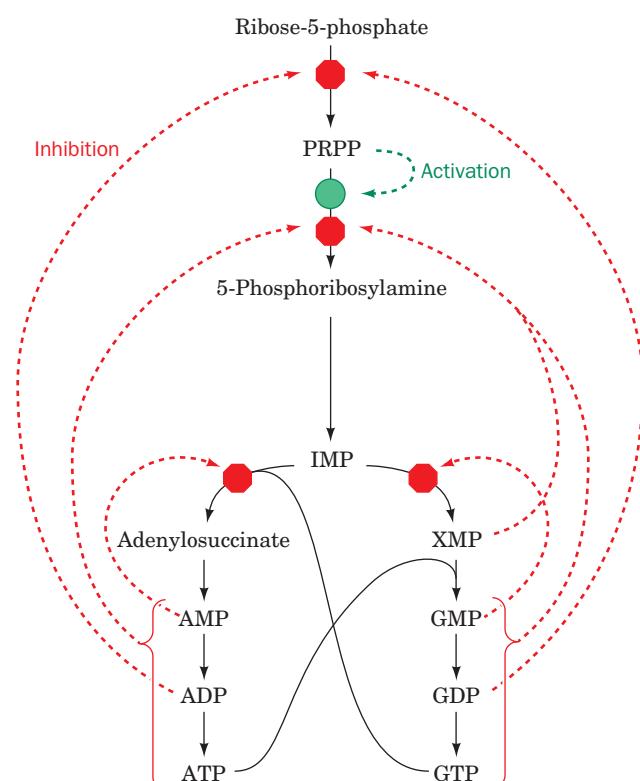


Figure 28-5 Control network for the purine biosynthesis pathway. Red octagons and green dots indicate control points. Feedback inhibition is indicated by dashed red arrows and feedforward activation is represented by dashed green arrows.

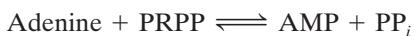
See the Animated Figures

sylamine. We have already seen that ribose phosphate pyrophosphokinase, the enzyme catalyzing Reaction 1 of the IMP pathway, is inhibited by both ADP and GDP (Section 28-1A). Amidophosphoribosyltransferase, the enzyme catalyzing the first committed step of the IMP pathway (Reaction 2), is likewise subject to feedback inhibition. In this case, however, the enzyme binds ATP, ADP, and AMP at one inhibitory site and GTP, GDP, and GMP at another. *The rate of IMP production is consequently independently but synergistically controlled by the levels of adenine nucleotides and guanine nucleotides.* Moreover, amidophosphoribosyltransferase is allosterically stimulated by PRPP (feedforward activation).

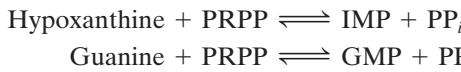
A second level of regulation occurs immediately below the branch point leading from IMP to AMP and GMP (Fig. 28-4). AMP and GMP are each competitive inhibitors of IMP in their own synthesis, so that excessive buildup of these products is impeded. In addition, the synthesis rates of adenine and guanine nucleotides are coordinated because GTP powers the synthesis of AMP from IMP, whereas ATP powers the synthesis of GMP from IMP. This reciprocity serves to balance the production of AMP and GMP (which are required in roughly equal amounts in nucleic acid biosynthesis): *The rate of synthesis of GMP increases with [ATP], whereas that of AMP increases with [GTP].*

D. Salvage of Purines

Most cells have an active turnover of many of their nucleic acids (particularly some types of RNA) which, through degradative processes described in Section 28-4A, result in the release of adenine, guanine, and hypoxanthine. These free purines are reconverted to their corresponding nucleotides through **salvage pathways**. In contrast to the *de novo* purine nucleotide synthesis pathway, which is virtually identical in all cells, salvage pathways are diverse in character and distribution. In mammals, purines are, for the most part, salvaged by two different enzymes. **Adenine phosphoribosyltransferase (APRT)** mediates AMP formation through the transfer of adenine to PRPP with the release of PP_i:



Hypoxanthine–guanine phosphoribosyltransferase (HGPRT) catalyzes the analogous reaction for both hypoxanthine and guanine:



a. Lesch–Nyhan Syndrome Results from HGPRT Deficiency

The symptoms of **Lesch–Nyhan syndrome**, which is caused by a severe HGPRT deficiency, indicate that purine salvage reactions have functions other than conservation of the energy required for *de novo* purine biosynthesis. This sex-linked congenital defect (affects almost only males) results in excessive uric acid production (uric acid is a purine

degradation product; Section 28-4A) and neurological abnormalities such as spasticity, mental retardation, and highly aggressive and destructive behavior, including a bizarre compulsion toward self-mutilation. For example, many children with Lesch–Nyhan syndrome have such an irresistible urge to bite their lips and fingers that they must be restrained. If the restraints are removed, communicative patients will plead that the restraints be replaced even as they attempt to injure themselves.

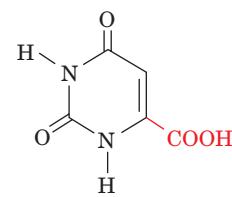
The excessive uric acid production in patients with Lesch–Nyhan syndrome is readily explained. The lack of HGPRT activity leads to an accumulation of the PRPP that would normally be used in the salvage of hypoxanthine and guanine. The excess PRPP activates amidophosphoribosyltransferase (which catalyzes Reaction 2 of the IMP biosynthesis pathway; Fig. 28-2), thereby greatly increasing the rate of synthesis of purine nucleotides and consequently that of their degradation product, uric acid. Yet the physiological basis of the associated neurological abnormalities remains obscure. That a defect in a single enzyme can cause such profound but well-defined behavioral changes nevertheless has important psychiatric implications.

2 SYNTHESIS OF PYRIMIDINE RIBONUCLEOTIDES

The biosynthesis of pyrimidines is a simpler process than that of purines. Isotopic labeling experiments have shown that atoms N1, C4, C5, and C6 of the pyrimidine ring are all derived from aspartic acid, C2 arises from HCO₃⁻, and N3 is contributed by glutamine (Fig. 28-6). In this section we discuss the pathways for pyrimidine ribonucleotide biosynthesis and how these processes are regulated.

A. Synthesis of UMP

The major breakthrough in the determination of the pathway for the *de novo* biosynthesis of pyrimidine ribonucleotides was the observation that mutants of the bread mold *Neurospora crassa*, which are unable to synthesize pyrimidines and therefore require both cytosine and uracil in their growth medium, grow normally when supplied instead with the pyrimidine **orotic acid** (uracil-6-carboxylic acid).



Orotic acid (uracil-6-carboxylic acid)

This observation led to the elucidation of the following six-reaction pathway for the biosynthesis of UMP (Fig. 28-7). Note that, in contrast to the case for purine nucleotides, the pyrimidine ring is coupled to the ribose-5-phosphate moiety *after* the ring has been synthesized.

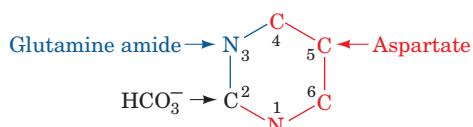


Figure 28-6 The biosynthetic origins of pyrimidine ring atoms.

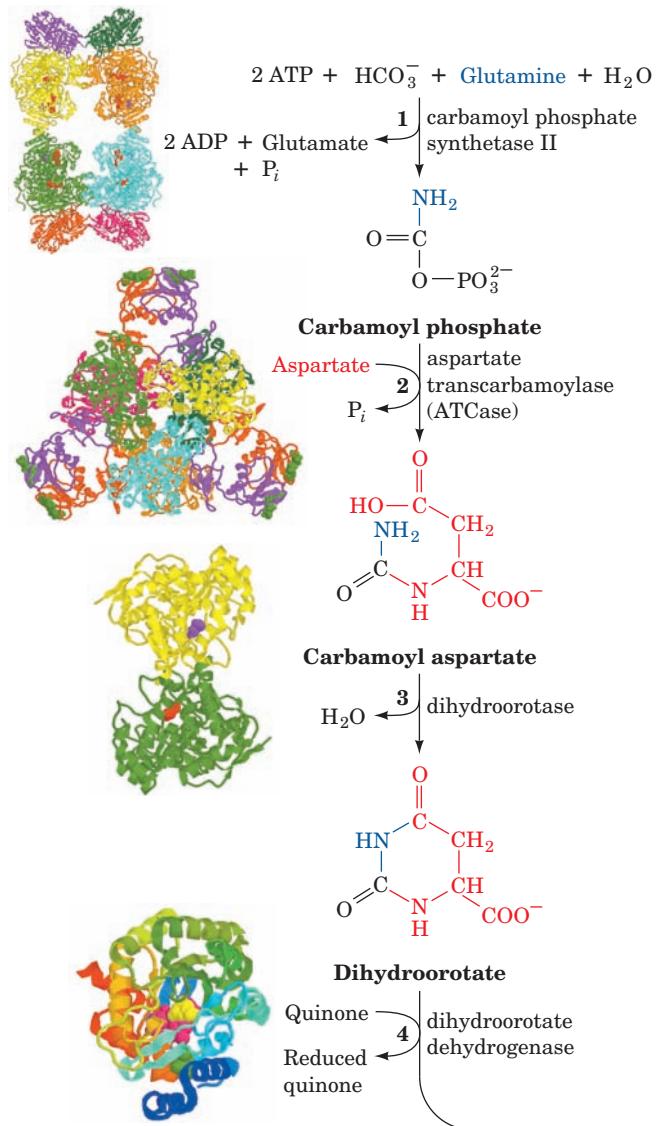
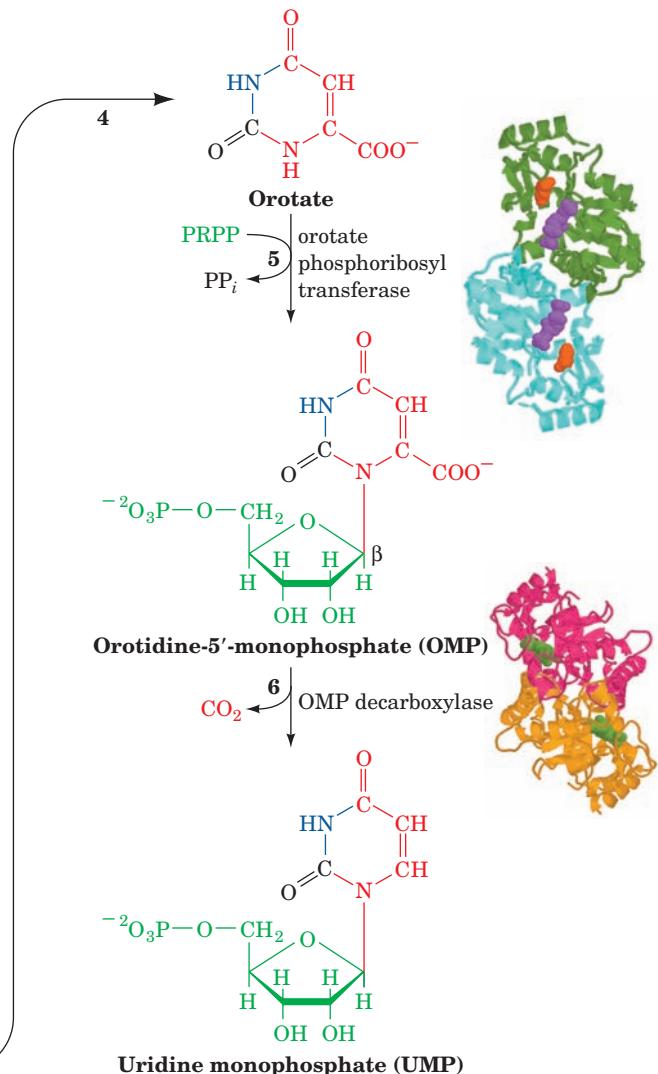


Figure 28-7 Metabolic pathway for the *de novo* synthesis of UMP. The pathway consists of six enzymatically catalyzed reactions. Note that, in contrast to the case for purine biosynthesis (Fig. 28-2), the pyrimidine ring is formed before its attachment to a ribose ring. The X-ray structures for the enzymes are drawn as described in the legend to Fig. 28-2. Enzyme 1, from *E. coli*, determined by Hazel Holden, University of Wisconsin, is an $\alpha_4\beta_4$ heterooctamer with D_2 symmetry, whose large subunits each bind two ADPNPs (red) and one ornithine (purple); PDBid 1D3H. Enzyme 2, from *E. coli*, determined by William Lipscomb, Harvard University, is a c_6r_6 heterododecamer with D_3 symmetry, whose regulatory (r) subunits each bind a CTP (green); PDBid

1. Synthesis of carbamoyl phosphate. The first reaction of pyrimidine biosynthesis is the synthesis of **carbamoyl phosphate** from HCO_3^- and the amide nitrogen of glutamine by the cytosolic enzyme **carbamoyl phosphate synthetase II (CPS II)**. This reaction is unusual in that it does not use biotin and consumes two molecules of ATP: One provides a phosphate group and the other energizes the reaction. We have previously discussed the synthesis of carbamoyl phosphate in connection with the formation of arginine (Section 26-2A). The carbamoyl phosphate that is used to



5AT1. Enzyme 3 from *E. coli*, determined by Hazel Holden, University of Wisconsin, is a C_2 dimer that binds carbamoyl aspartate (purple) in one subunit and orotate (red) in the other; PDBid 1J79. Enzyme 4, from humans, determined by Jon Clardy, Cornell University, is a monomer that binds orotate (yellow), FMN (magenta), and A77 1726 (green); PDBid 1D3H. Enzyme 5, from *Salmonella typhimurium*, determined by James Sacchettini, Albert Einstein College of Medicine, is a C_2 dimer that binds orotate (orange) and PRPP (purple); PDBid 1OPR. Enzyme 6, from *B. subtilis*, determined by Steven Ealick, Cornell University, is a C_2 dimer that binds UMP (green); PDBid 1DBT. See the Animated Figures

synthesize arginine via the urea cycle is synthesized by a separate mitochondrial enzyme, **carbamoyl phosphate synthetase I (CPS I)**, which uses ammonia as its nitrogen source. Prokaryotes only have one carbamoyl phosphate synthetase, which supplies both pyrimidine and arginine biosynthesis and utilizes glutamine. This latter enzyme, as we have seen, contains three different active sites that are connected by a remarkable 96-Å-long tunnel through which intermediate products diffuse (Fig. 26-9).

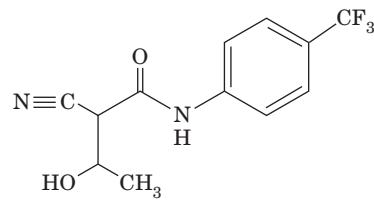
The pyrimidine biosynthetic pathway is a target for antiparasitic drugs. For example, the parasitic protozoan *Toxoplasma gondii*, which infects most mammals, causes **toxoplasmosis**, a disease whose complications include blindness, neurological dysfunction, and death in immunocompromised individuals (e.g., those with AIDS). Most parasites have evolved to take advantage of nutrients supplied by their hosts. However, *T. gondii* is unable to meet its needs exclusively through nucleotide salvage pathways and retains the ability to synthesize uracil *de novo*. Drugs that target the parasite's carbamoyl phosphate synthetase II (an enzyme whose structure and kinetics distinguish it from its mammalian counterpart) could therefore prevent *T. gondii* growth. Moreover, there is evidence that *T. gondii* strains that have been engineered to lack carbamoyl phosphate synthetase II are avirulent and could be useful as vaccines in humans and livestock.

2. Synthesis of carbamoyl aspartate. Condensation of carbamoyl phosphate with aspartate to form **carbamoyl aspartate** is catalyzed by **aspartate transcarbamylase (ATCase)**. This reaction, the pathway's flux-generating step, occurs without need of ATP because carbamoyl phosphate is intrinsically activated. The structure and regulation of *E. coli* ATCase is discussed in Section 13-4.

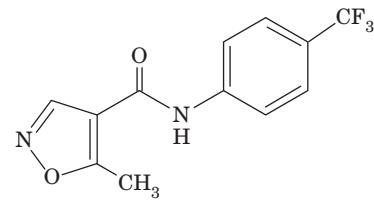
3. Ring closure to form dihydroorotate. The third reaction of the pathway was elucidated by Arthur Kornberg following his observation that microorganisms made to utilize orotic acid as a carbon source first reduce it to **dihydroorotate**. The reaction forming the pyrimidine ring yields dihydroorotate in an intramolecular condensation catalyzed by the zinc metalloenzyme **dihydroorotase**.

4. Oxidation of dihydroorotate. Dihydroorotate is irreversibly oxidized to orotate by **dihydroorotate dehydrogenase (DHODH)**. The eukaryotic enzyme, which contains FMN, is an integral membrane protein that is located on the outer surface of the inner mitochondrial membrane, where ubiquinone supplies its oxidizing power. The other five enzymes of pyrimidine nucleotide biosynthesis are cytosolic in animal cells. Many bacterial dihydroorotate dehydrogenases are NAD⁺-linked flavoproteins that contain FMN, FAD, and a [2Fe-2S] cluster. These enzymes normally function degradatively, that is, in the direction orotate → dihydroorotate, thereby permitting these bacteria to metabolize orotate and accounting for Kornberg's observation. The reaction mediated by eukaryotic DHODH involves two redox steps, as is indicated in Fig. 28-8. The X-ray structure of human DHODH in complex with orotate, determined by Jon Clardy, reveals that the pyrimidine ring of orotate is stacked over the FMN's flavin ring with

the orotate C6 and FMN N5 separated by 3.6 Å, a distance that is compatible with direct hydride transfer between these two centers. A tunnel leads from the opposite side of the flavin ring to a hydrophobic region on the enzyme surface. The enzyme presumably binds to the mitochondrial membrane surface via this hydrophobic patch, thereby permitting ubiquinone, which readily diffuses within the mitochondrial membrane, to approach and reoxidize the enzyme's bound FMNH₂. In the X-ray structure, this tunnel contains a tightly bound molecule named **A77 1726**, which is the primary metabolite of **leflunomide** (trade name **Arava**),



A77 1726



Leflunomide

a compound that is in clinical use for the treatment of rheumatoid arthritis. A77 1726 attenuates this autoimmune disease by blocking pyrimidine biosynthesis in T lymphocytes, thereby reducing their inappropriate proliferation. However, A77 1726 does not inhibit bacterial DHODHs.

5. Acquisition of the ribose phosphate moiety. Orotate reacts with PRPP to yield **orotidine-5'-monophosphate (OMP)** in a reaction catalyzed by **orotate phosphoribosyltransferase** and driven by hydrolysis of the eliminated PP_i. This reaction fixes the anomeric form of pyrimidine nucleotides in the β configuration. Orotate phosphoribosyltransferase also acts to salvage other pyrimidine bases, such as uracil and cytosine, by converting them to their corresponding nucleotides. Although the various phosphoribosyltransferases, including HGPRT, exhibit little sequence similarity, their X-ray structures indicate that they contain a common structural core that resembles the dinucleotide binding fold (Section 8-3Bi) but lacks one of its β strands.

6. Decarboxylation to form UMP. The final reaction of the pathway is the decarboxylation of OMP by **OMP decarboxylase (ODCase)** to form UMP. ODCase enhances the rate (k_{cat}/K_M) of OMP decarboxylation by a factor of 2×10^{23} over that of the uncatalyzed reaction, making it the most catalytically proficient enzyme known. Yet ODCase has no cofactors to help stabilize the reaction's putative carbanion intermediate. How is it able to do so? The X-ray structure, by Steven Ealick, of ODCase from *B. subtilis* in complex with UMP indicates that a bound OMP's C6

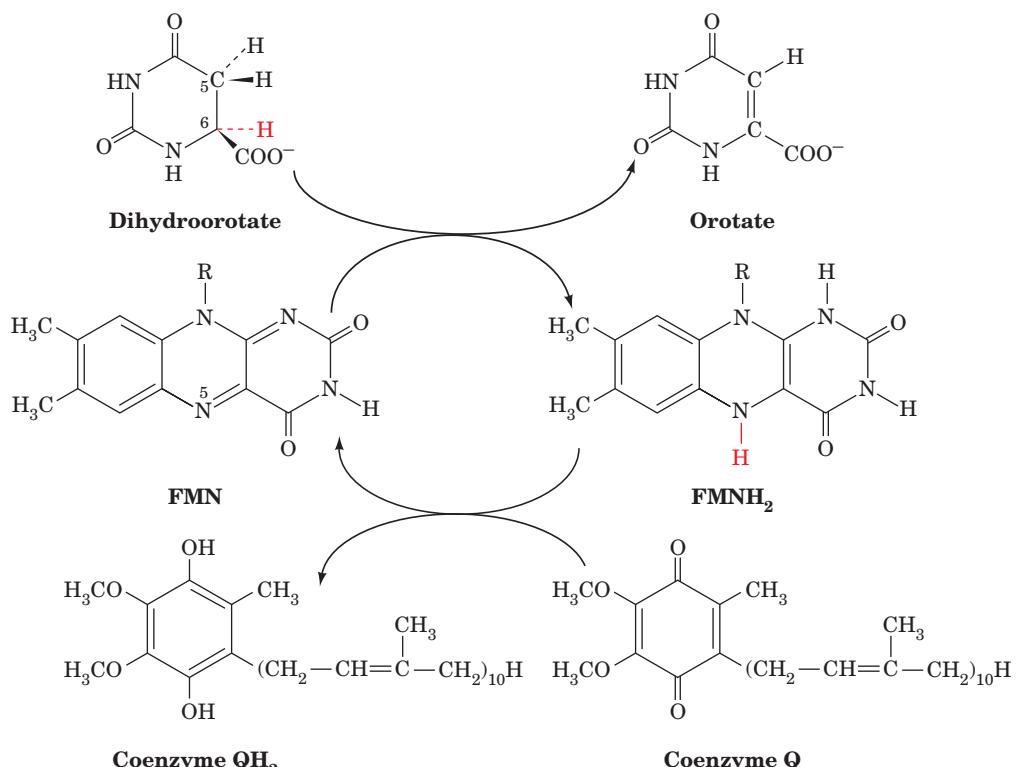


Figure 28-8 Reactions catalyzed by eukaryotic dihydroorotate dehydrogenase. The reaction is initiated by the enzyme-mediated abstraction of a proton from C5 of dihydroorotate followed by the direct hydride transfer from C6 of dihydroorotate to N5 of FMN to yield orotate and FMNH₂, which may then be

protonated to yield FMNH₂. The FMNH₂ (or FMNH⁻) then reacts with coenzyme Q acquired from the inner mitochondrial membrane to regenerate the enzyme in its FMN form and yield coenzyme QH₂, which then re-enters the inner mitochondrial membrane.

carboxyl group that is coplanar with its pyrimidine ring would be in close proximity to the side chains of both Asp 60 and Lys 62. Ealick has therefore proposed a mechanism (Fig. 28-9) in which the electrostatic interactions between the closely spaced carboxyl groups of OMP and Asp 60 destabilize OMP's ground state. This destabilization would be reduced in the transition state by the shift of OMP's negative charge from its carboxyl group toward C6, where it would be stabilized by the adjacent positively charged side chain of Lys 62. This side chain is also proposed to protonate the fragmenting C—C bond when it becomes sufficiently basic to accept the proton, thus avoiding the formation of a high-energy carbanion intermediate. The

unfavorable electrostatic interaction between OMP and Asp 60 occurs because the enzyme tightly binds OMP through extensive interactions with its other functional groups. Indeed, the removal of OMP's phosphate group, which is quite distant from the C6 carboxyl group, decreases the catalytic reaction's k_{cat}/K_M by a factor of 7×10^7 , thus providing a striking example of how binding energy can be applied to catalysis (preferential transition state binding).

In bacteria, the six enzyme activities mediating UMP biosynthesis occur on independent proteins (Fig. 28-7). In animals, however, as Mary Ellen Jones demonstrated, the first three enzymatic activities of the pathway, carbamoyl

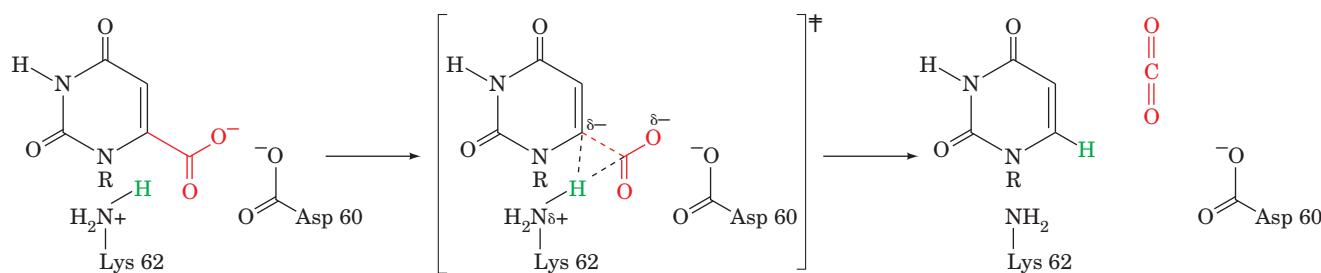


Figure 28-9 Proposed catalytic mechanism for OMP decarboxylase. [After Appleby, T.C., Kinsland, C., Begley, T.P., and Ealick, S.E., *Proc. Natl. Acad. Sci.* **97**, 2005 (2000).]

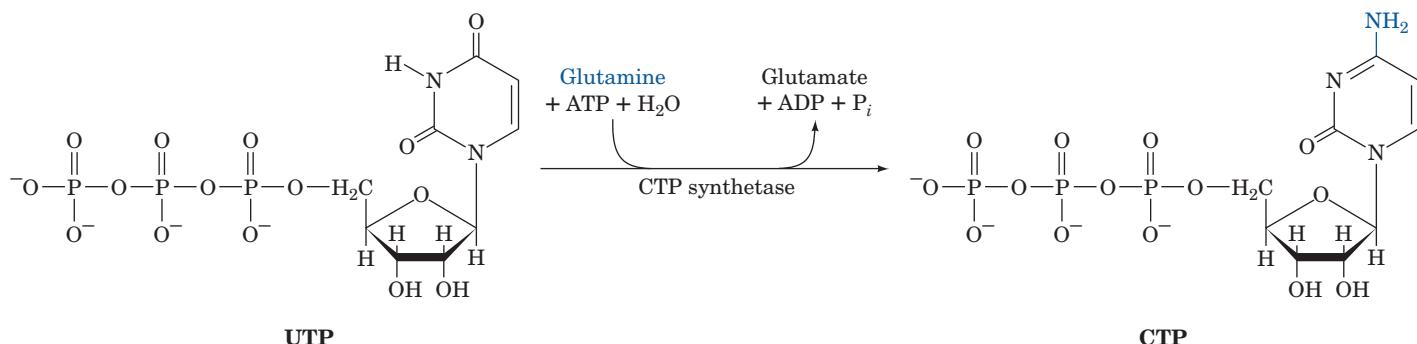
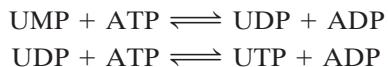


Figure 28-10 Synthesis of CTP from UTP.

phosphate synthetase II, ATCase, and dihydroorotate, occur on a single 2225-residue polypeptide chain known as **CAD**, which forms 1400-kD homohexamers. Similarly, Reactions 5 and 6 of the animal pyrimidine pathway are catalyzed by a single 480-residue polypeptide named **UMP synthase** that forms 102-kD homodimers.

B. Synthesis of UTP and CTP

The synthesis of UTP from UMP is analogous to the synthesis of purine nucleotide triphosphates (Section 28-1B). The process occurs by the sequential actions of a nucleoside monophosphate kinase and nucleoside diphosphate kinase:



CTP is formed by amination of UTP by **CTP synthetase** (Fig. 28-10). In animals, the amino group is donated by glutamine, whereas in bacteria it is supplied directly by ammonia.

C. Regulation of Pyrimidine Nucleotide Biosynthesis

In bacteria, the pyrimidine biosynthesis pathway is primarily regulated at Reaction 2, the ATCase reaction (Fig. 28-11a). In *E. coli*, control is exerted there through the allosteric stimulation of ATCase by ATP and its inhibition by CTP (Section 13-4). In many bacteria, however, UTP is the major ATCase inhibitor.

In animals, ATCase is not a regulatory enzyme. Rather, pyrimidine biosynthesis is controlled by the activity of carbamoyl phosphate synthetase II, which is inhibited by UDP and UTP and activated by ATP and PRPP (Fig. 28-11b). A second level of control in the mammalian pathway occurs at OMP decarboxylase, for which UMP and to a lesser extent CMP are competitive inhibitors. In all organisms, the rate of OMP production varies with the availability of its precursor, PRPP. The PRPP level, it will be recalled, depends on the activity of ribose phosphate pyrophosphokinase, which is inhibited by ADP and GDP (Section 28-1A).

a. Orotic Aciduria Results from an Inherited Enzyme Deficiency

Orotic aciduria, an inherited human disease, is characterized by the excretion of large amounts of orotic acid in the urine, retarded growth, and severe anemia. It results from a deficiency in the bifunctional enzyme catalyzing

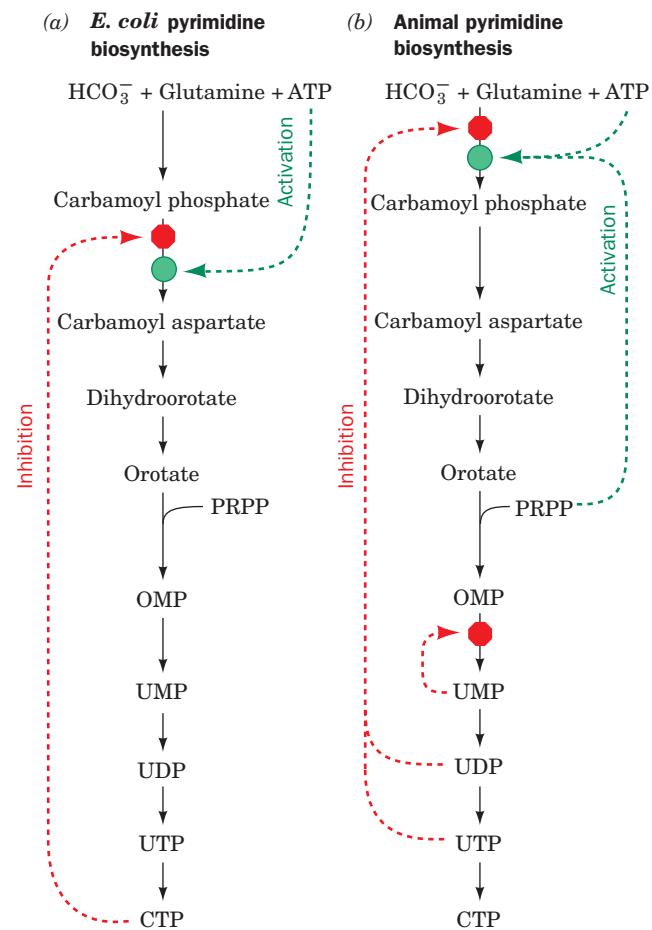


Figure 28-11 Regulation of pyrimidine biosynthesis. The control networks are shown for (a) *E. coli* and (b) animals. Red octagons and green dots indicate control points. Feedback inhibition is represented by dashed red arrows and activation is indicated by dashed green arrows.  See the Animated Figures

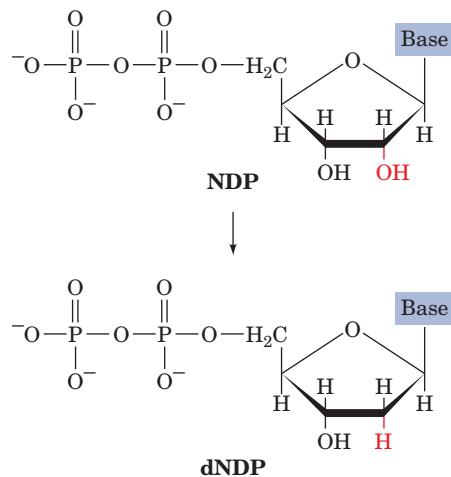
Reactions 5 and 6 of pyrimidine nucleotide biosynthesis. Consideration of the biochemistry of this situation led to its effective treatment: the administration of uridine and/or cytidine. The UMP formed through the phosphorylation of these nucleosides, besides replacing that normally synthesized, inhibits carbamoyl phosphate synthetase II so as to attenuate the rate of orotic acid synthesis. Few other genetic deficiencies in pyrimidine nucleotide biosynthesis are known in humans, presumably because most such defects are lethal *in utero*.

3 FORMATION OF DEOXYRIBONUCLEOTIDES

DNA differs chemically from RNA in two major respects: (1) its nucleotides contain 2'-deoxyribose residues rather than ribose residues, and (2) it contains the base thymine (5-methyluracil) rather than uracil. In this section we consider the biosynthesis of these DNA components.

A. Production of Deoxyribose Residues

Deoxyribonucleotides are synthesized from their corresponding ribonucleotides by the reduction of their C2' position rather than by their *de novo* synthesis from deoxyribose-containing precursors.



This pathway was established through Irwin Rose's study of how rats metabolize cytidine that is ¹⁴C-labeled in both its base and ribose components. The dCMP recovered from the rats' DNA had the same labeling ratio in its cytosine and deoxyribose residues as had the original cytidine, indicating that the DNA's components remained linked during DNA synthesis. If the cytosine and the ribose residues had become separated, dilution of the labeled cytosine and ribose residues with unlabeled residues, which are present in rat tissues in different amounts, would have altered this ratio.

The enzymes that catalyze the formation of deoxyribonucleotides by the reduction of the corresponding ribonucleotides are named **ribonucleotide reductases (RNRs)**. Three classes of RNRs are known that differ in their

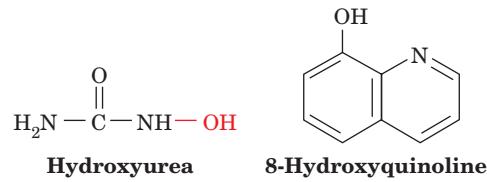
substrates (NDP or NTP), the cofactors they employ, and in the way they obtain reducing equivalents (see below). Class I and II RNRs are widely distributed among prokaryotes; some species have a Class I RNR, whereas other, sometimes related species have a Class II RNR. However, all eukaryotes except a few unicellular species have Class I RNRs. Class III RNRs occur in prokaryotes that can grow anaerobically. (Class III RNRs are O₂-sensitive whereas Class I RNRs require O₂ for activation; see below.) In fact, *E. coli*, which can grow both aerobically and anaerobically, expresses a Class I and a Class III RNR. In what follows, we shall mainly discuss the mechanism of Class I RNRs but end with a consideration of the evolutionary relationships among the different classes of RNRs.

a. Class I Ribonucleotide Reductase: Structure and Mechanism

The *E. coli* Class I RNR, as Peter Reichard demonstrated, is mainly present *in vitro* as a heterotetramer that can be decomposed to two catalytically inactive homodimers, R1₂ (761-residue subunits) and R2₂ (375-residue subunits), which together form the enzyme's two active sites (Fig. 28-12a). Each R1 subunit contains a substrate binding site as well as three independent effector binding sites that control both the enzyme's catalytic activity and its substrate specificity (see below). R1's catalytic residues include several redox-active thiol groups.

The X-ray structure of R2₂ (Fig. 28-12b), determined by Hans Eklund, reveals that each of its subunits contains a novel binuclear Fe(III) prosthetic group whose two Fe(III) ions are bridged by both an O²⁻ ion (a μ -oxo bridge) and the carboxyl group of Glu 115 (Fig. 28-12c). Each Fe(III) is further liganded by two carboxyl O atoms from Asp or Glu residues, a His N₈ atom, and a water molecule. The Fe(III) complex interacts with Tyr 122 to form, as EPR measurements indicate, an unusual tyrosyl free radical (TyrO[•]) that is 5 Å from the closest Fe atom and is buried 10 Å beneath the surface of the protein, where it is out of contact with solvent and any oxidizable side chain [tyrosyl radicals have also been observed in cytochrome *c* oxidase (Section 22-2C5c) and in Photosystem II (Section 24-2Cd)].

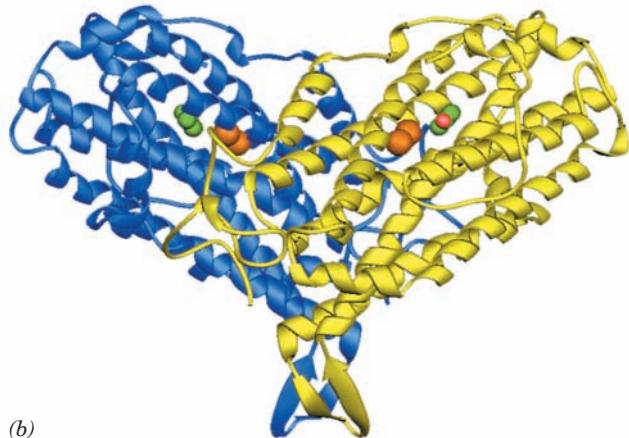
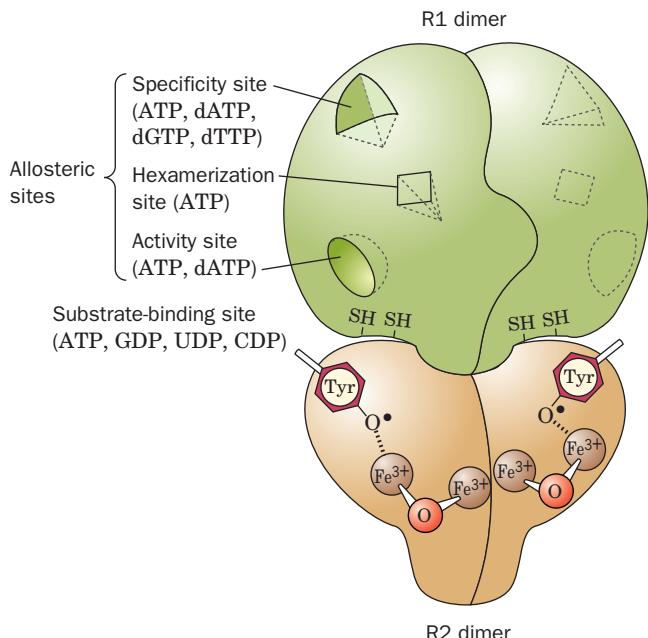
The *E. coli* RNR is inhibited by **hydroxyurea**, which specifically quenches (destroys) the tyrosyl radical, and by **8-hydroxyquinoline**, which chelates Fe³⁺ ions.



Mammalian RNRs have similar characteristics to the *E. coli* enzyme. Indeed, hydroxyurea is in clinical use as an antitumor agent.

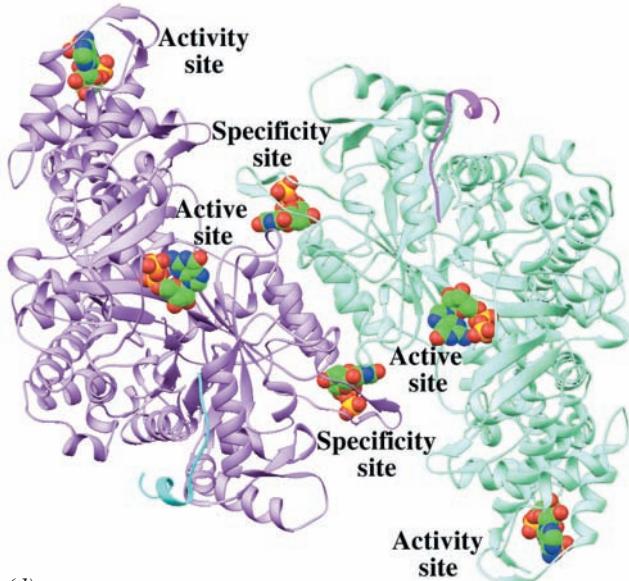
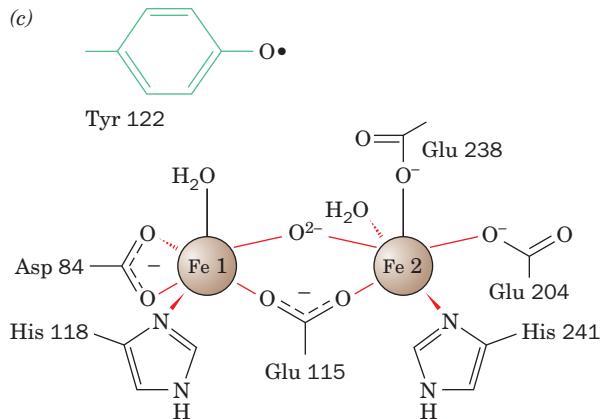
If *E. coli* RNR is incubated with [3'-³H]UDP, a small but reproducible fraction of the ³H is released as ³H₂O. This observation, together with kinetic, spectroscopic, and site-directed mutagenesis studies, led JoAnne Stubbe to

(a)



(b)

(c)



(d)

Figure 28-12 (Opposite) Class I ribonucleotide reductase from *E. coli*. (a) A schematic diagram of its quaternary structure. The enzyme consists of two pairs of identical subunits, R1₂ and R2₂. Each R2 subunit contains a binuclear Fe(III) complex that generates a phenoxy radical at its Tyr 122. The R1 subunits each contain three different allosteric effector sites and five catalytically important Cys residues. The enzyme's two active sites are located near the interface between neighboring R1 and R2 subunits. (b) The X-ray structure of R2₂ as viewed perpendicular to its 2-fold axis with the dimer's longest dimension in the horizontal plane. One subunit of the homodimeric protein is shown in blue and the other in yellow. The Fe(III) ions of its binuclear Fe complexes are represented by orange spheres and the radical-harboring Tyr 122 side chains are drawn in space-filling form with their C and O atoms green and red. Note that each subunit consists mainly of a bundle of eight unusually long helices. (c) The binuclear Fe(III) complex of R2.

Each Fe(III) ion is octahedrally coordinated by a His N₆ atom and five O atoms, including those of the O²⁻ ion and the Glu carboxyl group that bridge the two Fe(III) ions. (d) The X-ray structure of the R1 dimer, each subunit of which is in complex with the 20-residue C-terminal peptide of R2 together with GDP in the active site and dTTP in the specificity site. The ATP analog AMPPNP bound in the activity site of the closely similar complex of R1 with the 20-residue peptide and AMPPNP has been superimposed on this structure. The structure is viewed along its 2-fold axis with its two subunits lavender and light green, the two R2 peptides cyan and magenta, and the GDP, dTTP, and ATP shown in space-filling form colored according to atom type (C green, N blue, O red, and P gold). [Parts b and d based on X-ray structures by Hans Eklund, Swedish University of Agricultural Sciences, Uppsala, Sweden. PDBIDs (b) 1RIB and (d) 3R1R and 4R1R.] See the Animated Figures and Interactive Exercise 28

formulate the following catalytic mechanism for *E. coli* RNR (Fig. 28-13):

1. RNR's free radical ($X\cdot$) abstracts an H atom from C3' of the substrate in the reaction's rate-determining step.

2 and 3. Acid-catalyzed cleavage of the C2'—OH bond releases H_2O to yield a radical cation intermediate. The radical mediates the stabilization of the C2' cation by the 3'—OH group's unshared electron pair, thereby accounting for the radical's catalytic role.

4. The radical cation intermediate is reduced by the enzyme's redox-active sulfhydryl pair to yield a 3'-deoxynucleotide radical and a protein disulfide group.

5. The 3' radical reabstracts an H atom from the protein to yield the product deoxynucleoside diphosphate and restore the enzyme to its radical state. A small fraction of the originally abstracted H atom exchanges with solvent before it can be replaced, thus accounting for the release of 3H on reduction of $[3\text{-}^3H]\text{UDP}$.

The Tyr 122 radical in R2 is too far away ($>10\text{ \AA}$) from the enzyme's catalytic site to abstract an electron directly from the substrate. Evidently, the protein mediates electron transfer from this tyrosyl radical to some other group ($X\cdot$ in Fig. 28-13) that is in close proximity to the substrate C3'—H group. Site-directed mutagenesis studies suggest that Cys 439 of R1, in its thiol radical form ($-\text{S}\cdot$), is the most plausible candidate for $X\cdot$ (which makes RNR the only enzyme in which a Cys residue is known to reduce a carbohydrate substrate). Similar studies suggest that Cys 225 and Cys 462 of R1 form the redox-active sulfhydryl pair that directly reduces substrate. Moreover, the resulting disulfide bond is subsequently reduced to regenerate active enzyme via disulfide interchange with Cys 754 and Cys 759 on R1, which are apparently positioned to accept electrons from external reducing agents (see below). Thus, each R1 subunit contains at least five Cys residues that chemically participate in nucleotide reduction.

These observations are confirmed by the X-ray structure of R1 in complex with R2's 20-residue C-terminal

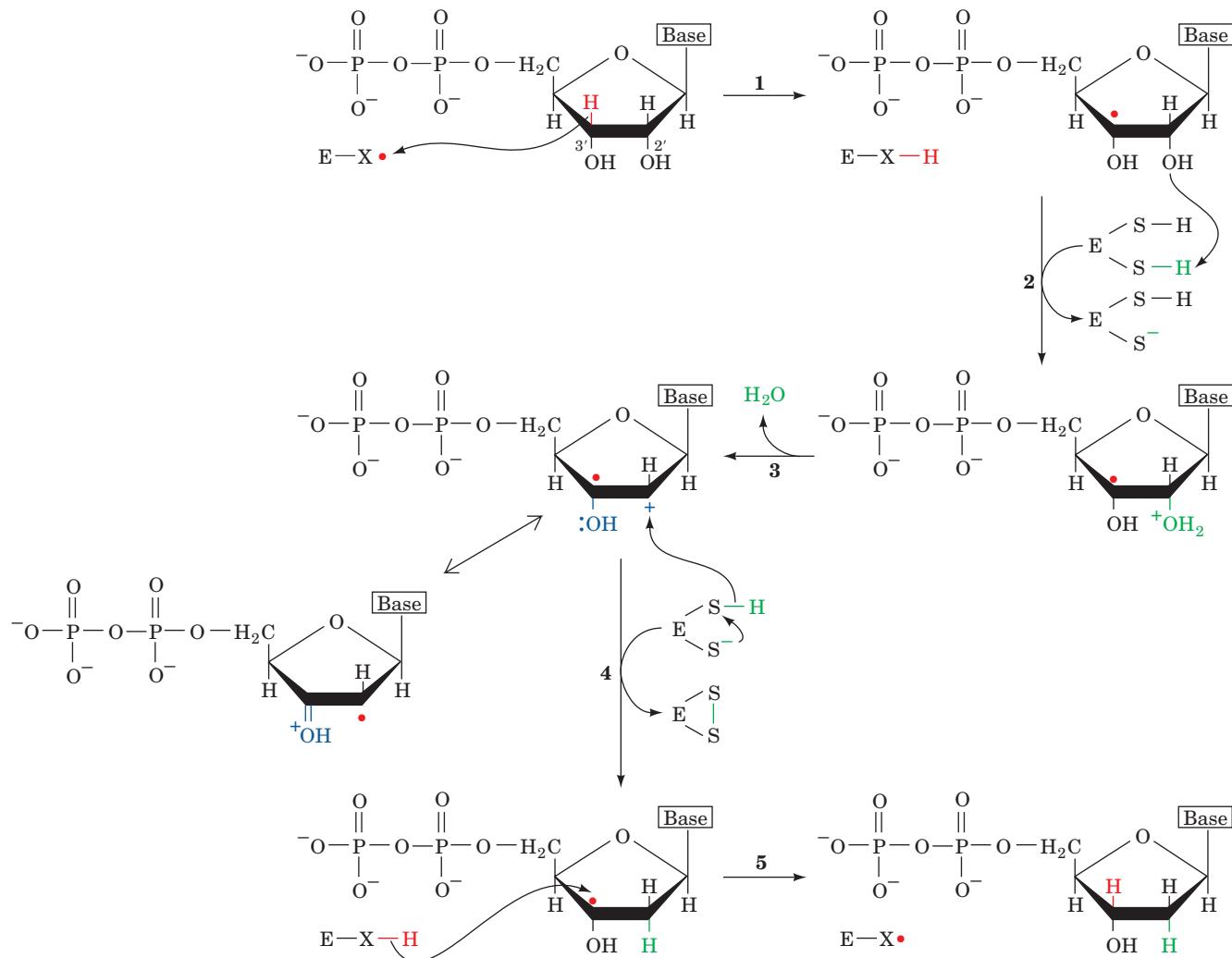


Figure 28-13 Enzymatic mechanism of ribonucleotide reductase. The reaction occurs via a free radical-mediated process in which reducing equivalents are supplied by the

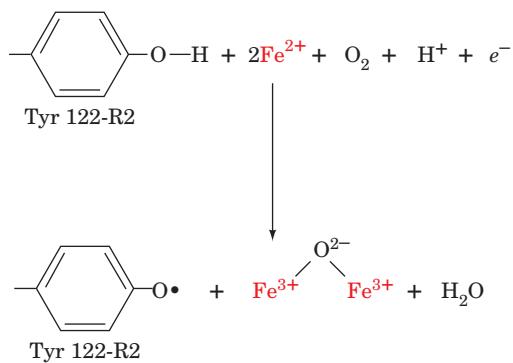
formation of an enzyme disulfide bond. [After Stubbe, J.A., *Biol. Chem.* **265**, 5330 (1990).]

polypeptide (R1 does not crystallize satisfactorily in the absence of this polypeptide), also determined by Eklund (Fig. 28-12d). The central domain of the three-domain R1 monomer consists of a novel 10-stranded α/β barrel that is formed by the antiparallel joining of two topologically similar half-barrels, each comprising five parallel β strands connected by four α helices. As with the similar 8-stranded α/β barrels that form the active sites of numerous enzymes (Section 8-3Bh), R1's active site Cys residues (439, 225, 462) are located in the mouth of the 10-stranded α/β barrel.

The two R1 Cys residues, 754 and 759, that are implicated in the regeneration of the active enzyme are components of R1's C-terminal segment, which is not visible in the X-ray structure of R1 and is presumably disordered. This observation supports the hypothesis that this C-terminal segment acts to flexibly shuttle reducing equivalents from the enzyme surface to its active site.

b. Radical Generation in Class I RNR Requires the Presence of O_2

One of the most remarkable aspects of Class I RNR is its ability to stabilize its normally highly reactive TyrO[•] radical (its half-life is 4 days in the protein vs milliseconds in solution). Yet quenching the radical, say, by hydroxyl-urea, inactivates the enzyme. How, then, is the radical generated in the first place? The radical may be restored *in vitro* by simply treating the inactive enzyme with Fe(II) and a reducing agent in the presence of O_2 .



This is a four-electron reduction of O_2 in which the reducing agent that supplies the electron represented by e^- may be ascorbate or even excess Fe^{2+} .

c. The Inability of Oxidized RNR to Bind Substrate Serves an Important Protective Function

Comparison of the X-ray structures of reduced R1 (in which the redox-active Cys 225 and Cys 462 residues are in their SH forms) with that of oxidized R1 (in which Cys 225 and Cys 462 are disulfide linked) reveals that Cys 462 in reduced R1 has rotated away from its position in oxidized R1 to become buried in a hydrophobic pocket, whereas Cys 225 moves into the region formerly occupied by Cys 462. The distance between the formerly disulfide-linked S atoms thereby increases from 2.0 to 5.7 Å. These movements are accompanied by small shifts of the surrounding polypeptide chain. Oxidized RNR does not bind substrate because its R1 Cys

225 would prevent the binding of substrate through steric interference of its S atom with the substrate NDP's O2' atom.

The inability of oxidized RNR to bind substrate has functional significance. In the absence of substrate, the enzyme's free radical is stored in the interior of the R2 protein, close to its dinuclear iron center. When substrate is bound, the radical is presumably transferred to it via a series of protein side chains in both R2 and R1. If the substrate is unable to properly react after accepting this free radical, as would be the case if RNR were in its oxidized state, this could result in the destruction of the substrate and/or the enzyme. Indeed, the mutation of the redox-active Cys 225 to Ser results in an enzyme that permits the formation of the substrate radical (Fig. 28-13); however, since the mutant enzyme is incapable of reducing it, the substrate radical instead decomposes followed by the release of its base and phosphate moieties. More importantly, a transient peptide radical forms, which cleaves and inactivates the R1 polypeptide chain while consuming the radical and thereby inactivating R2. Thus, an important role of the enzyme is to control the release of the radical's powerful oxidizing capability. It does so in part by preventing the binding of substrate while the enzyme is in its oxidized form.

d. Ribonucleotide Reductase Is Regulated by Effector-Induced Oligomerization

The synthesis of the four dNTPs in the amounts required for DNA synthesis is accomplished through feedback control. The maintenance of the proper intracellular ratios of dNTPs is essential for normal growth. Indeed, *a deficiency of any dNTP is lethal, whereas an excess is mutagenic because the probability that a given dNTP will be erroneously incorporated into a growing DNA strand increases with its concentration relative to those of the other dNTPs.*

The activities of both *E. coli* and mammalian Class I RNRs are allosterically responsive to the concentrations of various (d)NTPs. Thus, as Reichard has shown, ATP induces the reduction of CDP and UDP; dTTP induces the reduction of GDP and inhibits the reduction of CDP and UDP; dGTP induces the reduction of ADP and, in mammals but not *E. coli*, inhibits the reduction of CDP and UDP; and dATP inhibits the reduction of all NDPs.

Barry Cooperman has shown that the catalytic activity of mouse RNR varies with its state of oligomerization (that is, RNR is a morphein; Section 26-4Ac), which in turn is governed by the binding of nucleotide effectors to three independent allosteric sites on R1: (1) the specificity site, which binds ATP, dATP, dGTP, and dTTP; (2) the activity site, which binds ATP and dATP; and (3) the hexamerization site, which binds only ATP. On the basis of molecular mass, ligand binding, and activity studies on mouse RNR, Cooperman formulated a model that quantitatively accounts for the allosteric regulation of Class I RNR. It has the following features (Fig. 28-14a):

1. The binding of ATP, dATP, dGTP, or dTTP to the specificity site induces the catalytically inactive R1 monomers to form a catalytically active dimer, R1₂.

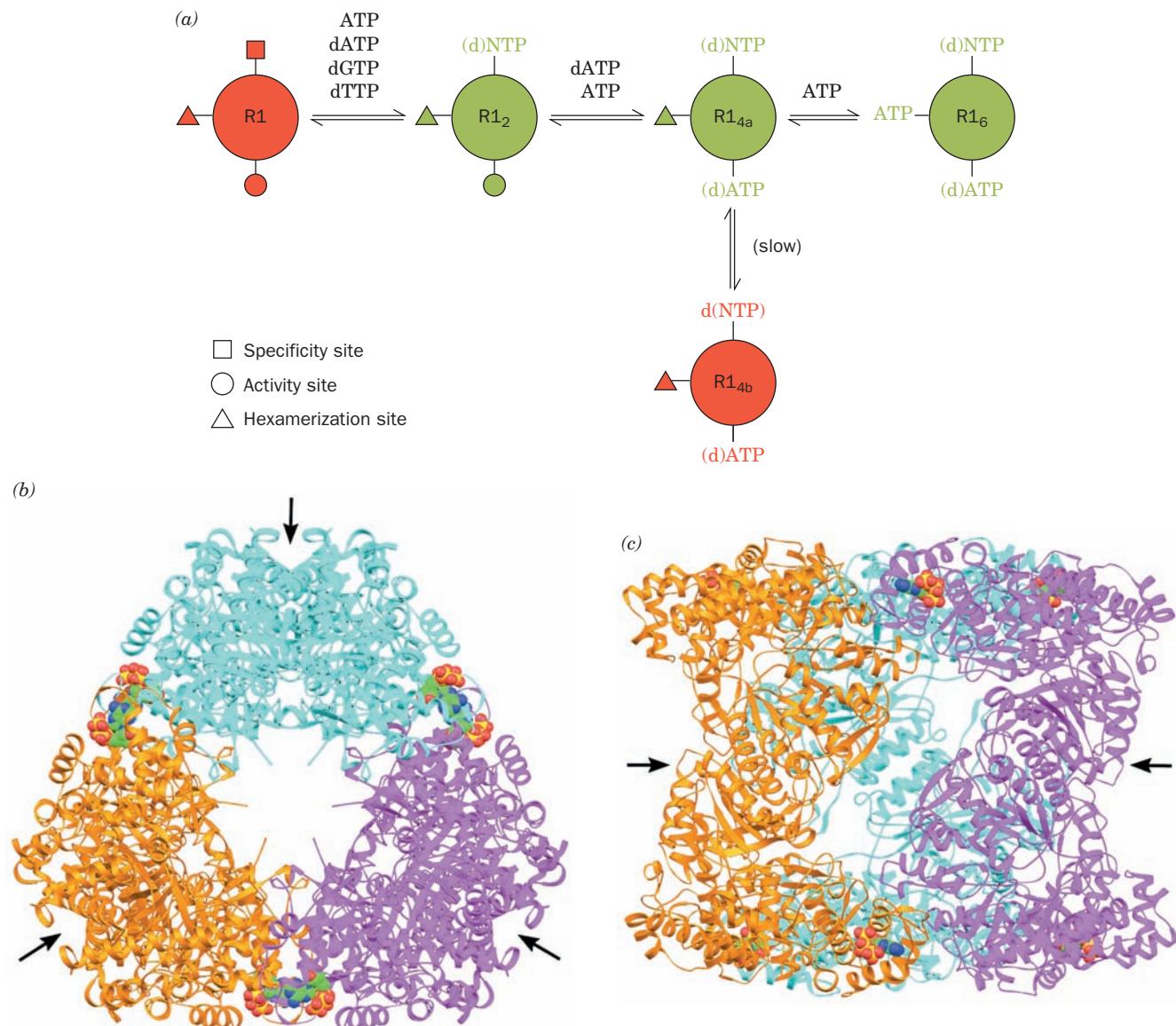


Figure 28-14 Ribonucleotide reductase regulation. **(a)** A model for the allosteric regulation of Class I RNR via its oligomerization. States shown in green have high activity and those shown in red have little or no activity. R2 has been omitted for simplicity. [After Kashlan, O.B., Scott, C.P., Lear, J.D., and Cooperman, B.S., *Biochemistry* **41**, 461 (2002).] **(b)** The X-ray structure of the R1 hexamer, which has D_3 symmetry, in complex with AMPPNP as viewed along its 3-fold axis. Each of its three dimers are differently colored (the X-ray structure of a dimer is

shown in Fig. 28-12d). The AMPPNP, which binds to the enzyme's activity sites, is drawn in space-filling form with C green, N blue, O red, and P gold. The black arrows point along the R1 dimers' 2-fold axes and indicate the probable docking sites for the binding of R2 dimers. **(c)** The R1 · AMPPNP hexamer as viewed along the vertical 2-fold axis in Part **b**. [Parts **b** and **c** based on an X-ray structure by Hans Eklund, Swedish University of Agricultural Sciences, Uppsala, Sweden. PDBid 3R1R.]

2. The binding of dATP or ATP to the activity site causes the dimers to form catalytically active tetramers, R1_{4a}, that slowly but reversibly change conformation to a catalytically inactive state, R1_{4b}.

3. The binding of ATP to the hexamerization site induces the tetramers to further aggregate to form catalytically active hexamers, R1₆, RNR's major active form.

The concentration of ATP in a cell is such that, *in vivo*, R1 is almost entirely in its tetrameric or hexameric forms.

As a consequence, ATP couples the overall rate of DNA synthesis to the cell's energy state.

The specificity and activity sites have been located in X-ray structures of *E. coli* R1 (Fig. 28-12d); the hexamerization site has not yet been identified. The R1 hexamer had, in fact, been previously observed in the X-ray structures of R1 (Fig. 28-14b,c), but the interactions between its contacting dimers are so tenuous that it was assumed that they are merely artifacts of crystallization with no physiological significance. Yet, since the activity site is located at this contact

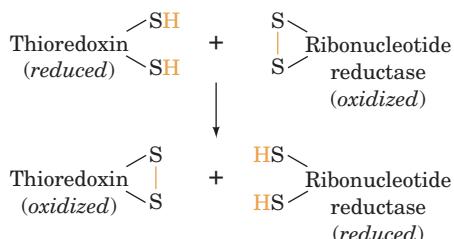
site, it now seems likely that its binding of (d)ATP induces R1 oligomerization through local conformational changes.

The foregoing model has, for simplicity, neglected the presence of R2 subunits although, of course, R1 and R2 must be present in equimolar amounts in the active enzyme. Presumably, the R1 and R2 dimers bind to one another such that their 2-fold axes coincide. The lack of space on the inside of the R1 hexamer dictates that the R2 dimers must contact the R1 dimers from outside the hexamer (Fig. 28-14b).

dCTP is not an effector of RNR. This is presumably because the intracellular balance between dCTP and dTTP is not controlled by RNR but, rather, is maintained by **deoxycytidine deaminase**, which converts dCTP to dUMP, the precursor of dTTP. This enzyme is activated by dCTP and inhibited by dTTP.

e. Thioredoxin and Glutaredoxin Are Class I Ribonucleotide Reductase's Physiological Reducing Agents

The final step in the RNR catalytic cycle is the reduction of the enzyme's newly formed disulfide bond to reform its redox-active sulphydryl pair. Dithiols such as that of 2-mercaptoethanol (Section 7-1B) can serve as the reducing agent for this process *in vitro* through a disulfide interchange reaction. One of the enzyme's physiological reducing agents, however, is **thioredoxin (Trx)**, a ubiquitous monomeric 105-residue protein that has a pair of closely proximal redox-active Cys residues, Cys 32 and Cys 35 (we have previously encountered thioredoxin in our study of the light-induced activation of the Calvin cycle; Section 24-3B). Thioredoxin reduces oxidized RNR via disulfide interchange.



The X-ray structure of reduced *E. coli* Trx (Fig. 28-15) reveals that the side chain of the redox-active Cys 32 is exposed on the protein's surface, where it is available for oxidation. Oxidized thioredoxin is, in turn, reduced by

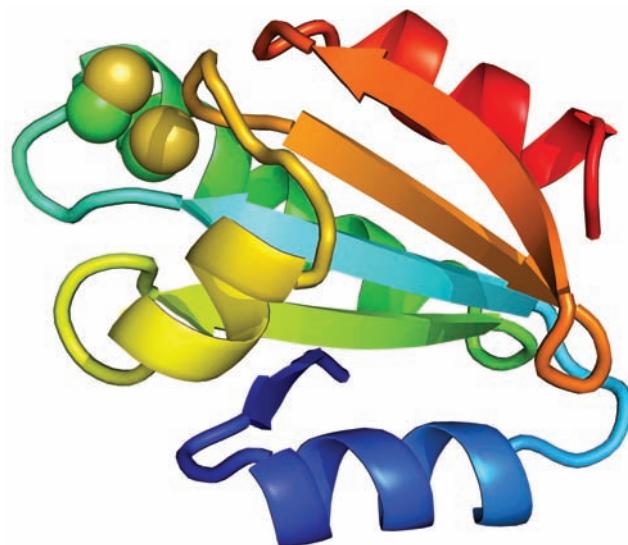


Figure 28-15 X-ray structure of human thioredoxin in its reduced (sulphydryl) state. The 105-residue polypeptide chain is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the redox-active residues, Cys 32 and Cys 35, are shown in space-filling form with C green and S yellow. This structure closely resembles those of the homologous α and α' domains of protein disulfide isomerase (PDI, Fig. 9-16). [Based on an X-ray structure by William Montfort, University of Arizona. PDBid 1ERT.]

NADPH in a reaction mediated by the flavoprotein **thioredoxin reductase**. NADPH therefore serves as the terminal reducing agent in the RNR-mediated reduction of NDPs to dNDPs (Fig. 28-16).

The existence of a viable *E. coli* mutant devoid of thioredoxin indicates that this protein is not the only substance capable of reducing oxidized RNR *in vivo*. This observation led to the discovery of **glutaredoxin**, a disulfide-containing, monomeric, 85-residue protein that can also reduce RNR (mutants devoid of both thioredoxin and glutaredoxin are nonviable). Oxidized glutaredoxin is reduced, via disulfide interchange, by the Cys-containing tripeptide glutathione which, in turn, is reduced by NADPH as catalyzed by glutathione reductase (GR; Section 21-2Ba). The relative

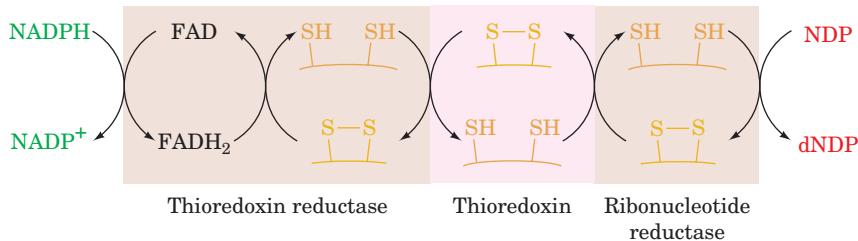


Figure 28-16 Electron-transfer pathway for nucleoside diphosphate (NDP) reduction. NADPH provides the reducing

equivalents for this process through the intermediacy of thioredoxin reductase, thioredoxin, and ribonucleotide reductase.

importance of thioredoxin and glutaredoxin in the reduction of RNRs remains to be established.

f. Thioredoxin Reductase Alternates Its Conformation with Its Redox State

Thioredoxin reductase (**TrxR**), a homodimer of 316-residue subunits, is a homolog of GR that catalyzes a similar reaction: the reduction of a substrate disulfide bond by NADPH as mediated by an FAD prosthetic group and a redox active sulphydryl pair (Cys 135 and Cys 138). However, the X-ray structure of the C138S mutant of *E. coli* TrxR in complex with NADP⁺ (Fig. 28-17a), determined by Charles Williams and John Kuriyan, reveals that TrxR and GR differ in their active site arrangements such that their redox-active sulphydryl pairs are on opposite sides of the flavin rings in the two enzymes. Nevertheless, TrxR's redox-active sulphydryl pair appears properly positioned to reduce the flavin ring. However, the NADP⁺'s nicotinamide ring is >17 Å from the flavin ring and the redox-active sulphydryl pair is buried such that it could not react with the enzyme's Trx substrate. How then does TrxR manage to transfer an electron pair from its bound NADPH via its flavin ring and redox-active sulphydryl pair to Trx?

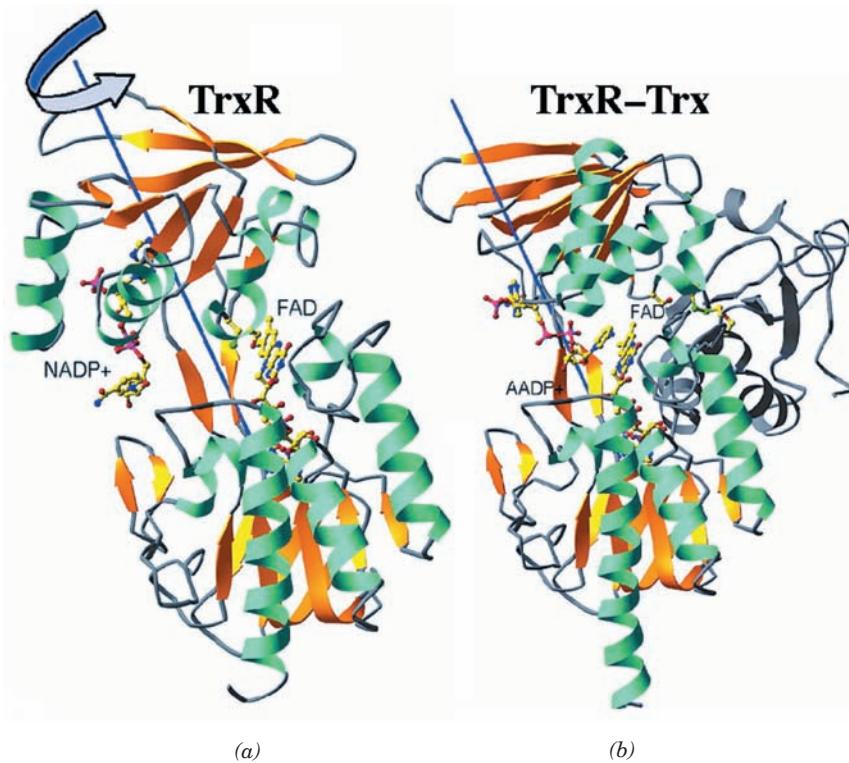
This question was answered by Williams and Martha Ludwig through their X-ray structure determination of the C135S mutant of TrxR, whose Cys 138 is disulfide-linked to Cys 32 of the C35S mutant of *E. coli* Trx (probably the physiologically relevant disulfide bond) and which is in complex with the NADP⁺ analog **3-aminopyridine adenine dinucleotide phosphate (AADP⁺)**. In this complex (Fig. 28-17b), TrxR's NADP⁺-binding domain has rotated by 67°

relative to the rest of the protein compared to its position in TrxR alone (Fig. 28-17a). This positions the AADP⁺'s pyridine ring to react with the flavin ring and positions TrxR's redox-active sulphydryl pair to undergo a disulfide interchange reaction with that of Trx. Moreover, in this latter conformation, the NADP⁺-binding domain appears to provide the recognition site for the substrate Trx. Evidently, TrxR alternates its conformation with each successive step in the process of transferring an electron pair from NADPH to the flavin to its redox-active sulphydryl pair to its bound Trx substrate. This added mechanistic complication relative to that of GR, which does not undergo a significant conformational change in reducing glutathione disulfide (Section 21-2Ba), has apparently evolved to permit TrxR to reduce its protein substrate: Trx would be too large for its redox-active sulphydryl pair to properly approach the active site sulphydryl pair in a GR-like enzyme.

g. The Three Classes of Ribonucleotide Reductases Are Evolutionarily Related

We have seen that the active forms of Class I RNRs are R1₂R2₂, R1₄R2₄, and R1₆R2₆ oligomers that have mechanistically essential tyrosyl radicals that are stabilized by oxo-bridged binuclear Fe(III) complexes, have NDPs for substrates, and obtain their reducing equivalents from thioredoxin and glutaredoxin. In contrast, Class II RNRs, which are α monomers or α_2 dimers, utilize a 5'-deoxyadenosylcobalamin cofactor (coenzyme B₁₂; Section 25-2E_b) for radical generation, have NDPs for substrates, and are reduced by thioredoxin and glutaredoxin; whereas Class III RNRs,

Figure 28-17 X-ray structures of *E. coli* thioredoxin reductase (TrxR). (a) The C138S mutant TrxR in complex with NADP⁺. The protein is shown in ribbon form colored according to its secondary structure. The NADP⁺, the FAD, and the side chains of Cys 135 and Ser 138 are drawn in ball-and-stick form with C yellow, N blue, O red, S green, and P magenta. (b) The C135S mutant TrxR in complex with AADP⁺ and covalently linked to the C35S mutant of Trx via a disulfide bond between TrxR Cys 138 and Trx Cys 32. The TrxR is represented as in Part a, the Trx ribbon is blue-gray, and its Cys 32 and Ser 35 side chains are drawn in ball-and-stick form. Comparison of these two structures reveals that TrxR's NADP⁺-binding domain (residues 120–243) undergoes a 67° rotation about the axis drawn in blue relative to the rest of the protein, which is shown in the same orientation in both structures. [Courtesy of Martha Ludwig, University of Michigan. PDBIDs (a) 1TDF and (b) 1F6M.]



which are α_2 dimers that interact with a radical-generating protein β_2 that contains a [4Fe–4S] cluster and requires S-adenosylmethionine (SAM; Section 26-3Ea) and NADPH for activity, have NTPs for substrates, and their reducing equivalents are provided by the oxidation of formate to CO_2 .

Since all known cellular life synthesizes its deoxyribonucleotides from ribonucleotides, the rise of an RNR must have preceded the evolutionary transition from the RNA world (Section 1-5Ca) to DNA-based life-forms. Did the three classes of RNRs arise independently or are they evolutionarily related? Despite the seemingly large differences between these different classes of RNRs, the reactions they catalyze are surprisingly similar. All replace the 2' OH group of ribose with H via a free radical mechanism involving a thiy radical with the reducing equivalents provided by a Cys sulphydryl group (Fig. 28-13; the second Cys residue of the redox-active sulphydryl pair in Class I and II RNRs is replaced by formate in Class III RNRs). They differ mainly in the way they generate the free radical. [In Class II RNRs, the radical is generated by the homolytic cleavage of its 5'-deoxyadenosylcobalamin cofactor's C—Co(III) bond (Section 25-2Ec). In Class III RNRs, it is generated by the NADPH-supplied and [4Fe—4S] cluster-mediated one-electron reductive cleavage of SAM by the β_2 protein to yield methionine and the 5'-deoxyadenosyl radical (the same radical generated by the homolytic cleavage of 5'-deoxyadenosylcobalamin), which then abstracts the H atom from a C_α —H group of a specific Gly on the α subunit to yield 5'-deoxyadenosine and a stable but O_2 -sensitive glycyl radical.] Moreover, the X-ray structures of both Class II and Class III RNRs reveal that their active sites are formed by 10-stranded α/β barrels that have the same connectivity as and are closely superimposable on that of Class I RNRs. It therefore appears that all three classes of RNRs are evolutionarily related. Reichard has proposed that, since life arose under anaerobic conditions and that formate, one of simplest organic reductants, was probably widely available on primitive Earth (Section 1-5B), the primordial RNR was a Class III-like enzyme. The rise of photosynthetic organisms that generated O_2 then promoted the evolution of Class II RNRs, which can function under both anaerobic and aerobic conditions. Class I RNRs, which require the presence of O_2 for activation, evolved last, presumably from a Class II RNR.

h. dNTPs Are Produced by Phosphorylation of dNDPs

In pathways involving Class I and Class II RNRs, the final step in the production of dNTPs is the phosphorylation of the corresponding dNDPs:



This reaction is catalyzed by nucleoside diphosphate kinase, the same enzyme that phosphorylates NDPs (Section 28-1Ba). As before, the reaction is written with ATP as the phosphoryl donor, although any NTP or dNTP can function in this capacity. In pathways involving Class III RNRs, the production of NTPs from NDPs precedes the reduction of NTPs to dNTPs.

B. Origin of Thymine

a. dUTP Diphosphohydrolase

The dTMP component of DNA is synthesized, as we discuss below, by methylation of dUMP. The dUMP is generated through the hydrolysis of dUTP by **dUTP diphosphohydrolase (dUTPase; also called dUTP pyrophosphatase)**:

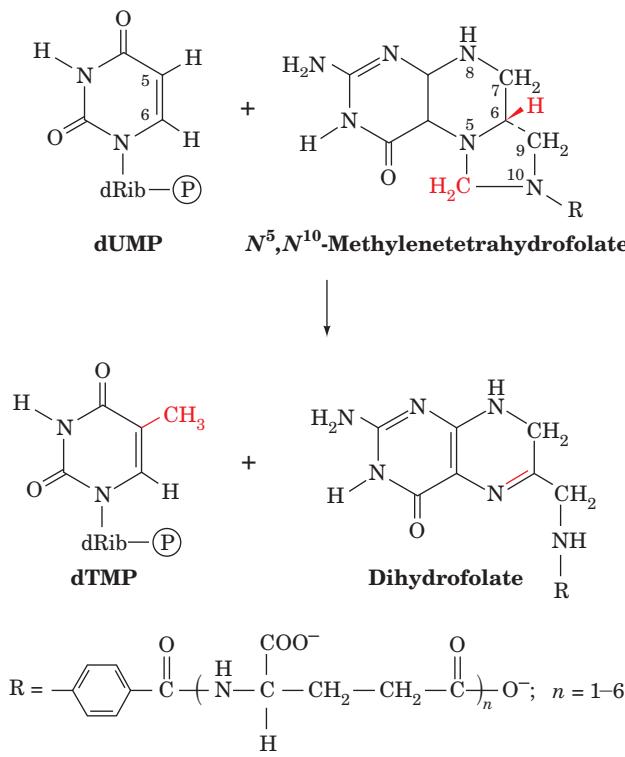


The reason for this apparently energetically wasteful process (dTMP, once formed, is rephosphorylated to dTTP) is that cells must minimize their concentration of dUTP in order to prevent incorporation of uracil into their DNA. This is because, as we discuss in Section 30-5Bd, DNA polymerase does not discriminate between dUTP and dTTP.

The X-ray structure of human dUTPase, determined by John Tainer, reveals the basis for this enzyme's exquisite specificity for dUTP. This homotrimer of 141-residue subunits binds dUTP in a snug-fitting cavity that sterically excludes thymine's C5 methyl group via the side chains of conserved residues (Fig. 28-18a). It differentiates uracil from cytosine via a set of hydrogen bonds from the protein backbone that mimic adenine's base-pairing interaction (Fig. 28-18b), and it differentiates dUTP from UTP by the steric exclusion of ribose's 2' OH group by the side chain of a conserved Tyr.

b. Thymidylate Synthase

dTMP is synthesized from dUMP by thymidylate synthase (TS) with N^5,N^{10} -methylenetetrahydrofolate (N^5,N^{10} -methylene-THF) as the methyl donor:



(THF cofactors are discussed in Section 26-4D). Note that the transferred methylene group (in which the carbon has the oxidation state of formaldehyde) is reduced to a methyl group (which has the oxidation state of methanol) at the

expense of the oxidation of the THF cofactor to dihydrofolate (DHF).

The catalytic mechanism of TS, a highly conserved homodimeric protein (with 264-residue subunits in *E. coli*), has been extensively investigated. On incubation of the enzyme with N^5,N^{10} -methylene-[6- 3 H]THF and dUMP, the 3 H is quantitatively transferred to the methyl group of the product dTMP. When [5- 3 H]dUMP is the substrate, however, the 3 H is released into the aqueous solvent. Such information, together with the knowledge that uracil C6, which occupies the β position of an α,β -unsaturated ketone, is susceptible to nucleophilic attack, led Daniel Santi to propose the following mechanistic scheme for the TS reaction (Fig. 28-19):

1. An enzyme nucleophile, identified as the thiolate group of Cys 146, attacks C6 of dUMP to form a covalent adduct.

2. C5 of the resulting enolate ion attacks the CH_2 group of the iminium cation in equilibrium with N^5,N^{10} -methylene-THF to form an enzyme-dUMP-THF ternary covalent complex.

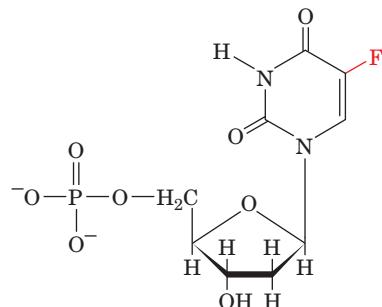
3. An enzyme base abstracts the acidic proton at the C5 position of the enzyme-bound dUMP, forming an exocyclic methylene group and eliminating the THF cofactor. The abstracted proton subsequently exchanges with solvent.

4. The redox change occurs via the migration of the N6—H atom of THF as a hydride ion to the exocyclic methylene group, converting it to a methyl group (thus

accounting for the above described transfer of 3 H) and yielding DHF. This reduction promotes displacement of the Cys thiolate group from the intermediate so as to release product, dTMP, and reform active enzyme.

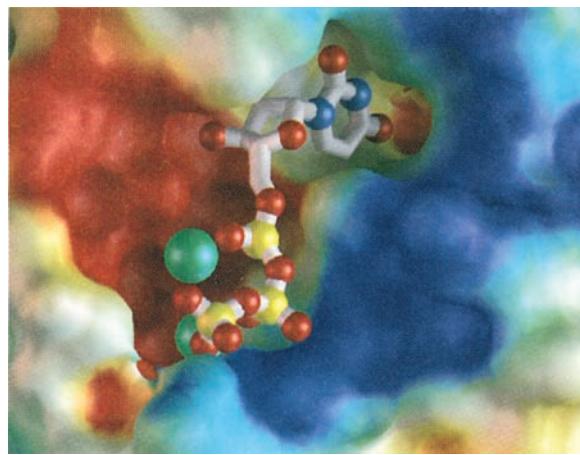
c. 5-Fluorodeoxyuridylate Is a Potent Antitumor Agent

The above mechanism is supported by the observation that **5-fluorodeoxyuridylate (FdUMP)**



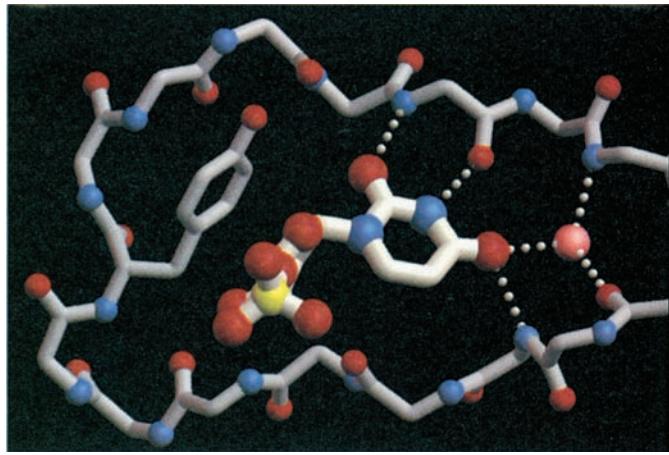
5-Fluorodeoxyuridylate (FdUMP)

is an irreversible inhibitor of TS. This substance, like dUMP, binds to the enzyme (an F atom has approximately the same radius as an H atom) and undergoes the first two steps of the normal enzymatic reaction. In Step 3, however, the enzyme cannot abstract the F atom as F^+ (recall that F is the most electronegative element), so that the enzyme is all but permanently immobilized as the enzyme-FdUMP-THF



(a)

Figure 28-18 X-ray structure of human dUTPase. (a) The molecular surface at the substrate binding site showing how the enzyme differentiates uracil from thymine. Bound dUTP is drawn in stick form with its N, O, and P atoms represented by blue, red, and yellow spheres. Mg^{2+} ions that were modeled into the structure are represented by green spheres. The protein's molecular surface is colored according to its electrostatic potential with positive, negative, and near neutral regions blue, red, and white, respectively. Note how the snug fit of the uracil ring into its binding site would sterically exclude thymine's C5 methyl group. (b) The substrate binding site indicating how the



(b)

enzyme differentiates uracil from cytosine and 2'-deoxyribose from ribose. dUTP bound at the active site is drawn as in Part a. The protein, mainly the backbone of a β hairpin motif, is similarly drawn but with thinner gray bonds. Hydrogen bonds are shown as dotted white lines, and a tightly bound, conserved water molecule is represented by a pink sphere. The pattern of hydrogen bonding donors and acceptors on the protein would prevent cytosine from binding in the active site pocket. The conserved Tyr side chain sterically excludes ribose's 2' OH group. [Courtesy of John Tainer, The Scripps Research Institute, La Jolla, California.]

Figure 28-19 Catalytic mechanism of thymidylate synthase. The methyl group is supplied by N^5,N^{10} -methylene-THF, which is concomitantly oxidized to dihydrofolate.

ternary covalent complex analogous to that after Step 2 in Fig. 28-19. Indeed, X-ray structural analysis by William Montfort revealed that crystals of *E. coli* TS that had been soaked in a solution containing FdUMP and N^5,N^{10} -methylene-THF contain precisely this complex (Fig. 28-20). Enzymatic inhibitors such as FdUMP, which inactivate an enzyme only after undergoing part or all of its normal catalytic reaction, are called **mechanism-based inhibitors** (alternatively, **suicide substrates** because they cause the enzyme to “commit suicide”). *Mechanism-based inhibitors, being targeted for particular enzymes, are among the most powerful, specific, and therefore useful enzyme inactivators.*

The strategic position of thymidylate synthase in DNA biosynthesis has led to the clinical use of FdUMP as an antitumor agent. Rapidly proliferating cells, such as cancer cells, require a steady supply of dTMP in order to survive and are therefore killed by treatment with FdUMP. In contrast, most normal mammalian cells, which grow slowly if at all, have a lesser requirement for dTMP, so that they are

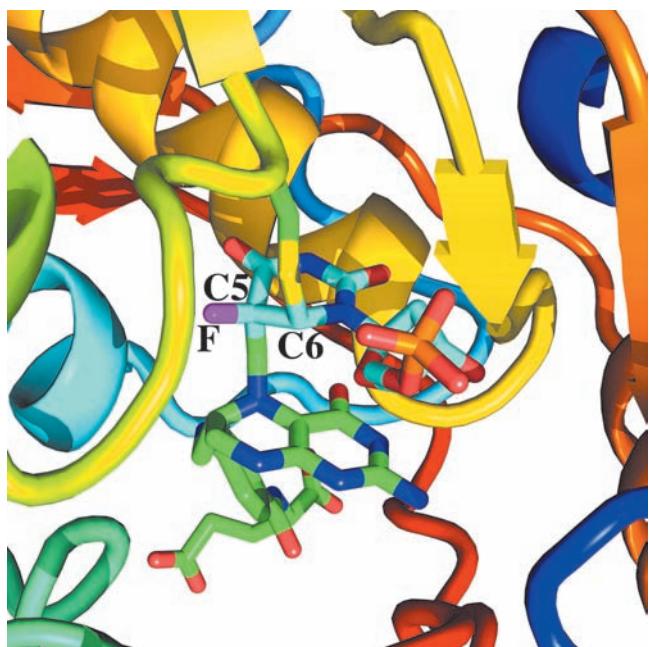
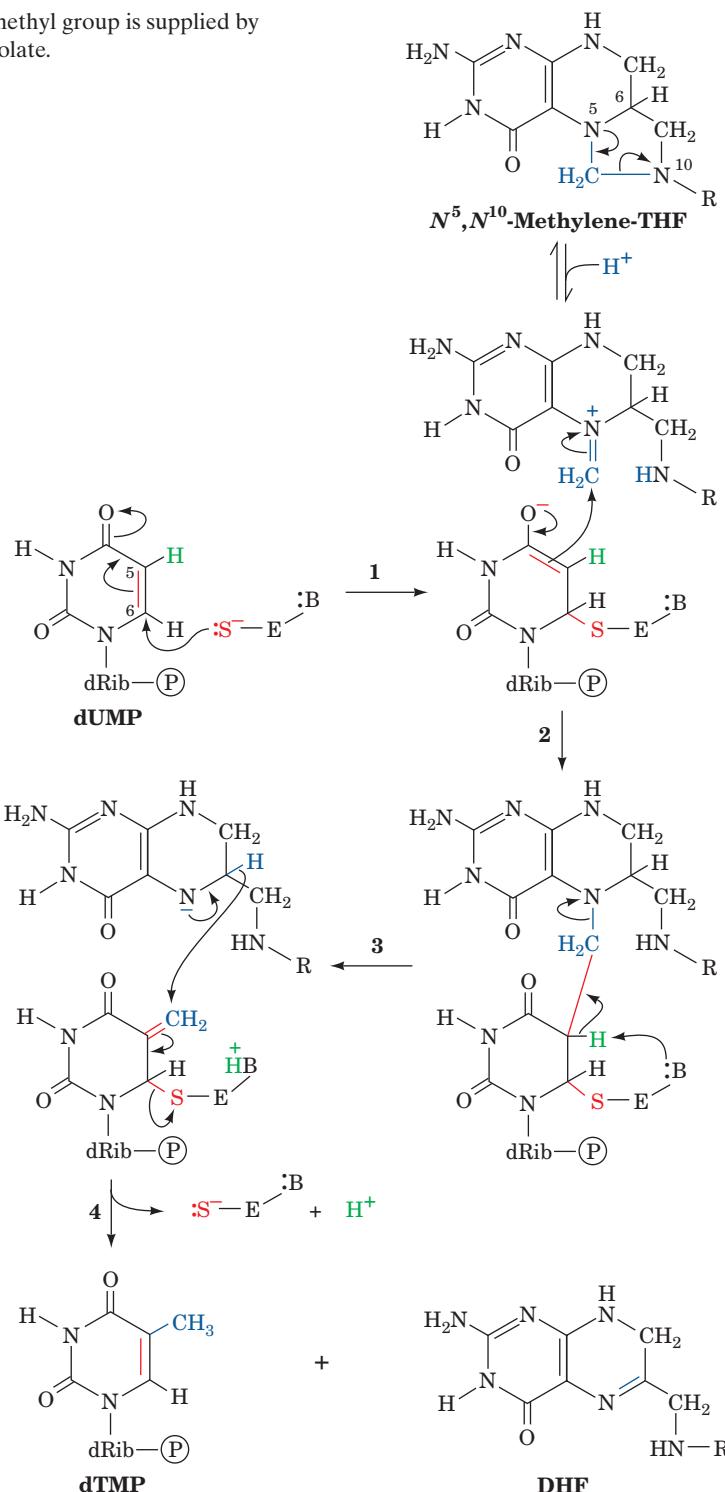


Figure 28-20 The X-ray structure of the *E. coli* thymidylate synthase in covalent complex with FdUMP and THF. The active site region of one subunit of this dimeric enzyme is shown in ribbon form with its polypeptide chain colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The FdUMP and N^5,N^{10} -methylene-THF are drawn in stick form with FdUMP C cyan, N^5,N^{10} -methylene-THF C green, N blue, O red, F purple, and P orange. The C5 and C6 atoms of FdUMP form covalent bonds with the CH_2 group substituent to N5 of THF and the S atom of Cys 146, whose side chain is drawn in stick form with C green and S yellow. [Based on an X-ray structure by William Montfort, University of Arizona, PDBid 1TSN.]



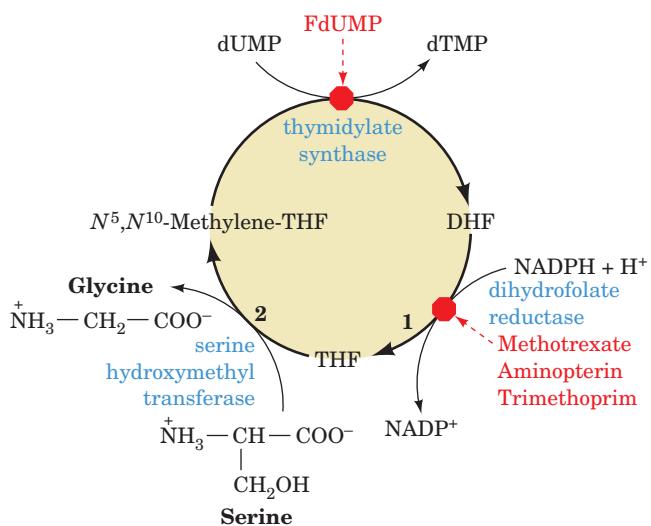


Figure 28-21 **Regeneration of N^5,N^{10} -methylenetetrahydrofolate.** The DHF product of the thymidylate synthase reaction is converted back to N^5,N^{10} -methylene-THF by the sequential actions of (1) dihydrofolate reductase and (2) serine hydroxymethyltransferase. Thymidylate synthase is inhibited by FdUMP, whereas dihydrofolate reductase is inhibited by the antifolates methotrexate, aminopterin, and trimethoprim.

relatively insensitive to FdUMP (some exceptions are the bone marrow cells that comprise the blood-forming tissues and much of the immune system, the intestinal mucosa, and hair follicles). **5-Fluorouracil** and **5-fluorodeoxyuridine** are also effective antitumor agents since they are converted to FdUMP through salvage reactions.

d. N^5,N^{10} -Methylene-THF Is Regenerated in Two Reactions

The thymidylate synthase reaction is biochemically unique in that it oxidizes THF to DHF; no other enzymatic reaction employing a THF cofactor alters this coenzyme's net oxidation state. The DHF product of the thymidylate synthase reaction is recycled to the enzyme's N^5,N^{10} -methylene-THF cofactor through two sequential reactions (Fig. 28-21):

1. DHF is reduced to THF by NADPH as catalyzed by **dihydrofolate reductase (DHFR)** (Section 26-4D). Although, in most organisms, DHFR is a monomeric monofunctional enzyme, in protozoa and at least some plants, DHFR and TS occur on the same polypeptide chain to form a bifunctional enzyme that has been shown to channel DHF from its TS to its DHFR active sites.

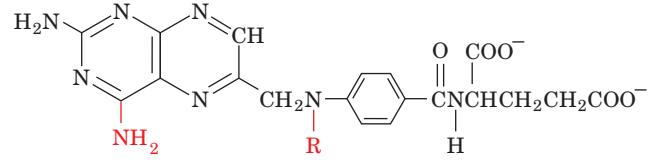
2. Serine hydroxymethyltransferase (Section 26-3Bb) transfers the hydroxymethyl group of serine to THF, yielding N^5,N^{10} -methylene-THF and glycine.

e. Antifolates Are Anticancer Agents

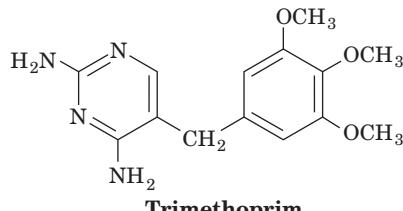
Inhibition of DHFR quickly results in all of a cell's limited supply of THF being converted to DHF by the thymidylate synthase reaction. Inhibition of DHFR therefore not only prevents dTMP synthesis (Fig. 28-21), but also blocks all other THF-dependent biological reactions such as the

synthesis of purines (Section 28-1A), methionine (Section 26-5Ba), and, indirectly, histidine (Section 26-5Be). DHFR (Fig. 28-22) therefore offers an attractive target for chemotherapy.

Methotrexate (amethopterin), aminopterin, and trimethoprim



$R = H$ **Aminopterin**
 $R = CH_3$ **Methotrexate (amethopterin)**



Trimethoprim

are DHF analogs that competitively although all but irreversibly bind to DHFR with an ~1000-fold greater affinity than does DHF. These **antifolates** (substances that interfere with the action of folate cofactors) are effective anticancer agents, particularly against childhood leukemias. In fact, a successful chemotherapeutic strategy is to treat a cancer

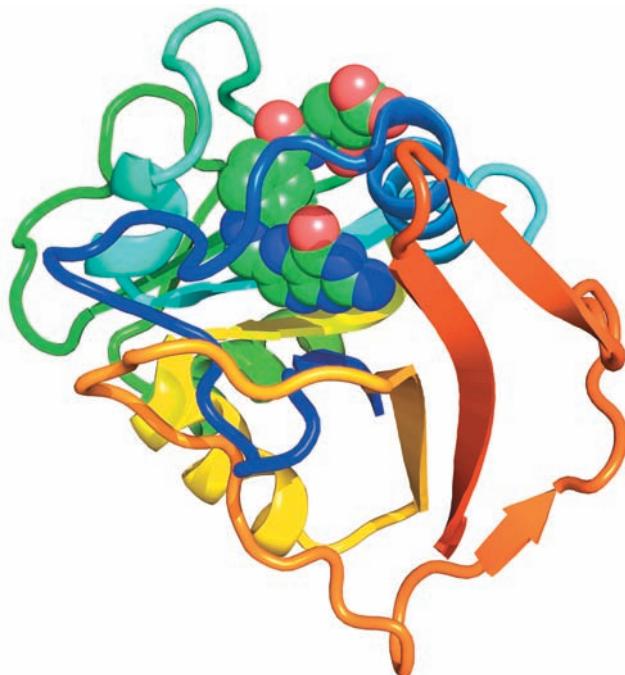


Figure 28-22 **X-ray structure of human dihydrofolate reductase in complex with folic acid.** The polypeptide is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The folic acid is drawn in space-filling form with C green, N blue, and O red. [Based on an X-ray structure by Joseph Kraut, University of California at San Diego. PDBid 1DHF.]

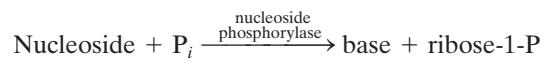
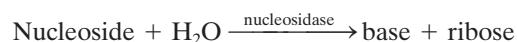
See Interactive Exercise 29.

victim with a lethal dose of methotrexate and some hours later “rescue” the patient (but hopefully not the cancer) by administering massive doses of 5-formyl-THF and/or thymidine. A low dose of methotrexate is also effective in the treatment of rheumatoid arthritis, inhibiting immune system activity and thus decreasing inflammation. Trimethoprim, which was discovered by George Hitchings and Gertrude Elion, binds much more tightly to bacterial DHFRs than to those of mammals and is therefore a clinically useful antibiotic.

4 NUCLEOTIDE DEGRADATION

Most foodstuffs, being of cellular origin, contain nucleic acids. Dietary nucleic acids survive the acid medium of the stomach; they are degraded to their component nucleotides, mainly in the duodenum, by pancreatic nucleases and intestinal phosphodiesterases. These ionic compounds, which cannot pass through cell membranes, are then hydrolyzed to nucleosides by a variety of group-specific nucleotidases and nonspecific phosphatases. Nucleosides may be directly absorbed by the intestinal mucosa or first undergo further degradation to free bases and ribose or

ribose-1-phosphate through the action of **nucleosidases** and **nucleoside phosphorylases**:



Radioactive labeling experiments have demonstrated that only a small fraction of the bases of ingested nucleic acids are incorporated into tissue nucleic acids. Evidently, the *de novo* pathways of nucleotide biosynthesis largely satisfy an organism’s need for nucleotides. Consequently, ingested bases, for the most part, are degraded and excreted. Cellular nucleic acids are also subject to degradation as part of the continual turnover of nearly all cellular components. In this section we outline these catabolic pathways and discuss the consequences of several of their inherited defects.

A. Catabolism of Purines

The major pathways of purine nucleotide and deoxynucleotide catabolism in animals are diagrammed in Fig. 28-23. Other organisms may have somewhat different

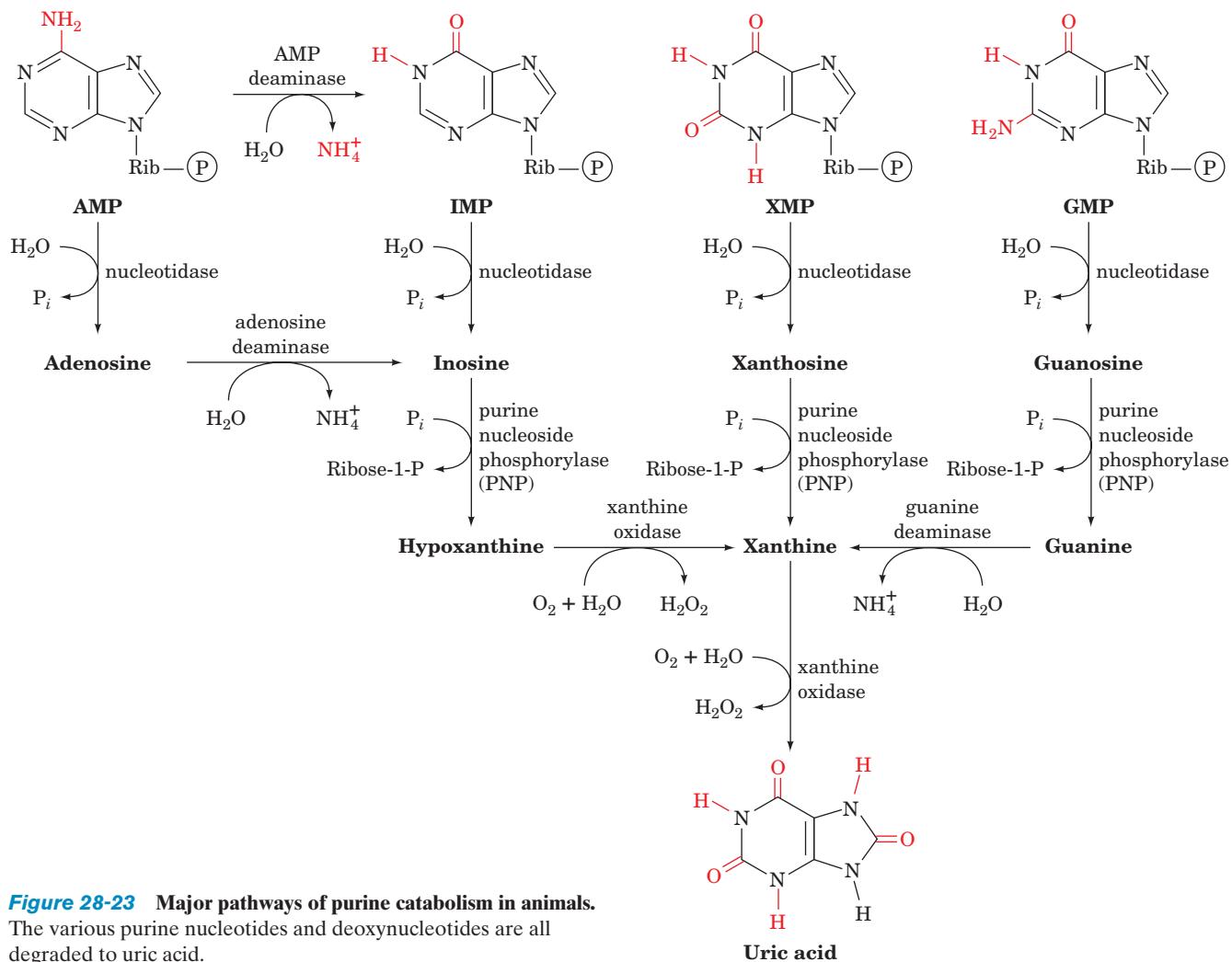


Figure 28-23 Major pathways of purine catabolism in animals.

The various purine nucleotides and deoxynucleotides are all degraded to uric acid.

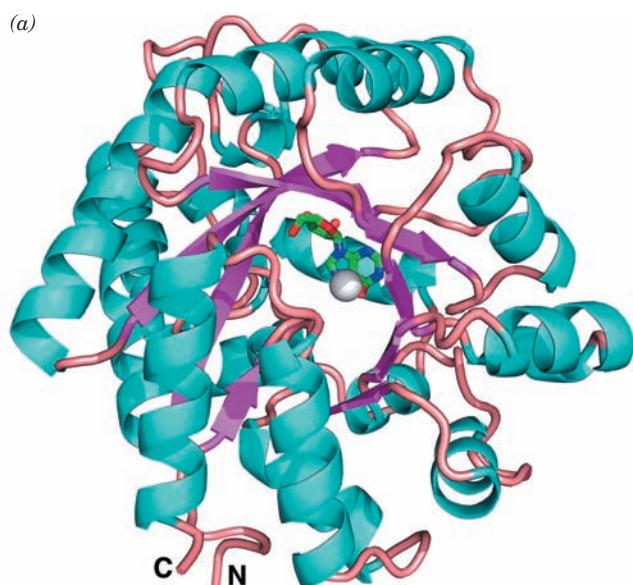
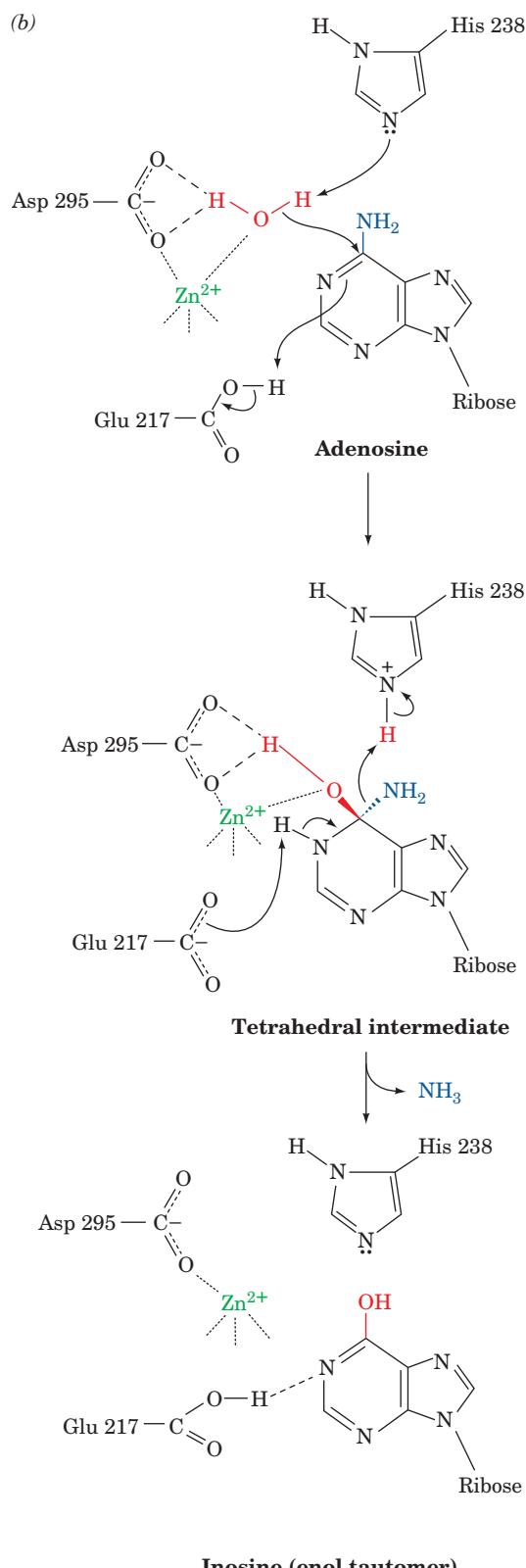


Figure 28-24 X-ray structure and mechanism of adenosine deaminase. (a) A ribbon diagram of murine adenosine deaminase in complex with its transition state analog 6-hydroxy-1,6-dihydropurine ribonucleoside (HDPR). The polypeptide is drawn in ribbon form colored according to its secondary structure (helices cyan, β strands magenta, and loops salmon) and viewed approximately down the axis of the enzyme's α/β barrel from the N-terminal ends of its β strands. The HDPR is shown in stick form with its C, N, and O atoms green, blue, and red. The enzyme-bound Zn^{2+} ion, which is coordinated by HDPR's 6-hydroxyl group, is represented by a silver sphere. [Based on an X-ray structure by Florante Quiocho, Baylor College of Medicine. PDBid 1ADA.] (b) The proposed catalytic mechanism of adenosine deaminase. A Zn^{2+} -polarized H_2O molecule (Section 15-1Cb) nucleophilically attacks C6 of the enzyme-bound adenosine molecule in a process that is facilitated by His 238 acting as a general base, Glu 217 acting as a general acid, and Asp 295 acting to orient the water molecule via hydrogen bonding. The resulting tetrahedral intermediate decomposes by the elimination of ammonia in a reaction that is aided by the now imidazolium and carboxyl side chains of His 238 and Glu 217 acting as a general acid and a general base, respectively. This yields inosine in its enol tautomeric form, which, on its release from the enzyme, largely assumes its dominant keto form. The Zn^{2+} is coordinated by three His side chains that are not shown. [After Wilson, D.K. and Quiocho, F.A., *Biochemistry* 32, 1692 (1993).]

Interactive Exercise 30

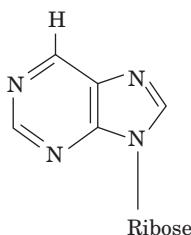
pathways among these various intermediates (including adenine), but all of these pathways lead to uric acid. Of course, the intermediates in these processes may instead be reused to form nucleotides via salvage reactions. In addition, ribose-1-phosphate, a product of the reaction catalyzed by **purine nucleoside phosphorylase (PNP)**, is isomerized by **phosphoribomutase** to the PRPP precursor ribose-5-phosphate.

Adenosine and deoxyadenosine are not degraded by mammalian PNP. Rather, adenine nucleosides and nucleotides are deaminated by **adenosine deaminase (ADA)** and **AMP deaminase** to their corresponding inosine derivatives, which, in turn, may be further degraded. The X-ray structure of murine ADA that was crystallized in the presence of its inhibitor **purine ribonucleoside** was determined by Florante Quiocho (Fig. 28-24a). The enzyme forms an

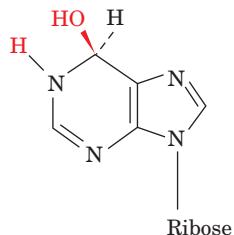


eight-stranded α/β barrel with its active site in a pocket at the C-terminal end of the β barrel, as occurs in nearly all known α/β barrel enzymes (Section 8-3Bh). Purine ribonucleoside

binds to ADA in a normally rare hydrated form, **6-hydroxy-1,6-dihydropurine ribonucleoside (HDPR)**,



Purine ribonucleoside



6-Hydroxy-1,6-dihydropurine ribonucleoside (HDPR)

a nearly ideal transition state analog of the ADA reaction. Although it had been previously reported that ADA does not require a cofactor, its X-ray structure clearly reveals that a zinc ion is bound in the deepest part of the active site pocket, where it is pentacoordinated by three His side chains, a carboxyl oxygen of Asp 295, and the O6 atom of HDPR. ADA's active site complex suggests a catalytic mechanism (Fig. 28-24b) reminiscent of that of carbonic anhydrase (Section 15-1Cb): His 238, which is properly positioned to act as a general base, abstracts a proton from a bound Zn^{2+} -activated water molecule, which nucleophilically attacks the adenine C6 atom to form a tetrahedral intermediate. Products are then formed by the elimination of ammonia.

a. Genetic Defects in ADA Result in Severe Combined Immunodeficiency Disease

Abnormalities in purine nucleoside metabolism arising from rare genetic defects in ADA selectively kill **lymphocytes** (a type of white blood cell). Since lymphocytes mediate much of the immune response (Section 35-2A), ADA deficiency results in **severe combined immunodeficiency disease (SCID)** that, without special protective measures, is invariably fatal in infancy due to overwhelming infection. The mutations in all eight known ADA variants obtained from SCID patients appear to structurally perturb the active site of ADA.

Biochemical considerations provide a plausible explanation of SCID's etiology (causes). In the absence of active ADA, deoxyadenosine is phosphorylated to yield levels of dATP that are 50-fold greater than normal. This high concentration of dATP inhibits ribonucleotide reductase (Section 28-3Ad), thereby preventing the synthesis of the other dNTPs, choking off DNA synthesis and thus cell proliferation. The tissue-specific effect of ADA deficiency on the immune system may be explained by the observation that lymphoid tissue is particularly active in deoxyadenosine phosphorylation.

SCID caused by ADA defects does not respond to treatment by the intravenous injection of ADA because the liver clears this enzyme from the bloodstream within minutes. If, however, several molecules of the biologically inert polymer **Polyethylene glycol (PEG)**

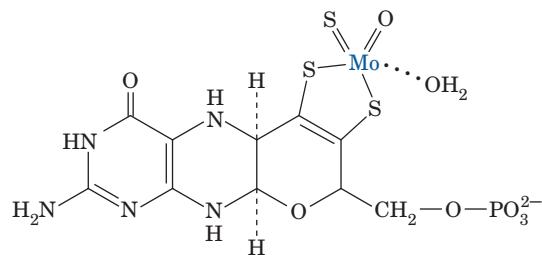
are covalently linked to surface groups on ADA, the resulting **PEG-ADA** remains in the blood for 1 to 2 weeks, thereby largely resuscitating the SCID victim's immune system. The protein-linked PEG only reduces the catalytic activity of ADA by ~40% but, evidently, masks it from the receptors that filter it out of the blood. SCID can therefore be treated effectively by PEG-ADA. This treatment, however, is expensive and not entirely satisfactory. Consequently, ADA deficiency was selected as one of the first genetic diseases to be treated by gene therapy (Section 5-5b): Lymphocytes were extracted from the blood of an ADA-deficient child and grown in the laboratory, had a normal ADA gene inserted into them via genetic engineering techniques (Section 5-5), and were then returned to the child. After 12 years, 20 to 25% of the patient's lymphocytes contained the introduced ADA gene. However, ethical considerations have mandated that the patient continue receiving injections of PEG-ADA so that the efficacy of this gene therapy protocol is unclear.

b. The Purine Nucleotide Cycle

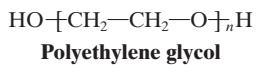
The deamination of AMP to IMP, when combined with the synthesis of AMP from IMP (Fig. 28-4, *left*), has the effect of deaminating aspartate to yield fumarate (Fig. 28-25). John Lowenstein demonstrated that this **purine nucleotide cycle** has an important metabolic role in skeletal muscle. An increase in muscle activity requires an increase in the activity of the citric acid cycle. This process usually occurs through the generation of additional citric acid cycle intermediates (Section 21-4). Muscles, however, lack most of the enzymes that catalyze these anaplerotic (filling up) reactions in other tissues. Rather, muscle replenishes its citric acid cycle intermediates as fumarate generated in the purine nucleotide cycle. The importance of the purine nucleotide cycle in muscle metabolism is indicated by the observation that the activities of the three enzymes involved are all severalfold higher in muscle than in other tissues. In fact, individuals with an inherited deficiency in muscle AMP deaminase (**myoadenylate deaminase deficiency**) are easily fatigued and usually suffer from cramps after exercise.

c. Xanthine Oxidase Is a Mini-Electron-Transport Protein

Xanthine oxidase (XO) converts hypoxanthine to xanthine, and xanthine to uric acid (Fig. 28-23, *bottom*). In mammals, this enzyme occurs mainly in the liver and the small intestinal mucosa. XO is a homodimer of ~1330-residue subunits, each of which binds a variety of electron-transfer agents: an FAD, two spectroscopically distinct [2Fe-2S] clusters, and a **molybdopterin complex (Mo-pt)**



Molybdopterin complex (Mo-pt)



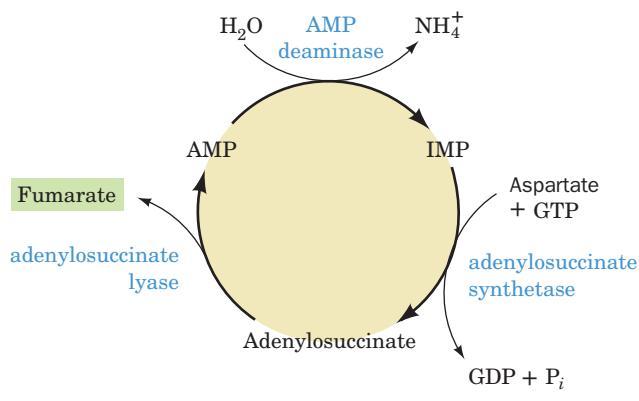
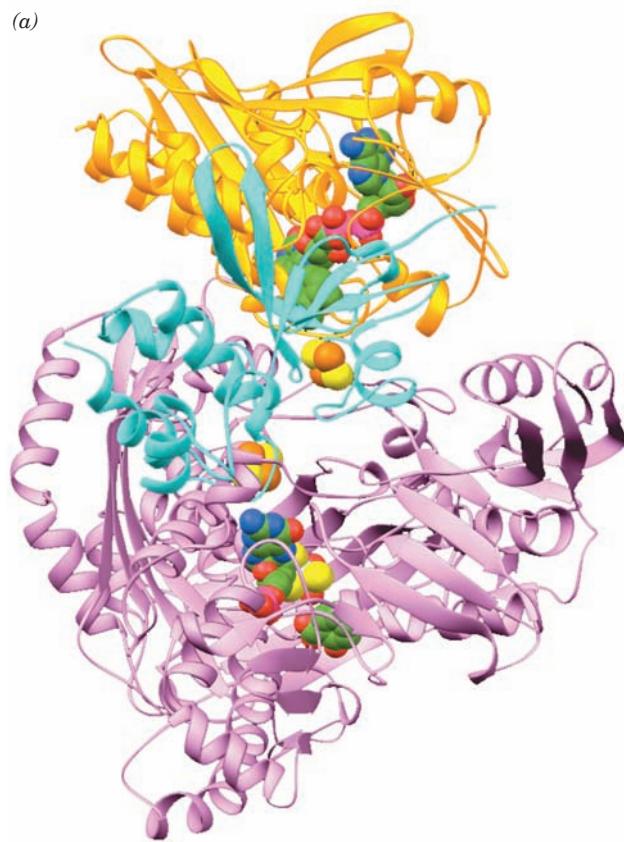


Figure 28-25 The purine nucleotide cycle. This pathway functions, in muscle, to prime the citric acid cycle by generating fumarate.

in which the Mo atom cycles between its Mo(VI) and Mo(IV) oxidation states. The final electron acceptor is O_2 , which is converted to H_2O_2 , a potentially harmful oxidizing agent that is subsequently disproportionated to H_2O and O_2 by catalase (Section 1-2Ad). In XO, the polypeptide has been proteolytically cleaved into three segments (the uncleaved enzyme, which is known as **xanthine dehydrogenase**, preferably uses NAD^+ as its electron acceptor, whereas XO does not react with NAD^+).



The X-ray structure of XO from cow's milk in complex with the competitive inhibitor salicylic acid (Fig. 25-74), determined by Emil Pai, reveals that the FAD and the molybdopterin complex are interposed by the two [2Fe–2S] clusters to form a mini-electron-transport chain (Fig. 28-26). Each of its three peptide segments forms a separate domain with the N-terminal domain binding the two [2Fe–2S] clusters, the central domain binding the FAD, and the C-terminal domain binding the Mo-pt complex. Although the salicylic acid does not contact the Mo-pt complex, it binds to XO in a way that blocks the approach of substrates to the metal center.

XO hydroxylates xanthine at its C8 position (and hypoxanthine at its C2 position), yielding uric acid in its enol form that tautomerizes to the more stable keto form:

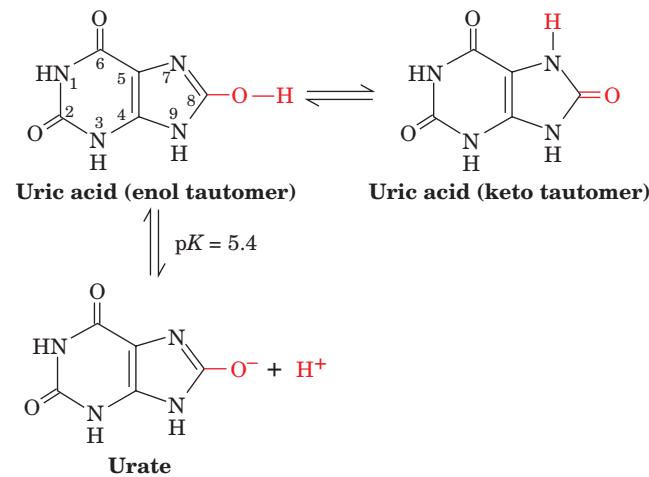


Figure 28-26 X-ray structure of xanthine oxidase from cow's milk in complex with salicylic acid. (a) Ribbon diagram of its 1332-residue subunit in which the N-terminal domain (residues 2–165) is cyan, the central domain (residues 224–528) is gold, and the C-terminal domain (residues 571–1315) is lavender. The enzyme's redox cofactors and bound salicylic acid are shown in space-filling form with C green, N blue, O red, S yellow, P magenta, Fe orange, and Mo light blue. The ~50-residue peptide segments spanning domains are disordered and are apparently highly flexible. (b) The enzyme's redox cofactors and salicylic acid (Sal) drawn in stick form with their S, Fe, and Mo atoms represented by spheres. The atoms are colored as in Part a and viewed from the same direction but with greater magnification. [Based on an X-ray structure by Emil Pai, University of Toronto, Toronto, Ontario, Canada. PDBid 1FIQ.]

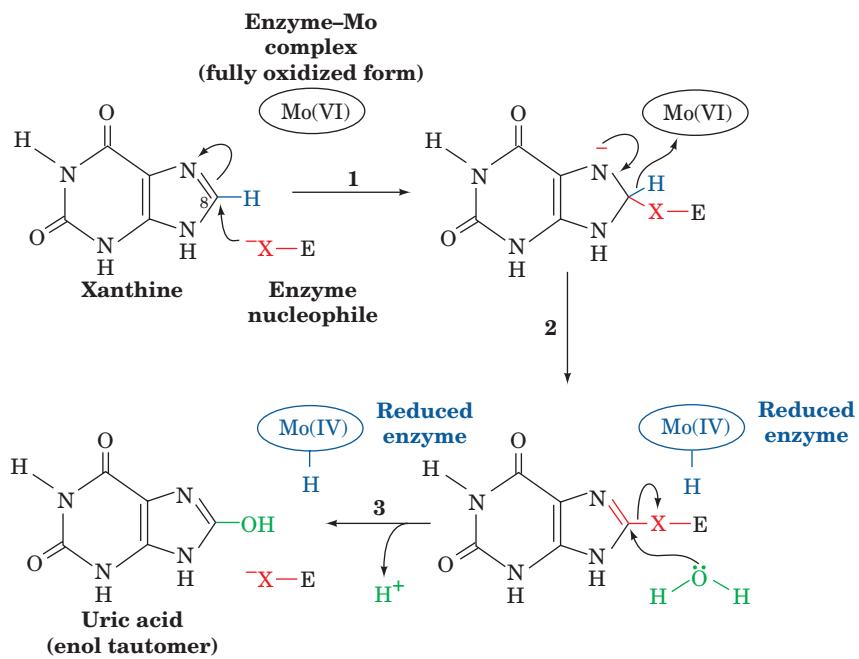


Figure 28-27 Mechanism of xanthine oxidase. The reduced enzyme is subsequently reoxidized by O₂, yielding H₂O₂.

(its enol form ionizes with a pK of 5.4; hence, the name uric acid). ¹⁸O-labeling experiments have demonstrated that the C8 keto oxygen of uric acid is derived from H₂O, whereas the oxygen atoms of H₂O₂ come from O₂. Chemical and spectroscopic studies suggest that the enzyme has the following mechanism (Fig. 28-27):

1. The reaction is initiated by the attack of an enzyme nucleophile, X, on the C8 position of xanthine.
2. The C8—H atom is eliminated as a hydride ion that combines with the Mo(VI) complex, thereby reducing it to the Mo(IV) state.
3. Water displaces the enzyme nucleophile producing uric acid.

In the second stage of the reaction, the now reduced enzyme is reoxidized to its original Mo(VI) state by reaction with O₂. This complex process, not surprisingly, is but poorly understood. EPR measurements indicate that electrons are funneled from the Mo(IV) through the two [2Fe-2S] clusters to the flavin and ultimately to O₂, yielding H₂O₂ and regenerated enzyme.

B. Fate of Uric Acid

In humans and other primates, the final product of purine degradation is uric acid, which is excreted in the urine. The same is true of birds, terrestrial reptiles, and many insects, but these organisms, which do not excrete urea, also catabolize their excess amino acid nitrogen to uric acid via purine biosynthesis. This complicated system of nitrogen excretion has a straightforward function: *It conserves water.*

Uric acid is only sparingly soluble in water, so that its excretion as a paste of uric acid crystals is accompanied by very little water. In contrast, the excretion of an equivalent amount of the much more water-soluble urea osmotically sequesters a significant amount of water.

In all other organisms, uric acid is further processed before excretion (Fig. 28-28). Mammals other than primates oxidize it to their excretory product, **allantoin**, in a reaction catalyzed by the Cu-containing enzyme **urate oxidase**. A further degradation product, **allantoic acid**, is excreted by teleost (bony) fish. Cartilaginous fish and amphibia further degrade allantoic acid to urea prior to excretion. Finally, marine invertebrates decompose urea to their nitrogen excretory product, NH₄⁺.

a. Gout Is Caused by an Excess of Uric Acid

Gout is a disease characterized by elevated levels of uric acid in body fluids. Its most common manifestation is excruciatingly painful arthritic joint inflammation of sudden onset, most often in the big toe (Fig. 28-29), caused by deposition of nearly insoluble crystals of sodium urate. Sodium urate and/or uric acid may also precipitate in the kidneys and ureters as stones, resulting in renal damage and urinary tract obstruction. Gout, which affects ~3 per 1000 persons, predominantly males, has been traditionally, although inaccurately, associated with overindulgent eating and drinking. The probable origin of this association is that in previous centuries, when wine was often contaminated with lead during its manufacture and storage, heavy drinking resulted in chronic lead poisoning, which, among other things, decreases the kidney's ability to excrete uric acid.

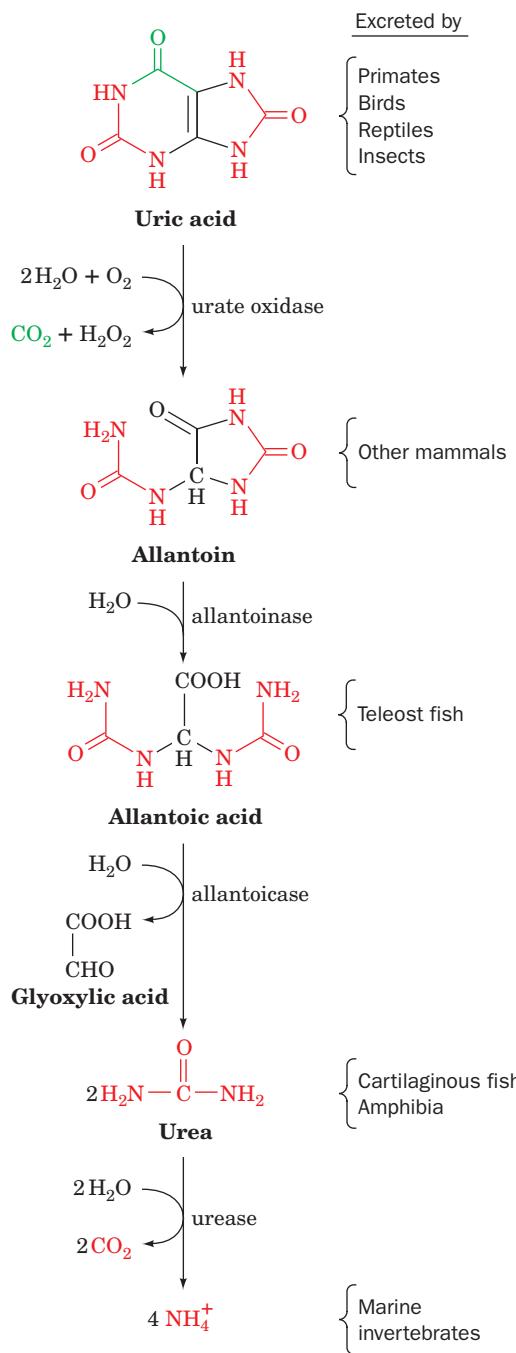


Figure 28-28 Degradation of uric acid to ammonia. The process is arrested at different stages in the indicated species and the resulting nitrogen-containing product is excreted.

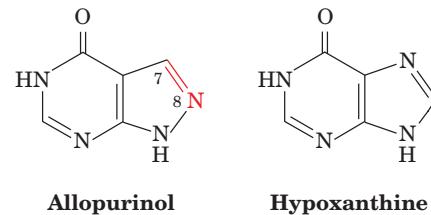
The most prevalent cause of gout is impaired uric acid excretion (although usually for other reasons than lead poisoning). Gout may also result from a number of metabolic insufficiencies, most of which are not well characterized. One well-understood cause is HGPRT deficiency (Lesch–Nyhan syndrome in severe cases), which leads to excessive uric acid production through PRPP accumulation (Section 28-1D). Uric acid overproduction is also



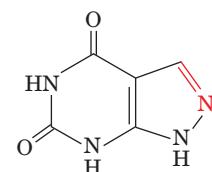
Figure 28-29 The Gout, a cartoon by James Gillray (1799). [Yale University Medical Historical Library.]

caused by glucose-6-phosphatase deficiency (von Gierke's glycogen storage disease; Section 18-4): The increased availability of glucose-6-phosphate stimulates the pentose phosphate pathway (Section 23-4), increasing the rate of ribose-5-phosphate production and consequently that of PRPP, which in turn stimulates purine biosynthesis.

Gout may be treated by administration of the xanthine oxidase inhibitor **allopurinol**, a hypoxanthine analog with interchanged N7 and C8 positions.



Xanthine oxidase hydroxylates allopurinol, as it does hypoxanthine, yielding **alloxanthine**,



which remains tightly bound to the reduced form of the enzyme, thereby inactivating it. Allopurinol consequently alleviates the symptoms of gout by decreasing the rate of uric acid production while increasing the levels of the more soluble hypoxanthine and xanthine. Although allopurinol controls the gouty symptoms of Lesch–Nyhan syndrome, it has no effect on its neurological symptoms.

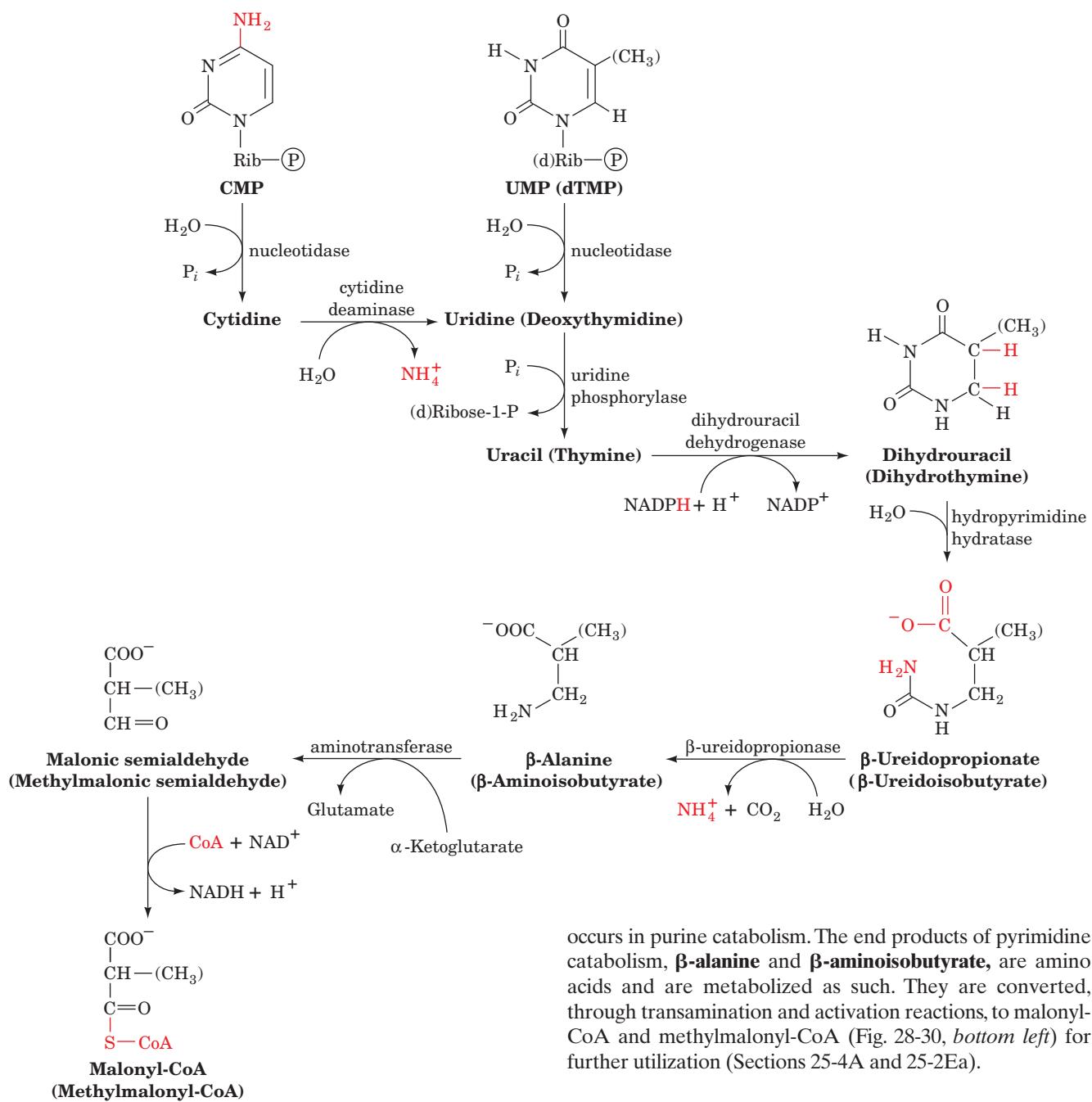


Figure 28-30 Major pathways of pyrimidine catabolism in animals. The amino acid products of these reactions are taken up in other metabolic processes. UMP and dTMP are degraded by the same enzymes; the pathway for dTMP degradation is given in parentheses.

C. Catabolism of Pyrimidines

Animal cells degrade pyrimidine nucleotides to their component bases (Fig. 28-30, top). These reactions, like those of purine nucleotides, occur through dephosphorylation, deamination, and glycosidic bond cleavages. The resulting uracil and thymine are then broken down in the liver through reduction (Fig. 28-30, middle) rather than by oxidation, as

occurs in purine catabolism. The end products of pyrimidine catabolism, **β -alanine** and **β -aminoisobutyrate**, are amino acids and are metabolized as such. They are converted, through transamination and activation reactions, to malonyl-CoA and methylmalonyl-CoA (Fig. 28-30, bottom left) for further utilization (Sections 25-4A and 25-2Ea).

5 BIOSYNTHESIS OF NUCLEOTIDE COENZYMES

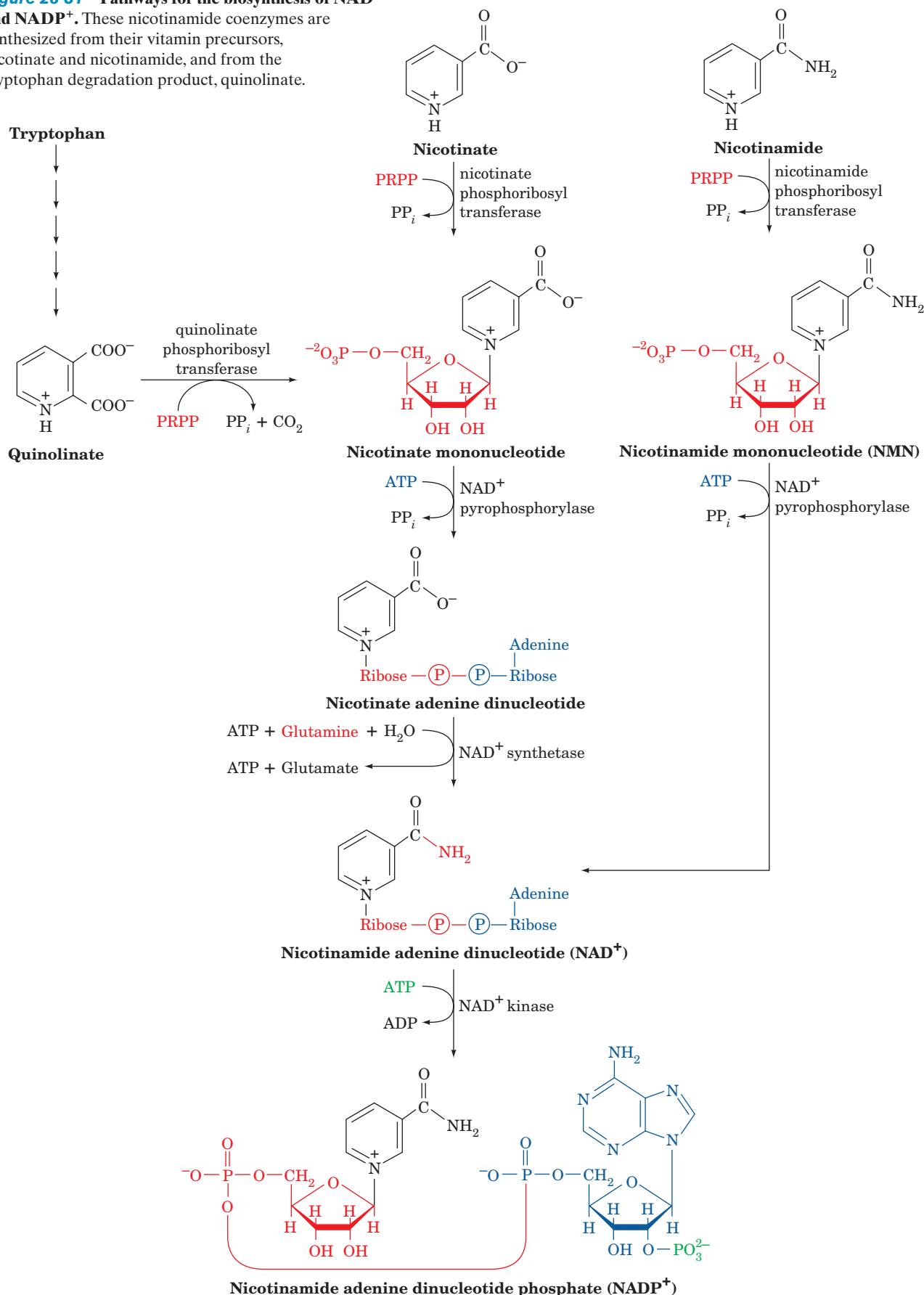
In this section we outline the assembly, in animals, of the nucleotide coenzymes NAD^+ and $NADP^+$, FMN and FAD, and coenzyme A, from their vitamin precursors. These vitamins are synthesized *de novo* only by plants and microorganisms.

A. Nicotinamide Coenzymes

The nicotinamide moiety of the nicotinamide coenzymes (NAD^+ and $NADP^+$) is derived, in humans, from dietary nicotinamide, nicotinic acid, or the essential amino acid tryptophan (Fig. 28-31). **Nicotinate phosphoribosyltransferase**,

Figure 28-31 Pathways for the biosynthesis of NAD^+ and NADP^+ .

These nicotinamide coenzymes are synthesized from their vitamin precursors, nicotinate and nicotinamide, and from the tryptophan degradation product, quinolinate.



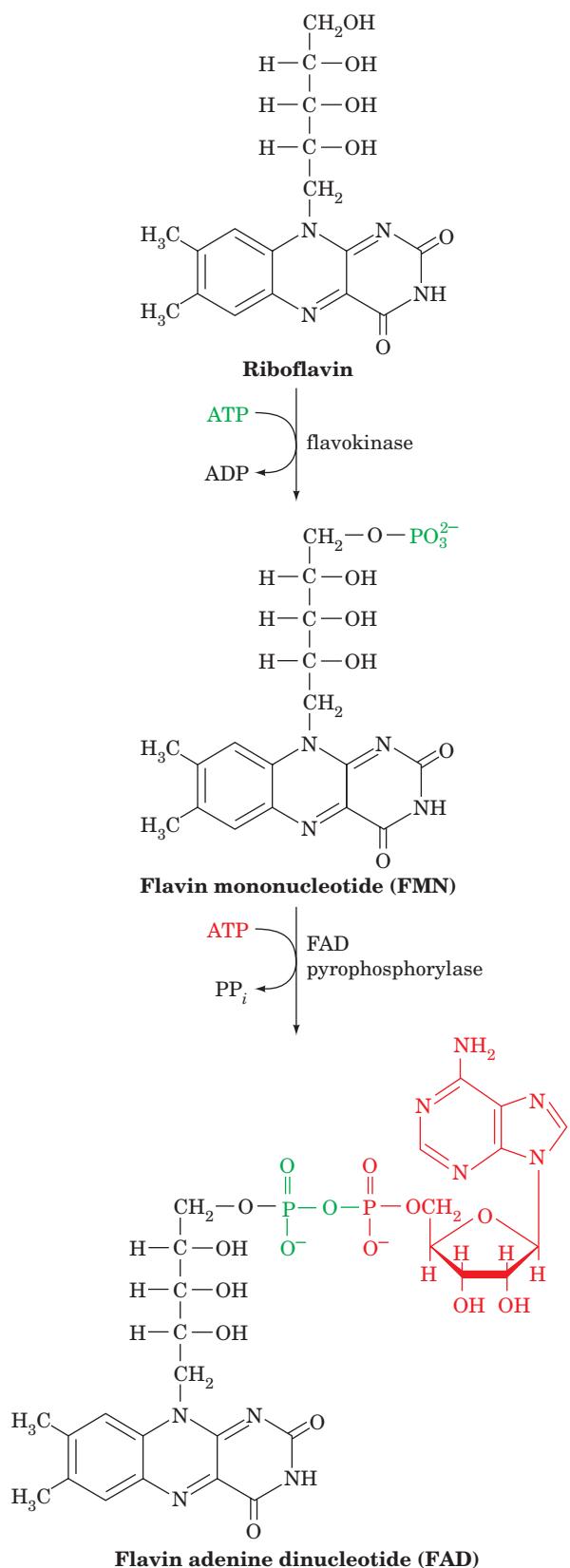


Figure 28-32 Biosynthesis of FMN and FAD from the vitamin precursor riboflavin.

which occurs in most mammalian tissues, catalyzes the formation of **nicotinate mononucleotide** from nicotinate and PRPP. This intermediate may also be synthesized from **quinolinate**, a degradation product of tryptophan (Section 26-3G), in a reaction mediated by **quinolinate phosphoribosyltransferase**, which occurs mainly in liver and kidney. A poor diet, nevertheless, may result in pellagra (nicotinic acid deficiency; Section 13-3), since, under such conditions, tryptophan will be almost entirely utilized in protein biosynthesis. Nicotinate mononucleotide is joined via a pyrophosphate linkage to an ATP-derived AMP residue by **NAD⁺ pyrophosphorylase** to yield **nicotinate adenine dinucleotide (desamido NAD⁺)**. Finally, **NAD⁺ synthetase** converts this intermediate to NAD⁺ by a transamidation reaction in which glutamine is the NH₂ donor.

NAD⁺ may also be synthesized from nicotinamide. This vitamin is converted to **nicotinamide mononucleotide (NMN)** by **nicotinamide phosphoribosyltransferase**, a widely occurring enzyme distinct from nicotinate phosphoribosyltransferase. However, NAD⁺ is synthesized from NMN and ATP by NAD pyrophosphorylase, the same enzyme that synthesizes nicotinate adenine dinucleotide.

NADP⁺ is formed via the ATP-dependent phosphorylation of the NAD⁺ adenosine residue's C2' OH group by **NAD⁺ kinase**.

B. Flavin Coenzymes

FAD is synthesized from riboflavin in a two-reaction pathway (Fig. 28-32). First, the 5'-OH group of riboflavin's ribityl side chain is phosphorylated by **flavokinase**, yielding flavin mononucleotide (FMN; not a true nucleotide since its ribityl residue is not a true sugar). FMN may then be formed by the coupling of FMN and ATP-derived AMP in a pyrophosphate linkage in a reaction catalyzed by **FAD pyrophosphorylase**. Both of these enzymes are widely distributed in nature.

C. Coenzyme A

Coenzyme A is synthesized in mammalian cells according to the pathway diagrammed in Fig. 28-33. Pantothenate, an essential vitamin, is phosphorylated by **pantothenate kinase** and then coupled to cysteine, the future business end of CoA, by **phosphopantothenoylcysteine synthetase**. After decarboxylation by **phosphopantothenoylcysteine decarboxylase**, the resulting **4'-phosphopantethiene** is coupled to AMP in a pyrophosphate linkage by **dephospho-CoA pyrophosphorylase** and then phosphorylated at its adenosine 3' OH group by **dephospho-CoA kinase** to form CoA. The latter two enzymatic activities occur on a single protein.

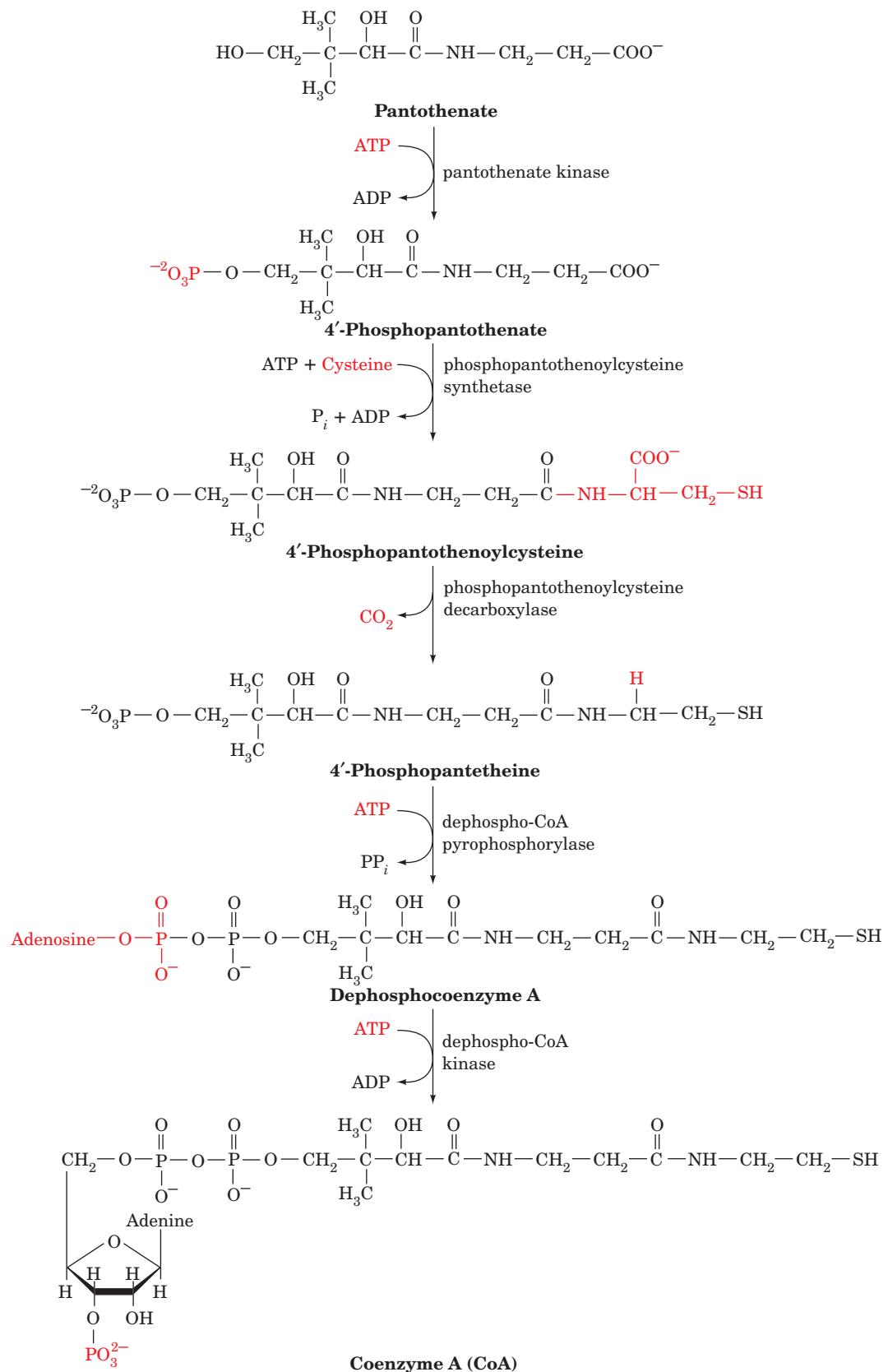


Figure 28-33 Biosynthesis of coenzyme A from pantothenate, its vitamin precursor.

CHAPTER SUMMARY

1 Synthesis of Purine Ribonucleotides Almost all cells synthesize purine nucleotides *de novo* via similar metabolic pathways. The purine ring is constructed in an 11-step reaction sequence that yields IMP. AMP and GMP are then synthesized from IMP in separate pathways. Nucleoside diphosphates and triphosphates are sequentially formed from these products via phosphorylation reactions. The rates of synthesis of these various nucleotides are interrelated through feedback inhibition mechanisms that monitor their concentrations. Purine nucleotides may also be synthesized from free purines salvaged from nucleic acid degradation processes. The importance of these salvage reactions is demonstrated, for example, by the devastating and bizarre consequences of Lesch–Nyhan syndrome.

2 Synthesis of Pyrimidine Ribonucleotides Cells also synthesize pyrimidines *de novo* but, in this six-step process, a free base is formed before it is converted to a nucleotide, UMP. UTP is then formed by phosphorylation of UMP, and CTP is synthesized by the amination of UTP. Pyrimidine biosynthesis is regulated by feedback inhibition as well as by the concentrations of purine nucleotides.

3 Formation of Deoxyribonucleotides Deoxyribonucleotides are formed by reduction of the corresponding ribonucleotides. Three classes of ribonucleotide reductase (RNR) have been characterized: Class I RNR, which occurs in nearly all eukaryotes and many prokaryotes, contains an Fe(III) $\text{—O}^2\text{—Fe(III)}$ group and a tyrosyl free radical; Class II and III RNRs, which occur only in prokaryotes, contain, respectively, a coenzyme B₁₂ cofactor, and a [4Fe–4S] cluster together with a glycyl radical. All of them catalyze free radical-

based reductions. The substrates for Class I and II RNRs are NDPs, whereas those for Class III RNRs are NTPs. Class I RNR has three independent regulatory sites that control its substrate specificity and its catalytic activity in part via its oligomerization state, thereby generating deoxynucleotides in the amounts required for DNA synthesis. The *E. coli* Class I RNR is reduced to its original state by electron-transport chains involving either thioredoxin, thioredoxin reductase, and NADPH; or glutaredoxin, glutathione, glutathione reductase, and NADPH. Thymine is synthesized by the methylation of dUMP by thymidylate synthase to form dTMP. The reaction's methyl source, N⁵,N¹⁰-methylene-THF, is oxidized in the reaction to yield dihydrofolate. N⁵,N¹⁰-Methylene-THF is subsequently regenerated through the sequential actions of dihydrofolate reductase and serine hydroxymethyltransferase. Since this sequence of reactions is required for DNA biosynthesis, it presents an excellent target for chemotherapy. FdUMP, a mechanism-based inhibitor of thymidylate synthase, and methotrexate, an antifolate that essentially irreversibly inhibits dihydrofolate reductase, are both highly effective anticancer agents.

4 Nucleotide Degradation Purine nucleotides are catabolized to yield uric acid. Depending on the species, the uric acid is either directly excreted or first degraded to simpler nitrogen-containing substances. Overproduction or underexcretion of uric acid in humans causes gout. Pyrimidines are catabolized in animal cells to amino acids.

5 Biosynthesis of Nucleotide Coenzymes The nucleotide coenzymes NAD⁺ and NADP⁺, FMN and FAD, and coenzyme A are synthesized in animals from vitamin precursors.

REFERENCES

General

- Nyhan, W.L., Disorders of purine and pyrimidine metabolism, *Mol. Genet. Metab.* **86**, 25–33 (2005).
 Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Part 11 contains chapters on defects in purine and pyrimidine metabolism.]

Purine Nucleotide Biosynthesis

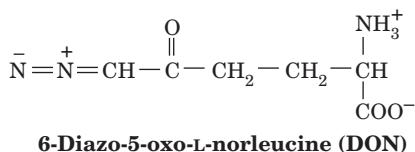
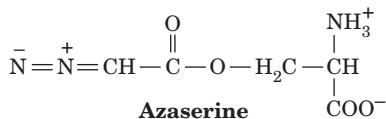
- Almassey, R.J., Janson, C.A., Kan, C.-C., and Hostomyska, Z., Structures of the apo and complexed *Escherichia coli* glycine-amide ribonucleotide transformylase, *Proc. Natl. Acad. Sci.* **89**, 6114–6118 (1992).
 Eriksen, T.A., Kadziola, A., Bentsen, A.-K., Harlow, K.W., and Larsen, S., Structural basis for the function of *Bacillus subtilis* phosphoribosylpyrophosphate synthetase, *Nature Struct. Biol.* **7**, 303–308 (2000).
 Greasley, S.E., Horton, P., Ramcharan, J., Beardsley, G.P., Benkovic, S.J., and Wilson, I.A., Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis, *Nature Struct. Biol.* **8**, 402–406 (2001).
 Kappock, T.J., Ealick, S.E., and Stubbe, J., Modular evolution of the purine biosynthetic pathway, *Curr. Opin. Chem. Biol.* **4**, 567–572 (2000).
 Levdikov, V.M., Barynin, V.V., Grebenko, A.I., Melik-Adamyan, W.R., Lamzin, V.S., and Wilson, K.S., The structure of SAICAR

- synthase: An enzyme in the *de novo* pathway of purine biosynthesis, *Structure* **6**, 363–376 (1998).
 Li, C., Kappock, T.J., Stubbe, J., Weaver, T.M., and Ealick, S.E., X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM) from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution, *Structure* **7**, 1155–1166 (1999).
 Löffler, M., Fairbanks, L.D., Zameitat, E., Marinaki, A.M., and Simmonds, H.A., Pyrimidine pathways in health and disease, *Trends Mol. Med.* **11**, 430–437 (2005).
 Mathews, I.I., Kappock, T.J., Stubbe, J., and Ealick, S.E., Crystal structure of *Escherichia coli* PurE, an unusual mutase in the purine biosynthetic pathway, *Structure* **7**, 1395–1406 (1999).
 Poland, B.W., Fromm, H.J., and Honzatko, R.B., Crystal structures of adenylosuccinate synthetase from *Escherichia coli* complexed with GDP, IMP, hadacidin, NO₃[–], and Mg²⁺, *J. Mol. Biol.* **264**, 1013–1027 (1996).
 Sintchak, M.D., Fleming, M.A., Futer, O., Raybuck, S.A., Chambers, S.P., Caron, P.R., Murcko, M.A., and Wilson, K.P., Structure and mechanism of inosine monophosphate dehydrogenase in complex with the immunosuppressant mycophenolic acid, *Cell* **85**, 921–930 (1996).
 Smith, J.L., Glutamine PRPP amidotransferase: Snapshots of an enzyme in action, *Curr. Opin. Struct. Biol.* **8**, 686–694 (1998).
 Tesmer, J.J., Klem, T.J., Deras, M.L., Davisson, V.J., and Smith, J.L., The crystal structure of GMP synthetase reveals a novel

- catalytic triad and is a structural paradigm for two enzyme families, *Nature Struct. Biol.* **3**, 74–86 (1996).
- Thoden, J.B., Kappock, T.J., Stubbe, J.A., and Holden, H.M., Three-dimensional structure of N^5 -carboxyaminoimidazole ribonucleotide synthetase: A member of the ATP grasp protein superfamily, *Biochemistry* **38**, 15480–15492 (1999). [X-ray structure of PurK.]
- Toth, E.A. and Yeates, T.O., The structure of adenylosuccinate lyase, an enzyme with dual activity in the *de novo* purine biosynthetic pathway, *Structure* **8**, 163–174 (2000).
- Wang, W., Kappock, T.J., Stubbe, J.A., and Ealick, S.E., X-ray structure of glycynamide ribonucleotide synthetase from *Escherichia coli*, *Biochemistry* **37**, 15647–15662 (1998).
- Zalkin, H. and Dixon, J.E., *De novo* purine nucleotide biosynthesis, *Prog. Nucleic Acid Res. Mol. Biol.* **42**, 259–285 (1992).
- Pyrimidine Nucleotide Biosynthesis**
- Begley, T.P., Appleby, T.C., and Ealick, S.E., The structural basis for the remarkable catalytic proficiency of orotidine 5'-monophosphate decarboxylase, *Curr. Opin. Struct. Biol.* **10**, 711–718 (2000).
- Evans, D.R. and Guy, H.I., Mammalian pyrimidine biosynthesis: Fresh insights into an ancient pathway, *J. Biol. Chem.* **279**, 33035–33038 (2004).
- Jones, M.E., Orotidylate decarboxylase of yeast and man, *Curr. Top. Cell Regul.* **33**, 331–342 (1992).
- Liu, S., Neidhardt, E.A., Grossman, T.H., Ocaí, T., and Clardy, J., Structures of human dihydroorotate dehydrogenase in complex with antiproliferative agents, *Structure* **8**, 25–33 (1999).
- Miller, B.G. and Wolfenden, R., Catalytic proficiency: The unusual case of OMP decarboxylase, *Annu. Rev. Biochem.* **71**, 847–885 (2002).
- Scapin, G., Ozturk, D.H., Grubmeyer, C., and Sacchettini, J.C., The crystal structure of the orotate phosphoribosyltransferase complexed with orotate and α -D-5-phosphoribosyl-1-pyrophosphate, *Biochemistry* **34**, 10744–10754 (1995).
- Thoden, J.B., Phillips, G.N., Jr., Neal, T.M., Raushel, F.M., and Holden, H.M., Molecular structure of dihydroorotase: A paradigm for catalysis through the use of a binuclear center, *Biochemistry* **40**, 6989–6997 (2001).
- Traut, T.W. and Jones, M.E., Uracil metabolism—UMP synthesis from orotic acid or uridine and conversion of uracil to β -alanine: Enzymes and cDNAs, *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 1–78 (1996).
- Synthesis of Deoxynucleotides**
- Carreras, C.W. and Santi, D.V., The catalytic mechanism and structure of thymidylate synthase, *Annu. Rev. Biochem.* **64**, 721–762 (1995).
- Eriksson, M., Uhlin, U., Ramaswamy, S., Ekberg, M., Regnström, K., Sjöberg, B.-M., and Eklund, H., Binding of allosteric effectors to ribonucleotide reductase protein R1: Reduction of active site cysteines promotes substrate binding, *Structure* **5**, 1077–1092 (1997).
- Finer-Moore, J.S., Santi, D.V., and Stroud, R.M., Lessons and conclusions from dissecting the mechanism of a bisubstrate enzyme: thymidylate synthase mutagenesis, function, and structure, *Biochemistry* **42**, 248–256 (2003).
- Kashlan, O.B., Scott, C.P., Lear, J.D., and Cooperman, B.S., A comprehensive model for the allosteric regulation of mammalian ribonucleotide reductase. Functional consequences of ATP- and dATP-induced oligomerization of the large subunit, *Biochemistry* **41**, 462–474 (2002).
- Knighton, D.R., Kan, C.-C., Howland, E., Janson, C.A., Hostomyska, Z., Welsh, K.M., and Matthews, D.A., Structure of and kinetic channeling in bifunctional dihydrofolate reductase-thymidylate synthase, *Nature Struct. Biol.* **1**, 186–194 (1994).
- Lennon, B.W., Williams, J.R., Jr., and Ludwig, M.L., Twists in catalysis: Alternating conformations in *Escherichia coli* thioredoxin reductase, *Science* **289**, 1190–1194 (2000).
- Logan, D.T., Andersson, J., Sjöberg, B.-M., and Nordlund, P., A glycol radical site in the crystal structure of a Class III ribonucleotide reductase, *Science* **283**, 1499–1504 (1999).
- Matthews, D.A., Villafranca, J.E., Janson, C.A., Smith, W.W., Welsh, K., and Freer, S., Stereochemical mechanisms of action for thymidylate synthase based on the X-ray structure of the covalent inhibitory ternary complex with 5-fluoro-2'-deoxyuridylate and 5,10-methylenetetrahydrofolate, *J. Mol. Biol.* **214**, 937–948 (1990); and Hyatt, D.C., Maley, F., and Montfort, W.R., Use of strain in a stereospecific catalytic mechanism: Crystal structure of *Escherichia coli* thymidylate synthase bound to FdUMP and methylenetetrahydrofolate, *Biochemistry* **36**, 4585–4594 (1997).
- Mol, C.D., Harris, J.M., McIntosh, E.M., and Tainer, J.A., Human dUTP pyrophosphatase: Uracil recognition by a β hairpin and active sites formed by three separate subunits, *Structure* **4**, 1077–1092 (1996).
- Nordlund, P. and Eklund, H., Structure and function of the *Escherichia coli* ribonucleotide reductase protein R2, *J. Mol. Biol.* **232**, 123–164 (1993).
- Nordlund, P. and Reichard, P., Ribonucleotide reductases, *Annu. Rev. Biochem.* **75**, 681–706 (2006).
- Powis, G. and Montfort, W.R., Properties and biological activities of thioredoxins, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 421–455 (2001).
- Sintchak, M.D., Arjara, G., Kellogg, B.A., Stubbe, J., and Drennan, C.L., The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer, *Nature Struct. Biol.* **9**, 293–300 (2002).
- Stubbe, J. and Riggs-Gelasco, P., Harnessing free radicals: Formation and function of the tyrosyl radical in ribonucleotide reductase, *Trends Biochem. Sci.* **23**, 438–443 (1998).
- Stubbe, J., Ge, J., and Yee, C.S., The evolution of ribonucleotide reduction revisited, *Trends Biochem. Sci.* **26**, 93–99 (2001); and Stubbe, J., Ribonucleotide reductases: The link between an RNA and a DNA world, *Curr. Opin. Struct. Biol.* **10**, 731–736 (2000).
- Uhlen, U. and Eklund, H., Structure of ribonucleotide reductase protein R1, *Nature* **370**, 533–539 (1994).
- Nucleotide Degradation**
- Enroth, C., Eger, B.T., Okamoto, K., Nishino, T., Nishino, T., and Pai, E., Crystal structure of bovine milk xanthine dehydrogenase and xanthine oxidase: Structure based mechanism of conversion, *Proc. Natl. Acad. Sci.* **97**, 10723–10728 (2000).
- Parkman, R., Weinberg, K., Crooks, G., Nolta, I., Kapoor, N., and Kohn, D., Gene therapy for adenosine deaminase deficiency, *Annu. Rev. Med.* **51**, 33–47 (2000).
- Wilson, D.K., Rudolph, F.B., and Quiocho, F.A., Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutations, *Science* **252**, 1278–1284 (1991); Wilson, D.K. and Quiocho, F.A., A pre-transition-state mimic of an enzyme: X-ray structure of adenosine deaminase with bound 1-deazaadenosine and zinc-activated water, *Biochemistry* **32**, 1689–1694 (1993); and Crystallographic observation of a trapped tetrahedral intermediate in a metalloenzyme, *Nature Struct. Biol.* **1**, 691–694 (1994).
- Biosynthesis of Nucleotide Coenzymes**
- Belenky, P., Bogan, K.L., and Brenner, C., NAD⁺ metabolism in health and disease, *Trends Biochem. Sci.* **32**, 12–19 (2007).

PROBLEMS

- 1. Azaserine (*O*-diazoacetyl-L-serine) and **6-diazo-5-oxo-L-norleucine (DON)****



are glutamine analogs. They form covalent bonds to nucleophiles at the active sites of enzymes that bind glutamine, thereby irreversibly inactivating these enzymes. Identify the nucleotide biosynthesis intermediates that accumulate in the presence of either of these glutamine antagonists.

- 2.** Suggest a mechanism for the AIR synthetase reaction (Fig. 28-2, Reaction 6).

- *3.** What is the energetic price, in ATPs, of synthesizing the hypoxanthine residue of IMP from CO_2 and NH_4^+ ?

- 4.** Why is deoxyadenosine toxic to mammalian cells?

- 5.** Indicate which of the following substances are mechanism-based inhibitors and explain your reasoning. (a) Tosyl-L-phenylalanine chloromethylketone with chymotrypsin (Section 15-3Ab). (b) Trimethoprim with bacterial dihydrofolate reductase. (c) The δ -lactone analog of $(\text{NAG})_4$ with lysozyme (Section 15-2Cb). (d) Allopurinol with xanthine oxidase.

- 6.** Why do individuals who are undergoing chemotherapy with cytotoxic (cell killing) agents such as FdUMP or methotrexate temporarily go bald?

- 7.** Normal cells die in a nutrient medium containing thymidine and methotrexate that supports the growth of mutant cells defective in thymidylate synthase. Explain.

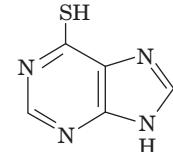
- 8.** FdUMP and methotrexate, when taken together, are less effective chemotherapeutic agents than when either drug is taken alone. Explain.

- 9.** Some microorganisms lack DHFR activity, but their thymidylate synthase has an FAD cofactor. What is the function of the FAD?

- 10.** Why is gout more prevalent in populations that eat relatively large amounts of meat?

- 11.** Gout resulting from the *de novo* overproduction of purines can be distinguished from gout caused by impaired excretion of uric acid by feeding a patient ^{15}N -labeled glycine and determining the distribution of ^{15}N in his or her excreted uric acid. What isotopic distributions are expected for each type of defect?

12. 6-Mercaptopurine,



6-Mercaptopurine

after its conversion to the corresponding nucleotide through salvage reactions, is a potent competitive inhibitor of IMP in the pathways for AMP and GMP biosynthesis. It is therefore a clinically useful anticancer agent. The chemotherapeutic effectiveness of 6-mercaptopurine is enhanced when it is administered with allopurinol. Explain the mechanism of this enhancement.

Schematic diagram of the eukaryotic preinitiation complex that is required for the transcription of DNA to messenger RNA.

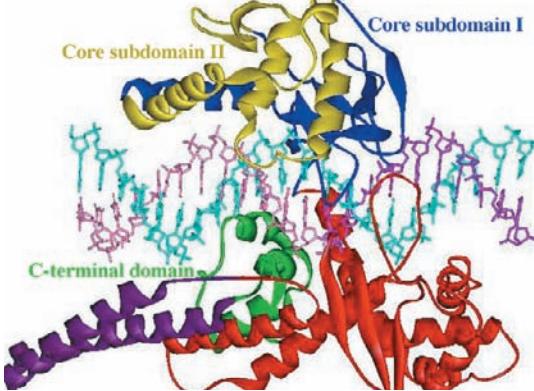
The TATA-box binding protein is shown in orange.



PART V

EXPRESSION AND TRANSMISSION OF GENETIC INFORMATION

This page intentionally left blank



Nucleic Acid Structures

CHAPTER 29

1 Double Helical Structures

- A. B-DNA
- B. Other Nucleic Acid Helices
- 2 Forces Stabilizing Nucleic Acid Structures**
 - A. Sugar-Phosphate Chain Conformations
 - B. Base Pairing
 - C. Base Stacking and Hydrophobic Interactions
 - D. Ionic Interactions
- 3 Supercoiled DNA**
 - A. Superhelix Topology
 - B. Measurements of Supercoiling
 - C. Topoisomerases

There are two classes of nucleic acids, **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the hereditary molecule in all cellular life-forms, as well as in many viruses. It has but two functions:

1. To direct its own **replication** during cell division.
2. To direct the **transcription** of complementary molecules of RNA.

RNA, in contrast, has more varied biological functions:

1. The RNA transcripts of DNA sequences that specify polypeptides, **messenger RNAs (mRNAs)**, direct the ribosomal synthesis of these polypeptides in a process known as **translation**.
2. The RNAs of ribosomes, which are about two-thirds RNA and one-third protein, have functional as well as structural roles.
3. During protein synthesis, amino acids are delivered to the ribosome by molecules of **transfer RNA (tRNA)**.
4. Certain RNAs are associated with specific proteins to form **ribonucleoproteins** that participate in the post-transcriptional processing of other RNAs.
5. A variety of short RNAs participate in the control of eukaryotic gene expression and in protection against viruses, a phenomenon known as **RNA interference (RNAi)**.
6. In many viruses, RNA, not DNA, is the carrier of hereditary information.

The structure and properties of DNA are introduced in Section 5-3. In this chapter we extend this discussion with emphasis on DNA; the structures of RNAs are detailed in Sections 31-4A and 32-2B. Methods of purifying, sequencing, and chemically synthesizing nucleic acids are discussed in Sections 6-6, 7-2, and 7-6, and recombinant DNA techniques are discussed in Section 5-5. Bioinformatics, as it concerns nucleic acids, is outlined in Section 7-4, and the Nucleic Acid Database is described in Section 8-3Cb.

1 DOUBLE HELICAL STRUCTURES

 See Guided Exploration 23: DNA structures Double helical DNA has three major helical forms, B-DNA, A-DNA, and Z-DNA, whose structures are depicted in Fig. 29-1. In this section we discuss the major characteristics of each of these helical forms as well as those of double helical RNA and DNA-RNA hybrid helices.

A. B-DNA

The structure of **B-DNA** (Fig. 29-1, middle panels), the biologically predominant form of DNA, is described in Section 5-3A. To recapitulate (Table 29-1), B-DNA consists of a right-handed double helix whose two antiparallel sugar-phosphate chains wrap around the periphery of the helix. Its aromatic bases (A, T, G, and C), which occupy the core of the helix, form complementary A · T and G · C Watson-Crick base pairs (Fig. 5-12), whose planes are nearly perpendicular to the axis of the double helix. Neighboring base pairs, whose aromatic rings are 3.4 Å thick, are stacked in van der Waals contact, with the helix axis passing through the middle of each base pair. B-DNA is ~20 Å in diameter and has two deep grooves between its sugar-phosphate chains: the relatively narrow **minor groove**, which exposes that edge of the base pairs from which the glycosidic bonds (the bonds from the base N to the ribose C1') extend (toward the bottom of Fig. 5-12), and the relatively wide **major groove**, which exposes the opposite edge of each base pair (toward the top of Fig. 5-12). Canonical (ideal) DNA has a helical twist of 10 base pairs (bp) per turn and hence a pitch (rise per turn) of 34 Å.

The Watson-Crick base pairs in either orientation are structurally interchangeable, that is, A · T, T · A, G · C, and

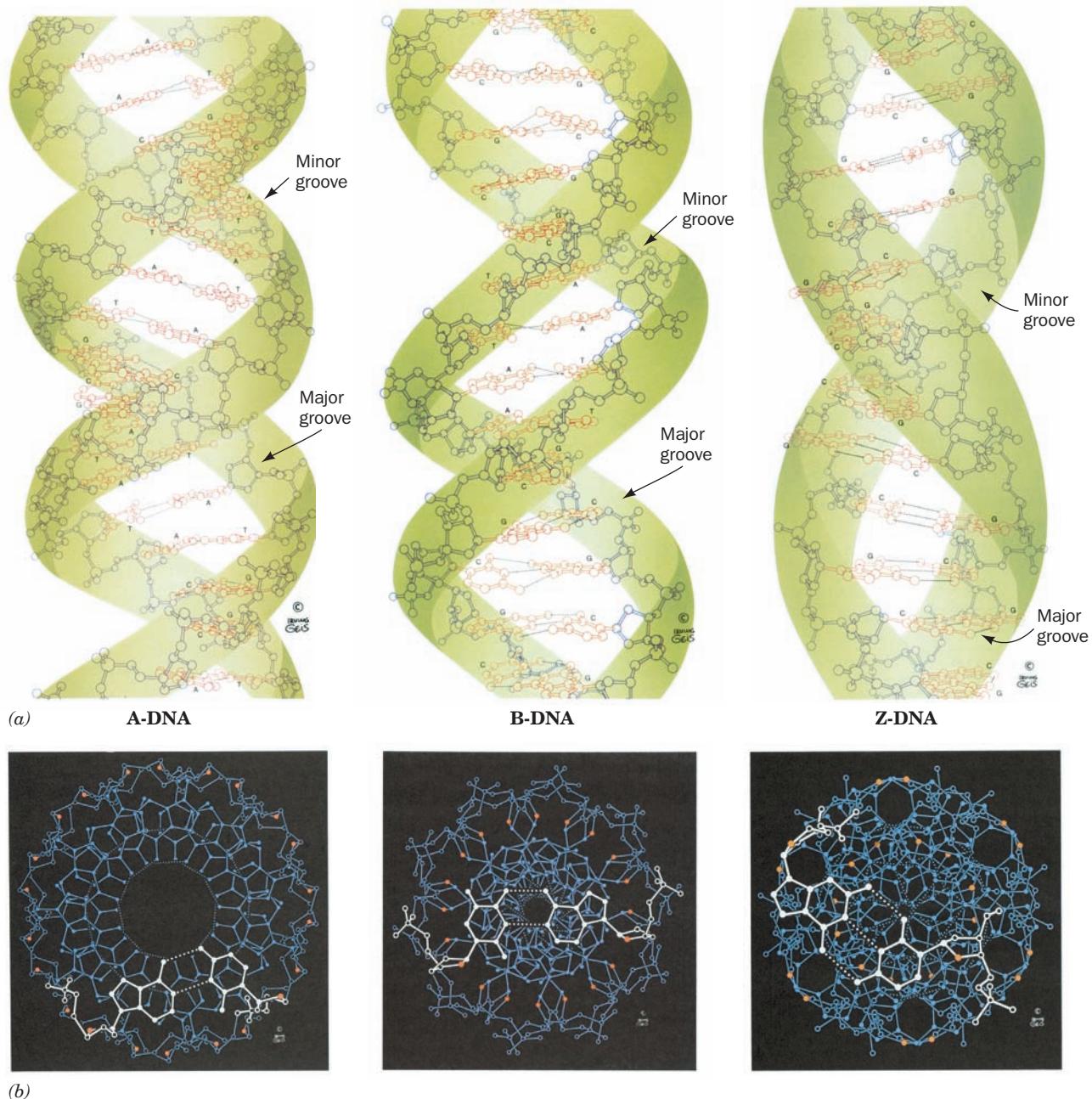


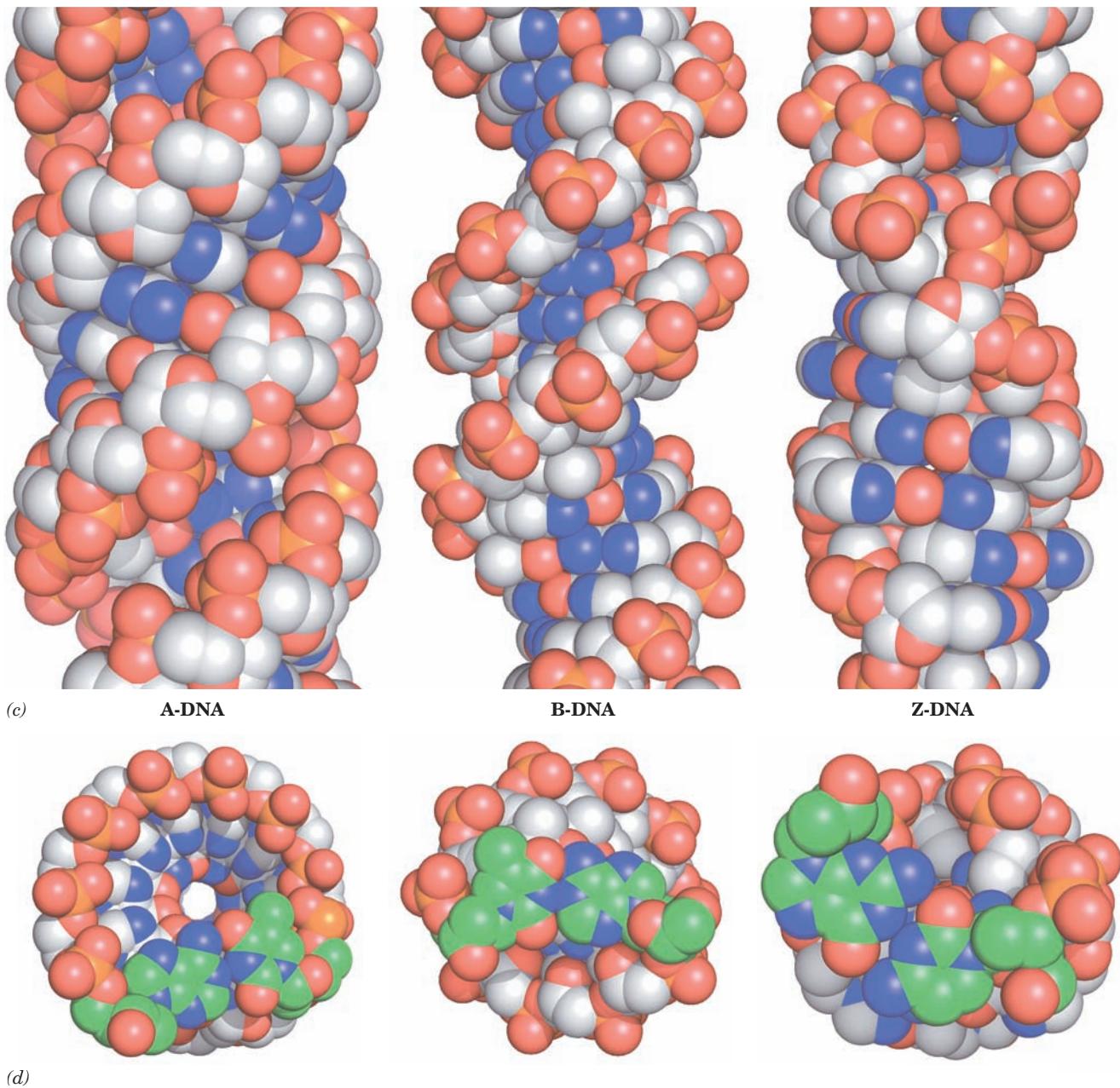
Figure 29-1 Structures of A-, B-, and Z-DNAs. (a) Ball-and-stick drawings viewed perpendicular to the helix axis. The sugar-phosphate backbones, which wind about the periphery of each molecule, are outlined by green ribbons, and the bases, which occupy its core, are red. Note that the two sugar-phosphate chains in each helix run in opposite directions so as to form right-handed double helices in A- and B-DNAs and a left-handed double helix in Z-DNA. (b) Views along the helix axis. The ribose

ring O atoms are red and the nucleotide pair closest to the viewer is white. Note that the helix axis passes far “above” the major groove of A-DNA, through the base pairs of B-DNA, and through the edge of the minor groove of Z-DNA. Consequently, A-DNA has a hollow core whereas B- and Z-DNAs have solid cores. Also note that the deoxyribose residues in A- and B-DNAs have the same conformation in each helix, but those in Z-DNA have two different conformations so that alternate ribose residues

C · G can replace each other in the double helix without altering the positions of the sugar-phosphate backbones’ C1’ atoms. In contrast, any other combination of bases would significantly distort the double helix since the formation of a non-Watson–Crick base pair would require considerable reorientation of the DNA’s sugar-phosphate backbones.

a. Real DNA Deviates from the Ideal Watson–Crick Structure

The DNA samples that were available when James Watson and Francis Crick formulated the Watson–Crick structure in 1953 were extracted from cells and hence consisted of molecules of heterogeneous lengths and base sequences.



lie at different radii. (c) Space-filling models viewed perpendicular to the helix axis and colored according to atom type (C white, N blue, O red, and P orange). (d) Space-filling models viewed along the helix axis and colored as in Part c but with the C atoms of the nucleotide pair closest to the viewer green. H atoms in Parts b, c, and d have been omitted for clarity. [Based on X-ray structures by the following: A-DNA, Olga Kennard, Dov Rabinovitch, Zippora Shakked, and Mysore Viswamitra,

Cambridge University, U.K. Nucleic Acid Database ID ADH010; B-DNA, Richard Dickerson and Horace Drew, Caltech. PDBid 1BNA; and Z-DNA, Andrew Wang and Alexander Rich, MIT. PDBid 2DCG. Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission. Model coordinates for Parts *c* and *d* generated by Helen Berman, Rutgers University.]  **See Kinemage**

sity is the average electron density of all the base pairs in the fiber. The Watson–Crick structure was based, in part, on the X-ray fiber diffraction pattern of B-DNA (Fig. 5-10).

By the late 1970s, advances in nucleic acid chemistry permitted the synthesis and crystallization of ever longer oligonucleotides of defined sequences (Section 7-6A),

Table 29-1 Structural Features of Ideal A-, B-, and Z-DNA

	A-DNA	B-DNA	Z-DNA
Helical sense	Right-handed	Right-handed	Left-handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11.6	10	12 (6 dimers)
Helical twist per base pair	31°	36°	9° for pyrimidine–purine steps; 51° for purine–pyrimidine steps
Helix pitch (rise per turn)	34 Å	34 Å	44 Å
Helix rise per base pair	2.9 Å	3.4 Å	7.4 Å per dimer
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines; C3'-endo for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines; syn for purines

Source: Mainly Arnott, S., in Neidle, S. (Ed.), *Oxford Handbook of Nucleic Acid Structure*, p. 35, Oxford University Press (1999).

many of which could be crystallized. Consequently, some 25 years after the Watson–Crick structure was formulated, its X-ray crystal structure was clearly visualized for the first time when Richard Dickerson and Horace Drew determined the first X-ray crystal structure of a B-DNA, that of the self-complementary dodecamer d(CGCGAATTCGCG), at near-atomic (1.9 Å) resolution. This molecule, whose structure was subsequently determined at significantly higher (1.4 Å) resolution by Loren Williams, has an average rise per residue of 3.3 Å and has 10.1 bp per turn (a helical twist of 35.5° per bp), values that are nearly equal to those of canonical B-DNA. However, individual residues depart significantly from this average conformation (Fig. 29-1a, middle panel). For example, the helical twist per base pair in this dodecamer ranges from 26° to 43°. Each base pair further deviates from its ideal conformation by such distortions as propeller twisting (the opposite rotation of paired bases about the base pair's long axis; in the 1.4-Å resolution structure, this quantity ranges from -23° to -7°) and base pair roll (the tilting of a base pair as a whole about its long axis; this quantity ranges from -14° to 17°).

X-ray and NMR studies of numerous other double helical DNA oligomers have amply demonstrated that *the conformation of DNA, particularly B-DNA, is irregular in a sequence-specific manner*, although the rules specifying how sequence governs conformation have proved to be surprisingly elusive. This is because *base sequence does not so much confer a fixed conformation on a double helix as it establishes the deformability of the helix*. Thus, 5'-R-Y-3' steps (where R and Y are the abbreviations for purines and pyrimidines, respectively) in B-DNA are easily bent because they exhibit relatively little ring–ring overlap between adjacent base pairs. In contrast, both Y-R steps and R-R steps (the latter, due to base pairing, are equivalent to

Y-Y steps), and most notably A-A steps, are more rigid because the extensive ring–ring overlap between their adjacent base pairs tends to keep these base pairs parallel. *This phenomenon, as we shall see, is important for the sequence-specific binding of DNA to proteins that process genetic information.* This is because many of these proteins wrap their target DNAs around them, in many cases by bending them by well over 90°. DNAs with different sequences than the target DNA would not bind so readily to the protein because they would resist deformation to the required conformation more than the target DNA.

B. Other Nucleic Acid Helices

X-ray fiber diffraction studies, starting in the mid-1940s, revealed that *nucleic acids are conformationally variable molecules*. Indeed, double helical DNA and RNA can assume several distinct structures that vary with such factors as the humidity and the identities of the cations present, as well as with base sequence. For example, fibers of B-DNA form in the presence of alkali metal ions such as Na⁺ when the relative humidity is 92%. In this subsection, we describe the other major conformational states of double-stranded DNA as well as those of double-stranded RNA and RNA–DNA hybrid helices.

a. A-DNA's Base Pairs Are Inclined to the Helix Axis

When the relative humidity is reduced to 75%, B-DNA undergoes a reversible conformational change to the so-called A form. Fiber X-ray studies indicate that **A-DNA forms a wider and flatter right-handed helix than does B-DNA** (Fig. 29-1, left panels; Table 29-1). A-DNA has 11.6 bp per turn and a pitch of 34 Å, which gives A-DNA an axial hole (Fig. 29-1b, d, left panels). A-DNA's most striking feature, however, is that the planes of its base pairs are

tilted 20° with respect to the helix axis. Since its helix axis passes “above” the major groove side of the base pairs (Fig. 29-1b, d, left panels) rather than through them as in B-DNA, A-DNA has a deep major groove and a very shallow minor groove; it can be described as a flat ribbon wound around a 6-Å-diameter cylindrical hole. Most self-complementary oligonucleotides of <10 base pairs, for example, d(GGCCGGCC) and d(GGTATACC), crystallize in the A-DNA conformation. Like B-DNA, these molecules exhibit considerable sequence-specific conformational variation although the degree of variation is less than that in B-DNA.

A-DNA has, so far, been observed in only three biological contexts: at the cleavage center of topoisomerase II (Section 29-3Cd), at the active site of DNA polymerase (Section 30-2Ae), and in certain Gram-positive bacteria that have undergone **sporulation** (the formation, under environmental stress, of resistant although dormant cell types known as **spores**; a sort of biological lifeboat). Such spores contain a high proportion (20%) of **small acid-soluble spore proteins (SASPs)**. Some of these SASPs induce B-DNA to assume the A form, at least *in vitro*. The DNA in bacterial spores exhibits a resistance to UV-induced damage that is abolished in mutants that lack these SASPs. This occurs because the B → A conformation change inhibits the UV-induced covalent cross-linking of pyrimidine bases (Section 30-5Aa), in part by increasing the distance between successive pyrimidines.

b. Z-DNA Forms a Left-Handed Helix

Occasionally, a seemingly well-understood or at least familiar system exhibits quite unexpected properties. Over 25 years after the discovery of the Watson–Crick structure, the crystal structure determination of the self-complementary hexanucleotide d(CGCGCG) by Andrew Wang and Alexander Rich revealed, quite surprisingly, a *left-handed double helix* (Fig. 29-1, right panels; Table 29-1). A similar helix is formed by d(CGCATGCG). This helix, which has been dubbed **Z-DNA**, has 12 Watson–Crick base pairs per turn, a pitch of 44 Å, and, in contrast to A-DNA, a deep minor groove and no discernible major groove (its helix axis passes “below” the minor groove side of its base pairs; Fig. 29-1b,d, right panels). Z-DNA therefore resembles a left-handed drill bit in appearance. The base pairs in Z-DNA are flipped 180° relative to those in B-DNA (Fig. 29-2) through conformational changes discussed in Section 29-2A. As a consequence, the repeating unit of Z-DNA is a dinucleotide, d(XpYp), rather than a single nucleotide as it is in the other DNA helices. The line joining successive phosphorus atoms on a polynucleotide strand of Z-DNA therefore follows a zigzag path around the helix (Fig. 29-1a,c, right panels; hence the name Z-DNA) rather than a smooth curve as it does in A- and B-DNA (Fig. 29-1a,c, left and middle panels).

Fiber diffraction and NMR studies have shown that complementary polynucleotides with alternating purines and pyrimidines, such as poly d(GC) · poly d(GC) and poly

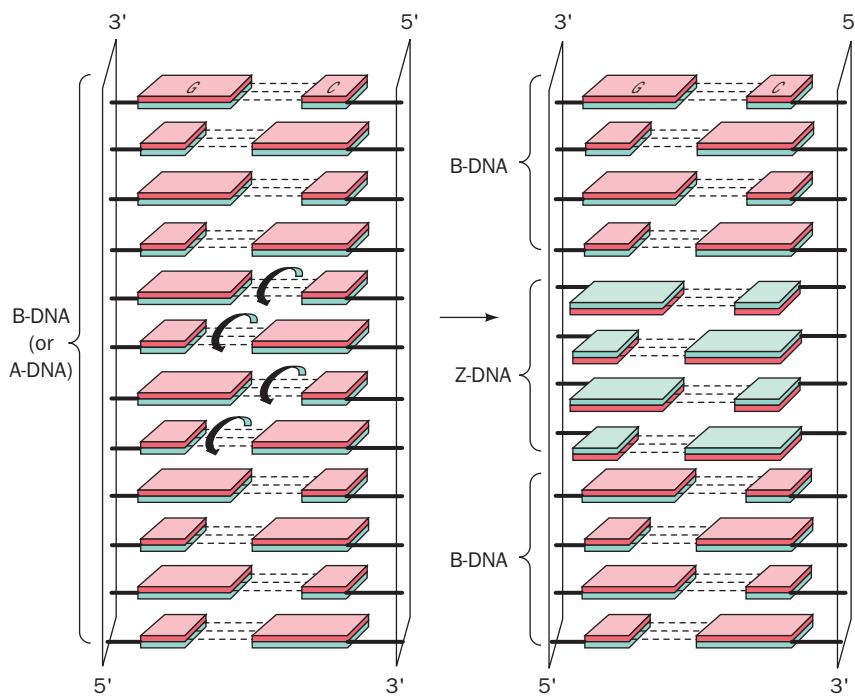


Figure 29-2 Conversion of B-DNA to Z-DNA. The conversion, here represented by a 4-bp DNA segment, involves a 180° flip of each base pair (curved arrows) relative to the sugar-phosphate chains (compare the base pair orientations of B- and Z-DNAs in Fig. 29-1b, d). Here, the different faces of the base pairs

are colored red and green. Note that if the drawing on the left is taken as looking into the minor groove of unwound A- or B-DNA, then in the drawing on the right, we are looking into the major groove of the unwound Z-DNA segment. [After Rich, A., Nordheim, A., and Wang, A.H.-J., *Annu. Rev. Biochem.* **53**, 799 (1984).]

$d(AC) \cdot poly\ d(GT)$, take up the Z-DNA conformation at high salt concentrations. Evidently, the Z-DNA conformation is most readily assumed by DNA segments with alternating purine–pyrimidine base sequences (for structural reasons explained in Section 29-2A). A high salt concentration stabilizes Z-DNA relative to B-DNA by reducing the otherwise increased electrostatic repulsions between closest approaching phosphate groups on opposite strands (8 Å in Z-DNA vs 12 Å in B-DNA). The methylation of cytosine residues at C5, a common biological modification (Section 30-7), also promotes Z-DNA formation since a hydrophobic methyl group in this position is less exposed to solvent in Z-DNA than it is in B-DNA.

Does Z-DNA have any biological function? Rich has proposed that the reversible conversion of specific segments of B-DNA to Z-DNA under appropriate circumstances acts as a kind of switch in regulating genetic expression, and there are indications that it transiently forms behind actively transcribing RNA polymerase (Section 31-

4As). It was nevertheless surprisingly difficult to prove the *in vivo* existence of Z-DNA. A major difficulty was demonstrating that a particular probe for detecting Z-DNA, for example, a Z-DNA-specific antibody, does not in itself cause what would otherwise be B-DNA to assume the Z conformation—a kind of biological uncertainty principle (the act of measurement inevitably disturbs the system being measured). However, Rich has discovered several proteins that specifically bind Z-DNA, including a family of Z-DNA-binding protein domains named **Z α** . The existence of these proteins strongly suggests that Z-DNA does, in fact, exist *in vivo*.

The X-ray structure of the 81-residue Z α domain from the RNA editing enzyme **ADAR1** (Section 31-4As) in complex with d(TCGCGCG) has been determined (Fig. 29-3a). The CGCGCG segment of this heptanucleotide is self-complementary, and therefore forms a 2-fold symmetric, 6-bp segment of Z-DNA with an overhanging dT at the 5' end of each strand (although these dT's are disordered in the

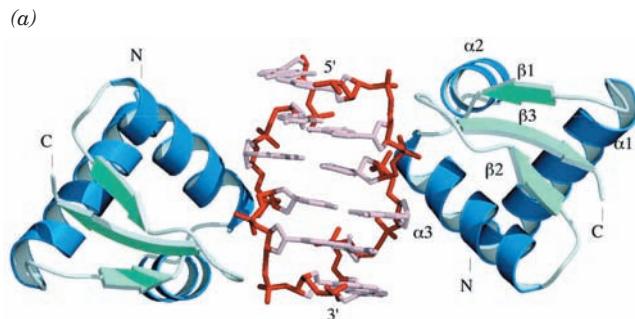
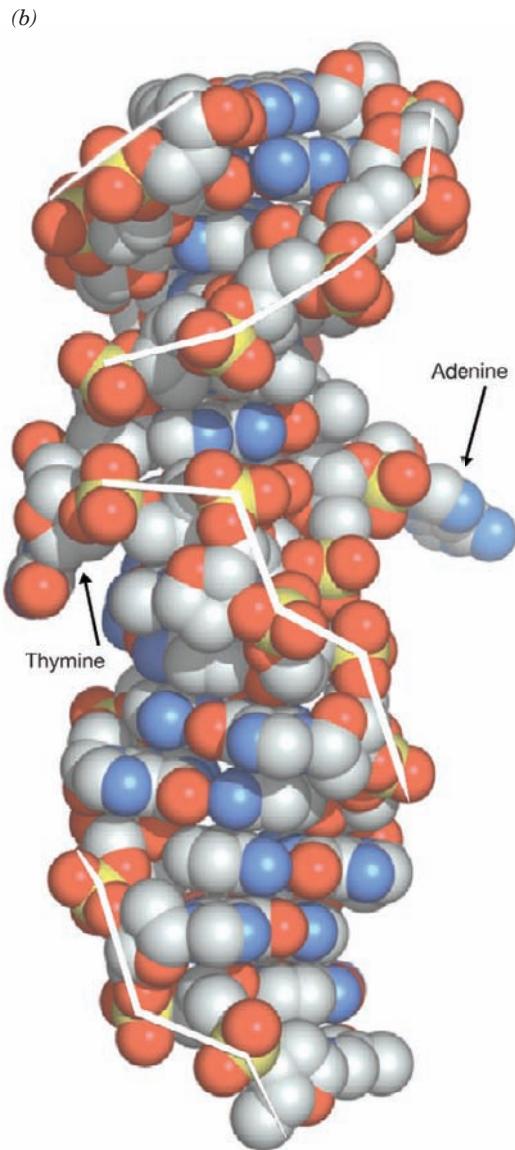


Figure 29-3 Structures of DNA in complex with ADAR1 Z α . (a) X-ray structure of the complex of two Z α domains with Z-DNA consisting of a duplex of the self-complementary hexamer d(CGCGCG) as viewed along its 2-fold axis of symmetry. The Z-DNA is shown in stick form with its backbones red and its remaining portions pink. The Z α domains are drawn in ribbon form with helices blue and sheets light green. Note that each Z α domain contacts only one strand of the Z-DNA.

(b) X-ray structure of a junction between Z-DNA and B-DNA. The complex consists of d(GTCGCGCGCCATAAACC) and d(ACGGTTTATGGCGCGCG) in complex with four Z α domains (not shown here for the sake of clarity). The DNA forms a duplex with a 2-nucleotide overhang at each end in which the lower 8 nucleotide pairs form Z-DNA, the upper 6 nucleotide pairs form B-DNA, and the A and T nucleotides at the interface between these two DNA segments have been extruded from the double helix. The DNA is drawn in space-filling form with C gray, N blue, O red, and P yellow. The white lines connect adjacent P atoms in the same polynucleotide, thereby showing the zigzag pattern of the left-handed Z-DNA and the smoother pattern of the right-handed B-DNA. [Courtesy of Alexander Rich, MIT. PDBids 1QBJ and 2ACJ.]



X-ray structure). A monomeric unit of Z α binds to each strand of the Z-DNA, out of contact with the Z α that binds to the opposite strand. The protein primarily interacts with Z-DNA via hydrogen bonds and salt bridges between polar and basic protein side chains and the Z-DNA's sugar-phosphate backbone. Note that none of the DNA's bases participate in these associations. The protein's DNA-binding surface, which is complementary in shape to the Z-DNA, is positively charged, as is expected for a protein that interacts with several closely spaced, anionic phosphate groups. It is postulated that ADAR1's Z α domain targets it to the Z-DNA upstream of actively transcribing genes (for reasons discussed in Section 31-4As).

In an effort to visualize the structure of the junction between B-DNA and Z-DNA, Rich and Kyeong Kyu Kim cocrystallized Z α with a duplex DNA that has two overhanging nucleotides at each end and whose two 17-nucleotide strands have the sequences d(GTCGCGCGC-CATAAAC) and d(ACGGTTTATGGCGCGCG). The X-ray structure of this complex reveals that 8 nucleotide pairs at one end of the double helix are Z-DNA, 6 nucleotide pairs at the other end of the helix are B-DNA, and the nucleotides that would otherwise form an A · T base pair at the junction between these segments have been expelled from the double helix (Fig. 29-3b). The base pairs of the B- and Z-DNA segments form a continuous stack, which stabilizes the structure (Section 29-2C). Four Z α 's are bound to the Z-DNA segment, two per polynucleotide strand, in a manner closely similar to that in the foregoing Z α -d(TCGCGCG) complex (for clarity, the Z α 's are not shown in Fig. 29-3b) and the B- and Z-DNA segments adopt their standard conformations. Evidently, under the proper conditions, the handedness of duplex DNA can be reversed by breaking one base pair and ejecting its nucleotides from the duplex.

c. RNA-11 and RNA-DNA Hybrids Have an A-DNA-Like Conformation

Double helical RNA is unable to assume a B-DNA-like conformation because of steric clashes involving its 2'-OH groups. Rather, it usually assumes a conformation resembling A-DNA (Fig. 29-1, left panels) known as **A-RNA** or **RNA-11**, which ideally has 11.0 bp per helical turn, a pitch of 30.9 Å, and its base pairs inclined to the helix axis by 16.7°. Many RNAs, for example, transfer and ribosomal RNAs (whose structures are detailed in Sections 32-2B and 32-3A), contain complementary sequences that form double helical stems.

Hybrid double helices, which consist of one strand each of DNA and RNA, are also predicted to have A-RNA-like conformations. In fact, the X-ray structure, by Nancy Horton and Barry Finzel, of a 10-bp complex of the DNA oligonucleotide d(GGCGCCCGAA) with the complementary RNA oligonucleotide r(UUCGGGCGCC) reveals (Fig. 29-4) that it forms a double helix with A-RNA-like character (Table 29-1) in that it has 10.9 bp per turn, a pitch of 31.3 Å, and its base pairs are, on average, inclined to the helix axis by 13.9°. Nevertheless, this hybrid helix also has B-DNA-like qualities in that the width of its minor groove

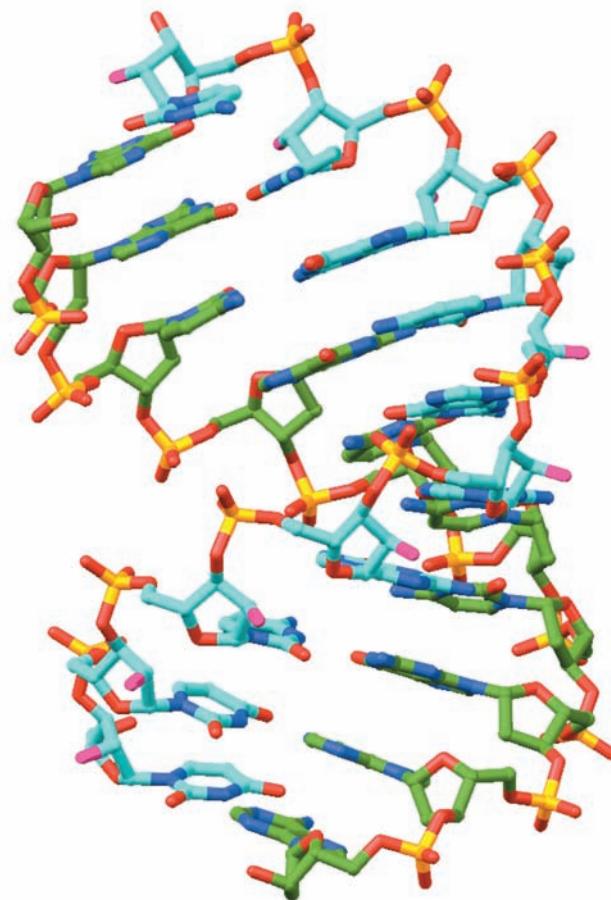


Figure 29-4 X-ray structure of a 10-bp RNA-DNA hybrid helix consisting of d(GGCGCCCGAA) in complex with r(UUCGGGCGCC). The structure is shown in stick form with RNA C atoms cyan, DNA C atoms green, N blue, O red except for RNA O2' atoms, which are magenta, and P yellow. [Based on an X-ray structure by Nancy Horton and Barry Finzel, Pharmacia & Upjohn, Inc., Kalamazoo, Michigan. PDBid 1FIX.]

See Interactive Exercise 31

(9.5 Å) is intermediate between those for canonical B-DNA (7.4 Å) and A-DNA (11 Å) and in that some of the ribose rings of its DNA strand have conformations characteristic of B-DNA (Section 29-2A), whereas others have conformations characteristic of A-RNA. Note that this structure is of biological significance because short segments of RNA · DNA hybrid helices occur in both the transcription of RNA on DNA templates (Section 31-2Ba) and in the initiation of DNA replication by short lengths of RNA (Section 30-1D). The RNA component of this helix is a substrate for **RNase H**, which specifically hydrolyzes the RNA strands of RNA · DNA hybrid helices *in vivo* (Section 30-4C).

2 FORCES STABILIZING NUCLEIC ACID STRUCTURES

Double-stranded DNA does not exhibit the structural complexity of proteins because it has only a limited repertoire of secondary structures and no comparable tertiary or

quaternary structures (although see Section 29-3). This is perhaps to be expected since there is a far greater range of chemical and physical properties among the 20 amino acid residues of proteins than there is among the four DNA bases. However, many RNAs have well-defined tertiary structures (Sections 31-4A, 32-4Ca, 32-2B, and 32-3A).

In this section we examine the forces that give rise to the structures of nucleic acids. These forces are, of course, much the same as those that are responsible for the structures of proteins (Section 8-4) but, as we shall see, the way they combine gives nucleic acids properties that are quite different from those of proteins.

A. Sugar-Phosphate Chain Conformations

The conformation of a nucleotide unit, as Fig. 29-5 indicates, is specified by the six torsion angles of the sugar-phosphate backbone and the torsion angle describing the orientation of the base about the glycosidic bond. It would seem that these seven degrees of freedom per nucleotide would render polynucleotides highly flexible. Yet, as we shall see, these torsion angles are subject to a variety of internal constraints that greatly restrict their conformational freedom.

a. Torsion Angles About Glycosidic Bonds Have Only One or Two Stable Positions

The rotation of a base about its glycosidic bond is greatly hindered, as is best seen by the manipulation of a space-filling molecular model. Purine residues have two sterically permissible orientations relative to the sugar known as the **syn** (Greek: with) and **anti** (Greek: against) conformations (Fig. 29-6). For pyrimidines, only the anti conformation is easily formed because, in the syn conformation, the sugar residue sterically interferes with the pyrimidine's C2 substituent. In most double helical nucleic acids, all bases are in the anti conformation (Fig. 29-1a,c, left and middle panels). The exception is Z-DNA (Section 29-1Bb), in which the alternating pyrimidine and purine residues are anti and syn (Fig. 29-1a,c, right panels). *This explains Z-DNA's pyrimidine-purine alternation.* Indeed, the base pair flips that convert B-DNA to Z-DNA (Fig. 29-2) are brought about by rotating each purine base about its gly-

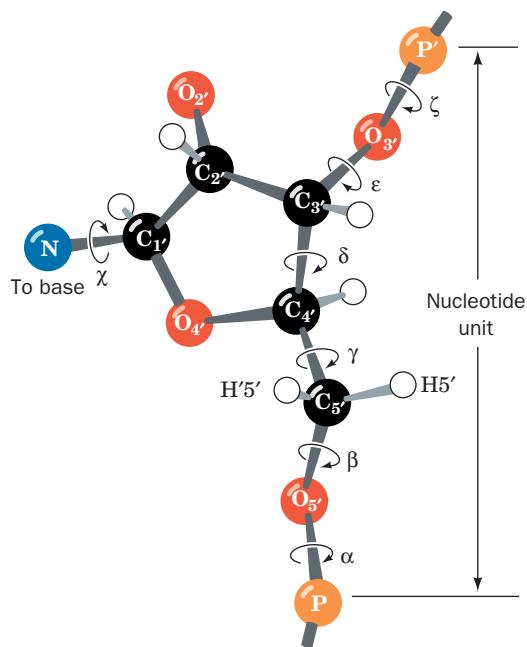


Figure 29-5 The conformation of a nucleotide unit is determined by the seven indicated torsion angles.

cosidic bond from the anti to syn conformation, whereas it is the sugars that rotate in the pyrimidine nucleotides, thereby maintaining them in their anti conformations.

b. Sugar Ring Pucker Is Largely Limited to Only a Few of Its Possible Arrangements

The ribose ring has a certain amount of flexibility that significantly affects the conformation of the sugar-phosphate backbone. The vertex angles of a regular pentagon are 108° , a value quite close to the tetrahedral angle (109.5°), so that one might expect the ribofuranose ring to be nearly flat. However, the ring substituents are eclipsed when the ring is planar. To relieve the resultant crowding, which even occurs between hydrogen atoms, the ring **puckers**; that is, it becomes slightly nonplanar, so as to reorient the ring substituents (Fig. 29-7; this is readily observed by the manipulation of a skeletal molecular model).

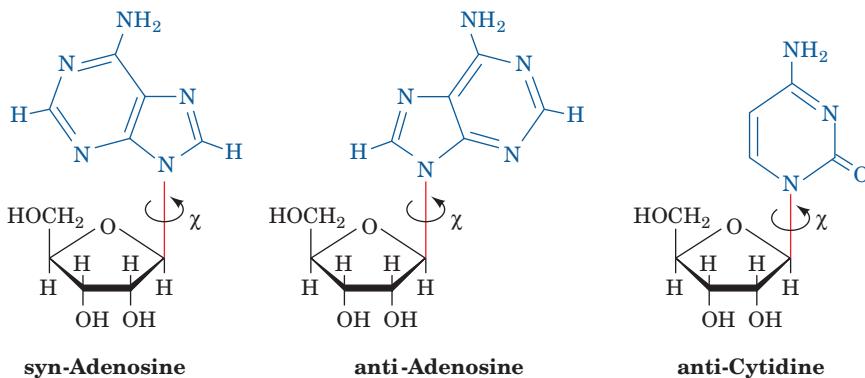


Figure 29-6 The sterically allowed orientations of purine and pyrimidine bases with respect to their attached ribose units.

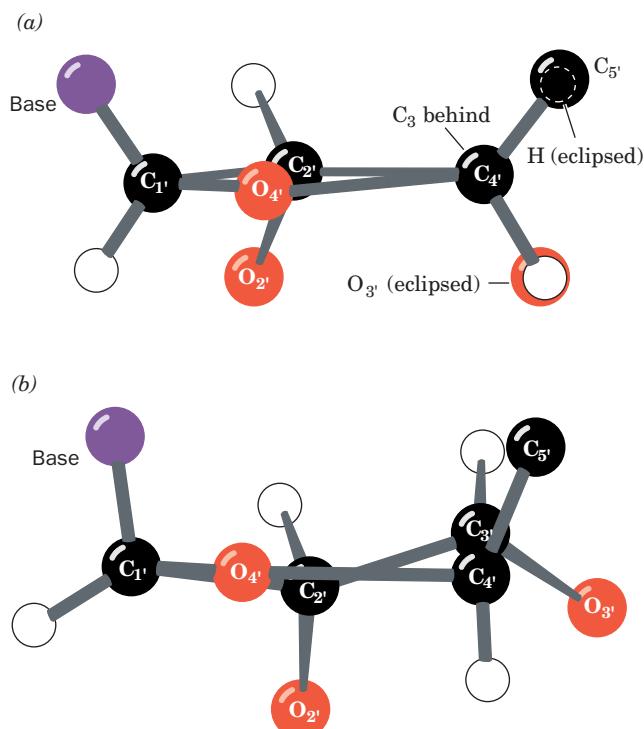


Figure 29-7 Ribose ring pucker. The substituents to (a) a planar ribose ring (here viewed down the C3'—C4' bond) are all eclipsed. The resulting steric strain is partially relieved by ring puckering such as in (b), a half-chair conformation in which C3' is the out-of-plane atom.

One would, in general, expect only three of a ribose ring's five atoms to be coplanar since three points define a plane. Nevertheless, in the great majority of the >50 nucleoside and nucleotide crystal structures that have been reported, four of the ring atoms are coplanar to within a few hundredths of an angstrom and the remaining atom is out of this plane by several tenths of an angstrom (the **half-chair** conformation).

If the out-of-plane atom is displaced to the same side of the ring as atom C5', it is said to have the **endo** conformation (Greek: *endon*, within), whereas displacement to the opposite side of the ring from C5' is known as the **exo** conformation (Greek: *exo*, out of). In the great majority of known nucleoside and nucleotide structures (molecules that are subject to few of the conformational constraints of double helices), the out-of-plane atom is either C2' or C3' (Fig. 29-8). C2'-*endo* is the most frequently occurring ribose pucker with C3'-*endo* and C3'-*exo* also being common. Other ribose conformations are rare.

The ribose pucker is conformationally important in nucleic acids because it governs the relative orientations of the phosphate substituents to each ribose residue (Fig. 29-8). For instance, it is difficult to build a regularly repeating model of a double helical nucleic acid unless the sugars are either C2'-*endo* or C3'-*endo*. In fact, canonical B-DNA has the C2'-*endo* conformation, whereas canonical A-DNA and RNA-11 are C3'-*endo*. In canonical Z-DNA, the purine nucleotides are all C3'-*endo* and the pyrimidine nucleotides are C2'-*endo*, which is another reason that the repeating unit of Z-DNA is a dinucleotide. The sugar pucksers observed in the X-ray structures of A-DNA are, in fact, almost entirely C3'-*endo*. However, those of B-DNA, although predominantly C2'-*endo*, exhibit significant variation including C4'-*exo*, O4'-*endo*, C1'-*exo*, and C3'-*exo*. This variation in B-DNA's sugar pucker is probably indicative of its greater flexibility relative to other types of DNA helices.

c. The Sugar-Phosphate Backbone Is Conformationally Constrained

If the torsion angles of the sugar-phosphate chain (Fig. 29-5) were completely free to rotate, there could probably be no stable nucleic acid structure. However, the comparison, by Muttaiya Sundaralingam, of some 40 nucleoside and nucleotide crystal structures has revealed that these angles are really quite restricted. For example, the torsion angle about the C4'—C5' bond (γ in Fig. 29-5) is rather narrowly distributed such that O4' usually has a gauche

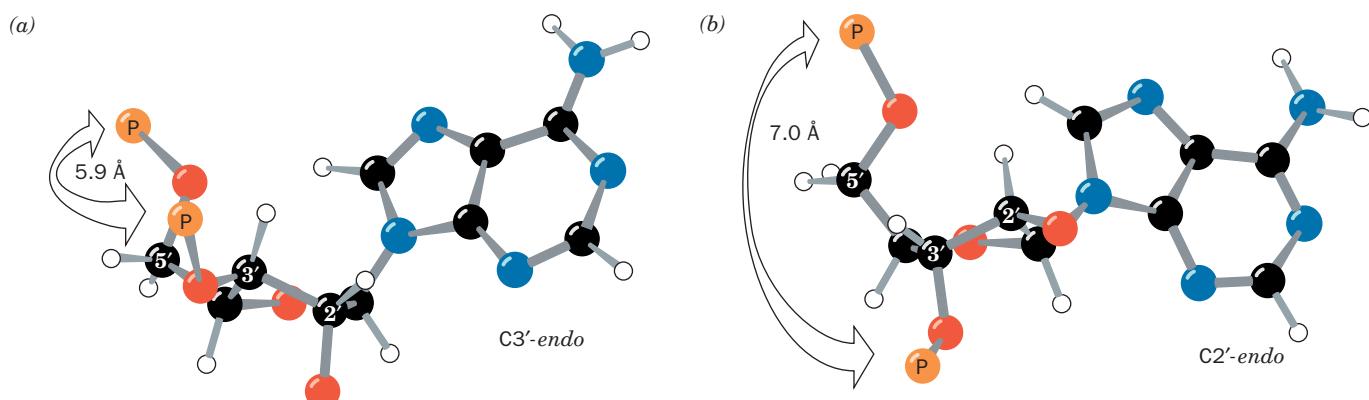


Figure 29-8 Nucleotide sugar conformations. (a) The C3'-*endo* conformation (on the same side of the sugar ring as C5'), which occurs in A-DNA and RNA-11. (b) The C2'-*endo* conformation, which occurs in B-DNA. The distances between adjacent P atoms

in the sugar-phosphate backbone are indicated. [After Saenger, W., *Principles of Nucleic Acid Structure*, p. 237, Springer-Verlag (1983).] See Kinemage Exercise 17-3

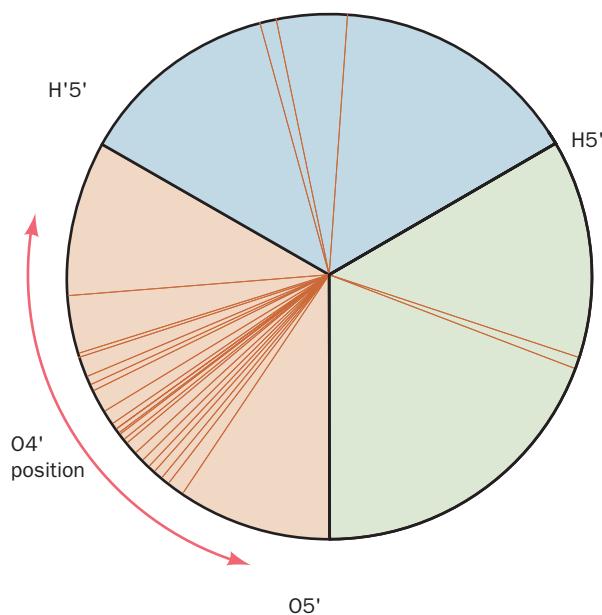


Figure 29-9 Conformational wheel showing the distribution of the torsion angle about the C4'—C5' bond. The torsion angle (γ in Fig. 29-5) was measured in 33 X-ray structures of nucleosides, nucleotides, and polynucleotides. Each radial line represents the position of the C4'—O4' bond in a single structure relative to the substituents of C5' as viewed from C5' to C4'. Note that most of the observed torsion angles fall within a relatively narrow range. [After Sundaralingam, M., *Biopolymers* 7, 838 (1969).]

conformation (having a torsion angle of $\sim 60^\circ$) with respect to O5' (Fig. 29-9). This is because the presence of the ribose ring together with certain noncovalent interactions of the phosphate group stiffens the sugar-phosphate chain by restricting its range of torsion angles. These restrictions are even greater in polynucleotides because of steric interference between residues.

The sugar-phosphate conformational angles of the various double helices are all reasonably strain free. *Double helices are therefore conformationally relaxed arrangements*.

ments of the sugar-phosphate backbone. Nevertheless, the sugar-phosphate backbone is by no means a rigid structure, so, on strand separation, it assumes a random coil conformation.

B. Base Pairing

Base pairing is apparently a “glue” that holds together double-stranded nucleic acids. Only Watson-Crick pairs occur in the crystal structures of self-complementary oligonucleotides. It is therefore important to understand how Watson-Crick base pairs differ from other doubly hydrogen bonded arrangements of the bases that have reasonable geometries (e.g., Fig. 29-10).

a. Unconstrained A · T Base Pairs Assume Hoogsteen Geometry

When monomeric adenine and thymine derivatives are cocrystallized, the A · T base pairs that form invariably have adenine N7 as the hydrogen bonding acceptor (**Hoogsteen geometry**; Fig. 29-10b) rather than N1 (Watson-Crick geometry; Fig. 5-12). This suggests that Hoogsteen geometry is inherently more stable for A · T pairs than is Watson-Crick geometry. Apparently steric and other environmental influences make Watson-Crick geometry the preferred mode of base pairing in double helices. A · T pairs with Hoogsteen geometry are nevertheless of biological importance; for example, they help stabilize the tertiary structures of tRNAs (Section 32-2Ba). In contrast, monomeric G · C pairs always cocrystallize with Watson-Crick geometry as a consequence of their triply hydrogen bonded structures.

b. Non-Watson-Crick Base Pairs Are of Low Stability

The bases of a double helix, as we have seen (Section 5-3A), associate such that any base pair position may interchangeably be A · T, T · A, G · C, or C · G without affecting the conformations of the sugar-phosphate chains. One might reasonably suppose that this requirement of **geometric complementarity** of the Watson-Crick base pairs, A with T and G with C, is the only reason that other

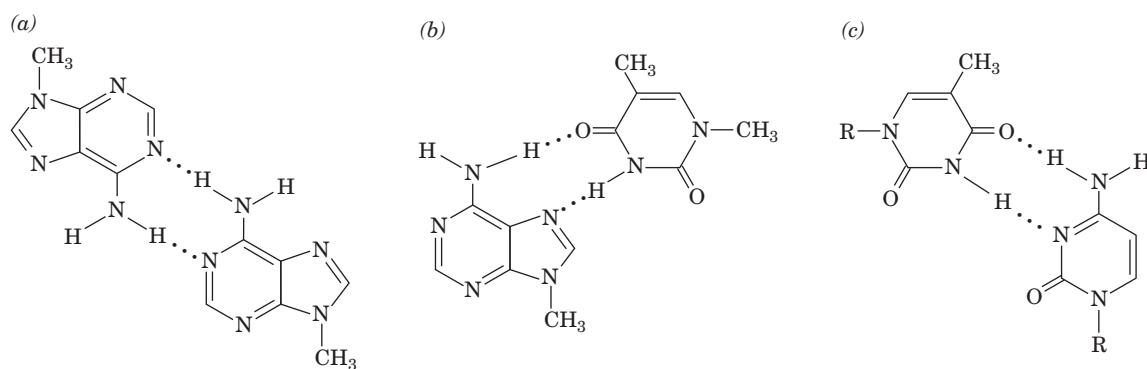


Figure 29-10 Some non-Watson-Crick base pairs. (a) The pairing of adenine residues in the crystal structure of 9-methyladenine. (b) Hoogsteen pairing between adenine and thymine residues in the crystal structure of 9-methyladenine ·

1-methylthymine. (c) A hypothetical pairing between cytosine and thymine residues. Compare these base pairs with the Watson-Crick base pairs in Fig. 5-12.

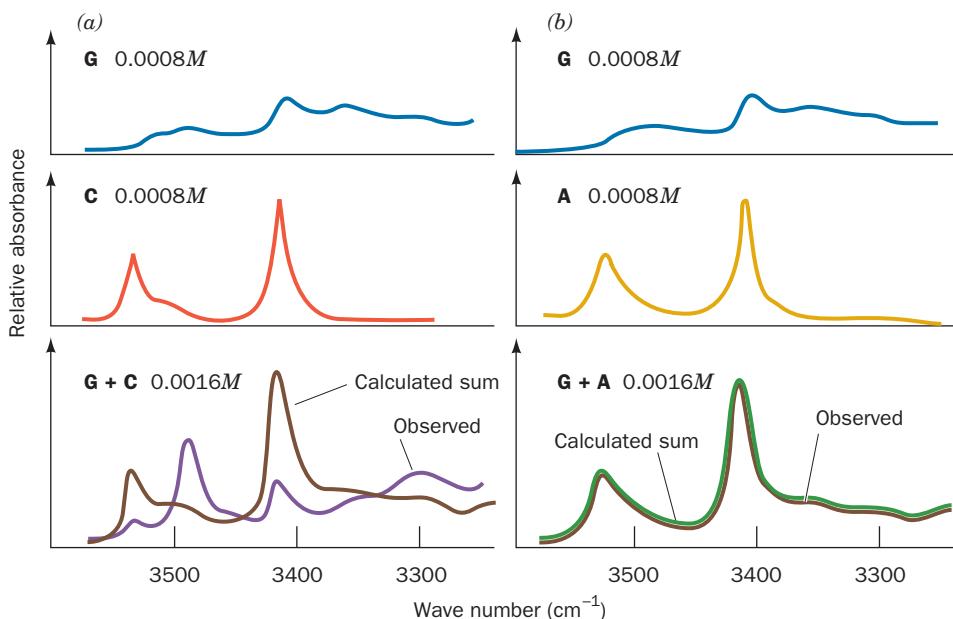


Figure 29-11 The infrared spectra, in the N—H stretch region, of guanine, cytosine, and adenine derivatives. The derivatives were analyzed both separately and in the indicated mixtures. The solvent, CDCl_3 , does not hydrogen bond with the bases and is relatively transparent in the frequency range of interest. (a) G + C. The brown curve in the lower panel, which is the sum of the spectra in the two upper panels, is the calculated spectrum of

G + C for noninteracting molecules. The band near 3500 cm^{-1} in the observed G + C spectrum (purple) is indicative of a specific hydrogen bonding association between G and C. (b) G + A. The close match between the calculated and observed spectra of the G + A mixture indicates that G and A do not significantly interact. [After Kyogoku, Y., Lord, R.C., and Rich, A., *Science* **154**, 5109 (1966).]

base pairs do not occur in a double helical environment. In fact, this was precisely what was believed for many years after the DNA double helix was discovered.

Eventually, the failure to detect pairs of different bases in nonhelical environments other than A with T (or U) and G with C led Richard Lord and Rich to demonstrate, through spectroscopic studies, that *only the bases of Watson–Crick pairs have a high mutual affinity*. Figure 29-11a shows the infrared (IR) spectrum in the N—H stretch region of guanine and cytosine derivatives, both separately and in a mixture. The band in the spectrum of the G + C mixture that is not present in the spectra of either of its components is indicative of a specific hydrogen bonding interaction between G and C. Such an association, which can occur between like as well as unlike molecules, may be described by ordinary mass action equations.



From analyses of IR spectra such as Fig. 29-11, the values of K for the various base pairs have been determined. The self-association constants of the Watson–Crick bases are given in the top of Table 29-2 (the hydrogen bonded association of like molecules is indicated by the appearance of new IR bands as the concentration of the molecule is increased). The bottom of Table 29-2 lists the association constants of the Watson–Crick pairs. Note that each of these latter quantities is larger than the self-association constants of either of their component bases, so that Watson–Crick

base pairs preferentially form from their constituents. In contrast, the non-Watson–Crick base pairs, A · C, A · G, C · U, and G · U, whatever their geometries, have association constants that are negligible compared with the self-pairing association constants of their constituents (e.g., Fig. 29-11b). *Evidently, a second reason that non-Watson–Crick base pairs do not occur in DNA double helices is that they have relatively little stability.* Conversely, the exclusive presence of Watson–Crick base pairs in DNA results, in part, from an **electronic complementarity** matching A to T and G to C. The theoretical basis of this electronic complementarity, which is an experimental observation, is obscure.

Table 29-2 Association Constants for Base Pair Formation

Base Pair	$K (M^{-1})^a$
Self-Association	
A · A	3.1
U · U	6.1
C · C	28
G · G	10^3 – 10^4
Watson–Crick Base Pairs	
A · U	100
G · C	10^4 – 10^5

^aData measured in deuteriochloroform at 25°C .

Source: Kyogoku, Y., Lord, R.C., and Rich, A., *Biochim. Biophys. Acta* **179**, 10 (1969).

This is because the approximations inherent in theoretical treatments make them unable to accurately account for the minor (few $\text{kJ} \cdot \text{mol}^{-1}$) energy differences between specific and nonspecific hydrogen bonding associations. The double helical segments of many RNAs, however, contain occasional non-Watson-Crick base pairs, most often G · U, which have functional as well as structural significance (e.g., Sections 32-2Ba and 32-2Db).

c. Hydrogen Bonds Only Weakly Stabilize DNA

It is clear that hydrogen bonding is required for the specificity of base pairing in DNA that is ultimately responsible for the enormous fidelity required to replicate DNA with almost no error (Section 30-3D). Yet, as is also true for proteins (Section 8-4Ba), *hydrogen bonding contributes little to the stability of the double helix*. For instance, adding the relatively nonpolar ethanol to an aqueous DNA solution, which strengthens hydrogen bonds, destabilizes the double helix, as is indicated by its decreased melting temperature (T_m ; Section 5-3Ca). This is because hydrophobic forces, which are largely responsible for DNA's stability (Section 29-2C), are disrupted by nonpolar solvents. In contrast, *the hydrogen bonds between the base pairs of native DNA are replaced in denatured DNA by energetically nearly equivalent hydrogen bonds between the bases and water*. This accounts for the thermodynamic observation that hydrogen bonding contributes only 2 to 8 kJ/mol to base pairing stability.

C. Base Stacking and Hydrophobic Interactions

Purines and pyrimidines tend to form extended stacks of planar parallel molecules. This has been observed in the structures of nucleic acids (e.g., Fig. 29-1) and in the several hundred reported X-ray crystal structures that contain nucleic acid bases. The bases in these structures are usually partially overlapped (e.g., Fig. 29-12). In fact, crystal structures of chemically related bases often exhibit similar stacking patterns. Apparently stacking interactions, which in the solid state are a form of van der Waals interaction (Section 8-4Ab), have some specificity, although certainly not as much as base pairing.

a. Nucleic Acid Bases Stack in Aqueous Solution

Bases aggregate in aqueous solution, as has been demonstrated by the variation of osmotic pressure with concentration. The van't Hoff law of osmotic pressure is

$$\pi = RTm \quad [29.2]$$

where π is the osmotic pressure, m is the molality of the solute (mol solute/kg solvent), R is the gas constant, and T is the temperature. The molecular mass, M , of an ideal solute can be determined from its osmotic pressure since $M = c/m$, where $c = \text{g solute/kg solvent}$.

If the species under investigation is of known molecular mass but aggregates in solution, Eq. [29.2] must be rewritten:

$$\pi = \phi RTm \quad [29.3]$$

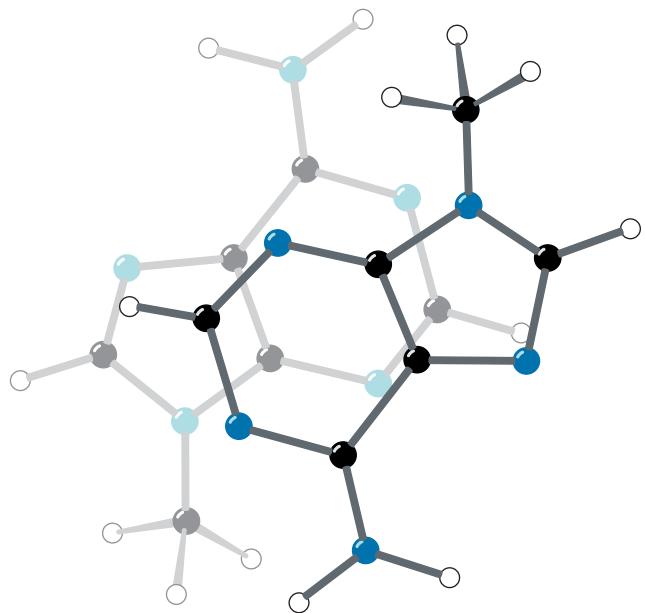
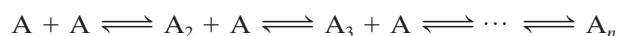
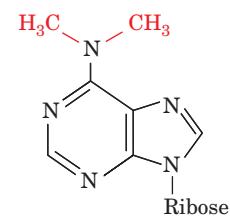


Figure 29-12 Stacking of adenine rings in the crystal structure of 9-methyladenine. The partial overlap of the rings is typical of the association between bases in crystal structures and in double helical nucleic acids. [After Stewart, R.F. and Jensen, L.H., *J. Chem. Phys.* **40**, 2071 (1964).]

where ϕ , the **osmotic coefficient**, indicates the solute's degree of association. ϕ varies from 1 (no association) to 0 (infinite association). The variation of ϕ with m for nucleic acid bases in aqueous solution (e.g., Fig. 29-13) is consistent with a model in which the bases aggregate in successive steps:



where n is at least 5 (if the reaction goes to completion, $\phi = 1/n$). This association cannot be a result of hydrogen bonding since **N^6, N^6 -dimethyladenosine**,



N^6, N^6 -Dimethyladenosine

which cannot form interbase hydrogen bonds, has a greater degree of association than does adenosine (Fig. 29-13). Apparently *the aggregation arises from the formation of stacks of planar molecules*. This model is corroborated by proton NMR studies: The directions of the aggregates' chemical shifts are compatible with a stacked but not a hydrogen bonded model. The stacking associations of monomeric bases are not observed in nonaqueous solutions.

Single-stranded polynucleotides also exhibit stacking interactions. For example, poly(A) shows a broad increase of

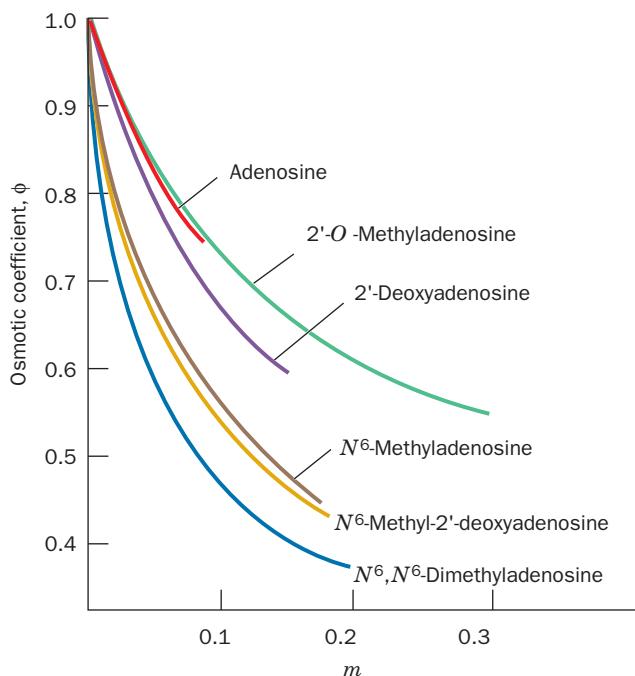
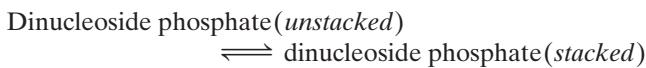


Figure 29-13 Variation of the osmotic coefficient ϕ with the molal concentrations m of adenosine derivatives in H_2O . The decrease of ϕ with increasing m indicates that these derivatives aggregate in solution. [After Broom, A.D., Schweizer, M.P., and Ts'o, P.O.P., *J. Am. Chem. Soc.* **89**, 3613 (1967).]

UV absorbance with temperature (Fig. 29-14a). This hyperchromism (which is indicative of nucleic acid denaturation; Section 5-3Ca) is independent of poly(A) concentration, so that it cannot be a consequence of intermolecular disaggregation. Likewise, it is not due to a reduction in intramolecular hydrogen bonding because poly(N^6,N^6 -dimethyladenosine) exhibits a greater degree of hyperchromism than does poly(A). The hyperchromism must therefore arise from some sort of stacking associations within a single strand that melt out with increasing temperature. This is not a very cooperative process, as is indicated by the broadness of the melting curve and the observation that short polynucleotides, including dinucleoside phosphates such as ApA, exhibit similar melting curves (Fig. 29-14b).

b. Nucleic Acid Structures Are Stabilized by Hydrophobic Forces

Stacking associations in aqueous solutions are largely stabilized by hydrophobic forces. One might reasonably suppose that hydrophobic interactions in nucleic acids are similar in character to those that stabilize protein structures. However, closer examination reveals that these two types of interactions are qualitatively different in character. Thermodynamic analysis of dinucleoside phosphate melting curves in terms of the reaction



(Table 29-3) indicates that base stacking is enthalpically

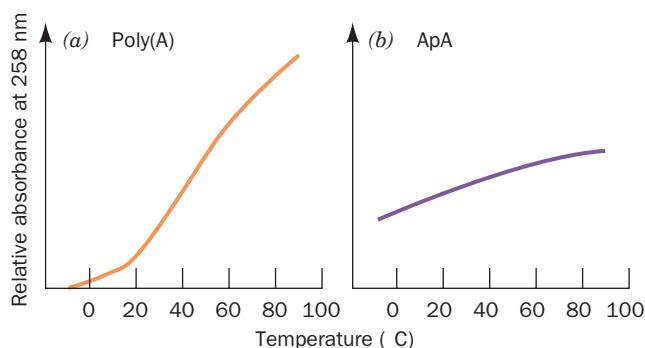
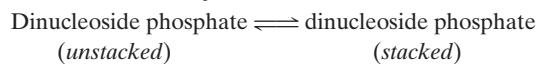


Figure 29-14 Melting curves for poly(A) and ApA. The broad temperature range of hyperchromic shifts at 258 nm of (a) poly(A) and (b) ApA is indicative of noncooperative conformational changes in these substances. Compare this figure with Fig. 5-16. [After Leng, M. and Felsenfeld, G., *J. Mol. Biol.* **15**, 457 (1966).]

driven and entropically opposed. Thus the hydrophobic interactions responsible for the stability of base stacking associations in nucleic acids are diametrically opposite in character to those that stabilize protein structures (which are enthalpically opposed and entropically driven; Section 8-4C). This is reflected in the differing structural properties of these interactions. For example, the aromatic side chains of proteins are almost never stacked and the crystal structures of aromatic hydrocarbons such as benzene, which resemble these side chains, are characteristically devoid of stacking interactions.

Hydrophobic forces in nucleic acids are poorly understood. The observation that they are different in character from the hydrophobic forces that stabilize proteins is nevertheless not surprising because the nitrogenous bases are considerably more polar than the hydrocarbon residues of proteins that participate in hydrophobic interactions. There is, however, no theory available that adequately explains the nature of hydrophobic forces in nucleic acids (our understanding of hydrophobic forces in proteins, it will be recalled, is similarly incomplete). They are complex interactions of which base stacking is probably a significant component. Whatever their origins, hydrophobic forces are of central importance in determining nucleic acid structures.

Table 29-3 Thermodynamic Parameters for the Reaction



Dinucleoside Phosphate	$\Delta H_{\text{stacking}}$ (kJ · mol ⁻¹)	$-\text{T} \Delta S_{\text{stacking}}$ (kJ · mol ⁻¹ at 25°C)
ApA	-22.2	24.9
ApU	-35.1	39.9
GpC	-32.6	34.9
CpG	-20.1	21.2
UpU	-32.6	36.2

Source: Davis, R.C. and Tinoco, I., Jr., *Biopolymers* **6**, 230 (1968).

D. Ionic Interactions

Any theory of the stability of nucleic acid structures must take into account the electrostatic interactions of their charged phosphate groups. Polyelectrolyte theory approximates the electrostatic interactions of DNA by considering the anionic double helix to be a homogeneously charged line or cylinder. We shall not discuss the details of this theory here, but note that it is often in reasonable agreement with experimental observations.

The melting temperature of duplex DNA increases with the cation concentration because these ions bind more tightly to duplex DNA than to single-stranded DNA due to the duplex DNA's higher anionic charge density. An increased salt concentration therefore shifts the equilibrium toward the duplex form, thus increasing the DNA's T_m . The observed relationship for Na^+ is

$$T_m = 41.1X_{G+C} + 16.6 \log[\text{Na}^+] + 81.5 \quad [29.4]$$

where X_{G+C} is the mole fraction of G + C base pairs (recall that T_m increases with the G + C content; Fig. 5-17); the equation is valid in the ranges $0.3 < X_{G+C} < 0.7$ and $10^{-3} M < [\text{Na}^+] < 1.0 M$. Other monovalent cations such as Li^+ and K^+ have similar nonspecific interactions with phosphate groups. Divalent cations, such as Mg^{2+} , Mn^{2+} , and Co^{2+} , in contrast, specifically bind to phosphate groups, so that *divalent cations are far more effective shielding agents for nucleic acids than are monovalent cations*. For example, an Mg^{2+} ion has an influence on the DNA double helix comparable to that of 100 to 1000 Na^+ ions. Indeed, enzymes that mediate reactions with nucleic acids or just nucleotides (e.g., ATP) usually require Mg^{2+} for activity. Moreover, Mg^{2+} ions play an essential role in stabilizing the complex structures assumed by many RNAs such as transfer RNAs (tRNAs; Section 32-2B) and ribosomal RNAs (rRNAs; Section 32-3A).

3 SUPERCOILED DNA

 See Guided Exploration 24: DNA supercoiling Genetic analyses indicate that numerous viruses and bacteria have circular genetic maps, which implies that their chromosomes are

likewise circular. This conclusion has been confirmed by electron micrographs in which circular DNAs are seen (Fig. 29-15). Some of these circular DNAs have a peculiar twisted appearance, a phenomenon that is known equivalently as **supercoiling**, **supertwisting**, and **superhelicity**. Supercoiling arises from a biologically important topological property of covalently closed circular duplex DNA that is the subject of this section. It is occasionally referred to as DNA's tertiary structure.

A. Superhelix Topology

Consider a double helical DNA molecule in which both strands are covalently joined to form a circular duplex molecule as is diagrammed in Fig. 29-16 (each strand can be joined only to itself because the strands are antiparallel). A geometric property of such an assembly is that the number of times one strand wraps about the other cannot be altered without first cleaving at least one of its polynucleotide strands. You can easily demonstrate this to yourself with a buckled belt in which each edge of the belt represents a strand of DNA. The number of times the belt is twisted before it is buckled cannot be changed without unbuckling or cutting the belt (cutting a polynucleotide strand).

This phenomenon, as James White proved mathematically in 1969, is expressed

$$L = T + W \quad [29.5]$$

in which:

1. L , the **linking number** (also symbolized Lk), is the number of times that one DNA strand winds about the other. This integer quantity is most easily counted when the molecule's duplex axis is constrained to lie in a plane (see below). However, *the linking number is invariant no matter how the circular molecule is twisted or distorted so long as both its polynucleotide strands remain covalently intact; the linking number is therefore a topological property of the molecule*.

2. T , the **twist** (also symbolized Tw), is the number of complete revolutions that one polynucleotide strand makes about the duplex axis in the particular conformation

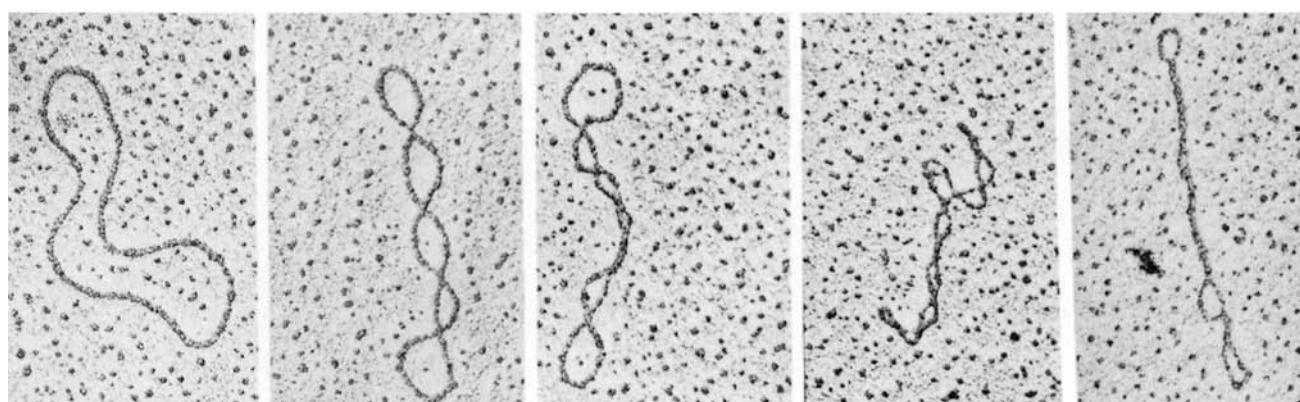


Figure 29-15 Electron micrographs of circular duplex DNAs. Their conformations vary from no supercoiling (left) to tightly supercoiled (right). [Electron micrographs by Laurien Polder.]

From Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 36, W.H. Freeman (1992). Used with permission.]

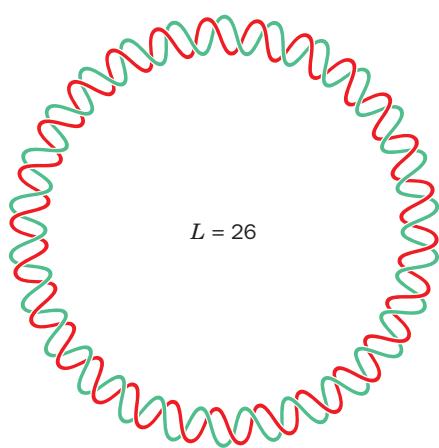


Figure 29-16 Schematic diagram of covalently closed circular duplex DNA that has 26 double helical turns. Its two polynucleotide strands are said to be **topologically bonded** to each other because, although they are not covalently linked, they cannot be separated without breaking covalent bonds.

under consideration. By convention, T is positive for right-handed duplex turns, so that, for B-DNA in solution, the twist is normally the number of base pairs divided by 10.5 (the number of base pairs per turn of the B-DNA double helix under physiological conditions; see Section 29-3Bc).

3. W , the **writhing number** (also symbolized Wr), is the number of turns that the duplex axis makes about the su-

perhelix axis in the conformation of interest. Its value is readily determined by projecting the DNA onto a plane and counting the number of times the duplex axis crosses itself (Fig. 29-17). The **writhing number** is a measure of the DNA's superhelicity. The difference between writhing and twisting is illustrated by the familiar example in Fig. 29-18. $W = 0$ when the DNA's duplex axis is constrained to lie in a plane (e.g., Fig. 29-16); then $L = T$, so L may be evaluated by counting the DNA's duplex turns.

The two DNA conformations diagrammed on the right of Fig. 29-19 are topologically equivalent; that is, they have the same linking number, L , but differ in their twists and writhing numbers. Note that T and W need not be integers (at least mathematically), only L . Although, strictly speaking, superhelicity is only defined for covalently closed circular duplex DNA, a linear segment of duplex DNA that is mechanically constrained from rotating at both ends (e.g., by protein anchors) has identical topological properties.

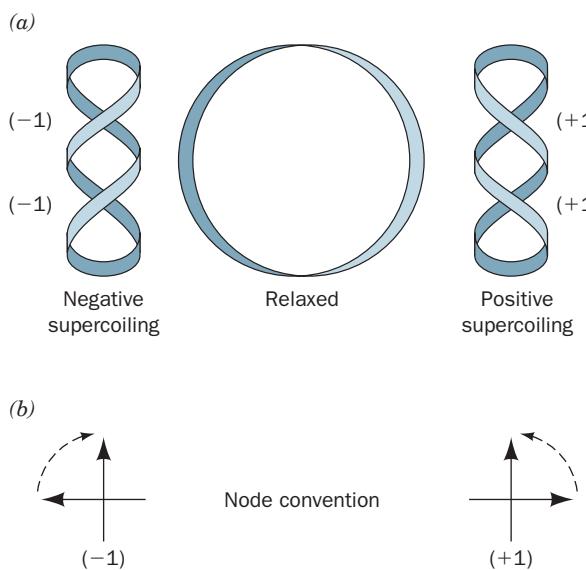


Figure 29-17 Topological relationships in covalently closed duplex DNA. (a) DNA molecules are represented by circular ribbons. DNA with no torsional strain is said to be relaxed. Underwinding or overwinding produces negative (-1) or positive $(+1)$ supercoils. (b) A negative writhe has a crossover in which a clockwise rotation of the front segment of $<180^\circ$ aligns it over the back segment, whereas a positive writhe has the corresponding counterclockwise rotation. [After Deweese, J.E., Osheroff, M.A., and Osheroff, N., *Biochem. Mol. Biol. Educ.* **37**, 2 (2009).]

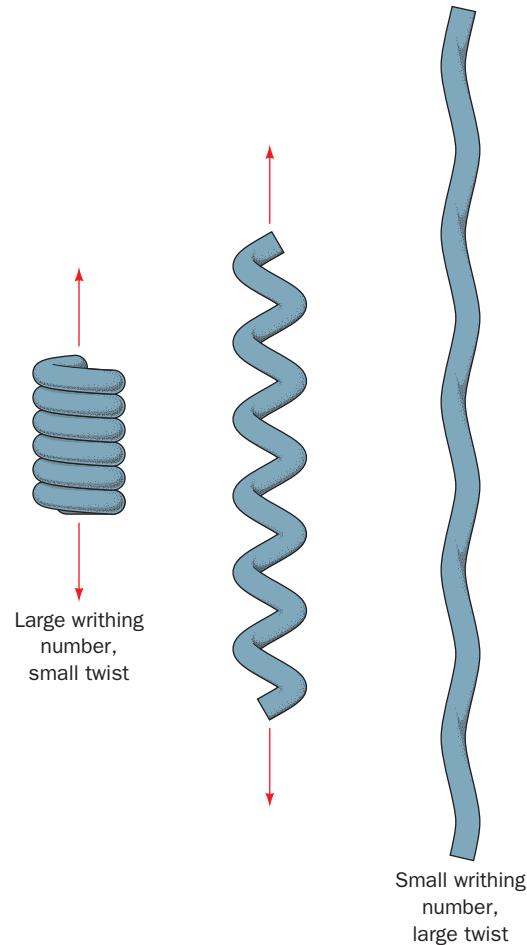


Figure 29-18 The difference between writhing and twist as demonstrated by a coiled telephone cord. Here the cord represents a double helical DNA molecule. In its relaxed state (left), the cord is in a helical form that has a large writhing number and a small twist. As the coil is pulled out (middle) until it is nearly straight (right), its writhing number becomes small as its twist becomes large.

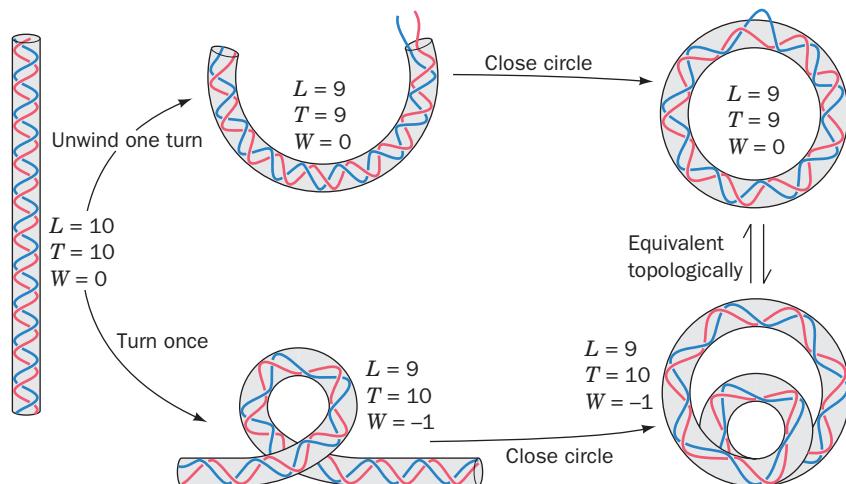


Figure 29-19 Two ways of introducing one supercoil into a DNA with 10 duplex turns.

The two closed circular forms shown (right) are topologically equivalent; that is, they are interconvertible without breaking any covalent bonds. The linking number L , twist T , and writhing number W are indicated for each form. Strictly speaking, the linking number is only defined for a covalently closed circle.

Since L is constant in an intact duplex DNA circle, for every new double helical twist, ΔT , there must be an equal and opposite superhelical twist, that is, $\Delta W = -\Delta T$. For example, a closed circular DNA without supercoils (Fig. 29-19, upper right) can be converted to a negatively supercoiled conformation (Fig. 29-18, lower right) by winding the duplex helix the same number of positive (right-handed) turns.

a. Supercoils May Be Toroidal or Interwound

A supercoiled duplex may assume two topologically equivalent forms:

1. **A toroidal helix** in which the duplex axis is wound as if about a cylinder (Fig. 29-20a).

2. **An interwound helix** in which the duplex axis is twisted around itself (Fig. 29-20b).

Note that these two interconvertible superhelical forms have opposite handedness. Since left-handed toroidal turns may be converted to left-handed duplex turns (e.g., Fig. 29-19), left-handed toroidal turns and right-handed interwound turns both have negative writhing numbers. Thus an underwound duplex ($T <$ number of bp/10.5), for example, will tend to develop right-handed interwound or left-handed toroidal superhelical turns when the constraints causing it to be underwound are released (the molecular forces in a DNA double helix promote its winding to its normal number of helical turns).

b. Supercoiled DNA Is Relaxed by Nicking One Strand

Supercoiled DNA may be converted to **relaxed circles** (as appears in the leftmost panel of Fig. 29-15) by treatment with **pancreatic DNase I**, an **endonuclease** (an enzyme that cleaves phosphodiester bonds within a polynucleotide strand) that cleaves only one strand of a duplex DNA. *One single-strand nick is sufficient to relax a supercoiled DNA.* This is because the sugar-phosphate chain opposite the nick is free to swivel about its backbone bonds (Fig. 29-5) so as to change the molecule's linking number and thereby alter its superhelicity. Supercoiling builds up elastic strain in a

DNA circle, much as it does in a rubber band. This is why the relaxed state of a DNA circle is not supercoiled.

B. Measurements of Supercoiling

Supercoiled DNA, far from being just a mathematical curiosity, has been widely observed in nature. In fact, its discovery in polyomavirus DNA by Jerome Vinograd stimulated the elucidation of the topological properties of superhelices rather than *vice versa*.

a. Intercalating Agents Control Supercoiling by Unwinding DNA

All naturally occurring DNA circles are underwound; that is, their linking numbers are less than those of their

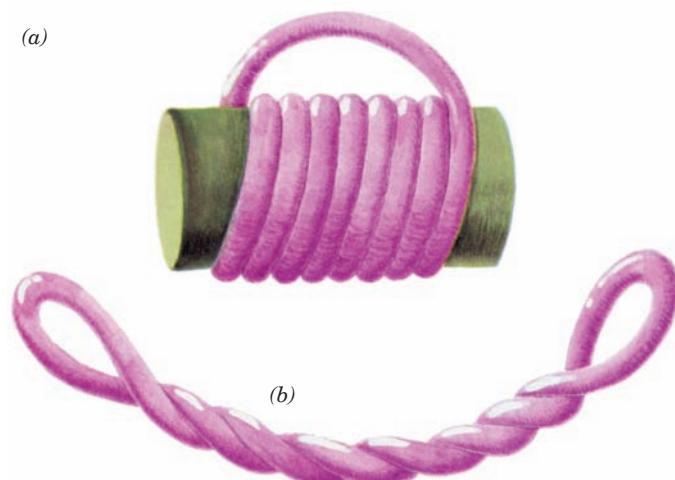


Figure 29-20 Toroidal and interwound supercoils. A rubber tube that has been (a) toroidally coiled in a left-handed helix around a cylinder with its ends joined such that it has no twist jumps to (b) an interwound helix with the opposite handedness when the cylinder is removed. Neither the linking number, the twist, nor the writhing number are changed in this transformation.

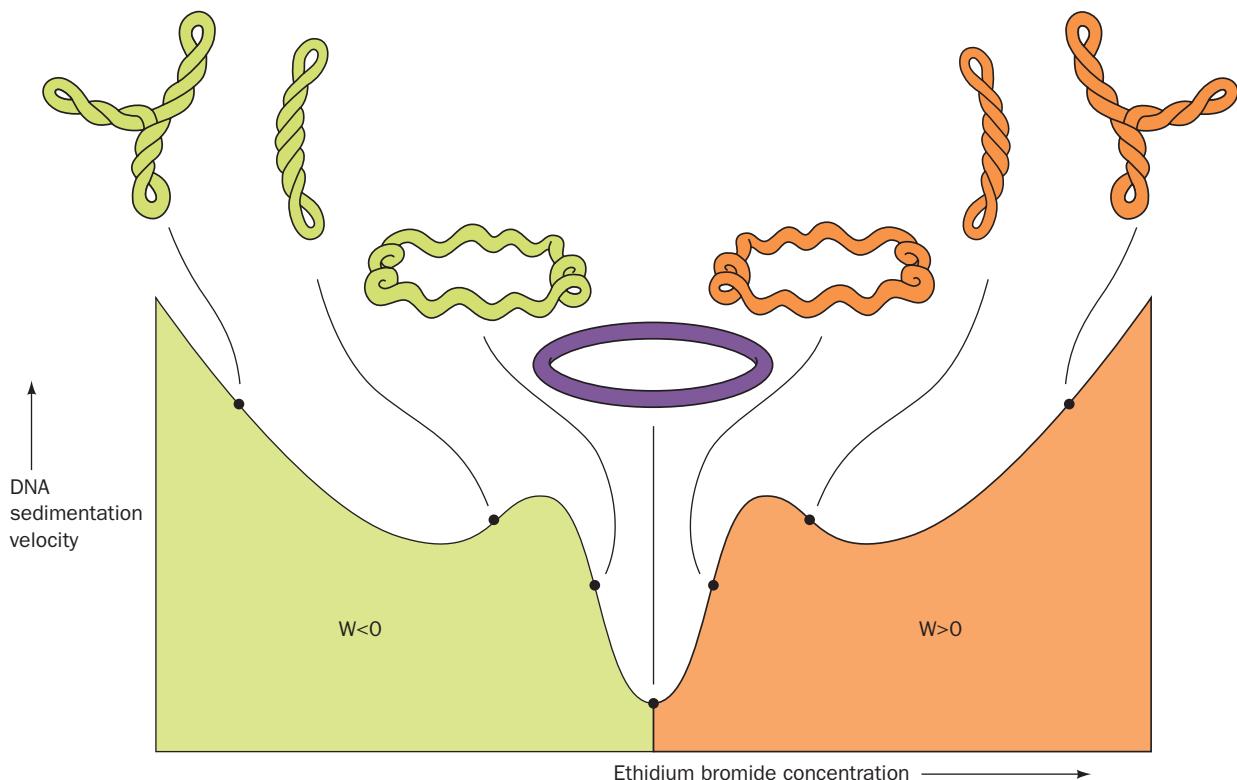


Figure 29-21 Sedimentation rate of underwound closed circular duplex DNA as a function of ethidium bromide concentration.

concentration. The intercalation of ethidium between the base pairs locally untwists the double helix (Fig. 29-22), which, since the linking number of the circle is constant, is accompanied by an equivalent increase in the writhing number. As the negatively coiled superhelix untwists, it becomes less compact and hence sediments more slowly. At the low point on the curve, the DNA

circles have bound sufficient ethidium to become fully relaxed. As the ethidium concentration is further increased, the DNA supercoils in the opposite direction, yielding a positively coiled superhelix. The supertwisted appearances of the depicted DNAs have been verified by electron microscopy. [After Bauer, W.R., Crick, F.H.C., and White, J.H., *Sci. Am.* **243**(1), 129 (1980). Copyright © 1981 by Scientific American, Inc.]

corresponding relaxed circles. This phenomenon has been established by observing the effect of ethidium ion binding on the sedimentation rate of circular DNA (Fig. 29-21). Intercalating agents such as ethidium (a planar aromatic cation; Section 6-6Ca) alter a circular DNA's degree

of superhelicity because they cause the DNA double helix to unwind (untwist) by $\sim 26^\circ$ at the site of the intercalated molecule (Fig. 29-22). $W < 0$ in an unconstrained underwound circle because of the tendency of a duplex DNA to maintain its normal twist of 1 turn per 10.5 bp. The titra-

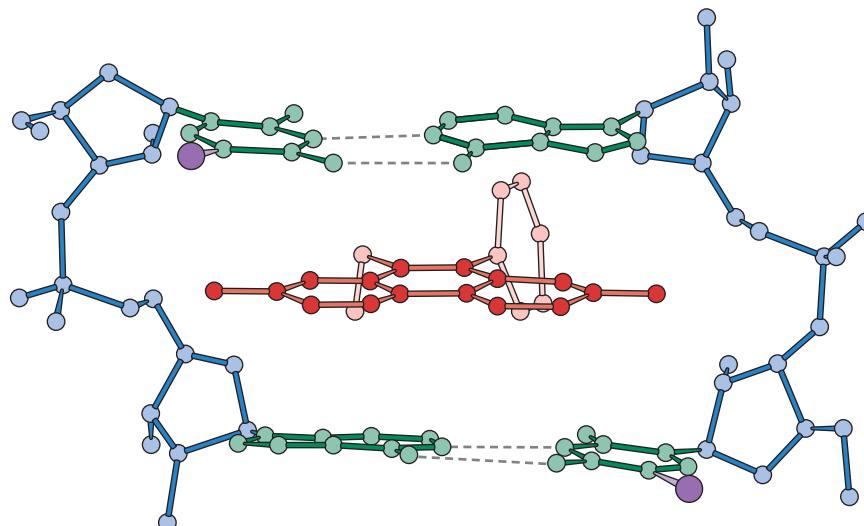


Figure 29-22 X-ray structure of a complex of ethidium with 5-iodo-UdA. Ethidium (red) intercalates between the base pairs (green with I purple) of the double helically paired dinucleoside phosphate and thereby provides a model for the binding of ethidium to duplex DNA. [After Tsai, C.-C., Jain, S.C., and Sobell, H.M., *Proc. Natl. Acad. Sci.* **72**, 629 (1975).]

tion of a DNA circle by ethidium unwinds the duplex (decreases T), which must be accompanied by a compensating increase in W . This, at first, lessens the superhelicity of an underwound circle. However, as the circle binds more and more ethidium, its value of W passes through zero (relaxed circles) and then becomes positive, so that the circle again becomes superhelical. Thus the sedimentation rate of underwound DNAs, which is a measure of their compactness and therefore their superhelicity, passes through a minimum as the ethidium concentration increases. This is what is observed with native DNAs (Fig. 29-21). In contrast, the sedimentation rate of an overwound circle would only increase with increasing ethidium concentration.

b. DNAs Are Separated According to Their Linking Number by Gel Electrophoresis

Gel electrophoresis (Sections 6-4 and 6-6C) also separates similar molecules on the basis of their compactness, so that the rate of migration of a circular duplex DNA increases with its degree of superhelicity. The agarose gel electrophoresis pattern of a population of chemically identical DNA molecules with different linking numbers therefore consists of a series of discrete bands (Fig. 29-23). The molecules in a given band all have the same linking number and differ from those in adjacent bands by $\Delta L \pm 1$.

Comparison of the electrophoretic band patterns of **simian virus 40 (SV40)** DNA that had been enzymatically relaxed to varying degrees and then resealed (Fig. 29-23) reveals that 26 bands separate native from fully relaxed SV40 DNAs. Native SV40 DNA therefore has $W = -26$ (although it is somewhat heterogeneous in this quantity). Since SV40 DNA consists of 5243 bp, it has 1 negative superhelical turn per ~ 19 duplex turns. Such a **superhelix density** (W/T) is typical of circular DNAs from various biological sources.

c. DNA in Physiological Solution Has 10.5 Base Pairs per Turn

The insertion, using genetic engineering techniques (Section 5-5C), of an additional x base pairs into a superhelical DNA with a given linking number will increase the DNA's twist and hence decrease its writhing number by x/h° , where h° is the number of base pairs per duplex turn. Such an insertion shifts the position of each band in the DNA's gel electrophoretic pattern by x/h° of the spacing between bands. By measuring the effects of several such insertions, James Wang established that $h^\circ = 10.5 \pm 0.1$ bp for B-DNA in solution under physiological conditions.

C. Topoisomerases

The normal biological functioning of DNA occurs only if it is in the proper topological state. In such basic biological processes as RNA transcription and DNA replication, the recognition of a base sequence requires the local separation of complementary polynucleotide strands. The negative supercoiling of naturally occurring DNAs results in a

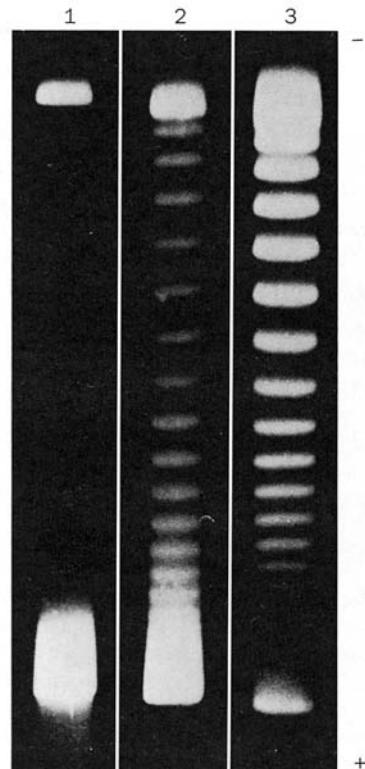


Figure 29-23 Agarose gel electrophoresis pattern of SV40 DNA. Lane 1 contains the negatively supercoiled native DNA (lower band; the DNA was applied to the top of the gel). In lanes 2 and 3, the DNA was exposed for 5 and 30 min, respectively, to an enzyme, known as a type IA topoisomerase (Section 29-3C), that relaxes negative supercoils one at a time by increasing the DNA's linking number (L). The DNAs in consecutively higher bands of a given gel have successively increasing linking numbers ($\Delta L = +1$). [From Keller, W., Proc. Natl. Acad. Sci. **72**, 2553 (1975).]

torsional strain that promotes such separations since it tends to unwind the duplex helix (an increase in T must be accompanied by a decrease in W). If DNA lacks the proper superhelical tension, the above vital processes (which themselves supercoil DNA; Sections 30-2C and 31-2Ca) occur quite slowly, if at all.

The supercoiling of DNA is controlled by a remarkable group of enzymes known as **DNA topoisomerases** (or simply **topoisomerases**). They are so named because they alter the topological state (linking number) of circular DNA but not its covalent structure. There are two classes of topoisomerases:

1. Type I topoisomerases, the first of which was discovered by James Wang in 1971, act by creating transient single-strand breaks in DNA. Type I enzymes are subclassified into **types IA, IB, and IC topoisomerases** on the basis of their amino acid sequences and reaction mechanisms (see below). Type I topoisomerases are denoted by odd Roman numerals (e.g., topoisomerase I, III, etc.).

2. Type II topoisomerases, the first of which was discovered by Martin Gellert in 1976, act by making transient double-strand breaks in DNA with the accompanying hydrolysis of ATP to ADP + P_i . Type II enzymes are subclassified into **types IIA** and **IIB topoisomerases** on the basis of their amino acid sequences. Type II topoisomerases are denoted by even Roman numerals (e.g., topoisomerase II, IV, etc.).

a. Type I Topoisomerases Incrementally Relax Supercoiled DNA

Type I topoisomerases catalyze the relaxation of supercoils in DNA by changing their linking number in increments of one turn until the supercoil is entirely relaxed. In most organisms, type IA enzymes, which are present in all cells, relax only negatively supercoiled DNA, whereas type IB enzymes, which are present in all eukaryotes and many prokaryotes (but not *E. coli*), relax both negatively and positively coiled DNA. However, many hyperthermophiles, both eubacteria and archaea, have a type IA

topoisomerase known as **reverse gyrase** that induces positive supercoiling in DNA through the ATP-driven action of a **helicase** domain that is fused to the N-terminus of the topoisomerase domain (helicases are discussed in Section 30-2C). This suggests that positive supercoiling, which tightens the DNA double helix, protects DNA from thermal denaturation. Although types IA and IB topoisomerases are both monomeric, ~100-kD enzymes, they share no apparent sequence or structural similarities and function, as we shall see, via different enzymatic mechanisms.

A clue to the mechanism of type IA topoisomerase was provided by the observation that it reversibly **catenates** (interlinks) single-stranded circles (Fig. 29-24a). Apparently the enzyme operates by cutting a single strand, passing a single-strand loop through the resulting gap, and then resealing the break (Fig. 29-24b), thereby twisting double helical DNA by one turn. In support of this **strand passage** mechanism, the denaturation of type IA enzyme that has been incubated with single-stranded circular DNA yields a linear DNA that has its 5'-terminal phosphoryl group linked to the enzyme via a phosphoTyr diester linkage.

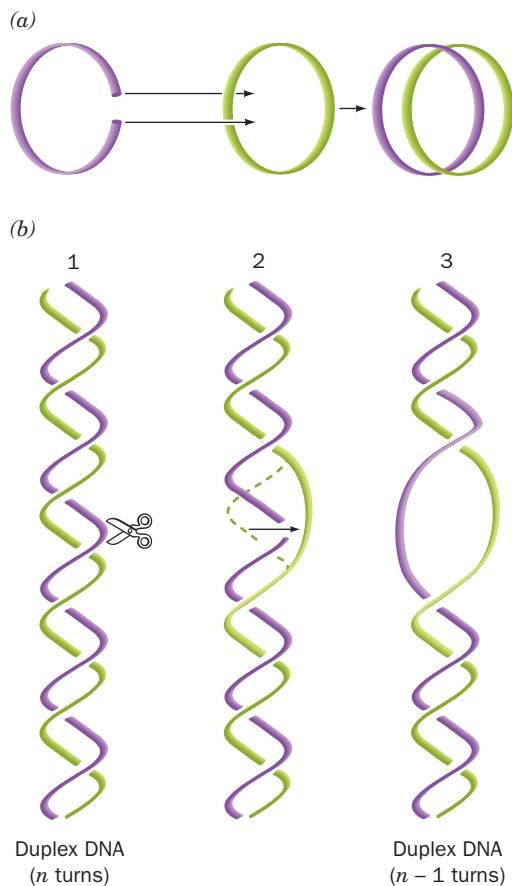
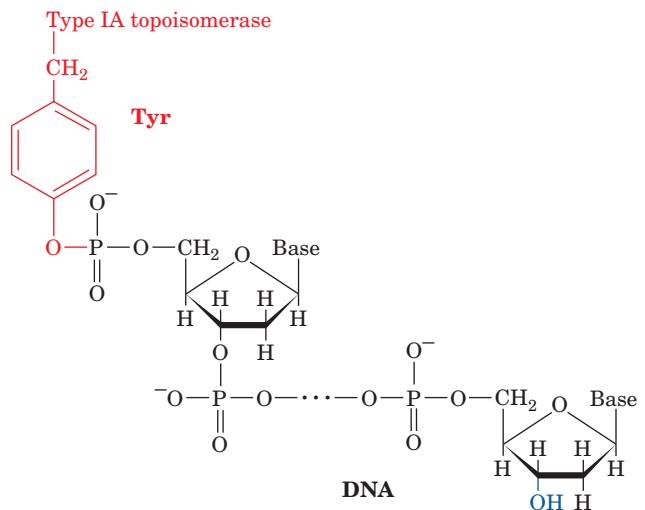


Figure 29-24 Type IA topoisomerase action. By cutting a single-stranded DNA, passing a loop of a second strand through the break, and then resealing the break, a type IA topoisomerase can (a) catenate two single-stranded circles or (b) unwind duplex DNA by one turn.



In contrast, denatured type IB enzyme is linked to the 3' end of DNA via a phosphoTyr linkage. By forming such covalent enzyme-DNA intermediates, the free energy of the cleaved phosphodiester bond is preserved, so that no energy input is required to reseal the nick.

b. Type IA Topoisomerase Functions via a Strand Passage Mechanism

Cells of *E. coli* contain two type IA topoisomerases named **topoisomerase I** (also called ω protein) and **topoisomerase III**. Topoisomerase III's Tyr 328 is the active site residue that forms a 5'-phosphoTyr linkage with the cleaved DNA. The X-ray structure of the inactive Y328F mutant of topoisomerase III in complex with the single-stranded octanucleotide d(CGCAACTT), determined by Alfonso

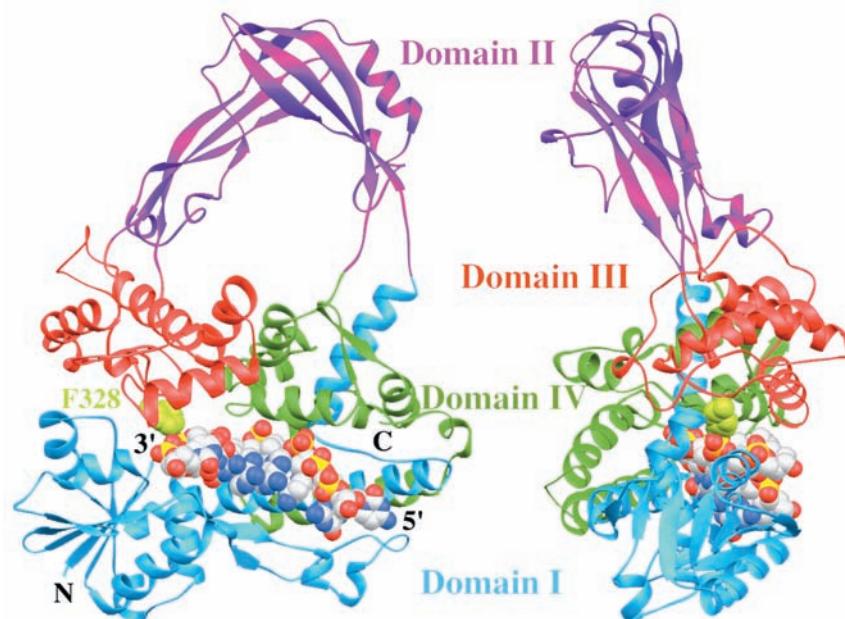


Figure 29-25 X-ray structure of the Y328F mutant of *E. coli* topoisomerase III, a type IA topoisomerase, in complex with the single-stranded octanucleotide d(CGCAACTT). The two views shown are related by a 90° rotation about a vertical axis. The protein's four domains are drawn in different colors. The DNA is

drawn in space-filling form with C white, N blue, O red, and P yellow. The enzyme's active site is marked by the side chain of Phe 328, which is shown in space-filling form in yellow-green. [Based on an X-ray structure by Alfonso Mondragón, Northwestern University. PDBid 1I7D.]

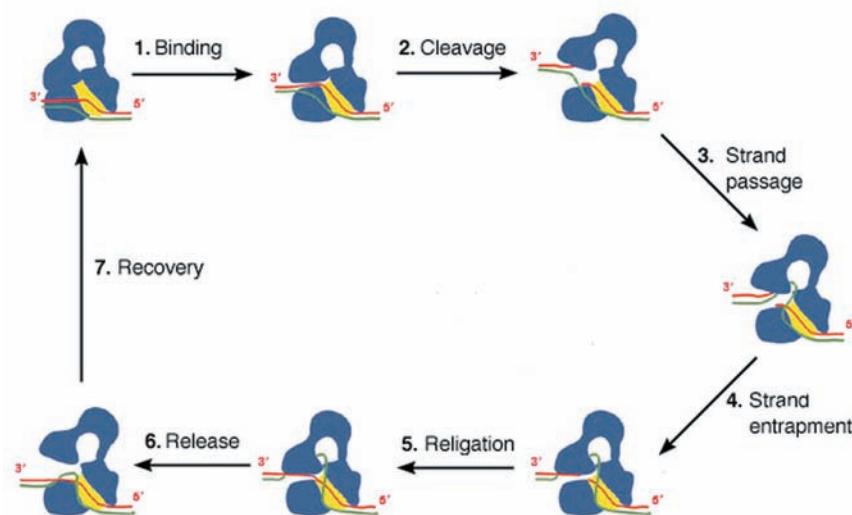


Figure 29-26 Proposed mechanism for the strand passage reaction catalyzed by type IA topoisomerases. The enzyme is shown in blue with the yellow patch representing the binding groove for single-stranded (ss) DNA. The two DNA strands, which are drawn in red and green, could represent the two strands of a covalently closed circular duplex or two ss circles. (1) The protein recognizes a ss region of the DNA, here the red strand, and binds it in its binding groove. This is followed by or occurs simultaneously with the opening of a gap between domains I and III. (2) The DNA is cleaved with the newly formed 5' end becoming covalently linked to the active site Tyr and the segment with the newly formed 3' end remaining tightly but noncovalently bound in the binding groove. (3) The unbroken (green) strand is passed through the opening or gate formed by

the cleaved (red) strand to enter the protein's central hole. (4) The unbroken strand is trapped by the partial closing of the gap. (5) The two cleaved ends of the red strand are rejoined in what is probably a reversal of the cleavage reaction. (6) The gap between domains I and III reopens to permit the escape of the red strand, yielding the reaction product in which the green strand has been passed through a transient break in the red strand. (7) The enzyme returns to its initial state. If the two strands form a negatively supercoiled duplex DNA, its linking number, L , has increased by 1; if they are separate ss circles, they have been catenated or decatenated. For duplex DNA, this process can be repeated until all of its supercoils have been removed ($W = 0$). [After a drawing by Alfonso Mondragón, Northwestern University.]

Mondragón (Fig. 29-25), reveals that this 659-residue monomer folds into four domains which enclose an \sim 20 by 28 Å hole that is large enough to contain a duplex DNA and which is lined with numerous Arg and Lys side chains. The octanucleotide binds in a groove that is also lined with Arg and Lys side chains with its sugar-phosphate backbone in contact with the protein and with most of its bases exposed for possible base pairing. Curiously, this single-stranded DNA assumes a B-DNA-like conformation even though its complementary strand would be sterically excluded from the groove. The DNA strand is oriented with its 3' end near the active site, where, if the mutant Phe 328 were the wild-type Tyr, its side chain would be properly positioned to nucleophilically attack the phosphate group bridging the DNA's C6 and T7 to form a 5'-phosphoTyr linkage with T7 and release C6 with a free 3'-OH. This structure and that of the homologous and structurally similar *E. coli* topoisomerase I suggest the mechanism for the type IA topoisomerase-catalyzed strand passage reaction that is diagrammed in Fig. 29-26.

c. Type IB Topoisomerase Functions via a Controlled Rotation Mechanism

Human topoisomerase I is a 765-residue type IB topoisomerase (and hence is unrelated to *E. coli* topoisomerase I). It mediates the transient cleavage of one strand of a duplex DNA through the nucleophilic attack of Tyr 723 on a DNA P atom to yield a 3'-linked phosphoTyr diester bond and a free 5'-OH group on the succeeding nucleotide. Limited proteolysis studies revealed that topoisomerase I consists of four major regions: its N-terminal, core, linker, and C-terminal domains. The \sim 210-residue, highly polar, N-terminal domain, which is poorly conserved, contains several nuclear targeting signals and is dispensable for enzymatic activity.

The X-ray structure of the catalytically inactive Y723F mutant of topoisomerase I lacking its N-terminal 214

residues and in complex with a 22-bp palindromic duplex DNA was determined by Wim Hol (Fig. 29-27). The core domain of this bilobal protein is wrapped around the DNA in a tight embrace. If the mutant Phe 723 were the wild-type Tyr, its OH group would be colinear with the scissile P—O5' bond and hence ideally positioned to nucleophilically attack this P atom so as to form a covalent linkage with the 3' end of the cleaved strand. As expected, the protein interacts with the DNA in a largely sequence independent manner: Of the 41 direct contacts that the protein makes to the DNA, 37 are protein-phosphate interactions and only one is base-specific. The protein interacts to a much greater extent with the five base pairs of the DNA's upstream segment (which would contain the cleaved strand's newly formed 5' end; 29 of the 41 contacts) than it does with the base pairs of the DNA's downstream segment (to which Tyr 723 would be covalently linked; 12 of the 41 contacts).

Topoisomerase I does not seem sterically capable of unwinding supercoiled DNA via the strand passage mechanism that type IA topoisomerases appear to follow (Fig. 29-26). Rather, as is diagrammed in Fig. 29-28, it is likely that topoisomerase I relaxes DNA supercoils by permitting the cleaved duplex DNA's loosely held downstream segment to rotate relative to the tightly held upstream segment. This rotation can only occur about the sugar-phosphate bonds in the uncleaved strand (α , β , γ , ϵ , and ζ in Fig. 29-5) that are opposite the cleavage site because the cleavage frees these bonds to rotate. In support of this mechanism, the protein region surrounding the downstream segment contains 16 conserved, positively charged residues that form a ring about this duplex DNA, which would presumably hold the DNA in the ring but not in any specific orientation. Nevertheless, the downstream segment is unlikely to rotate freely because the cavity containing it is shaped so as to interact with the downstream segment during some portions of its rotation. Hence,

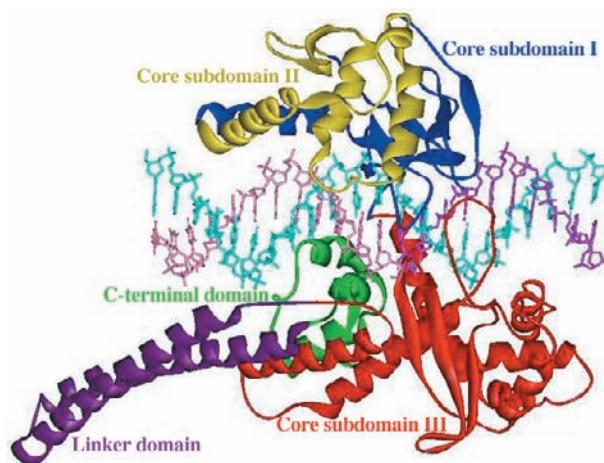
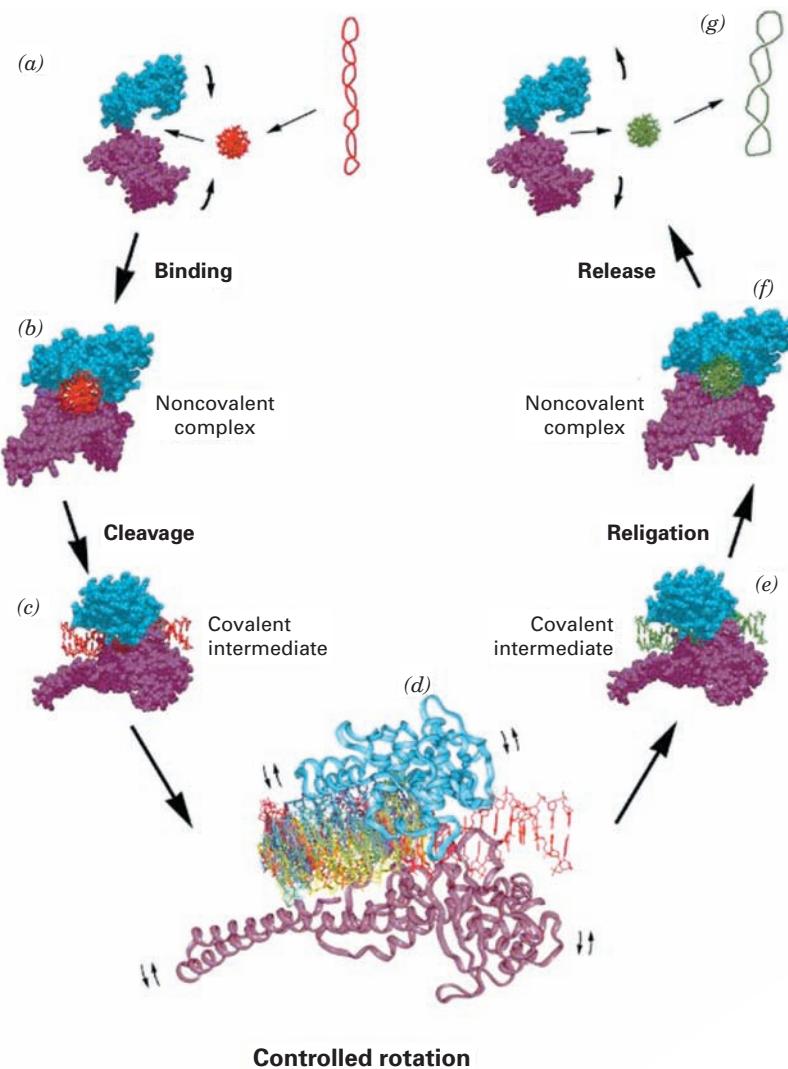


Figure 29-27 X-ray structure of the N-terminally truncated, Y723F mutant of human topoisomerase I in complex with a 22-bp duplex DNA. The protein's various domains and subdomains are drawn in different colors. The DNA's uncleaved

strand is cyan, and the upstream and downstream portions of the scissile strand are purple and pink, respectively. [Courtesy of Wim Hol, University of Washington. PDBid 1A36.]

Figure 29-28 Controlled rotation mechanism for type IB topoisomerases. A highly negatively supercoiled DNA (red, with a right-handed writhe) is converted, via stages (a) through (g), to a less supercoiled form (green). Topoisomerase I is drawn as a bilobal space-filling structure, in which the cyan lobe is formed by core subdomains I and II (Fig. 29-27) and the purple lobe is formed by core subdomain III, the linker domain, and the C-terminal domain. The structure shown in (d), which is expanded by a factor of 2, shows the downstream portion of the rotating DNA (that containing the cleaved strand's new 5' end) at 30° intervals, all differently colored. Since the enzyme is not always in direct contact with the rotating DNA, small rocking motions of the protein (small curved arrows) may accompany the controlled rotation. [Courtesy of Wim Hol, University of Washington.]



type IB topoisomerases are said to mediate a **controlled rotation** mechanism in relaxing supercoiled DNA. This unwinding is driven by the superhelical tension in the DNA and hence requires no other energy input. Eventually, the DNA is religated by a reversal of the cleavage reaction and the now less supercoiled DNA is released.

Type IC topoisomerase, whose only known family member, **topoisomerase V**, occurs exclusively in archaea, resembles type IB topoisomerases (which do not occur in archaea) in that it forms 3'-phosphoTyr intermediates and appears to function via a controlled rotation mechanism. However, it has no sequence or structural resemblance to type IB topoisomerases.

d. Type II Topoisomerases Function via a Strand Passage Mechanism

Bacteria have two types of type IIA topoisomerases: **DNA gyrase** (or just **gyrase**) and **topoisomerase IV**, both of which are A_2B_2 heterotetramers. Eukaryotic type IIA topoisomerases, which are named **topoisomerase II**, are homologous to bacterial type IIA topoisomerases but with

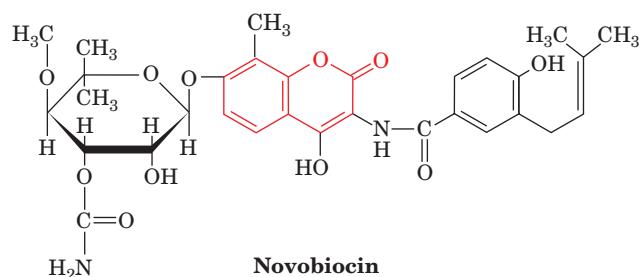
their A and B subunits fused so that they are homodimers. The type IIB topoisomerase family has only one member, **topoisomerase VI**, an A_2B_2 heterotetramer that occurs mainly in archaea (although some archaea express both type IIA and IIB topoisomerases). The A subunits of types IIA and IIB topoisomerases share a common ATPase module, but their B subunits are unrelated.

Gyrase is unique among topoisomerases in that it generates negative supercoils in DNA. All other topoisomerases but reverse gyrase only relax supercoils (DNA supercoiling in eukaryotes is generated differently from that in prokaryotes; Section 34-1Ba). It's A and B subunits are named **GyrA** and **GyrB**.

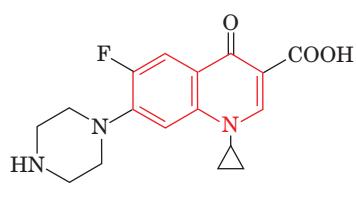
Type II topoisomerases can also catenate and decatenate double-stranded circles as well as tie and untie knots in them. The importance of this function can be seen as follows. The ~6.1 billion base pairs of DNA in a diploid human cell have an aggregate length of ~2 m and are confined to a nucleus that is 5 to 10 μm in diameter. Imagine that the ~20- \AA -wide DNA was expanded by a factor of 5 million to the width of a 1-cm-diameter rope. It would

then be \sim 10,000 km long (the distance from San Francisco to Rome) and confined to an expanded nucleus that is 25 to 50 m in diameter. Then imagine trying to manipulate the ropes in this system without generating tangles or knots. Yet such difficulties normally occur in cells during DNA replication, repair, and recombination (Chapter 30), as well as the segregation of the daughter chromosomes in dividing cells (Section 1-4B). It is the function of type II topoisomerases to untangle the DNA during these processes.

Bacterial but not eukaryotic type IIA topoisomerases are inhibited by a variety of substances including **novobiocin**, a member of the *Streptomyces*-derived **coumarin** family of antibiotics, and **ciprofloxacin** (trade name **Cipro**), a member of the synthetically generated **quinolone** family of antibiotics (their coumarin and quinolone groups are drawn in red):



Novobiocin



Ciprofloxacin

These agents profoundly inhibit bacterial DNA replication and RNA transcription, thereby demonstrating the importance of properly supercoiled DNA in these processes. Studies using *E. coli* gyrase mutants resistant to these substances have demonstrated that ciprofloxacin associates with GyrA and novobiocin binds to GyrB.

The gel electrophoretic pattern of duplex circles that have been exposed to gyrase shows a band pattern in which the linking numbers differ by increments of 2 rather than 1, as occurs with type I topoisomerases. Evidently, gyrase acts by cutting both strands of a duplex, passing the duplex through the break, and resealing it (Fig. 29-29). This hypothesis is corroborated by the observation that when gyrase is incubated with DNA and ciprofloxacin, and subsequently denatured with guanidinium chloride, a GyrA subunit remains covalently linked to the 5' end of each of the two cut strands through a phosphoTyr linkage. These cleavage sites are staggered by 4 bp, thereby yielding sticky ends.

Saccharomyces cerevisiae (baker's yeast) topoisomerase II is a homodimer of 1428-residue subunits whose N- and C-terminal segments are homologous to *E. coli*'s GyrB (804 residues) and GyrA (878 residues) subunits, respectively. The breakage/reunion domain, which encompasses residues 410 to 1202, can, by itself, cleave duplex DNA but cannot transport it through the break without the action of the enzyme's ATPase domain (residues 1–409). However, the C-terminal segment (residues 1203–1428), which is poorly conserved, appears to be dispensable.

Although the structure of a full length type IIA topoisomerase is as yet unknown, those of its ATPase and breakage/reunion domains from both *E. coli* gyrase and yeast topoisomerase II have been determined. The X-ray structure of the homodimeric topoisomerase II ATPase in complex with the nonhydrolyzable ATP analog AMPPNP, determined by James Berger, consists of two domains (Fig.

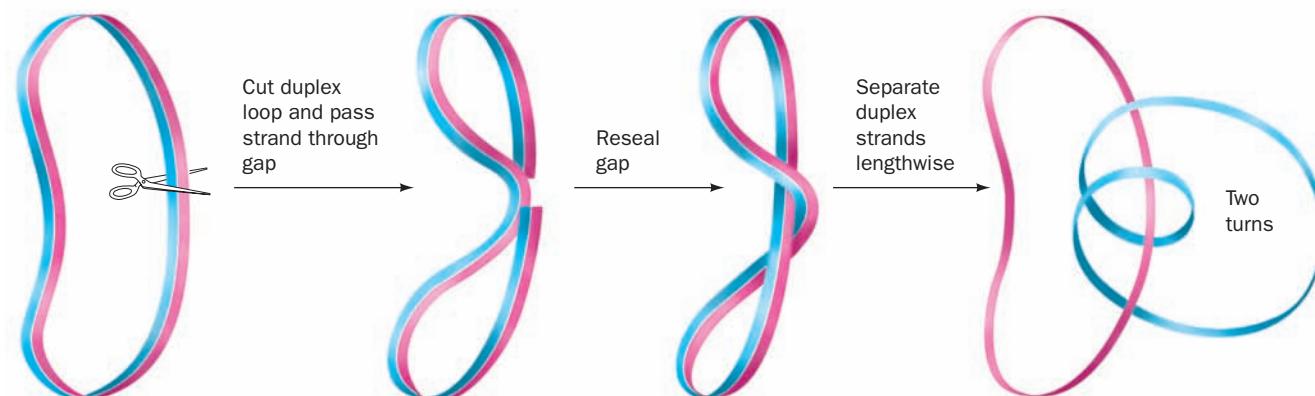


Figure 29-29 A demonstration, in which DNA is represented by a ribbon, that cutting a duplex circle, passing the double helix through the resulting gap, and then resealing the break changes

the linking number by 2. Separating the resulting single strands (slitting the ribbon along its length; *right*) indicates that one single strand makes two complete revolutions about the other.

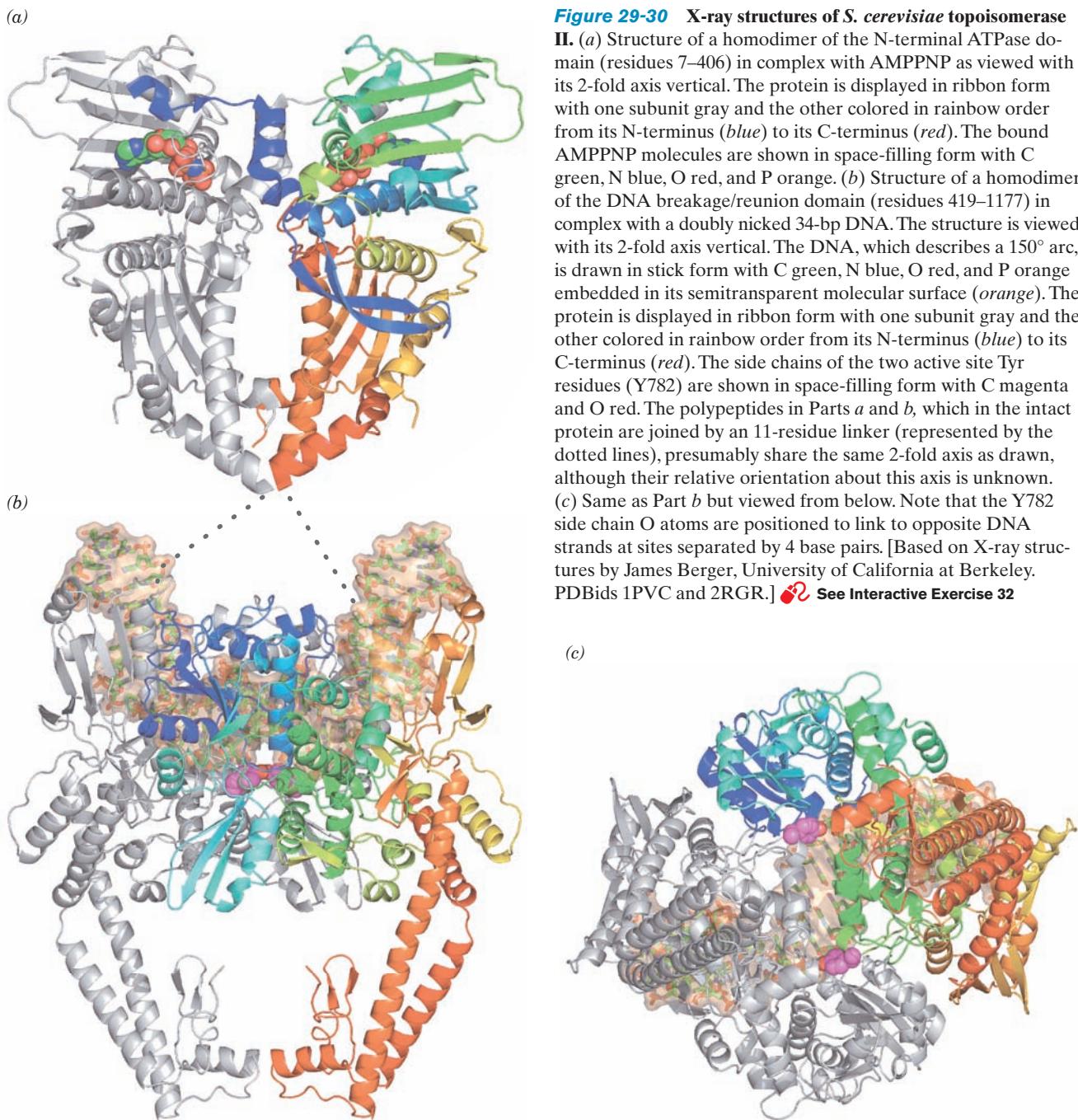


Figure 29-30 X-ray structures of *S. cerevisiae* topoisomerase II.

(a) Structure of a homodimer of the N-terminal ATPase domain (residues 7–406) in complex with AMPPNP as viewed with its 2-fold axis vertical. The protein is displayed in ribbon form with one subunit gray and the other colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The bound AMPPNP molecules are shown in space-filling form with C green, N blue, O red, and P orange. (b) Structure of a homodimer of the DNA breakage/reunion domain (residues 419–1177) in complex with a doubly nicked 34-bp DNA. The structure is viewed with its 2-fold axis vertical. The DNA, which describes a 150° arc, is drawn in stick form with C green, N blue, O red, and P orange embedded in its semitransparent molecular surface (orange). The protein is displayed in ribbon form with one subunit gray and the other colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the two active site Tyr residues (Y782) are shown in space-filling form with C magenta and O red. The polypeptides in Parts *a* and *b*, which in the intact protein are joined by an 11-residue linker (represented by the dotted lines), presumably share the same 2-fold axis as drawn, although their relative orientation about this axis is unknown. (c) Same as Part *b* but viewed from below. Note that the Y782 side chain O atoms are positioned to link to opposite DNA strands at sites separated by 4 base pairs. [Based on X-ray structures by James Berger, University of California at Berkeley. PDBids 1PVC and 2RGR.]  See Interactive Exercise 32

29-30*a*). The N-terminal domain binds AMPPNP and the C-terminal domains form the walls of a large hole through the dimer, which in the X-ray structure of the structurally similar *E. coli* GyrB is 20 Å across, the same width as the B-DNA double helix.

Berger also determined the X-ray structure of the yeast topoisomerase breakage/reunion domain in complex a 15-bp DNA that has a self-complementary 4-nucleotide overhang on the 5' end of one its strands. The DNA thereby forms a 2-fold symmetric 34-bp duplex with nicks on opposite strands separated by 4 bp (Fig. 29-30*b,c*). These are precisely the sites at which the enzyme would cleave an intact DNA by linking its newly formed 5'-ending strands to the

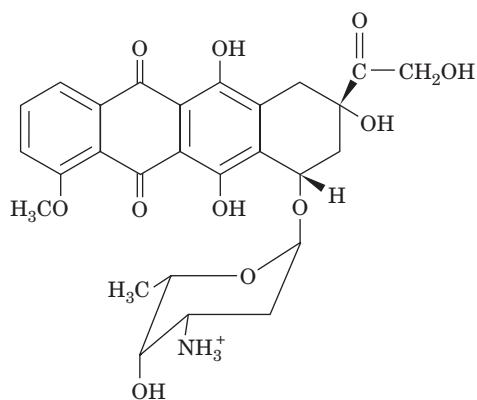
active site Tyr 782 residues. The protein binds the DNA in a positively charged groove that spans the width of the dimer and, in doing so, bends it through an arc of 150° (we shall see in later chapters that DNA-binding proteins often deform their bound DNA, although such an extreme deformation is unusual). Interestingly, the DNA between the two cleavage sites is essentially in the A form. There are almost no direct contacts between the protein and the DNA bases, as is expected for a protein with little sequence specificity. Note also that the C-terminal portions of the protein come together to enclose a large centrally located empty space.

Consideration of the foregoing two structures and those of the corresponding portions of *E. coli* gyrase suggests a

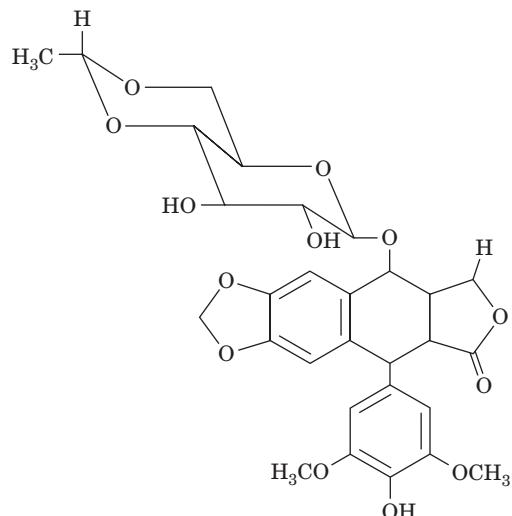
type of strand passage model for the mechanism of type IIA topoisomerases (Fig. 29-31) in which the DNA duplex to be cleaved, the so-called G-segment (G for gate), binds in the above-described groove across the top of the cleavage/reunion domain. ATP binding to the ATP-binding domain then induces a sequence of conformational changes in which the G-segment is cleaved and the resulting two fragments are spread apart by at least 20 Å through the action of the protein. This permits the passage of the DNA's so-called T-segment (T for transported) through the break in the DNA and breakage/reunion domain's upper gate (which may also contain portions of the ATPase domain) into its central hole, thereby incrementing the DNA's linking number by 2. Then, in a process that is accompanied by ATP hydrolysis, the upper gate closes to reseal the cleaved DNA, and the T-segment passes through the breakage/reunion domain's bottom gate. Finally, the resulting ADP and P_i are released and the bottom gate closes to yield recycled enzyme. Many of these enzymatic states have been observed in the several known X-ray structures of type IIA topoisomerase components.

e. Topoisomerase Inhibitors Are Effective Antibiotics and Cancer Chemotherapy Agents

Coumarin derivatives such as novobiocin, and quinolone derivatives such as ciprofloxacin, specifically inhibit gyrase and are therefore antibiotics. In fact, ciprofloxacin is the most efficacious oral antibiotic against gram-negative bacteria presently in clinical use (novobiocin's adverse side effects and the rapid generation of bacterial resistance to it have resulted in the discontinuation of its use in the treatment of human infections). A number of substances, including **doxorubicin** (also called **adriamycin**; a product of *Streptomyces peucetius*) and **etoposide** (a synthetic derivative),



Doxorubicin (Adriamycin)



Etoposide

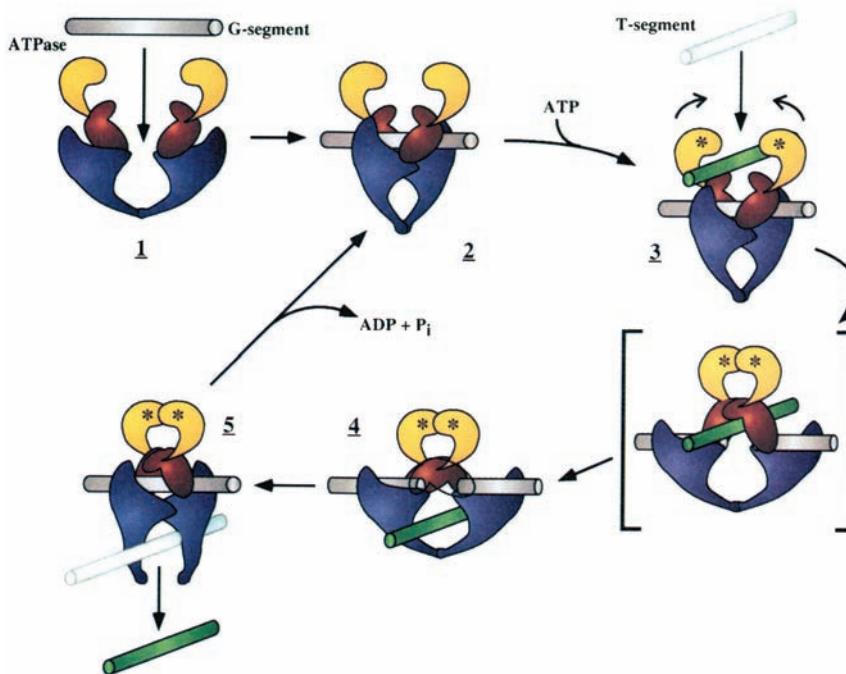


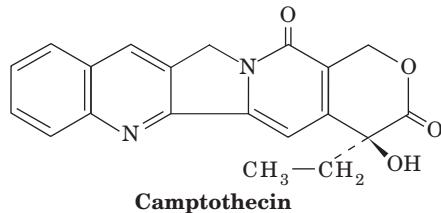
Figure 29-31 Model for the enzymatic mechanism of type II topoisomerases. The protein's ATPase domain and the upper and lower portions of the breakage/reunion domain are colored yellow, red, and purple, respectively, and the DNA's G- and T-segments are colored gray and green, respectively. In 1, the G-segment binds to the enzyme, thereby inducing the conformational change drawn in 2. The binding of ATP (represented by asterisks) and a T-segment (3) induces a series of conformational changes in which the G-segment is cleaved as the upper gate opens. The ATPase domains dimerize, and the T-segment is transported through the break into the central hole (4). The DNA transport step is shown as proceeding through the hypothetical intermediate in square brackets. The G-segments are then resealed and the T-segment is released through the lower gate (5). This gate then closes as the ATP is hydrolyzed and the resulting ADP and P_i released to yield the enzyme in its starting state (2). [Courtesy of James Wang, Harvard University.]

inhibit eukaryotic type IIA topoisomerases and are therefore widely used in cancer chemotherapy. Indeed, ~50% of cancer chemotherapy regimens contain at least one drug targeted to type IIA topoisomerases.

Type IIA topoisomerase inhibitors act in either of two ways. Many of them, including novobiocin, inhibit their target enzyme's ATPase activity (novobiocin is a competitive inhibitor of ATP because it tightly binds to GyrB in a way that prevents the binding of ATP's adenine ring). They therefore kill cells by blocking topoisomerase activity, which results in the arrest of DNA replication and RNA transcription. However, other substances, including ciprofloxacin, doxorubicin, and etoposide, enhance the rate at which their target type IIA topoisomerases cleave double-stranded DNA and/or reduce the rate at which these enzymes reseal these breaks. Consequently, these agents induce higher than normal levels of transient protein-bridged breaks in the DNA of treated cells. These protein bridges are easily ruptured by the passage of the replication and transcription machinery, thereby rendering the breaks permanent. Although all cells have elaborate enzymatic systems to repair damaged DNA (Section 30-5), a sufficiently high level of DNA damage overwhelms these

systems and hence results in cell death. Consequently, since rapidly replicating cells such as cancer cells have elevated levels of type IIA topoisomerases, they are far more likely to incur lethal DNA damage through the poisoning of their type IIA topoisomerases than are slow-growing or quiescent cells.

Type IB topoisomerases are specifically inhibited by the quinoline-based alkaloid **camptothecin**



(a product of the Chinese yew tree, *Camptotheca acuminata*) and its derivatives, which act by stabilizing the covalent topoisomerase I–DNA complex. These compounds, the only known naturally occurring topoisomerase IB inhibitors, are potent anticancer agents.

CHAPTER SUMMARY

1 Double Helical Structures B-DNA consists of a right-handed double helix of antiparallel sugar–phosphate chains with ~10 bp per turn of 34 Å and with its bases nearly perpendicular to the helix axis. Bases on opposite strands hydrogen-bond in a geometrically complementary manner to form A · T and G · C Watson–Crick base pairs. At low humidity, B-DNA undergoes a reversible transformation to a wider, flatter right-handed double helix known as A-DNA. Z-DNA, which is formed at high salt concentrations by polynucleotides of alternating purine and pyrimidine base sequences, is a left-handed double helix. Double helical RNA and RNA · DNA hybrids have A-DNA-like structures. The conformation of DNA, particularly that of B-DNA, varies with its base sequence largely because DNA's flexibility varies with its base sequence.

2 Forces Stabilizing Nucleic Acid Structures The orientations about the glycosidic bond and the various torsion angles in the sugar–phosphate chain are sterically constrained in nucleic acids. Likewise, only a few of the possible sugar pucker conformations are commonly observed. Watson–Crick base pairing is both geometrically and electronically complementary. Yet hydrogen bonding interactions do not greatly stabilize nucleic acid structures. Rather, the structures are largely stabilized by hydrophobic interactions. Nevertheless, the hydrophobic forces in nucleic acids are qualitatively different in character from those that stabilize proteins. Electrostatic interactions between charged phosphate groups are also important structural determinants of nucleic acids.

3 Supercoiled DNA The linking number (L) of a covalently closed circular DNA is topologically invariant. Consequently, any change in the twist (T) of a circular duplex must be balanced by an equal and opposite change in its writhing number (W), which indicates its degree of supercoiling. Supercoiling can be induced by intercalation agents. The gel electrophoretic mobility of DNA increases with its degree of superhelicity. Naturally occurring DNAs are all negatively supercoiled and must be so in order to participate in DNA replication and RNA transcription.

Type IA topoisomerases relax negatively supercoiled DNAs via a strand passage mechanism in which they cleave a single strand of DNA to form a 5'-phosphoTyr bond, pass a single-strand DNA segment through the gap, and then reseal the gap. Type IB topoisomerases relax both negatively and positively supercoiled DNAs via a controlled rotation mechanism involving a single-strand cleavage in which a transient phosphoTyr bond is formed with the newly generated 3' end. Type II topoisomerases relax duplex DNA in increments of two supertwists at the expense of ATP hydrolysis by making a double-strand scission in the DNA so as to form two transient 5'-phosphoTyr linkages, passing the duplex through the break, and resealing it. DNA gyrase also generates negative supertwists in an ATP-dependent manner. Topoisomerases are the targets of several antibiotics and chemotherapeutic agents.

REFERENCES

General

- Bloomfield, V.A., Crothers, D.M., and Tinoco, I., Jr., *Nucleic Acids: Structures, Properties, and Functions*, University Science Books (2000).
- Creighton, T.E., *The Biophysical Chemistry of Nucleic Acids & Proteins*, Chapters 2–6, Helvetian Press (2010).
- Neidle, S. (Ed.), *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press (1999).
- Saenger, W., *Principles of Nucleic Acid Structure*, Springer-Verlag (1984).
- The double helix—50 years, *Nature* **421**, 395–453 (2003). [A supplement containing a series of articles on the historical, cultural, and scientific influences of the DNA double helix celebrating the fiftieth anniversary of its discovery.]

Structures and Stabilities of Nucleic Acids

- Arnott, S., Historical article: DNA polymorphism and the early history of the double helix, *Trends Biochem. Sci.* **31**, 349–364 (2006).
- Bacolla, A. and Wells, R.D., Non-B DNA conformations, genomic rearrangements, and human disease, *J. Biol. Chem.* **46**, 47411–47414 (2004).
- Dickerson, R.E., Sequence-dependent B-DNA conformation in crystals and in protein complexes, in Sarma, R.H. and Sarma, M.H. (Eds.), *Structure, Motion, Interaction and Expression in Biological Molecules*, pp. 17–35, Adenine Press (1998); and DNA bending: the prevalence of kinkiness and the virtues of normality, *Nucleic Acids Res.* **26**, 1906–1926 (1998).
- Fairhead, H., Setlow, B., and Setlow, P., Prevention of DNA damage in spores and *in vitro* by small, acid-soluble proteins from *Bacillus* species, *J. Bacteriol.* **175**, 1367–1374 (1993).
- Ha, C.H., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K.K., Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases, *Nature* **437**, 1183–1186 (2005).
- Horton, N.C. and Finzel, B.C., The structure of an RNA/DNA hybrid: A substrate of the ribonuclease activity of HIV-1 reverse transcriptase, *J. Mol. Biol.* **264**, 521–533 (1996).
- Rich, A., The double helix: tale of two pockers, *Nature Struct. Biol.* **10**, 247–249 (2003). [A historical review.]
- Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A., and Rich, A., Crystal structure of the Z α domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA, *Science* **284**, 1841–1845 (1999).
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E., Crystal structure analysis of a complete turn of B-DNA, *Nature* **287**, 755–758 (1980); and Shui, X., McFail-Isom, L., Hu, G.G., and Williams, L.D., The B-DNA decamer at high resolution reveals a spine of sodium, *Biochemistry* **37**, 8341–8355 (1998). [The Dickerson dodecamer at its original 2.5 Å resolution and at its later-determined 1.4 Å resolution.]
- Supercoiled DNA**
- Bates, A.D. and Maxwell, A., *DNA Topology* (2nd ed.) Oxford University Press (2005).
- Champoux, J.J., DNA topoisomerases: Structure, function, and mechanism, *Annu. Rev. Biochem.* **70**, 369–413 (2001).
- Changela, A., DiGate, R., and Mondragón, A., Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA, *Nature* **411**, 1077–1081 (2001); and Mondragón, A. and DiGate, R., The structure of *Escherichia coli* DNA topoisomerase III, *Structure* **7**, 1373–1383 (1999).
- Classen, S., Olland, S., and Berger, J., Structure of the topoisomerase II ATPase region and its mechanism of inhibition by the chemotherapeutic agent ICRF-187, *Proc. Natl. Acad. Sci.* **100**, 10629–10634 (2003).
- Deweese, J.E., Osheroff, M.A., and Osheroff, N., DNA topology and topoisomerases, *Biochem. Mol. Biol. Educ.* **37**, 2–10 (2009).
- Dong, K.C. and Berger, J.M., Structural basis for the gate-DNA recognition and bending by type IIA topoisomerases, *Nature* **450**, 1201–1205 (2007). [X-ray structure of the yeast topoisomerase II breakage/reunion domain in complex with DNA.]
- Dong, K.C. and Berger, J.M., Structure and function of DNA topoisomerases, in Rice, P.A. and Correll, C.C. (Eds.), *Protein–Nucleic Acid Interactions*, RSC Publishing (2008); and Corbett, K.D. and Berger, J.M., Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 95–118 (2004).
- Gadelle, D., Filée, J., Buhler, C., and Forterre, P., Phylogenomics of type II DNA topoisomerases, *Bioessays* **25**, 232–242 (2003).
- Lebowitz, J., Through the looking glass: The discovery of supercoiled DNA, *Trends Biochem. Sci.* **15**, 202–207 (1990). [An informative eyewitness account of how DNA supercoiling was discovered.]
- Li, T.-K. and Liu, L.F., Tumor cell death induced by topoisomerase-targeting drugs, *Annu. Rev. Pharmacol. Toxicol.* **41**, 53–77 (2001).
- Maxwell, A., DNA gyrase as a drug target, *Biochem. Soc. Trans.* **27**, 48–53 (1999).
- Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J., Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA, *Science* **279**, 1504–1513 (1998); Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G.J., and Champoux, J.J., A model for the mechanism of human topoisomerase I, *Science* **279**, 1534–1541 (1998); and Redinbo, M.R., Champoux, J.J., and Hol, W.G.J., Structural insights into the function of type IB topoisomerases, *Curr. Opin. Struct. Biol.* **9**, 29–36 (1999).
- Taneja, B., Patel, A., Slesarev, A., and Mondraón, A., Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases, *EMBO J.* **25**, 398–408 (2006).
- Wang, J.C., Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine, *Q. Rev. Biophys.* **31**, 107–144 (1998).
- Wang, J.C., Cellular roles of DNA topoisomerases: a molecular perspective, *Nature Rev. Mol. Cell Biol.* **3**, 430–440 (2002).

PROBLEMS

- 1.** A · T base pairs in DNA exhibit greater variability in their propeller twisting than do G · C base pairs. Suggest the structural basis of this phenomenon.

- *2.** At Na^+ concentrations $>5M$, the T_m of DNA decreases with increasing $[\text{Na}^+]$. Explain this behavior. (Hint: Consider the solvation requirements of Na^+ .)

***3.** Why are the most commonly observed conformations of the ribose ring those in which either atom C2' or atom C3' is out of the plane of the other four ring atoms? (*Hint:* In puckering a planar ring such that one atom is out of the plane of the other four, the substituents about the bond opposite the out-of-plane atom remain eclipsed. This is best observed with a ball-and-stick model.)

4. Polyomavirus DNA can be separated by sedimentation at neutral pH into three components that have sedimentation coefficients of 20, 16, and 14.5S and that are known as Types I, II, and III DNAs, respectively. These DNAs all have identical base sequences and molecular masses. In 0.15M NaCl, both Types II and III DNA have melting curves of normal cooperativity and a T_m of 88°C. Type I DNA, however, exhibits a very broad melting curve and a T_m of 107°C. At pH 13, Types I and III DNAs have sedimentation coefficients of 53 and 16S, respectively, and Type II separates into two components with sedimentation coefficients of 16 and 18S. How do Types I, II, and III DNAs differ from one another? Explain their different physical properties.

5. When the helix axis of a closed circular duplex DNA of 2310 bp is constrained to lie in a plane, the DNA has a twist (T) of 207. When released, the DNA takes up its normal twist of 10.5 bp per turn. Indicate the values of the linking number (L), writhing number (W), and twist for both the constrained and unconstrained conformational states of this DNA circle. What is the superhelix density, σ , of both the constrained and unconstrained DNA circles?

6. A covalently closed circular duplex DNA has a 100-bp segment of alternating C and G residues. On transfer to a solution containing a high salt concentration, this segment undergoes a transition from the B conformation to the Z conformation. What is the accompanying change in its linking number, writhing number, and twist?

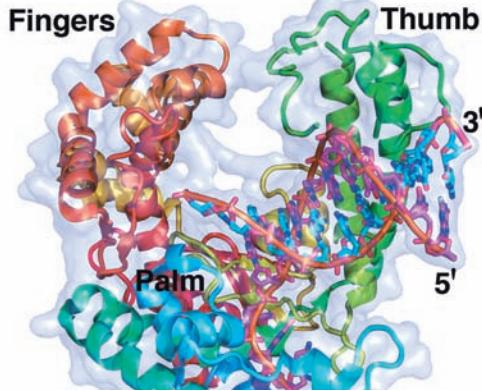
7. You have discovered an enzyme secreted by a particularly virulent bacterium that cleaves the C2'—C3' bond in the deoxyribose residues of duplex DNA. What is the effect of this enzyme on supercoiled DNA?

8. A bacterial chromosome consists of a protein–DNA complex in which its single DNA molecule appears to be supercoiled, as demonstrated by ethidium bromide titration. However, in contrast to the case with naked circular duplex DNA, the light single-strand nicking of chromosomal DNA does not abolish this supercoiling. What does this indicate about the structure of the bacterial chromosome, that is, how do its proteins constrain its DNA?

9. Although types IA and IIA topoisomerases exhibit no significant sequence similarity, it has been suggested that they are distantly related based on the similarities of certain aspects of their enzymatic mechanisms. What are these similarities?

10. Draw the mechanism of DNA strand cleavage and rejoining mediated by topoisomerase IA.

DNA Replication, Repair, and Recombination



CHAPTER 30

1 DNA Replication: An Overview

- A. Replication Forks
- B. Role of DNA Gyrase
- C. Semidiscontinuous Replication
- D. RNA Primers

2 Enzymes of Replication

- A. DNA Polymerase I
- B. DNA Polymerase III
- C. Unwinding DNA: Helicases and Single-Strand Binding Protein
- D. DNA Ligase
- E. Primase

3 Prokaryotic Replication

- A. Bacteriophage M13
- B. Bacteriophage ϕ X174
- C. *Escherichia coli*
- D. Fidelity of Replication

4 Eukaryotic Replication

- A. The Cell Cycle
- B. Eukaryotic Replication Mechanisms
- C. Reverse Transcriptase
- D. Telomeres and Telomerase

5 Repair of DNA

- A. Direct Reversal of Damage
- B. Excision Repair
- C. Mismatch Repair
- D. The SOS Response
- E. Double-Strand Break Repair
- F. Identification of Carcinogens

6 Recombination and Mobile Genetic Elements

- A. Homologous Recombination
- B. Transposition and Site-Specific Recombination

7 DNA Methylation and Trinucleotide Repeat Expansions

Here we begin a three-chapter series on the basic processes of gene expression: DNA replication (this chapter), transcription (Chapter 31), and translation (Chapter 32). These processes have been outlined in Section 5-4. We shall now discuss them in greater depth with an emphasis on how we have come to know what we know.

1 DNA REPLICATION: AN OVERVIEW

Watson and Crick's seminal paper describing the DNA double helix ended with the statement: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." In a succeeding paper they expanded on this rather cryptic remark by pointing out that a DNA strand could act as a template to direct the synthesis of its complementary strand. Although Meselson and Stahl demonstrated, in 1958, that DNA is, in fact, semiconservatively replicated (Section 5-3B), it was not until some 20 years later that the mechanism of DNA replication in prokaryotes was understood in reasonable detail. This is because, as we shall see in this chapter, the DNA replication process rivals translation in its complexity but is mediated by often loosely associated protein assemblies that are present in only a few copies per cell. *The surprising intricacy of DNA replication compared to the chemically similar transcription process (Section 31-2) arises from the need for extreme accuracy in DNA replication so as to preserve the integrity of the genome from generation to generation.*

A. Replication Forks

DNA is replicated by enzymes known as **DNA-directed DNA polymerases** or simply **DNA polymerases**. These enzymes utilize single-stranded DNA as templates on which to catalyze the synthesis of the complementary strand from the appropriate deoxynucleoside triphosphates (Fig. 30-1). The incoming nucleotides are selected by their ability to form Watson-Crick base pairs with the template DNA so that the newly synthesized DNA strand forms a double helix with the template strand. *Nearly all known DNA polymerases can only add a nucleotide donated by a nucleoside triphosphate to the free 3'-OH group of a base paired polynucleotide so that DNA chains are extended only in the 5' \rightarrow 3' direction.* DNA polymerases are discussed further in Sections 30-2A, 30-2B, and 30-4B.

a. Duplex DNA Replicates Semiconservatively at Replication Forks

John Cairns obtained the earliest indications of how chromosomes replicate through the autoradiography of replicating DNA. Autoradiograms of circular chromosomes grown in a medium containing [3 H]thymidine show

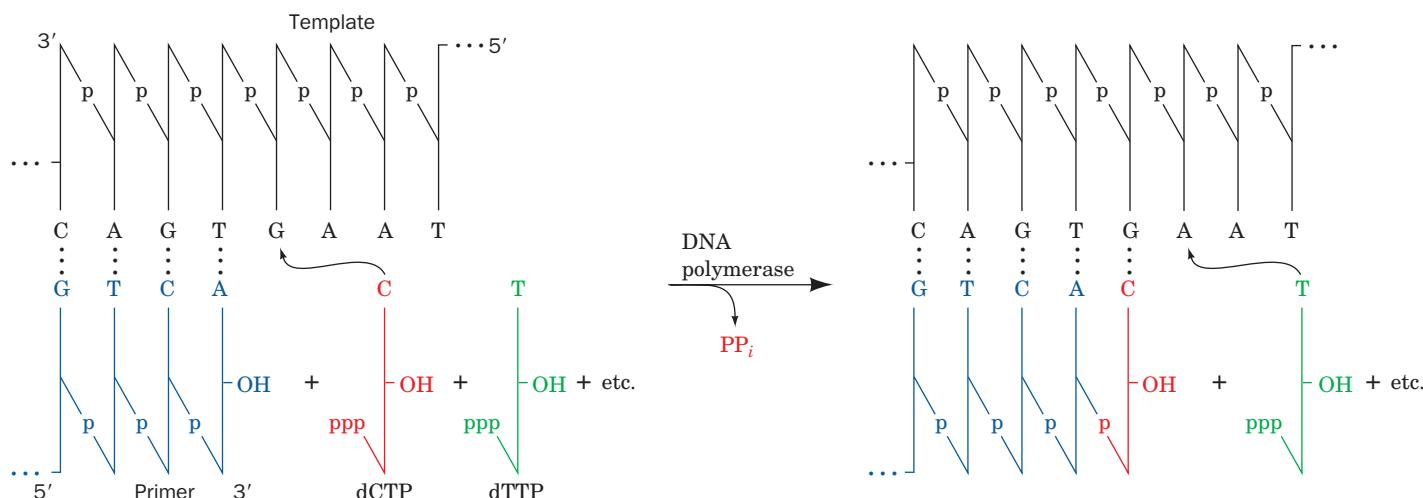


Figure 30-1 Action of DNA polymerase. DNA polymerases assemble incoming deoxynucleoside triphosphates on

single-stranded DNA templates such that the growing strand is elongated in its $5' \rightarrow 3'$ direction.

the presence of replication “eyes” or “bubbles” (Fig. 30-2). These so-called **θ structures** (after their resemblance to the Greek letter theta) indicate that *double-stranded DNA (dsDNA)* replicates by the progressive separation of its two *parental strands* accompanied by the synthesis of their *complementary strands* to yield two *semiconservatively replicated duplex daughter strands* (Fig. 30-3). DNA replication involving θ structures is known as **θ replication**.

A branch point in a replication eye at which DNA synthesis occurs is called a **replication fork**. A replication bubble may contain one or two replication forks (**unidirectional** or **bidirectional replication**). Autoradiographic studies have demonstrated that θ replication is almost always bidirectional (Fig. 30-4). Moreover, such experiments, together with genetic evidence, have established that prokaryotic and bacteriophage DNAs have but one **replication origin** (point where DNA synthesis is initiated).

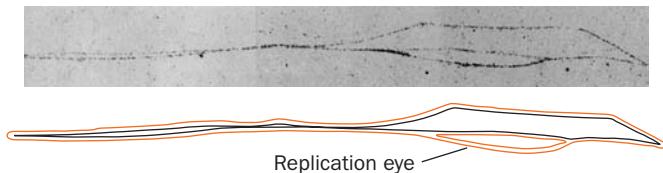


Figure 30-2 Autoradiogram and its interpretive drawing of a replicating *E. coli* chromosome. The bacterium had been grown for somewhat more than one generation in a medium containing [^3H]thymidine, thereby labeling the subsequently synthesized DNA so that it appears as a line of dark grains in the photographic emulsion (red lines in the interpretive drawing). The size of the replication eye indicates that the circular chromosome is about one-sixth duplicated in the present round of replication. [Courtesy of John Cairns, Cold Spring Harbor Laboratory, New York.]

B. Role of DNA Gyrase

The requirement that the parent DNA unwind at the replication fork (Fig. 30-3) presents a formidable topological obstacle. For instance, *E. coli* DNA is replicated at a rate of ~ 1000 nucleotides/s. If its $1300\text{-}\mu\text{m}$ -long chromosome were linear, it would have to flail around within the confines of a $3\text{-}\mu\text{m}$ -long *E. coli* cell at ~ 100 revolutions/s

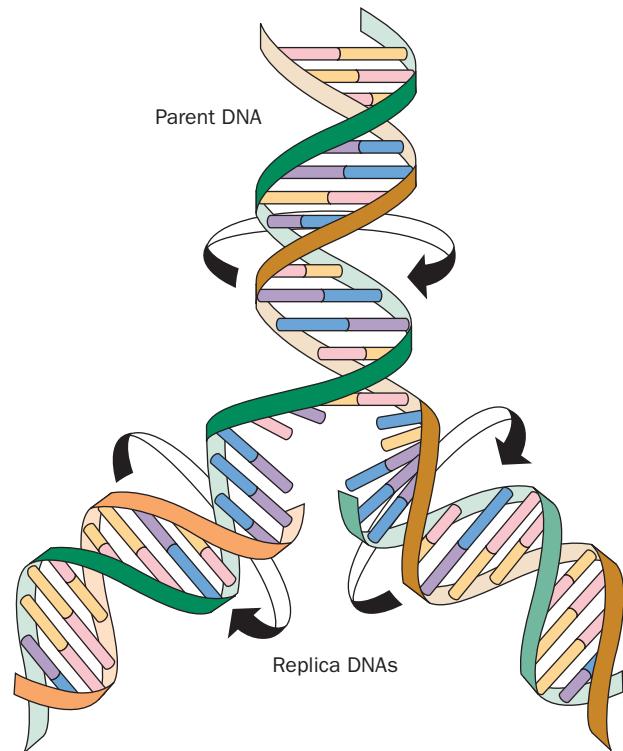


Figure 30-3 Replication of DNA.

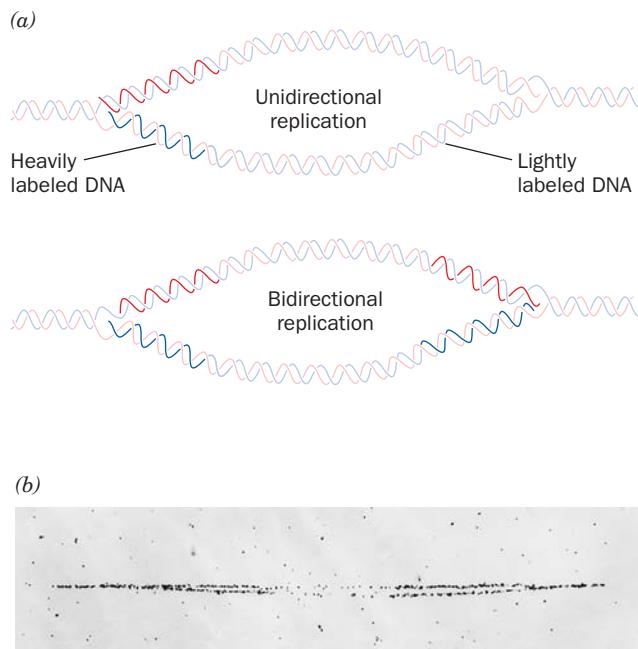


Figure 30-4 Autoradiographic differentiation of unidirectional and bidirectional θ replication of DNA. (a) An organism is grown for several generations in a medium that is lightly labeled with [3 H]thymidine so that all of its DNA will be visible in an autoradiogram. A large amount of [3 H]thymidine is then added to the medium for a few seconds before the DNA is isolated (**pulse labeling**) in order to label only those bases near the replication fork(s). Unidirectional DNA replication will exhibit only one heavily labeled branch point (*above*), whereas bidirectional DNA replication will exhibit two such branch points (*below*). (b) An autoradiogram of *E. coli* DNA so treated, demonstrating that it is bidirectionally replicated. [Courtesy of David M. Prescott, University of Colorado.]

(recall that B-DNA has \sim 10 bp per turn). But since the *E. coli* chromosome is, in fact, circular, even this could not occur. Rather, the DNA molecule would accumulate $+100$ supercoils/s (see Section 29-3A for a discussion of supercoiling) until it became too tightly coiled to permit further unwinding. Naturally occurring DNA's negative supercoiling promotes DNA unwinding but only to the extent of \sim 5% of its duplex turns (recall that naturally occurring DNAs are typically underwound by one supercoil per \sim 19 duplex turns; Section 29-3Bb). In prokaryotes, however, negative supercoils may be introduced into DNA through the action of a type IIA topoisomerase (DNA gyrase; Section 29-3Cd) at the expense of ATP hydrolysis. This process is essential for prokaryotic DNA replication as is demonstrated by the observation that DNA gyrase inhibitors, such as novobiocin, arrest DNA replication except in mutants whose DNA gyrase does not bind these antibiotics.

C. Semidiscontinuous Replication

The low-resolution images provided by autoradiograms such as Figs. 30-2 and 30-4b suggest that dsDNA's two antiparallel strands are simultaneously replicated at an advancing replication fork. Yet, all known DNA polymerases can only extend DNA strands in the $5' \rightarrow 3'$ direction. How, then, does DNA polymerase copy the parent strand that extends in the $5' \rightarrow 3'$ direction past the replication fork? This question was answered in 1968 by Reiji Okazaki through the following experiments. If a growing *E. coli* culture is pulse-labeled for 30 s with [3 H]thymidine, much of the radioactive and hence newly synthesized DNA has a sedimentation coefficient in alkali of 7S to 11S. These so-called **Okazaki fragments** evidently consist of only 1000 to 2000 nucleotides (nt; 100–200 nt in eukaryotes). If, however, following the 30 s [3 H]thymidine pulse, the *E. coli* are transferred to an unlabeled medium (a **pulse-chase** experiment), the resulting radioactively labeled DNA sediments at a rate that increases with the time that the cells had grown in the unlabeled medium. The Okazaki fragments must therefore become covalently incorporated into larger DNA molecules.

Okazaki interpreted his experimental results in terms of the **semidiscontinuous replication** model (Fig. 30-5). The two parent strands are replicated in different ways. *The newly synthesized DNA strand that extends $5' \rightarrow 3'$ in the direction of replication fork movement, the so-called **leading strand**, is essentially continuously synthesized in its $5' \rightarrow 3'$ direction as the replication fork advances. The other newly synthesized strand, the **lagging strand**, is also synthesized in its $5' \rightarrow 3'$ direction but discontinuously as Okazaki fragments. The Okazaki fragments are only covalently joined together sometime after their synthesis in a reaction catalyzed by the enzyme **DNA ligase** (Section 30-2D).*

The semidiscontinuous model of DNA replication is corroborated by electron micrographs of replicating DNA showing single-stranded regions on one side of the

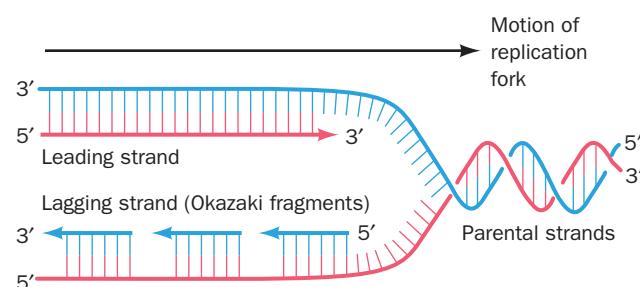


Figure 30-5 Semidiscontinuous DNA replication. In DNA replication, both daughter strands (*leading strand* red, *lagging strand* blue) are synthesized in their $5' \rightarrow 3'$ directions. The leading strand is synthesized continuously, whereas the lagging strand is synthesized discontinuously.

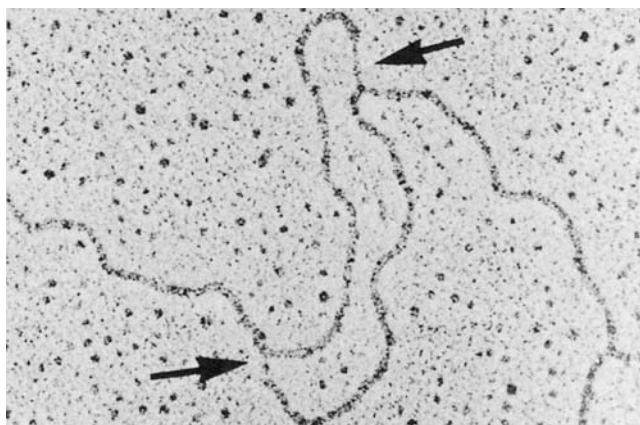


Figure 30-6 Electron micrograph of a replication eye in *Drosophila melanogaster* DNA. Note that the single-stranded regions (arrows) near the replication forks have the trans configuration consistent with the semidiscontinuous model of DNA replication. [From Kriegstein, H.J. and Hogness, D.S., *Proc. Natl. Acad. Sci.* **71**, 173 (1974).]

replication fork (Fig. 30-6). In bidirectionally replicating DNA, moreover, the two single-stranded regions occur, as expected, on diagonally opposite sides of the replication bubble.

D. RNA Primers

DNA polymerases' all but universal requirement for a free 3'-OH group to extend a DNA chain poses a question that was emphasized by the establishment of the semidiscontinuous model of DNA replication: How is DNA synthesis initiated? Careful analysis of Okazaki fragments revealed

that their 5' ends consist of RNA segments of 1 to 60 nt (a length that is species dependent) that are complementary to the template DNA chain (Fig. 30-7). *E. coli* has two enzymes that can catalyze the formation of these **RNA primers**: **RNA polymerase**, the ~459-kD multisubunit enzyme that mediates transcription (Section 31-2), and the much smaller primase (60 kD), the monomeric product of the *dnaG* gene.

Primase is insensitive to the RNA polymerase inhibitor **rifampicin** (Section 31-2Bb). The observation that rifampicin inhibits only leading strand synthesis therefore indicates that *primase initiates the Okazaki fragment primers*. The initiation of leading strand synthesis in *E. coli*, a much rarer event than that of Okazaki fragments, can be mediated *in vitro* by either RNA polymerase or primase alone but is greatly stimulated when both enzymes are present. It is therefore thought that these enzymes act synergistically *in vivo* to prime leading strand synthesis.

Mature DNA does not contain RNA. The RNA primers are eventually removed and the resulting single-strand gaps are filled in with DNA by a mechanism described in Section 30-2A(j).

2 ENZYMES OF REPLICATION

DNA replication is a complex process involving a great variety of enzymes. It requires, to list only its major actors in their order of appearance: (1) DNA topoisomerases, (2) enzymes known as helicases that separate the DNA strands at the replication fork, (3) proteins that prevent them from reannealing before they are replicated, (4) enzymes that synthesize RNA primers, (5) a DNA polymerase, (6) an enzyme to remove the RNA primers, and (7) an enzyme to covalently link successive Okazaki fragments. In this section, we describe the properties and functions of many of these proteins.

A. DNA Polymerase I

In 1957, Arthur Kornberg reported that he had discovered an enzyme that catalyzes the synthesis of DNA in extracts of *E. coli* through its ability to incorporate the radioactive label from [¹⁴C]thymidine triphosphate into DNA. This enzyme, which has since become known as **DNA polymerase I** or **Pol I**, consists of a monomeric 928-residue polypeptide.

Pol I couples deoxynucleoside triphosphates on DNA templates (Fig. 30-1) in a reaction that occurs through the nucleophilic attack of the growing DNA chain's 3'-OH group on the α -phosphoryl of an incoming nucleoside triphosphate. The reaction is driven by the resulting elimination of PP_i and its subsequent hydrolysis by inorganic pyrophosphatase. The overall reaction resembles that catalyzed by RNA polymerase (Fig. 5-23) but differs from it by the strict requirement that the incoming nucleoside be linked to a free 3'-OH group of a polynucleoside that is base paired to the template (RNA polymerase initiates

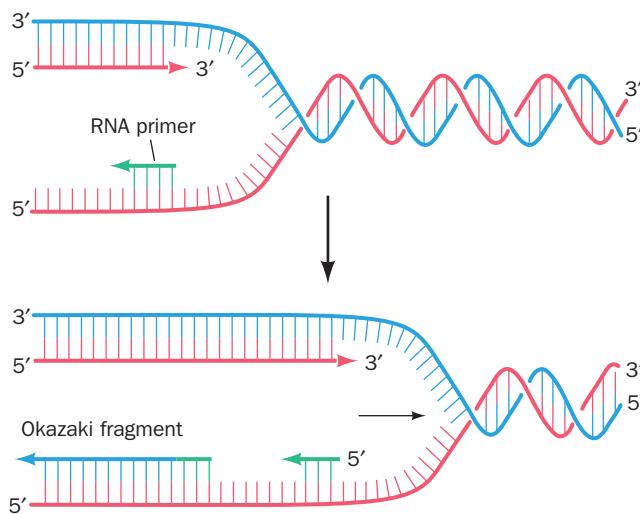


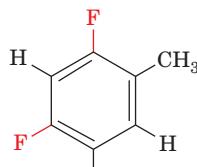
Figure 30-7 Priming of DNA synthesis by short RNA segments.

transcription by linking together two ribonucleoside triphosphates on a DNA template; Section 31-2C). The complementarity between the product DNA and the template was at first inferred through base composition and hybridization studies but was eventually directly established by base sequence determinations. The error rate of Pol I in copying the template is extremely low, as was first demonstrated by its *in vitro* replication of the 5386-nt DNA from bacteriophage ϕ X174 to yield fully infective phage DNA. In fact, its measured error rate is around one wrong base per 10 million.

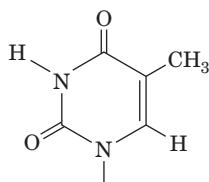
Pol I is said to be **processive** in that it catalyzes a series of successive polymerization steps, typically 20 or more, without releasing the template (the opposite of “processive” is “**distributive**”). Pol I can, of course, work in reverse by degrading DNA through pyrophosphorolysis. This reverse reaction, however, probably has no physiological significance because of the low *in vivo* concentration of PP_i resulting from the action of inorganic pyrophosphatase.

a. Pol I Recognizes the Incoming dNTP According to the Shape of the Base Pair It Forms with the Template DNA

The specificity of Pol I for an incoming base arises from the requirement that it form a Watson–Crick base pair with the template rather than direct recognition of the incoming base (recall that the four base pairs, A · T, T · A, G · C, and C · G, have nearly identical shapes; Fig. 5-12). Thus, as Eric Kool demonstrated, when the “base” **2,4-difluorotoluene (F)**,



2,4-Difluorotoluene base (F)



Thymine (T)

which is isosteric with (has the same shape as) thymine but does not accept hydrogen bonds, is synthetically inserted into a template DNA, Pol I incorporates A opposite the F with a similar rate of mismatches as it incorporates A opposite T. Likewise, dFTP is incorporated opposite template A with a similar fidelity as is dTTP. Yet the incorporation of F opposite an A in DNA destabilizes the double helix by 15 kJ/mol relative to T opposite this A. Evidently, *Pol I selects an incoming dNTP largely according to its ability to form a Watson–Crick-shaped pair with the template base but with little regard for its hydrogen bonding properties*. Indeed, the NMR structure of a 12-bp DNA containing a centrally located F opposite an A reveals that it assumes a B-DNA conformation in which the F–A pair closely resembles a T · A base pair in the same position of an otherwise identical DNA.

b. Pol I Can Edit Its Mistakes

In addition to its polymerase activity, Pol I has two independent hydrolytic activities:

1. It can act as a 3' → 5' exonuclease.
2. It can act as a 5' → 3' exonuclease.

The 3' → 5' exonuclease reaction differs chemically from the pyrophosphorolysis reaction (the reverse of the polymerase reaction) only in that H₂O rather than PP_i is the nucleotide acceptor. Kinetic and crystallographic studies, however, indicate that these two catalytic activities occupy separate active sites (see below). The 3' → 5' exonuclease function is activated by an unpaired 3'-terminal nucleotide with a free OH group. If Pol I erroneously incorporates a wrong (unpaired) nucleotide at the end of a growing DNA chain, the polymerase activity is inhibited and the 3' → 5' exonuclease excises the offending nucleotide (Fig. 5-36). The polymerase activity then resumes DNA replication. *Pol I therefore has the ability to proofread or edit a DNA chain as it is synthesized so as to correct its mistakes*. This explains the great fidelity of DNA replication by Pol I: The overall fraction of bases that the enzyme misincorporates, $\sim 10^{-7}$, is the product of the fraction of bases that its polymerase activity misincorporates and the fraction of misincorporated bases that its 3' → 5' exonuclease activity fails to excise. The price of this high fidelity is that $\sim 3\%$ of correctly incorporated nucleotides are also excised.

The Pol I 5' → 3' exonuclease binds to dsDNA at single-strand nicks with little regard to the character of the 5' nucleotide (5'-OH or phosphate group; base paired or not). It cleaves the DNA in a base paired region beyond the nick such that the DNA is excised as either mononucleotides or oligonucleotides of up to 10 residues (Fig. 5-33). In contrast, the 3' → 5' exonuclease removes only unpaired mononucleotides with 3'-OH groups.

c. Pol I's Polymerase and Two Exonuclease Functions Each Occupy Separate Active Sites

The 5' → 3' exonuclease activity of Pol I is independent of both its 3' → 5' exonuclease and its polymerase activities. In fact, as we saw in Section 7-2A, proteases such as subtilisin or trypsin cleave Pol I into two fragments: a larger C-terminal or **Klenow fragment** (residues 324–928), which contains both the polymerase and the 3' → 5' exonuclease activities; and a smaller, N-terminal fragment (residues 1–323), which contains the 5' → 3' exonuclease activity. Thus Pol I contains three active sites on a single polypeptide chain.

d. The X-Ray Structure of Klenow Fragment Indicates How It Binds DNA

The X-ray structure of Klenow fragment in complex with dsDNA, determined by Thomas Steitz, reveals that

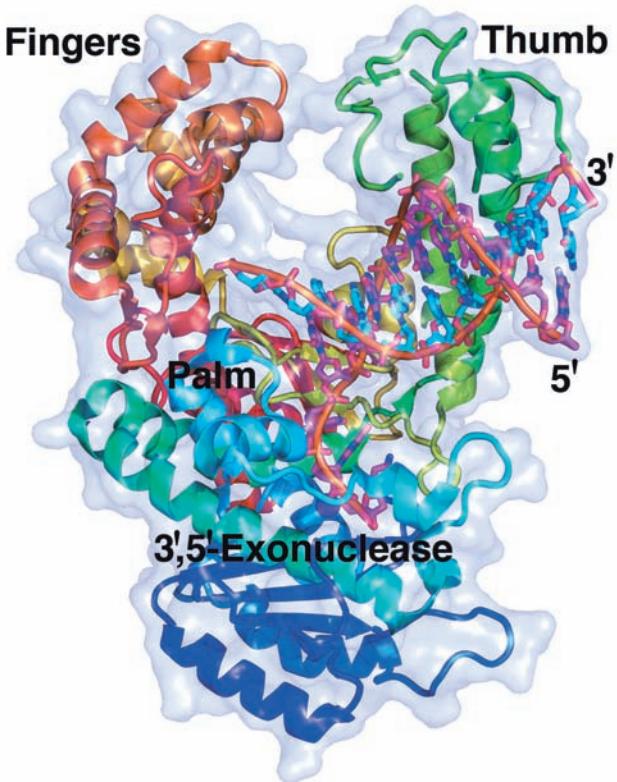


Figure 30-8 X-ray structure of *E. coli* DNA polymerase I Klenow fragment in complex with a dsDNA. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and embedded in its semitransparent molecular surface. The DNA is shown in stick form with the C atoms of its 10-nt template strand cyan, the C atoms of its 13-nt primer strand magenta, N blue, O red, and P orange, and with successive P atoms in each polynucleotide strand connected by orange rods. A Zn^{2+} ion (purple sphere) marks the $3' \rightarrow 5'$ -exonuclease active site. [Based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1KLN.]

See Interactive Exercise 33

the enzyme consists of two domains (Fig. 30-8). The smaller domain (residues 324–517) contains the $3' \rightarrow 5'$ exonuclease site, as was demonstrated by the absence of this function but not polymerase activity in a genetically engineered Klenow fragment mutant that lacks the divalent metal ion-binding sites known to be essential for $3' \rightarrow 5'$ exonuclease activity but which otherwise has a normal structure. The larger domain (residues 521–928) contains the polymerase active site at the bottom of a prominent cleft, a surprisingly large distance (~ 25 Å) from the $3' \rightarrow 5'$ exonuclease site. The cleft, which is lined with positively charged residues, has the appropriate size (~ 22 Å wide by ~ 30 Å deep) and shape to bind a B-DNA molecule in a manner resembling a right hand grasping a rod. Consequently, Klenow fragment's polymerase domains are known as its “fingers,” “thumb,” and “palm” (Fig. 30-8). Indeed, all DNA polymerases of known structure have similar shapes with a fingers domain that binds the incoming nucleotide and the

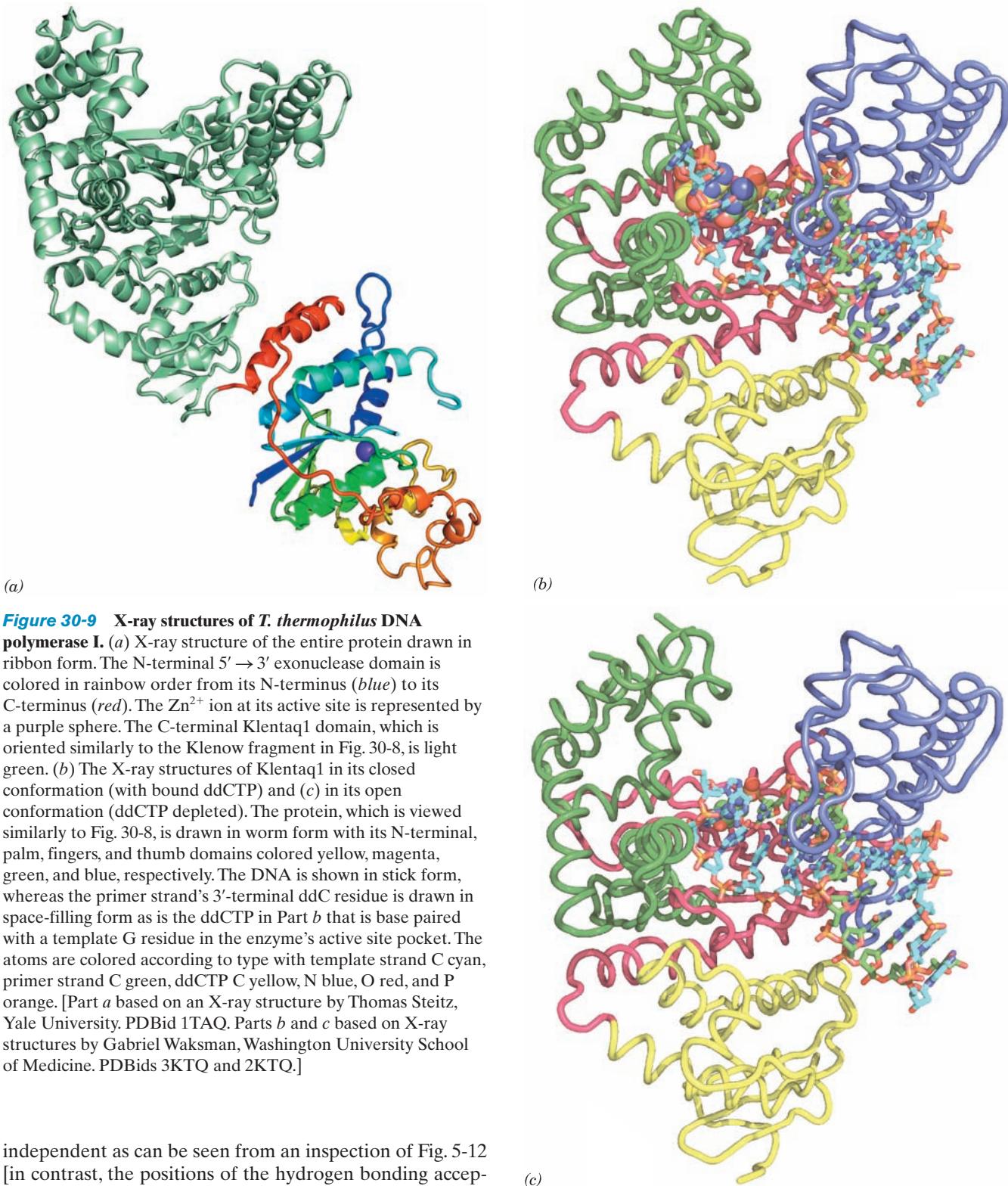
templating base, a thumb domain that guides the newly formed dsDNA as it leaves the active site, and a palm domain that contains the active site at the bottom of the cleft between the fingers and thumb domains (Sections 30-2Ba, 30-4Ba, and 30-4Ca).

e. DNA Polymerase Distinguishes Watson–Crick Base Pairs via Sequence-Independent Interactions That Induce Domain Movements

DNA polymerase I from the thermophile *Thermus aquaticus* (Taq), which is 51% identical to its *E. coli* homolog, lacks a functional $3' \rightarrow 5'$ exonuclease. Its X-ray structure (Fig. 30-9a), also determined by Steitz, reveals that its C-terminal domain closely resembles Klenow fragment (Fig. 30-8), although the metal ion-binding carboxylate residues that are essential for the $3' \rightarrow 5'$ exonuclease function of Klenow fragment are absent in Taq polymerase. Its N-terminal $5' \rightarrow 3'$ exonuclease domain, which appears to be only loosely tethered to the C-terminal polymerase domain, contains a conserved cluster of metal ions at the bottom of a cleft that is ~ 70 Å from the polymerase active site. It is therefore unclear how the polymerase and $5' \rightarrow 3'$ exonuclease functions work in concert, as we shall see (Section 30-2Ai), to produce dsDNA with a single nick.

Gabriel Waksman crystallized the C-terminal domain of Taq DNA polymerase I [Klenetaq1; which is often used in PCR experiments (Section 5-5F)] in complex with an 11-bp dsDNA that has a GGAAA-5' overhang at the 5' end of its template strand, and the crystals were incubated with 2',3'-dideoxy-CTP (ddCTP; which lacks a 3'-OH group). The X-ray structure of these crystals (Fig. 30-9b) reveals that a ddC residue had been covalently linked to the 3' end of the primer and formed a Watson–Crick pair with the template overhang's 3' G. Moreover, a second ddCTP molecule (to which the primer's new 3'-terminal ddC residue is incapable of forming a covalent bond) occupies the enzyme's active site where it forms a Watson–Crick pair with the template's next G that is largely out of contact with the surrounding aqueous solution. Clearly, Klenetaq1 retains its catalytic activity in this crystal.

A DNA polymerase must distinguish correctly paired bases from mismatches and yet do so via sequence-independent interactions with the incoming dNTP. The foregoing X-ray structure reveals that this occurs through an active site pocket that is complementary in shape to Watson–Crick base pairs. The pocket is formed by the stacking of a conserved Tyr side chain on the template base, as well as by van der Waals interactions with the protein and with the preceding base pair. In addition, although the dsDNA is mainly in the B conformation, the 3 base pairs nearest the active site assume the A conformation, as has also been observed in the X-ray structures of several other DNA polymerases in their complexes with DNA. The resulting wider and shallower minor groove (Section 29-1Ba) permits protein side chains to form hydrogen bonds with the otherwise inaccessible N3 atoms of the purine bases and O2 atoms of the pyrimidine bases. The positions of these hydrogen bond acceptors are sequence-



(a)

(b)

(c)

Figure 30-9 X-ray structures of *T. thermophilus* DNA polymerase I. (a) X-ray structure of the entire protein drawn in ribbon form. The N-terminal 5' → 3' exonuclease domain is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The Zn²⁺ ion at its active site is represented by a purple sphere. The C-terminal Klenetaq1 domain, which is oriented similarly to the Klenow fragment in Fig. 30-8, is light green. (b) The X-ray structures of Klenetaq1 in its closed conformation (with bound ddCTP) and (c) in its open conformation (ddCTP depleted). The protein, which is viewed similarly to Fig. 30-8, is drawn in worm form with its N-terminal, palm, fingers, and thumb domains colored yellow, magenta, green, and blue, respectively. The DNA is shown in stick form, whereas the primer strand's 3'-terminal ddC residue is drawn in space-filling form as is the ddCTP in Part b that is base paired with a template G residue in the enzyme's active site pocket. The atoms are colored according to type with template strand C cyan, primer strand C green, ddCTP C yellow, N blue, O red, and P orange. [Part a based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1TAQ. Parts b and c based on X-ray structures by Gabriel Waksman, Washington University School of Medicine. PDBids 3KTQ and 2KTQ.]

independent as can be seen from an inspection of Fig. 5-12 [in contrast, the positions of the hydrogen bonding acceptors in the major groove vary with both the identity (A · T vs G · C) and the orientation (e.g., A · T vs T · A) of the base pair]. However, with a non-Watson–Crick pairing, these hydrogen bonds would be greatly distorted if not completely disrupted. The protein also makes extensive sequence-independent hydrogen bonding and van der Waals interactions with the DNA's sugar–phosphate backbone.

The above Klenetaq1 · DNA · ddCTP crystals were partially depleted of ddCTP by soaking them in a stabilizing solution that lacks ddCTP. The X-ray structure of the ddCTP-depleted crystals (Fig. 30-9c) reveals that Klenetaq1

assumes a so-called open conformation, which differs significantly from that in the so-called closed conformation described above by a 46° hingelike motion of the fingers domain away from the polymerase active site (Fig. 30-9c). Evidently, the formation of a Watson–Crick base pair at the polymerase active site triggers the formation of a productive ternary complex that buries the incoming nucleotide. In particular, the foregoing Tyr side chain, which extends from the rightmost helix of the fingers domain (Fig. 30-9b), stacks on the incoming base, an interaction that is absent in the open conformation. Moreover, a Lys and an Arg side chain, that also extend from the rightmost helix of the fingers domain, form salt bridges with the α - and β -phosphate groups of the incoming dNTP. These observations are consistent with kinetic measurements on Pol I indicating that the binding of the correct dNTP to the enzyme induces a rate-limiting conformational change that yields a tight ternary complex. It therefore appears that the enzyme rapidly samples the available dNTPs in its open conformation but only when it binds the correct dNTP in a Watson–Crick pairing with the template base does it form the catalytically competent closed conformation. The subsequent reaction steps then rapidly yield the product complex which, following a second conformational change, releases the product PP_i . Finally, the DNA is translocated in the active site, probably via a linear diffusion mechanism, so as to position it for the next reaction cycle.

The comparison of the above X-ray structures with that of Klenetaq1 alone indicates that on binding DNA, the thumb domain moves to wrap around the DNA. It is likely that this conformational change is largely responsible for Pol I's processivity. In both Klenetaq1 · DNA structures, neither the dsDNA nor the single-stranded DNA (**ssDNA**) passes through the cleft between the thumb and fingers domain as the shape and position of the cleft suggest. Rather, the template strand makes a sharp bend at the first unpaired base, thereby unstacking this base and positioning this ssDNA on the same side of the cleft as the dsDNA. Similar arrangements have been observed in X-ray structures of other DNA polymerases in their complexes with DNA.

f. The DNA Polymerase Catalytic Mechanism Involves Two Metal Ions

The X-ray structures of a variety of DNA polymerases suggest that they share a common catalytic mechanism for nucleotidyl transfer (Fig. 30-10). Their active sites all contain two metal ions, usually Mg^{2+} , that are liganded by two invariant Asp side chains in the palm domain. Metal ion B in Fig. 30-10 is liganded by all three phosphate groups of the bound dNTP, whereas metal ion A bridges the α -phosphate group of this dNTP and the primer's 3'-OH group. Metal ion A presumably activates the primer's 3'-OH group for an in-line nucleophilic attack on the α -phosphate group (Fig. 16-6b), whereas metal ion B functions to orient its bound triphosphate group and to electrostatically shield their negative charges as well as the additional negative charge on the transition state leading to the release of the PP_i ion (Section 16-2B).

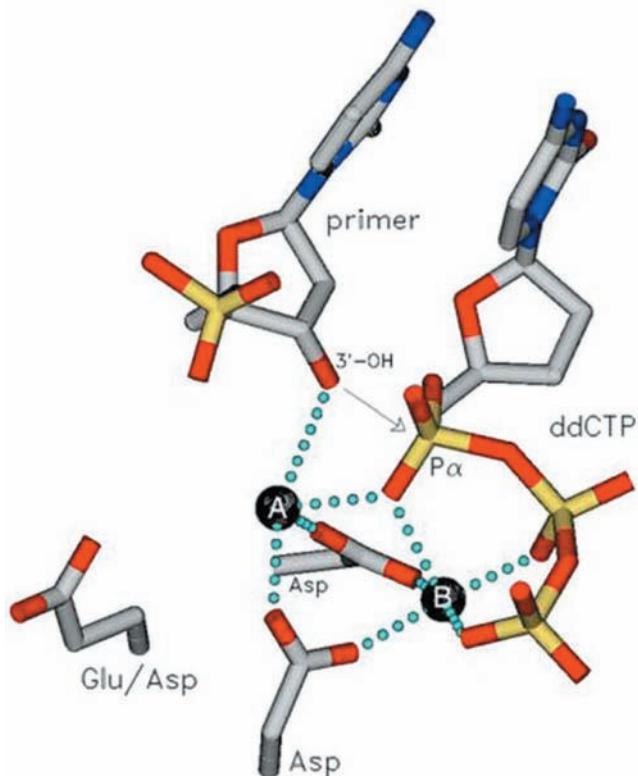


Figure 30-10 Schematic diagram for the nucleotidyl transferase mechanism of DNA polymerases. A and B represent enzyme-bound metal ions that usually are Mg^{2+} . Atoms are colored according to type (C gray, N blue, O red, and P yellow) and metal ion coordination is represented by dotted lines. Metal ion A activates the primer's 3'-OH group for in-line nucleophilic attack on the incoming dNTP's α -phosphate group (arrow), whereas metal ion B acts to orient and electrostatically stabilize the negatively charged triphosphate group. [Courtesy of Tom Ellenberger, Harvard Medical School.]

g. Editing Complexes Contain the Primer Strand in the $3' \rightarrow 5'$ Exonuclease Site

The complex of Klenow fragment with DNA shown in Fig. 30-8 contains a 13-nt primer strand [d(GCCTCGCG-GCGGC)] and a 10-nt template strand [d(GCCGC-GAGGC)] that is complementary to the 10-nt segment at the 5' end of the primer strand. The 3'-terminal nucleotide of the primer strand (the last one that an active polymerase would have added) is bound at the $3' \rightarrow 5'$ exonuclease active site. This arrangement is made possible by the opening up of the G · C base pair that would otherwise be formed by the 5' nucleotide of the template strand, which remains bound near the entrance of the polymerase active site. Evidently, the Klenow fragment has bound the primer strand in an “editing” complex rather than in the polymerase cleft.

In *E. coli* Pol I, how does the 3' end of the primer strand transfer between the polymerase active site and the $3' \rightarrow 5'$ exonuclease active site? This appears to occur through the competition of these sites for the 3' end of the primer strand, which base pairs to form dsDNA in the polymerase

site and binds as a single strand to the exonuclease site. Thus, the formation of a Watson–Crick base pair facilitates the binding of the primer strand to the polymerase site preparatory for the next round of chain extension, whereas a mismatched base pair greatly slows the polymerase reaction while promoting the binding of the primer strand to the exonuclease site. Comparison of the editing complex with those of the Klenetaq1 · DNA complexes suggests that the transfer of the primer strand from the polymerase to the editing sites of Klenow fragment requires that the dsDNA translocate backward (toward the 3' end of the template strand) by several angstroms along the helix axis.

h. Pol I Functions Physiologically to Repair DNA

For some 13 years after Pol I's discovery, it was generally assumed that this enzyme was *E. coli*'s DNA replicase because no other DNA polymerase activity had been detected in *E. coli*. This assumption was made untenable by Cairns and Paula De Lucia's isolation, in 1969, of a mutant *E. coli* whose extracts exhibit <1% of the normal Pol I activity (although it has nearly normal levels of the 5' → 3' exonuclease activity) but which nevertheless reproduce at the normal rate. This mutant strain, however, is highly susceptible to the damaging effects of UV radiation and **chemical mutagens** (substances that chemically induce mutations; Section 32-1A). *Pol I* evidently plays a central role in the repair of damaged (chemically altered) DNA.

Damaged DNA, as we discuss in Section 30-5, is detected by a variety of DNA repair systems. Many of them endonucleolytically cleave the damaged DNA on the 5' side of the lesion, thereby activating Pol I's 5' → 3' exonuclease. While excising this damaged DNA, Pol I simultaneously fills in the resulting single-strand gap through its polymerase activity. In fact, its 5' → 3' exonuclease activity increases 10-fold when the polymerase function is active. Perhaps the simultaneous excision and polymerization activities of Pol I protect DNA from the action of cellular nucleases that would further damage the otherwise gapped DNA.

i. Pol I Catalyzes Nick Translation

Pol I's combined 5' → 3' exonuclease and polymerase activities can replace the nucleotides on the 5' side of a single-strand nick on otherwise undamaged DNA. These reactions, in effect, translate (move) the nick toward the DNA strand's 3' end without otherwise changing the molecule (Fig. 30-11). This **nick translation** process, in the presence of labeled deoxynucleoside triphosphates, is synthetically employed to prepare highly radioactive DNA (the required nicks may be generated by treating the DNA with a small amount of pancreatic **Dnase I**).

j. Pol I's 5' → 3' Exonuclease Functions

Physiologically to Excise RNA Primers

Pol I's 5' → 3' exonuclease also removes the RNA primers at the 5' ends of newly synthesized DNA while its DNA polymerase activity fills in the resulting gaps (Fig. 5-34). The importance of this function was demonstrated by the isolation

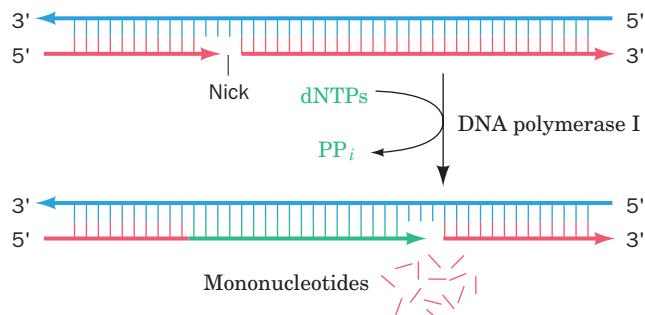


Figure 30-11 Nick translation as catalyzed by Pol I.

of temperature-sensitive *E. coli* mutants that are neither viable nor exhibit any 5' → 3' exonuclease activity at the restrictive temperature of ~43°C (the low level of polymerase activity in the Pol I mutant isolated by Cairns and De Lucia is apparently sufficient to carry out this essential gap-filling process during chromosome replication). Thus Pol I has an indispensable role in *E. coli* DNA replication although a different one than was first supposed.

B. DNA Polymerase III

The discovery of normally growing *E. coli* mutants that have very little Pol I activity stimulated the search for an additional DNA polymerizing activity. This effort was rewarded by the discovery of two more enzymes, designated, in the order they were discovered, **DNA polymerase II (Pol II)** and **DNA polymerase III (Pol III)**. The properties of these enzymes are compared with that of Pol I in Table 30-1. Pol II and Pol III had not previously been detected because their combined activities in the assays used are normally <5% that of Pol I.

A mutant *E. coli* lacking measurable Pol II activity grows normally. However, Pol II has been implicated as a participant in repairing DNA damage via the **SOS response** (Section 30-5D), as have two additional *E. coli* enzymes that

Table 30-1 Properties of *E. coli* DNA Polymerases

	Pol I	Pol II	Pol III
Mass (kD)	103	90	130
Molecules/cell	400	?	10–20
Turnover number ^a	600	30	9000
Structural gene	<i>polA</i>	<i>polB</i>	<i>polC</i>
Conditionally lethal mutant	+	–	+
Polymerization: 5' → 3'	+	+	+
Exonuclease: 3' → 5'	+	+	+
Exonuclease: 5' → 3'	+	–	–

^aNucleotides polymerized min⁻¹ · molecule⁻¹ at 37°C.

Source: Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 167, Freeman (1992).

were more recently discovered: **DNA polymerase IV (Pol IV)** and **DNA polymerase V (Pol V)** (Section 30-5Db).

a. Pol III Is *E. coli*'s DNA Replicase

The cessation of DNA replication in temperature-sensitive *polC* mutants above the restrictive (high) temperature demonstrates that *Pol III is E. coli's DNA replicase*. Its **Pol III core** has the subunit composition $\alpha\epsilon\theta$ where α , the *polC* gene product (Table 30-2), contains the polymerase function. The catalytic properties of Pol III core resemble those of Pol I (Table 30-1) except for Pol III core's inability to replicate primed ssDNA or nicked dsDNA. Rather, Pol III core acts *in vitro* at single-strand gaps of <100 nucleotides, a situation that probably resembles the state of DNA at the replication fork. The Pol III $3' \rightarrow 5'$ exonuclease function, which resides on the enzyme's ϵ subunit, is DNA's primary editor during replication; it enhances the enzyme's replication fidelity by up to 200-fold. However, the Pol III $5' \rightarrow 3'$ exonuclease acts only on single-stranded DNA, so it cannot catalyze nick translation. θ is an accessory protein that stimulates the editing function of ϵ .

The X-ray structure of residues 1 to 917 of the 1160-residue *E. coli* Pol III α subunit, determined Mike O'Donnell and John Kuriyan, reveals that this protein has the expected thumb, fingers, and palm domains (Fig. 30-12). Nevertheless, it exhibits no significant sequence similarity and a different fold from all but two other DNA polymerases of known structure (both from gram-positive bacteria). In addition, Pol III α has an N-terminal PHP (for polymerases and histidinol phosphatase) domain that buttresses both the palm and thumb domains.

*Pol III core ($\alpha\epsilon\theta$) functions in vivo as part of a complicated and labile multisubunit enzyme, the **Pol III holoenzyme**, which consists of at least 10 types of subunits (Table*

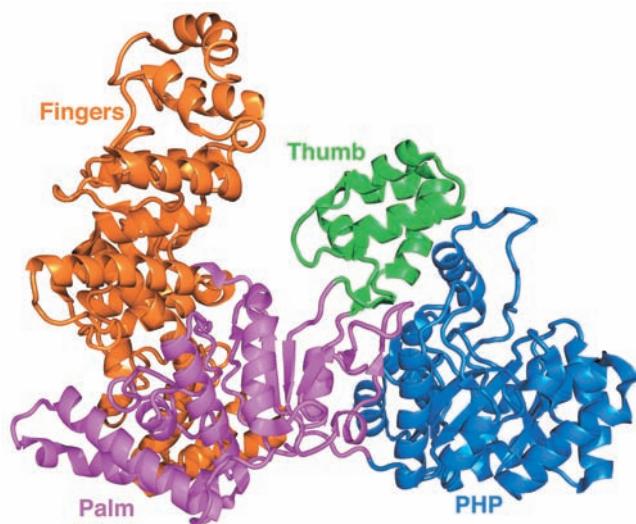


Figure 30-12 X-ray structure of the *E. coli* Pol III α subunit.

The protein is drawn in ribbon form with its thumb, PHP, palm, and finger domains green, blue, magenta, and orange, respectively. Note the handlike shape of the protein but its entirely different fold from that of Klentaq1 (Fig. 30-9). [Based on an X-ray structure by Mike O'Donnell, The Rockefeller University, and John Kuriyan, University of California at Berkeley. PDBid 2HQ4.]

30-2). The latter 7 subunits in Table 30-2 act to modulate Pol III core's activity. For example, Pol III core has a processivity of 10 to 15 residues; it can only fill in short single-stranded regions of DNA. However, Pol III core is rendered processive by association with the **β subunit** in the presence of the 7-subunit **γ complex** ($\gamma\tau_2\delta\delta'\chi\Psi$). Assembly of the processive enzyme is a two-stage process in which the γ complex transfers the β subunit to the primed template in an ATP-dependent reaction followed by the assembly of Pol III core with the β subunit on the DNA (Section 30-3Cc). The β subunit confers essentially unlimited processivity (>5000 residues) on the core enzyme even if the γ complex is subsequently removed. In fact, the β subunit is very strongly bound to the DNA, although it can freely slide along it.

b. The β Subunit Forms a Ringlike Sliding Clamp

The observation that a β subunit clamped to a cut circular DNA slides to the break and falls off suggests that the β subunit forms a closed ring around the DNA, thereby preventing its escape. Kuriyan and O'Donnell determined the X-ray structure of the β subunit in complex with a primer-template DNA (dsDNA with a single-stranded extension on the $5'$ end of one of its strands, the template strand). The protein forms a homodimer of C-shaped, 366-residue monomer units that associate to form an ~ 80 - \AA -diameter doughnut-shaped structure (Fig. 30-13a) that is equivalently known as the **sliding clamp** and the **β clamp**. The sliding clamp's central hole is ~ 35 \AA in diameter, which is larger than the 20- and 26- \AA diameters of B- and A-DNAs (recall that the hybrid helices which RNA

Table 30-2 Components of *E. coli* DNA Polymerase III Holoenzyme

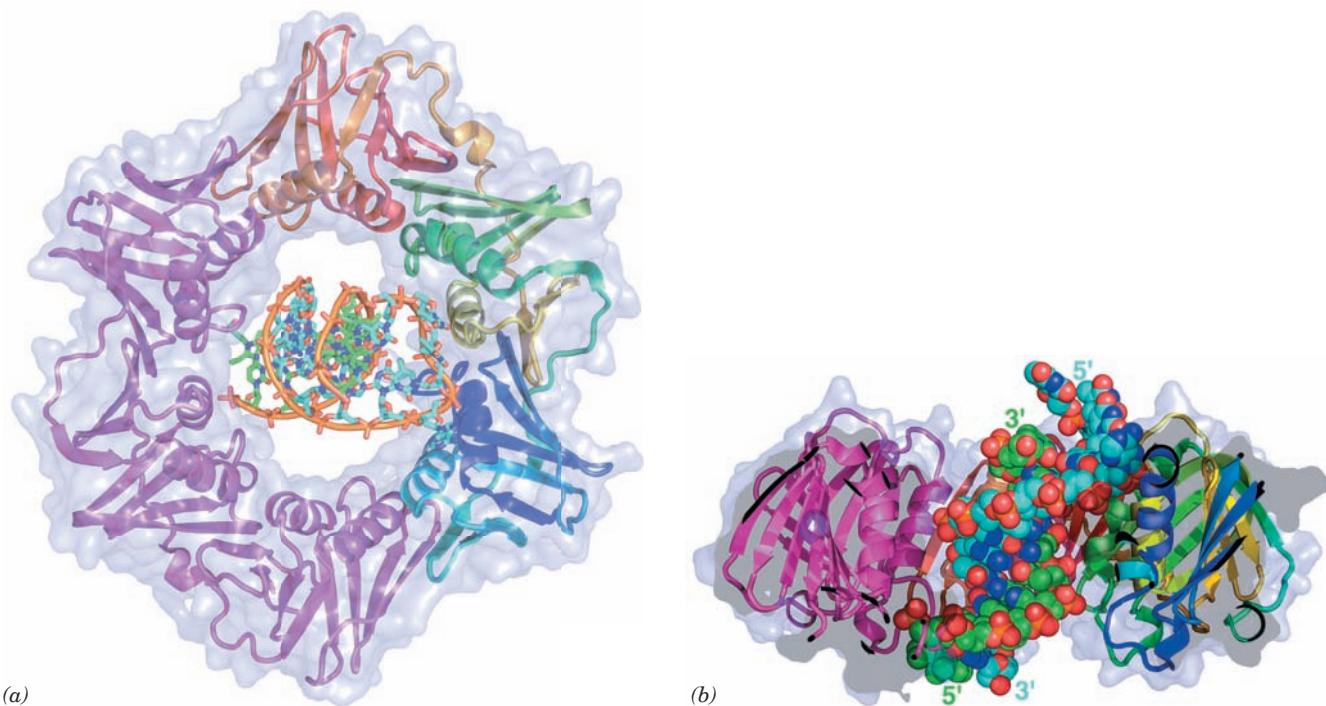
Subunit	Mass (kD)	Structural Gene
α^a	130	<i>polC</i> (<i>dnaE</i>)
ϵ^a	27.5	<i>dnaQ</i>
θ^a	10	<i>holE</i>
τ^b	71	<i>dnaX^c</i>
γ^b	45.5	<i>dnaX^c</i>
δ^b	35	<i>holA</i>
δ'^b	33	<i>holB</i>
χ^b	15	<i>holC</i>
ψ^b	12	<i>holD</i>
β	40.6	<i>dnaN</i>

^aComponents of the Pol III core.

^bComponents of the γ complex.

^cThe γ and τ subunits are encoded by the same gene sequence; the γ subunit comprises the N-terminal end of the τ subunit.

Sources: Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 169, Freeman (1992); and Baker, T.A. and Wickner, S.H., *Annu. Rev. Genet.* **26**, 450 (1992).



(a)

(b)

Figure 30-13 X-ray structure of the β subunit of *E. coli* Pol III holoenzyme in complex with DNA. (a) The homodimeric sliding clamp is drawn in ribbon form embedded in its semitransparent surface diagram and viewed along its 2-fold axis with one subunit magenta and the other colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA, which consists of a 10-bp double-stranded segment with a 4-nt single-stranded extension at the 5' end of one of the strands, is drawn in stick form with the C atoms of the template strand cyan, those of the primer strand green, N blue, O red, and P orange, and with an

orange rod connecting successive P atoms in each strand. (b) Cutaway diagram of the structure in Part a rotated 90° about the horizontal axis. The DNA, which is shown in space-filling form, is inclined by $\sim 22^\circ$ to the protein's 2-fold axis, which is vertical in this diagram. Note that in a replicating system, the primer strand would be extended toward the top of the diagram. [Based on an X-ray structure by John Kuriyan, University of California at Berkeley, and Mike O'Donnell, The Rockefeller University. PDBid 3BEP.]

primers make with DNA have A-DNA-like conformations; Section 29-1Bc). Each β subunit consists of six tandem $\beta\alpha\beta\beta\beta$ motifs of identical topology, which associate in pairs to form three pseudo-2-fold symmetric domains of very similar structures (although with $<20\%$ sequence identity). The dimeric ring therefore has the shape of a 6-pointed star in which the 12 helices line the central hole and the β strands associate in six β sheets that form the protein's outer surface. Electrostatic calculations indicate that the interior surface of the ring is positively charged, whereas its outer surface is negatively charged.

The α helices lining the protein's central hole are all approximately perpendicular to their radially adjacent segments of the sugar-phosphate backbone. These helices therefore span the major and minor grooves of the DNA rather than entering into them as do many helices that make sequence-specific interactions with dsDNA (e.g., Section 31-3Da). Since A- and B-DNAs have 11 and 10.5 bp per turn, whereas the sliding clamp has a pseudo-12-fold symmetry, it appears that the sliding clamp largely minimizes its associations with its threaded DNA, which facilitates the unencumbered passage of the DNA through the

sliding clamp. Nevertheless, the primer-template DNA's helix axis is inclined to the homodimeric protein's 2-fold axis by $\sim 22^\circ$ such that its ssDNA segment is in van der Waals contact with a specific portion of the β clamp's inner wall (Fig. 30-13b; in previous model building studies based on the structure of the sliding clamp alone, these axes were assumed to be coincident). In fact, primer-template DNA binds to the sliding clamp ~ 4 -fold more tightly than does dsDNA and the mutation of the residues that interact with the ssDNA segment significantly reduces the efficiency of DNA replication. Indeed, as is explained in Section 30-3Cc, the interaction between the ssDNA and the sliding clamp is physiologically significant.

C. Unwinding DNA: Helicases and Single-Strand Binding Protein

Pol III holoenzyme, unlike Pol I, cannot unwind dsDNA. Rather, *three proteins*, DnaB protein (the product of the *dnaB* gene; proteins may be assigned the name of the gene specifying them but in roman letters with the first letter capitalized), Rep helicase, and single-strand binding protein

Table 30-3 Unwinding and Binding Proteins of *E. coli* DNA Replication

Protein	Subunit Structure	Subunit Mass (kD)
DnaB protein	Hexamer	50
SSB	Tetramer	19
Rep protein	Monomer	68
PriA protein	Monomer	76

Source: Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 366, Freeman (1992).

(SSB) (Table 30-3), work in concert to unwind the DNA before an advancing replication fork (Fig. 30-14) in a process that is driven by ATP hydrolysis.

a. Hexameric Helicases Mechanically Separate the Strands of dsDNA by Climbing Up One Strand

Access to the genetic information encoded in a double helical nucleic acid requires that its base-paired strands be separated. The proteins that do so, which are known as **helicases**, form a diverse group of enzymes that facilitate a variety of functions including DNA replication, recombination, and repair, as well as transcription termination (Section 31-2Da), RNA splicing, and RNA editing (Section 31-4A). Indeed, all forms of life contain helicases, 12 varieties of which occur in *E. coli*. A helicase functions by translocating along one strand of a double helical nucleic acid so as to separate the strands in its path. This, of course, requires free energy, and hence helicases are driven by NTP hydrolysis. Helicases have been classified into six superfamilies that vary in their characteristics, including their direction of translocation along their bound single strand ($5' \rightarrow 3'$ or $3' \rightarrow 5'$) and whether they function as hexameric rings or dimers.

E. coli **DnaB** protein, a **hexameric helicase** of identical 471-residue subunits, separates the strands of dsDNA by translocating along the lagging strand template in the $5' \rightarrow 3'$ direction while hydrolyzing ATP (it can also use GTP and CTP but not UTP). Electron microscopy and X-ray studies reveal that DnaB forms a hexameric ring that, depending on conditions, exhibits C_3 or C_6 symmetry and which encloses an $\sim 30\text{-}\text{\AA}$ -diameter central channel. Hexameric DnaB binds three primase molecules via the latter's helicase binding domain (Section 30-2E). Similarly, the bacteriophage **T7 gene 4 helicase/primase** (bacteriophage T7 infects *E. coli*) forms a two-tiered hexagonal ring (Fig. 30-15) whose N-terminal domains (residues 1–271) contain its primase activity and whose C-terminal domains (residues 272–566) carry out its helicase function. T7 gene 4 helicase/primase (also called **T7 gp4**; gp for gene product) translocates along ssDNA in the $5' \rightarrow 3'$ direction while preferentially hydrolyzing dTTP (but also hydrolyzes dATP and ATP).

Leemor Joshua-Tor determined the only known X-ray structure of a hexameric helicase in complex with DNA, that of the **E1 protein** of bovine papillomavirus, which translocates along ssDNA in the $3' \rightarrow 5'$ direction (the

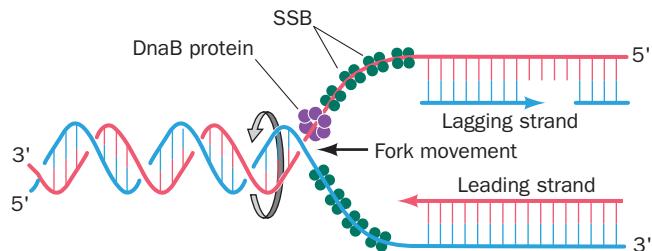


Figure 30-14 Unwinding of DNA by the combined action of DnaB and SSB proteins. The hexameric DnaB protein moves along the lagging strand template in the $5' \rightarrow 3'$ direction. The resulting separated DNA strands are prevented from reannealing by SSB binding.

opposite direction of DnaB and T7 gp4). The protein in the structure, which contains the C-terminal 274 residues of the 605-residue E1 protein, consists of two domains: a 74-residue N-terminal oligomerization domain and a 200-residue C-terminal **AAA+** domain (AAA+ for ATPases associated with cellular activities; a functionally diverse protein family). There are two families of hexameric helicases, the RecA family and the AAA+ family. RecA-family hexameric helicases (RecA catalyzes homologous recombination; Section 30-6Ab), such as DnaB and T7 gp4,

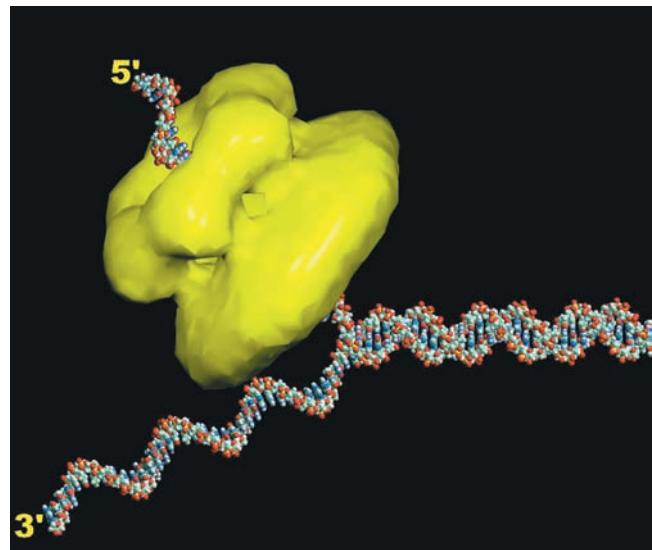


Figure 30-15 Electron microscopy-based image reconstruction of T7 gene 4 helicase/primase. In this two-tiered hexameric ring (yellow), the smaller lobe of each subunit forms the N-terminal primase domain and the larger lobe forms the C-terminal helicase domain. The protein is postulated to interact with DNA as is depicted by this model of a DNA fork consisting of a 30-bp duplex segment and two 25-nt single-stranded segments with the 5' tail threaded through the hexameric ring. The way in which the 3' tail interacts with the protein, if at all, is unknown. [Courtesy of S.S. Patel and K.M. Picha, University of Medicine and Dentistry of New Jersey.]

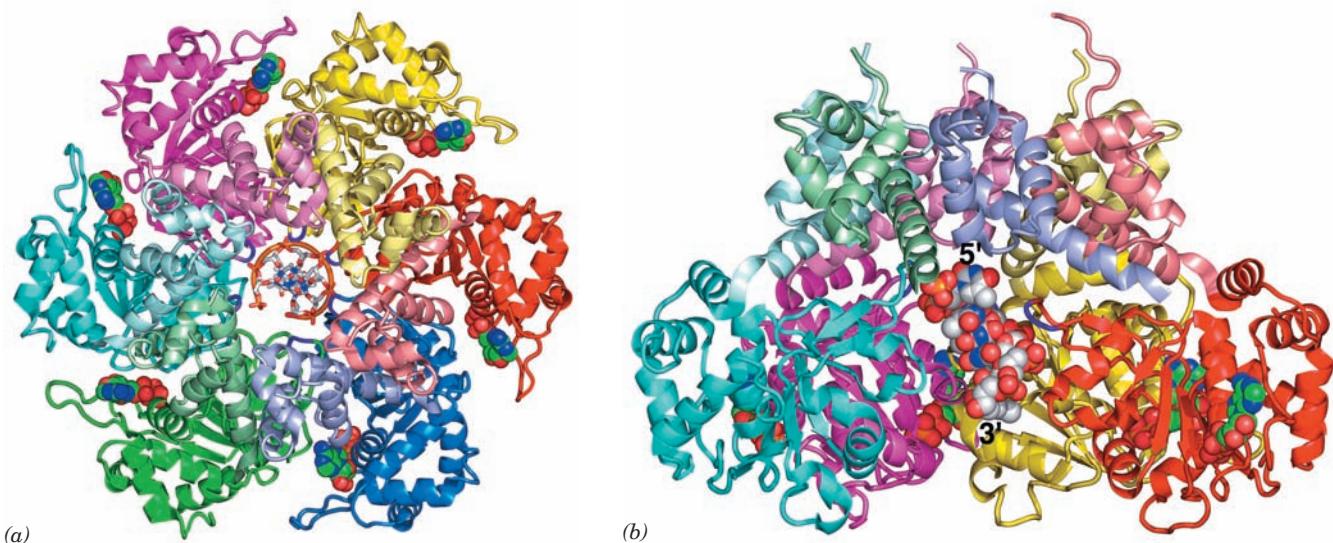


Figure 30-16 X-ray structure of bovine papillomavirus E1 helicase in complex with poly(dT) and ADP. (a) The protein is drawn in ribbon form viewed along the homohexameric's pseudo-6-fold axis with each protein subunit differently colored and with the oligomerization domain of each subunit more lightly shaded than its AAA+ domain. The protein loops extending radially inward from each subunit to interact with the DNA's phosphate groups are purple. The poly(dT), 6 nt of which are visible, is drawn in stick form with C gray, N blue, O red, and P orange and

with successive P atoms joined by orange rods. Its 5' end is closest to the viewer. The ADP is shown in space-filling form with C green, N blue, O red, and P orange. (b) Side view of the protein related to that in Part a by a 90° rotation about the horizontal axis. The blue and green AAA+ domains in Part a, except for their DNA-interacting loops, have been deleted to expose the DNA, which is drawn in space-filling form. [Based on an X-ray structure by Leemor Joshua-Tor, Cold Spring Harbor Laboratory, New York. PDBid 2GXA.]

translocate in the 5' → 3' direction and occur mainly in eubacteria and their phages, whereas AAA+ hexameric helicases, such as the E1 protein, translocate in the 3' → 5' direction and occur mainly in archaea, eukaryotes, and their viruses.

The E1 structure reveals that this helicase, which was crystallized with ADP and a 13-nt poly(dT) (although only 6 nt are visible in the X-ray structure), forms a two-layered hexagonal ring in which the oligomerization domains form a rigid collar with nearly perfect 6-fold symmetry. In contrast, the AAA+ domains deviate significantly from this symmetry (Fig. 30-16a). An ADP is bound at a radially peripheral site between each neighboring pair of AAA+ domains. The poly(dT) forms a right-handed helix that binds in the minimally ~13-Å-diameter central channel of the AAA+ domain hexamer (which is too narrow to admit dsDNA) with its 5' end toward the top of the hexamer in Fig. 30-16. The DNA's phosphate groups each interact with a positively charged loop (residues 505–508) that extends radially inward from each AAA+ domain and hence these loops form an arrangement that resembles a right-handed spiral staircase that tracks the ssDNA's sugar-phosphate backbone. Apparently, the protein steps through a series of ATP-driven conformational changes that, via interactions with the loops, pushes the ssDNA through the channel from bottom to top in Fig. 30-16b. During this process, each loop maintains its grip on the same phosphate group. ATP

hydrolysis occurs toward the bottom of the spiral staircase and ADP release occurs between subunits located toward its top. A new ATP then binds to this site, which causes the topmost loop to drop to the bottom of the staircase, where it binds the next available phosphate group and repeats the catalytic cycle. Thus the E1 helicase mechanically separates the strands of dsDNA by pulling itself along the groove of one strand in its 3' → 5' direction but without turning relative to the DNA.

b. Rep Helicase Dimers Separate the Strands of dsDNA via an “Active, Rolling” Mechanism

Two other helicases, **Rep helicase** and **PriA protein**, have been implicated in the replication of various *E. coli* phage DNAs (Section 30-3B) and also participate in certain aspects of *E. coli* DNA replication (Section 30-3C). Both proteins translocate along DNA in the 3' → 5' direction (and hence along the opposite strand from DnaB) while hydrolyzing ATP. Rep helicase is not essential for *E. coli* DNA replication but the rate at which *E. coli* replication forks propagate is reduced ~2-fold in *rep*⁻ mutants.

Rep helicase is a 673-residue monomer in solution but dimerizes on binding to DNA. Both subunits of the Rep dimer bind to ssDNA or dsDNA such that DNA binding to one subunit strongly inhibits DNA binding to the other (negative cooperativity). This observation led Timothy Lohman to propose the “active, rolling” mechanism for

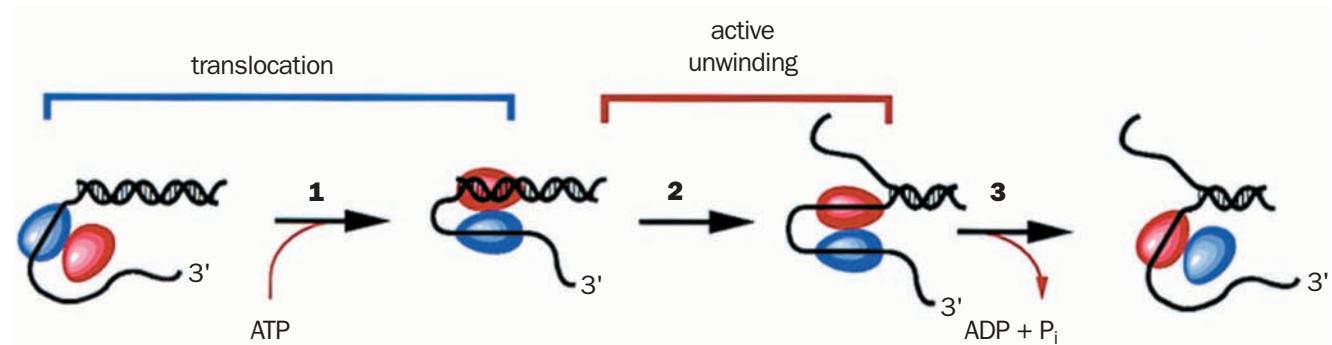


Figure 30-17 Active, rolling mechanism for DNA unwinding by Rep helicase. (1) The subunit of dimeric Rep helicase that is not bound to ssDNA binds to dsDNA accompanied by ATP binding. (2) The subunit bound to dsDNA unwinds the double strand and remains bound to the 3'-ending strand. (3) In a

process that is accompanied by the release of the ATP hydrolysis products, the subunit closer to the 3' end of the bound ssDNA releases it preparatory for a new cycle of dsDNA unwinding. [Courtesy of Gabriel Waksman, Washington University School of Medicine.]

Rep-mediated DNA unwinding in which the two subunits of the dimer alternate in binding dsDNA and the 3' end of the ssDNA at the ssDNA/dsDNA junction (Fig. 30-17). The two subunits then “walk” up the DNA while unwinding it in an ATP-dependent manner via a subunit switching mechanism in which the helicase subunit that is bound to the dsDNA displaces its 5'-starting strand while remaining bound to its 3'-starting strand. Release of the other subunit from the 3'-starting ssDNA then permits this subunit to bind to and unwind the new end of the dsDNA, thereby continuing the cycle.

The X-ray structure of *E. coli* Rep helicase in complex with the short ssDNA dT(pT)₁₅ and ADP (Fig. 30-18), determined by Lohman and Waksman, reveals that the relatively straight ssDNA molecule binds two contacting Rep monomers. A Rep monomer consists of two domains, 1 and 2, each of which is formed by two subdomains, A and B, with the two N-terminal subdomains (1A and 2A) homologous to each other. In the two Rep monomers that are bound to the same ssDNA, subdomain 2B exhibits strikingly different orientations with respect to the other three subdomains (Fig. 30-18). The Rep monomer that is bound to the 5' end of the ssDNA (which it contacts between bases 1 and 8) assumes the “open” conformation in which the four subdomains form an assembly that is reminiscent of a crab claw with one pincer (subdomain 2B) larger than the other (subdomain 1B). The DNA is bound at the bottom of the resulting cleft, whose floor is formed by subdomains 1A and 2A. In the Rep monomer that binds to the 3' end of the ssDNA (which it contacts between bases 9 and 16), subdomain 2B has reoriented relative to the other subdomains via a 130° rotation about a hinge region between subdomains 2A and 2B, thereby closing the cleft about the DNA to form the “closed” conformation. This conformation change is consistent with the active, rolling mechanism even though the way in which two Rep monomers form the dimer observed in solution remains unknown. The ADP binds to Rep between its subdomains 1A and 2A in close proximity to the DNA, suggesting that conformation changes at the ATP-binding site arising from ATP hydroly-

sis are transmitted to the DNA-binding site via the secondary structural elements that contact both sites. The way in which Rep separates the two strands of dsDNA is, as yet, unknown.

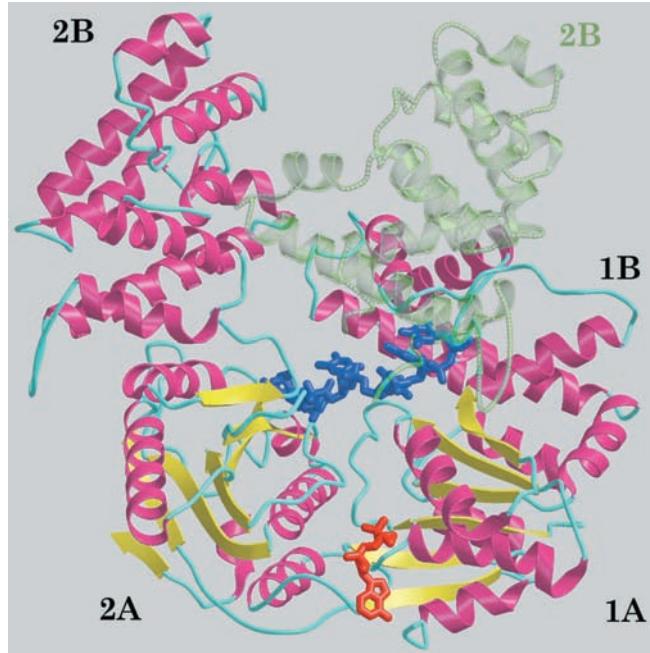


Figure 30-18 Superposition of the X-ray structures of Rep helicase in its open and closed forms in complex with dT(pT)₁₅ and ADP. The monomer in the open conformation is drawn in ribbon form colored according to secondary structure (helices magenta, β sheets yellow, and coil cyan) with its bound ssDNA segment and ADP drawn in stick form in blue and in red. In the closed conformation, subdomain 2B (transparent green ribbon) has rotated via a 130° hinge motion relative to subdomains 1A, 1B, and 2A so as to close over the ssDNA. [Courtesy of Gabriel Waksman, Washington University School of Medicine. PDBid 1UAA.]

c. Single-Strand Binding Protein Prevents ssDNA from Reannealing

If left to their own devices, the separated DNA strands behind an advancing helicase would rapidly reanneal to reform dsDNA. What prevents them from doing so is the binding of **single-strand binding protein (SSB)**. It also prevents ssDNA from forming fortuitous intramolecular secondary structures (helical stems) and protects it from nucleases. Numerous copies of SSB cooperatively coat ssDNA, thereby maintaining it in an unpaired state. Note, however, that ssDNA must be stripped of SSB before it can be replicated by Pol III holoenzyme.

E. coli SSB is a homotetramer of 177-residue subunits. SSB binds ssDNA in several distinct modes referred to as $(SSB)_n$, which differ by the number of nucleotides (n)

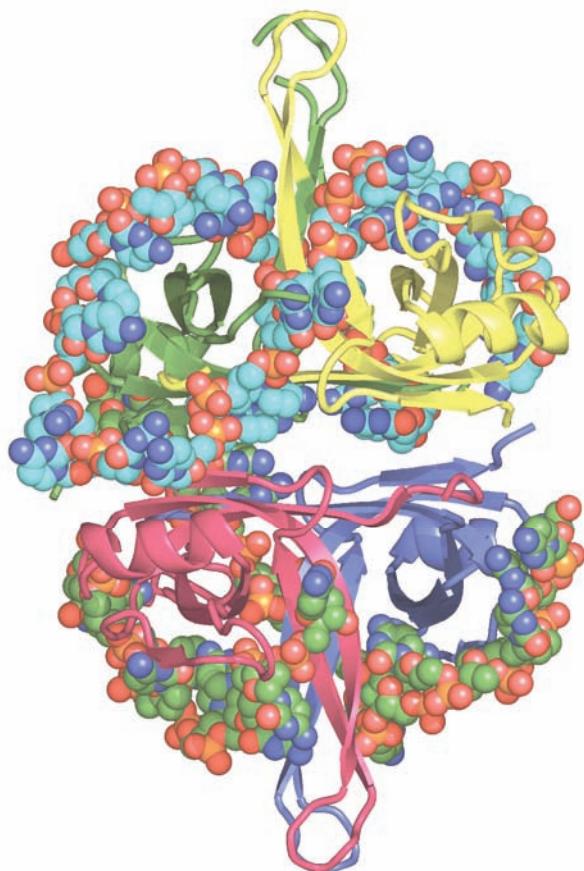


Figure 30-19 X-ray structure of *E. coli* SSB in complex with $dC(pC)_{34}$. The homotetramer, which has D_2 symmetry, is viewed along one of its 2-fold axes with its other 2-fold axes horizontal and vertical. Each of its subunits (which include the N-terminal 134 residues of the 177-residue polypeptide) are differently colored. Its two bound ssDNA molecules are drawn in space-filling form colored according to atom type with the upper strand C cyan, lower strand C green, N blue, O red, and P orange. (The lower strand is partially disordered and hence appears to consist of two fragments.) [Based on an X-ray structure by Timothy Lohman and Gabriel Waksman, Washington University School of Medicine. PDBid 1EYG.]

bound to each tetramer. The two major modes are $(SSB)_{35}$, in which only two of the tetramer's subunits strongly interact with the ssDNA, and $(SSB)_{65}$, in which all four subunits interact with the ssDNA. The $(SSB)_{35}$ mode displays unlimited cooperativity in that it forms extended strings of contacting tetramers along the length of a bound ssDNA, whereas the $(SSB)_{65}$ mode has limited cooperativity in that it forms beaded clusters on ssDNA that consist of only a few contacting tetramers.

Proteolysis studies have shown that SSB's ssDNA-binding site is contained within its 115 N-terminal residues. The X-ray structure of *E. coli* SSB's chymotryptic fragment (residues 1–135) in complex with $dC(pC)_{34}$, determined by Lohman and Waksman, reveals that the tetrameric protein has D_2 symmetry and binds two molecules of $dC(pC)_{34}$ (Fig. 30-19). For one of these 35-mers, 28 nucleotides (residues 3–30) were visible and these assumed the shape of an elongated horseshoe that wrapped around two SSB subunits with approximate 2-fold symmetry and with its apex contacting a third subunit. The other bound ssDNA was partially disordered such that only two segments were visible, one with 14 nt (residues 3–16) and the other with 9 nt (residues 19–27). The paths of the ssDNA segments along the surface of the SSB suggested models that rationalize the different properties of $(SSB)_{35}$ and $(SSB)_{65}$. In the $(SSB)_{65}$ model, the two ends of a 65-nt segment emerge from the same side of the tetramer, which would limit the number of SSB tetramers that can bind to contiguous 65-nt segments of ssDNA. However, in the $(SSB)_{35}$ model, the two ends of a 35-nt segment emerge from opposite ends of the tetramer, thereby permitting an unlimited series of SSB tetramers to interact end-to-end along the length of a ssDNA.

D. DNA Ligase

Pol I, as we saw in Section 30-2A, replaces the Okazaki fragments' RNA primers with DNA through nick translation. *The resulting single-strand nicks between adjacent Okazaki fragments, as well as the nick on circular DNA after leading strand synthesis, are sealed in a reaction catalyzed by DNA ligase.* The free energy required by this reaction is obtained, in a species-dependent manner, through the coupled hydrolysis of either NAD^+ to $NMN^+ + AMP$ or ATP to $PP_i + AMP$. The *E. coli* enzyme, which is also known as **LigA**, is a 671-residue monomer that utilizes NAD^+ and catalyzes a three-step reaction (Fig. 30-20):

1. The adenylyl group of NAD^+ is transferred to the ϵ -amino group of an enzyme Lys residue to form an unusual phosphoamide adduct that is, nevertheless, readily isolated.

2. The adenylyl group of this activated enzyme is transferred to the 5'-phosphoryl terminus of the nick to form an adenylylated DNA. Here, AMP is linked to the 5'-nucleotide via a pyrophosphate rather than the usual phosphodiester bond.

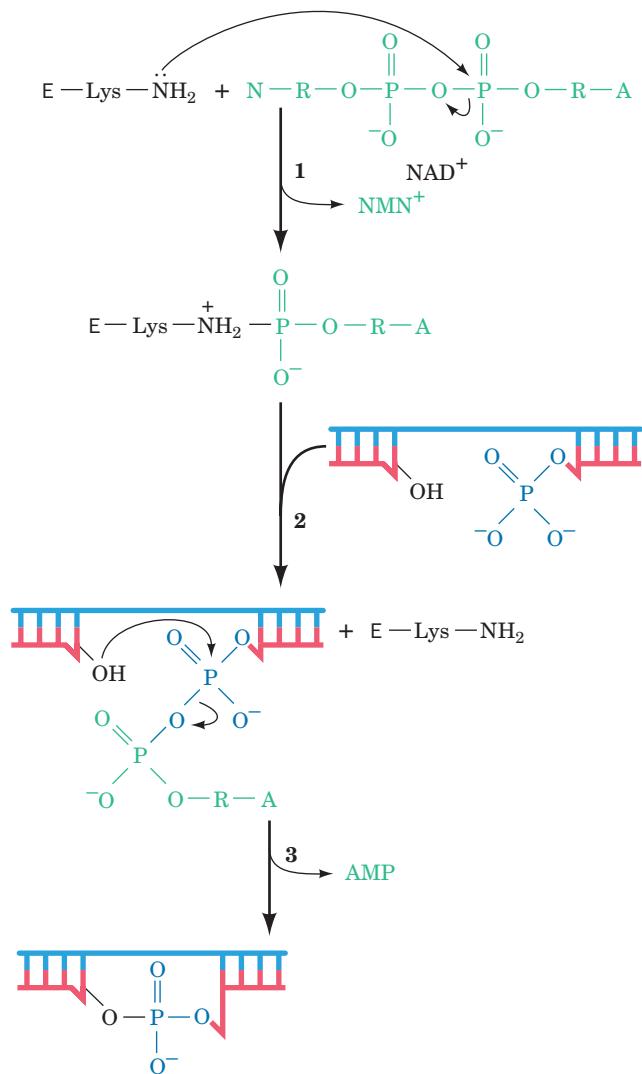


Figure 30-20 The reactions catalyzed by *E. coli* DNA ligase. In eukaryotic and T4 DNA ligases, NAD⁺ is replaced by ATP so that PP_i rather than NMN⁺ is eliminated in the first reaction step. The numbered steps are described in the text.

3. DNA ligase catalyzes the formation of a phosphodiester bond by attack of the 3'-OH on the 5'-phosphoryl group, thereby sealing the nick and releasing AMP.

ATP-requiring DNA ligases, such as those of all eukaryotes and bacteriophage T4, release PP_i in the first step of the reaction rather than NMN⁺. T4 ligase is also noteworthy in that, at high DNA concentrations, it can link together two duplex DNAs (**blunt end ligation**) in a reaction that is a boon to genetic engineering (Section 5-5C).

The X-ray structure of *E. coli* DNA ligase in complex with a singly nicked 26-bp dsDNA and AMP was determined by Stewart Shuman. The complex was formed by reacting the protein with NAD⁺ in the presence of Mg²⁺ (thus forming the phosphoamidate product of step 1 of the DNA ligase reaction; Fig. 30-20), removing the Mg²⁺, and then adding the nicked dsDNA. The X-ray structure of

crystals of this complex revealed that the protein forms a C-shaped clamp that encircles a 19-bp segment of the DNA centered on the nick (Fig. 30-21). Moreover, the complex had progressed through step 2 of the reaction, that is, the adenylyl group had formed a pyrophosphate linkage with the 5'-phosphate group at the nick. The reason that the enzyme did not complete its catalytic cycle by sealing the nick is presumably due to the absence of Mg²⁺.

Residues 587 to 671 form a domain that is not visible in this X-ray structure although it is poorly resolved in the X-ray structure of DNA ligase from *Thermus filiformis*. Apparently, this domain has high mobility, which suggests that it folds out to allow the enzyme's nicked dsDNA substrate to bind to the active site and then folds back to help immobilize the DNA.

E. Primase

The primases from bacteria and several bacteriophages track the moving replication fork in close association with its DNA helicase. Thus, the N-terminal domain of T7 gene 4 helicase/primase forms its primase function (Fig. 30-15), whereas *E. coli* primase (**DnaG**) forms a noncovalent complex with DnaB. Since these DNA helicases translocate along the lagging strand template DNA in its 5' → 3' direction (Fig. 30-14), the primase must reverse its direction of travel in order to synthesize an RNA primer in its 5' → 3' direction. DnaG, which is held to the RNA primed site by its association with SSB, can synthesize up to 60-nt primers *in vitro*, although *in vivo*, primers have the length of 11 ± 1 nt. Since a replication fork in *E. coli* moves at ~1000 nt per second and Okazaki fragments are ~1000 nt in length, about one RNA primer must be synthesized per second at each replication fork. Primases tend to initiate synthesis at specific 3-nt sequences on the template. In *E. coli* this sequence is GTA, which is overrepresented in templates for lagging strand synthesis.

DnaG is a 581-residue monomeric protein. Proteolysis studies have shown that it consists of three domains: an N-terminal Zn²⁺-binding domain (residues 1–110), which tetrahedrally ligands a Zn²⁺ ion via three Cys residues and a His residue and is implicated in recognizing ssDNA; a central RNA polymerase domain (residues 111–433) that carries out primer synthesis; and a C-terminal helicase binding domain (residues 434–581) that interacts with DnaB. Isolated DnaG is only weakly active *in vitro*; it synthesizes primers at a maximum rate of three per hour. However, in the presence of DnaB, it synthesizes primers at the rate observed *in vivo*. Since *E. coli* have 50 to 100 DnaG molecules per cell, this presumably limits primer synthesis to the replication fork. The importance of this function is underscored by T7 gp4, whose helicase and primase functions reside on the same polypeptide (Fig. 30-15).

The X-ray structure of the DnaG catalytic domain in complex with a 15-nt ssDNA (Fig. 30-22), determined by James Berger, reveals a cashew-shaped protein whose fold is unrelated to those of any other DNA or RNA polymerases. It contains an ~100-residue segment that is similar in both sequence and structure to segments in types IA

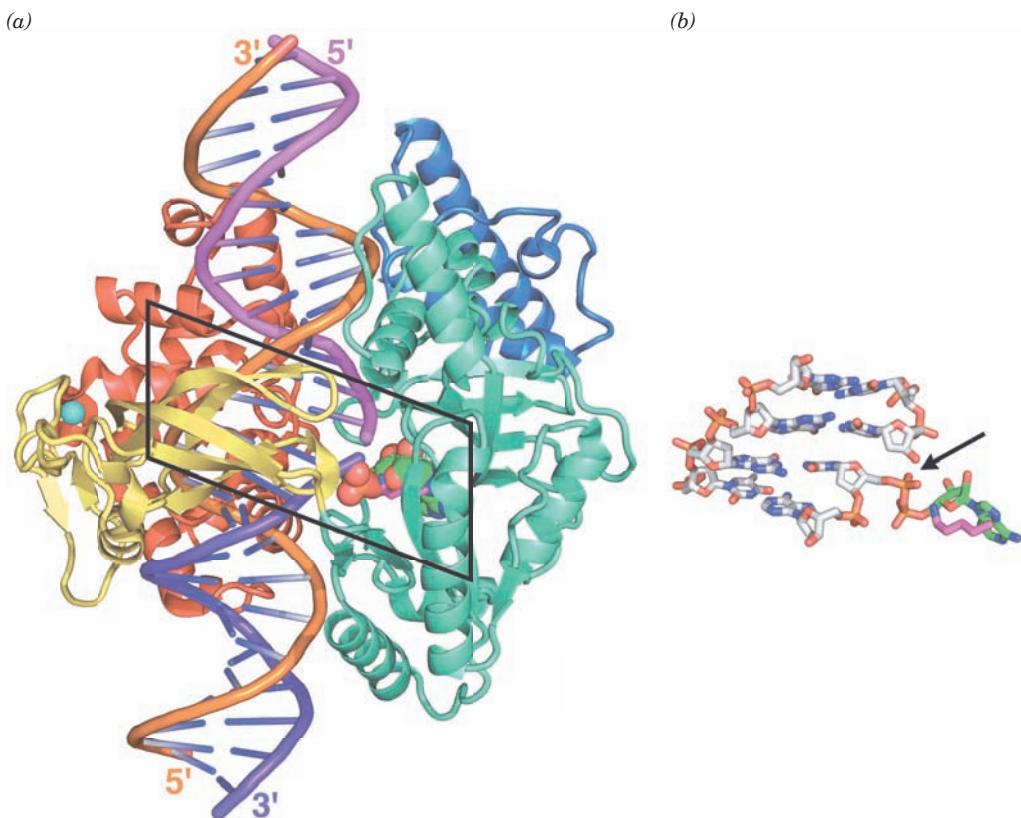


Figure 30-21 X-ray structure of *E. coli* DNA ligase in complex with a singly nicked 26-bp dsDNA and AMP. (a) The enzyme is drawn in ribbon form with its four domains colored, from N- to C-terminus, blue, aquamarine, yellow, and red. The dsDNA is shown in ladder form with the sugar-phosphate backbone of its 26-nt strand orange and those of its complementary two 13-nt strands magenta and purple. The AMP, which is covalently bound in pyrophosphate linkage to the phosphate group at the 5'-end of the purple strand, is drawn in space-filling form with C green, N blue, O red, and P orange. The side chain of Lys 115, which forms a phosphamide adduct in step 1 of the DNA ligase reaction (Fig. 30-20) is shown in stick form with C magenta and N blue. A Zn²⁺ ion, represented by a cyan sphere, is tetrahedrally liganded by four Cys residues. It is distant from the active site and therefore appears to have structural rather than catalytic function. (b) The 4 bp of nicked DNA in the boxed area of Part a in pyrophosphate linkage with the AMP together with the side chain of Lys 115 are all shown in stick form. The structure is viewed and colored as in Part a but with DNA C gray. The arrow points to the DNA's single-strand nick. [Based on an X-ray structure by Stewart Shuman, Sloan-Kettering Institute, New York, New York. PDBid 2OWO.]

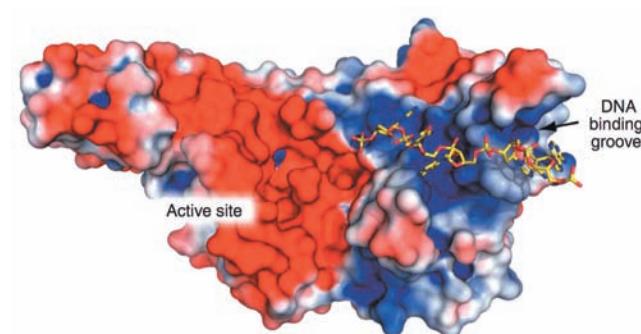


Figure 30-22 X-ray structure of *E. coli* primase in complex with ssDNA. The protein is represented by its molecular surface colored according to its electrostatic potential with red negative, white nearly neutral, and blue positive. The DNA is drawn in stick form with C and P yellow, N blue, and O red. Note the strongly basic character of the DNA binding groove and the highly acidic nature of the active site region. [Courtesy of James Berger, University of California at Berkeley. PDBid 3B39.]

of the DNA ligase reaction (Fig. 30-20) is shown in stick form with C magenta and N blue. A Zn²⁺ ion, represented by a cyan sphere, is tetrahedrally liganded by four Cys residues. It is distant from the active site and therefore appears to have structural rather than catalytic function. (b) The 4 bp of nicked DNA in the boxed area of Part a in pyrophosphate linkage with the AMP together with the side chain of Lys 115 are all shown in stick form. The structure is viewed and colored as in Part a but with DNA C gray. The arrow points to the DNA's single-strand nick. [Based on an X-ray structure by Stewart Shuman, Sloan-Kettering Institute, New York, New York. PDBid 2OWO.]

and IIA topoisomerases (Section 29-3C) and has therefore been named the **Toprim fold** (for topoisomerase and primase). The Toprim fold consists of a 4-stranded parallel β sheet flanked by three helices that resembles the nucleotide-binding (Rossmann) fold (Section 8-3Bi). The active site is marked by several residues that are highly conserved in DnaG-type primases, and in particular, a Glu and two Asp residues, which are invariant in all known Toprim folds and which, in the X-ray structure of a type IIA topoisomerase, coordinate an Mg²⁺ ion.

The ssDNA in the structure, only 5 nt of which are visible, occupies a positively charged groove on the surface of the catalytic subunit that feeds into its active site. The protein makes only a few hydrogen bonding and van der Waals interactions with the DNA's sugar-phosphate backbone and no specific interactions with its bases. Apparently, this DNA binding groove functions to nonspecifically capture a DNA template strand and direct it to the enzyme's active site.

3 PROKARYOTIC REPLICATION

Bacteriophages are among the simplest biological entities and their DNA replication mechanisms reflect this fact. Much of what we know about how DNA is replicated therefore stems from the study of this process in various phages. In this section we examine DNA replication in the **coliphages** (bacteriophages that infect *E. coli*) **M13** and ϕ X174 and then consider DNA replication in *E. coli* itself. Eukaryotic DNA replication is discussed in Section 30-4.

A. Bacteriophage M13

Bacteriophage M13 carries a 6408-nt single-stranded circular DNA known as its **viral** or (+) strand. On infecting an *E. coli* cell, this strand directs the synthesis of its complementary or (–) strand to form the circular duplex **replicative form (RF)**, which may be either nicked (**RF II**) or supercoiled (**RF I**). This replication process (Fig. 30-23)

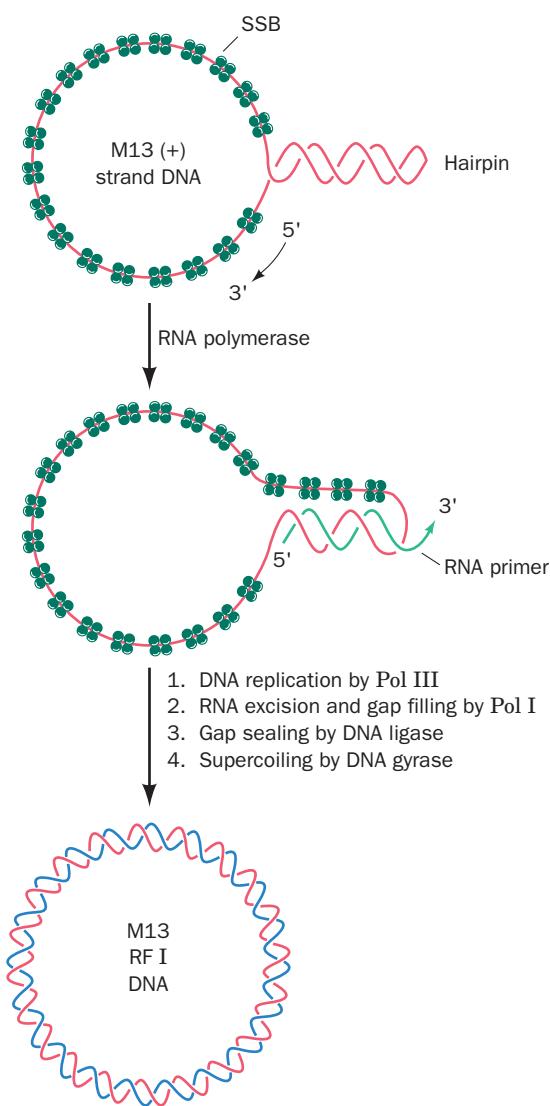


Figure 30-23 The synthesis of the M13 (–) strand DNA on a (+) strand template to form M13 RF I DNA.

may be taken as a paradigm for leading strand synthesis in duplex DNA.

As the M13 (+) strand enters the *E. coli* cell, it becomes coated with SSB except at a palindromic 57-nt segment that forms a hairpin. RNA polymerase commences primer synthesis 6 nt before the start of the hairpin and extends the RNA 20 to 30 residues to form a segment of RNA–DNA hybrid duplex. The DNA that is displaced from the hairpin becomes coated with SSB so that when RNA polymerase reaches it, primer synthesis stops. Pol III holoenzyme then extends the RNA primer around the circle to form the (–) strand. The primer is removed by Pol I-catalyzed nick translation, thereby forming RF II, which is converted to RF I by the sequential actions of DNA ligase and DNA gyrase.

B. Bacteriophage ϕ X174

Bacteriophage ϕ X174, as does M13, carries a small (5386 nt) single-stranded circular DNA. Curiously, the *in vivo* conversion of the ϕ X174 viral DNA to its replicative form is a much more complex process than that for M13 DNA in that ϕ X174 replication requires the participation of a nearly 600-kD protein assembly known as a **primosome** (Table 30-4).

a. ϕ X174 (–) Strand Replication Is a Paradigm for Lagging Strand Synthesis

ϕ X174 (–) strand synthesis occurs in a six-step process (Fig. 30-24):

1. The reaction sequence begins in the same way as that for M13: The (+) strand is coated with SSB except for a 44-nt hairpin. A 70-nt sequence containing this hairpin, known as **pas** (for primosome assembly site), is then recognized and bound by the **PriA**, **PriB**, and **PriC** proteins.

2. DnaB and **DnaC** proteins in the form of a **DnaB₆ · DnaC₆** complex add to the DNA with the help of **DnaT** protein in an ATP-requiring process. DnaC protein is then released yielding the **preprimosome**. The preprimosome, in turn, binds primase yielding the primosome.

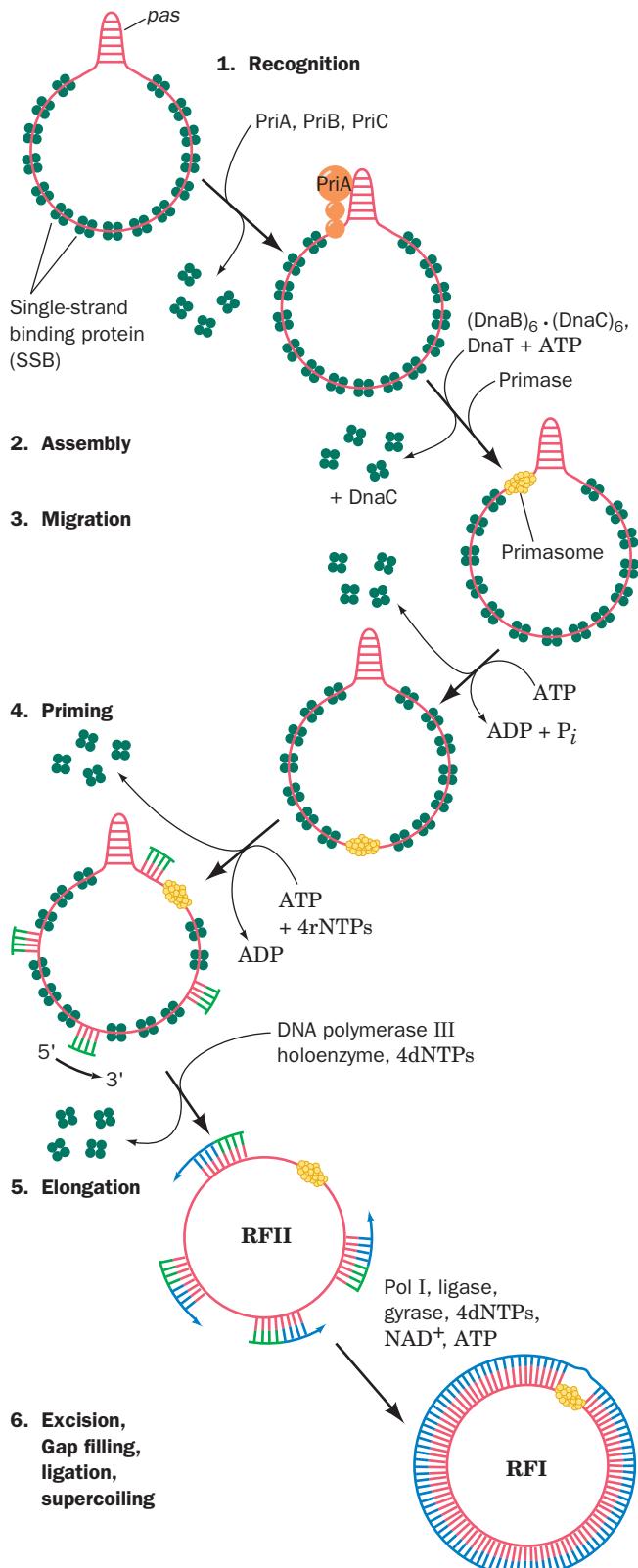
Table 30-4 Proteins of the Primosome^a

Protein	Subunit Structure	Subunit Mass (kD)
PriA	Monomer	76
PriB	Dimer	11.5
PriC	Monomer	23
DnaT	Trimer	22
DnaB	Hexamer	50
DnaC ^b	Monomer	29
Primase (DnaG)	Monomer	60

^aThe complex of all primosome proteins but primase is known as the preprimosome.

^bNot part of the preprimosome or the primosome.

Source: Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), pp. 286–288, Freeman (1992).



3. The primosome is propelled in the $5' \rightarrow 3'$ direction along the (+) strand by the PriA and DnaB helicases at the

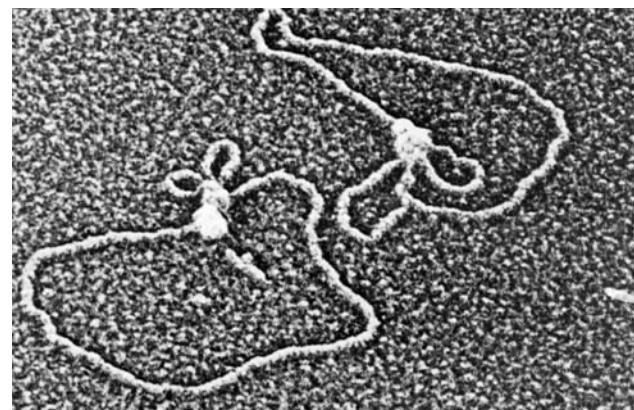


Figure 30-25 Electron micrograph of a primosome bound to a ϕ X174 RF I DNA. Such complexes always contain a single primosome with one or two associated small DNA loops. [Courtesy of Jack Griffith, Lineberger Cancer Research Center, University of North Carolina.]

expense of ATP hydrolysis. This motion, which displaces the SSB in its path, is opposite in direction to that of template reading during DNA chain propagation.

4. At randomly selected sites, the primosome reverses its migration while primase synthesizes an RNA primer. The initiation of primer synthesis requires the participation of DnaB protein which, through concomitant ATP hydrolysis, is thought to alter template DNA conformation in a manner required by primase.

5. Pol III holoenzyme extends the primers to form Okazaki fragments.

6. Pol I excises the primers and replaces them by DNA. The fragments are then joined by DNA ligase and supercoiled by DNA gyrase to form the ϕ X174 RF I.

The primosome remains complexed with the DNA (Fig. 30-25) where it participates in (+) strand synthesis (see below).

b. ϕ X174 (+) Strand Replication Serves as a Model for Leading Strand Synthesis

One strand of a circular duplex DNA may be synthesized via the **rolling circle** or **σ -replication** mode (so called because of the resemblance of the replicating structure to the Greek letter sigma; Fig. 30-26). The ϕ X174 (+) strand is synthesized on an RF I template by a variation on this process, the **looped rolling circle mode** (Fig. 30-27):

1. (+) strand synthesis begins with the primosome-aided binding of the phage-encoded 513-residue enzyme **gene A protein** to its ~30-bp recognition site. There, gene A protein cleaves a specific phosphodiester bond on the (+) strand nucleotide (near the beginning of gene A) by forming a covalent bond between a Tyr residue and the DNA's 5'-phosphoryl group, thereby conserving the cleaved bond's energy.

2. Rep helicase (Section 30-2Cb) subsequently attaches to the (-) strand at the gene A protein and, with the aid of

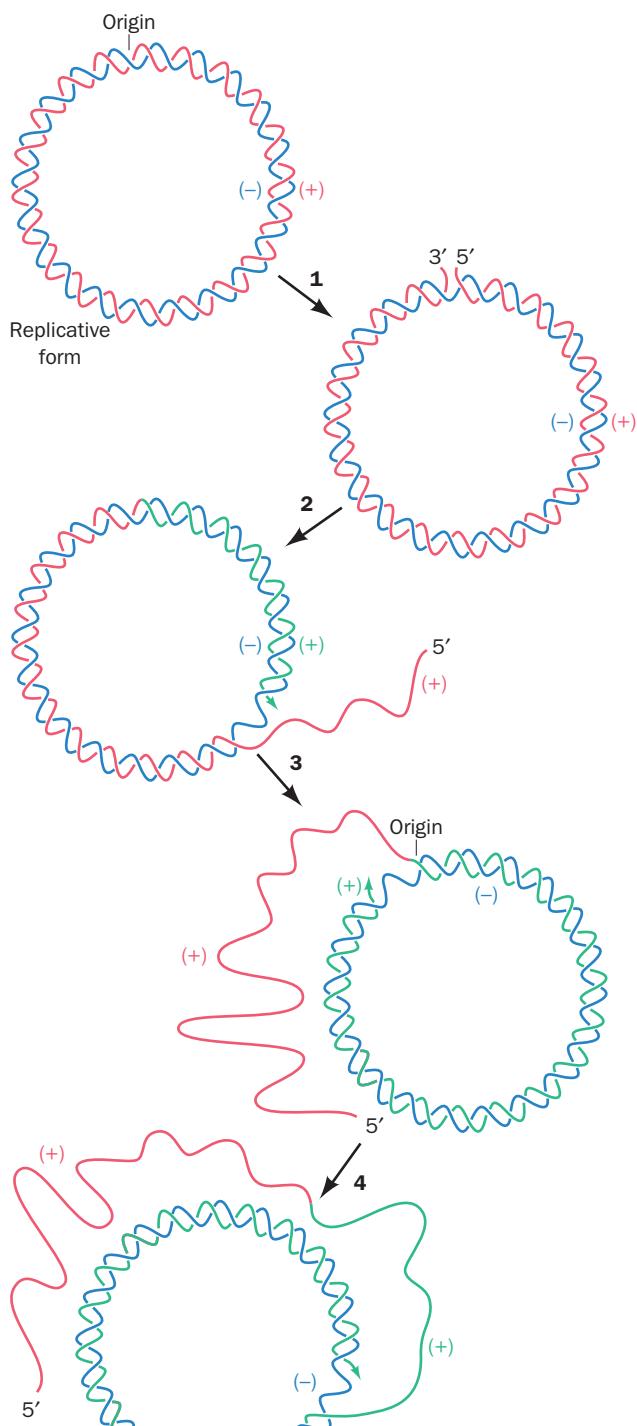


Figure 30-26 The rolling circle mode of DNA replication. The (+) strand being synthesized is extended from a specific cut made at the replication origin (1) so as to strip away the old (+) strand (2 and 3). The continuous synthesis of the (+) strand on a circular (-) strand template produces a series of tandemly linked (+) strands (4), which may later be separated by a specific endonuclease.

the primosome still associated with the (+) strand, commences unwinding the duplex DNA from the (+) strand's

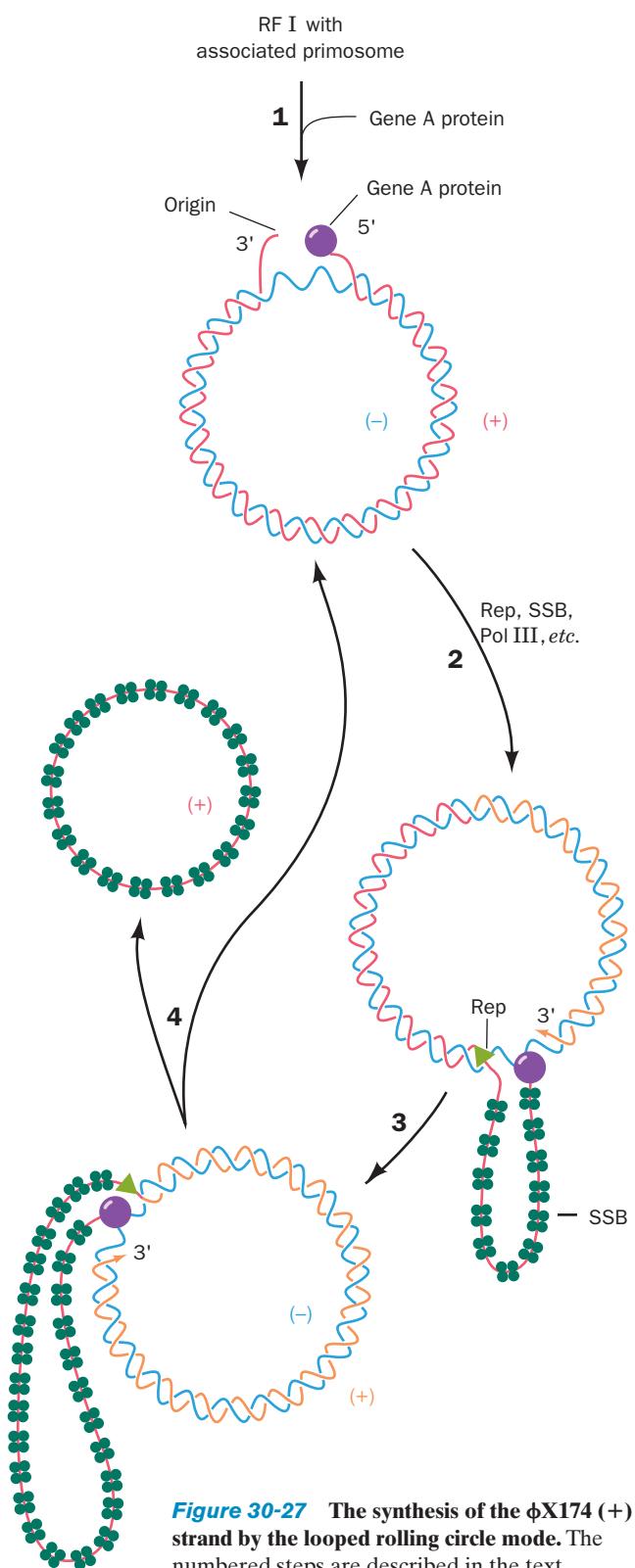


Figure 30-27 The synthesis of the φX174 (+) strand by the looped rolling circle mode. The numbered steps are described in the text.

5' end. The displaced (+) strand is coated with SSB, which prevents it from reannealing to the (-) strand. Rep helicase is essential for the replication of φX174 DNA, but not for the *E. coli* chromosome, as is demonstrated by the inability of φX174 to multiply in *rep*⁻ *E. coli*. Pol III holoenzyme extends the (+) strand from its free 3'-OH group.

3. The extension process generates a **looped rolling circle** structure in which the 5' end of the old (+) strand remains linked to the gene A protein at the replication fork. It is thought that as the old (+) strand is peeled off the RF, the primosome synthesizes the primers required for the later generation of a new (-) strand.

4. When it has come full circle around the (-) strand, the gene A protein again makes a specific cut at the replication origin so as to form a covalent linkage with the new (+) strand's 5' end. Simultaneously, the newly formed 3'-terminal OH group of the old, looped-out (+) strand nucleophilically attacks its 5'-phosphoryl attachment to the gene A protein, thereby liberating a covalently closed (+) strand. This is possible because the gene A protein has two closely spaced Tyr residues that alternate in their attachment to the 5' ends of successively synthesized (+) strands.

The replication fork continues its progress about the duplex circle, producing new (+) strands in a manner reminiscent of linked sausages being pulled off a reel.

In the intermediate stages of a ϕ X174 infection, each newly synthesized (+) strand directs the synthesis of the (-) strand to form RF I as described above. In the later stages of infection, however, the newly formed (+) strands are packaged into phage particles.

C. *Escherichia coli*

See Guided Exploration 25. The replication of DNA in *E. coli*. The *E. coli* chromosome replicates by the bidirectional θ mode from a single replication origin (Section 30-1Aa). The most plausible model for events at the *E. coli* replication fork (Fig. 30-28) is largely derived from studies on the simpler and more experimentally accessible DNA replication mechanisms of coliphages such as M13 and ϕ X174. Duplex DNA is unwound by DnaB helicase on the lagging strand template,

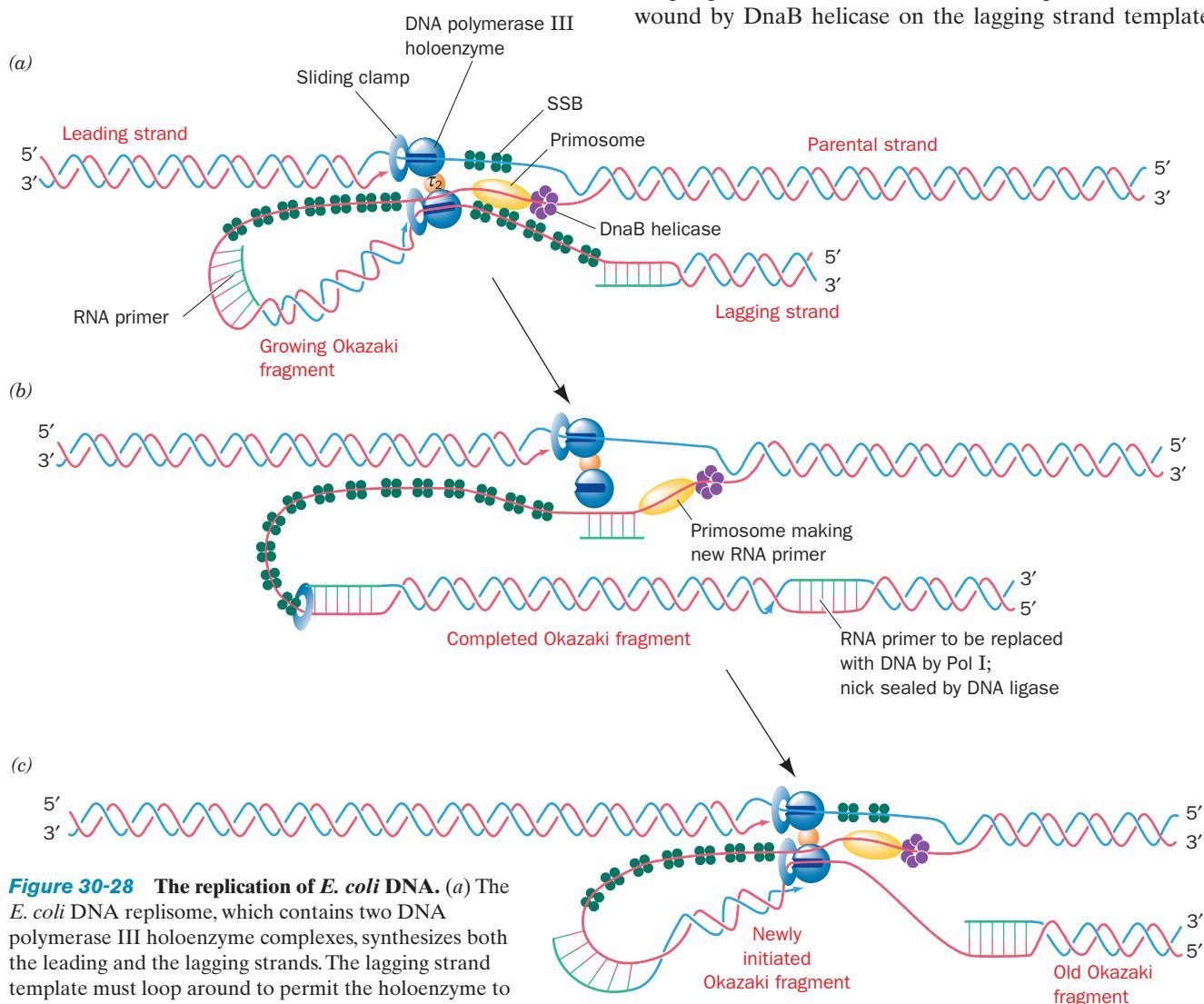
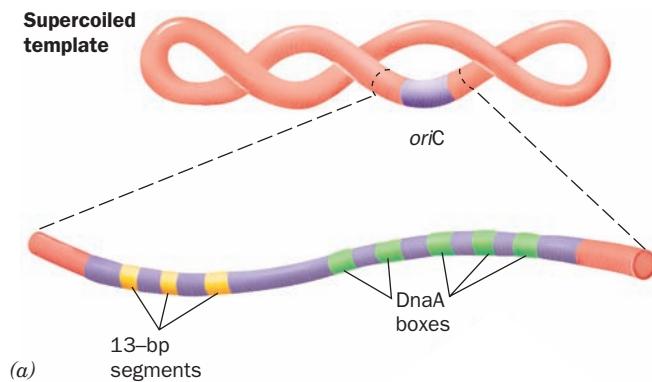


Figure 30-28 The replication of *E. coli* DNA. (a) The *E. coli* DNA replisome, which contains two DNA polymerase III holoenzyme complexes, synthesizes both the leading and the lagging strands. The lagging strand template must loop around to permit the holoenzyme to extend the primosome-primed lagging strand. Although not shown here, the DnaB helicase binds to τ_2 and hence moves with the replisome. (b) The holoenzyme releases the lagging strand template when it encounters the previously synthesized Okazaki fragment. This possibly signals the primosome to initiate the

synthesis of lagging strand RNA primer. (c) The holoenzyme rebinds the lagging strand template and extends the RNA primer to form a new Okazaki fragment. Note that in this model, leading strand synthesis is always ahead of lagging strand synthesis.

where it is joined by the primosome. The separated single strands are immediately coated by SSB. Leading strand synthesis is catalyzed by Pol III holoenzyme, as is that of the lagging strand after priming by primosome-associated primase. Both leading and lagging strand syntheses occur on a single ~900-kD multisubunit particle, the **replisome**, which contains two Pol III cores ($\alpha\epsilon\theta$) that are joined together by a dimer of τ subunits that bridges the α subunits. Hence, the lagging strand template must be looped around (Fig. 30-28). The τ_2 dimer also binds the DnaB helicase (an interaction that is not indicated in Fig. 30-28), thereby stimulating its helicase action while holding it to the replication fork. After completing the synthesis of an Okazaki fragment, the lagging strand holoenzyme relocates to a new primer near the replication fork, the primer heading the previously synthesized Okazaki fragment is excised by Pol I-catalyzed nick translation, and the nick is sealed by DNA ligase. Since lagging strand synthesis is more complex and hence more time-consuming than leading strand synthesis, the replisome functions to coordinate these two processes.



a. *E. coli* DNA Replication Is Initiated at *oriC* in a Process Mediated by DnaA Protein

The replication origin of the *E. coli* chromosome consists of a unique 250-bp segment known as the *oriC* locus. This sequence, segments of which are highly conserved among gram-negative bacteria, supports the bidirectional replication of the various plasmids into which it has been inserted. The *oriC* locus contains five highly conserved 9-bp segments with consensus sequence 5'-TTATCCACA-3' known as **DnaA boxes** because they are specifically bound by **DnaA protein** (Fig. 30-29a). These are interspersed with several so-called **I-sites** that deviate from this consensus sequence and are bound by DnaA with lesser affinity. In addition, the “left” boundary region of *oriC* contains three tandemly repeated, 13-bp, AT-rich segments (consensus sequence 5'-GATCTNTTNTTT-3' where N marks nonspecific positions) that are known as **DNA unwinding elements (DUEs)**.

DnaA (467 residues in *E. coli*) consists of four domains that are, from N- to C-terminus, a helicase interaction domain that mediates interactions with DnaB helicase (see below), a

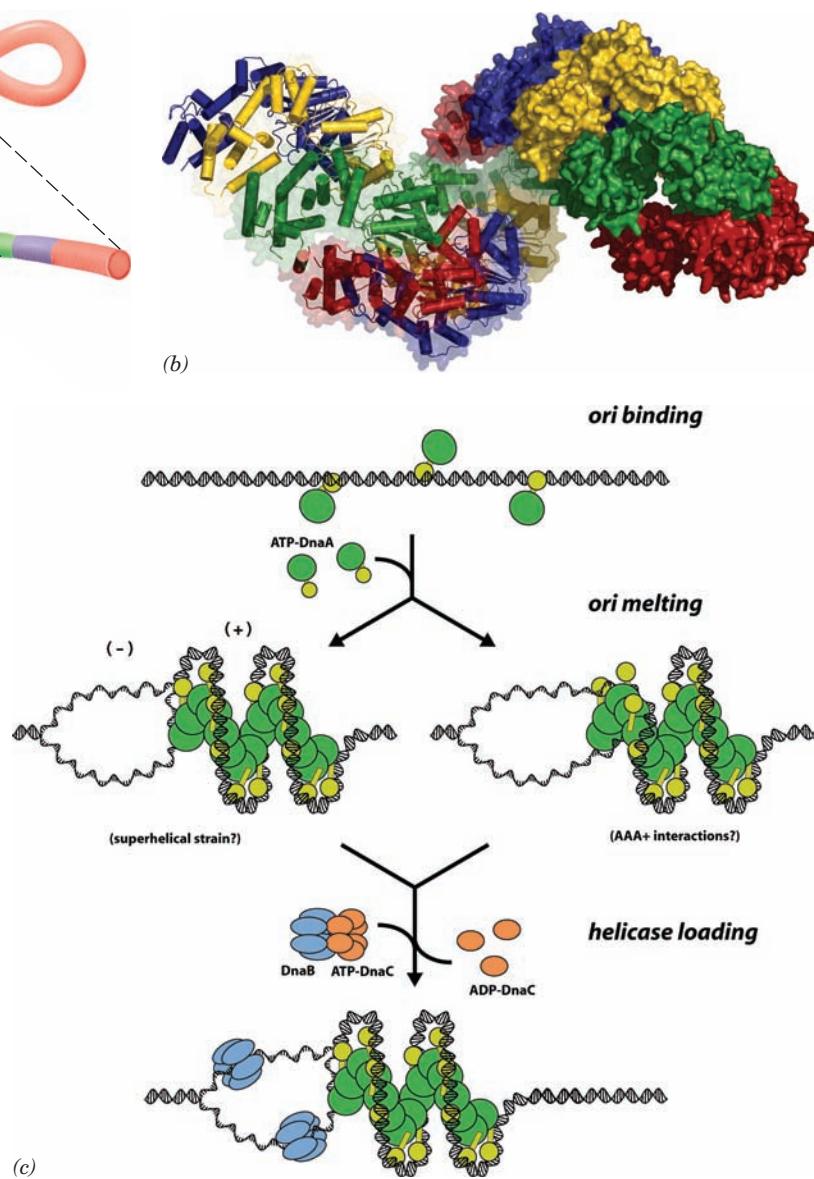


Figure 30-29 DNA replication initiation at *oriC*. (a) Diagram of *oriC* showing the relative positions of its DnaA boxes (green) and its DNA unwinding elements (DUEs; yellow). (b) X-ray structure of the right-handed helical filament formed by the two C-terminal domains of *A. aeolicus* DnaA. It has eight subunits per turn and a pitch of 178 Å. Twelve subunits are shown, from right to left, in the alternating colors red, green, yellow, and blue. The right subunits are drawn as surface diagrams and the left subunits are represented by their polypeptide backbones with helices in tube form. (c) Model for initiation at *oriC*. The green ovals represent the N-terminal three domains of DnaA and the yellow ovals represent the associated C-terminal DNA-binding domains. See the text for an explanation. [Parts b and c courtesy of James Berger, University of California at Berkeley. PDBid 2HCB.]

flexible and poorly conserved linker, an ATPase domain that is a member of the AAA+ family (Section 30-2Ca), and a DNA binding domain. The X-ray structure of the C-terminal two domains of DnaA from *Aquifex aeolicus* (a thermophilic bacterium), determined by Berger, unexpectedly revealed that it forms a multisubunit right-handed helix (Fig. 30-29b).

Experiments with *oriC*-containing plasmids, pioneered by Kornberg, together with the X-ray structure of DnaA, indicate that replication initiation in *E. coli* occurs via the following process (Fig. 30-29c):

1. In the presence of ATP, DnaA, which is normally bound to three of *oriC*'s five DnaA boxes throughout *E. coli*'s lifetime, recruits additional DnaA subunits to the remaining DnaA boxes and to the I-sites so as to form a right-handed helix of DnaA subunits that is bound to the DNA. This generates local positive supercoils in the DNA. The superhelical strain resulting from the compensating negative supercoils (recall that the linking number of a covalently closed circular DNA such as an *E. coli* chromosome is invariant; Section 29-3A) melts the DUE-containing segment [Fig. 30-29c, *middle left*; recall that bacterial chromosomes are normally already negatively supercoiled (Section 29-2Bb)]. Alternatively, or in addition, the DnaA's ATPase domains may actively unwind the DNA (Fig. 30-29c, *middle right*). This process is facilitated by two homologous DNA-binding proteins, **HU** and **integration host factor (IHF)**, that induce DNA bending (IHF is discussed in Section 33-3Ca).

2. The *oriC*-DnaA complex recruits two DnaB₆ · DnaC₆ complexes to opposite ends of the melted region to form the **prepriming complex**. DnaC, an ATPase that is a homolog of DnaA, functions to facilitate the loading of the DnaB hexamers onto the DNA. Its X-ray structure, also determined by Berger, shows that it forms a helical assembly similar to that

of DnaA. The AAA+ domains of DnaA and DnaC interact in an ATP-dependent manner to recruit and properly position the DnaB helicases, following which the DnaC is released.

In the presence of SSB and gyrase, DnaB helicase further unwinds the DNA in the prepriming complex in both directions so as to permit the entry of primase and RNA polymerase. The participation of both these enzymes in leading strand primer synthesis (Section 30-1D), together with the limitation of this process to the *oriC* site, suggests that the RNA polymerase activates primase to synthesize the primer. This perhaps explains the similarity of *oriC*'s DUEs to RNA polymerase's transcriptional promoters (Section 31-2Ba). The stage is thereby set for bidirectional DNA replication by Pol III holoenzyme as described above.

b. The Initiation of *E. coli* DNA Replication Is Strictly Regulated

Chromosome replication in E. coli occurs only once per cell division, so this process must be tightly controlled. The doubling (cell generation) time of *E. coli* at 37°C varies with growth conditions from <20 min to ~10 h. Yet the constant ~1000 nt/s rate of movement of each replication fork fixes the 4.6×10^6 -bp *E. coli* chromosome's replication time, *C*, at ~40 min. Moreover, the segregation of cellular components and the formation of a septum between them, which must precede cell division, requires a constant time, *D* = 20 min, after the completion of the corresponding round of chromosome replication. *Cells with doubling times less than C + D = 60 min must consequently initiate chromosome replication before the end of the preceding cell division cycle.* This results in the formation of **multiforked chromosomes** as is diagrammed in Fig. 30-30 for a cell division time of 35 min.

Even in cells that contain multiple *oriC* sites, DNA replication is initiated at each such site once and only once per cell generation. However, after initiation has occurred, chain elongation proceeds at a uniform, largely uncontrolled rate. This suggests that a post-initiation *oriC* site is somehow sequestered from (prevented from interacting with) the replication initiation machinery, a phenomenon called **sequestration**. There is extensive morphological evidence, such as shown in Fig. 30-31, that the *E. coli* chromosome is associated with the cell membrane. This attachment

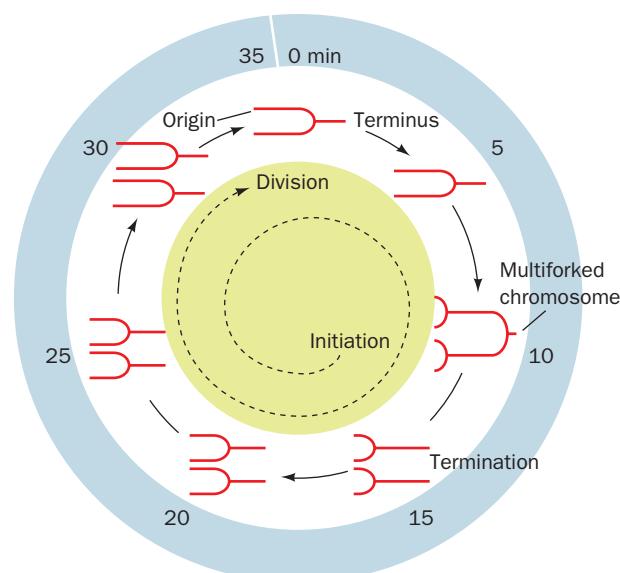


Figure 30-30 Multiforked chromosomes in *E. coli*. In cells that are dividing every 35 min, the fixed 60-min interval between the initiation of replication and cell division results in the production of multiforked chromosomes. [After Lewin, B., *Genes VII*, p. 370, Oxford University Press (2000).]

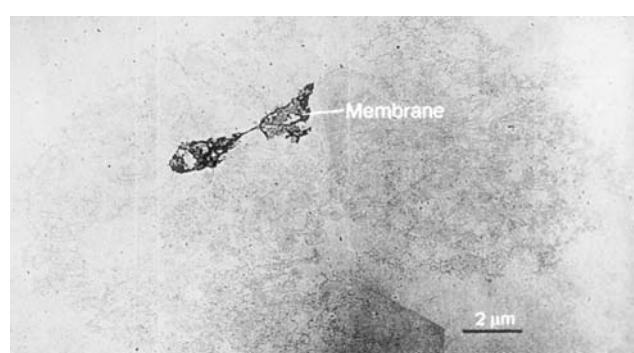


Figure 30-31 Electron micrograph of an intact and supercoiled *E. coli* chromosome attached to two fragments of the cell membrane. [From Delius, H. and Worcel, A., *J. Mol. Biol.* 82, 108 (1974).]

would help explain how replicated chromosomes are segregated into different cells during cell division. But what is the mechanism of sequestration?

The sequence most commonly methylated in *E. coli* is the palindrome GATC, which is methylated at N6 of both its A bases by **Dam methyltransferase** (Section 30-7). GATC occurs 11 times in *oriC*, including at the beginning of all four of its 13-bp DUEs (see above). Newly replicated GATC segments are hemimethylated, that is, the GATC sequences on the newly synthesized strand are unmethylated. Although Dam methyltransferase begins methylating most hemimethylated GATC segments immediately after their synthesis (within ~ 1.5 min), those on *oriC* remain hemimethylated for around one-third of a cell generation. Consequently, the observation that membranes bind hemimethylated *oriC*, but not unmethylated or fully methylated *oriC*, suggests that hemimethylated *oriC* is bound to the membrane in a way that makes it inaccessible to both the initiation machinery and Dam methyltransferase.

The association of hemimethylated *oriC* with membrane requires the presence of the 181-residue **SeqA** protein, the product of *seqA* gene. Thus in *seqA*⁻ cells: (1) the time to fully methylate hemimethylated GATC sites in *oriC* is reduced to 5 min, whereas the time to do so for other GATC sites is unaffected; (2) the synchrony of initiation of multiple *oriC* sites is lost; and (3) in the absence of functional Dam methyltransferase, fully methylated *oriC*-containing plasmids are replicated numerous times per cell generation, whereas in the presence of SeqA they are replicated only once. Evidently, sequestration occurs via the SeqA-mediated binding of hemimethylated *oriC* to the membrane. The hemimethylated promoter of the *dnaA* gene is similarly sequestered so as to repress its transcription, thereby providing an additional mechanism for preventing promiscuous initiation of DNA replication.

c. The Clamp Loader Loads the Sliding Clamp onto the DNA

Extensive investigations in many laboratories have led to the model of the *E. coli* replisome drawn in Fig. 30-32. The sliding clamp, which is responsible for Pol III's high processivity, is a ring-shaped dimer of β subunits through which the DNA strand being replicated is threaded (Section 30-2Bb). The two tightly associated β subunits ($K_D < 50$ nM) that form the sliding clamp dissociate with a half-life of ~ 100 min at 37°C. Yet, since each replisome synthesizes around one Okazaki fragment per second, a sliding clamp must be loaded onto the lagging strand template at this frequency. This loading function is carried out in an ATP-dependent process by the γ complex ($\gamma\tau_2\delta'\chi\Psi$). The τ and γ subunits are both encoded by the *dnaX* gene with τ (643 residues) the full-length product and γ (431 residues) its C-terminally truncated form; the C-terminal 122 residues of τ are known as τ_c . The γ complex, of which only τ_2 is diagrammed in Fig. 30-28, bridges the replisome's two Pol III cores via its τ_c segments, which also bind the DnaB helicase (Fig. 30-32). The χ and Ψ subunits form a heterodimer in which χ competes with primase for its binding site on SSB and hence functions to accelerate the dissociation of primase from the RNA primer it synthesized as well as link the γ complex to SSB. However, χ and Ψ are not essential participants in the clamp loading process and, therefore, we shall refer to the $\gamma\tau_2\delta'$ complex as the **clamp loader**. How does the clamp loader do its job?

Of the clamp loader's five subunits, only δ is capable of binding to and opening up the sliding clamp on its own. Kuriyan and O'Donnell determined the X-ray structure of the δ subunit in 1:1 complex with a β subunit that had two residues in its dimerization interface mutated so as to prevent its dimerization. The structure reveals (Fig. 30-33) that δ , which consists of three domains, inserts its β interaction element, a hydropho-

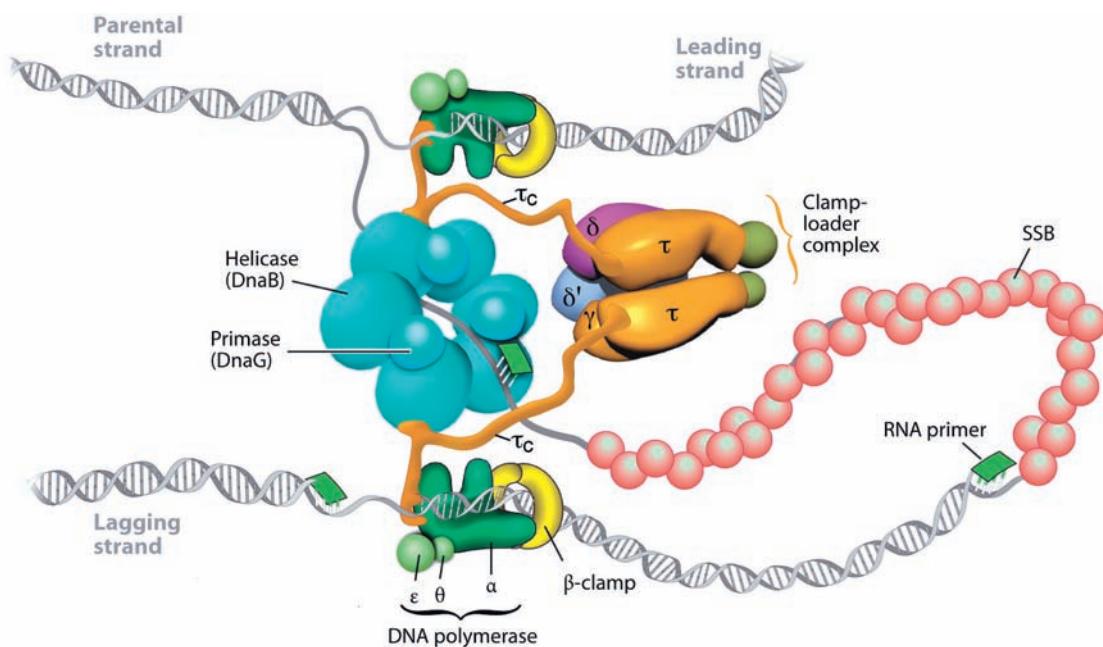


Figure 30-32 Architecture of the *E. coli* replisome. See the text for details. Compare this to Fig. 30-28. [Courtesy of Charles Richardson, Harvard Medical School.]

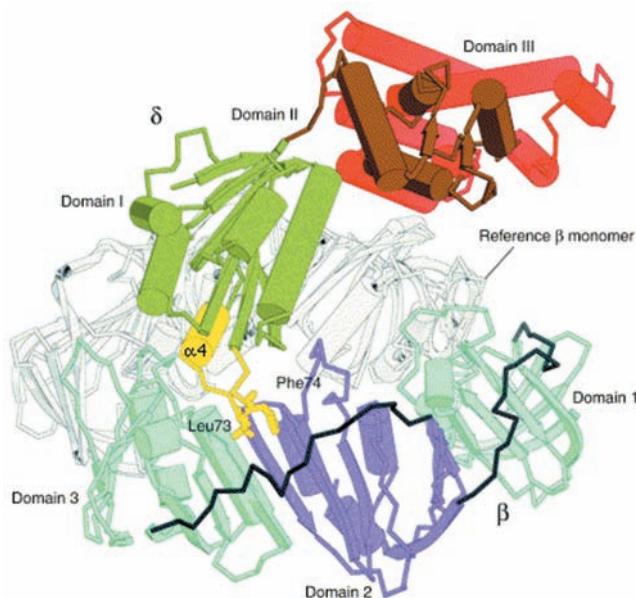


Figure 30-33 X-ray structure of the β - δ complex. A second β subunit taken from the X-ray structure of the sliding clamp (Fig. 30-13), the “Reference β monomer,” is drawn in gray. The view is along the edge of the β ring. The δ subunit’s β interaction element (yellow) consists largely of the $\alpha 4$ helix and two hydrophobic residues, Leu 73 and Phe 74, whose side chains are drawn in stick form. [Courtesy of John Kuriyan, University of California at Berkeley. PDB 1JQJ.]

bic plug that forms the tip of its N-terminal domain, into a hydrophobic pocket on one face of β . Comparison of δ in this structure with that in the $\gamma_3\delta\delta'$ complex (see below) reveals that the β interaction element undergoes a dramatic conformational change on binding to β in which its $\alpha 4$ helix rotates by 45° and translates by 5.5 \AA . Moreover, in forming the β - δ

complex, the β subunit increases its radius of curvature relative to that in the β dimer (Fig. 30-13) such that the β - δ interaction would induce the opening of one of the sliding clamp’s β - β interfaces by $\sim 15 \text{ \AA}$. Such a gap is large enough to permit the passage of ssDNA but not dsDNA. Apparently, the clamp loader functions by trapping one β subunit of the sliding clamp in a conformation that prevents ring closure rather than actively pulling apart the two halves of the ring. This is corroborated by molecular dynamics simulations (Section 9-4a) suggesting that a β_2 dimer has a stable conformation but that an isolated β subunit with the conformation it has in the β_2 dimer rapidly (in $\sim 1.5 \text{ ns}$) converts to a conformation resembling that in the β - δ complex. Thus, the conformational change of the δ subunit’s β interaction element on binding to a β subunit is reminiscent of the action of a plumber’s wrench in unlatching the nearby β - β interface so as to allow the sliding clamp to spring open.

The X-ray structure of the $\gamma_3\delta\delta'$ complex (the clamp loader with both its τ subunits lacking τ_c ; γ and τ are interchangeable in terms of their clamp loading functions) in complex with a primer-template DNA and ADP · BeF₃ (an ATP analog), also determined by Kuriyan and O’Donnell, suggests how the clamp loader functions. The γ , δ , and δ' subunits all have similar folds; their N-terminal domains are all members of the widely distributed AAA+ family (DnaA and DnaC proteins are also members of this family) even though only the γ (and τ) subunits bind and hydrolyze ATP. The conserved regions of AAA+ proteins consist of two domains, an N-terminal ATP-binding domain and a smaller domain composed of a 3-helix bundle, whose relative orientations vary with ATP binding. The $\gamma_3\delta\delta'$ pentamer’s C-terminal domains form a ring-shaped collar (Fig. 30-34a) in which the subunits are arranged in clockwise order δ' - $\gamma 1$ - $\gamma 2$ - $\gamma 3$ - δ (Fig. 30-34b). The AAA+ domains are arranged in a right-handed spiral that tracks the minor groove of the dsDNA. Nevertheless, the clamp

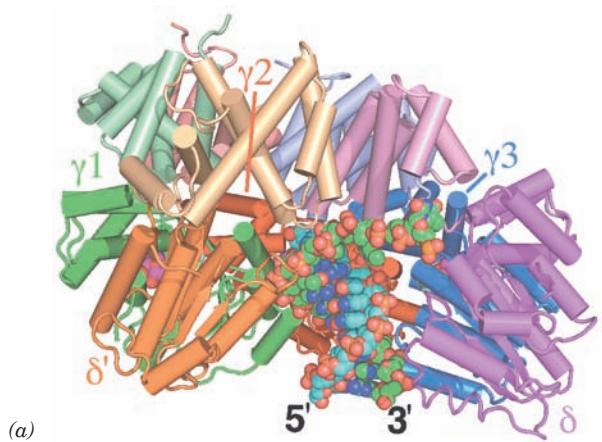
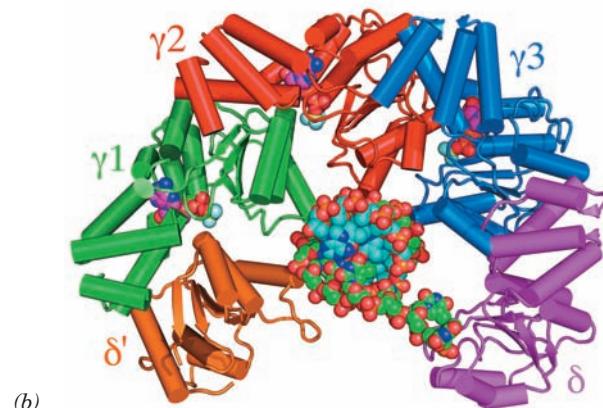


Figure 30-34 X-ray structure of the $\gamma_3\delta\delta'$ clamp loader in complex with a primer-template DNA and ADP · BeF₃. (a) View between the δ' and δ subunits approximately perpendicular to the dsDNA’s helical axis. The protein is drawn in tube-and-arrow form with its subunits colored as indicated and with the C-terminal domain of each subunit a lighter shade. The DNA consists of 10 bp with a 5' overhang of 5 nt and, together with the ADP · BeF₃, is drawn in space-filling form with primer C cyan, template C green, ADP C magenta, N blue, O red, and P orange,



Be light green, and F light blue. (b) View rotated 90° about the horizontal axis relative to Part a. The C-terminal domain of each subunit has been deleted for clarity. Note how the single-stranded portion of the template strand turns by $\sim 90^\circ$ to avoid colliding with the collar formed by the clamp loader’s C-terminal domains. The stiffness of dsDNA makes it unlikely that it could make such a turn. [Based on an X-ray structure by Mike O’Donnell, The Rockefeller University, and John Kuriyan, University of California at Berkeley. PDBid 3GLF.]

loader associates with the DNA almost entirely through contacts with the phosphate groups of the template strand alone. Thus, this structure is reminiscent of that of the E1 helicase in complex with ssDNA (Section 30-2Ca) with one of its six subunits missing.

The clamp loader must tightly bind the sliding clamp prior to its loading on the template DNA but must subsequently release the clamp to avoid interfering with its binding to the Pol III core ($\alpha\epsilon\theta$). The structures of the clamp loader and the β - δ complex, together with a variety of biochemical evidence, suggest a model of how this might occur (Fig. 30-35): The binding of ATP to $\gamma 1$ (the γ subunit that contacts δ') results in a conformational change that exposes the otherwise occluded ATP-binding site of $\gamma 2$; ATP binding to $\gamma 2$ likewise exposes $\gamma 3$; and ATP binding to $\gamma 3$ exposes the δ subunit's β interaction element, thereby permitting it to bind to a β subunit so as to spring open the sliding clamp. Primer-template DNA then inserts itself through the resulting gap in the sliding clamp. This process is facilitated by the gap between the AAA+ domains of δ and δ' subunits, which permits the clamp loader to track the template strand while avoiding contact with the primer strand. Eventually, β - and DNA-stimulated hydrolysis of the bound ATPs releases the β subunit from the clamp loader, whereon the sliding clamp closes around the DNA.

The departure of the clamp loader permits the Pol III core to bind to the sliding clamp. However, when the synthesis of an Okazaki fragment has been completed, the Pol III core must dissociate from the sliding clamp so that it can initiate the synthesis of the next Okazaki fragment. How does this occur?

Pol III's α subunit binds to the same hydrophobic pocket on the β_2 sliding clamp as does the δ subunit. This was shown by the observations that the phosphorylation of a kinase recognition sequence that had been engineered into the C-terminal segment of β is inhibited by both α and δ . The β subunit has an ~30-fold greater affinity for the γ complex in the presence of ATP than it has for the Pol III core. However, when primer-template DNA is also present, this order of affinity is reversed with β preferring to bind to the Pol III core (possibly due to the additional con-

tacts between the core and the DNA). Thus once the sliding clamp has been loaded onto the primer-template, the clamp loader is replaced by the Pol III core, which thereby blocks the clamp loader from unloading the clamp. Instead, the clamp loader loads a new clamp onto the lagging strand template in association with the primer that the primosome had synthesized in preparation for the next round of Okazaki fragment synthesis (Fig. 30-28b).

The X-ray structure of the sliding clamp in complex with primer-template DNA (Fig. 30-13) indicates that its ssDNA segment binds to the same site as do the α and δ subunits. This may serve to attract the primer-template DNA to the inside of the open clamp, which in turn may facilitate the release of the clamp loader and hence the closure of the sliding clamp. The binding of ssDNA to the sliding clamp may also prevent it from sliding away before it can be bound by the α subunit.

When the Pol III core has completed its synthesis of the Okazaki fragment, that is, when the gap between the two successively synthesized Okazaki fragments has been reduced to a nick, it releases the DNA and the sliding clamp. The Pol III core then binds to the newly primed template and its associated clamp (displacing the clamp loader), where it commences the synthesis of the next Okazaki fragment. Thus, a series of switches that are activated by ATP and DNA structure ensure the vectorial progression of lagging strand replication. Throughout this process, the Pol III holoenzyme is held at the replication fork by the leading strand Pol III core, which remains tethered to the DNA by its associated sliding clamp.

The sliding clamp that remains around the completed Okazaki fragment probably functions to recruit Pol I and DNA ligase so as to replace the RNA primer on the previously synthesized Okazaki fragment with DNA and seal the remaining nick. However, the sliding clamp must eventually be recycled. It was initially assumed that this was the job of the clamp loader. However, it is now clear that the release of the sliding clamp from its associated DNA is largely carried out by free δ subunit (the "wrench" in the clamp loader that cracks apart the β subunits forming the sliding clamp), which is synthesized in 5-fold excess over that required to populate the cell's few clamp loaders.

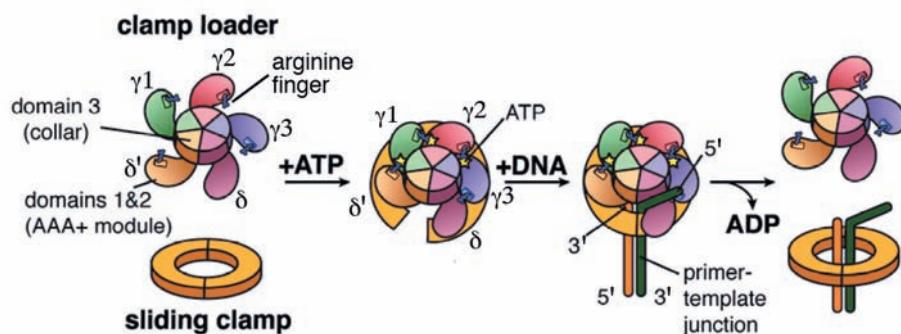


Figure 30-35 Schematic diagram of the clamp loader cycle.

This speculative model is based on a combination of structural and biochemical information. The "arginine finger," an Arg side chain that interacts with the γ -phosphate group of an ATP bound

to a neighboring subunit, is a common feature of AAA+ ATPases that form ringlike structures. [Modified from a drawing by Mike O'Donnell, The Rockefeller University, and John Kuriyan, University of California at Berkeley.]

d. Replication Termination Is Facilitated by Tus Protein

The *E. coli* replication terminus is a large (350 kb) region flanked by ten nearly identical nonpalindromic ~23-bp terminator sites, **TerH**, **TerI**, **TerE**, **TerD**, and **TerA** on one side and **TerJ**, **TerG**, **TerF**, **TerB**, and **TerC** on the other (Fig. 30-36; note that *oriC* is directly opposite the terminus region on the *E. coli* chromosome). A replication fork traveling counterclockwise as drawn in Fig. 30-36 passes through *TerJ*, *TerG*, *TerF*, *TerB*, and *TerC* but stops on encountering either *TerA*, *TerD*, *TerE*, *TerI*, or *TerH* (*TerD*, *TerE*, *TerI*, and *TerH* are presumably backup sites for *TerA*). Similarly, a clockwise-traveling replication fork transits *TerH*, *TerI*, *TerE*, *TerD*, and *TerA* but halts at *TerC* or, failing that, *TerB* or *TerF* or *TerG* or *TerI*. Thus, these termination sites act as one-way valves that allow replication forks to enter the termination region but not to leave it. This arrangement guarantees that the two replication forks generated by bidirectional initiation at *oriC* will meet in the replication terminus even if one of them arrives there well ahead of its counterpart.

The arrest of replication fork motion at *Ter* sites requires the action of **Tus** protein, a 309-residue monomer that is the product of the **tus** gene (for *terminator utilization substance*). Tus specifically binds to a *Ter* site, where it prevents strand displacement by DnaB helicase, thereby arresting replication fork motion. The X-ray structure of Tus in complex with a 15-bp *Ter* sequence-containing DNA with a single T overhang at each 5' end, determined by Kosuke Morikawa, reveals that Tus consists of two domains that

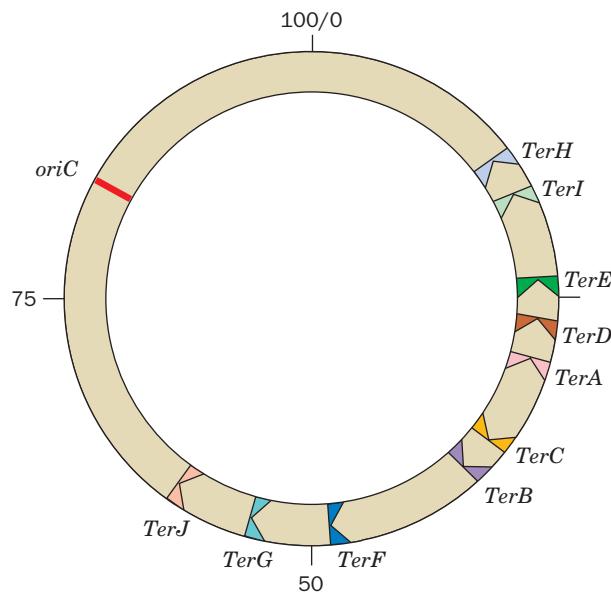


Figure 30-36 Map of the *E. coli* chromosome showing the positions of the *Ter* sites and the *oriC* site. The *TerJ*, *TerG*, *TerF*, *TerB*, and *TerC* sites, in combination with Tus protein, allow a counterclockwise-moving replisome to pass but not a clockwise-moving replisome. The opposite is true of the *TerH*, *TerI*, *TerE*, *TerD*, and *TerA* sites. Consequently, two replication forks that initiate bidirectional DNA replication at *oriC* will meet between the oppositely facing *Ter* sites.

form a deep positively charged cleft that largely envelops the bound DNA (Fig. 30-37). A 5-bp segment of the DNA near the side of Tus that permits the passage of the replication fork (the lower side of Fig. 30-37) is deformed and underwound relative to canonical (ideal) B-DNA such that its major groove becomes deeper and its minor groove is significantly expanded. The protein makes polar contacts with more than two-thirds of the phosphate groups in a 13-bp region and its interdomain β sheet penetrates the deepened major groove to make sequence-specific contacts with the exposed bases. The importance of this interdomain region for Tus function is demonstrated by the observation that most single residue mutations that reduce the ability of Tus to arrest replication occur in this interdomain region.

When Tus is fused to another DNA-binding protein, replication is inhibited at the other protein's binding site. This suggests that Tus does not act as a simple DNA-binding clamp, but interacts with DnaB helicase, the leading component of a replication fork (Fig. 30-32), to inhibit its helicase action. Apparently, Tus prevents the progress of DnaB in unwinding DNA from one side of Tus but not the other. Indeed, the encounter of DnaB with a Tus–Ter complex in the permissive direction causes Tus to rapidly dissociate from the DNA, whereas such an encounter from the nonpermissive direction generates a so-called locked Tus–Ter complex. Nevertheless, the way Tus and DnaB interact is unknown. Curiously, however, this termination system is not essential for termination. When the replication

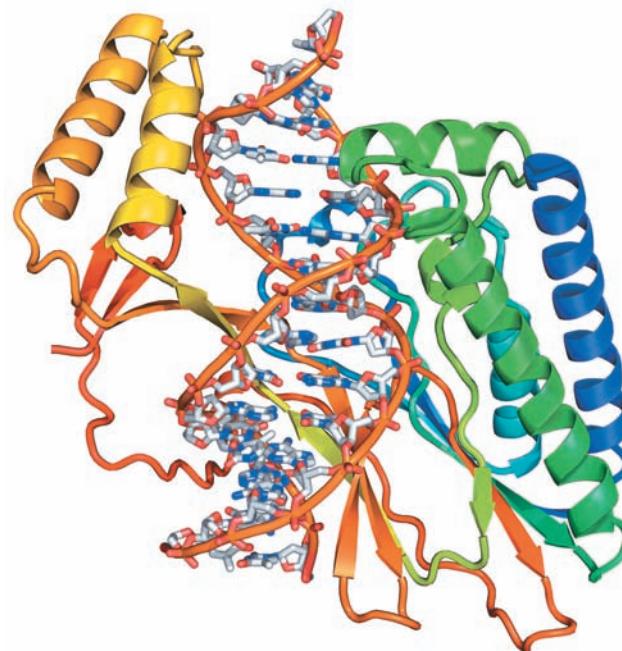


Figure 30-37 X-ray structure of *E. coli* Tus protein in complex with a 15-bp *Ter*-containing DNA. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA is shown in stick form with C gray, N blue, O red, and P orange and with successive P atoms in the same strand joined by orange rods. [Based on an X-ray structure by Kosuke Morikawa, Protein Engineering Research Institute, Osaka, Japan. PDBid 1ECR.] See Interactive Exercise 34

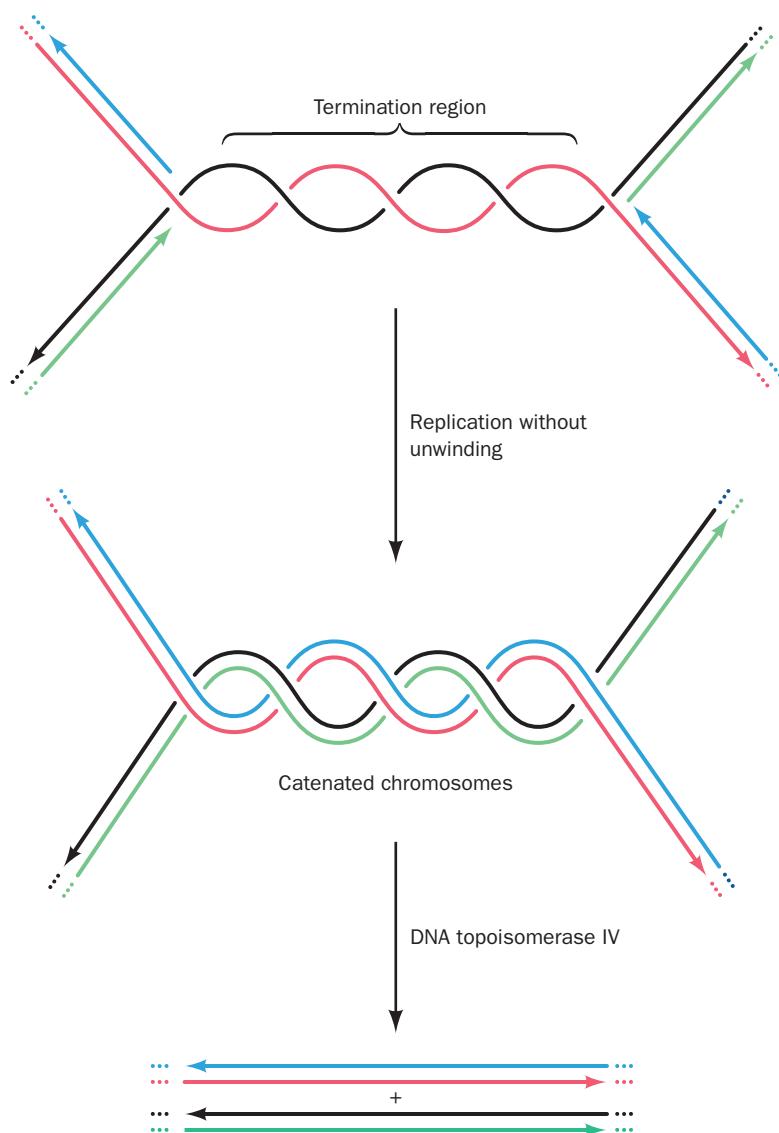


Figure 30-38 The formation and separation of catenated dsDNAs at the replication termination site. The parental strands are red and black and the daughter strands are green and blue.

For clarity, the double helical character of the newly formed dsDNA molecules is not shown.

terminus is deleted, replication simply stops, apparently through the collision of opposing replication forks. Nevertheless, this termination system is highly conserved in gram-negative bacteria.

As two oppositely moving replication forks collide at the termination site, the newly synthesized strands become covalently linked to yield two covalently closed double-stranded chromosomes. However, since the parental DNA strands remain wound about each other by several turns (presumably, DNA gyrase cannot gain access to the DNA when the colliding replication forks closely approach each other), the product dsDNA strands must be wound about each other by the same number of turns (Fig. 30-38). The resulting catenated circular dsDNAs must be separated so that each can be passed to a different daughter cell. This is the job of the type II topoisomerase named topoisomerase IV (Section 29-3Cd).

D. Fidelity of Replication

Since a single polypeptide as small as the Pol I Klenow fragment can replicate DNA by itself, why does *E. coli* maintain a battery of >20 intricately coordinated proteins to replicate its chromosome? The answer apparently is to ensure the nearly perfect fidelity of DNA replication required to preserve the genetic message's integrity from generation to generation.

The rates of reversion of mutant *E. coli* or T4 phage to the wild type indicates that only one mispairing occurs per 10^8 to 10^{10} base pairs replicated. This corresponds to ~ 1 error per 1000 bacteria per generation. Such high replication accuracy arises from four sources:

1. Cells maintain balanced levels of dNTPs through the mechanism discussed in Section 28-3Ad. This is an important aspect of replication fidelity because a dNTP present

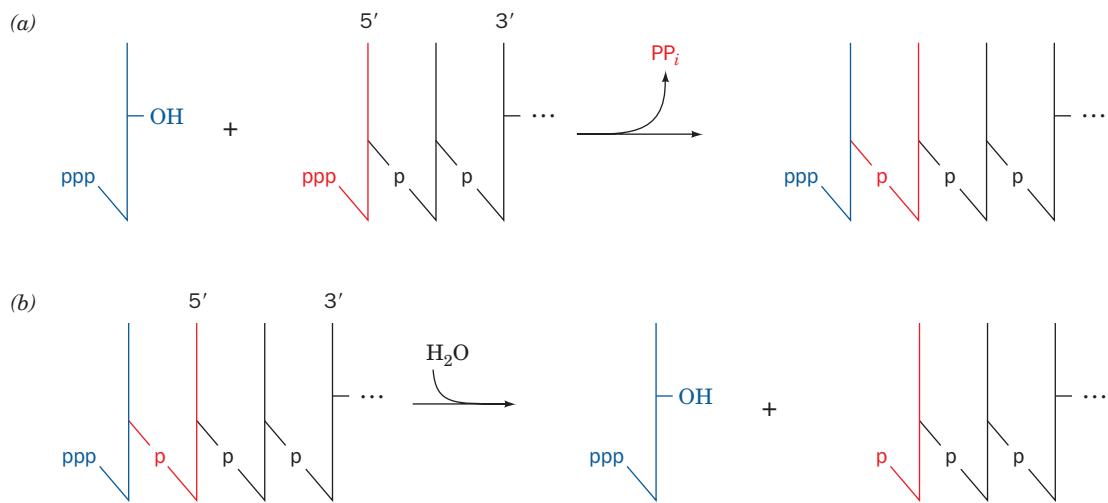


Figure 30-39 Chemical consequences if a DNA polymerase could synthesize DNA in its 3' → 5' direction. (a) The coupling of each nucleoside triphosphate to the growing chain would be driven by the hydrolysis of the previously appended nucleoside triphosphate. (b) The editorial removal of an incorrect 5'-terminal nucleoside triphosphate would render the DNA chain incapable of further extension.

at aberrantly high levels is more likely to be misincorporated and, conversely, one present at low levels is more likely to be replaced by the dNTPs present at higher levels.

2. The polymerase reaction itself has extraordinary fidelity. This is because, as we have seen (Section 30-2Ae), the polymerase reaction occurs in two stages: (1) a binding step in which the incoming dNTP base-pairs with the template while the enzyme is in an open conformation that cannot catalyze the polymerase reaction; and (2) a catalysis step in which the polymerase forms a closed conformation about the newly formed base pair, which properly positions its catalytic residues (induced fit). Since the formation of the closed conformation requires that the incoming dNTP form a Watson–Crick-shaped base pair with the template, the conformation change constitutes a double check for correct base pairing.

3. The 3' → 5' exonuclease functions of Pol I and Pol III detect and eliminate the occasional errors made by their polymerase functions. In fact, mutations that increase a DNA polymerase's proofreading exonuclease activity decrease the rates of mutation of other genes.

4. A remarkable battery of enzyme systems, contained in all cells, function to repair residual errors in the newly synthesized DNA as well as any damage that it may incur after its synthesis through chemical and/or physical insults. We discuss these DNA repair systems in Section 30-5.

In addition, the inability of a DNA polymerase to initiate chain elongation without a primer is a feature that increases DNA replication fidelity. The first few nucleotides of a chain to be coupled together are those most likely to be mispaired because of the cooperative nature of base pairing interactions (Section 29-2). The editing of a short duplex oligonucleotide is similarly an error-prone process. The use of RNA primers eliminates this source of error since the

RNA is eventually replaced by DNA under conditions that permit accurate base pairing to be achieved.

One might wonder why cells have evolved the complex system of discontinuous lagging strand synthesis rather than a DNA polymerase that could simply extend DNA chains in their 3' → 5' direction. Consideration of the chemistry of DNA chain extension also leads to the conclusion that this system promotes high-fidelity replication. The linking of 5'-deoxyribonucleotide triphosphates in the 3' → 5' direction would require the retention of the growing chain's 5'-terminal triphosphate group to drive the next coupling step (Fig. 30-39a). On editing a mispaired 5'-terminal nucleotide (Fig. 30-39b), this putative polymerase would—in analogy with Pol I, for example—excise the offending nucleotide, leaving either a 5'-OH or a 5'-phosphate group. Neither of these terminal groups is capable of energizing further chain extension. A proofreading 3' → 5' DNA polymerase would therefore have to be capable of reactivating its edited product. The inherent complexity of such a system has presumably selected against its evolution.

4 EUKARYOTIC REPLICATION

There is a remarkable degree of similarity between eukaryotic and prokaryotic DNA replication mechanisms. Nevertheless, there are important differences between these two replication systems as a consequence of the vastly greater complexity of eukaryotes in comparison to prokaryotes. For example, eukaryotic chromosomes are structurally complicated and dynamic complexes of DNA and protein (Section 34-1) with which the replication machinery must interact in carrying out its function. Consequently, as is true of most aspects of biochemistry, our knowledge of how DNA is replicated in eukaryotes has lagged well behind that for prokaryotes, although in recent years there has

been significant progress in our understanding of this essential process. In this section, we outline what is known about DNA replication in eukaryotes. We also discuss two DNA polymerases that are peculiar to eukaryotic systems: reverse transcriptase and telomerase.

A. The Cell Cycle

The **cell cycle**, the general sequence of events that occur during the lifetime of a eukaryotic cell, is divided into four distinct phases (Fig. 30-40):

1. Mitosis and cell division occur during the relatively brief **M phase** (for mitosis).
2. This is followed by the **G₁ phase** (for gap), which covers the longest part of the cell cycle. This is the main period of cell growth.
3. G₁ gives way to the **S phase** (for synthesis), which in contrast to events in prokaryotes, *is the only period in the cell cycle when DNA is synthesized*.
4. During the relatively short **G₂ phase**, the now tetraploid cell prepares for mitosis. It then enters M phase once again and thereby commences a new round of the cell cycle.

The cell cycle for cells in culture typically occupies a 16- to 24-h period. In contrast, cell cycle times for the different types of cells of a multicellular organism may vary from as little as 8 h to >100 days. Most of this variation occurs in the G₁ phase. Moreover, many terminally differentiated cells, such as neurons or muscle cells, never divide; they assume a quiescent state known as the **G₀ phase**.

A cell's irreversible "decision" to proliferate is made during G₁. Quiescence is maintained if, for example, nutrients are in short supply or the cell is in contact with other cells (**contact inhibition**). Conversely, DNA synthesis may be induced by various agents such as carcinogens or tumor viruses, which trigger uncontrolled cell proliferation (can-

cer; Sections 19-3B and 34-4C); by the surgical removal of a tissue, which results in its rapid regeneration; or by proteins known as **mitogens**, which bind to cell-surface receptors and induce cell division (Section 34-4D).

a. The Cell Cycle Is Controlled by Cyclins and Cyclin-Dependent Protein Kinases

The progression of a cell through the cell cycle is regulated by proteins known as **cyclins** and **cyclin-dependent protein kinases (Cdks)**. Cyclins are so named because they are synthesized during one phase of the cell cycle and completely degraded during a succeeding phase (protein degradation is discussed in Section 32-6). A particular cyclin specifically binds to and thereby activates its corresponding Cdk(s) to phosphorylate its target proteins, thus activating these proteins to carry out the processes comprising that phase of the cell cycle. In order to enter a new phase in the cell cycle, a cell must satisfy a corresponding **checkpoint**, which monitors whether the cell has satisfactorily completed the preceding phase [e.g., the attachment of all chromosomes to the mitotic spindle must precede mitosis (Section 1-4Aa); if this were not the case for even one chromosome, one daughter cell would lack this chromosome and the other would have two, both deleterious if not lethal conditions]. If the cell has not met the criteria of the checkpoint, the cell cycle is slowed or even arrested until it does so. We further discuss cell cycle control in Section 34-4C.

B. Eukaryotic Replication Mechanisms

Much of what we know about eukaryotic DNA replication has been learned from studies on budding yeast (*Saccharomyces cerevisiae*) and fission yeast (*Schizosaccharomyces pombe*), the simplest eukaryotes, and on simian virus 40 (SV40), which has a 5243-bp circular DNA chromosome that has only one replication origin. However, studies of DNA replication in the cells of **metazoa** (multicellular animals), particularly *Drosophila*, *Xenopus laevis* (an African clawed toad, whose eggs are easily studied), and humans, have also led to important advances in our knowledge.

a. Eukaryotic Cells Contain Numerous DNA Polymerases

The many known DNA polymerases can be classified into six families based on phylogenetic relationships. Members of families A (e.g., *E. coli* Pol I), B (e.g., *E. coli* Pol II), and C (e.g., *E. coli* Pol III) encompass all replicative polymerases as well as some repair polymerases, family D occurs only in archaea where its functions are poorly understood, and families X and Y participate in DNA repair. The fingers and thumb domains have structures that are unique to each family, whereas the catalytic residue-containing palm domains are similar in families A, B, and Y. Animal cells express at least four distinct types of DNA polymerases that are implicated in DNA replication (Table 30-5). They are designated, in the order of their discovery, DNA polymerases (pol) α , γ , δ , and ϵ (alternatively, POLA, POLG, POLD1, and POLE), of which pol γ is a member of family A and the others are members of family B.

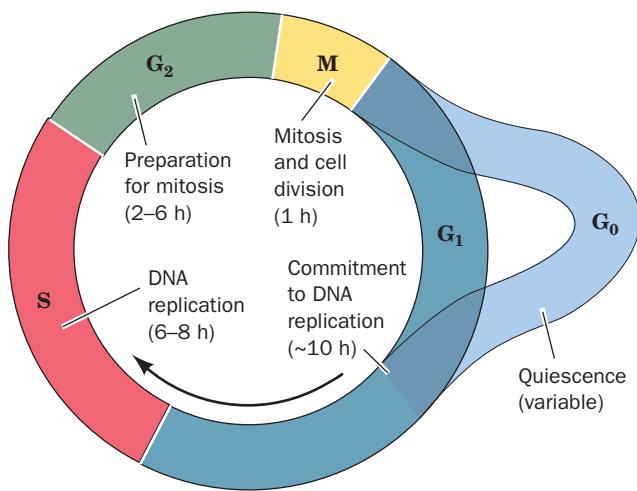


Figure 30-40 The eukaryotic cell cycle. Cells in G₁ may enter a quiescent phase (G₀) rather than continuing about the cycle.

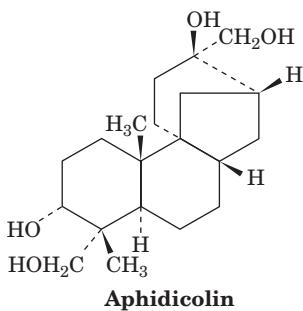
Table 30-5 Properties of Eukaryotic DNA Polymerases That Participate in DNA Replication

	α	γ	δ	ε
Location	Nucleus	Mitochondrion	Nucleus	Nucleus
Subunit masses (kD) ^a	167, 79, 62, 48 (166, 66, 59, 50)	144 (140, 55)	125, 55, 40 (124, 51, 51)	256, 78, 23, 22 (262, 60, 17, 12)
Family	B	A	B	B

^aYeast *S. cerevisiae* (human cells).

Source: Mainly Johnson, A. and O'Donnell, M., *Annu. Rev. Biochem.* **74**, 283 (2005).

Pol α occurs only in the cell nucleus where it participates in the replication of chromosomal DNA. This function was largely established through the use of its specific inhibitor **aphidicolin**.



and by the observation that pol α activity varies with the rate of cellular proliferation. Pol α , as do all DNA polymerases, replicates DNA by extending a primer 5' \rightarrow 3' under the direction of a single-stranded DNA template. This heterotetramer, which lacks exonuclease activity, consists of a 167-kD polymerase subunit, a 48-kD primase subunit, a 62-kD subunit that is required for full primase activity, and a 79-kD subunit that is implicated in the regulation of initiation, all of which are collectively known as **pol α /primase**.

Pol δ is a heterotrimer whose 125-kD catalytic subunit lacks an associated primase but contains a proofreading 3' → 5' exonuclease domain. The X-ray structure of the yeast pol δ catalytic subunit (also called pol δ), determined by Aneel Aggarwal, reveals that this enzyme consists of five domains arranged around a central hole that is near its polymerase active site (Fig. 30-41). It has the right-hand-like architecture first seen in A-family DNA polymerases (Figs. 30-8 and 30-9), and its palm domain has a structurally similar core that contains the two invariant Asp residues implicated in the nucleotidyl transfer mechanism (Fig. 30-10). However, there are major differences between A-family polymerases (e.g., Fig. 30-8) and pol δ, which is representative of B-family polymerases. Most notably, in pol δ:

1. The fingers domain, which consists of only a pair of antiparallel helices, is rotated by $\sim 60^\circ$ relative to that in A-family polymerases.
 2. The exonuclease domain projects from the top of the fingers domain rather than from the bottom of the palm domain, as it does in A-family polymerases.

3. There is a large N-terminal domain that A-family polymerases lack.

4. The template strand enters the active site from a cleft between the N-terminal and exonuclease domains, whereas in A-family polymerases, it does so from the fingers domain.

5. The newly formed dsDNA nearest the active site has a B-DNA-like conformation instead of the A-DNA-like conformation observed in A-family polymerases (Section 30-2Ae).

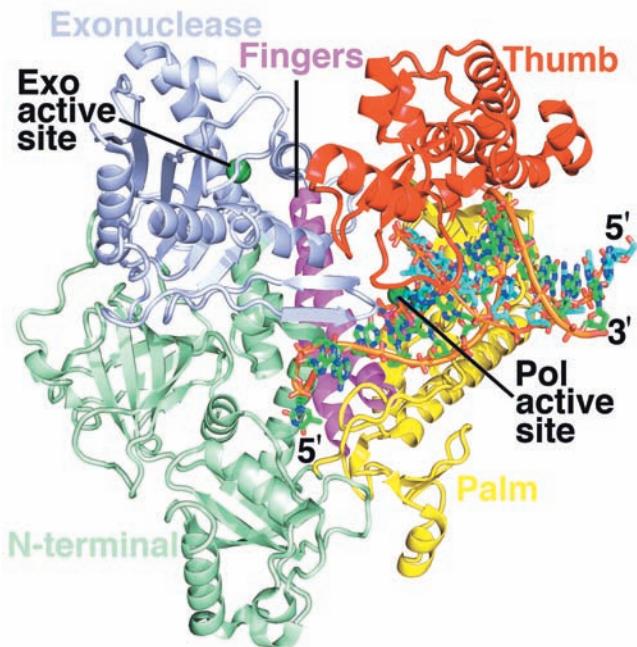


Figure 30-41 X-ray structure of yeast DNA polymerase δ (pol δ) in complex with primer-template DNA and dCTP. The protein is drawn in ribbon form with its five domains differently colored as indicated. The DNA, whose primer and template strands consist of 12 and 16 nt, together with the incoming dCTP, is drawn in stick form with template C green, primer C cyan, dCTP C magenta, N blue, O red, and P orange and with successive P atoms in each DNA strand connected by orange rods. The Ca^{2+} ions at the polymerase (Pol) active site are represented by dark green spheres as is the Ca^{2+} ion at the exonuclease (Exo) active site. [Based on an X-ray structure by Aneel Aggarwal, Mount Sinai School of Medicine, New York, New York. PDBid 3IAY.]

In contrast to pol α , which exhibits only moderate processivity (\sim 100 nucleotides), that of pol δ is essentially unlimited (replicates the entire length of a template), but only when it is in complex with a protein named **proliferating cell nuclear antigen (PCNA)**; so named because it occurs only in the nuclei of proliferating cells and reacts with antibodies produced by a subset of patients with the autoimmune disease systemic lupus erythematosus). The X-ray structure of PCNA (Fig. 30-42), determined by Kuriyan, reveals that it forms a trimeric ring with almost identical structure (and presumably function) as the *E. coli* β_2 sliding clamp (Fig. 30-13). Thus, each PCNA subunit consists of four rather than six of the structurally similar $\beta\alpha\beta\beta\beta$ motifs from which the *E. coli* β subunit is constructed. Intriguingly, PCNA and the β subunit exhibit no significant sequence identity, even when their structurally similar portions are aligned. Archaea also have sliding clamps with pseudohexagonal symmetry.

Pol δ in complex with PCNA is required for lagging strand synthesis. In contrast, pol α /primase functions to synthesize \sim 12-nt RNA primers, which it extends by an additional \sim 20 nt of DNA. Then, in a process called **polymerase switching**, the eukaryotic counterpart of the *E. coli* clamp loader (Section 30-3C), **replication factor C (RFC)**, displaces the pol α and loads PCNA on the template DNA

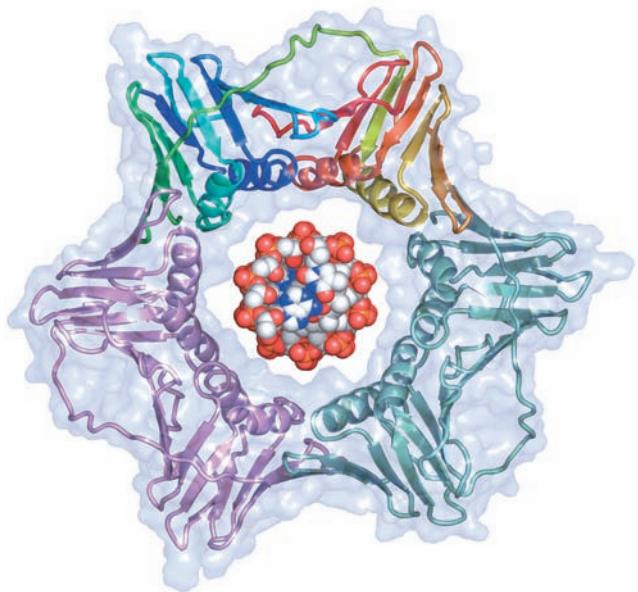


Figure 30-42 X-ray structure of human PCNA. Its three subunits, which form a 3-fold symmetric ring, are drawn in ribbon form embedded in their semitransparent surface diagram. One of these subunits is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), another is pink, and the third is light green. A space-filling model of B-DNA viewed along its helix axis has been drawn in the center of the PCNA ring. Compare this structure with that of the β_2 sliding clamp of the *E. coli* Pol III holoenzyme (Fig. 30-13). [Based on an X-structure by John Kuriyan, University of California at Berkeley. PDBid 1AXC.]  See Interactive Exercise 35

near the primer strand, following which pol δ binds to the PCNA and processively extends the DNA strand.

RFC, like the *E. coli* clamp loader, is a heteropentamer of AAA+ family subunits, but whereas the clamp loader contains three identical subunits ($\gamma_3\delta\delta'$), RFC consists of five different subunits, RFC-A through RFC-E (alternatively, RFC1–RFC5). The X-ray structure of yeast RFC in complex with PCNA (Fig. 30-43), determined by O'Donnell and Kuriyan, reveals that RFC's A, B, and C subunits bind in conserved hydrophobic grooves on the face of PCNA on which its C-terminal residues are located (since PCNA's three identical subunits are linked head-to-tail, its two faces are different). In fact, this so-called C side of PCNA, which faces away from the direction of polymerase motion, binds many of the proteins that participate in replicative processes, including most DNA polymerases, and hence PCNA plays a major role in

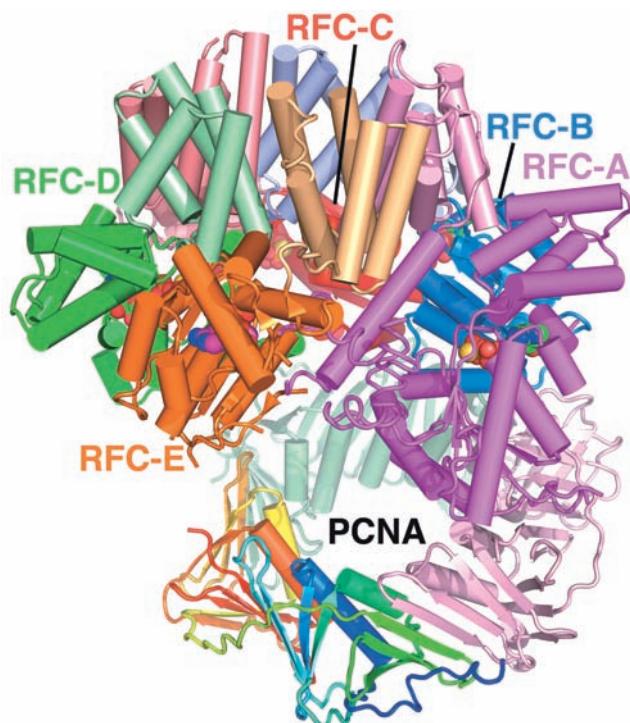


Figure 30-43 X-ray structure of yeast RFC in complex with PCNA, ADP, and ATP γ S. Both proteins are drawn in tube-and-arrow form with the RFC subunits colored as their homologs in Fig. 30-34a (with the RFC-A, B, C, D, and E subunits corresponding to the *E. coli* clamp loader's δ , γ_3 , γ_2 , γ_1 , and δ' subunits, respectively) and the PCNA colored as in Fig. 30-42. The view is similar to that of the *E. coli* clamp loader in Fig. 30-34a. ADP, which binds to only RFC-E, and ATP γ S, which binds to the other four subunits, are shown in space-filling form with ATP γ S C green, ADP C magenta, N blue, O red, P orange, and S yellow. Note that all five RFC subunits have a nucleotide binding site, whereas the *E. coli* δ and δ' subunits do not. [Based on an X-ray structure by Mike O'Donnell, The Rockefeller University, and John Kuriyan, University of California at Berkeley. PDBid 1SXJ.]

recruiting the components of the various types of replication forks. The C-terminal domains of each RFC subunit associate to form a ring-shaped collar, as do the C-terminal domains of the *E. coli* clamp loader (Fig. 30-34). Likewise, RFC's AAA+ domains are arranged in a right-handed spiral that matches the helical path of the sugar-phosphate backbone of B-DNA. Presumably, prokaryotic and eukaryotic clamp loaders interact with primer-template DNA and their corresponding sliding clamps in similar ways.

Pol ε, a heterotetrameric nuclear enzyme, is the most enigmatic participant in DNA replication. Pol ε is highly processive in the absence of PCNA and has a 3' → 5' exonuclease activity that degrades single-stranded DNA to 6- or 7-residue oligonucleotides rather than to mononucleotides, as does that of pol δ. Although pol ε is necessary for the viability of yeast, its essential function can be carried out by only the noncatalytic C-terminal half of its 256-kD catalytic subunit, which is unique among B-family DNA polymerases. This suggests that the C-terminal half of the pol ε catalytic subunit is required for the assembly of the replication complex. Nevertheless, Thomas Kunkel has shown that pol ε is probably the leading strand replicase, although it may also contribute to lagging strand synthesis. Moreover, pol δ may also participate in leading strand synthesis.

Pol γ, a monomer, occurs exclusively in the mitochondrion, where it presumably replicates the mitochondrial DNA. Chloroplasts contain a similar enzyme.

Eukaryotic cells contain batteries of DNA polymerases. These include the DNA polymerases that participate in chromosomal DNA replication (pol α, δ, and ε) and several that take part in DNA repair processes (Section 30-5) including **pol β, η, ι, κ**, and **ζ** (alternatively, POLB, POLH, POLI, POLK, and POLZ). Pol β, an X-family enzyme, is remarkable for its small size (a 335-residue monomer in humans).

b. Eukaryotic Chromosomes Consist of Numerous Replicons

Eukaryotic and prokaryotic DNA replication systems differ most obviously in that eukaryotic chromosomes have multiple replication origins in contrast to the single replication origin of prokaryotic chromosomes. Eukaryotic cells replicate DNA at the rate of ~50 nt/s (~20 times slower than does *E. coli*) as was determined by autoradiographically measuring the lengths of pulse-labeled sections of eukaryotic chromosomes. Since a eukaryotic chromosome typically contains 60 times more DNA than those of prokaryotes, its bidirectional replication from a single origin would require ~1 month to complete. Electron micrographs such as Fig. 30-44, however, reveal that eukaryotic chromosomes contain multiple origins, one every 3 to 300 kb depending on both the species and the tissue, so that S phase usually occupies only a few hours.

Cytological observations indicate that the various chromosomal regions are not all replicated simultaneously; rather, clusters of 20 to 80 adjacent **replicons**

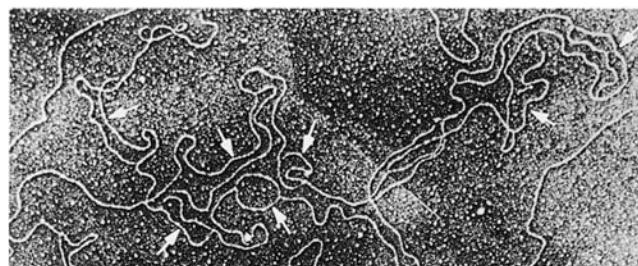


Figure 30-44 Electron micrograph of a fragment of replicating *Drosophila* DNA. The arrows indicate its multiple replication eyes. [From Kriegstein, H.J. and Hogness, D.S., *Proc. Natl. Acad. Sci.* **71**, 136 (1974).]

(replication units; DNA segments that are each served by a replication origin) are activated simultaneously. New replicons are activated throughout S phase until the entire chromosome has been replicated. During this process, replicons that have already been replicated are distinguished from those that have not; that is, *a cell's chromosomal DNA is replicated once and only once per cell cycle*.

c. The Assembly of the Eukaryotic Initiation Complex Occurs in Two Stages

The once-and-only-once replication of eukaryotic DNA per cell cycle is conferred by a type of binary switch. A **pre-replicative complex (pre-RC)** is assembled at each replication origin during the G₁ phase of the cell cycle. This is the only period of the cell cycle during which the pre-RC can form and hence this process is known as **licensing**. However, a licensed pre-RC cannot initiate DNA replication. Rather, it must be activated to do so, a process that occurs only during S phase. *This temporal separation of pre-RC assembly and origin activation ensures that a new pre-RC cannot assemble on an origin that has already "fired" (commenced replication) so that an origin can only fire once per cell cycle.* How does this occur?

The licensing process and how the pre-RC is activated to form an initiation complex are still incompletely understood. Thus, although it appears that most of the proteins forming these complexes have been identified, their structures, interactions, and, in many cases, their functions are largely unknown. Keeping this in mind, let us consider what is known about these processes.

Replication origins are surprisingly variable among species, often within the same organism, and even vary with a given organism's developmental stage. Thus, whereas *S. cerevisiae* origins, which are known as **autonomously replicating sequences (ARS)**, contain a highly conserved 11-bp AT-rich sequence within a less well defined ~125-bp region, some metazoan origins are dispersed over 10 to 50 kb "initiation zones" that contain multiple origins and, in some cases, require no specific DNA sequence at all. Despite this disparity, the proteins that participate in eukaryotic DNA replication are highly conserved from yeast to humans.

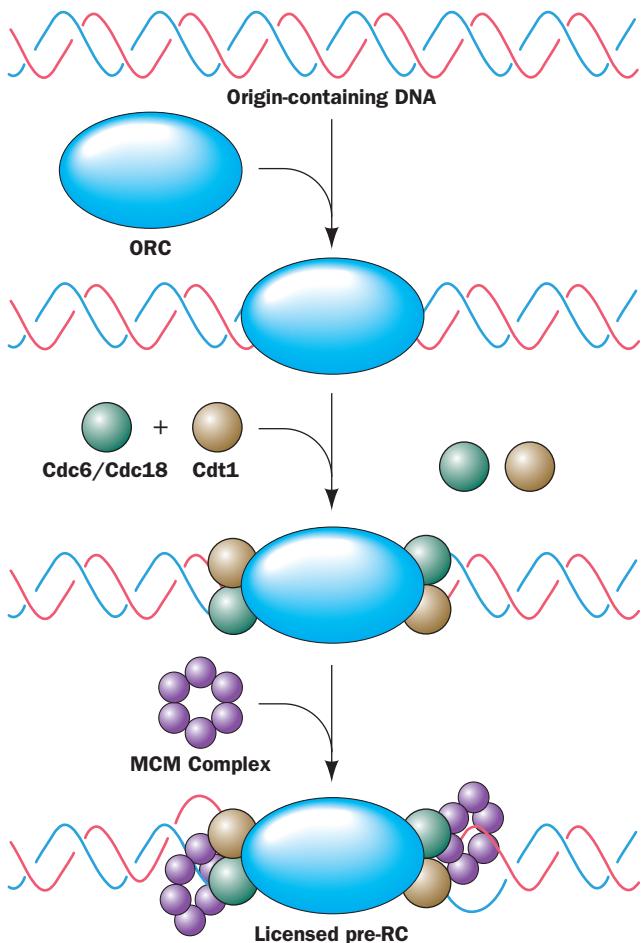


Figure 30-45 Schematic diagram for the assembly of the eukaryotic pre-replicative complex (pre-RC). The actual stoichiometries, positions, and interactions of its various components are largely unknown. The pre-RC only forms during the G_1 phase of the cell cycle.

The assembly of the pre-RC (Fig. 30-45) begins late in M phase or early in G_1 phase with the binding of the **origin recognition complex (ORC)**, a hexamer of related proteins (**Orc1** through **Orc6**), to the origin, where it remains bound during most or all of the cell cycle. ORC, the functional analog of DnaA protein in *E. coli* replication initiation (Section 30-3Ca), then recruits two proteins, **Cdc6** in *S. cerevisiae* (**Cdc18** in *S. pombe*; Cdc for cell division cycle) and **Cdt1**. These proteins then cooperate with the ORC to load the **MCM complex** [named for its minichromosome (plasmid) maintenance functions], a hexamer of related subunits (**Mcm2** through **Mcm7**), onto the DNA to yield the licensed pre-RC. The MCM complex, a ring-shaped ATP-driven helicase, is the analog of *E. coli* DnaB helicase, whereas Cdc6/Cdc18 together with Cdt1 appears to be an analog of *E. coli* DnaC (which facilitates DnaB loading). With the exception of Cdt1, all of these proteins, Orc1 through Orc6, Cdc6/Cdc18, Mcm2 through Mcm7, as well as *E. coli* DnaA, DnaB, and DnaC, are AAA+ ATPases.

The conversion of a licensed pre-RC to an active initiation complex requires the addition of pol α /primase, pol ϵ , and several accessory proteins, which only occurs at the onset of S phase. This process begins with addition of **Mcm10** protein (which shares no sequence similarity with any of the subunits of the MCM complex) to the pre-RC, which probably displaces Cdt1. This is followed by the addition of at least two protein kinases, a Cdk and **Ddk**, the latter being a heterodimer of the protein kinase **Cdc7** with its activating subunit **Dbf4** (Ddk stands for *Dbf4*-dependent kinase). Ddk acts to phosphorylate five of the six MCM subunits (all but Mcm2) so as to activate the MCM complex as a helicase. In contrast, the way in which Cdks activate the pre-RC is poorly understood, although several ORC and MCM proteins as well as Cdc6/Cdc18 are phosphorylated by Cdks. Ddk together with a Cdk also recruits **Cdc45** to the growing initiation complex. Cdc45, in turn, is required for the assembly of the initiating synthetic machinery at the replication fork, including pol α /primase, pol ϵ , PCNA, and **replication protein A (RPA)**, the heterotrimeric eukaryotic counterpart of SSB, thereby forming an active initiation complex.

d. Re-Replication Is Prevented through the Actions of Cdks and Geminin

Once initiation (priming) has occurred, the initiation complex is joined by RFC and pol δ and, as is described above, is converted to an active replicative complex by polymerase switching. DNA replication then proceeds bidirectionally until each replication fork has collided with an oppositely traveling replication fork, thereby completing the replication of the replicon. An active replication fork will destroy any licensed pre-RCs and unfired initiation complexes in its path, thereby preventing the DNA at such sites from being replicated twice. Eukaryotes appear to lack termination sequences and proteins analogous to the *Ter* sites and *Tus* protein in *E. coli*.

Several redundant mechanisms ensure that a pre-RC can initiate DNA synthesis only once. Cdks are active from late G_1 phase through late M phase. These elevated Cdk levels, which are required to activate initiation, also prevent reinitiation. The Cdk-mediated phosphorylation of Cdc6/Cdc18, which occurs late in G_1 after the pre-RCs have formed, causes Cdc6/Cdc18 to be proteolytically degraded in yeast and exported from the nucleus in mammalian cells. Evidently, Cdc6/Cdc18 is only required for the assembly of the pre-RC, not its activation. The helicase activity of the MCM complex is inhibited by phosphorylation, at least *in vitro*. Moreover, MCM proteins are exported from the nucleus in G_2 and M phases, a process that is interrupted by Cdk inactivation. However, the function of Cdk-mediated phosphorylation of ORC proteins is unclear.

Metazoan cells have yet another mechanism to prevent the assembly of a licensed pre-RC on already replicated DNA. High levels of a protein named **geminin** appear in S phase and continue to accumulate until late M phase, when geminin is degraded. Geminin associates with Cdt1 (which together with Cdc6/Cdc18 loads the MCM complex onto

the ORC) so as to inhibit the assembly of the pre-RC. This inhibition can be reversed by the addition of excess Cdt1. It therefore seems likely that the presence of geminin provides protection against DNA re-replication under conditions when Cdks are inhibited by checkpoint activation. In addition, re-replication is also prevented by the degradation of Cdt1 after the origin with which it is associated has fired. The requirement for DNA replication in this process is indicated by its blockage by the DNA polymerase inhibitor aphidicolin.

Finally, cells that have shifted to the G_0 (quiescent) phase of the cell cycle (Fig. 30-40)—the majority of cells in the human body—cease making DNA. Such cells are characterized by the absence of Cdk activity. In proliferating cells, this would permit the re-replication of DNA. However, cells in G_0 also lack the proteins of the MCM complex and are therefore incapable of assembling licensed pre-RCs. Since cancerous cells are characterized by being in a state of rapid proliferation (Section 19-3B), the presence of MCM complex proteins in what should be quiescent cells is a promising diagnostic marker for cancer.

e. Primers Are Removed by RNase H1 and Flap Endonuclease-1

In lagging strand synthesis, when pol δ reaches the previously synthesized Okazaki fragment, it partially displaces its RNA primer through DNA synthesis, thereby generating an RNA flap. The primer is then removed through the actions of two enzymes: **RNase H1** removes most of the RNA leaving only a 5' ribonucleotide adjacent to the DNA, which is then removed by **flap endonuclease-1 (FEN1)**. However, as we have seen, pol α /primase extends the RNA primers it has made by ~ 20 nt of DNA before it is displaced by pol δ . Since pol α lacks proofreading ability, this primer extension is more likely to contain errors than the DNA synthesized by pol δ . However, FEN1, which is recruited to the replication fork by its binding to the C side of PCNA, provides what is, in effect, pol α 's proofreading function: It is also an endonuclease that excises mismatch-containing oligonucleotides up to 15 nt long from the 5' end of an annealed DNA strand. Moreover, FEN1 can make several such excisions in succession to remove more distant mismatches. The excised segment is later replaced by pol δ as it synthesizes the succeeding Okazaki fragment.

f. Mitochondrial DNA Is Replicated via RNA Okazaki Fragments

Mammalian mitochondria contain two to ten copies of their ~ 16 -kb circular chromosome. The pol γ -mediated replication of this chromosome occurs unidirectionally from a single origin. In this process, as Ian Holt showed, the lagging strand Okazaki fragments are entirely synthesized as RNA (Fig. 30-46). The RNA is then replaced by DNA, although the way this occurs is poorly understood. One possibility is that this RNA is excised in much the same way as the primers in the nucleus, that is, through the actions of RNase H1 and FEN1, followed by the synthesis of lagging strand DNA by pol γ .

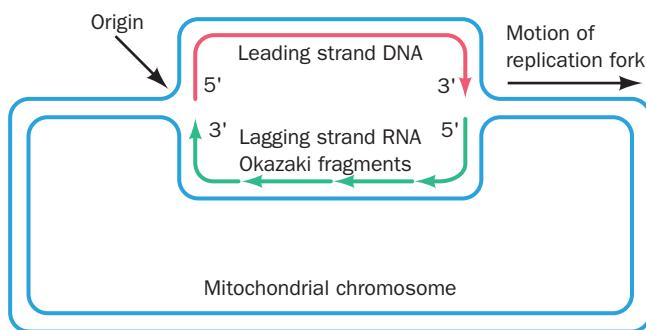


Figure 30-46 Replication of the mammalian mitochondrial chromosome. This circular chromosome is unidirectionally replicated from a single origin by a process in which the Okazaki fragments are entirely RNA.

C. Reverse Transcriptase

The **retroviruses**, which are RNA-containing eukaryotic viruses such as certain tumor viruses and human immunodeficiency virus (HIV), contain an **RNA-directed DNA polymerase (reverse transcriptase)**. This enzyme, which was independently discovered in 1970 by Howard Temin and David Baltimore, acts much like Pol I in that it synthesizes DNA in the 5' \rightarrow 3' direction from primed templates. In the case of reverse transcriptase, however, RNA is the template. The discovery of reverse transcriptase caused a mild sensation in the biochemical community because it was perceived by some as being heretical to the central dogma of molecular biology (Section 5-4). There is, however, no thermodynamic prohibition to the reverse transcriptase reaction; in fact, under certain conditions, Pol I can likewise copy RNA templates.

Reverse transcriptase transcribes the retrovirus's single-stranded RNA genome to a double-stranded DNA as follows (Fig. 30-47):

1. The retroviral RNA acts as a template for the synthesis of its complementary DNA (RNA-directed DNA polymerase activity), yielding an RNA–DNA hybrid helix. The DNA synthesis is primed by a host cell tRNA whose 3' end partially unfolds to base pair with a complementary segment of the viral RNA.

2. The RNA strand is then nucleolytically degraded (RNase H activity; H for hybrid).

3. The DNA strand acts as a template for the synthesis of its complementary DNA (DNA-directed DNA polymerase activity), yielding double-stranded DNA.

The DNA is then integrated into a host cell chromosome.

Reverse transcriptase has been a particularly useful tool in genetic engineering because of its ability to transcribe mRNAs to complementary strands of DNA (cDNA). In transcribing eukaryotic mRNAs, which have poly(A) tails (Section 31-4Ab), the primer can be oligo(dT). cDNAs have been used, for example, as probes in Southern blotting (Section 5-5D) to identify the genes coding for their

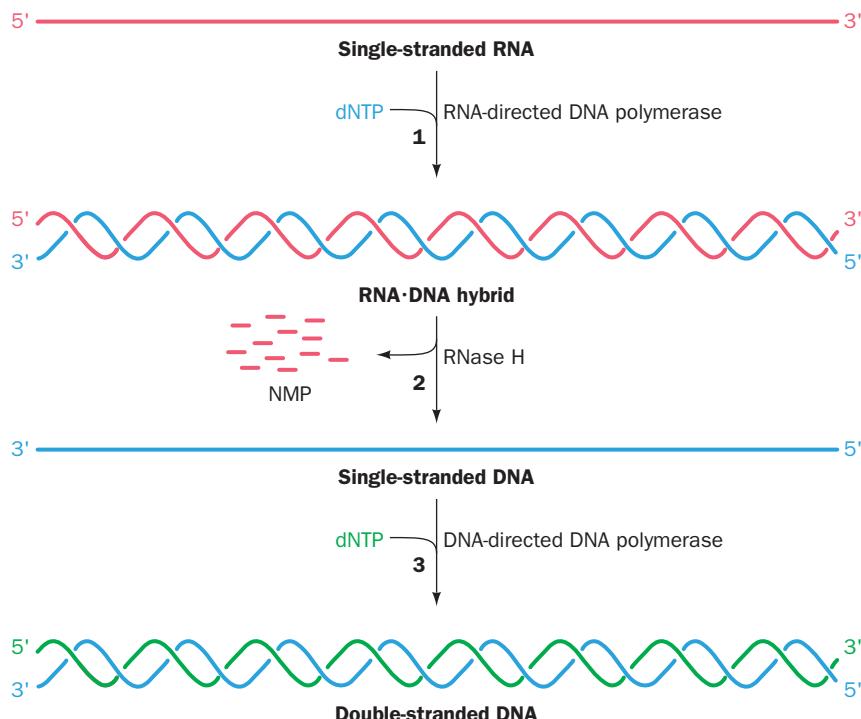


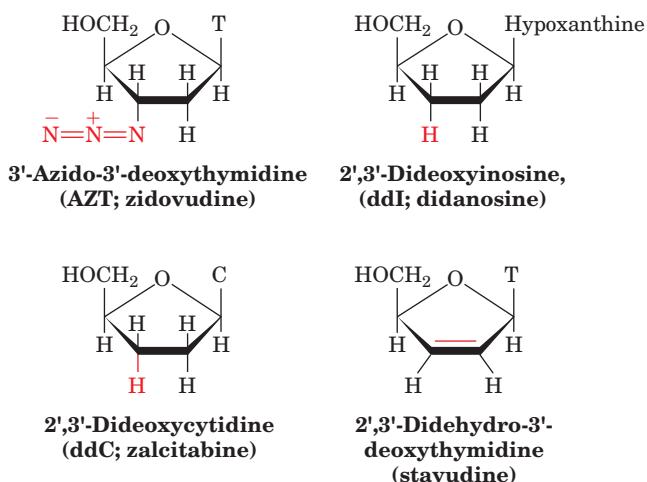
Figure 30-47 The reactions catalyzed by reverse transcriptase.

corresponding mRNAs. An RNA's base sequence can be easily determined by sequencing its cDNA (Section 7-2A).

a. X-Ray Structure of HIV-1 Reverse Transcriptase

HIV-1 reverse transcriptase (RT) is a dimeric protein whose subunits are synthesized as identical 66-kD polypeptides, known as **p66** (p for protein), that each contain a polymerase domain and an RNase H domain. However, the RNase H domain of one of the two subunits is proteolytically excised, thereby yielding a 51-kD polypeptide named **p51**. Thus, RT is dimer of p66 and p51.

The first drugs to be clinically approved to treat AIDS, **3'-azido-3'-deoxythymidine (AZT; zidovudine)**, **2',3'-dideoxyinosine (ddI; didanosine)**, **2',3'-dideoxycytidine (ddC; zalcitabine)**, and **2',3'-didehydro-3'-deoxythymidine (stavudine)**,



are RT inhibitors. Unfortunately, resistant strains of HIV-1 arise quite rapidly because RT lacks a proofreading exonuclease function and hence is highly error prone. Thus, as we have seen (Section 15-4Cd), effective long-term anti-HIV therapy requires the concurrent administration of at least one RT inhibitor and an HIV protease inhibitor.

Edward Arnold determined the X-ray structure of RT complexed to a 21-bp primer-template DNA with a 5-nt overhang at the 5' end of its template strand, a dideoxy-dG (ddG) residue at the 3' end of its primer strand (which prevents its further extension), and in complex with dATP (Fig. 30-48). The polymerase domains of p66 and p51 each contain four subdomains, which, because of their collective resemblance in p66 to DNA polymerases, are named, from N- to C-terminus, fingers, palm, thumb, and connection. Indeed, reverse transcriptases form a separate family of DNA polymerases, family RT. In p66, the RNase H domain follows the connection.

p51 has undergone a remarkable conformational change relative to p66: The connection has rotated by 155° and translated by 17 Å to bring it from a position in p66 in which it contacts only the RNase H domain (Fig. 30-48a) to one in p51 in which it contacts all three other polymerase subdomains (Fig. 30-48b). This permits p66 and p51 to bring different surfaces of their connections into juxtaposition to form, in part, RT's DNA-binding groove. Thus, the chemically identical polymerase domains of p66 and p51 are not related by 2-fold molecular symmetry (a rare but not unprecedented phenomenon), but instead, associate in a sort of head-to-head and tail-to-tail arrangement (Fig. 30-48c). Consequently, RT has only one polymerase active site and one RNase H active site. This is an example of viral genetic economy: HIV-1, with its limited genome size, has

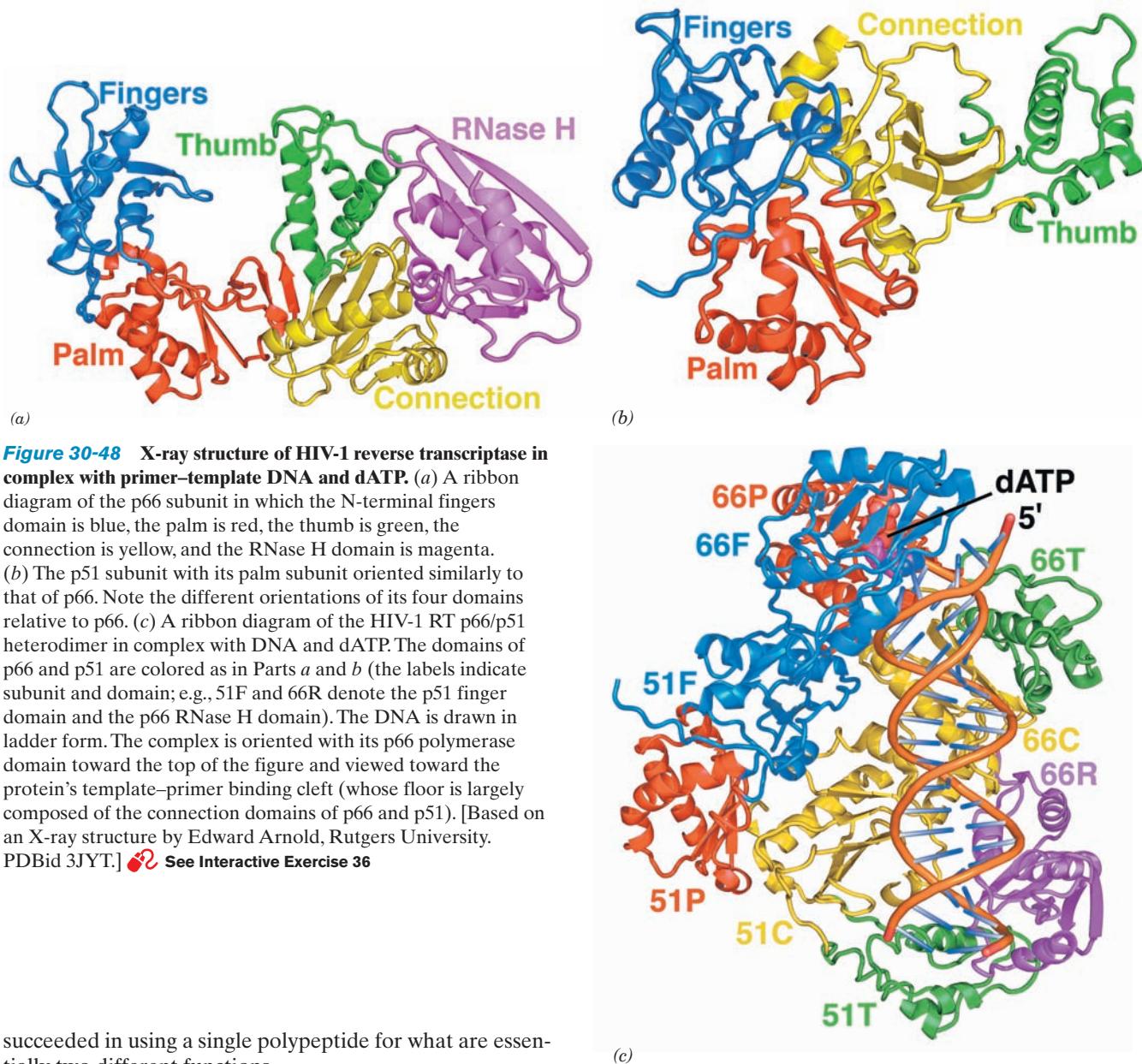


Figure 30-48 X-ray structure of HIV-1 reverse transcriptase in complex with primer-template DNA and dATP. (a) A ribbon diagram of the p66 subunit in which the N-terminal fingers domain is blue, the palm is red, the thumb is green, the connection is yellow, and the RNase H domain is magenta. (b) The p51 subunit with its palm subunit oriented similarly to that of p66. Note the different orientations of its four domains relative to p66. (c) A ribbon diagram of the HIV-1 RT p66/p51 heterodimer in complex with DNA and dATP. The domains of p66 and p51 are colored as in Parts a and b (the labels indicate subunit and domain; e.g., 51F and 66R denote the p51 finger domain and the p66 RNase H domain). The DNA is drawn in ladder form. The complex is oriented with its p66 polymerase domain toward the top of the figure and viewed toward the protein's template-primer binding cleft (whose floor is largely composed of the connection domains of p66 and p51). [Based on an X-ray structure by Edward Arnold, Rutgers University. PDBid 3JYT.]  See Interactive Exercise 36

succeeded in using a single polypeptide for what are essentially two different functions.

The dATP binds at the 3' end of the primer strand, near p66's three catalytically essential Asp side chains, where it pairs with a template dT base. The DNA assumes a conformation that, near the polymerase active site, resembles A-DNA (note the A-DNA-like tilt of the bases with respect to the helix axis below the dATP in Fig. 30-48c), but elsewhere more closely resembles B-DNA (in which the bases are nearly perpendicular to the helix axis), a phenomenon that also has been observed in several structures of A-family DNA polymerases in their complexes with DNA (Section 30-2Ae). Most of the protein-DNA interactions involve the DNA's sugar-phosphate backbone.

The RT active site region contains the few sequence motifs that are conserved among the various polymerases. Indeed, this region of p66 has a striking structural resemblance to DNA polymerases of known structure (Sections 30-2A and 30-4B). This suggests that other DNA polymerases are likely to bind DNA in a similar manner.

D. Telomeres and Telomerase

The ends of linear chromosomes cannot be replicated by any of the mechanisms we have yet considered. This is because the RNA primer at the 5' end of a completed lagging strand cannot be replaced with DNA; the primer required to do this would have no place to bind. How, then, are the DNA sequences at the ends of eukaryotic chromosomes, the **telomeres** (Greek: *telos*, end), replicated?

Telomeric DNA has an unusual sequence: It consists of up to several thousand tandem repeats of a simple, species-dependent, G-rich sequence concluding the 3'-ending strand of each chromosomal terminus. For example, the ciliated protozoan *Tetrahymena* has the repeating telomeric sequence TTGGGG, whereas in all vertebrates it is

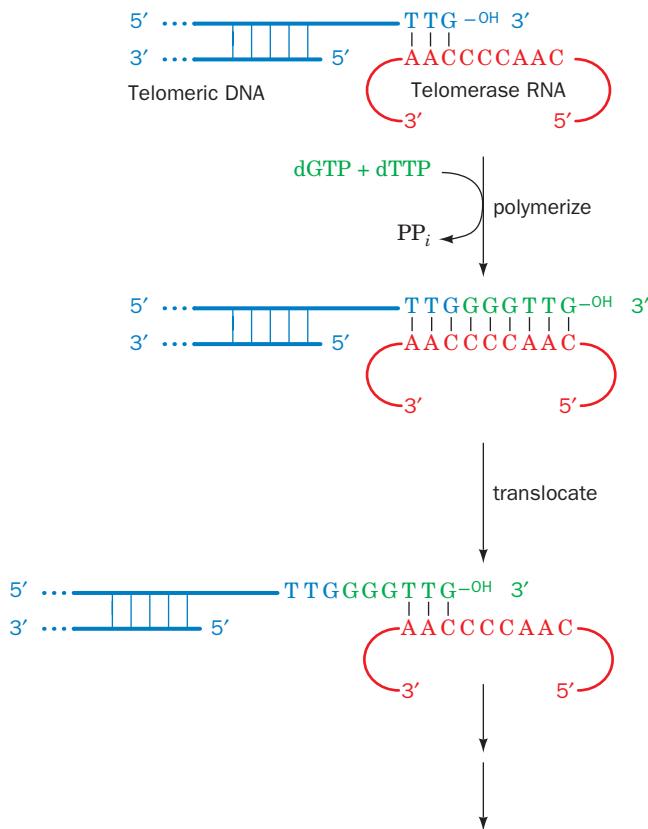


Figure 30-49 Proposed mechanism for the synthesis of telomeric DNA by *Tetrahymena* telomerase. The telomere's 5'-ending strand is later extended by normal lagging strand synthesis. [After Greider, C.W. and Blackburn, E.H., *Nature* **337**, 336 (1989).]

TTAGGG. Moreover, this strand ends with an overhang that varies from ~20 nt in yeast to ~200 bp in humans.

Elizabeth Blackburn, Carol Greider, and Jack Szostak have shown that telomeric DNA is synthesized by a novel mechanism. The enzyme that synthesizes the G-rich strand of telomeric DNA is named **telomerase**. *Tetrahymena* telomerase, for example, adds tandem repeats of the telomeric sequence TTGGGG to the 3' end of any G-rich telomeric oligonucleotide independently of any exogenously added template. A clue as to how this occurs came from the discovery that telomerases are ribonucleoproteins whose RNA components contain a segment that is complementary to the repeating telomeric sequence. This sequence apparently acts as a template in a kind of reverse transcriptase reaction that synthesizes the telomeric sequence, translocates to the DNA's new 3' end, and repeats the process (Fig. 30-49). This hypothesis is confirmed by the observation that mutationally altering the telomerase RNA gene segment complementary to telomere DNA results in telomere DNA with the corresponding altered sequence. In fact, telomerase's highly conserved protein component, which is named **TERT**, is homologous to known reverse transcriptases (its RNA component is called **TER**). The DNA strand complementary to the telomere's G-rich

strand is apparently synthesized by the normal cellular machinery for lagging strand synthesis, thereby accounting for the 3' overhang of the G-rich strand.

a. TERT Resembles Other DNA Polymerases

The X-ray structure of the 596-residue TERT from the red flour beetle *Tribolium castaneum*, determined by Emmanuel Skordalakes, reveals that this subunit contains the familiar fingers–palm–thumb domain organization of other DNA polymerases (Fig. 30-50) and, in particular, resembles the corresponding domains of the HIV-1 reverse transcriptase p66 subunit (Fig. 30-48). In addition, TERT has an N-terminal RNA-binding domain named TRBD (for *telomere repeat binding domain*). TRBD closes the gap between the thumb and fingers, thereby yielding a ringlike protein with a hole that is ~26 Å wide and ~21 Å deep. This is sufficient to accommodate an ~8-bp segment of double-stranded nucleic acid.

b. Telomeres Must Be Capped

Without the action of telomerase, a chromosome would be shortened at both ends by 50 to 100 nt with every cycle of DNA replication and cell division. It was therefore initially assumed that, in the absence of active telomerase, essential genes located near the ends of chromosomes would eventually be lost, thereby killing the descendants of the originally affected cells. However, it is now evident that telomeres serve a vital chromosomal function that is compromised before this can happen. Free DNA ends, which are subject to nuclease degradation, trigger DNA damage repair systems that normally function to rejoin the ends of broken chromosomes (as well as cell cycle arrest until this has happened). Thus exposed telomeric DNA would result in the end-to-end fusion of chromosomes, a process that leads to chromosomal instability and eventual cell death [fused chromosomes often break in mitosis (their two centromeres may cause them to be pulled in opposite directions), activating DNA damage checkpoints]. However, in a process known as **capping**, telomeric DNA is specifically bound by proteins that sequester the DNA ends. There is mounting evidence that capping is a dynamic process in which the probability of a telomere spontaneously upcapping increases as telomere length decreases. Since most somatic cells in multicellular organisms have very low levels of telomerase activity, this explains why such cells in culture can only undergo a limited number of doublings (20–60) before they reach senescence (a stage in which they cease dividing) and eventually die (Section 19-3B). Indeed, otherwise immortal *Tetrahymena* cultures with mutationally impaired telomerasess exhibit characteristics reminiscent of senescent mammalian cells before dying off. Apparently, *the loss of telomerase function in somatic cells is a basis for aging in multicellular organisms*.

c. Telomere Length Correlates with Aging

There is strong experimental evidence in support of this theory of aging. The analysis of cultured human fibroblasts

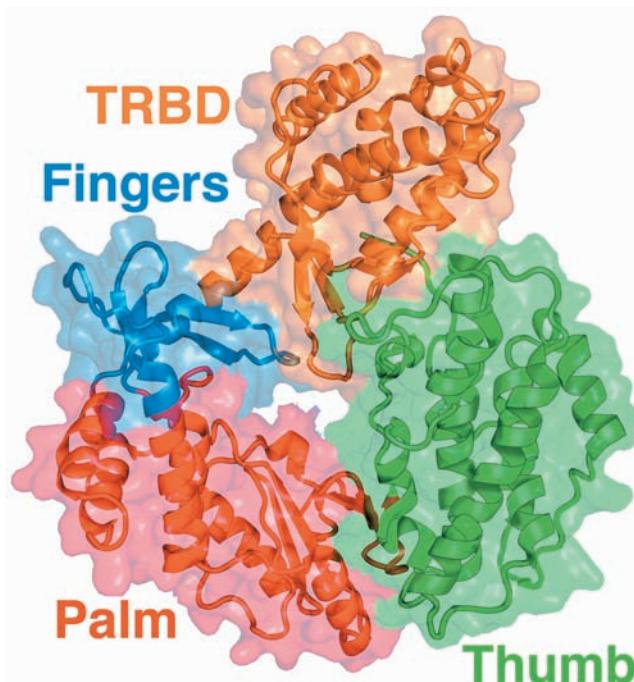


Figure 30-50 X-ray structure of TERT from *Tribolium castaneum*. The protein is shown ribbon form with its TRBD, fingers, palm, and thumb colored orange, blue, red, and green, respectively, and embedded in its like-colored semitransparent

surface diagram. Compare this structure with that of p66 subunit of HIV-1 reverse transcriptase (Fig. 30-48a). [Based on an X-ray structure by Emmanuel Skordalakes, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 3DU5.]

from a number of donors between 0 and 93 years old indicates that there is only a weak correlation between the proliferative capacity of a cell culture and the age of its donor. There is, however, a strong correlation, valid over the entire donor age range, between the initial telomere length in a cell culture and its proliferative capacity. Thus, cells that initially have relatively short telomeres undergo significantly fewer doublings than cells with longer telomeres. Moreover, fibroblasts from individuals with **progeria** (a rare disease characterized by rapid and premature aging resulting in childhood death) have short telomeres, an observation that is consistent with their known reduced proliferative capacity in culture. In contrast, sperm (which, being germ cells, are in effect immortal) from donors ranging in age from 19 to 68 years had telomeres that did not vary in length with donor age, which indicates that telomerase is active at some stage of germ cell growth. Likewise, those few cells in a culture that become immortal (capable of unlimited proliferation) exhibit an active telomerase and a telomere of stable length, as do the cells of unicellular eukaryotes (which are also immortal). It therefore appears that telomere erosion is a significant cause of cellular senescence and hence aging. Indeed transgenic mice that constitutively (at a constant rate) express telomerase have increased lifespans (although, in contrast to humans, mice in which telomerase has been knocked out survive without significant problems for several generations before they become infertile).

d. Cancer Cells Have Active Telomerases

What advantage might multicellular organisms gain by eliminating telomerase activity in their somatic cells? An intriguing possibility is that cellular senescence is a mechanism that protects multicellular organisms from cancer. The two defining characteristics of cancer cells are that they are immortal and grow uncontrollably (Sections 19-3B and 34-4C). If mammalian cells were normally immortal, the incidence of cancer would probably be far greater than it is since immortalization, which requires an active telomerase, is a major step toward **malignant transformation** (cancer formation), which requires several independent genetic changes (Section 19-3B). Indeed, nearly all human cancers exhibit high telomerase activity. Moreover, as Robert Weinberg demonstrated, human fibroblasts in culture can be malignantly transformed by the acquisition of only three genes, those encoding: (1) TERT (its TER subunit is normally expressed in somatic cells), (2) an oncogenic variant of H-Ras (an essential participant in intracellular signal transduction pathways; Section 19-3C), and (3) the SV40 **large-T antigen** [SV40 is a tumor virus whose large-T antigen binds and functionally inactivates the tumor suppressor proteins known as **Rb** and **p53** (Section 34-4C; it also functions as a helicase in viral DNA replication)]. The age-related decline in telomere length in humans does not occur in mice, which suggests that telomere loss evolved to suppress tumor formation in long-lived animals, such as humans, but not in short-lived animals,

such as mice. Thus, telomerase inhibitors may be effective antitumor agents.

e. Telomeric DNA Can Dimerize via G-Quartets

It has long been known that guanine forms strong Hoogsteen-type base pairs (Table 29-2) that can further associate to form cyclic tetramers known as **G-quartets** (Fig. 30-51a). Indeed, G-rich polynucleotides are notoriously difficult to work with because of their propensity to aggregate. The G-rich overhanging strands of telomeres dimerize to form stable complexes in solution, presumably via the formation of G-quartet-containing structures.

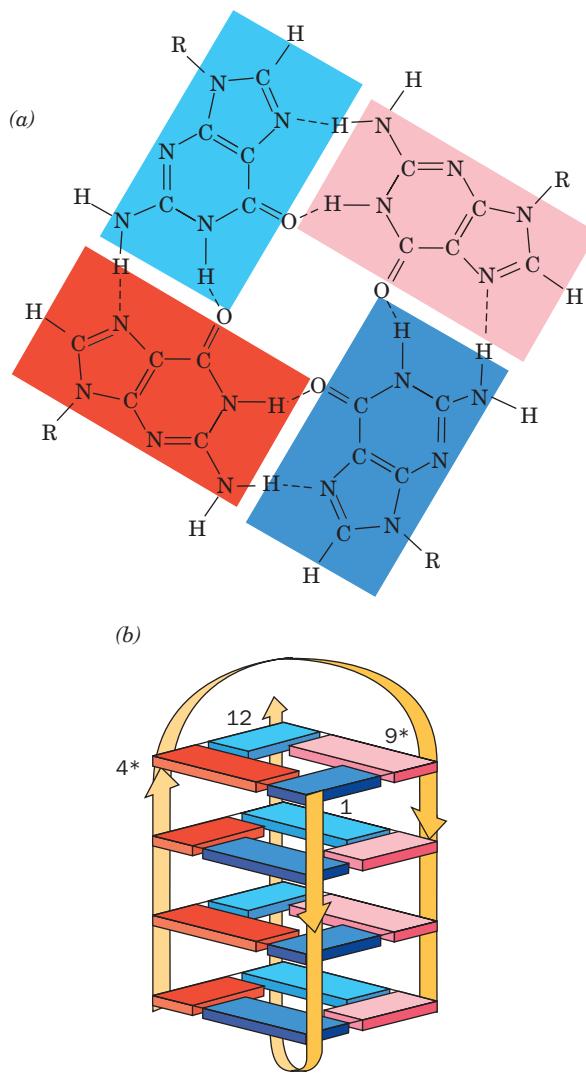


Figure 30-51 NMR structure of the telomeric oligonucleotide d(GGGGTTTTGGGG). (a) The base-pairing interactions in the G-quartet at the end of the quadruplex in solution. (b) Schematic diagram of the NMR solution structure, in which the strand directions are indicated by arrows. The nucleotides are numbered 1 to 12 in one strand and 1* to 12* in the symmetry-related strand. Guanine residues G1 to G4 are represented by dark blue rectangles, G8 to G12 are light blue, G1* to G4* are red, and G9* to G12* are pink. [After Schultze, P., Smith, F.W., and Feigon, J., *Structure* 2, 227 (1994). PDBid 156D.]

The 3'-terminal telomeric overhang of the ciliated protozoan *Oxytricha nova* has the sequence d(T₄G₄)₂, which resembles the repeating telomeric sequences of other organisms. The NMR structure of the dodecamer d(G₄T₄G₄), determined by Juli Feigon (Fig. 30-51b), reveals that each oligonucleotide folds back on itself to form a hairpin, two of which associate in an antiparallel fashion to form a structure that contains four stacked G-quartets, with the T₄ sequences forming the loops at the ends of each stack.

The **telomere end binding protein (TEBP)** of *O. nova* is a heterodimeric capping protein that binds to and protects the foregoing 3' overhang. The X-ray structure of TEBP in complex with d(G₄T₄G₄), determined by Steve Schultz, reveals that the DNA binds in a deep cleft between the protein's α and β subunits, where it adopts an irregular nonhelical conformation (Fig. 30-52). In addition, two other d(G₄T₄G₄) molecules form a G-quartet-linked dimer with the same conformation they adopt in solution (Fig. 30-51). The G-quartet assembly fits snugly into a small positively

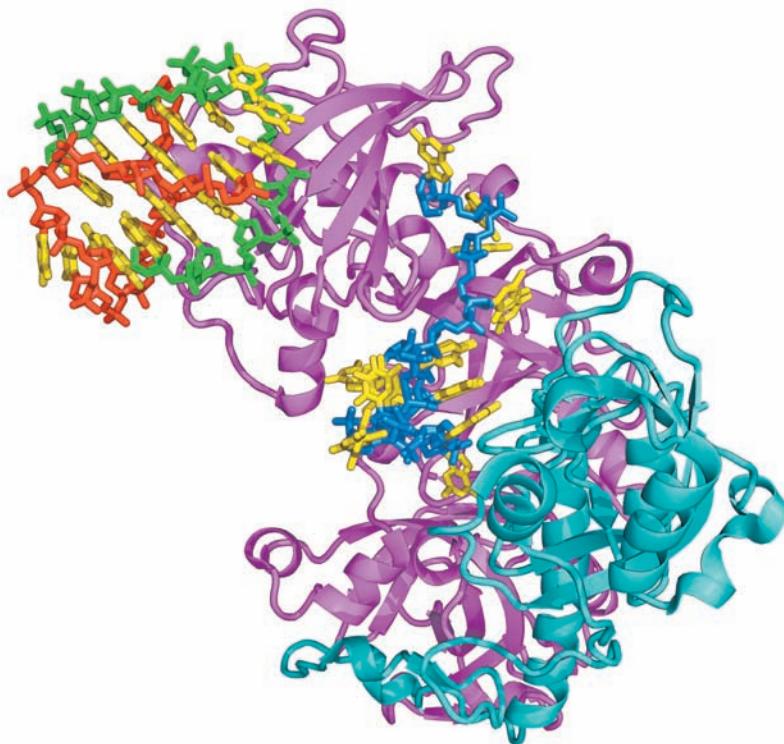


Figure 30-52 X-ray structure of *Oxytricha nova* telomere end binding protein (TEBP) in complex with d(G₄T₄G₄). The TEBP is drawn in ribbon form with its α and β subunits magenta and cyan. The DNA is drawn in stick form with its bases yellow, the sugar-phosphate backbone of the single strand that binds in a cleft between the protein's α and β subunits blue, and the backbones of two strands that form a G-quartet-linked dimer red and green. The G-quartet-linked dimer binds in a cavity formed by the N-terminal domains of three symmetry-related α chains, although only one of them is shown here. [Based on an X-ray structure by Steve Schultz, University of Colorado. PDBid 1JB7.]

charged cavity formed by the N-terminal domains of three symmetry-related (in the crystal) α subunits at sites distinct from their ssDNA binding sites. The α subunit (TEBP α) contains three so-called **OB folds** (OB for *oligomer binding*), common oligonucleotide/oligosaccharide binding motifs that each contain a characteristic 5-stranded β barrel. Two of these OB folds participate in DNA binding and a third interacts with TEBP β , which also contains an OB fold. The presence of both the ssDNA and the G-quartet assembly in the X-ray structure supports the hypothesis that multiple DNA structures and, in particular, G-quartets, play a role in telomere biology.

Although TEBP is not present in yeast or vertebrates, both humans and fission yeast express a telomere end-binding protein named **POT1** (for *protection of telomeres-1*) and its binding partner **TPP1** (so named because it had been previously called *TINT1*, *PTOP*, and *PIP1*), which bind to the single-stranded overhang at the ends of telomeres. POT1 consists mainly of two OB folds in which the N-terminal OB fold is homologous to that of TEBP α . The deletion of POT1 causes rapid loss of telomeric DNA and chromosomal end-joining. TPP1 is structurally similar to TEBP β , despite their only 11% sequence identity, and hence the POT1-TPP1 complex appears to be a homolog of TEBP.

f. Telomeres Form T-Loops

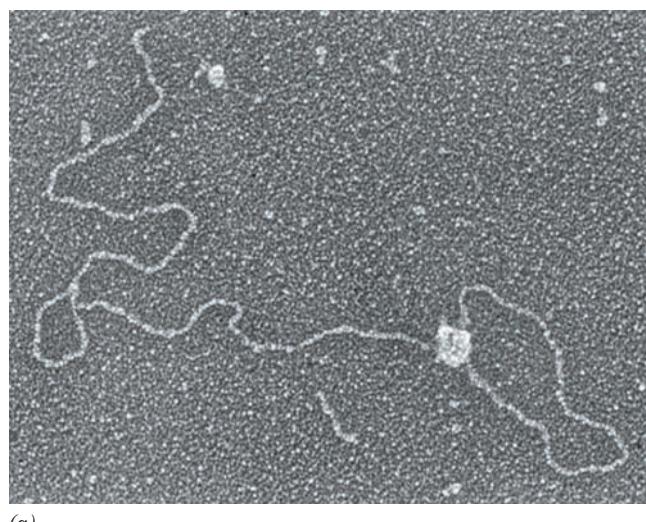
Mammalian telomeric DNA is also capped by two related proteins, **TRF1** and **TRF2** (TRF for *telomere repeat-binding factor*). Jack Griffith and Titia de Lange have shown through electron microscopy (EM) studies that, in the presence of TRF2, otherwise linear telomeric DNA forms large duplex end loops named **T-loops** (T for *telomere*;

Fig. 30-53a). Moreover, the EM of DNA from mammalian telomeres, whose strands had been chemically cross-linked to preserve their structural relationships on deproteinization, likewise revealed the presence of abundant T-loops of varying sizes. These observations suggest that T-loops are formed by the TRF2-induced invasion of the 3' telomeric overhang into the repeating telomeric dsDNA (Fig. 30-53b) to form a **D-loop** (D for *displacement*); a segment of dsDNA whose two strands are separated). T-loops have also been observed in protozoa, suggesting that T-loops are a conserved feature of eukaryotic telomeres. TRF1 is implicated in controlling telomeric length, presumably by somehow limiting the number of TRF1 molecules that can bind to a telomere.

POT1, which binds to the single-stranded 3' extension at chromosome ends, and TRF1 and TRF2, which bind double-stranded telomere DNA, are bridged by TPP1 and **TIN2** (for *TRF1 interacting protein 2*). A sixth protein, **RAP1** (for *repressor activator protein 1*), binds mainly to TRF2. The complex formed by these six proteins, which is known as **shelterin**, apparently functions both to protect the mammalian telomere from being mistaken for a broken chromosome and hence being subject to DNA repair, and to limit telomere length by preventing the extension of its bound telomere by telomerase.

5 REPAIR OF DNA

DNA is by no means the inert substance that might be supposed from naive consideration of genome stability. Rather, the reactive environment of the cell, the presence



(a)

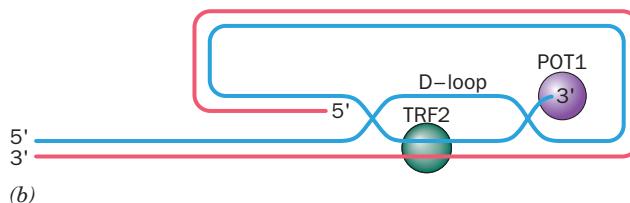


Figure 30-53 The telomeric T-loop. (a) An electron micrograph of a dsDNA consisting of a 3-kb unique sequence followed by ~2 kb of the repeating sequence TTAGGG on the strand that ends with a 150- to 200-nt 3' overhang. This model telomeric DNA was then incubated with human TRF2. [Courtesy of Jack Griffith, University of North Carolina at Chapel Hill.] (b) The proposed structure of a T-loop. In a process that is mediated by TRF2, the repeating TTAGGG sequence in the DNA's 3' overhang displaces a portion of the same strand (blue) in the double-stranded region of the DNA to form a duplex segment with the complementary strand (red), thereby generating a D-loop. The telomere end-binding protein POT1 specifically binds to the end of the 3' overhang.

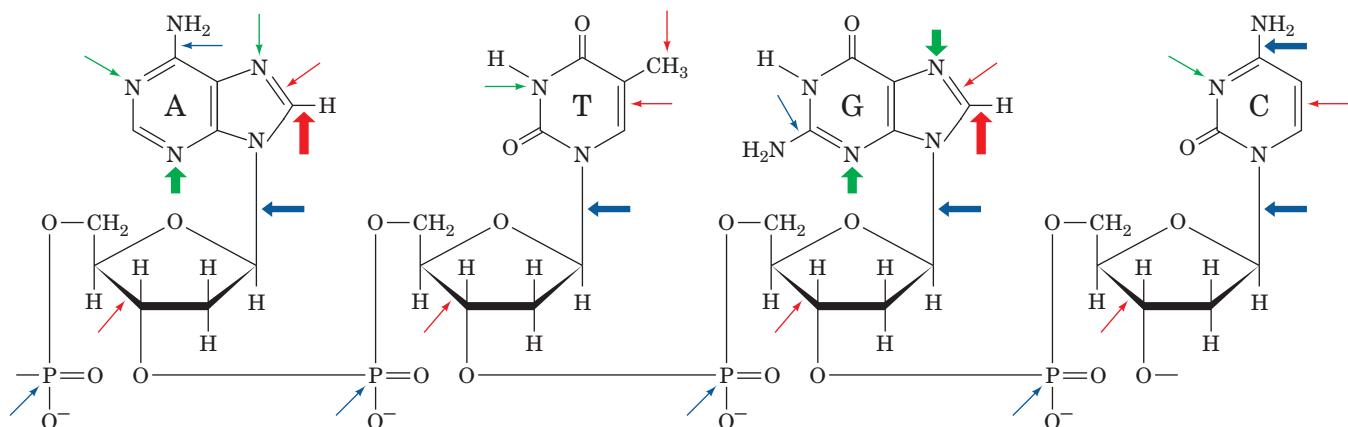


Figure 30-54 The types and sites of chemical damage to which DNA is normally susceptible *in vivo*. Red arrows indicate sites subject to oxidative attack, blue arrows indicate sites subject to spontaneous hydrolysis, and green arrows indicate sites subject

to nonenzymatic methylation by *S*-adenosylmethionine. The width of an arrow is indicative of the relative frequency of the reaction. [After Lindahl, T., *Nature* **362**, 709 (1993).]

of a variety of toxic substances, and exposure to UV or ionizing radiation subjects it to numerous chemical insults that excise or modify bases and alter sugar–phosphate groups (Fig. 30-54). Indeed, some of these reactions occur at surprisingly high rates. For example, under normal physiological conditions, the glycosidic bonds of $\sim 10,000$ of the 3 billion purine nucleotides in each human cell hydrolyze spontaneously each day.

Any DNA damage must be repaired if the genetic message is to maintain its integrity. Such repair is possible because of duplex DNA's inherent information redundancy. The biological importance of DNA repair is indicated by the identification of at least 130 genes in the human genome that participate in DNA repair and by the great variety of DNA repair pathways possessed by even relatively simple organisms such as *E. coli*. In fact, the major DNA repair processes in eukaryotic cells and *E. coli* are chemically quite similar. These processes are outlined in this section.

The two redundant copies of genetic information carried by dsDNA ideally suit it for the repair of damage to one of its strands: The damaged strand can be repaired under the direction of the undamaged strand. The importance of dsDNA to the storage of genetic information is indicated by the fact that only a few small viruses carry single-stranded DNA or RNA as their genetic material (e.g., ϕ X174 and HIV). The DNA repair systems discussed below do not operate on single-stranded nucleic acids and hence these viruses have very high rates of mutation. Thus it appears that only organisms with very small genomes can afford the economy of not encoding their genomes on dsDNA.

A. Direct Reversal of Damage

a. Pyrimidine Dimers Are Split by DNA Photolyase

UV radiation of 200 to 300 nm promotes the formation of a cyclobutyl ring between adjacent thymine residues on the same DNA strand to form an intrastrand thymine

dimer (Fig. 30-55). Similar cytosine and thymine–cytosine dimers are likewise formed but at lesser rates. Such **cyclobutane pyrimidine dimers (CPDs)** locally distort DNA's base-paired structure such that it can be neither transcribed nor replicated. Indeed, a single thymine dimer, if unrepaired, is sufficient to kill an *E. coli*.

Pyrimidine dimers may be restored to their monomeric forms through the action of light-absorbing enzymes named **photoreactivating enzymes** or **DNA photolyases** that are present in many prokaryotes and eukaryotes (including goldfish, rattlesnakes, and marsupials, but not placental mammals). These enzymes are 55- to 65-kD monomers that bind to a pyrimidine dimer in DNA, a process that can occur in the dark. A noncovalently bound chromophore, in

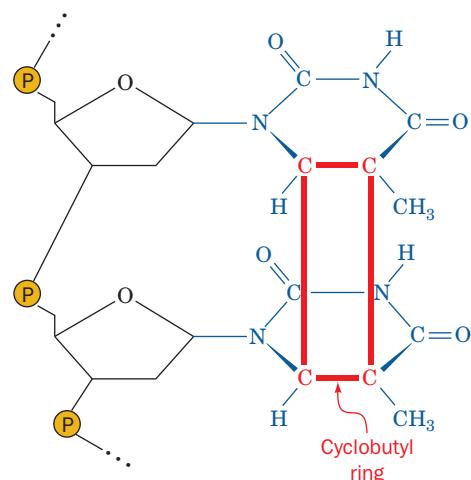
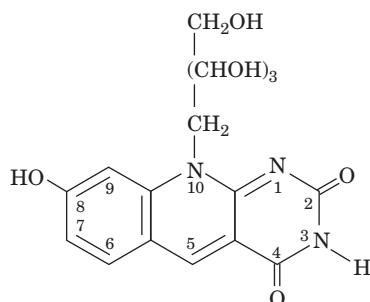


Figure 30-55 The cyclobutylthymine dimer that forms on UV irradiation of two adjacent thymine residues on a DNA strand. The $\sim 1.6\text{-}\text{\AA}$ -long covalent bonds joining the thymine rings (red) are much shorter than the normal $3.4\text{-}\text{\AA}$ spacing between stacked rings in B-DNA, thereby locally distorting the DNA.

some species an N^5,N^{10} -methenyltetrahydrofolate (MTHF; Fig. 26-49) and in others a **5-deazafavin**,



8-Hydroxy-7,8-didemethyl-5-deazariboflavin

then absorbs 300- to 500-nm light and transfers the excitation energy to a noncovalently bound FADH^- , which in turn transfers an electron to the pyrimidine dimer, thereby splitting it. Finally, the resulting pyrimidine anion re-reduces the FADH^\cdot and the now unblemished DNA is released, thereby completing the catalytic cycle. DNA photolyases bind either dsDNA or ssDNA with high affinity but without regard to base sequence.

Thomas Carell and Lars-Oliver Essen determined the X-ray structure of the 474-residue DNA photolyase from the cyanobacterium *Anacystis nidulans* in complex with a 9-bp dsDNA containing a synthetic thymine dimer whose bridging phosphate group was replaced by a $-\text{O}-\text{CH}_2-\text{O}-$ group (which does not affect the enzyme's ability to split the dimer). The DNA binds to a highly positively charged surface on the protein with its thymine dimer flipped out of the double helix and bound in a deep cavity (Fig. 30-56). This flip-out is probably facilitated by the relatively weak base pairing interactions of the thymine dimer and the distortions it imposes on the double helix. In the following discussions we shall see that this so-called **base flipping** (really nucleotide flipping since entire nucleotides flip out of the double helix) is by no means an unusual process for enzymes that perform chemistry on the bases of dsDNA. The DNA outside of the thymine dimer assumes the B conformation but at the thymine dimer is bent by 50° away from the protein, thereby unstacking the adenine bases complementary to the dimerized thymine bases. The "hole" in the DNA helix left by the flipped-out thymine dimer is partially occupied by an irregular protein ridge.

In the X-ray structure, the thymine dimer's C5—C5 and C6—C6 bonds are broken. Yet, in the crystal, the enzyme-bound thymine dimer is stable in the dark for at least a year. Apparently, the X-rays used to generate the diffraction data mimic the effects of the light that normally ruptures these bonds. Moreover, the FAD's isoalloxazine ring exhibits a 9° "butterfly" bend about its N5—N10 axis (the isoalloxazine ring's atomic numbering scheme is given in Fig. 16-8), which is indicative that it is in the fully reduced FADH^- form. The isoalloxazine ring and adenine ring of the FAD $^-$, which has a folded conformation, are in van der Waals contact with one or the other bases of the thymine dimer and the isoalloxazine ring is ~ 10 Å distant from the

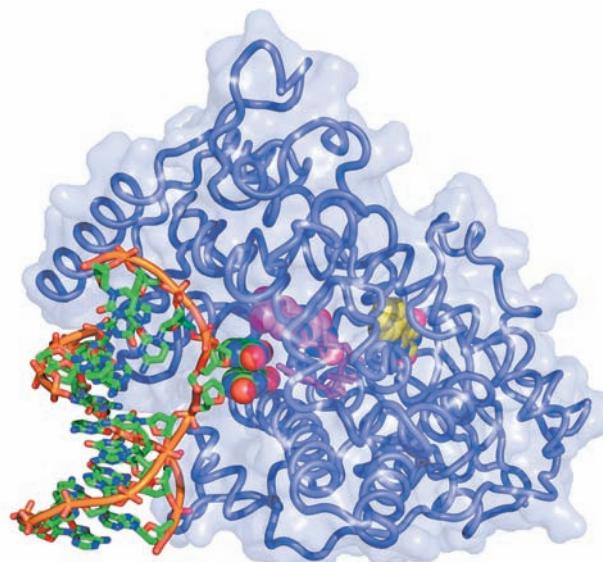


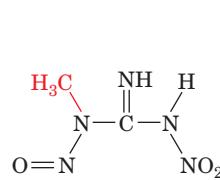
Figure 30-56 X-ray structure of *A. nidulans* DNA photolyase in complex with dsDNA containing a synthetic thymine dimer.

The protein is drawn in worm form embedded in its semitransparent surface diagram. The DNA, in which the phosphate group bridging the nucleotides of the thymine dimer is replaced by a $-\text{O}-\text{CH}_2-\text{O}-$ group, is drawn mainly in stick form but with the bases of the thymine dimer in space-filling form, all colored according atom type (C green, N blue, O red, and P orange) and with successive P atoms in each polynucleotide chain connected by orange rods. The FAD and MTHF are drawn in stick form with their flavin and flavinlike rings in space-filling form and with FAD C magenta and MTHF C yellow. [Based on an X-ray structure by Thomas Carell, Ludwig Maximilians University, Munich, Germany, and Lars-Oliver Essen, Philipps University, Marburg, Germany. PDBid 1TEZ.]

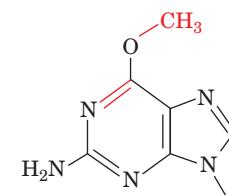
flavin-like ring of the MTHF. This permits the observed efficient energy transfer in the photolyase reaction (which has a quantum yield of ~ 0.9).

b. Alkyltransferases Dealkylate Alkylated Nucleotides

The exposure of DNA to alkylating agents such as **N-methyl-N'-nitro-N-nitrosoguanidine (MNNG)**



N-Methyl-N'-nitro-N-nitrosoguanidine (MNNG)



O⁶-Methylguanine residue

yields, among other products, **O⁶-alkylguanine** residues. The formation of these derivatives is highly mutagenic because on replication, they frequently cause the incorporation of thymine instead of cytosine.

***O*⁶-Methylguanine** and ***O*⁶-ethylguanine** lesions of DNA in all species tested are repaired by ***O*⁶-alkylguanine–DNA alkyltransferase**, which directly transfers the offending alkyl group to one of its own Cys residues. The reaction inactivates this protein, which therefore cannot be strictly classified as an enzyme. The alkyltransferase reaction has elicited considerable attention because carcinogenesis induced by methylating and ethylating agents is correlated with deficient repair of *O*⁶-alkylguanine lesions.

The *E. coli* *O*⁶-alkylguanine–DNA alkyltransferase activity occurs on the 178-residue C-terminal segment of the 354-residue **Ada protein** (the product of the *ada* gene). Its X-ray structure (Fig. 30-57a), determined by Eleanor Dodson and Peter Moody, reveals, unexpectedly, that its active site Cys residue, Cys 321, is buried inside the protein. Apparently, the protein must undergo a significant conformational change on DNA binding in order to effect the methyl transfer reaction.

Ada protein's 92-residue N-terminal segment has an independent function: It repairs methyl phosphotriesters in DNA (methylated phosphate groups) by irreversibly transferring the offending methyl group to its Cys 69. The NMR structure of Ada's N-terminal domain (Fig. 30-57b), determined by Gregory Verdine and Gerhard Wagner, reveals

that Cys 69, together with three other Cys residues, tetrahedrally coordinates a Zn²⁺ ion. This presumably stabilizes the thiolate form of Cys 69 over its thiol form, thereby facilitating its nucleophilic attack on the methyl group.

Intact Ada protein that is methylated at its Cys 69 binds to a specific DNA sequence, which is located upstream of the *ada* gene and several other genes encoding DNA repair proteins, thereby inducing their transcription. Evidently, Ada also functions as a chemosensor of methylation damage.

B. Excision Repair

Cells employ two types of excision repair mechanisms: (1) **nucleotide excision repair (NER)**, which functions to repair relatively bulky DNA lesions; and (2) **base excision repair (BER)**, which repairs nonbulky lesions involving a single base.

a. Nucleotide Excision Repair

NER is a DNA repair mechanism found in all cells that eliminates damage to dsDNA by excising an oligonucleotide containing the lesion and filling in the resulting single-strand

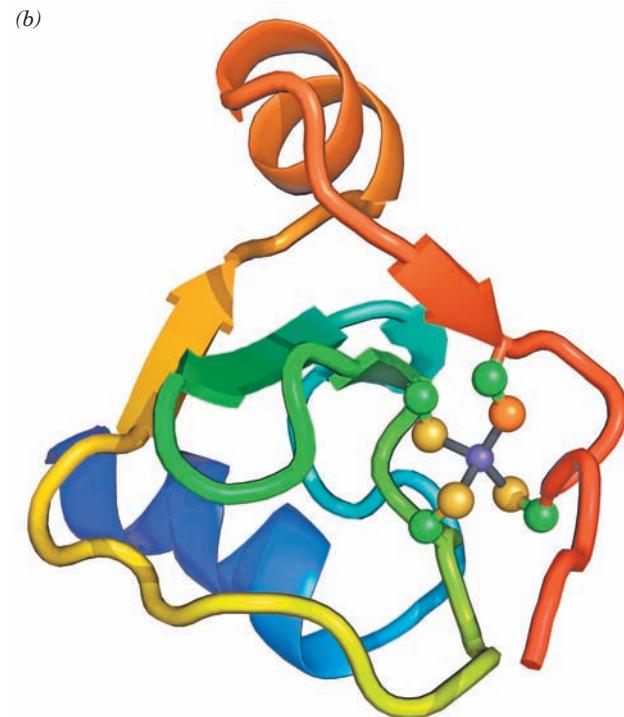
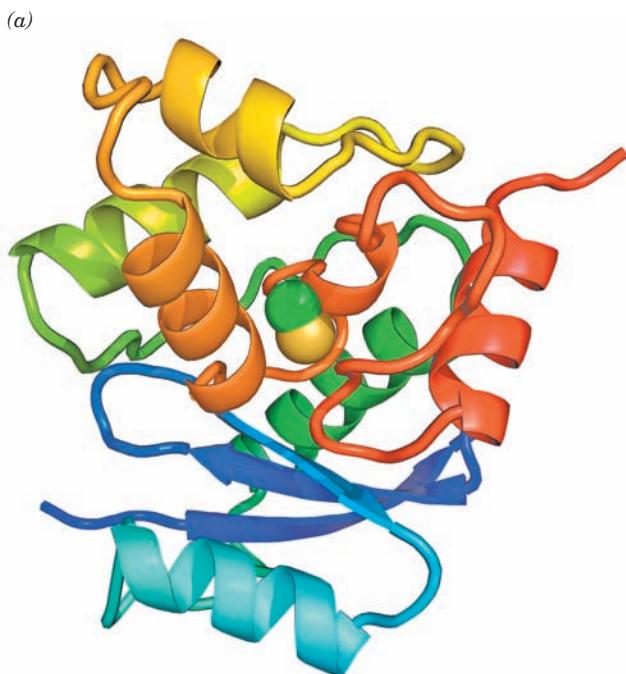


Figure 30-57 Structure of the *E. coli* Ada protein. (a) The X-ray structure of Ada's 178-residue C-terminal segment, which contains its *O*⁶-alkylguanine–DNA alkyltransferase function. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chain of Cys 146 (Cys 321 in the intact protein), to which the methyl group is irreversibly transferred, is shown in space-filling form with C green and S yellow. Note that this residue is almost entirely buried within the protein. [Based on an X-ray structure by Eleanor Dodson and Peter Moody, University of York, U.K. PDBid 1SFE.] (b) The NMR structure of Ada's 92-residue,

N-terminal segment, which mediates its methyl phosphotriester repair function. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). Its bound Zn²⁺ ion is represented by a purple sphere and its four tetrahedrally coordinating Cys side chains are shown in ball-and-stick form, with C green and S yellow except for the orange S atom of Cys 69, which becomes irreversibly methylated when the protein encounters a methylated phosphate group on DNA. The coordinate bonds to the Zn²⁺ ion are represented by gray sticks. [Based on an NMR structure by Gregory Verdine and Gerhard Wagner, Harvard University. PDBid 1ADN.]

gap. NER repairs lesions that are characterized by the displacement of bases from their normal positions, such as pyrimidine dimers, or by the addition of a bulky substituent to a base. This system appears to be activated by a helix distortion rather than by the recognition of any particular group. In humans, NER is the major defense against two important carcinogens, sunlight and cigarette smoke. The mechanism of NER in prokaryotes is similar to that in eukaryotes. However, prokaryotic NER employs 3 subunits, whereas eukaryotic NER involves the actions of 16 subunits. The eukaryotic proteins are conserved from yeast to humans but none of them exhibit any sequence similarity to the prokaryotic proteins, suggesting that the two NER systems arose by convergent evolution.

In *E. coli*, NER is carried out in an ATP-dependent process through the actions of the **UvrA**, **UvrB**, and **UvrC** proteins (the products of the *uvrA*, *uvrB*, and *uvrC* genes). This system, which is often referred to as the **UvrABC endonuclease** (although, as we shall see, there is no complex that contains all three subunits), cleaves the damaged DNA strand at the seventh and at the third or fourth phosphodiester bonds from the lesion's 5' and 3' sides, respectively (Fig. 30-58; this system is also known as an **excinuclease** to indicate that it excises a DNA segment rather than cleaving it in only one place as do most endonucleases). The excised 11- or 12-nt oligonucleotide is displaced by the binding of **UvrD** (also called **helicase II**) and replaced through the actions of Pol I and DNA ligase.

The mechanism of prokaryotic NER was elucidated mainly by Aziz Sancar. It begins with the damage recognition step in which a (UvrA)₂UvrB heterotrimer binds

tightly although nonspecifically to dsDNA, which it probes for damage according to its local propensity for bending and unwinding. The presence of a lesion activates the helicase function of UvrB to unwind 5 bp around the lesion in an ATP-driven process. This conformation change induces the dissociation of the UvrA from the complex, which allows the binding of UvrC. UvrB then makes the incision on the 3' side of the lesion following which UvrC makes the incision on its 5' side. UvrD binds to the resulting nicks in the DNA, which displaces UvrC and the lesion-containing oligomer. This makes the 5' incision site accessible to Pol I, which fills in the gap and displaces UvrB. Finally, DNA ligase seals the remaining nick yielding refurbished DNA.

b. Xeroderma Pigmentosum and Cockayne Syndrome Are Caused by Genetically Defective NER

In humans, the rare inherited disease **xeroderma pigmentosum (XP; Greek: *xeros*, dry + *derma*, skin)** is mainly characterized by the inability of skin cells to repair UV-induced DNA lesions. Individuals suffering from this autosomal recessive condition are extremely sensitive to sunlight. During infancy they develop marked skin changes such as dryness, excessive freckling, and keratoses (a type of skin tumor; the skin of these children is described as resembling that of farmers with many years of sun exposure), together with eye damage, such as opacification and ulceration of the cornea. Moreover, they develop often fatal skin cancers at a 2000-fold greater rate than normal and internal cancers at a 10 to 20-fold increased rate. Curiously, many individuals with XP also have a bewildering variety of seemingly unrelated symptoms including progressive neurological degeneration and developmental deficits.

Cultured skin fibroblasts from individuals with xeroderma pigmentosum are defective in the NER of pyrimidine dimers. Cell-fusion experiments with cultured cells taken from various patients have demonstrated that this disease results from defects in any of 8 complementation groups (Section 1-4Cc), indicating that there must be at least 8 gene products, **XPA** through **XPG** and **XPV**, involved in this clearly important UV damage repair pathway.

What is the biochemical basis for the diverse group of symptoms associated with impaired NER? The reactive oxygen species (ROS; Section 22-4Cg) produced by oxidative metabolism readily damages DNA. Some of these oxidative lesions are repaired via NER. Since neurons have high rates of respiration and are long-lived nondividing cells, it seems likely that they would be particularly susceptible to oxidative damage in the absence of NER. This explains the progressive neurological deterioration in XP.

The need to repair a DNA lesion is particularly urgent if the damaged gene is being transcribed because RNA polymerase cannot transcribe through damaged DNA. Cells therefore recruit their DNA repair machinery to such genes in a process known as **transcription-coupled repair (TCR)**, which operates only on the DNA strand that is being transcribed (lesions on the complementary strand are repaired at normal rates). For example, pyrimidine dimers are more rapidly removed from transcribed portions of DNA than from unexpressed sequences.

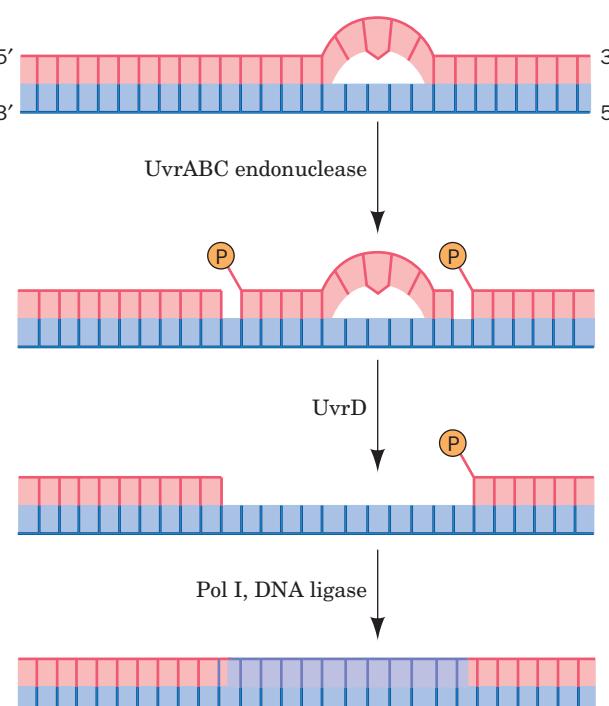


Figure 30-58 The mechanism of nucleotide excision repair (NER) of pyrimidine photodimers.

Cockayne syndrome (CS) is an inherited disease caused by defective TCR. Individuals with CS are hypersensitive to UV radiation (although they have a normal incidence of skin cancer) and exhibit stunted growth as well as neurological dysfunction due to neuron demyelination leading to death in childhood. CS is most often caused by mutations in two complementation groups, **CSA** and **CSB**, although certain defects in **XPB**, **XPD**, and **XPG** can also cause CS in addition to XP.

The retarded development typical of XPB and XPD defects and perhaps the demyelination that occurs in CS appear to be due more to impaired transcription than to defective NER. This is explained by the fact that the DNA helicases XPB and XPD are subunits of the ten-subunit eukaryotic transcription factor **TFIIF**, whose proper functioning is required for transcription initiation by RNA polymerase II (Section 34-3Bb) as well as NER and TCR. A eukaryotic RNA polymerase that is stalled at a DNA lesion is recognized by CSB, which then recruits the other TCR components. CSA interacts with both CSB and the **p44** subunit of TFIIF. Once the damage has been repaired, the RNA polymerase resumes transcription. Thus, in CS, RNA polymerase molecules become permanently stalled on DNA lesions.

TCR also occurs in bacteria. For example, in *E. coli*, the protein **TCRF** (for transcription repair coupling factor; also known as **Mfd** protein, for *m*utation *f*requency *d*cline) is an ATP-powered DNA translocase that displaces stalled RNA polymerase from the damaged template strand, following which TCRF recruits the proteins of the UvrABC system to the damage site. The repaired gene is then transcribed from its beginning. Since prokaryotic genes are much shorter than eukaryotic genes (most of which contain several long noncoding segments known as introns; Section 31-4Ac), this is a more efficient use of resources than generating the complex machinery necessary to restart transcription as occurs in eukaryotes.

c. Base Excision Repair

DNA bases are modified by reactions that occur under normal physiological conditions as well as through the action of environmental agents. For example, adenine and cytosine residues spontaneously deaminate at finite rates to yield hypoxanthine and uracil residues, respectively. *S*-Adenosylmethionine (SAM), a common metabolic methylating agent (Section 26-3Ea), occasionally nonenzymatically methylates a base to form derivatives such as 3-methyladenine and 7-methylguanine residues (Fig. 30-54). Ionizing radiation can promote ring opening reactions in bases. Such changes modify or eliminate base pairing properties.

DNA containing a damaged base may be restored to its native state through base excision repair (BER). Cells contain a variety of **DNA glycosylases** that each cleave the glycosidic bond of a corresponding specific type of altered nucleotide (Fig. 30-59), thereby leaving a deoxyribose residue in the backbone. Such **apurinic** or **apyrimidinic (AP)** sites (also called **abasic sites**) are also generated under normal physiological conditions by the spontaneous hydrolysis of a glycosidic bond. The deoxyribose residue is then cleaved on one side by an **AP endonuclease**, the deoxyribose and several adjacent residues are removed by the action of a

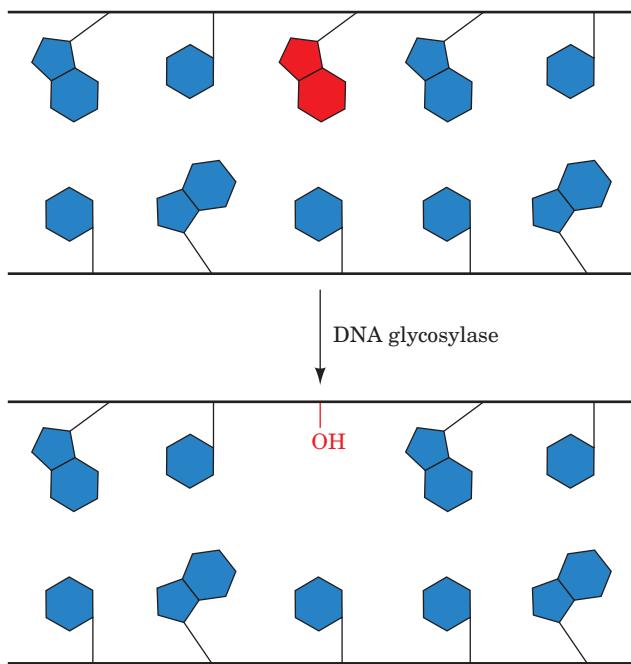


Figure 30-59 Action of DNA glycosylases. These enzymes hydrolyze the glycosidic bond of their corresponding altered base (red) to yield an AP site.

cellular exonuclease (possibly associated with a DNA polymerase), and the gap is filled in and sealed by a DNA polymerase and DNA ligase.

d. Uracil in DNA Would Be Highly Mutagenic

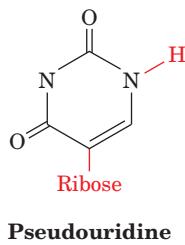
For some time after the essential functions of nucleic acids had been elucidated, there seemed no apparent reason for nature to go to the considerable metabolic effort of using thymine in DNA and uracil in RNA when these substances have virtually identical base pairing properties. This enigma was solved by the discovery of cytosine's penchant for conversion to uracil by deamination, either via spontaneous hydrolysis (Fig. 30-54), which is estimated to occur ~120 times per day in each human cell, or by reaction with nitrates (Section 32-1Aa). If U were the normal DNA base, the deamination of C would be highly mutagenic because there would be no indication of whether the resulting mismatched G · U base pair had originally been G · C or A · U. Since T is DNA's normal base, however, any U in DNA is almost certainly a deaminated C. U's that occur in DNA are efficiently excised by **uracil-DNA glycosylase [UDG; also called uracil N-glycosylase (UNG)]** and then replaced by C through BER.

UDG also has an important function in DNA replication. dUTP, an intermediate in dTTP synthesis, is present in all cells in small amounts (Section 28-3Ba). DNA polymerases do not discriminate well between dUTP and dTTP (recall that DNA polymerases select a base for incorporation into DNA according to its ability to form a Watson-Crick-shaped base pair with the template; Section 30-2Aa) so that, despite the low dUTP level that cells maintain, newly synthesized DNA contains an occasional U. These U's are rapidly replaced by T through BER. However, since excision occurs more rapidly than repair, all

newly synthesized DNA is fragmented. When Okazaki fragments were first discovered (Section 30-1C), it therefore seemed that all DNA was synthesized discontinuously. This ambiguity was resolved with the discovery of *E. coli* defective in UDG. In these *ung*[−] mutants, only about half of the newly synthesized DNA is fragmented, strongly suggesting that DNA's leading strand is synthesized continuously.

e. Uracil-DNA Glycosylase Induces Uridine Nucleotides to Flip Out

The X-ray structure of human UDG in complex with a 10-bp DNA containing a U · G mismatch (which can form a doubly hydrogen-bonded base pair whose shape differs from that of Watson–Crick base pairs; Section 32-2Db), determined by John Tainer, reveals that the UDG has bound the DNA with its uridine nucleotide flipped out of the ds-DNA (Fig. 30-60). Moreover, the enzyme has hydrolyzed uridine's glycosidic bond yielding the free uracil base and an AP site on the DNA, although both remain bound to the enzyme. The cavity in the DNA's base stack that would otherwise be occupied by the flipped-out uracil is filled by the side chain of Arg 272, which intercalates into the DNA from its minor groove side. The X-ray structure of a similar complex in which the U · G mismatch was replaced by a U · A base pair contained essentially identical features. However, when the U in the U · A-containing complex was replaced by **pseudouridine** (in which the “glycosidic” bond is made to uracil's C5 atom rather than to N1),



Pseudouridine

the uracil remained covalently linked to the DNA because the UDG could not hydrolyze its now C—C “glycosidic” bond.

How does UDG detect a base-paired uracil in the center of DNA and how does it discriminate so acutely between uracil and other bases, particularly the closely similar thymine? The above X-ray structures indicate that the phosphate groups flanking the flipped-out nucleotide are 4 Å closer together than they are in B-DNA (8 Å vs 12 Å), which causes the DNA to kink by ~45° in the direction parallel to the view in Fig. 30-60. These distortions arise from the binding of three rigid protein loops to the DNA, which would be unable to simultaneously bind to undistorted B-DNA. This led Tainer to formulate the “pinch–push–pull” mechanism for uracil detection in which he postulated that UDG rapidly scans a DNA for uracil by periodically binding to it so as to compress and thereby slightly bend the DNA's backbone (pinch). The DNA's presumed low resistance to bending at a uracil-containing site (a U · G base pair is smaller than C · G and hence leaves a space in the base stack, whereas a U · A base pair is even weaker than T · A) permits the enzyme to flip out the uracil by intercalating Arg 272 into the minor groove (push), thereby fully bending and kinking the DNA.

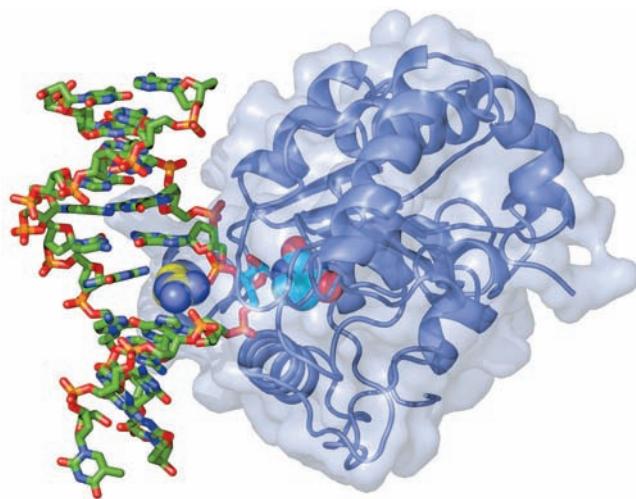


Figure 30-60 X-ray structure of human uracil–DNA glycosylase (UDG) in complex with a 10-bp DNA containing a U · G base pair. The protein (the C-terminal 223 residues of the 304-residue monomer) is represented by its ribbon diagram embedded in its transparent molecular surface. The DNA, viewed looking into its major groove, is drawn in stick form colored according to atom type (C green, N blue, O red, P orange). The **U · G base pair**'s uridine nucleotide has flipped out of the double helix (to the right of the DNA) and has been hydrolyzed to yield an AP nucleotide (stick form with C cyan) and uracil (space-filling form with C cyan), which remains bound in the enzyme's binding pocket. The side chain of Arg 272 (space-filling form with C yellow) has intercalated into the DNA base stack to fill the space vacated by the flipped-out uracil base. [Based on an X-ray structure by John Tainer, The Scripps Research Institute, La Jolla, California. PDBid 4SKN.]

This process is aided by the tight binding of the flipped-out uracil to the enzyme (pull). The exquisite specificity of this binding pocket for uracil prevents the binding and hence hydrolysis of any other base that the enzyme may have induced to flip out. Thus the overall shapes of adenine and guanine exclude them from this pocket, whereas thymine's 5-methyl group is sterically blocked by the rigidly held side chain of Tyr 147. Cytosine, which has approximately the same shape as uracil, is excluded through a set of hydrogen bonds emanating from the protein that mimic those made by adenine in a Watson–Crick A · U base pair.

AP sites in DNA are highly cytotoxic because they irreversibly trap mammalian topoisomerase I in its covalent complex with DNA (Section 29-3Ca). Moreover, since the ribose at the AP site lacks a glycosidic bond, it can readily convert to its linear form (Section 11-1B), whose reactive aldehyde group can cross-link to other cell components. This rationalizes why AP sites remain tightly bound to UDG in solution as well as in crystals. UDG activity is enhanced by AP endonuclease, the next enzyme in the BER pathway, but the two enzymes do not interact in the absence of DNA. This suggests that UDG remains bound to an AP site it generated until it is displaced by the more tightly binding AP endonuclease, thereby protecting the cell from the AP site's cytotoxic effects. It seems likely that other damage-specific DNA glycosylases function similarly.

C. Mismatch Repair

Any replicational mispairing that has eluded the editing functions of the various participating DNA polymerases may still be corrected by a process known as **mismatch repair (MMR)**. For example, *E. coli* Pol I and Pol III have error rates of 10^{-6} to 10^{-7} per base pair replicated but the observed mutational rates in *E. coli* are 10^{-9} to 10^{-10} per base pair replicated. In addition, the MMR system can correct insertions or deletions of up to 4 nt (which arise from the slippage of one strand relative to the other in the active site of DNA polymerase). The importance of MMR is indicated by the fact that defects in the human MMR system result in a high incidence of cancer, most notably **hereditary nonpolyposis colorectal cancer (HNPCC)** (which affects several organs and may be the most common inherited predisposition to cancer).

If an MMR system is to correct errors in replication rather than perpetuate them, it must distinguish the parental DNA, which has the correct base, from the daughter strand, which has an incorrect although normal base. In *E. coli*, as we have seen (Section 30-3C), this is possible because newly replicated GATC palindromes remain hemimethylated until the Dam methyltransferase has had sufficient time to methylate the daughter strand.

E. coli mismatch repair, which was elucidated in large part by Paul Modrich, requires the participation of three proteins and occurs as follows (Fig. 30-61):

1. **MutS** (853 residues) binds to a mismatched base pair or unpaired bases as a homodimer.

2. The MutS–DNA complex binds **MutL** (615 residues), also as a homodimer.

3. The MutS–MutL complex translocates along the DNA in both directions, thereby forming a loop in the DNA. The translocation appears to be driven by the ATPase function of MutS.

4. On encountering a hemimethylated GATC palindrome, the MutS–MutL complex recruits **MutH** (228 residues) and activates this single-strand endonuclease to make a nick on the 5' side of the unmethylated GATC. This GATC may be located on either side of the mismatch and over 1000 bp distant from it, although repair efficiency decreases with the distance between the nick and the mismatch.

5. MutS–MutL recruits UvrD helicase, which in concert with an exonuclease separates the strands and degrades the nicked strand from the nick to beyond the mismatch. If the nick is on the 3' side of the mismatch as shown, the exonuclease is **exonuclease I** (a 3' \rightarrow 5' exonuclease), whereas if the nick is on the 5' side of the mismatch, the exonuclease can be either **RecJ** or **exonuclease VII** (both 5' \rightarrow 3' exonucleases).

The resulting gap is filled in by Pol III and sealed by DNA ligase, thereby correcting the mismatch. MutL is also an ATPase, which, it is postulated, functions to coordinate the various steps of mismatch repair.

Eukaryotic MMR systems are, not surprisingly, more complicated than those of *E. coli*. Eukaryotes express six homologs of MutS and five homologs of MutL that form

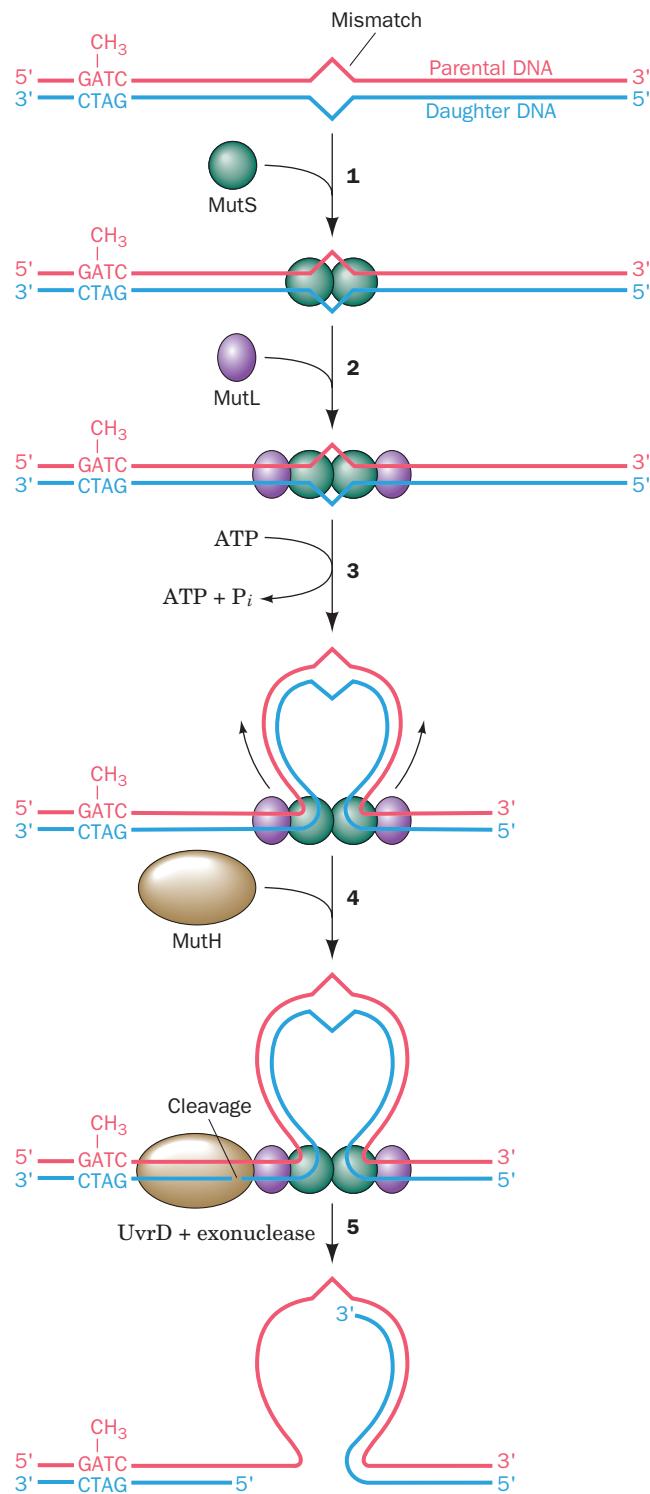


Figure 30-61 The mechanism of mismatch repair in *E. coli*.

heterodimers on mismatched DNA. However, homologs of MutH only occur in gram-negative bacteria. Eukaryotes must have some other way of differentiating the parental and daughter DNA strands. Perhaps a newly synthesized daughter strand is identified by its as-yet unsealed nicks. DNA resynthesis is probably mediated by pol δ .

D. The SOS Response

Agents that damage DNA, such as UV radiation, alkylating agents, and cross-linking agents, induce a complex system of cellular changes in *E. coli* known as the **SOS response**. *E. coli* so treated cease dividing and increase their capacity to repair damaged DNA.

a. LexA Protein Represses the SOS Response

Clues as to the nature of the SOS response were provided by the observations that *E. coli* with mutant *recA* or *lexA* genes have their SOS response permanently switched on. **RecA**, a 353-residue protein that coats DNA as a multimeric helical filament, plays a central role, as we shall see, in homologous recombination (Section 30-6Ab). When *E. coli* are exposed to agents that damage DNA or inhibit DNA replication, their RecA specifically mediates the proteolytic cleavage

of **LexA** (202 residues) between its Asp 84 and Gly 85. RecA is activated to do so on binding to ssDNA (it was initially assumed that RecA catalyzes the proteolysis of LexA but subsequent experiments by John Little indicate that activated RecA stimulates LexA to cleave itself). Further investigations indicated that LexA functions as a repressor of 43 genes that participate in DNA repair and the control of cell division, including *recA*, *lexA*, *uvrA*, and *uvrB*. DNA sequence analyses of the LexA-repressible genes revealed that they are all preceded by a homologous 20-nt sequence, the so-called **SOS box**, that has the palindromic symmetry characteristic of operators (control sites to which repressors bind so as to interfere with transcriptional initiation by RNA polymerase; Section 5-4A). Indeed, LexA has been shown to specifically bind the SOS boxes of *recA* and *lexA*.

The preceding observations suggest a model for the regulation of the SOS response (Fig. 30-62). During normal

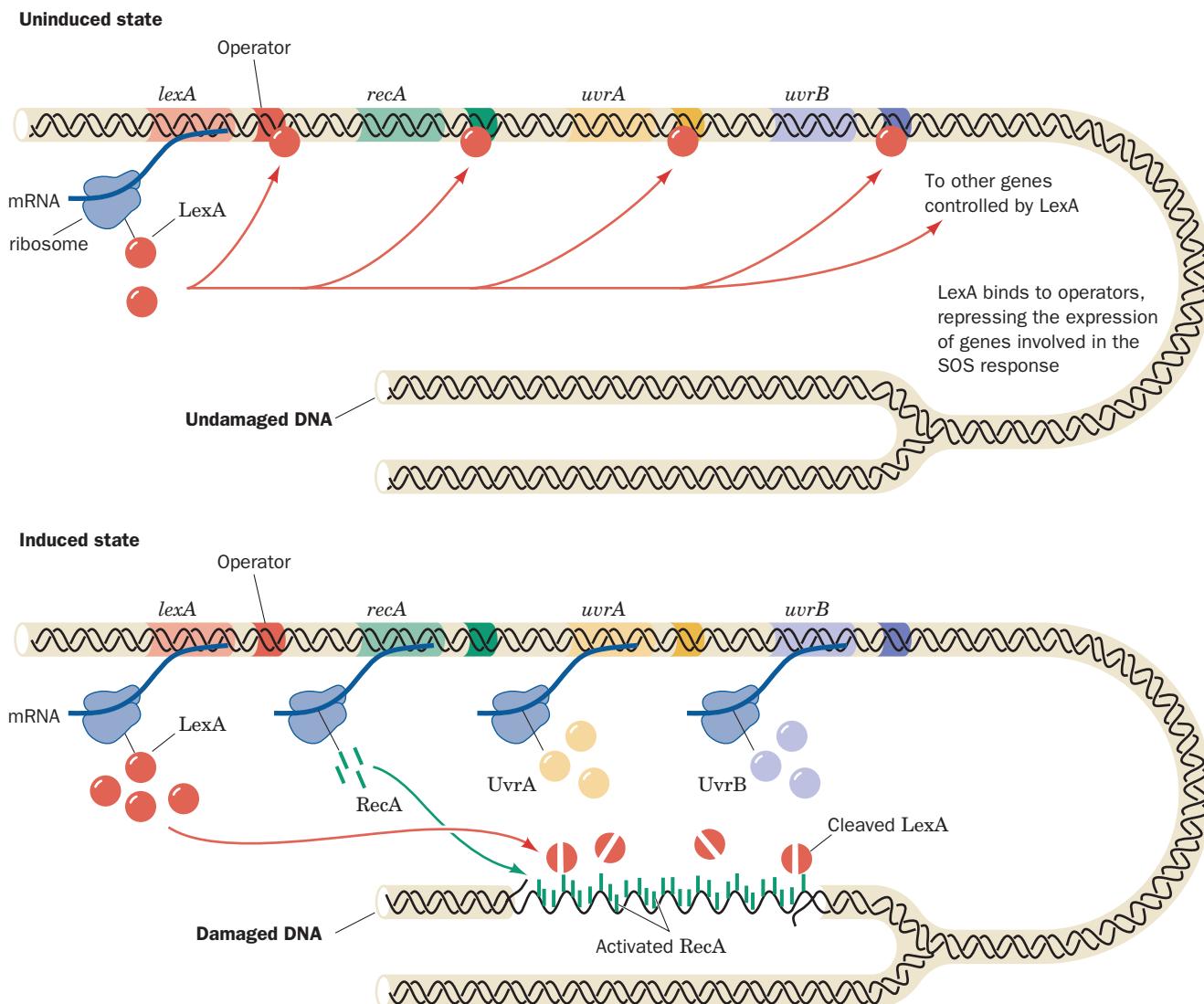


Figure 30-62 Regulation of the SOS response in *E. coli*. In a cell with undamaged DNA (above), LexA largely represses the synthesis of LexA, RecA, UvrA, UvrB, and other proteins involved in the SOS response. When there has been extensive

DNA damage (below), RecA is activated by binding to the resulting single-stranded DNA to stimulate LexA self-cleavage. The consequent synthesis of the SOS proteins results in the repair of the DNA damage.

growth, LexA largely represses the expression of the SOS genes, including the *lexA* gene, by binding to their SOS boxes so as to inhibit RNA polymerase from initiating the transcription of these genes. When DNA damage has been sufficient to produce postreplication gaps, however, this ssDNA binds to RecA so as to stimulate LexA cleavage. The LexA-repressible genes are consequently released from repression and direct the synthesis of SOS proteins including that of LexA (although this repressor continues to be cleaved through the influence of RecA). When the DNA lesions have been eliminated, RecA ceases stimulating LexA's autoproteolysis. The newly synthesized LexA can then function as a repressor, which permits the cell to return to normality.

b. SOS Repair Is Error Prone

The *E. coli* Pol III holoenzyme is unable to replicate through a variety of lesions such as AP sites and thymine dimers. On encountering such lesions, the replisome stalls and disassembles by releasing its Pol III cores, a process that is called replication fork "collapse." Cells have two general modes for restoring collapsed replication forks, **recombination repair** and **SOS repair**. Recombination repair circumvents the damaged template by using a homologous chromosome as its template DNA in a process known as **homologous recombination**, which also functions to generate genetic diversity. Hence we shall postpone our discussion of recombination repair until after our consideration of homologous recombination in Section 30-6A. In the following paragraphs we discuss SOS repair.

In SOS repair, the Pol III core lost from the collapsed replication fork is replaced by one of two so-called **bypass DNA polymerases**, whose synthesis is induced by the SOS response: **DNA polymerase IV (Pol IV)**, the 336-residue product of the *dinB* gene) or **DNA polymerase V [Pol V]**; the heterotrimeric product of the *umuD* and *umuC* genes, **UmuD'2C** (umu for *UV mutagenesis*), where UmuD' is produced by the RecA-assisted self-cleavage of the 139-residue **UmuD** to remove its N-terminal 24 residues, and UmuC consists of 422 residues]. Both of these enzymes are Y-family DNA polymerases, all of whose members lack $3' \rightarrow 5'$ proofreading exonuclease activity and replicate undamaged DNA with poor fidelity and low processivity and hence are also known as **error-prone DNA polymerases**.

Translesion synthesis (TLS) by Pol V, which was characterized in large part by O'Donnell and Myron Goodman, requires the simultaneous presence of the β_2 sliding clamp, the γ complex (clamp loader), and SSB, together with a RecA filament in complex with the ssDNA arising from the action of helicase on the dsDNA ahead of the stalled replication fork. This so-called **Pol V mutasome** tends to incorporate G about half as often as A opposite thymine dimers and AP sites, with pyrimidines being installed infrequently. This process is, of course, highly mutagenic. But even in replicating undamaged DNA, Pol V is at least 1000-fold more error prone than is Pol I or Pol III holoenzyme. However, after synthesizing ~ 7 nt, the Pol V mutasome is replaced by Pol III holoenzyme, which commences normal

DNA replication after the now bypassed lesion. Pol II, a TLS participant that accurately replicates DNA, is also induced by the SOS response but it is synthesized well before Pol V appears (see below). The role of Pol II appears to be the mediation of error-free TLS, and only if this process fails is it replaced by Pol V to carry out error-prone TLS.

There are numerous types of DNA lesions besides AP sites and thymine dimers that interfere with normal DNA replication. Depending on the type of lesion, Pol IV, which is also error prone, may instead be recruited to carry out TLS. With many lesions, TLS may skip over the altered nucleotide, resulting in deletion of one or two bases in the daughter strand opposite the lesion (yielding a **frameshift mutation**, so called because it would change a structural gene's reading frame from that point onward; Section 5-4Bd). Moreover, Pol IV is prone to generating frameshift mutations even when replicating undamaged DNA.

The Y-family DNA polymerase **Dpo4** from the archaeabacterium *Sulfolobus solfataricus* P2, a homolog of *E. coli* Pol IV and Pol V, misincorporates ~ 1 base per 500 replicated nucleotides. The X-ray structure of a complex of Dpo4 with a primer-template DNA that had been incubated with ddATP (which is complementary to the template base), determined by Wei Yang, reveals the structural basis for this low fidelity (Fig. 30-63). The 352-residue protein contains the fingers, palm, and thumb domains

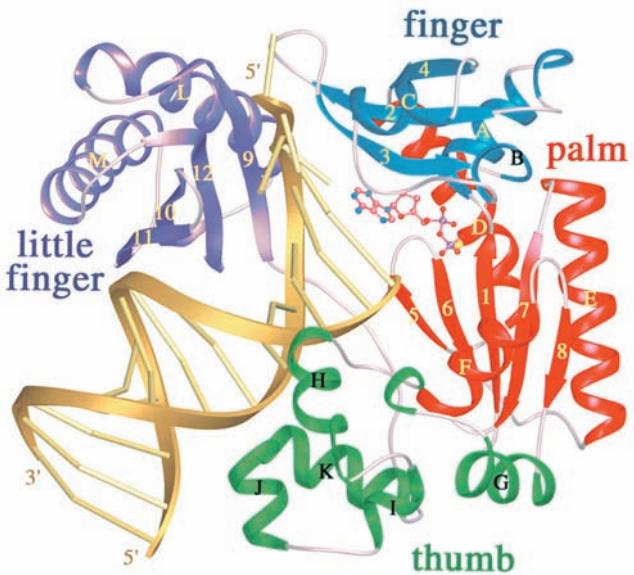


Figure 30-63 X-ray structure of the bypass DNA polymerase Dpo4 from *Sulfolobus solfataricus* P2 in complex with a primer-template DNA and ddADP. The protein is drawn in ribbon form with its fingers, palm, thumb, and little finger domains blue, red, green, and purple, respectively. The DNA is gold with its backbones drawn as ribbons and its bases represented by rods. The ddADP, which is base-paired to a template T in the enzyme's active site, is shown in ball-and-stick form colored according to atom type (C pink, N blue, O red, and P magenta). [Courtesy of Wei Yang, NIH, Bethesda, Maryland. PDBid 1JX4.]

common to all known DNA polymerases (although their orders differ in the sequences of the different families of DNA polymerases) and, in addition, has a C-terminal domain unique to Y-family DNA polymerases that has been dubbed the “little finger” domain. The enzyme, as expected, has incorporated a ddA residue at the 3' end of the primer and, in addition, binds a ddADP in base-paired complex to the new template T. The little finger domain binds in the major groove of the DNA. However, the fingers and thumb domains are small and stubby compared to those of replicative DNA polymerases such as Klentaq1 (Fig. 30-9) and pol δ (Fig. 30-41), and the residues that contact the base pair in the active site are all Gly and Ala rather than the Phe, Tyr, and Arg that mainly do so in the replicative DNA polymerases. Moreover, the bound DNA is entirely in the B form rather than in the A form at the active site as occurs in many replicative DNA polymerases. Since the minor groove is more accessible in A-DNA than in B-DNA (Section 29-1B), this suggests that error-prone DNA polymerases have relatively little facility to monitor the base-pairing fidelity of the incoming nucleotide. This accounts for the ability of error-prone DNA polymerases to accommodate distorted template DNA as well as non-Watson–Crick base pairs at their active sites.

SOS repair is an error-prone and hence mutagenic process. It is therefore a process of last resort that is only initiated \sim 50 min after SOS induction if the DNA has not already been repaired by other means. Yet, DNA damage that normally activates the SOS response is nonmutagenic in the *recA*⁺ *E. coli* that survive. This is, as we saw, because bypass DNA polymerases will replicate over a DNA lesion even when there is no information as to which bases were originally present. Indeed, *most mutations in E. coli arise from the actions of the SOS repair system*, which is therefore a testimonial to the proposition that survival with a chance of loss of function (and the possible gain of new ones) is advantageous, in the Darwinian sense, over death, although only a small fraction of cells actually survive this process. It has therefore been suggested that, under conditions of environmental stress, the SOS system functions to increase the rate of mutation so as to increase the rate at which the *E. coli* adapt to the new conditions. Finally, it should be noted that the eukaryotic pols η , ι , and κ , all Y-family members, and pol ζ , an X-family member, are implicated in TLS and that pol η , the product of the *XPV* gene, is defective in the XPV form of xeroderma pigmentosum (Section 30-5Bb).

E. Double-Strand Break Repair

Double-strand breaks (DSBs) in DNA are produced when a replication fork encounters a nick and by the reactive oxygen species (ROS) by-products of oxidative metabolism and ionizing radiation (which also produces ROS). In fact, around 5 to 10% of dividing cells in culture exhibit at least one chromosome break at any given time. Moreover, DSBs are normal intermediates in certain specialized cellular processes such as recombination during meiosis (Sec-

tion 1-4Ab) and **V(D)J recombination** in lymphoid cells, which helps generate the vast diversity of antigen-binding sites in antibodies and T-cell receptors (Section 35-2C). Unrepaired or misrepaired DSBs can be lethal to cells or cause chromosomal aberrations that may lead to cancer. Hence, the efficient repair of DSBs is essential for cell viability and genomic integrity.

Cells have two general modes to repair DSBs: recombination repair, which only occurs during the late S and G₂ phases of the cell cycle (when sister chromatids are present to serve as templates), and **nonhomologous end-joining (NHEJ)**, which functions throughout the cell cycle. Here we discuss NHEJ, a process which, as its name implies, directly rejoins DSBs. The recombination repair of DSBs is discussed in Section 30-6Ag.

In NHEJ, the broken ends of the DSB must be aligned, its frayed ends trimmed and/or filled in, and their strands ligated. The core NHEJ machinery in eukaryotes includes the DNA end-binding protein **Ku** (a heterodimer of homologous 70- and 83-kD subunits, **Ku70** and **Ku80**), **DNA ligase IV**, and the accessory protein **Xrc4**. Ku, an abundant nuclear protein, binds to a DSB, whether blunt or with an overhang, and hence appears to be the cell's primary DSB sensor. The X-ray structure of Ku in complex with a 14-bp DNA, determined by Jonathan Goldberg, reveals that the protein cradles the dsDNA segment along its entire length and encircles its central \sim 3 bp segment (Fig. 30-64). The protein ring is also present in the closely

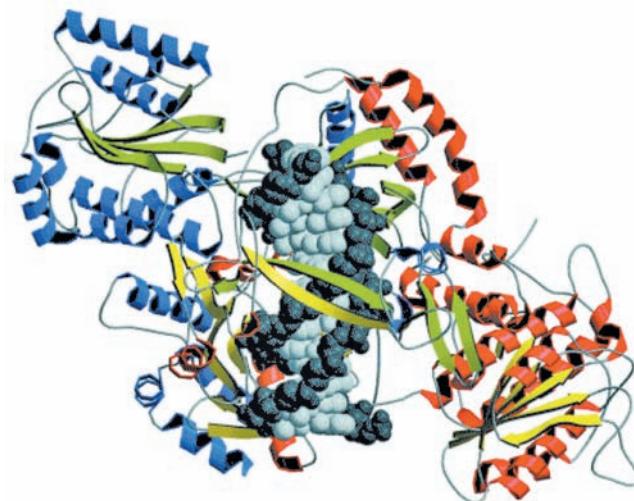


Figure 30-64 X-ray structure of human Ku protein in complex with DNA containing 14 bp. The subunits of Ku70 (red helices and yellow strands) and Ku80 (blue helices and green strands) are viewed along the pseudo-2-fold axis relating them. The DNA, viewed with its DSB pointing upward, is drawn in space-filling form with its sugar-phosphate backbone dark gray and its base pairs light gray. Note that the DNA is surrounded by a ring of protein. [Courtesy of John Tainer, The Scripps Research Institute, La Jolla, California. Based on an X-ray structure by Jonathan Goldberg, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1JEY.]

similar X-ray structure of Ku alone, thereby explaining why Ku that is bound to a dsDNA, which is then circularized, becomes permanently associated with it. Ku makes no specific contacts with the DNA's bases and few with its sugar-phosphate backbone, but instead fits snugly into the DNA's major and minor grooves so as to precisely orient it.

Ku–DNA complexes have been shown to dimerize so as to align the members of a DSB, both blunt ended and with short (1–4 bp) complementary single strands, for ligation as is diagrammed in Fig. 30-65. The DNA ends are exposed along one face of each Ku–DNA complex, presumably making them accessible to polymerases that fill in gaps and to nucleases that trim excess and inappropriate ends preparatory for ligation by DNA ligase IV in complex with Xrcc4. Nucleotide trimming, which of course generates mutations, appears to be carried out in

an ATP-dependent manner by the evolutionarily conserved **Mre11 complex**, which consists of two **Mre11** nuclelease subunits and two **Rad50** ATPase subunits. Ku is eventually released from the rejoined DNA, perhaps by proteolytic cleavage.

The reason that the mutations generated by NHEJ are usually not unacceptably deleterious is that only a small fraction of the mammalian genome is expressed (Section 34-2A). In fact, the genome in a somatic cell of a 70-year-old human typically contains ~2000 “scars” caused by NHEJ.

F. Identification of Carcinogens

Many forms of cancer are known to be caused by exposure to certain chemical agents that are therefore known as **carcinogens**. It has been estimated that as much as 80% of human cancer arises in this fashion. There is considerable evidence that the primary event in carcinogenesis is often damage to DNA (carcinogenesis is discussed in Section 34-4C). Carcinogens are consequently also likely to induce the SOS response in bacteria and thus act as indirect mutagenic agents. In fact, there is a high correlation between carcinogenesis and mutagenesis (recall, e.g., the progress of xeroderma pigmentosum; Section 30-5B).

There are presently over 80,000 man-made chemicals of commercial importance and ~1000 new ones are introduced each year. The standard animal tests for carcinogenesis, exposing rats or mice to high levels of the suspected carcinogen and checking for cancer, are expensive and require ~3 years to complete. Thus, relatively few substances have been tested in this manner.

a. The Ames Test Assays for Probable Carcinogenicity

Bruce Ames devised a rapid and effective bacterial assay for carcinogenicity that is based on the high correlation between carcinogenesis and mutagenesis. He constructed special tester strains of *Salmonella typhimurium* that are *his*[−] (cannot synthesize histidine so that they are unable to grow in its absence), have cell envelopes that lack the lipopolysaccharide coating that renders normal *Salmonella* impermeable to many substances (Section 11-3Bc), and have inactivated excision repair systems. Mutagenesis in these tester strains is indicated by their reversion to the *his*⁺ phenotype.

In the **Ames test**, ~10⁹ tester strain bacteria are spread on a culture plate that contains only a small amount of histidine to permit the bacteria to initially grow and mutate. Usually a mixture of several *his*[−] strains is used so that mutations due to both base changes and nucleotide insertions or deletions can be detected. A mutagen placed in the culture medium causes some of these *his*[−] bacteria to revert to the *his*⁺ phenotype, which is detected by their growth into visible colonies after 2 days at 37°C (Fig. 30-66). The mutagenicity of a substance is scored as the number of such colonies less the few spontaneously revertant colonies that occur in the absence of the mutagen.

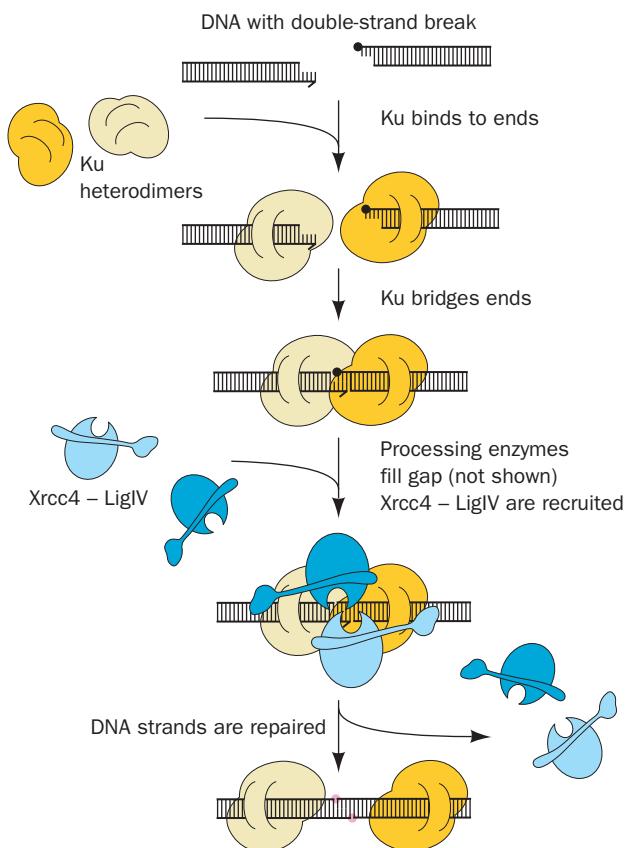


Figure 30-65 Schematic diagram of nonhomologous end-joining (NHEJ). The left dsDNA fragment is missing a base and the right fragment is blocked by a nonligatable group (filled black circle). The two Ku heterodimers are drawn in two shades of yellow and the Xrcc4–DNA ligase IV complexes are drawn in two shades of blue. The newly repaired links in the DNA are represented by pink circles. [After Jones, J.M., Gellert, M., and Yang, W., *Structure* 9, 881 (2001).]

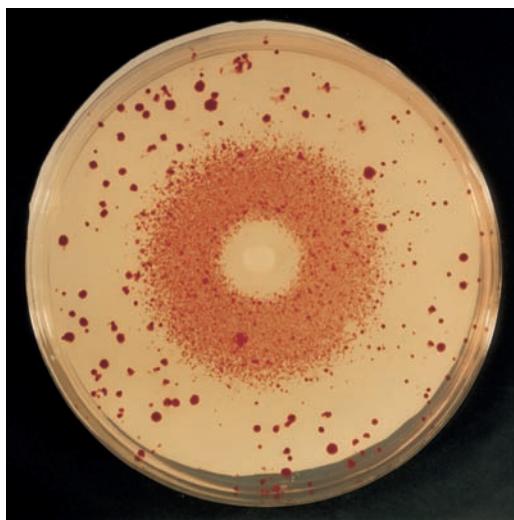


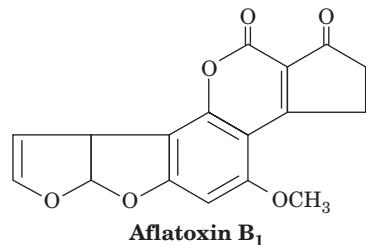
Figure 30-66 The Ames test for mutagenesis. A filter paper disk containing a mutagen, in this case the alkylating agent ethyl methanesulfonate, is centered on a culture plate containing *his*[−] tester strains of *Salmonella typhimurium* in a medium that initially contained only a small amount histidine. A dense halo of revertant bacterial colonies appears around the disk from which the mutagen diffused. The larger colonies distributed about the culture plate are spontaneous revertants. The bacteria near the disk have been killed by the toxic mutagen's high concentration. [Courtesy of Raymond Devoret, Institut Curie, Orsay, France.]

Many noncarcinogens are converted to carcinogens in the liver or in other tissues via a variety of detoxification reactions (e.g., those catalyzed by the cytochromes P450; Section 15-4Bc). A small amount of rat liver homogenate is therefore included in the Ames test medium in an effort to approximate the effects of mammalian metabolism.

b. Both Man-Made and Naturally Occurring Substances Can Be Carcinogenic

There is an ~80% correspondence between the compounds determined to be carcinogenic by animal tests and those found to be mutagenic by the Ames test. Dose-response curves, which are generated by testing a given compound at a number of concentrations, are almost always linear and extrapolate back to zero, indicating that *there is no threshold concentration for mutagenesis*. Several compounds to which humans have been extensively exposed that were found to be mutagenic by the Ames test were later found to be carcinogenic in animal tests. These include tris(2,3-dibromopropyl)phosphate, which was used as a flame retardant on children's sleepwear in the mid-1970s and can be absorbed through the skin; and furylfuranide, which was used in Japan in the 1960s and 1970s as an antibacterial additive in many prepared foods (and had passed two animal tests before it was found to be mutagenic). Carcinogens are not confined to man-made compounds but also occur in nature. For example, carcinogens are contained in many plants that are

common in the human diet, including alfalfa sprouts. **Aflatoxin B₁,**



one of the most potent carcinogens known, is produced by fungi that grow on peanuts and corn. Charred or browned food, such as occurs on broiled meats and toasted bread, contains a variety of DNA-damaging agents. Thus, with respect to carcinogenesis, as Ames has written, "Nature is not benign."

6 RECOMBINATION AND MOBILE GENETIC ELEMENTS

The chromosome is not just a simple repository of genetic information. If this were so, the unit of mutation would have to be an entire chromosome rather than a gene because there would be no means of separating a mutated gene from the other genes of the same chromosome. Chromosomes would therefore accumulate deleterious mutations until they became nonviable.

It has been known from some of the earliest genetic studies that pairs of allelic genes may exchange chromosomal locations by a process known as **genetic recombination** (Section 1-4Cb). Mutated genes can thereby be individually tested, since their propagation is then not absolutely dependent on the propagation of the genes with which they had been previously associated. In this section, we consider the mechanisms by which genetic elements can move, both between chromosomes and within them.

A. Homologous Recombination

Homologous recombination (also called **general recombination**) is defined as the exchange of homologous segments between two DNA molecules. Both genetic and cytological studies have long indicated that such a crossing-over process occurs in higher organisms during meiosis (Fig. 1-27). Bacteria, which are normally haploid, likewise have elaborate mechanisms for the interchange of genetic information. They can acquire foreign DNA through transformation (Section 5-2A), through a process called **conjugation** (mating) in which DNA is directly transferred from one cell to another via a cytoplasmic bridge (Section 31-1Ac), and via **transduction** in which a defective bacteriophage that has erroneously acquired a segment of bacterial DNA rather than the viral chromosome transfers this DNA to another bacterial cell. In all of these processes, the foreign DNA is installed in the recipient's chromosome or plasmid

through homologous recombination (to be propagated, a DNA segment must be part of a replicon; that is, be associated with a replication origin such as occurs in a chromosome, a plasmid, or a virus).

a. Recombination Occurs via a Crossed-Over Intermediate

The prototypical model for homologous recombination (Fig. 30-67) was proposed by Robin Holliday in 1964 on the basis of genetic studies on fungi. The corresponding strands of two aligned homologous DNA duplexes are nicked, and the nicked strands cross over to pair with the nearly complementary strands of the homologous duplex after which the nicks are sealed (Fig. 30-67a–e), thereby yielding a four-way junction known as a **Holliday junction** (Fig. 30-67e). A Holliday junction has, in fact, been observed in the X-ray structure of d(CCGGTACCGG), determined Shing Ho (Fig. 30-68), in which, perhaps unexpectedly, all the bases

form normal Watson–Crick base pairs without any apparent strain. The crossover point can move in either direction, often thousands of nucleotides, in a process known as **branch migration** (Fig. 30-67e,f) in which the four strands exchange base-pairing partners.

A Holliday junction can be resolved into two duplex DNAs in two equally probable ways (Fig. 30-67g–l):

1. The cleavage of the strands that did not cross over (right branch of Fig. 30-67j–l) exchanges the ends of the original duplexes to form, after nick sealing, the traditional recombinant DNA (Fig. 1-27b).

2. The cleavage of the strands that crossed over (left branch of Fig. 30-67j–l) exchanges a pair of homologous single-stranded segments.

The recombination of circular duplex DNAs results in the types of structures diagrammed in Fig. 30-69. Electron microscopic evidence for the existence of the postulated

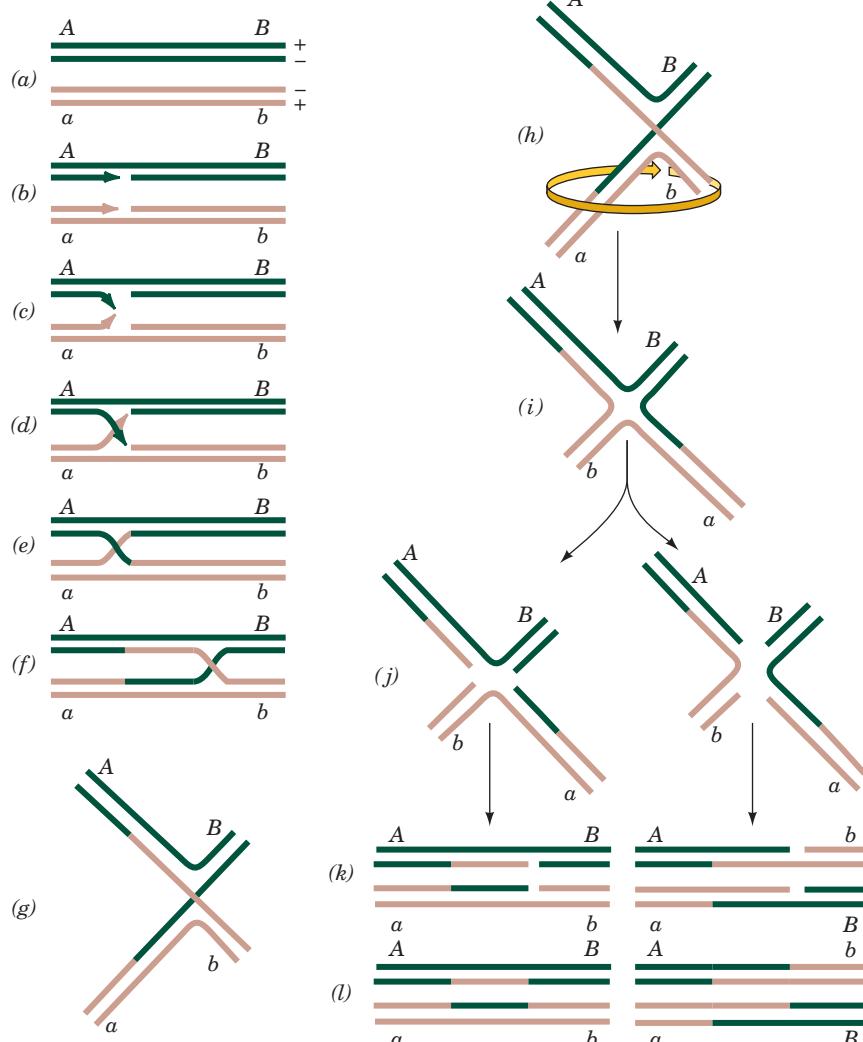


Figure 30-67 The Holliday model of homologous recombination between homologous DNA duplexes.

See the Animated Figures

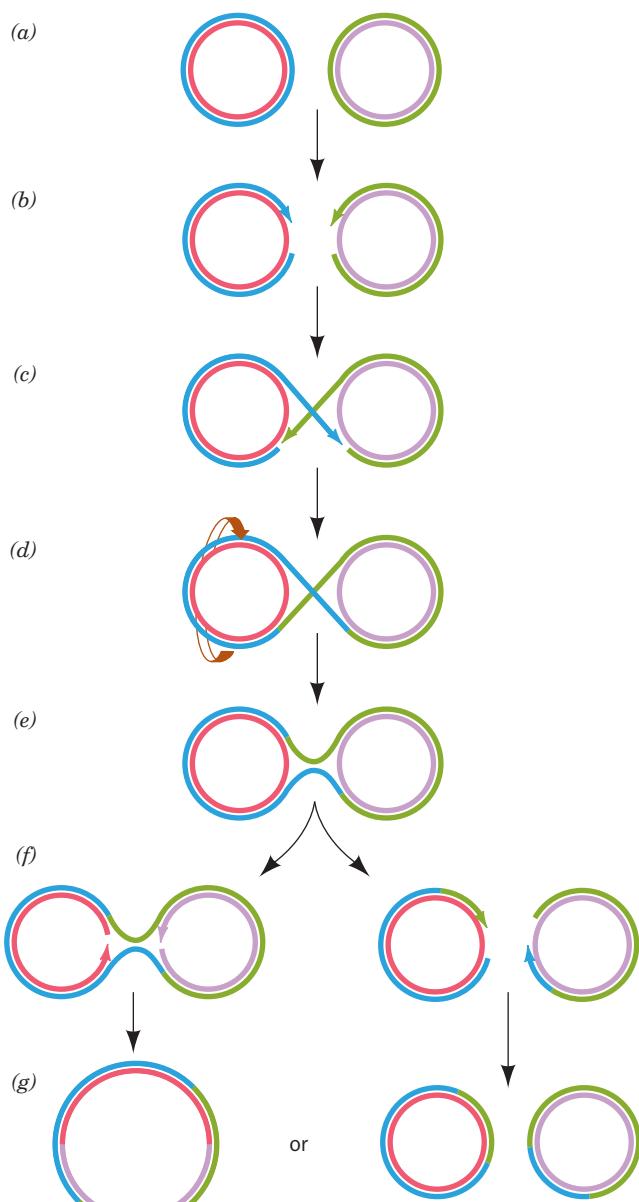
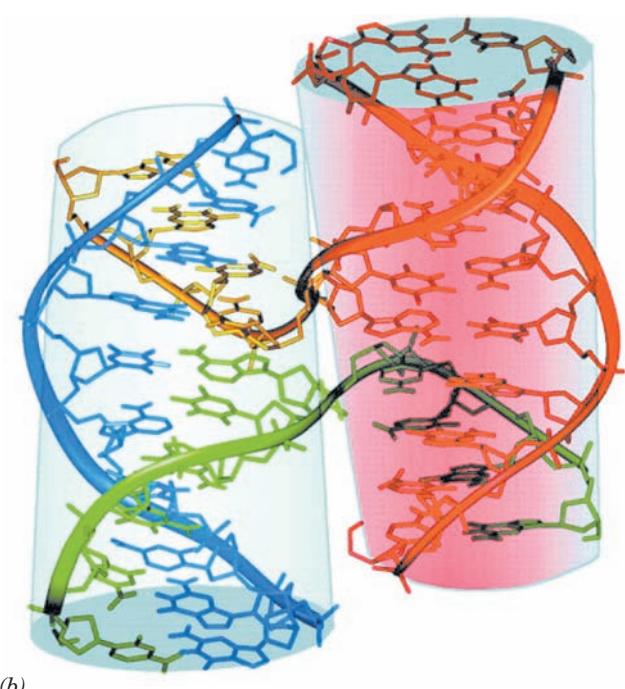
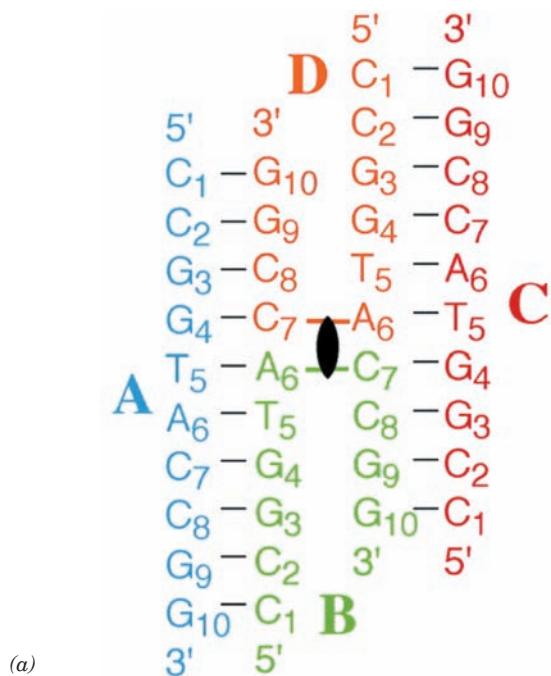
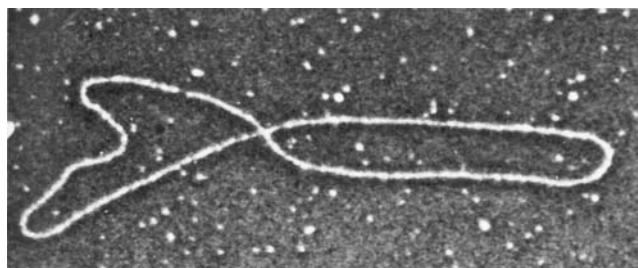


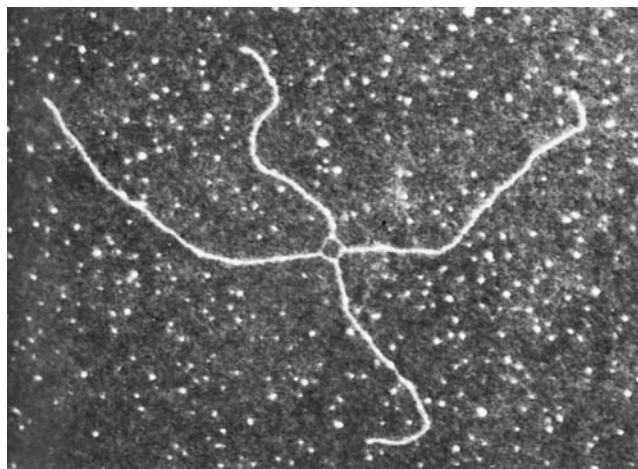
Figure 30-69 Homologous recombination between two circular DNA duplexes. This process can result either in two circles of the original sizes or in a single composite circle.

Figure 30-68 X-ray structure of the self-complementary decamerid DNA d(CCGGTACCGG).

(a) The secondary structure of the four-stranded Holliday junction formed by this sequence in which the four strands, A, B, C, and D, are differently colored, their nucleotides are numbered 1 to 10 from their 5' to 3' termini, and Watson-Crick base-pairing interactions are represented by black dashes. The 2-fold axis relating the two helices of this so-called **stacked-X conformation** is represented by the black lenticular symbol. (b) The observed three-dimensional structure of the Holliday junction, as viewed along its 2-fold axis, in which the oligonucleotides are represented in stick form with their backbones traced by ribbons, all colored as in Part a. With the exception of the backbones of strands B and D at the crossovers, the two arms of this structure each form an undistorted B-DNA helix, including the stacking of the base pairs flanking the crossovers. The two helices are inclined to each other by 41°. Note that Fig. 30-67g is a schematic representation of the stacked-X conformation as viewed perpendicular to both helices (from the side in this drawing and hence having the projected appearance of the letter X). A Holliday junction can also assume a so-called **open-X conformation**, which is represented by Fig. 30-67i. [Courtesy of Shing Ho, Oregon State University. PDBid 1DCW.]



(a)



(b)

Figure 30-70 Electron micrographs of intermediates in the homologous recombination of two plasmids. (a) A figure-8 structure. This corresponds to Fig. 30-69d. (b) A chi structure that results from the treatment of a figure-8 structure with a restriction endonuclease. Note the thinner single-stranded connections in the crossover region. [Courtesy of Huntington Potter, University of South Florida, and David Dressler, Oxford University, U.K.]

“figure-8” structures is shown in Fig. 30-70a. These figure-8 structures were shown not to be just twisted circles by cutting them with a restriction endonuclease to yield **chi structures** (after their resemblance to the Greek letter χ) such as that pictured in Fig. 30-70b.

b. Homologous Recombination in *E. coli* Is Catalyzed by RecA

The observation that *recA*⁻ *E. coli* have a 10⁴-fold lower recombination rate than the wild-type indicates that *RecA* protein has an important function in recombination. Indeed, RecA greatly increases the rate at which complementary strands renature *in vitro*. This versatile protein (recall it also stimulates the autoproteolysis of LexA to trigger the SOS response and is an essential participant in the translesion synthesis of DNA; Section 30-5D) polymerizes cooperatively without regard to base sequence on ssDNA or on dsDNA that has a single-stranded gap. The resulting filaments, which may contain up to several thousand RecA monomers, specifically bind the homologous dsDNA, and, in an ATP-dependent reaction, catalyze strand exchange.

EM studies by Edward Egelman revealed that RecA filaments bound to ssDNA or dsDNA form a right-handed helix with \sim 6.2 RecA monomers per turn and a pitch (rise per turn) of 95 Å. The DNA in these filaments binds to the protein with 3 nt (or bp) per RecA monomer and hence is underwound with \sim 18.5 nt (or bp) per turn (vs 10 bp per turn for canonical B-DNA).

The formation of RecA–DNA filaments is highly cooperative; it requires five or six RecA protomers to form a stable assembly. Consequently, attempts to crystallize RecA–DNA filaments over many years were unsuccessful. Nikola Pavletich ingeniously solved this conundrum by linking five or six *E. coli* RecA genes (each corresponding to residues 1–335 of this 353-residue protein) in tandem via 14-residue linkers and mutating the first and last RecA so as to prevent them from forming longer filaments. These fusion proteins, which had DNA-dependent ATPase and strand-exchange activities comparable to that of monomeric RecA, formed crystals containing both ssDNA and dsDNA.

The X-ray structure of the $\text{RecA}_5\text{--}(\text{ADP--AlF}_4^-)_5\text{--}(\text{dT})_{15}\text{--}(\text{dA})_{12}$ complex (Fig. 30-71; ADP–AlF₄⁻ is a nonhy-

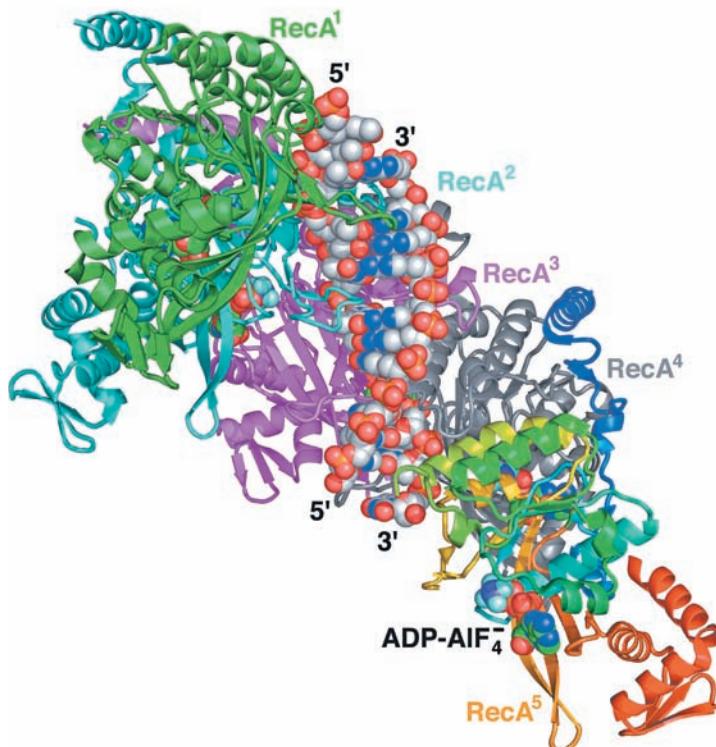


Figure 30-71 X-ray structure of the $\text{RecA}_5\text{--}(\text{ADP--AlF}_4^-)_5\text{--}(\text{dT})_{15}\text{--}(\text{dA})_{12}$ complex viewed with its filament axis vertical. The RecA units RecA¹ (the N-terminal unit) through RecA⁴ are colored green, cyan, magenta, and gray, respectively, with the C-terminal unit, RecA⁵, colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA and ADP–AlF₄⁻ are drawn in space-filling form with DNA C gray, ADP C green, N blue, O red, P orange, F light blue, and Al purple. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 3CMX.]

drolyzable ATP analog) exhibits a straight filament axis with overall helical parameters that are closely similar to those derived from EM studies. Each RecA unit consists of a largely helical 30-residue N-terminal segment, a 240-residue α/β ATPase core, and a 64-residue C-terminal globular domain. The linkers connecting adjacent RecA units are disordered. Each RecA unit makes extensive contacts with its nearest neighbors so as to form a filament with a deep helical groove that exposes the DNA bound in its interior (Fig. 30-71 is viewed looking into this groove).

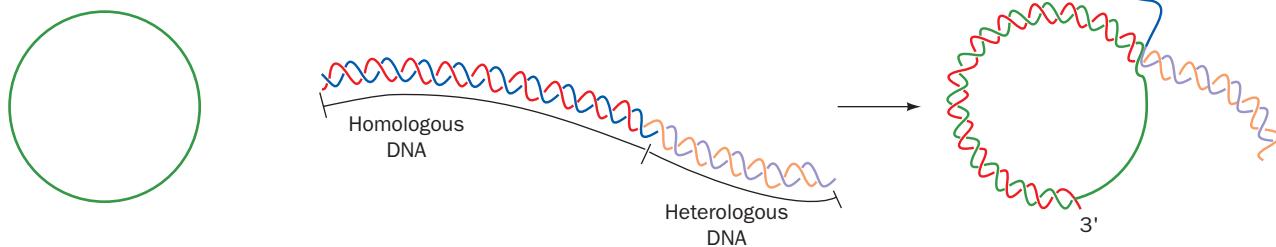
The two DNA strands, which lie close to the filament axis, form a complete set of Watson–Crick base pairs. However, rather than being smoothly stretched out, as had been expected, the dsDNA assumes an irregular conformation in which each 3-bp segment that is bound to a RecA unit closely resembles B-DNA with the steps between successive base pairs in this triplet having an axial rise of ~ 3.4 Å and a helical twist of $\sim 30^\circ$ (vs 3.4 Å and 36° for canonical B-DNA; Table 29-1). In contrast, the step between successive base pair triplets has an axial rise of 8.4 Å and a helical twist of -4° , thereby forming a 5-Å-high gap between successive triplets that is partially filled by the side chain of the conserved Ile 199. The sugar–phosphate backbone of the DNA strand furthest from the viewer in Fig. 30-71 [the (dT)₁₅]

makes extensive contacts with RecA. In contrast, the other strand [the (dA)₁₂] makes few contacts with the protein; it is held in place almost entirely by base pairing with the first strand. The ADP–AlF₄⁻ is sandwiched between adjacent α/β ATPase cores, where it is completely buried.

The X-ray structure of the ssDNA-containing RecA₆–(ADP–AlF₄⁻)₆–(dT)₁₈ complex closely resembles that of the foregoing dsDNA-containing complex but with the absence of the DNA strand closest to the viewer in Fig. 30-71. Thus, each RecA unit binds a (dT)₃ segment that is held in a B-DNA-like conformation with successive (dT)₃ segments separated by a 7.8-Å axial rise.

How does RecA mediate DNA strand exchange between single-stranded and duplex DNAs? On encountering a dsDNA with a strand that is complementary to its bound ssDNA, RecA partially unwinds the duplex and, in a reaction driven by RecA-catalyzed ATP hydrolysis, exchanges the ssDNA with the corresponding strand on the duplex. *This process tolerates only a limited degree of mispairing and requires that one of the participating DNA strands have a free end.* The assimilation (exchange) of a single-stranded circle with a strand on a linear duplex (Fig. 30-72) cannot proceed past the 3' end of a highly mismatched segment in the complementary strand. *The invasion of the single strand must*

Assimilation of 3' end of homologous DNA



No assimilation of noncomplementary DNA

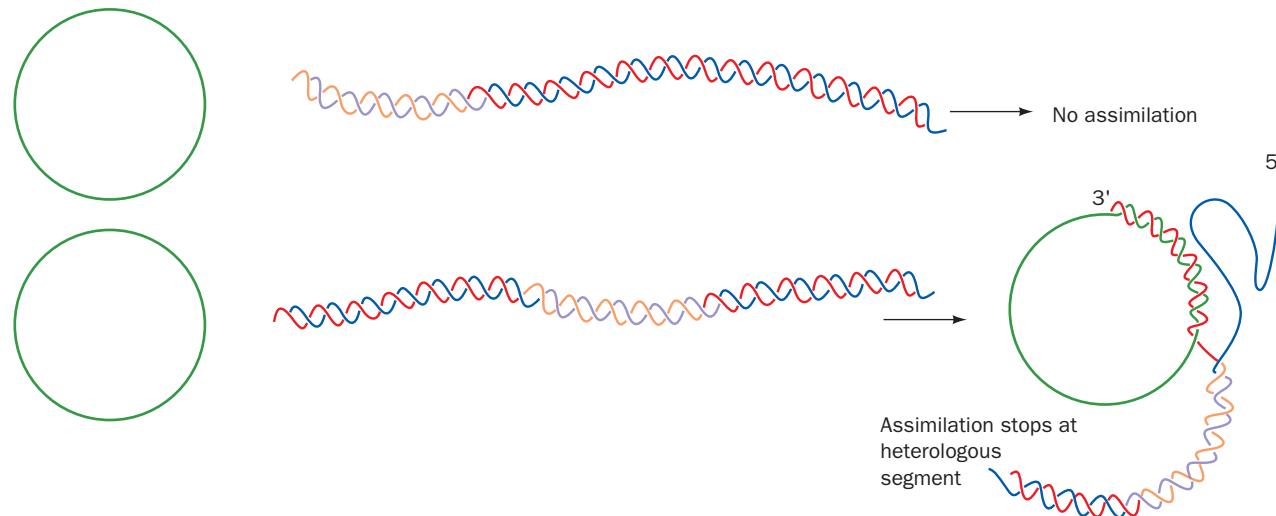


Figure 30-72 The RecA-catalyzed assimilation of a single-stranded circle by a dsDNA can occur only if the dsDNA has a 3' end that can base-pair with the circle (red strand).

Strand assimilation cannot proceed through a noncomplementary segment (purple and orange strands).

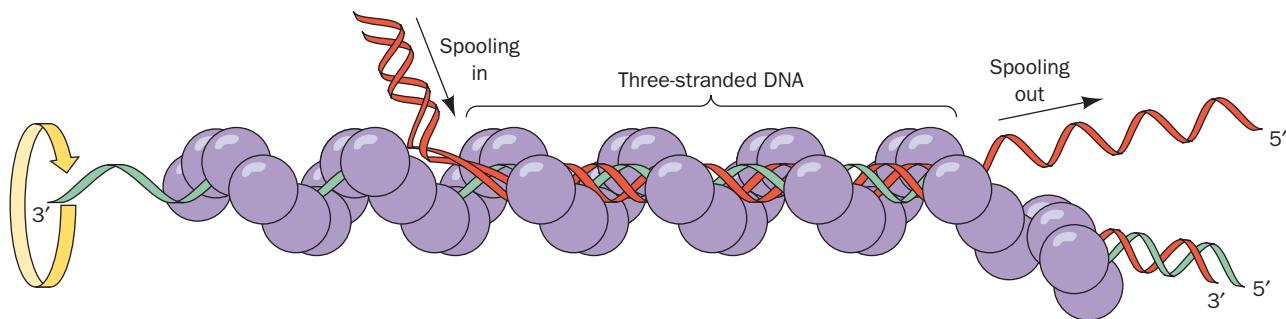


Figure 30-73 Hypothetical model for the RecA-mediated strand exchange reaction. Homologous DNA molecules are paired in advance of strand exchange in a three-stranded helix. The ATP-driven rotation of the RecA filament about its helix

axis would cause duplex DNA to be “spooled in” to the filament, right to left as drawn. [After West, S.C., *Annu. Rev. Biochem.* **61**, 617 (1992).]

therefore begin at its 5' end. A model for the consequent branch migration process is diagrammed in Fig. 30-73. Of course, two such strand exchange processes must occur simultaneously in a Holliday junction (Figs. 30-67 and 30-69).

The above structures suggest that the fidelity of homologous recombination arises from the B-DNA-like conformation that RecA imposes on the otherwise flexible bound ssDNA strand, which would exclude non-Watson–Crick base pairs. Strand exchange, of course, requires the separation of the two strands of the incoming dsDNA to permit one of its strands to sample base pairing with the ssDNA substrate. The above structures suggest that this is facilitated by the disruption of base stacking between base pair triplets in the RecA–DNA complex. However, the structure of the triple helical DNA intermediate in the strand exchange reaction is, as yet, unknown.

c. Eukaryotes Have RecA-Like Proteins

Yeast **RAD51** (339 residues) functions in the ATP-dependent repair and recombination of DNA in much the same way as does the 30% homologous *E. coli* RecA protein. The electron micrograph-based image reconstruction of RAD51 in complex with double-stranded DNA is nearly identical to that of RecA at low resolution: Both complexes form helical filaments in which the DNA has an $\sim 5.1\text{-}\text{\AA}$ rise per base pair and 18.6 bp per turn. Since RAD51 homologs occur in chickens, mice, and humans, it is very likely that such filaments universally mediate DNA repair and recombination.

d. RecBCD Initiates Recombination by Making Single-Strand Nicks

The single-strand nicks to which RecA binds are made by the **RecBCD** protein, the 330-kD heterotrimeric product of the SOS genes *recB*, *recC*, and *recD*. RecB is both a 3' \rightarrow 5' helicase and a nuclease, whereas RecD is a 5' \rightarrow 3' helicase. The formation of a RecA binding site begins with RecBCD binding to the end of a dsDNA and then unwinding it via its two ATP-driven helicase functions (Fig. 30-74). As it does so, RecB nucleolytically degrades the unwound single strands behind it, with the 3'-ending strand being

cleaved more often and hence broken down to smaller fragments than the 5'-ending strand. However, on RecC encountering the sequence GCTGGTGG from its 3' end (the so-called **Chi sequence**, which occurs about every ~ 5 kb in the *E. coli* genome), the enzyme pauses and ceases its cleavage of the 3'-ending strand but increases the rate at which it cleaves the 5'-ending strand. RecBCD then helps load RecA onto the 3'-ending strand before dissociating from the DNA.

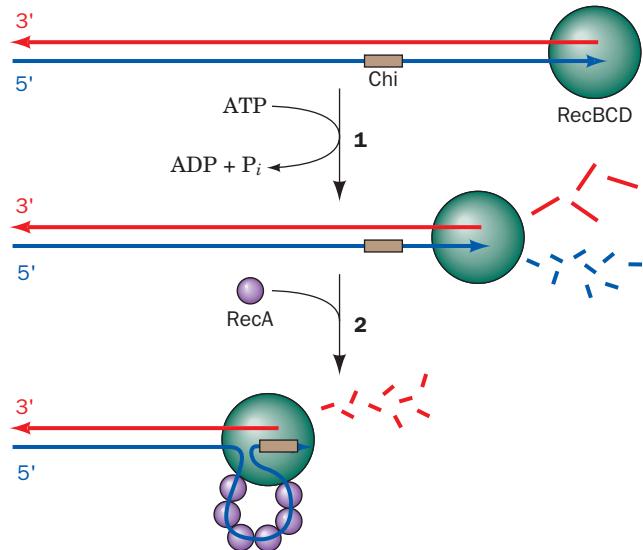


Figure 30-74 The generation of a 3'-ending single-strand DNA segment by RecBCD to initiate recombination. (1) RecBCD binds to a free end of a dsDNA and, in an ATP-driven process, advances along the helix, unwinding the DNA and degrading the resulting single strands behind it, with the 3'-ending strand cleaved more often than the 5'-ending strand. (2) When RecBCD encounters a properly oriented Chi sequence, it binds it and thus stops cleaving the 3'-ending strand but increases the frequency at which it cleaves the 5'-ending strand. This generates the potentially invasive 3'-ending strand segment to which RecA binds.

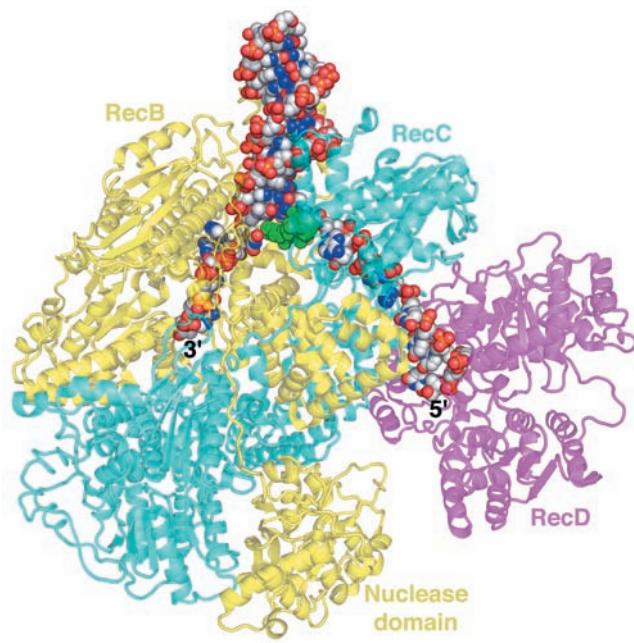


Figure 30-75 X-ray structure of *E. coli* RecBCD in complex with a 51-nt DNA capable of forming a 21-bp hairpin loop. The protein is drawn in semitransparent ribbon form with RecB yellow, RecC cyan, and RecD magenta. Note how the RecB nuclelease domain is linked to the rest of the subunit by an extended polypeptide tether. The DNA is shown in space-filling form with C gray, N blue, O red, and P orange. A loop from RecC, which is drawn in space-filling form in green, is situated so as to wedge apart the two strands of the incoming dsDNA with the 3'-ending strand binding to the 3' → 5' helicase of RecB and the 5'-ending strand passing through RecC to bind to the 5' → 3' helicase of RecD. [Based on an X-ray structure by Dale Wigley, The London Research Institute, Herts, U.K. PDBid 3K70.]

Dale Wigley determined the X-ray structure of *E. coli* RecBCD in complex with a 51-nt DNA that could form a hairpin loop containing an up to 21-bp dsDNA stem (Fig. 30-75). The structure shows that RecB (1180 residues) and RecC (1122 residues) are intimately intertwined with RecB's C-terminal nuclelease domain connected to the rest of the subunit by an extended 21-residue polypeptide tether. A 15-bp segment of dsDNA enters the protein through a tunnel between RecB and RecC. There it encounters a loop from RecC that appears to wedge the two strands apart, with the 6-nt 3'-ending single strand of the DNA binding to RecB and the 10-nt 5'-ending single strand binding to RecD (608 residues; the 5-nt loop connecting the two strands of the dsDNA at the top of Fig. 30-75 is disordered). The structure explains the different rates of cleavage of the two DNA strands. The 3'-ending strand emerges from a tunnel through RecC in the vicinity of the RecB nuclelease domain, which is positioned to processively cleave it. The 5'-ending strand competes with the 3'-ending strand for the nuclease site, but since the 5'-ending strand is less favorably located, it is cleaved less frequently. However, after RecD has bound a Chi sequence,

the 3'-ending strand is no longer available for cleavage, which permits the nuclease to cleave the 5'-ending strand more frequently.

RecBCD can only commence unwinding DNA at a free duplex end. Such ends are not normally present in *E. coli*, which has a circular genome, but become available during such recombinational processes as bacterial transformation, conjugation, and viral transduction, as well as at collapsed replication forks.

e. RuvABC Mediates the Branch Migration and the Resolution of the Holliday Junction

The branch migration of the RecA-generated Holliday junction (Fig. 30-67e,f) requires the breaking and reforming of base pairs as the bases exchange partners in passing from one double helical stem to the other. Since $\Delta G = 0$ for this process, it was initially assumed that it occurs spontaneously. However, such a process would move forward and backward at random and, moreover, would be blocked by as little as a single mismatched base pair. In *E. coli*, and most other bacteria, branch migration is an ATP-driven unidirectional process that is mediated by two proteins whose synthesis is induced by the SOS response (Section 30-5D): **RuvB** (336 residues; Ruv for repair of UV damage), an ATP-powered pump that drives branch migration but binds only weakly to DNA; and **RuvA** (203 residues), which binds to both a Holliday junction and to RuvB, thereby targeting **RuvB** to the DNA.

The X-ray structure of *Mycobacterium leprae* (the cause of leprosy) RuvA in complex with a synthetic and immobile Holliday junction (Fig. 30-76a), determined by Morikawa, reveals that RuvA forms a homotetramer to which the Holliday junction binds in its open-X conformation (Fig. 30-76b). The RuvA tetramer, which has the appearance of a four-petaled flower (it has C_4 symmetry rather than the D_2 symmetry of the vast majority of homotetramers), is relatively flat ($80 \times 80 \times 45 \text{ \AA}$) with one square face concave and the other convex. The concave face (that facing the viewer in Fig. 30-76b), which is highly positively charged and is studded with numerous conserved residues, has four symmetry-related grooves that bind the Holliday junction's four arms. This face's centrally located projection or "pin" is formed by the side chains of Glu 55 and Asp 56 from each subunit, and hence the repulsive forces between them and the Holliday junction's anionic phosphate groups probably facilitate the separation of the single-stranded DNA segments and guide them from one double helix to another.

RuvB is a member of the AAA+ family of ATPases (Section 30-2Ca). The X-ray structure of *Thermus thermophilus* RuvB crystallized in the presence of both ADP and AMPPNP, determined by Morikawa, reveals two molecules of RuvB with somewhat different conformations: one binding ADP and the other binding AMPPNP. Each RuvB molecule consists of three consecutive domains arranged in a crescentlike configuration with the adenine nucleotides binding at the interface between its N-terminal and middle domains. EM studies indicate that, in the presence of dsDNA, RuvB oligomerizes to form a hexamer (Fig. 30-77a),

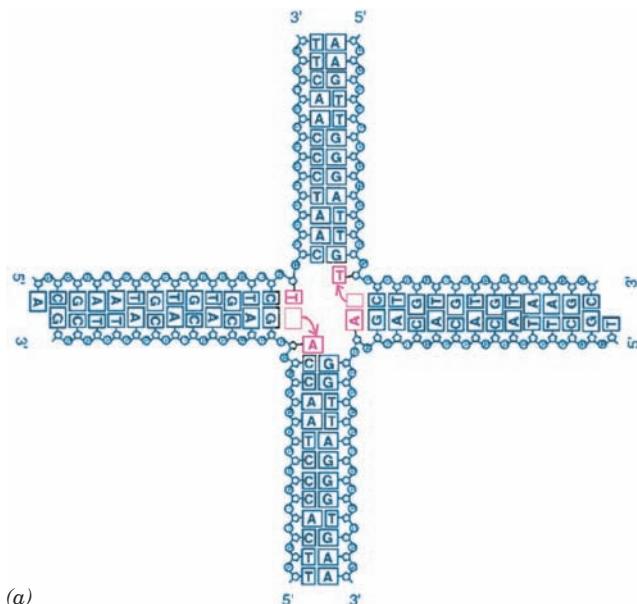
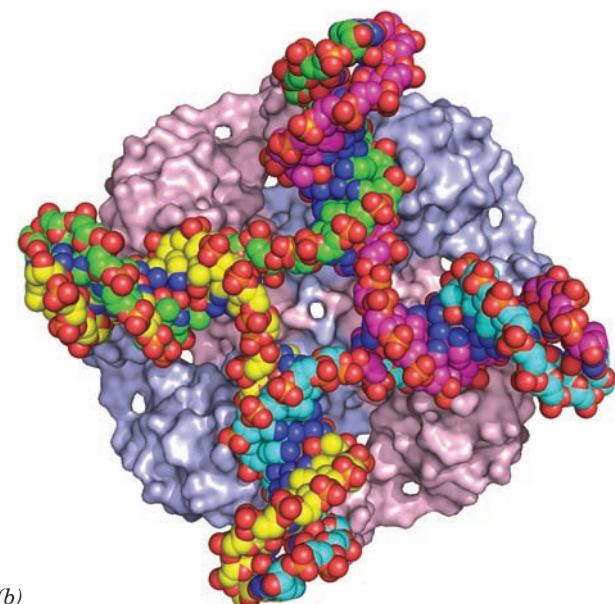


Figure 30-76 X-ray structure of a RuvA tetramer in complex with a Holliday junction. (a) A schematic drawing of the synthetic and immobile Holliday junction in this structure showing its base sequence. The two A · T base pairs that are disrupted at the crossover (and which, if the Holliday junction consisted of two homologous dsDNAs, as it normally does, would exchange base pairing partners) are magenta. (b) The RuvA–Holliday junction complex as viewed along the protein



tetramer's 4-fold axis. The protein is represented as its molecular surface with its subunits alternately colored pink and light blue. The DNA is drawn in space-filling form colored according to atom type with C atoms in different chains in different colors, N blue, O red, and P orange. [Part a courtesy of and Part b based on an X-ray structure by Kosuke Morikawa, Biomolecular Engineering Research Institute, Osaka, Japan. PDBid 1C7Y.]

as do most other AAA+ family members, including the D2 domain of NSF (Fig. 12-78). A hexameric model of RuvB (Fig. 30-77b), constructed by superimposing the N-terminal

domain of the RuvB monomer on the ATPase domains of the NSF D2 hexamer, agrees well with the EM-based image and contains no serious steric clashes. This 130-Å-diameter

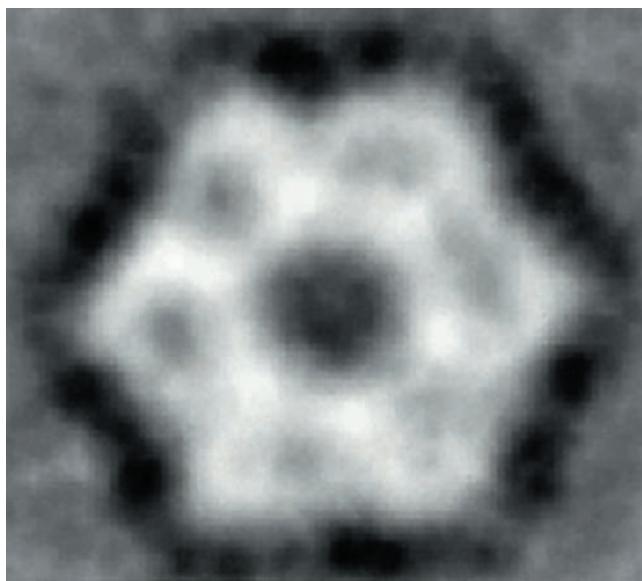
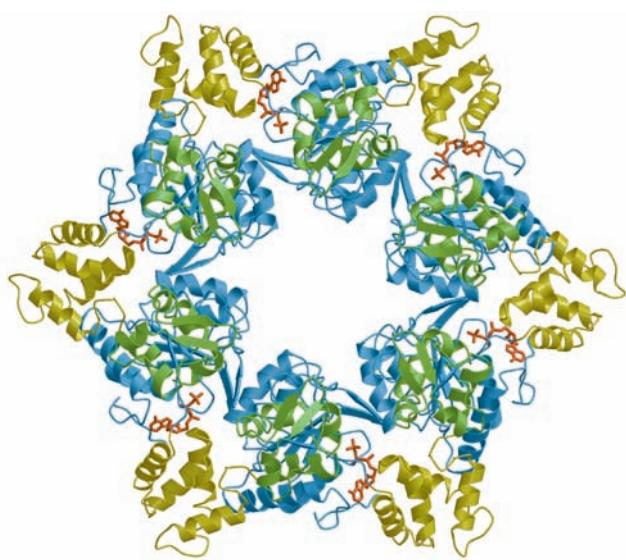


Figure 30-77 Proposed structure of the *T. thermophilus* RuvB hexamer. (a) An EM-based image reconstruction of RuvB complexed with a 30-bp DNA (not visible) as viewed along its 6-fold axis. The image resolution is 30 Å. (b) A model of the RuvB hexamer that was constructed from the X-ray structure of RuvB monomers by superimposing their N-terminal domains on



the homologous ATPase domains of the NSF D2 homohexamer (Fig. 12-78). The N-terminal, middle, and C-terminal domains are blue, yellow, and green, respectively, and its bound AMPPNP is drawn in stick form in red. [Courtesy of Kosuke Morikawa, Biomolecular Engineering Research Institute, Osaka, Japan. PDBid 1HQC.]

hexameric model contains a 30-Å-diameter hole through which a single dsDNA can readily be threaded (see below). Moreover the six β hairpins, one per monomer, that have been implicated in binding to RuvA are located on the top face of the hexamer (as pictured in Fig. 30-77b).

The EM images of the RuvAB–Holliday junction complex indicate that RuvA binds two oppositely located RuvB hexamers. This has led to the model of their interaction depicted in Fig. 30-78 in which RuvA binds the Holliday junction and helps load the RuvB hexameric rings onto two opposing arms of the Holliday junction. The two hexameric rings are postulated to counter-rotate, each in the anticlockwise direction looking toward the center of the junction, so as to screw the horizontal DNA strands through the center of the junction and into the top and bottom double helices, thereby effecting branch migration (although rather than actually rotating relative to RuvA, a RuvB hexamer might pull the dsDNA through its central hole by “walking” up its grooves in a manner resembling that postulated for hexagonal helicases; Section 30-2Ca). The direction of branch migration depends on which pair of arms the RuvB hexamers are loaded.

The final stage of homologous recombination is the resolution of the Holliday junction into its two homologous dsDNAs. This process is carried out by **RuvC**, a homodimeric endonuclease of 173-residue subunits whose X-ray structure indicates that its active sites are located ~30 Å apart on the same face of the protein. This suggests that RuvC sits down on the open face of the RuvAB–Holliday junction complex, that facing the viewer in Fig. 30-78, to

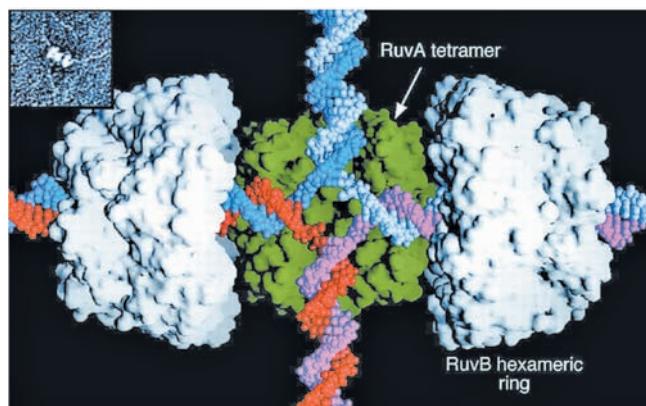


Figure 30-78 Model of the RuvAB–Holliday junction complex. The model is based on electron micrographs such as that in the inset. The proteins are represented by their surface diagrams with the RuvA tetramer, as seen in its X-ray structure, green and the two oppositely oriented RuvB hexamers white. The DNA of the Holliday junction is drawn in space-filling form with its homologous blue and pink strands complementary to its light blue and red strands. The complex is postulated to drive branch migration via the ATP-driven counter-rotation of the RuvB hexamers relative to the RuvA tetramer. This pumps (screws) the horizontal dsDNAs through the RuvB hexamers to the center of the Holliday junction, where their strands separate and then base-pair with their homologs to form new dsDNAs, which are pumped out vertically. [Courtesy of Peter Artymiuk, University of Sheffield, U.K.]

cleave oppositely located strands at the Holliday junction. The resulting single-strand nicks in the now resolved dsDNAs are sealed by DNA ligase.

The X-ray structure of RuvC in complex with DNA has not been determined, although model building studies suggest that it binds Holliday junction DNA in its stacked-X conformation. However, Dietrich Suck determined the X-ray structure of bacteriophage **T4 endonuclease VII** in complex with a Holliday junction in the stacked-X conformation (Fig. 30-79). RuvC and the 157-residue T4 endonuclease VII exhibit no structural similarity but both are homodimers of relatively small subunits that have similar

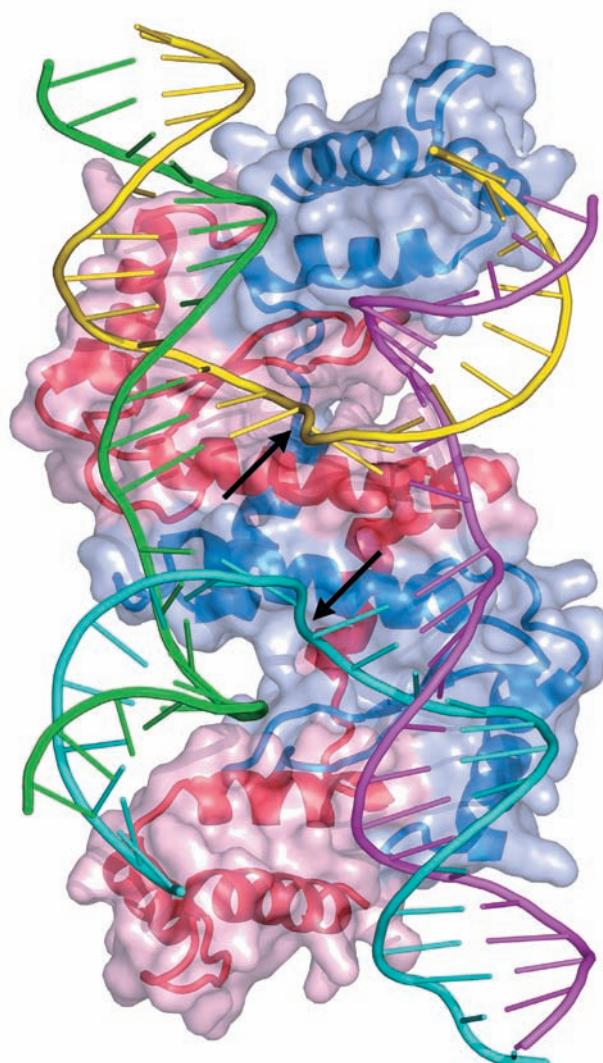


Figure 30-79 X-ray structure of bacteriophage T4 endonuclease VII resolving a Holliday junction as viewed along its pseudo-2-fold axis. The Holliday junction DNA is drawn in ladder form with each of its four different 24-nt strands differently colored. The protein, a homodimer of 157-residue subunits, is shown in ribbon form embedded in its semitransparent molecular surface with one subunit red and the other blue. The arrows indicate the symmetrically located DNA cleavage sites. Compare the DNA in this structure to that in Fig. 30-68b. [Based on an X-ray structure by Dietrich Suck, European Molecular Biology Laboratory, Heidelberg, Germany. PDBid 2QNC.]

functions: the resolution of Holliday junctions into two duplex DNAs by introducing symmetrically placed nicks in equivalent strands (Fig. 30-67j).

The forgoing model of the **RuvABC resolvosome** provides a satisfying mechanism for branch migration and Holliday junction resolution. However, there is a fly in this particular ointment. The X-ray structure of an *M. leprae* RuvA–Holliday junction complex crystallized under conditions different from that in Fig. 30-76, determined by Laurence Pearl, resembles the complex in Fig. 30-76b but with a second RuvA tetramer in face-to-face contact with the concave (DNA-binding) side of the first. Hence, the Holliday junction is contained in two intersecting tunnels running through the resulting RuvA octamer. Are both RuvA–Holliday junction structures biologically relevant, or is one an artifact of crystallization? Pearl argues that the extensive complementary contacts between the two RuvA tetramers, which are strongly conserved, are unlikely to be artifactual and that a single RuvA tetramer is unlikely to withstand the torque exerted by the two (in effect) counter-rotating RuvB hexamers. However, if the RuvA octamer is biologically relevant, one of its tetramers would at some point have to dissociate in order to allow RuvC access to the Holliday junction. Yet, modeling studies indicate that the RuvC dimer cannot properly contact the RuvB tetramer-bound Holliday junction without it changing from its open-X to its stacked-X conformation. Further investigations are necessary to resolve these inconsistencies.

f. Recombination Repair Reconstitutes Damaged Replication Forks

Transformation, transduction, and conjugation are such rare events that the vast majority of bacterial cells never participate in these processes. Similarly, the only place in the metazoan life cycle at which gene shuffling through homologous recombination occurs is in meiosis (Section 1-4A). Why then do nearly all cells have elaborate systems for mediating homologous recombination? It is because damaged replication forks occur at a frequency of at least once per bacterial cell generation and perhaps 10 times per eukaryotic cell cycle. The DNA lesions that damage the replication forks can be circumvented via homologous recombination in a process named **recombination repair** [translesion synthesis, which is highly mutagenic, is a process of last resort (Section 30-5Db)]. Indeed, the rates of synthesis of RuvA and RuvB are greatly enhanced by the SOS response. Thus, as Michael Cox pointed out, *the primary function of homologous recombination is to repair damaged replication forks*. In what follows, we describe recombination repair as it occurs in *E. coli*.

Recombination repair is called into play when a replication fork encounters an unrepaired single-strand lesion (Fig. 30-80):

1. DNA replication is arrested at the lesion but continues on the opposing undamaged strand for some distance before the replisome fully collapses (Section 30-5Db).

2. The replication fork regresses to form a type of Holliday junction dubbed a “chicken foot.” This process may occur spontaneously as driven by the positive supercoiling that has built up ahead of the replication fork, it may be mediated by RecA, or it may be promoted by **RecG**, an

ATP-driven helicase that catalyzes branch migration at DNA junctions with three or four branches.

3. The single-strand gap at the collapsed replication fork, now an overhang, is filled in by Pol I.

4. Reverse branch migration mediated by RuvAB or RecG yields a reconstituted replication fork, which supports replication restart (see below).

Note that this process does not actually repair the single-strand lesion that has caused the problem but instead reconstructs the replication fork in a way that permits the previously discussed DNA repair systems (Section 30-5) to eventually eliminate the lesion.

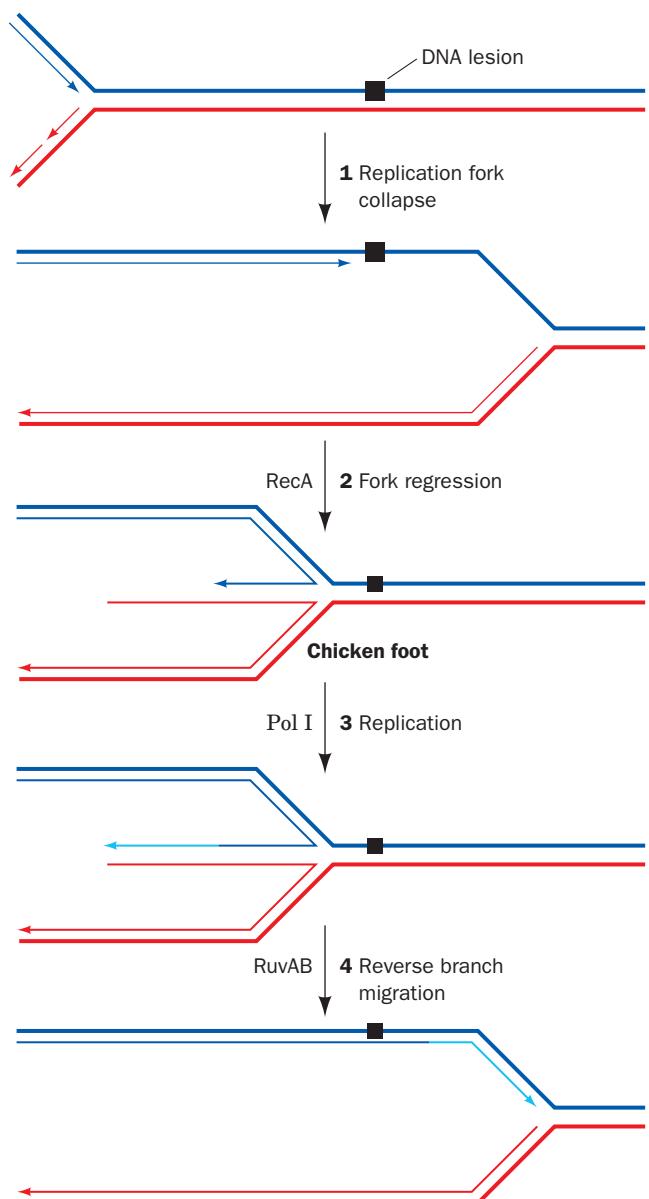


Figure 30-80 The recombination repair of a replication fork that has encountered a single-strand lesion. Thick lines indicate parental DNA, thin lines indicate newly synthesized DNA, the cyan lines indicate DNA that was synthesized by Pol I, and the arrows point in the 5' → 3' direction. [After Cox, M.M., *Annu. Rev. Genet.* 35, 53 (2001).]

A second situation that requires recombination repair is the encounter of a replication fork with an unrepaired single-strand nick (Fig. 30-81):

1. When a single-strand nick is encountered, the replication fork collapses.
2. The repair process begins via the RecBCD plus RecA-mediated invasion of the newly synthesized and undamaged 3'-ending strand into the homologous dsDNA starting at its broken end.
3. Branch migration, as mediated by RuvAB, then yields a Holliday junction, which exchanges the replication fork's 3'-ending strands.

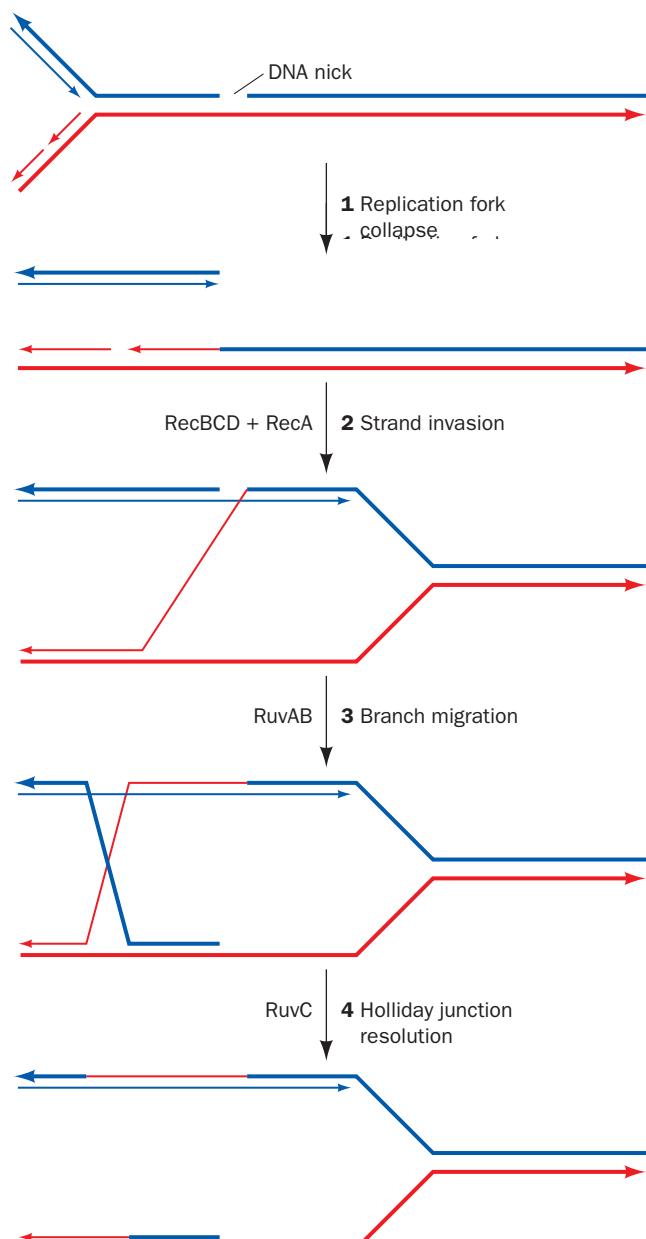


Figure 30-81 The recombination repair of a replication fork that has encountered a single-strand nick. Thick lines indicate parental DNA, thin lines indicate newly synthesized DNA, and the arrows point in the 5' → 3' direction. [After Cox, M.M., *Annu. Rev. Genet.* **35**, 53 (2001).]

4. RuvC then resolves the Holliday junction yielding a reconstituted replication fork ready for replication restart.

Thus, the 5'-ending strand of the nick has, in effect, become the 5' end of an Okazaki fragment.

The final step in the recombination repair process is the restart of DNA replication. This process is, of necessity, distinct from the replication initiation that occurs at *oriC* (Section 30-3Ca). **Origin-independent replication restart** is mediated by the same seven-protein primosome that initiates the minus strand replication of bacteriophage ϕ X174 (Table 30-4), which has therefore been named the **restart primosome**.

g. Recombination Repair Reconstitutes Double-Strand Breaks

We have seen that double-strand breaks (DSBs) in DNA can be rejoined, often mutagenically, by nonhomologous end-joining (NHEJ; Section 30-5E). DSBs may also be nonmutagenically repaired through a recombination repair process known as homologous end-joining, which occurs via two Holliday junctions (Fig. 30-82):

1. The DSB's double-stranded ends are resected to produce single-stranded ends. One of the 3'-ending strands invades the corresponding sequence of a homologous

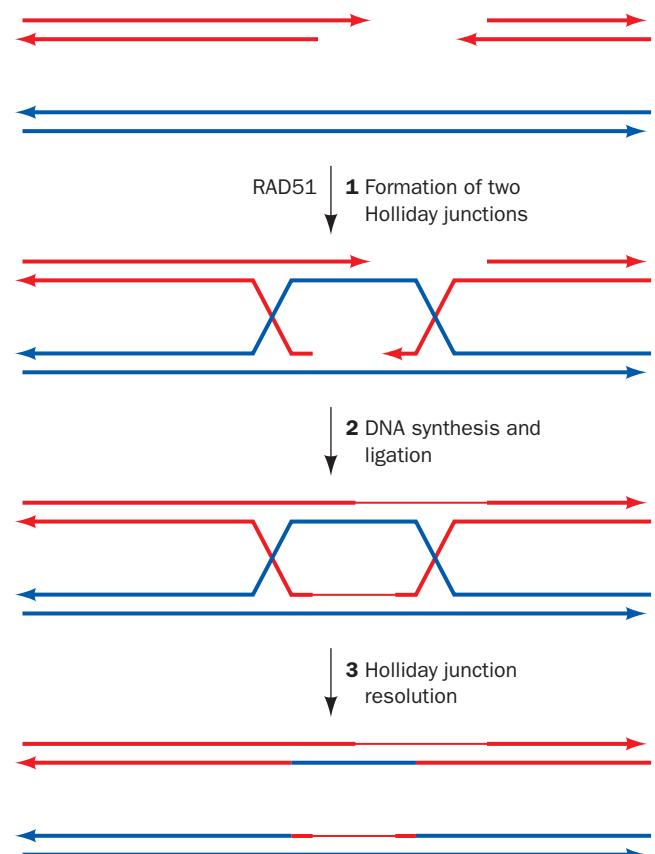


Figure 30-82 The repair of a double-strand break in DNA by homologous end-joining. Thick lines indicate parental DNA, thin lines indicate newly synthesized DNA, and the arrows point in the 5' → 3' direction. [After Haber, J.E., *Trends Genet.* **16**, 259 (2000).]

Table 30-6 Properties of Some Insertion Elements

Insertion Element	Length (bp)	Inverted Terminal Repeat (bp)	Direct Repeat at Target (bp)	Number of Copies in <i>E. coli</i> Chromosome
IS1	768	23	9	5–8
IS2	1327	41	5	5
IS4	1428	18	11–13	5
IS5	1195	16	4	1–2

Source: Mainly Lewin, B., *Genes IX*, p. 524, Oxford University Press (2008).

chromosome to form a Holliday junction, a process that, in eukaryotes, is mediated by the RecA homolog RAD51. The other 3'-ending strand pairs with the displaced strand segment on the homologous chromosome to form a second Holliday junction.

2. DNA synthesis and ligation fills in the gaps and seals the joints.

3. Both Holliday junctions are resolved to yield two intact double strands.

Thus, the sequences that may have been expunged in the formation of the DSB are copied from the homologous chromosome. Of course, a limitation of homologous end-joining, particularly in haploid cells, is that a homologous chromosomal segment may not be available.

The importance of recombination repair in humans is demonstrated by the observation that defects in the proteins **BRCA1** (1863 residues) and **BRCA2** (3418 residues), both of which interact with RAD51, are associated with a greatly increased incidence of breast, ovarian, prostate, and pancreatic cancers. Indeed, individuals with mutant *BRCA1* or *BRCA2* genes have up to an 80% lifetime risk of developing cancer. Recombination can also function to elongate shortened telomeres without the need for telomerase.

B. Transposition and Site-Specific Recombination

In the early 1950s, on the basis of genetic analysis, Barbara McClintock reported that the variegated pigmentation pattern of maize (Indian corn) kernels results from the action of genetic elements that can move about the maize genome. This proposal was resoundingly ignored because it was contrary to the then held genetic orthodoxy that chromosomes consist of genes linked in fixed order. Another 20 years were to pass before evidence of mobile genetic elements was found in another organism, *E. coli*.

It is now known that **transposable elements** or **transposons** are common in both prokaryotes and eukaryotes, where they influence the variation of phenotypic expression over the short term and evolutionary development over the long term. Each transposon codes for the enzymes that specifically insert it into the recipient DNA. This process has been described as **illegitimate recombination** because it requires no homology between donor and

recipient DNAs. Since the insertion site is chosen largely at random, transposition is a potentially dangerous process; the insertion of a transposon into an essential gene will kill a cell together with its resident transposons. Hence transposition is tightly regulated; it occurs at a rate of only 10^{-5} to 10^{-7} events per element per generation. The conditions that trigger transposition are, for the most part, unknown.

a. Prokaryotic Transposons

Prokaryotic transposons with three levels of complexity have been characterized:

1. The simplest transposons, and the first to be characterized, are named **insertion sequences** or **IS elements**. They are designated by “IS” followed by an identifying number. IS elements are normal constituents of bacterial chromosomes and plasmids. For example, a common *E. coli* strain has eight copies of **IS1** and five copies of **IS2**. IS elements generally consist of <2000 bp. These comprise a so-called **transposase** gene, and in some cases a regulatory gene, flanked by short inverted (having opposite orientation) terminal repeats (Fig. 30-83 and Table 30-6). The inverted repeats are essential for transposition; their genetic alteration invariably prevents this process. An inserted IS element is flanked by a directly (having the same orientation) repeated segment of host DNA (Fig. 30-83). This suggests that an IS element is inserted in the host DNA at a staggered cut that is later filled in (Fig. 30-84). The length of this target sequence (most commonly 5 to 9 bp), but not its sequence, is characteristic of the IS element.

2. More complex transposons carry genes not involved in the transposition process, for example, antibiotic resistance genes. Such transposons are designated “Tn” followed

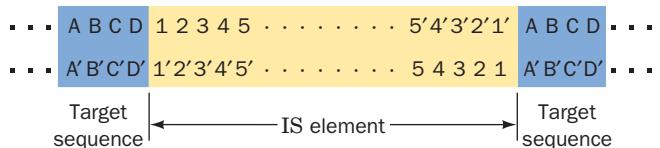


Figure 30-83 Structure of IS elements. These and other transposons have inverted terminal repeats (numerals) and are flanked by direct repeats of host DNA target sequences (letters).

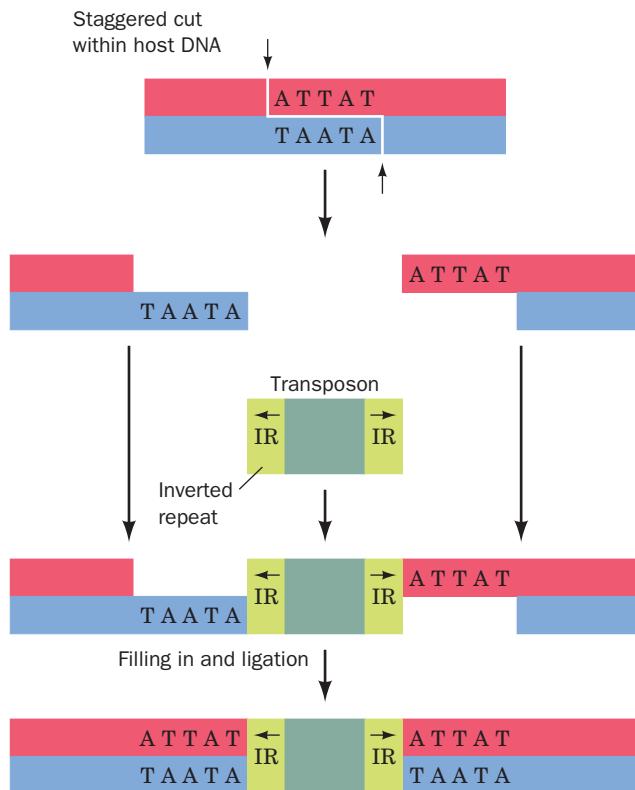


Figure 30-84 A model for the generation of direct repeats of the target sequence by transposon insertion.

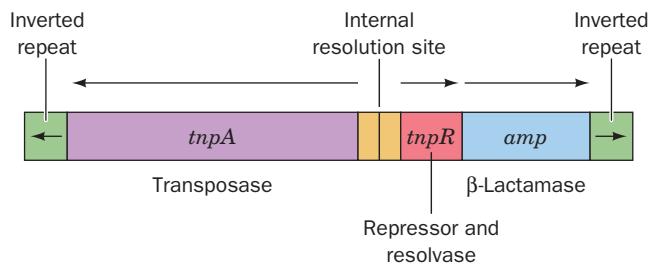


Figure 30-85 A map of transposon Tn3.

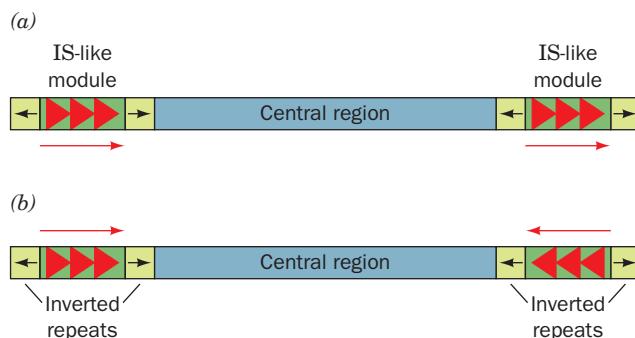


Figure 30-86 A composite transposon. This element consists of two identical or nearly identical IS-like modules (green) flanking a central region carrying various genes. The IS-like modules may have either (a) direct or (b) inverted relative orientations.

by an identifying number. For example, **Tn3** (Fig. 30-85) consists of 4957 bp and has inverted terminal repeats of 38 bp each. The central region of Tn3 codes for three proteins: (1) a 1015-residue transposase named **TnpA**; (2) a 185-residue protein known as **TnpR**, which mediates the **site-specific recombination** reaction necessary to complete the transposition process (see below) and also functions as a repressor for the expression of both *tnpA* and *tnpR*; and (3) a **β-lactamase** that inactivates ampicillin (Section 11-3Bb). The site-specific recombination occurs in an AT-rich region known as the **internal resolution site** that is located between *tnpA* and *tnpR*.

3. The so-called **composite transposons** (Fig. 30-86) consist of a gene-containing central region flanked by two identical or nearly identical IS-like modules that have either the same or an inverted relative orientation. It therefore seems that composite transposons arose by the association of two originally independent IS elements. Since the IS-like modules are themselves flanked by inverted repeats, the ends of either type of composite transposon must also be inverted repeats. Experiments demonstrate that composite transposons can transpose any sequence of DNA in their central region.

There are two modes of transposition: (1) **direct** or **simple transposition**, in which the transposon, as the name implies, physically moves from one DNA site to another; and (2) **replicative transposition**, in which the transposon remains at its original site and a copy of it is inserted at a target site. The two modes, as we shall see, have similar mechanistic features and, indeed, some transposons can move by either mode.

b. Direct Transposition of Tn5 Occurs by a Cut-and-Paste Mechanism

Tn5 is a 5.8-kb composite transposon that contains the gene encoding the 476-residue **Tn5 transposase** together with three antibiotic resistance genes. It is flanked by inverted IS-like modules ending in 19-bp sequences called outside end (OE) sequences. Tn5 undergoes direct transposition via a “cut-and-paste” mechanism that was elucidated in large part by William Reznikoff (Fig. 30-87):

1. Each of Tn5’s two OE sequences on the donor DNA is bound by a monomer of Tn5 transposase.
2. The transposase dimerizes to form a catalytically active **synaptic complex** in which the transposon is held between the two transposase subunits.
3. Each transposase subunit activates a water molecule to nucleophilically attack the outermost nucleotide of its bound OE sequence, yielding a free 3'-OH group. This 3'-OH group is then activated to attack the opposite strand on the DNA to form a hairpin structure, thereby excising the transposon from the DNA. The hairpin is then hydrolyzed to yield a blunt-ended dsDNA at each end of the transposon, thus completing the “cut” portion of the transposition mechanism.
4. The synaptic complex binds to the target DNA.

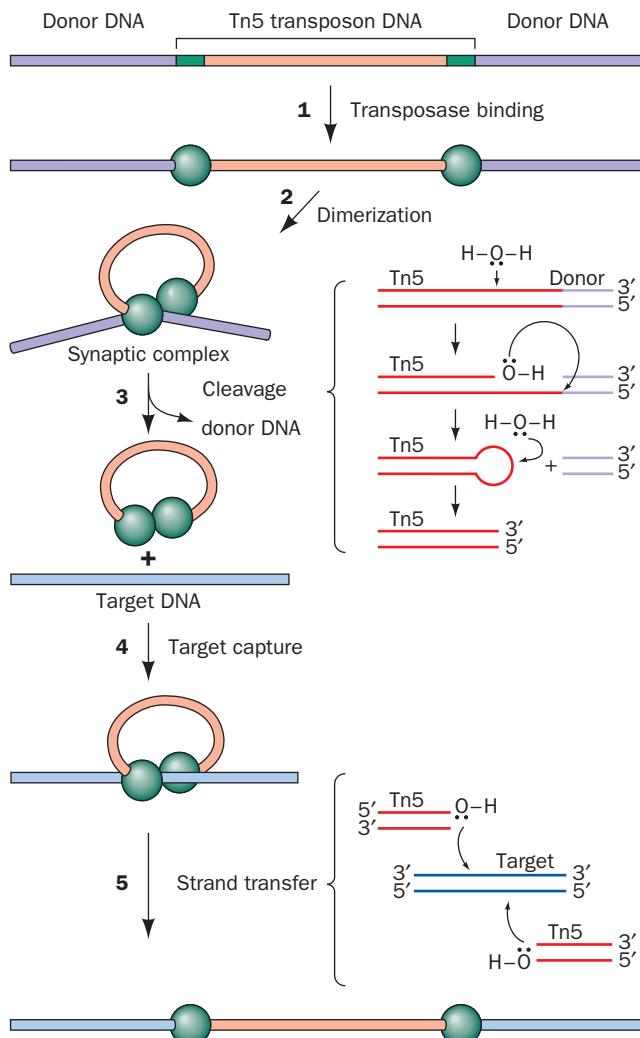


Figure 30-87 The cut-and-paste transposition mechanism catalyzed by Tn5 transposase. The reactions comprising Steps 3 and 5 are indicated beside the braces to the right of these steps. [After Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I., *Science* **289**, 77 (2000).]

5. The transposon's 3'-OH groups nucleophilically attack the target DNA on opposite strands spaced 9 bp apart, thereby installing the transposon at the target site. Remarkably, this reaction and the three preceding lytic reactions are all mediated by the same catalytic site. The repair of the oppositely located single-strand gaps (Fig. 30-84) completes the “paste” portion of the mechanism.

Although, strictly speaking, not part of the transposition process, the double-strand break in the donor DNA left by the excision of the transposon must be repaired if the donor DNA is to be propagated (in bacteria, the donor DNA is often a plasmid so that its loss has little effect on the cell since plasmids are generally present in multiple copies).

The X-ray structure of a Tn5 synaptic complex (Fig. 30-88), determined by Reznikoff and Ivan Rayment, provides a model of the synaptic complex at the stage following its

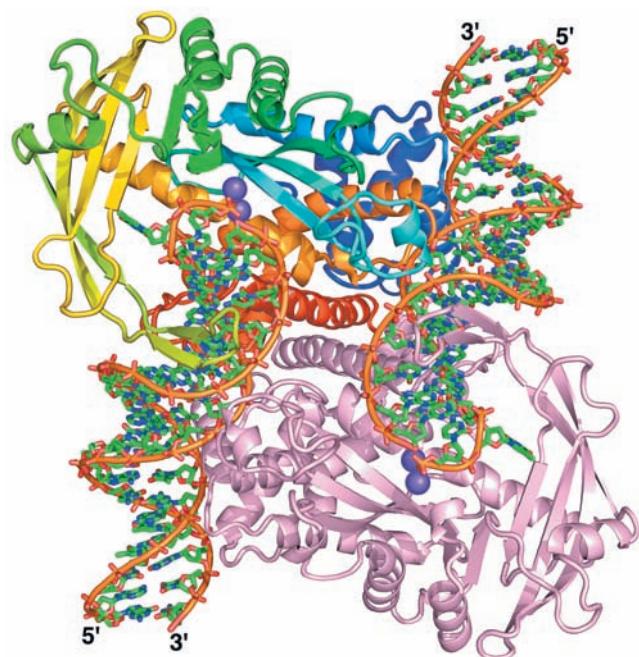


Figure 30-88 X-ray structure of Tn5 transposase in complex with a 20-bp DNA containing the OE sequence. The complex, which represents the product of Step 3 in Fig. 30-87, is viewed along its 2-fold axis with one of its two identical subunits colored in rainbow order from N-terminus (blue) to C-terminus (red) and the other subunit pink. The DNA is drawn in stick form with C green, N blue, O red, and P orange and with successive P atoms on the same polynucleotide connected by orange rods. The bound Mn²⁺ ions, which mark the enzyme's active site, are represented by purple spheres. The DNAs' reactive 3'-OH groups are located at these active sites. [Based on an X-ray structure by William Reznikoff and Ivan Rayment, University of Wisconsin. PDBid 1MUS.]

cleavage from the donor DNA (the product of Step 3 in Fig. 30-87). This 2-fold symmetric complex consists of a dimer of Tn5 transposase subunits binding two 20-bp DNA segments containing the Tn5 transposon's 19-bp OE sequence with the outer end of each OE sequence bound to the protein (and whose opposite ends would, *in vivo*, be connected by the looped around transposon; Fig. 30-87). Both transposase subunits extensively participate in binding each DNA segment, thereby explaining why the individual subunits cannot cleave their bound DNA segments before forming the synaptic complex. The protein holds the DNA in a distorted B-DNA conformation with its two end pairs of nucleotides no longer base paired. Indeed, the penultimate base on the nontransferred strand is flipped out of the double helix and binds in a hydrophobic pocket. The transferred strand's free 3'-OH group, which occupies the active site, is bound in the vicinity of a cluster of three catalytically essential acidic residues, the so-called **DDE motif**, which is shared with other transposases. In the X-ray structure the DDE motif binds two Mn²⁺ ions, although physiologically it probably binds two Mg²⁺ ions. This suggests that transposases employ a metal-activated catalytic

mechanism similar to that of the DNA polymerases (Section 30-2A). The facing surface of the protein in Fig. 30-88 is positively charged with a prominent groove running from upper left to lower right that forms the apparent binding site for the target DNA.

Wild-type Tn5 transposase has such low catalytic activity that it is undetectable *in vitro*. However, that in the X-ray structure is a hyperactive mutant form that contains the mutations E54K and L372P (an unusual circumstance in that it is far more common to mutationally inhibit an enzyme under crystallographic study so as to trap it at some specific stage along its reaction pathway). Lys 54 is hydrogen bonded to O4 of a thymine base on the transferred strand. In the wild-type transposase, Glu 54 would probably have an unfavorable charge–charge repulsion with a nearby phosphate group, thus providing a structural basis for the increased activity of the E54K mutant. The L372P mutation disorders the peptide segment between residues 373 and 391 (it is ordered in the X-ray structure of wild-type Tn5 transposase lacking its N-terminal 55 residues), thereby suggesting that this mutation facilitates a conformational change required for substrate binding.

c. Replicative Transposition Occurs via Cointegrates

If a plasmid carrying a transposon resembling Tn3 is introduced into a bacterial cell carrying a plasmid that lacks the transposon, in some of the progeny cells both types of plasmid will contain the transposon (Fig. 30-89). Evidently, *such transposition involves the replication of the transposon into the recipient plasmid rather than its transfer from donor to recipient*.

Two plasmids, one containing a replicative transposon, will occasionally fuse to form a so-called **cointegrate** containing like-oriented copies of the transposon at both junctions of the original plasmids (Fig. 30-90). Yet, some of the progeny of a cointegrate-containing cell lack the cointegrate and instead contain both original plasmids, each with one copy of the transposon (Fig. 30-89). The cointegrate must therefore be an intermediate in the transposition process.

Although the mechanism of replicative transposition has not been fully elucidated, a plausible model for this process (and there are several) that accounts for the foregoing observations consists of the following steps (Fig. 30-91):

1. A pair of staggered single-strand cuts, such as is diagrammed in Fig. 30-84, is made by the transposon-encoded transposase at the target sequence of the recipient plasmid so as to liberate 3'-OH ends. Similarly, single-strand cuts are made on opposite strands to either side of the transposon. Note that these reactions resemble those catalyzed by Tn5 transposase (Fig. 30-87).

2. Each of the transposon's free ends is ligated to a protruding single strand at the insertion site. This forms a replication fork at each end of the transposon.

3. The transposon is replicated, thereby yielding a cointegrate.

4. Through a site-specific recombination between the internal resolution sites of the two transposons, the cointe-

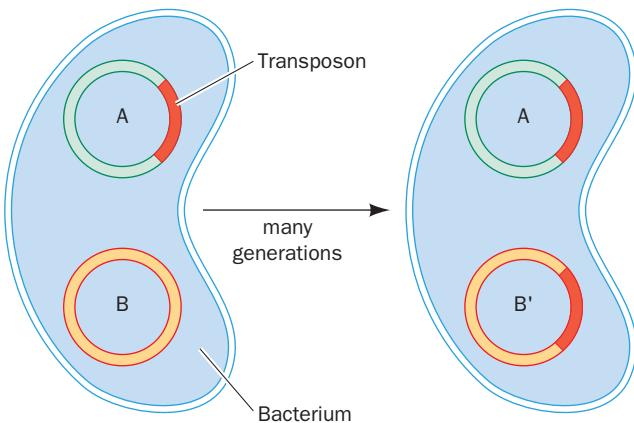


Figure 30-89 **Relicative transposition.** This type of transposition inserts a copy of the transposon at the target site while another copy remains at the donor site.

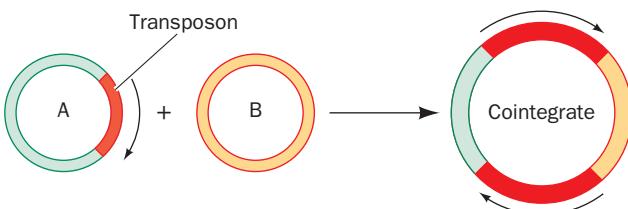


Figure 30-90 **A cointegrate.** This structure forms by the fusion of two plasmids, one carrying a transposon, such that both junctions of the original plasmid are spanned by transposons with the same orientation (arrows).

grate is resolved into the two original plasmids, each of which contains a transposon. This crossover process is catalyzed by transposon-encoded **recombinases** (TnpR in Tn3; also known as **resolvases**) rather than RecA; transposition proceeds normally in *recA*⁻ cells (although RecA will resolve a cointegrate containing a transposon with a mutant resolvase and/or an altered internal resolution site, albeit at a much reduced rate).

Site-specific recombinases fall into only two protein families, **serine recombinases** and **tyrosine recombinases**, which are named after the amino acid residue that forms a transient covalent linkage to the DNA during the recombinase reaction. As we shall, these two types of recombinases function via different mechanisms.

d. $\gamma\delta$ Resolvase Catalyzes Site-Specific Recombination

The **$\gamma\delta$ resolvase**, a serine recombinase that forms a homodimer in solution, is a TnpR homolog that is encoded by the **$\gamma\delta$ transposon** (a member of the Tn3 family of replicative transposons; Fig. 30-85). It catalyzes a site-specific recombination event in which a cointegrate containing two copies of the $\gamma\delta$ transposon is resolved, via double-strand DNA cleavage, strand exchange, and

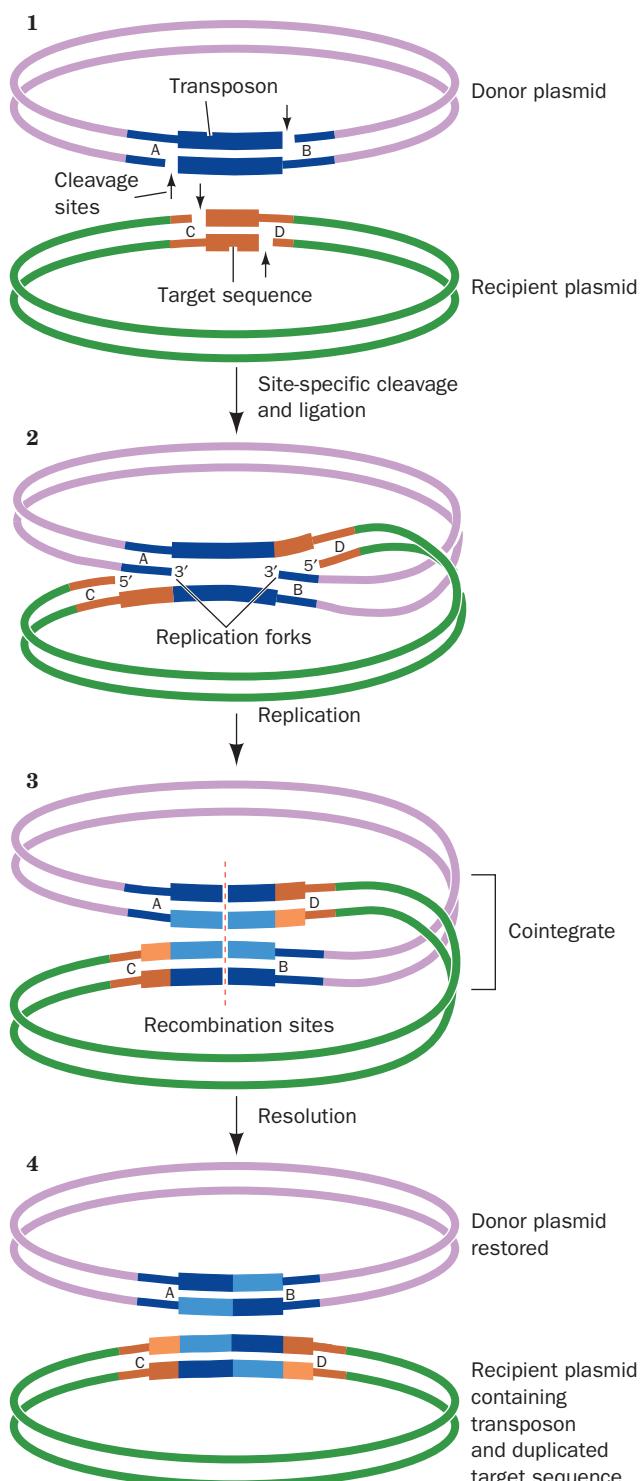


Figure 30-91 A model for transposition involving the intermediacy of a cointegrate. Here more lightly shaded bars represent newly synthesized DNA. [After Shapiro, J.A., *Proc. Natl. Acad. Sci.* **76**, 1934 (1979).]

religation (the last step in Fig. 30-91), into two catenated (linked) dsDNA circles that each contain one copy of the $\gamma\delta$ transposon (it also serves as its own transcriptional repressor as does TnpR). The $\gamma\delta$ transposon contains a 114-bp

res site that includes three binding sites for $\gamma\delta$ resolvase dimers, each of which contains an inverted repeat of the $\gamma\delta$ resolvase's 12-bp recognition sequence. The resolution of the cointegrate involves the binding of a $\gamma\delta$ resolvase homodimer to all six of these binding sites in the cointegrate (three from each of its two transposons) as is diagrammed in Fig. 30-92. The reaction proceeds via the formation of a transient phosphoSer bond between Ser 10 and the 5'-phosphate at each cleavage site.

The X-ray structure of the $\gamma\delta$ resolvase homotetramer in complex with two 34-bp palindromic dsDNA segments containing an inverted repeat of the 12-bp recognition sequence separated by an 4-bp spacer (Fig. 30-93), determined by Nigel Grindley and Steitz, reveals that this synaptic tetramer has D_2 symmetry. Each 183-residue resolvase monomer consists of an N-terminal catalytic domain (residues 1–120) and a C-terminal DNA-binding domain (residues 148–183) connected by an extended arm (residues 121–147). Both dsDNAs, which are located at the periphery of the protein core, have been cleaved with each of the resulting four 5' ends in phosphoSer linkage to the resolvase. This preserves the free energy of the cleaved phosphodiester bond so that it can later be reformed with a different partner, much as occurs with topoisomerases (Section 29-3C).

Each centrally located catalytic domain approaches its bound DNA from its minor groove side with its C-terminal helix (helix E) binding over the minor groove (the segment of the E helix that contacts the DNA is disordered in the absence of the DNA). Each C-terminal domain binds over the major groove of its recognition sequence on the opposite side of the DNA from its attached catalytic domain with the extended arm that connects them running more or less along the DNA's minor groove. The two C-terminal domains of the resolvase subunits labeled L and R (and the symmetry-related L' and R' subunits) in Fig. 30-93 are thereby separated by two helical turns along the cleaved DNA, the segments of which closely assume the B-DNA conformation. Each C-terminal helix binds in the DNA's major groove and, together with its preceding helix, forms a **helix-turn-helix (HTH) motif**, a common sequence-specific DNA-binding motif that occurs mainly in prokaryotic transcriptional repressors and activators (Section 31-3D).

The structure of the L–R dimer closely resembles that in the X-ray structure of the dimer bound to uncleaved site I DNA (the presynaptic dimer). This, and the short (17 Å) distance between the free 3'-OH group in the L subunit and the phosphoSer bond in the R subunit compared to other such distances in the complex (L–L' and L–R'), indicates that the L–R and L'–R' dimers correspond to the initial site I-bound dimers soon after cleavage or just before religation (Fig. 30-92). Consequently, the interface between the L–R and L'–R' dimers must be the newly formed synaptic interface.

How are the DNA strands in the synaptic complex exchanged, that is, how is the DNA bound to the L subunit ligated to the DNA bound to either the L' subunit or the R' subunit (and R to either R' or L')? In either case, the free 3'-OH group on each subunit is ~50 Å from the

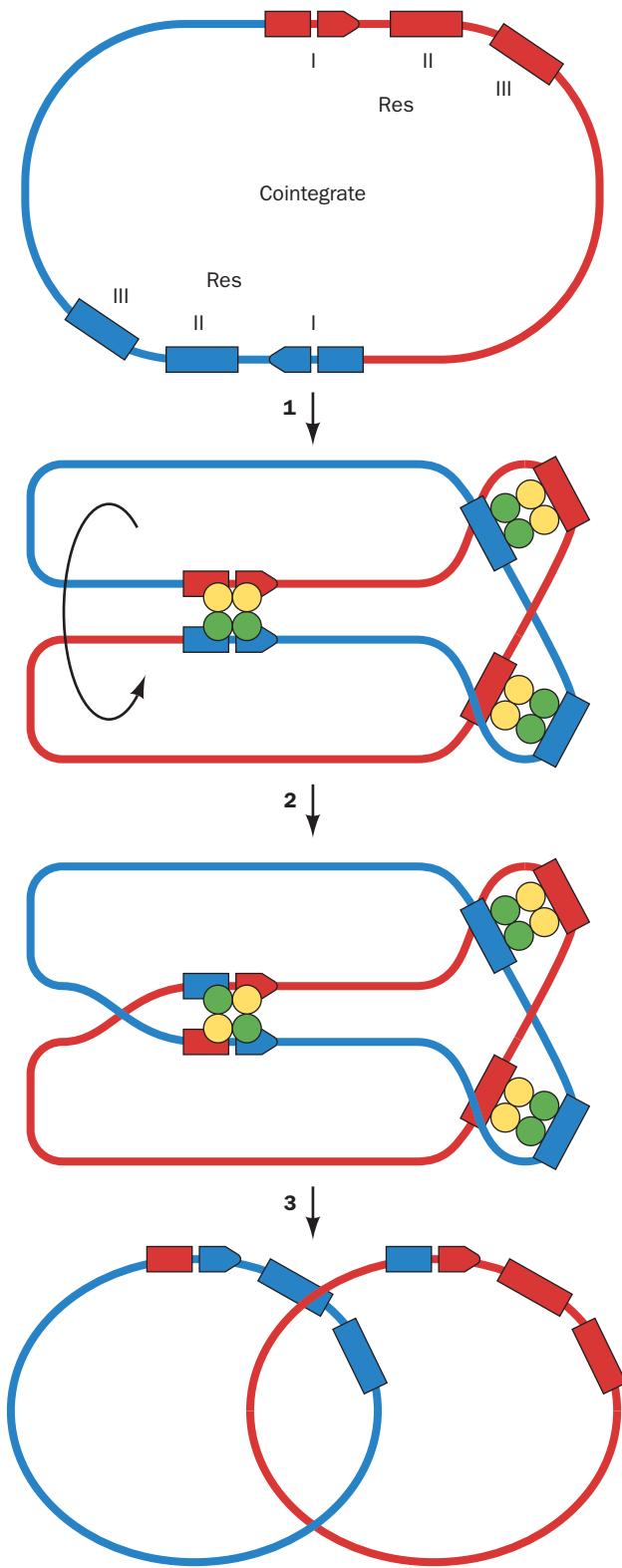


Figure 30-92 A model for the resolution of a cointegrate containing two $\gamma\delta$ transposons to form two catenated dsDNA circles. (1) The $\gamma\delta$ resolvase binds as six homodimers to its binding sites, I, II, and III, in each of the cointegrate's two *res* sites (yellow and green circles represent the $\gamma\delta$ resolvase monomers initially bound to the red and blue *res* sites, respectively), which then associate to form synaptic tetramers. Although not shown as such, the synaptic tetramers bound to sites I associate with the synaptic tetramers bound to sites II and III to form, as seen in the electron microscope, a compact globule of unknown structure known as a **synapsosome**. (2) The dsDNA at sites I both undergo staggered (by 2 bp) double-strand scissions via the transient formation of phosphoSer bonds between Ser 10 and the 5'-phosphates at the cleavage sites. The cleaved strands then exchange places (cross over) in a process that apparently requires the rotation of one of the pairs of resolvase monomers with respect to the other and are then ligated. (3) The dissociation of the synapsosome yields the catenated dsDNA circles. [Courtesy of Gregory Mullen, University of Connecticut Health Center.]

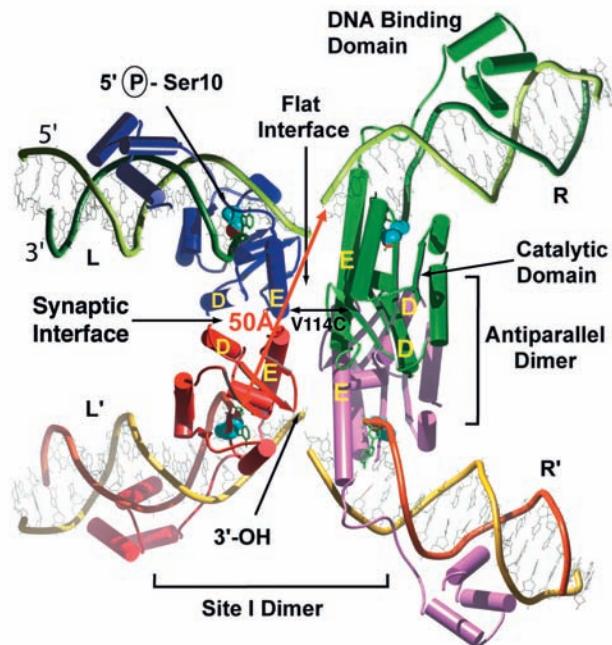


Figure 30-93 X-ray structure of a $\gamma\delta$ resolvase synaptic tetramer in complex with two 34-bp palindromic, site I-containing dsDNAs. The D_2 -symmetric complex is viewed with one of its 2-fold axes horizontal (the other 2-fold axes lie in the vertical plane labeled "Flat interface"). The DNAs (dark and light green and yellow and orange) have been cleaved into half-sites labeled L, R, L', and R' through the nucleophilic attack of the Ser 10 side chain (drawn in space-filling form with C blue and O red) on the 5'-phosphate group of A20 (drawn in stick form in green). The subunits of the resolvase tetramer, whose helices are drawn in tube form, are colored blue, green, red, and pink. The intact DNAs initially bind to the L-R and L'-R' dimers, the so-called site I dimers. The L-L' dimer and the symmetrically equivalent R-R' dimer form so-called antiparallel dimers whose D and E helices associate as four-helix bundles. In addition, a Val 114 \rightarrow Cys mutation (V114C) disulfide-cross-links the E helices of the L-R and L'-R' dimers across the so-called flat interface. [Courtesy of Thomas Steitz, Yale University. PDBID 1ZR4.]

phosphoSer group that it would have to nucleophilically attack in the religation step and the space between them is filled with protein. Clearly, the synaptic complex must undergo a dramatic structural change to accomplish the religation step.

The observation that the interface between the L–L' and R–R' dimers is largely hydrophobic and unusually flat (Fig. 30-93) strongly suggests that the religation occurs after a 180° rotation of these dimers with respect to one another (about the horizontal 2-fold axis in Fig. 30-93), thus exchanging the positions of the R and R' subunits with respect to the L and L' subunits. The rotation is presumably driven by the superhelical tension in the naturally negatively supercoiled cointegrate. This model is supported by energy calculations and the observation that mutating Val 114 of the E helix to Cys, which, under oxidizing conditions, disulfide-links the E helices on the L and R (and L' and R') subunits (a mutation that was present in the foregoing structure), yields a complex that can form the covalent intermediate in Fig. 30-93 but cannot carry out the religation step—presumably because the disulfide bonds prevent the above rotation. However, reducing these disulfide bonds restores the complex's recombinational activity. Furthermore, mutation of Lys 136, an E helix residue located at the flat interface, to Cys prevents religation of the cleaved DNA when subjected to oxidizing conditions, even though it requires a 75° rotation of the L–L'/R–R' interface in the structure in Fig. 30-93 to bring the Cys 136 side chains on opposing subunits close enough to form a disulfide bond. Of course, a detailed understanding of the mechanism of the $\gamma\delta$ resolvase reaction will require the knowledge of how all six $\gamma\delta$ resolvase dimers that form the synapsosome participate in the reaction (Fig. 30-92).

e. Replicative Transposons Are Responsible for Much Genetic Remodeling in Prokaryotes

In addition to mediating their own insertion into DNA, *replicative transposons promote inversions, deletions, and rearrangements of the host DNA*. Inversions can occur

when the host DNA contains two copies of a transposon in inverted orientation. The recombination of these transposons inverts the region between them (Fig. 30-94a). If, instead, the two transposons have the same orientation, the resolution of this cointegrate-like structure deletes the segment between the two transposons (Fig. 30-94b; if the deleted segment lacks a replication origin, it will not be propagated). The deletion of a chromosomal segment in this manner, followed by its integration into the chromosome at a different site by a separate recombinational event, results in chromosomal rearrangement.

Transposition appears to be important in chromosomal and plasmid evolution. Indeed, it has been suggested that transposons are nature's genetic engineering "tools." For example, the rapid evolution, since antibiotics came into common use, of plasmids that confer resistance to several antibiotics (Section 5-5Ba) has resulted from the accumulation of the corresponding antibiotic-resistance transposons in these plasmids. Transposon-mediated rearrangements may well have been responsible for organizing originally distant genes into coordinately regulated operons (Section 5-4Aa) as well as for forming new proteins by linking two formerly independent gene segments. Moreover, *the occurrence of identical transposons in unrelated bacteria indicates that the transposon-mediated transfer of genetic information between organisms is not limited to related species, in contrast to genetic transfers mediated by homologous recombination*.

f. Phase Variation Is Mediated by Site-Specific Recombination

Phenotypic expression in bacteria can be regulated by site-specific recombination. For example, certain strains of

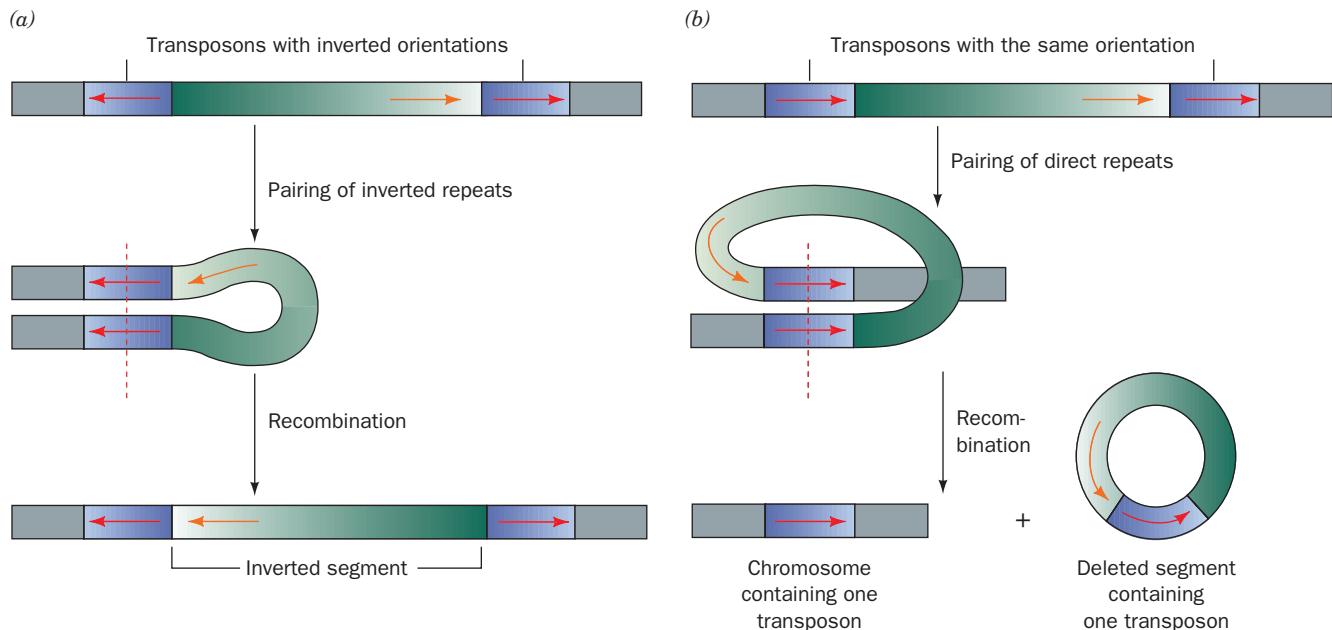


Figure 30-94 Chromosomal rearrangement via recombination.

(a) The inversion of a DNA segment between two identical transposons with inverted orientations. (b) The deletion of a

DNA segment between two identical transposons with the same orientation. This process parcels one transposon each to the resulting two DNA segments.

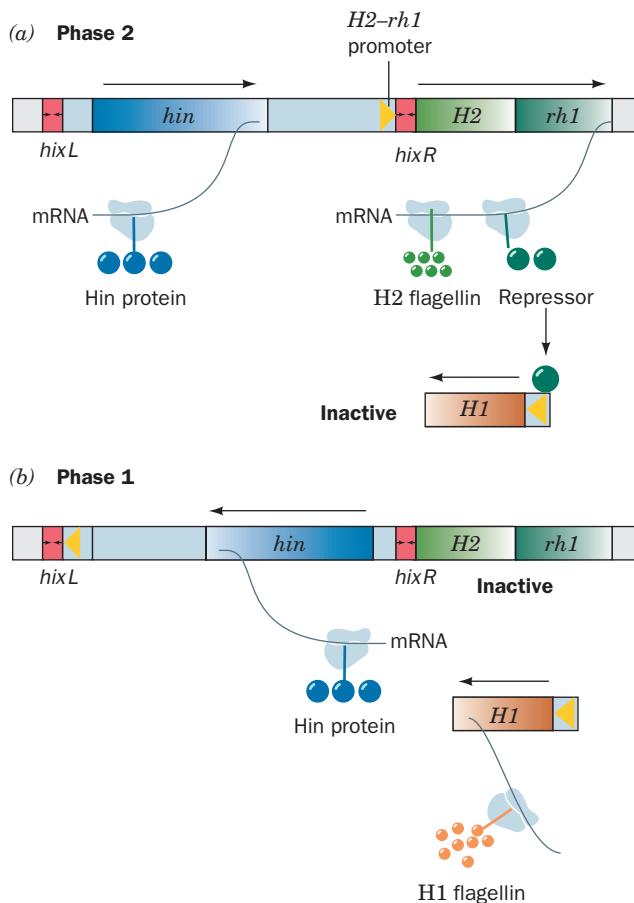


Figure 30-95 The mechanism of phase variation in *Salmonella*. (a) In Phase 2 bacteria, the *H2-rh1* promoter is oriented so that *H2* flagellin and repressor are synthesized. Repressor binds to the *H1* gene, thereby preventing its expression. (b) In Phase 1 bacteria, the segment preceding the *H2-rh1* transcription unit has been inverted relative to its orientation in Phase 2 bacteria. Hence this transcription unit cannot be expressed because it lacks a promoter. This releases *H1* from repression and results in the synthesis of *H1* flagellin. The inversion of the segment preceding the *H2-rh1* transcription unit is mediated by the Hin protein, which is expressed in either orientation by the *hin* gene.

Salmonella typhimurium make two antigenically distinct versions of the protein **flagellin** (the major component of the whiplike flagella with which bacteria propel themselves; Section 35-3I) that are designated **H1** and **H2**. Only one of these proteins is expressed by any particular cell but about once every 1000 cell divisions, in a process known as **phase variation**, a cell switches the type of flagellin it synthesizes. It is thought that phase variation helps *Salmonella* evade its host's immunological defenses.

What is the mechanism of phase variation? The two flagellin genes reside on different parts of the bacterial chromosome. *H2* is linked to the *rh1* gene that encodes a repressor of *H1* expression (Fig. 30-95; *rh1*, *H2*, and *H1* are also known as *fliA*, *fliB*, and *fliC*, respectively). Hence, when the *H2-rh1* transcription unit is expressed, *H1* synthesis is

repressed; otherwise *H1* is synthesized. Melvin Simon has shown that the expression of the *H2-rh1* unit is controlled by the orientation of a 995-bp segment that lies upstream of *H2* (Fig. 30-95) and that contains the following elements:

1. A promoter for *H2-rh1* expression.
2. The *hin* gene, which encodes the 190-residue **Hin DNA invertase**. Hin, a serine recombinase, mediates the inversion of the DNA segment in a manner similar to that diagrammed in Fig. 30-94a. In fact, Hin is ~40% identical in sequence with the $\gamma\delta$ resolvase, which strongly suggests that these proteins have similar structures.
3. Two closely related 26-bp sites, *hixL* and *hixR*, that form the boundaries of the segment and hence contain its cleavage sites. They each consist of two imperfect 12-bp inverted repeats separated by 2 nt.

In the Phase 2 orientation (Fig. 30-95a), the properly oriented promoter is just upstream of *H2*, so this gene and *rh1* are coordinately expressed, thereby repressing *H1* synthesis. In Phase 1 bacteria (Fig. 30-95b), however, this segment has the opposite orientation. Consequently, neither *H2* nor *rh1*, which then lacks a promoter, is expressed so that *H1* is synthesized.

g. Cre-Mediated Site-Specific Recombination Occurs via 3'-PhosphoTyr Intermediates

Bacteriophages, as we have seen (Fig. 1-31), replicate themselves within their host bacterial cells which, in most cases, they then lyse to release the progeny phage, a lifestyle that is therefore known as the **lytic** mode. However, certain bacteriophages can assume an alternative, nondestructive lifestyle, the **lysogenic** mode, in which they install their DNA, usually in the host chromosome via site-specific recombination, so that the phage DNA is passively replicated with the host DNA. However, if the bacterial host encounters conditions in which it is unlikely to survive, the phage DNA is excised from the bacterial chromosome via a reversal of the site-specific recombination reaction and it reenters the lytic mode so as to escape the doomed host. We discuss the genetic factors that maintain the balance between the lytic and lysogenic lifestyles in **bacteriophage λ** in Section 33-3.

The enzymes that mediate the foregoing site-specific recombination reactions are members of the **λ integrase** (**λ Int**; alternatively, tyrosine recombinase) family, whose ~1000 known members also occur in prokaryotes and eukaryotes. These include the **XerC** and **XerD** proteins of *E. coli* which, operating in concert, function to decatenate the two linked circular dsDNA products of homologous recombination (Fig. 30-69g, left), as well as type IB topoisomerases (Section 29-3Cc).

The structurally best characterized member of the λ integrase family is the **Cre recombinase** of *E. coli* **bacteriophage P1**. In its lysogenic state, bacteriophage P1 is a single-copy circular plasmid (rather than being inserted in the host chromosome as is bacteriophage λ), but in the phage head (the lytic mode), P1 DNA is a linear dsDNA that has a 34-bp *loxP* site at each end. The main function of Cre,

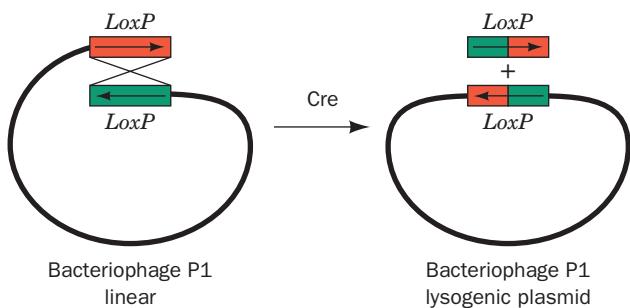


Figure 30-96 The circularization of linear bacteriophage P1 DNA.

DETAILED EXPLANATION This occurs through the Cre-mediated site-specific recombination between its two terminally located *loxP* sites (red and green) to yield its lysogenic plasmid.

which is encoded by bacteriophage P1, is to mediate the site-specific recombination between these two *loxP* sites so as to circularize the linear DNA (Fig. 30-96).

The *loxP* site is palindromic except for its central 8-bp crossover region, which confers directionality on the site. In carrying out the recombination reaction, the 343-residue Cre subunits form a homotetramer that binds two *loxP* sites in an antiparallel orientation, with each Cre subunit binding half of a *loxP* site. Then, as is diagrammed in Fig. 30-97, oppositely located Cre subunits catalyze single-strand scissions on the 5' side of the crossover region on one strand of each of the two dsDNAs. This occurs through the nucleophilic attack of each of these active Cre subunit's conserved Tyr 324 residues on the DNA's scissile phosphoester bond to yield a 3'-phosphoTyr intermediate on one side of the cleaved bond and a free 5'-OH group on the other side (as similarly occurs in the reactions catalyzed by type IB topoisomerases; Section 29-3Cc). Each of the liberated 5'-OH groups then nucleophilically attacks the 3'-phosphoTyr group on the opposite duplex to form a Holliday junction, thereby releasing the Tyr residues. The Holliday

junction is resolved into two recombined dsDNAs when the two Cre subunits that had not yet participated in the reaction mediate the same cleavage and strand exchange reactions on the two heretofore unreacted single strands. This latter process must be preceded by a structural rearrangement (isomerization) of the Cre tetramer that positions the catalytic Tyr residues in the latter pair of subunits to participate in the reaction while those in the former pair of subunits are similarly removed from the scene of the action. Note that this mechanism differs from that mediated by serine recombinases in that the latter cleave all four DNA strands prior to initiating their exchange and hence do not have a Holliday junction intermediate (Section 30-6Bd).

The X-ray structures of Cre tetramers in their complexes with several *loxP* model DNAs, determined by Gregory Van Duyne, have helped elucidate its mechanism. When the DNA had a single-strand nick past the second nucleotide from the 5' end of the crossover region, Cre-catalyzed strand scission yielded a free nucleotide (a CMP) that diffused away. Since this nucleotide contained the otherwise reactive 5'-OH group, the 3'-phosphoTyr intermediate was irreversibly trapped, that is, Cre could not carry out the strand exchange reaction in Fig. 30-97 (this nicked DNA is a suicide substrate for Cre; Section 28-3Bc). The X-ray structure of the Cre complex of this nicked DNA confirmed the presence of the 3'-phosphoTyr intermediate and indicated, through model building, that the 5'-OH group on the missing CMP residue would be well positioned to nucleophilically attack the 3'-phosphoTyr bond on the opposite strand (Fig. 30-98a). Note that this complex is only 2-fold symmetric although its four Cre subunits and much of the DNA are related by pseudo-4-fold symmetry. When the DNA was, instead, an immobile Holliday junction (Fig. 30-98b), the complex was also pseudo-4-fold symmetric with the single strands that had crossed over noticeably kinked at their centers. These structures revealed that the conformational changes necessary to carry out the strand exchange and isomerization reactions (Fig. 30-97) required surprisingly small movements on the part of the Cre subunits and that only the sugar-phosphate backbones of the strand-exchanged nucleotides needed to move in order to form the Holliday junction.

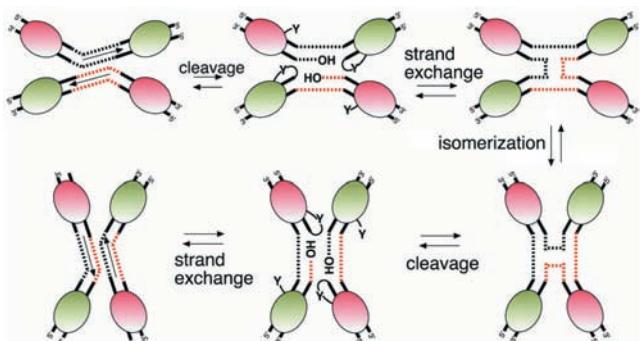


Figure 30-97 The mechanism of Cre-*loxP* site-specific recombination.

DETAILED EXPLANATION The dashed lines represent the nonpalindromic crossover regions of the *loxP* sites. The green and magenta Cre subunits are active for cleavage in the top and bottom parts of the diagram, respectively, with their roles being switched by the isomerization step. Note that the mechanism does not require branch migration of the Holliday junction intermediate. [Courtesy of Gregory Van Duyne, University of Pennsylvania School of Medicine.]

h. Most Transpositions in Eukaryotes Involve RNA Intermediates

Transposons similar to those in prokaryotes also occur in eukaryotes, including yeast, maize, *Drosophila*, and humans. In fact, ~3% of the human genome consists of DNA-based transposons although, in most cases, their sequences have mutated so as to render them inactive, that is, these transposons are evolutionary fossils. However, many eukaryotic transposons exhibit little similarity to those of prokaryotes. Rather, their base sequences resemble those of retroviruses (see below), which suggests that these transposons are degenerate retroviruses. The transposition of these so-called **retrotransposons** occurs via a pathway that resembles the replication of retroviral DNA (Section 15-4C): (1) their transcription to RNA, (2) the reverse transcriptase-mediated copying of this RNA to cDNA (Section 30-4C), and (3) the largely random insertion of this DNA into the host

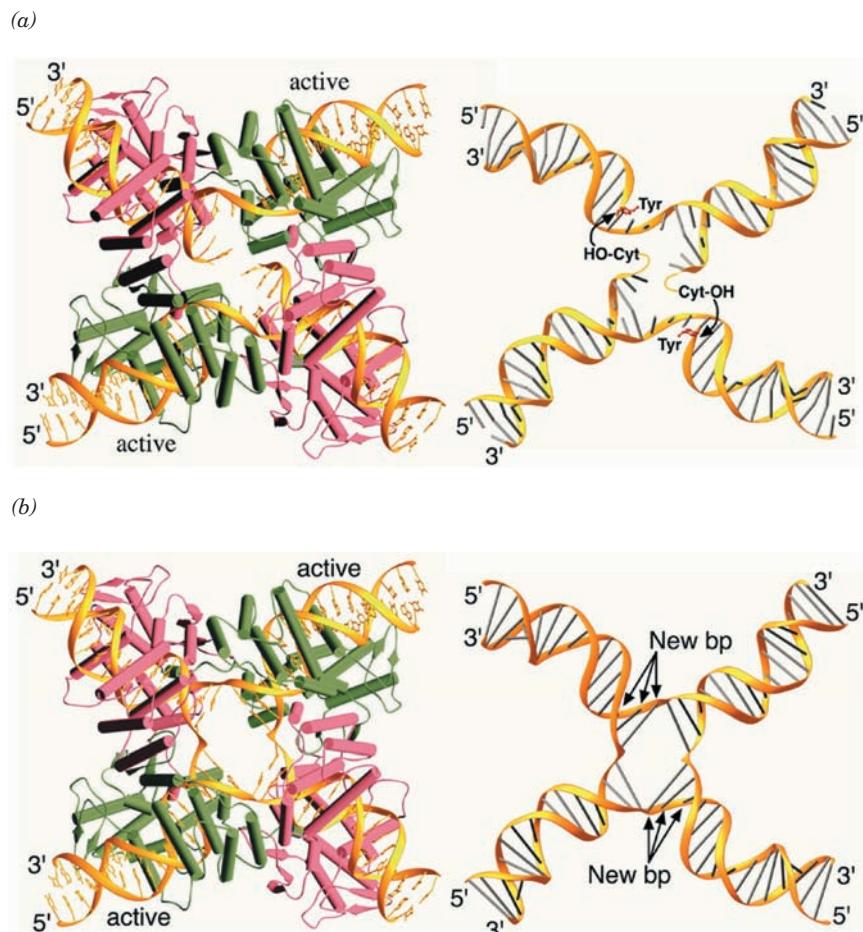


Figure 30-98 X-ray structures of the Cre homotetramer in its complexes with model *loxP* DNAs. (a) Two identical dsDNAs that were nicked past the second nucleotide from the 5' end of their crossover regions; and (b) an immobile Holliday junction. The left panels show the Cre-DNA complexes as viewed along their exact 2-fold and pseudo-4-fold axes, with the active and inactive subunits green and magenta, respectively (as in Fig. 30-97), and with the DNA gold. The right panels show only the DNAs in the X-ray structures as viewed from below the left panels. In the right panel of Part a, the active site Tyr that is covalently linked

to the 3'-OH group of the cleaved DNA strand is shown in stick form (red) and the modeled-in position of the cleaved CMP's 5'-OH group is shown positioned to nucleophilically attack the 3'-phosphoTyr group on the opposite dsDNA (curved arrows). In the right panel of Part b, the three base pairs that form as a consequence of strand exchange are indicated. Note that the vertical strands in the crossovers but not the horizontal strands are distinctly kinked at their centers. [Courtesy of Gregory Van Duyne, University of Pennsylvania School of Medicine. PDBIDs 2CRX, 3CRX, 4CRX, and 5CRX.]

organism's genome as mediated by enzymes known as **integases** (which catalyze reactions similar to and structurally resemble cut-and-paste DNA transposases).

The involvement of RNA in retrotransposon-mediated transposition was ingeniously shown by Gerald Fink through his remodeling of **Ty1**, the most common transposable element in budding yeast (which has ~35 copies of this 6.3-kb element comprising ~13% of its 1700 kb genome; Ty stands for *Transposon yeast*), so that it contained a yeast intron (a sequence that is excised from an RNA transcript and hence is absent in the mature RNA; Section 5-4Ac) and was preceded by a galactose-sensitive yeast promoter. The transposition rate of this remodeled Ty1 element varied with the galactose concentration in the medium and the transposed elements all lacked the intron, thereby demonstrating the participation of an RNA intermediate.

A retroviral genome (Fig. 30-99a) is flanked by direct long terminal repeats (**LTRs**) of 250 to 600 bp and typically contains the genes encoding three polyproteins: **gag**, which is

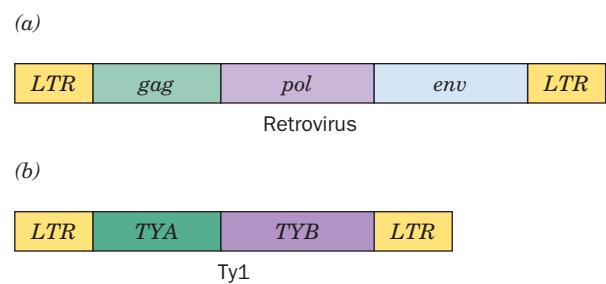


Figure 30-99 Gene sequences of (a) retroviruses and (b) the Ty1 retrotransposon.

cleaved to the proteins comprising the viral core (Fig. 15-34); **pol**, which is cleaved to the above-mentioned reverse transcriptase and integrase, as well as the protease that catalyzes these cleavages; and **env**, which is cleaved to viral outer envelope proteins. Ty1 (Fig. 30-99b) is likewise flanked by LTRs (of 330 bp) but expresses only two polyproteins: **TYA** and **TYB**, the counterparts of *gag* and *pol*. Moreover, TYA and TYB, together with Ty1 RNA, form viruslike particles in the yeast cytoplasm. However, Ty1 lacks a counterpart of the retroviral *env* gene. Hence Ty1 is an “internal virus” that can only replicate within a genome, albeit at an extremely low rate compared to that of real retroviral infections.

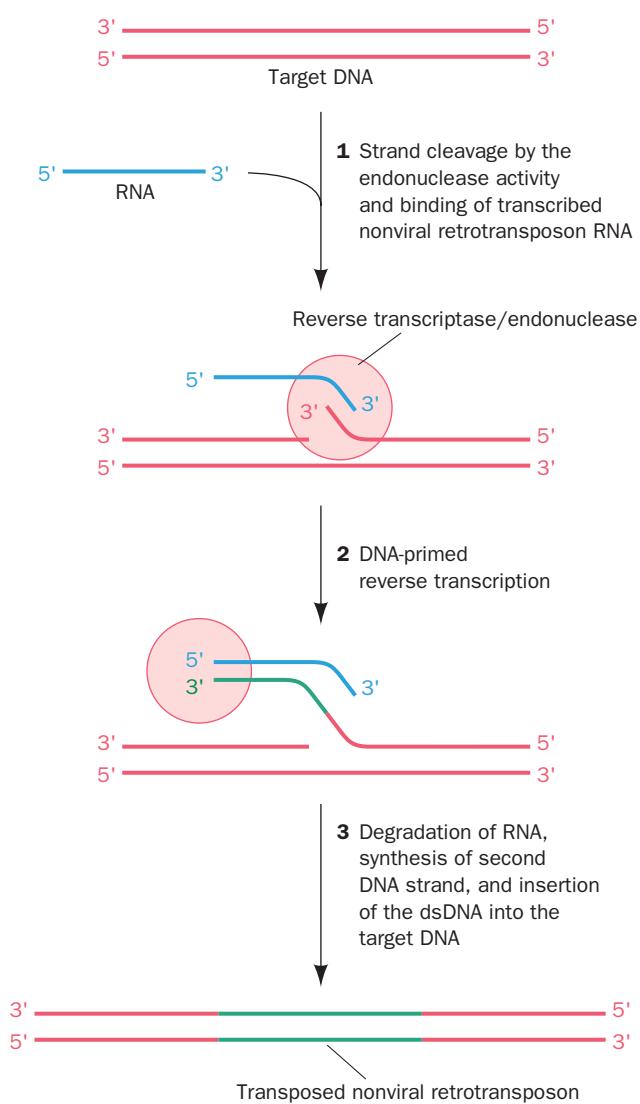


Figure 30-100 Proposed mechanism for the transposition of nonviral retrotransposons. (1) The retrotransposon-encoded reverse transcriptase/endonuclease nicks one strand of the target DNA and then recruits the RNA transcript of the retrotransposon to this site. (2) The DNA-primed reverse transcription of the retrotransposon RNA. (3) The RNA is degraded, and the second DNA strand is synthesized using the first strand as its template (normal reverse transcriptase reactions; Section 30-4C), followed by the insertion of the resulting nonviral retrotransposon into the target DNA via a poorly understood process.

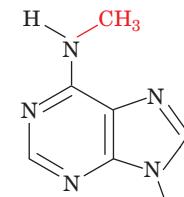
Copia (Latin for abundance), the most abundant retrotransposon in the *Drosophila* genome (which contains 20–60 copies of copia), resembles Ty1.

The LTRs in retroviruses and retrotransposons such as Ty1 and *copia* are essential elements for their transcription and hence for their transposition. Yet, vertebrate genomes also contain retrotransposons that lack LTRs and hence cannot be transcribed analogously to retroviruses. A common family of these **nonviral retrotransposons**, the 1- to 7-kb **long interspersed nuclear elements (LINEs)**, each contain two open reading frames: *ORF1*, which contains sequences similar to those in *gag*; and *ORF2*, which contains sequences similar to those encoding reverse transcriptase. A proposed mechanism for the transposition of LINEs is diagrammed in Fig. 30-100.

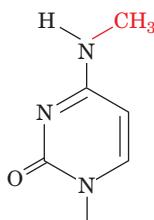
Different types of transposons, DNA-only, retroviral, and nonviral, predominate in different organisms. Thus bacteria, as we have seen, contain nearly exclusively DNA-only transposons, yeast have mainly retroviral retrotransposons, *Drosophila* have all three types, and in humans LINEs predominate. In fact, the human genome contains an estimated 1.4 million LINEs or LINE fragments that comprise ~20% of the 3.0-billion-bp human genome (genomic organization is discussed in Section 34-2). The great majority of these molecular parasites have mutated to the point of inactivity but a few still appear capable of further transposition. Indeed, several hereditary diseases are caused by the insertion of a LINE into a gene. Several other types of retrotransposons also comprise significant fractions of the human genome as we shall see in Section 34-2.

7 DNA METHYLATION AND TRINUCLEOTIDE REPEAT EXPANSIONS

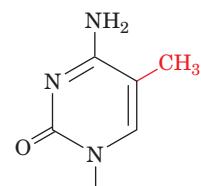
The A and C residues of DNA may be methylated, in a species-specific pattern, to form ***N*⁶-methyladenine (m⁶A)**, ***N*⁴-methylcytosine (m⁴C)**, and ***5*-methylcytosine (m⁵C)** residues, respectively.



***N*⁶-Methyladenine (m⁶A) residue**

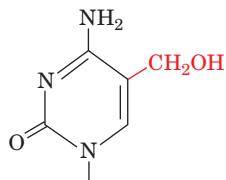


***N*⁴-Methylcytosine (m⁴C) residue**



5-Methylcytosine (m⁵C) residue

These are the only types of modifications to which DNA is subjected in cellular organisms (although all the C residues of T-even phage DNAs are converted to **5-hydroxymethylcytosine** residues,



5-Hydroxymethylcytosine residue

which may, in turn, be glycosylated). These methyl groups project into B-DNA's major groove, where they can interact with DNA-binding proteins. In most cells, only a few percent of the susceptible bases are methylated, although this figure rises to >30% of the C residues in some plants.

Bacterial DNAs are methylated at their own particular restriction sites, thereby preventing the corresponding restriction endonucleases from degrading the DNA (Section 5-5A). These restriction-modification systems, however, account for only part of the methylation of bacterial DNAs. In *E. coli*, most DNA methylation is catalyzed by the products of the *dam* and *dcm* genes. The **Dam methyltransferase (Dam MTase)** methylates the A residue in all GATC sequences, whereas the **Dcm MTase** methylates both C residues in CC_TGG at their C5 positions. Note

that both of these sequences are palindromic. We have seen that *E. coli* uses Dam Mtase-mediated methylation to differentiate parental from newly synthesized DNA in mismatch repair (Section 30-5C) and in limiting *oriC*-based DNA replication initiation to once per cell generation via sequestration (Section 30-3Cb).

a. The MTase Reaction Occurs via a Covalent Intermediate in Which the Target Base Is Flipped Out

The Dam and Dcm MTases, as do all known DNA MTases, use *S*-adenosylmethionine (SAM) as their methyl donor. Indeed, all m⁵C-MTases share a set of conserved sequence motifs. Daniel Santi has proposed that the catalytic mechanism of these m⁵C-MTases (Fig. 30-101) is similar to that of thymidylate synthase (Fig. 28-19) in that both types of enzymes transfer methyl groups to pyrimidine C5 atoms via a reaction that is initiated by the nucleophilic attack of a Cys thiolate group on the pyrimidine's C6 position. The pyrimidine's C5 atom is thereby activated as a resonance-stabilized carbanion that nucleophilically attacks the methyl donor's methyl group (which in thymidylate synthase is donated by *N*⁵,*N*¹⁰-methylene-THF rather than SAM) to yield a covalent intermediate. This intermediate subsequently decomposes to products through the enzymatic abstraction of the proton substituent to C5 and elimination of the enzyme. The Cys thiolate nucleophile is a component of a Pro-Cys dipeptide that is invariant in all known m⁵C-MTases and thymidylate synthases.

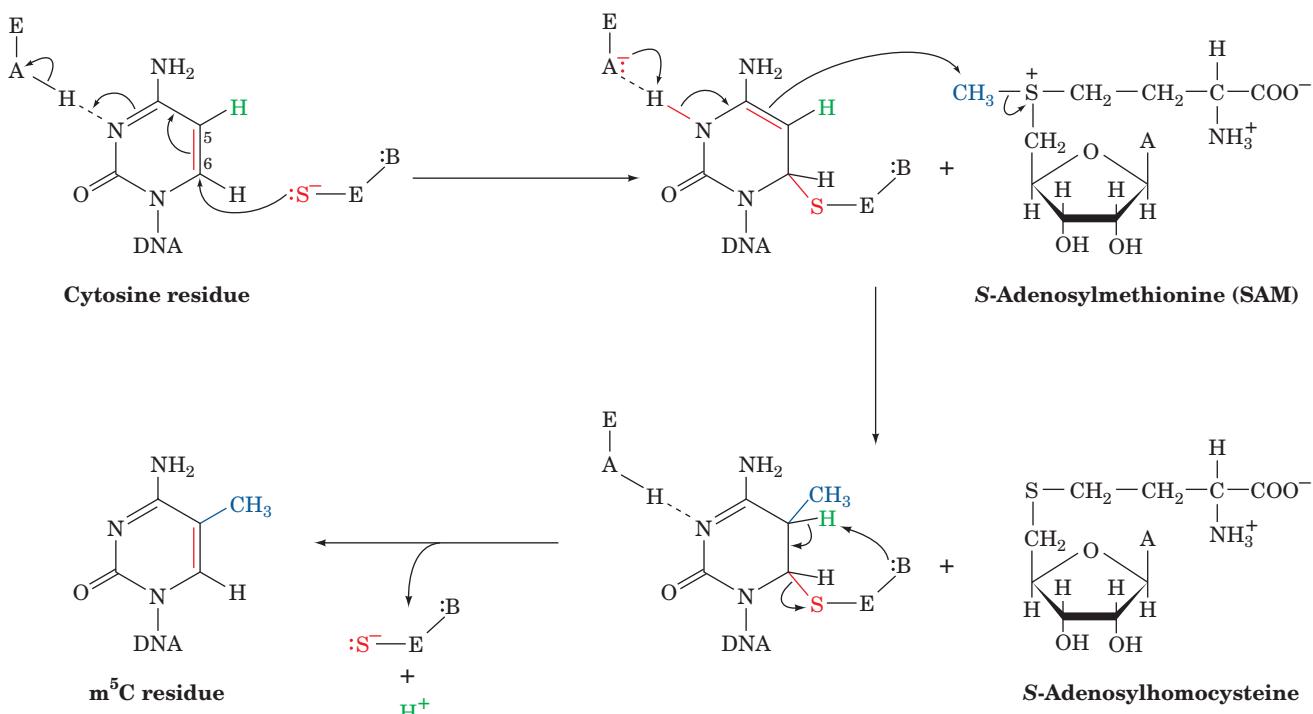
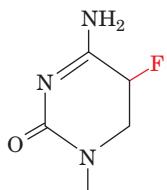


Figure 30-101 The catalytic mechanism of 5-methylcytosine methyltransferases (m⁵C-MTases). The methyl group is supplied by SAM, which thereby becomes *S*-adenosylhomocysteine. In M.HhaI, the DNA MTase from *Haemophilus haemolyticus*, the

active site thiolate group, $\text{S}^{\text{-}}\text{E}$, is on Cys 81, the enzyme general acid, $\text{E}-\text{A}$, is Glu 119, and the enzyme general base, $\text{E}-\text{B}$, has not been identified. [After Verdine, G.L., *Cell* **76**, 198 (1994).]

This mechanism is supported by the observation that the action of m⁵C-MTases on a **5-fluorocytosine (f⁵C) residue**



5-Fluorocytosine (f⁵C) residue

irreversibly traps the covalent intermediate (and hence inactivates the enzyme) because the enzyme cannot abstract fluorine, the most electronegative element, as an F⁺ ion (5-fluorodeoxyuridylate is likewise a suicide substrate for thymidylate synthase; Section 28-3Bc). Stereochemical principles dictate that the enzyme's Cys thiolate group can nucleophilically attack cytosine's C5 position only from above or below the ring. This is possible because, as we shall see below, the enzyme induces its cytosine target to flip out of the DNA double helix.

The DNA MTase from *Haemophilus haemolyticus* (**M.HhaI**), a 327-residue monomer, is a component of this bacterium's restriction-modification system. M.HhaI methylates its recognition sequence, 5'-GC_nGC-3' in double-stranded DNA, to yield 5'-G-m⁵C-GC-3'. Richard Roberts and Xiaodong Cheng determined the X-ray structure of the inactivated M.HhaI-DNA complex formed by incubating the enzyme with the self-complementary sequence d(TGATAG-**f⁵C**-GCTATC) (in which the enzyme's recognition sequence is in bold) in the presence of SAM. The DNA binds to the enzyme in a large cleft between its two unequally sized domains (Fig. 30-102). The structure's most striking feature is that the f⁵C nucleotide has flipped out of the minor groove in the otherwise largely undistorted B-DNA helix and has inserted into the enzyme's active site.

There, the f⁵C has reacted with SAM so as to yield adenosylhomocysteine (SAM without its methyl group) and the methylated intermediate covalently linked to Cys 81. The side chain of Gln 237 fills the cavity in the DNA double helix left by the departure of the f⁵C by hydrogen-bonding to the opposing G base. Comparison of this structure with that of M.HhaI in complex only with SAM indicates that on binding the DNA the protein's so-called active site loop (residues 80–99) swings around to contact the DNA, a movement of up to 25 Å. Nearly all base-specific interactions are made in the major groove by two Gly-rich loops (residues 233–240 and 250–257), the so-called recognition loops. The protein also makes extensive sequence-nonspecific contacts with DNA phosphate groups.

Base flipping was first observed in the above X-ray structure. However, as is now clear from the structures of other MTases as well as those of variety of DNA repair enzymes (e.g., Sections 30-5Aa and 30-5Be), *base flipping is a common mechanism through which enzymes gain access to the bases in dsDNA on which they perform chemistry*.

b. DNA Methylation in Eukaryotes Functions in Gene Regulation

5-Methylcytosine is the only methylated base in most eukaryotic DNAs, including those of vertebrates. This modification occurs largely in the CG dinucleotide of various palindromic sequences. CG is present in the vertebrate genome at only about one-fifth its randomly expected

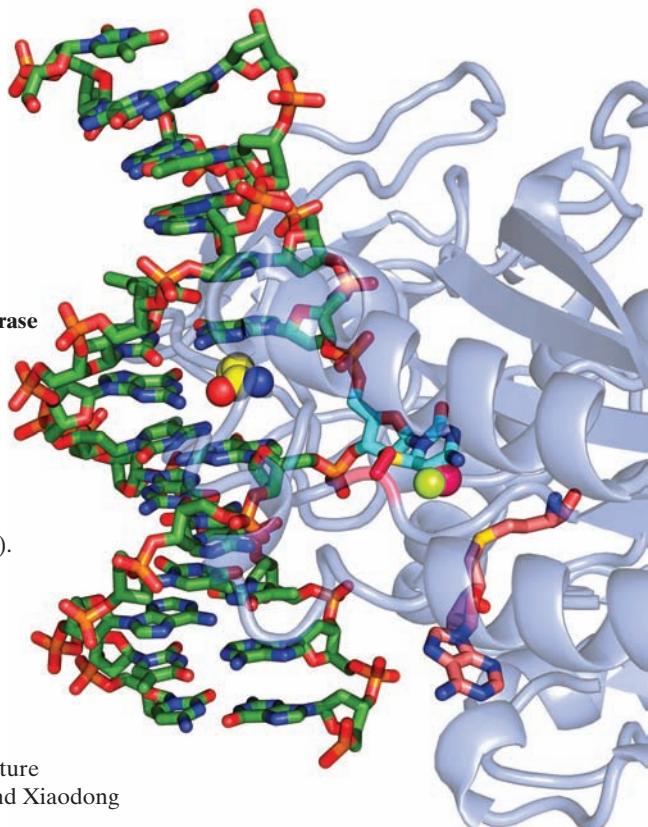


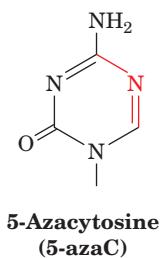
Figure 30-102 X-ray structure of the M.HhaI DNA methyltransferase in complex with S-adenosylhomocysteine and a dsDNA containing a methylated 5-fluorocytosine base at the enzyme's target site. The protein is represented by a semitransparent ribbon. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P orange). Its methylated 5-fluorocytosine residue (C atoms cyan) has swung out of the DNA helix into the enzyme's active site pocket, where its C6 forms a covalent bond with the S atom of an enzyme Cys residue (C atoms magenta and S yellow). The methyl group and a fluorine atom at C5 (which prevents the methylation reaction from going to completion) are represented by magenta and yellow-green spheres, respectively. The position of the flipped-out cytosine base in the DNA double helix is occupied by the side chain of a Gln residue (shown in space-filling form with C yellow), which hydrogen bonds to the "orphaned" guanine base. The S-adenosylmethionine, which has given up its methyl group, is drawn in stick form with its C atoms pink. [Based on an X-ray structure by Richard Roberts, New England Biolabs, Beverly, Massachusetts, and Xiaodong Cheng, Cold Spring Harbor Laboratory, New York. PDBid 1MHT.]

frequency. The upstream regions of many genes, however, have normal CG frequencies and are therefore known as **CpG islands**.

The degree of eukaryotic DNA methylation and its pattern are conveniently assessed by comparing the Southern blots (Section 5-5D) of DNA cleaved by the restriction endonucleases HpaII (which cleaves CCGG but not C-m⁵C-GG) and MspI (which cleaves both). Such studies indicate that eukaryotic DNA methylation varies with the species, the tissue, and the position along a chromosome.

The m⁵C residues in a given DNA segment can be identified through **bisulfite sequencing**, in which the DNA is reacted with **bisulfite ion** (HSO₃⁻), which selectively deaminates C (but not m⁵C) residues to U, followed by PCR amplification (Section 5-5F), which copies these U's to T's and the m⁵C's to C's. Comparison of the sequences of the amplified DNA with that of untreated DNA reveals which C's in the untreated DNA are methylated.

There is clear evidence that *DNA methylation switches off eukaryotic gene expression, particularly when it occurs in the promoter regions upstream of a gene's transcribed sequences*. For example, globin genes are less methylated in erythroid cells than they are in nonerythroid cells and, in fact, the specific methylation of the control region in a recombinant globin gene inhibits its transcription in transfected cells. In further support of the inhibitory effect of DNA methylation is the observation that **5-aza-cytosine (5-azaC)**,



a base analog that cannot be methylated at its N5 position and that inhibits DNA MTases, stimulates the synthesis of several proteins and changes the cellular differentiation patterns of cultured eukaryotic cells. The observation that repetitive intragenic parasites such as LINEs are highly methylated in somatic tissues has led to the hypothesis that CpG methylation in mammals arose to prevent the spurious transcriptional initiation of these retrotransposons.

The way in which DNA methylation prevents gene expression is poorly understood. One possibility is that DNA methylation can directly block the binding of transcriptional activators to their target sequences. However, in many cases, DNA methylation is recognized by a family of proteins that contain a conserved **methyl-CpG binding domain (MBD)**. Since the methyl groups of m⁵C residues extend into dsDNA's major groove, MBDs can bind to them without perturbing DNA's double helical structure. MBD-containing proteins inhibit the transcription of their bound promoter-methylated genes by recruiting protein complexes that induce the alteration of the local chromo-

some structure in a way that prevents the transcription of the associated gene (eukaryotic chromosome structure is discussed in Section 34-1). Another possibility has been raised by the observation that the methylation of synthetic poly(GC) stabilizes its Z-DNA conformation. Quite possibly, the formation of Z-DNA, which has been detected *in vivo* (Section 29-1Bb), acts as a conformational switch to turn off local gene expression.

c. DNA Methylation in Eukaryotes Is Self-Perpetuating

The palindromic nature of DNA methylation sites in eukaryotes permits the methylation pattern on a parental DNA strand to direct the generation of the same pattern in its daughter strand (Fig. 30-103). This **maintenance methylation** would result in the stable “inheritance” of a methylation pattern in a cell line and hence cause these cells to all have the same differentiated phenotype. Such changes to the genome are described as being **epigenetic** (Greek: *epi*,

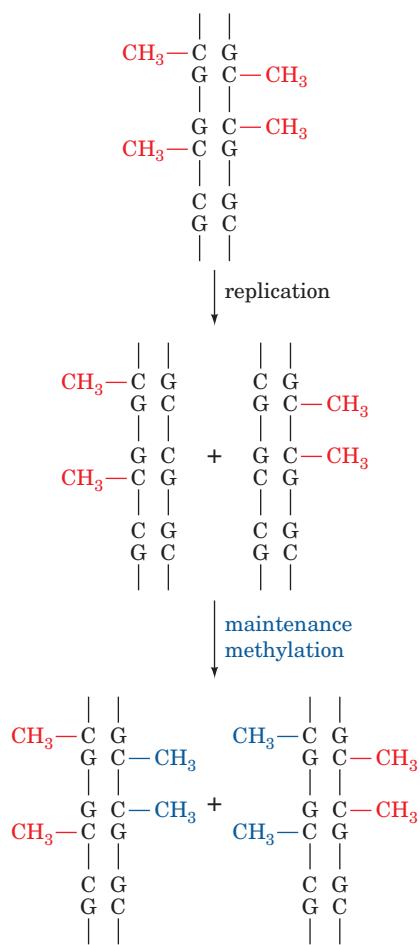


Figure 30-103 Maintenance methylation. The pattern of methylation on a parental DNA strand induces the corresponding methylation pattern in the complementary strand. In this way, a stable methylation pattern may be maintained in a cell line.

upon or beside) because they provide an additional layer of information that specifies when and where specific portions of the otherwise fixed genome are expressed (an epigenetic change that we have already encountered is the lengthening of telomeres in germ cells; Section 30-4D). Epigenetic characteristics, as we shall see, are not bound by the laws of Mendelian inheritance.

There is considerable experimental evidence favoring the existence of maintenance methylation, including the observation that artificially methylated viral DNA, on transfection into eukaryotic cells, maintains its methylation pattern for at least 30 cell generations. Maintenance methylation in mammals appears to be mediated mainly by the protein **DNMT1** (for *DNA methyltransferase 1*), which has a strong preference for methylating hemimethylated substrate DNAs. In contrast, prokaryotic DNA MTases such as M.HhaI do not differentiate between hemimethylated and unmethylated substrate DNAs. The importance of maintenance methylation is demonstrated by the observation that mice homozygous for deletion of the *DNMT1* gene die early in embryonic development.

The interaction energy provided a single methyl group in the major groove of DNA seems insufficient to permit a protein to reliably differentiate between an m^5C and a C residue. Nevertheless, the **SRA** (for *SET* and *RING*-associated) domain of the protein **UHRF1** (for *ubiquitin-like*, containing *PHD* and *RING* finger domains 1) does so and thereon recruits DNMT1 to the site. How does the SRA domain distinguish a single m^5C residue from unmethylated C?

The X-ray structure of the 210-residue SRA domain in complex with a 12-bp DNA containing a centrally located hemimethylated CpG was independently determined by Masahiro Shirakawa and Sirano Dhe-Paganon. It reveals that the DNA's m^5C residue has flipped out from the minor groove side of the largely straight B-form DNA to tightly bind deep within in a protein pocket (Fig. 30-104). Two loops reach around the resulting gap in the DNA helix from the major and minor grooves to interact with the other three bases of the hemimethylated CpG site. The m^5C base is replaced in the double helix by the side chain of an Arg residue extending from the minor groove loop, which hydrogen-bonds to the "orphaned" G residue without greatly disturbing its conformation.

DNMT1 has both a catalytic domain and a UHRF1-binding domain. The catalytic domain is thought to have a similar structure and mode of DNA binding as bacterial MTases such as M.HhaI (Fig. 30-102). However, a model of the DNMT1 catalytic domain–DNA–SRA domain complex based on the structures of the M.HhaI–DNA and SRA–DNA complexes indicates that it is unlikely that both the m^5C and C residues of a double helical CpG island could be simultaneously flipped out of the DNA. This suggests that the binding of the UHRF1-binding domain of DNMT1 to the SRA–DNA complex causes the m^5C to flip back into the DNA double helix while the C to be methylated flips out to bind to the catalytic domain of DNMT1.

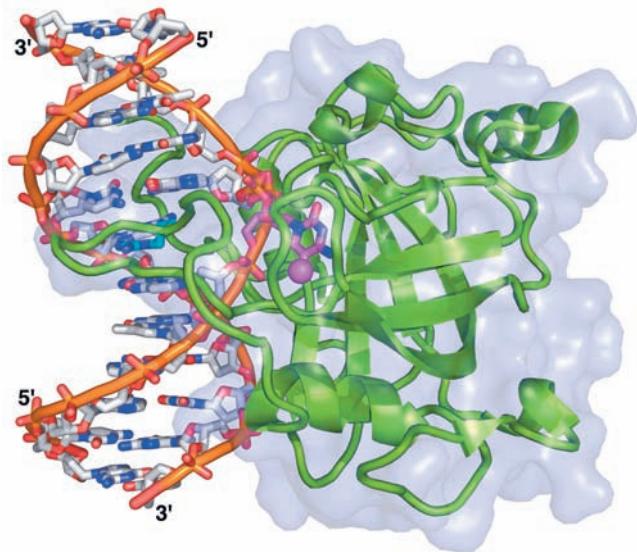


Figure 30-104 X-ray structure of the SRA domain of mouse **UHRF1** in complex with a hemimethylated 12-bp DNA. The protein is drawn in ribbon form (green) embedded in its semitransparent molecular surface. The DNA is shown in stick form colored according to atom type (DNA C gray except for the m^5C residue C which is magenta, N blue, O red, and P orange) with successive P atoms in the same strand connected by orange rods. The methyl group of the m^5C residue is represented by a magenta sphere. An Arg side chain (C cyan) fills the space in the DNA double helix vacated by the flipped-out m^5C . [Based on an X-ray structure by Masahiro Shirakawa, Kyoto University, Japan. PDBid 2ZDK.]

The pattern of DNA methylation in mammals varies in early embryological development. DNA methylation levels are high in mature gametes (sperm and ova) but are nearly eliminated by the time a fertilized ovum has become a **blastocyst** (a hollow ball of cells, the stage at which the embryo implants into the uterine wall; embryonic development is discussed in Section 34-4A). After this stage, however, the embryo's DNA methylation levels globally rise until, by the time the embryo has reached the developmental stage known as a **gastrula**, its DNA methylation levels have risen to adult levels, where they remain for the lifetime of the animal. This *de novo* (new) methylation appears to be mediated by two DNA MTases distinct from DNMT1 named **DNMT3a** and **DNMT3b**. An important exception to this remethylation process is that the CpG islands of germline cells (cells that give rise to sperm or ova) remain unmethylated. This ensures the faithful transmission of the CpG islands to the succeeding generation in the face of the strong mutagenic pressure of m^5C deamination (which yields T, a mutation that mismatch repair occasionally fails to correct).

The change in DNA methylation levels (epigenetic reprogramming) during embryonic development suggests that the pattern of genetic expression differs in embryonic and somatic cells. This explains the observed high failure

rate in cloning mammals (sheep, mice, cattle, etc.) by transferring the nucleus of an adult cell into an enucleated oocyte (immature ovum). Few of these animals survive to birth, many of those that do so die shortly thereafter, and most of the ~1% that do survive have a variety of abnormalities, most prominently an unusually large size. However, the survival of any embryos at all is indicative that the oocyte has the remarkable capacity to epigenetically reprogram somatic chromosomes (although it is rarely entirely successful in doing so) and that mammalian embryos are relatively tolerant of epigenetic abnormalities. Presumably, the reproductive cloning of humans from adult nuclei would result in similar abnormalities and for this reason (in addition to social and ethical prohibitions) should not be attempted.

d. Genomic Imprinting Results from Differential DNA Methylation

It has been known for thousands of years that maternal and paternal inheritance can differ. For example, a mule (the offspring of a mare and a male donkey) and a hinny (the offspring of a stallion and a female donkey) have obviously different physical characteristics, a hinny having shorter ears, a thicker mane and tail, and stronger legs than a mule. This is because, in mammals only, certain maternally and paternally supplied genes are differentially expressed, a phenomenon termed **genomic imprinting**. The genes that are subject to genomic imprinting are, as Rudolph Jaenisch has shown, differentially methylated in the two parents during gametogenesis and the resulting different methylation patterns are resistant to the wave of demethylation that occurs during the formation of the blastocyst and to the wave of *de novo* methylation that occurs thereafter.

The importance of genomic imprinting is demonstrated by the observation that an embryo derived from the transplantation of two male or two female pronuclei into an ovum fails to develop (pronuclei are the nuclei of mature sperm and ova before they fuse during fertilization). Inappropriate imprinting is also associated with certain diseases. For example, **Prader-Willi syndrome (PWS)**, which is characterized by the failure to thrive in infancy, small hands and feet, marked obesity, and variable mental retardation, is caused by a >5000-kb deletion in a specific region of the paternally inherited chromosome 15. In contrast, **Angelman syndrome (AS)**, which is manifested by severe mental retardation, a puppetlike ataxic (uncoordinated) gait, and bouts of inappropriate laughter, is caused by a deletion of the same region from the maternally inherited chromosome 15. These syndromes are also exhibited by those rare individuals who inherit both their chromosomes 15 from their mothers for PWS and from their fathers for AS. Evidently, certain genes on the deleted chromosomal region must be paternally inherited to avoid PWS and others must be maternally inherited to avoid AS. Several other human diseases are also associated with either maternal or paternal inheritance or lack thereof.

e. DNA Methylation Is Associated with Cancer

The mutation of an m^5C residue to T (with its associated G to A mutation on the complementary strand) is, by far, the most prevalent mutational change in human cancers. Such mutations usually convert proto-oncogenes to oncogenes (Section 19-3B) or inactivate tumor suppressors (Section 34-4Ca). In addition, the hypomethylation of proto-oncogenes and the hypermethylation of genes encoding tumor suppressors are associated with cancers, although it is unclear whether these are initiating or consolidating events for malignancies.

f. Several Neurological Diseases Are Associated with Trinucleotide Repeat Expansions

Fragile X syndrome, whose major symptoms include mental retardation and a characteristic long, narrow face with large ears, afflicts 1 in 4500 males and 1 in 9000 females. Fragile X syndrome is so named because, in affected individuals, the tip of the X chromosome's long arm is connected to the rest of the chromosome by a slender thread that is easily broken. The genetics of this condition are bizarre. The maternal grandfathers of individuals having fragile X syndrome may be asymptomatic, both clinically and cytogenetically. Their daughters are likewise asymptomatic, but these daughters' children of either sex may have the syndrome. Evidently, the fragile X defect is activated by passage through a female. Moreover, the **penetrance** of the disease, that is, the probability of a child having fragile X syndrome, and its severity, increase with each succeeding generation, a phenomenon termed **genetic anticipation**.

The affected gene in fragile X syndrome, *FMRI* (for *fragile X mental retardation 1*), encodes a 632-residue RNA-binding protein named **FMRP** (for *FMR protein*), which apparently functions in the transport of certain mRNAs from the nucleus to the cytoplasm (Section 34-3C), where it probably regulates their translation. FMRP, which is highly conserved in vertebrates, is expressed in most tissues but most heavily in brain neurons, where a variety of evidence indicates that its participation is required for the proper formation and/or function of synapses.

In the general population, the 5' untranslated region of *FMRI* contains a polymorphic $(CGG)_n$ sequence with n ranging from 6 to 60 and often punctuated by one or two AGG interruptions. However, in certain asymptomatic individuals, n has increased from 60 to 200, a so-called premutation that males transmit in unchanged form to their daughters (they transmit a Y rather than an X chromosome to their sons). In the daughters' children, however, ~80% of the individuals inheriting a premutant *FMRI* gene exhibit an astonishing expansion (amplification) of the triplet repeat with n ranging from >200 to several thousand, as well as the symptoms of the disease, a so-called full mutation. These triplet repeats differ in size among siblings and often exhibit heterogeneity within an individual, suggesting that they are somatically generated.

These **dynamic mutations**, which expand more often than they contract, perhaps arise through slippage of the DNA during replication. One way that slippage might occur

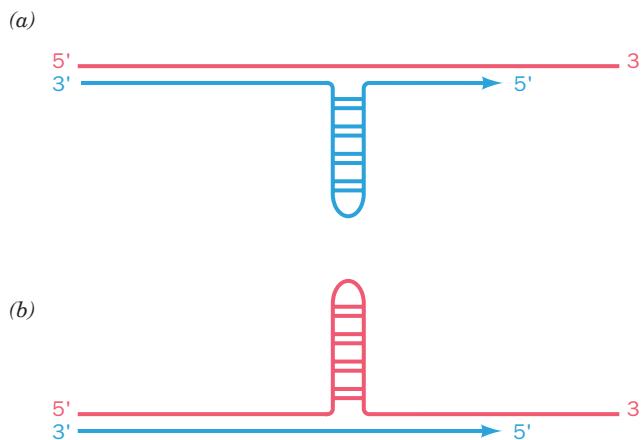


Figure 30-105 The loop-out mechanism for the alteration of the number of consecutive triplet repeats in DNA through its replication. Here the template strand is red and its nascent (newly synthesized) daughter strand is blue. With long tracts of repeating sequences, the probability of a loop-out occurring increases because its neighboring sequences will remain base-paired. The loop-out of a $(CNG)_n$ repeat is stabilized by its ability to form a partially base-paired structure. (a) If the daughter strand loops out, the number of repeats increases. (b) If the template strand loops out, the number of repeats decreases.

is through the formation of loop-outs (Fig. 30-105) on either the newly synthesized strand (causing expansions) or on the template strand (causing contractions). Such $(CNG)_n$ loops are likely stabilized by their ability to form imperfectly base-paired stems. As expected, the frequency of slippage increases with the number of repeats.

The peculiar genetics of fragile X syndrome is a consequence of genomic imprinting through methylation. The *FMRI* gene is unmethylated in normal individuals. However, it is hypermethylated in individuals with a maternally transmitted full mutation. This maintains the *FMRI* gene in a transcriptionally silent (inactive) state, thereby accounting for the symptoms of the disease. The lesser frequency and severity of fragile X syndrome in females are accounted for by the fact that females have two X chromosomes, one of which is unlikely to be mutated.

Over 30 other pathological instances of the expansion of DNA repeats, most of which are also neurological diseases, are known including the following:

1. Myotonic dystrophy (DM), the most common adult form of **muscular dystrophy** (its estimated incidence is 1 in 8000). It is a multisystem autosomal dominant disorder that is mainly characterized by progressive muscle weakness and wasting, the severity of which increases with successive generations while the age of onset decreases (genetic anticipation). Its most severe form, congenital DM, is

exclusively maternally transmitted. DM arises from a tri-nucleotide expansion in the 3' untranslated region of the gene encoding **myotonic dystrophy protein kinase (MDPK)**, which is expressed in the neurons affected by DM. The repeating triplet, $(CAG)_n$, is present in between 5 and 30 copies in the MDPK gene of normal individuals but expands from at least 50 repeats in minimally affected individuals to ~ 2000 repeats in severely affected individuals.

2. Huntington's disease (HD; previously called Huntington's chorea), a devastating neurodegenerative disorder characterized by progressively choreic (disordered) movements, cognitive decline, and emotional disturbances over an average 18-year course that is inevitably fatal. This dominant autosomal disorder, which affects ~ 1 in 10,000 individuals and has an average age of onset of ~ 40 years, is a consequence of the selective loss of certain groups of neurons in the brain. The *HD* gene, which encodes a widely expressed 3145-residue protein of unknown function named **huntingtin**, contains a polymorphic trinucleotide repeat, $(CAG)_n$, within its polypeptide coding sequence. The *HD* genes from 150 independent families with HD all contained between 37 and 86 repeat units, whereas those from normal individuals had 11 to 34 repeats. Moreover, the *HD* repeat length is unstable: $>80\%$ of meiotic transmissions show either increases or decreases, with the largest increases occurring in paternal transmissions (genomic imprinting). The number of repeats in afflicted individuals is inversely correlated with the age of onset of HD.

CAG is the codon for Gln (Table 5-3) and hence mutant huntingtin contains a long poly(Gln) tract. Synthetic poly(Gln) aggregates as β sheets that are linked by hydrogen bonds involving both their main chain and side chain amide groups. Indeed, the nuclei of HD-affected neurons contain inclusions that presumably consist of aggregates of huntingtin or its proteolytic products. It is these inclusions, as Max Perutz pointed out, that apparently kill the neurons in which they are contained, although the mechanism of how they do so is unknown. The long incubation period before the symptoms of HD become evident is attributed to the lengthy nucleation time for aggregate formation, much as occurs in the formation of amyloid fibrils (Section 9-5A).

3. Spinocerebellar ataxia (SCA) type 1, a progressive neurodegenerative disease whose age of onset is typically in the third or fourth decade, although it exhibits genetic anticipation. Like HD, it is caused by selective neuronal loss and is associated with an expansion of a CAG repeat in a coding region, in this case of a neuronal protein named **ataxin-1**. There it expands from ~ 28 to between 43 and 81 copies, thereby yielding a poly(Gln) tract of increased length (and tendency to aggregate). Four similar diseases, **SCA types 2, 3, 6, and 7**, are caused by $(CAG)_n$ expansions in different neuronal proteins.

CHAPTER SUMMARY

1 and 2 DNA Replication DNA is replicated in the 5' → 3' direction by the assembly of deoxynucleoside triphosphates on complementary DNA templates. Replication is initiated by the generation of short RNA primers, as mediated in *E. coli* by primase and RNA polymerase. The DNA is then extended from the 3' ends of the primers through the action of a DNA polymerase (Pol III in *E. coli*). The leading strand at a replication fork is synthesized continuously, whereas the lagging strand is synthesized discontinuously by the formation of Okazaki fragments. RNA primers on newly synthesized DNA are excised and replaced by DNA through Pol I-catalyzed (in *E. coli*) nick translation. The single-strand nicks are then sealed by DNA ligase. Mispairing errors during DNA synthesis are corrected by the 3' → 5' exonuclease functions of both Pol I and Pol III. The Klenow fragment of Pol I and other DNA polymerases of known structure have a right-hand-like structure with the active site located in the palm domain. Pol I recognizes the incoming nucleotide according to the shape of the base pair it forms with the template base and catalyzes the formation of a phosphodiester bond via a mechanism involving two metal ions. DNA synthesis by *E. coli* Pol III requires the participation of many auxiliary proteins including DNA gyrase, DnaB helicase, single-strand binding protein (SSB), primase, the β_2 sliding clamp, and DNA ligase.

3 Prokaryotic Replication DNA synthesis commences from specific sites known as replication origins. In the synthesis of the bacteriophage M13 (−) strand on the (+) strand template, the origin is recognized and primer synthesis is initiated by RNA polymerase. The analogous process in bacteriophage φX174, as well as in *E. coli*, is mediated by a complex primase-containing particle known as the primosome. φX174 (+) strands are synthesized according to the looped rolling circle mode of DNA replication on (−) strand templates of the replicative form in a process that is directed by the virus-specific gene *A* protein.

The *E. coli* chromosome is bidirectionally replicated in the θ mode from a single origin, *oriC*. It is recognized by DnaA protein, which together with DnaC protein, loads the DnaB helicase onto the developing replication forks. Leading strand synthesis is probably primed by RNA polymerase and primase working together, whereas Okazaki fragments are primed by primase in the primosome. The uncontrolled initiation of DNA replication is prevented by the sequestration of newly synthesized and hence hemimethylated *oriC* by membrane-associated SeqA protein, which prevents the *oriC* from becoming fully methylated at its multiple GATC sites. The β_2 sliding clamp, which is responsible for Pol III's processivity, is loaded onto the DNA by the $\gamma_3\delta\delta'$ clamp loader in an ATP-driven process. The δ subunit, when unmasked by ATP binding to the γ subunits, acts as a molecular “wrench” to spring open the sliding clamp, thereby permitting the entry of a single-stranded template DNA. Replication termination is facilitated by Tus protein which, on binding to an appropriately oriented *Ter* site, arrests the motion of a replication fork by binding to DnaB helicase. The great complexity of the DNA replication process functions to ensure the enormous fidelity necessary to maintain genome integrity.

4 Eukaryotic Replication Progression through the eukaryotic cell cycle is mediated by cyclins complexed to their

cognate cyclin-dependent protein kinases (Cdks). Chromosomal DNA replication is initiated by pol α/primase, which synthesizes a primer followed by a short length of DNA. Then, via polymerase switching mediated by replication factor C (RFC), the eukaryotic clamp loader, pol δ processively synthesizes both the lagging and leading strands in complex with PCNA, the eukaryotic sliding clamp.

Eukaryotic chromosomal DNA is synthesized in multiple origin-containing segments known as replicons. Nevertheless, chromosomal DNA is synthesized once and only once per cell cycle. The re-replication of DNA is prevented because replication initiation is licensed only in the G₁ phase of the cell cycle by the formation of the pre-replicative complex (pre-RC) but DNA is synthesized only in S phase by the activation of the pre-RC. The pre-RC is assembled in G₁ phase by the binding of the origin recognition complex (ORC) to an origin, which recruits Cdc6/Cdc18 and Cdt1 followed by the MCM complex, the replicative helicase. The activation of the pre-RC begins in S phase with the addition of Mcm10 followed by the phosphorylation of many of the pre-RC's subunits by Cdks and Ddk. Cdc45 then binds followed by pol α/primase, pol ε, PCNA, and the replication protein A (RPA), the SSB counterpart, to yield the active initiation complex. Re-replication is prevented through the actions of Cdks, which cause the elimination of Cdc6/Cdc18 and inhibit the helicase activity of the MCM complex. In metazoan cells, re-replication is also prevented by the binding of geminin to Cdt1. Mitochondrial DNA is replicated by polymerase γ via a process in which the Okazaki fragments are entirely RNA.

Retroviruses produce DNA on RNA templates in a reaction sequence catalyzed by reverse transcriptase. Telomeric DNA, a G-rich repeating octamer on the 3'-ending strand, is synthesized by the RNA-containing enzyme telomerase. Telomerase is active in germ cells but not somatic cells, a phenomenon that is at least in part responsible for cellular senescence and aging. The observation that telomerase is active in nearly all cancer cells suggests that telomerase inactivation is a defense against the development of cancer. The free DNA ends of telomeres are capped to prevent them from triggering DNA damage checkpoints. The *O. nova* telomere heterodimer TEBP binds both single strands of telomere DNA and a G-quartet-containing dimer. Its counterparts in humans and yeast are POT1 and TPP1. Telomeric DNA forms T-loops which are formed by the TRF2-mediated invasion of the 3' telomeric overhang into repeating telomeric dsDNA to form a D-loop. POT1, TPP1, TRF1, together with TRF2, TIN2, and RAP1, form a complex on the telomere called shelterin that functions to protect telomeres from DNA repair processes and limit the size of the telomere.

5 Repair of DNA Cells have a great variety of DNA repair mechanisms. DNA damage may be directly reversed such as in the photoreactivation of UV-induced pyrimidine dimers or in the repair of O⁶-alkylguanine lesions by the transfer of the offending alkyl group to a repair protein. Pyrimidine dimers, as well as many other types of DNA lesions, may also be removed by nucleotide excision repair (NER), which in *E. coli* involves the UvrABC system. Xeroderma pigmentosum, an inherited human disease characterized by marked UV-induced skin changes and a greatly increased incidence of

cancer, is caused by defects in any of seven complementation groups that participate in NER. Cockayne syndrome is associated with defects in the proteins responsible for transcription-coupled repair (TCR), which repairs lesions in actively transcribing genes. In base excision repair (BER), DNA glycosylases specifically remove the corresponding chemically altered bases, including uracil, through mechanisms that involve base flip-outs to form AP sites. The AP sites are cleaved on one side by an AP endonuclease, removed together with adjacent residues by an exonuclease, and replaced through the actions of a DNA polymerase and a DNA ligase. In mismatch repair (MMR), base-pairing mismatches arising from replication errors are corrected. In *E. coli* MMR, MutS and MutL bind to the mismatch and then identify the daughter strand, which contains the error, according to which strand of the nearest hemimethylated GATC palindrome is unmethylated. MutH then cleaves this strand, which is excised past the mismatch and replaced.

DNA damage in *E. coli* induces the SOS response, a LexA- and RecA-mediated process in which the error-prone bypass DNA polymerases Pol IV and Pol V replicate a damaged template DNA even if it provides no information as to which base to incorporate. Double-strand break (DSB) repair by nonhomologous end-joining (NHEJ) is facilitated by Ku protein, which holds two dsDNA ends together for ligation by DNA ligase IV in complex with XrcC4. The high correlation between mutagenesis and carcinogenesis permits the detection of carcinogens by the Ames test.

6 Recombination and Mobile Genetic Elements Genetic information may be exchanged between homologous DNA sequences through homologous recombination, a process that occurs according to the Holliday model. In *E. coli*, strand invasion to form Holliday junctions is mediated by RecA after the RecBCD-mediated generation of the single-strand nicks to which RecA binds. Branch migration is mediated by RuvAB, which consists of a homotetramer (or a homooctamer) of RuvA, which binds both a Holliday junction and two oppositely located RuvB hexamers. In an ATP-driven process, the RuvB hexamers (in effect) counter-rotate to pump the dsDNA stems into the center of the RuvA-bound Holliday junction, where each of its single strands exchange base-pairing partners to form new dsDNA stems, which are translocated toward the periphery of the complex. The Holliday junction is eventually resolved to its component dsDNAs by RuvC and sealed by DNA ligase.

The primary function of homologous recombination is to repair damaged replication forks resulting from the encounters of replisomes with unrepaired single-strand lesions or breaks. DSBs may be rejoined via a recombination repair process called homologous end-joining.

Chromosomes and plasmids may be rearranged through the action of transposons. These DNA segments carry the genes that encode the proteins that mediate the transposition process as well as other genes. Tn5 transposase catalyzes the cut-and-paste transposition of the Tn5 transposon. Replicative transposition proceeds via the intermediacy of cointegrates, which are resolved through the action of enzymes such as the $\gamma\delta$ resolvase. Transposition may be important in chromosomal and plasmid evolution and has been implicated in the control of phenotypic expression such as phase alternation in *Salmonella*, a process that is catalyzed by the Hin DNA invertase, a homolog of the $\gamma\delta$ resolvase. Members of the λ integrase family of proteins, such as Cre recombinase, insert dsDNA segments into their target sites via a Holliday junction intermediate in which transient covalent bonds are formed between active site Tyr side chains and the 3'-OH groups at the cleavage sites. Retrotransposons undergo transposition through an RNA intermediate. Many retrotransposons, such as yeast Ty1, are “internal” retroviruses that can only replicate within a genome. Nonviral retrotransposons, such as LINEs, the dominant transposons in the human genome, have a different transpositional mechanism.

7 DNA Methylation and Trinucleotide Repeat Expansions Prokaryotic DNA may be methylated at its A or C bases. This prevents the action of restriction endonucleases and permits the correct mismatch repair of newly replicated DNA. In most eukaryotes, DNA methylation, which occurs, mainly at CpG islands, through the formation of m⁵C, has been implicated in the control of gene expression and, via maintenance methylation, in genomic imprinting.

Over 30 inherited, mostly neurological diseases, including fragile X syndrome, myotonic dystrophy, and Huntington’s disease, are characterized by the genetically bizarre expansion of segments of (CNG)_n triplets. If an expanded triplet repeat occurs in an upstream noncoding region of a gene, its aberrant methylation, perhaps through genomic imprinting, may lead to the gene’s transcriptional silencing; if the expanded repeat is instead manifested as a poly(Gln) tract in a protein, the resulting protein aggregates may kill the neurons in which it occurs.

REFERENCES

General

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *The Molecular Biology of the Cell* (5th ed.), Chap. 5, Garland Science (2008).
- Kornberg, A., *For Love of Enzymes: The Odyssey of a Biochemist*, Harvard University Press (1989). [A scientific autobiography.]
- Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), Freeman (1992). [A compendium of information about DNA replication whose first author is the founder of the field.]
- Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R., *Molecular Biology of the Gene* (6th ed.), Chapters 8–11, Cold Spring Harbor Laboratory Press (2008).

Prokaryotic DNA Replication

- Bailey, S., Eliason, W.K., and Steitz, T.A., Structure of hexameric DnaB helicase and its complex with a domain of DnaG protein, *Science* **318**, 459–463 (2007).
- Beese, L.S., Derbyshire, V., and Steitz, T.A., Structure of DNA polymerase I Klenow fragment bound to duplex DNA, *Science* **260**, 352–355 (1993).
- Benkovic, J.J., Valentine, A.M., and Salinas, F., Replisome-mediated DNA replication, *Annu. Rev. Biochem.* **70**, 181–208 (2001).
- Carr, K.M. and Kaguni, J.M., Stoichiometry of DnaA and DnaB protein in initiation at the *Escherichia coli* chromosomal origin, *J. Biol. Chem.* **276**, 44919–44925 (2001).

- Caruthers, J.M. and McKay, D.B., Helicase structure and mechanism, *Curr. Opin. Struct. Biol.* **12**, 123–133 (2002).
- Corn, J.E. and Berger, J.M., Regulation of bacterial priming and daughter strand synthesis through helicase–primase interactions, *Nucleic Acids Res.* **34**, 4082–4087 (2006).
- Corn, J.E., Pelton, J.G., and Berger, J.M., Identification of a DNA primase template tracking site redefines the geometry of primer synthesis, *Nature Struct. Mol. Biol.* **15**, 16–169 (2008).
- Davey, M.J., Jeruzalmi, D., Kuriyan, J., and O'Donnell, M., Motors and switches: AAA⁺ machines within the replisome, *Nature Rev. Mol. Cell Biol.* **3**, 1–10 (2002).
- Doublie, S., Sawaya, M.R., and Ellenberger, T., An open and closed case for all polymerases, *Structure* **7**, R31–R35 (1999). [Reviews the mechanisms of DNA polymerases.]
- Duerstadt, K.E. and Berger, J.M., AAA⁺ ATPases in the initiation of DNA replication, *Crit. Rev. Biochem. Mol. Biol.* **43**, 163–187 (2008).
- Enemark, E.J. and Joshua-Tor, L., Mechanism of DNA translocation in a replicative hexameric helicase, *Nature* **442**, 270–275 (2006).
- Frick, D.N. and Richardson, C.C., DNA primases, *Annu. Rev. Biochem.* **70**, 39–80 (2001).
- Georgescu, R.E., Kim, S.-S., Yurieva, O., Kuriyan, J., Kong, X.-P., and O'Donnell, M., Structure of a sliding clamp on DNA, *Cell* **132**, 43–54 (2008).
- Hamdan, S.M. and Richardson, C.C., Motors, switches, and contacts in the replisome, *Annu. Rev. Biochem.* **78**, 205–243 (2009).
- Jeruzalmi, D., O'Donnell, M., and Kuriyan, J., Clamp loaders and sliding clamps, *Curr. Opin. Struct. Biol.* **12**, 217–224 (2002).
- Kaguni, J.M., DnaA: Controlling the initiation of bacterial DNA replication and more, *Annu. Rev. Microbiol.* **60**, 351–371 (2006).
- Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., and Morikawa, K., Structure of a replication-terminator protein complexed with DNA, *Nature* **383**, 598–603 (1996).
- Keck, J.L., Roche, D.D., Lynch, A.S., and Berger, J.M., Structure of the RNA polymerase domain of *E. coli* primase, *Science* **287**, 2482–2486 (2000); and Podobnik, M., McInerney, P., O'Donnell, M., and Kuriyan, J., A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases, *J. Mol. Biol.* **300**, 353–362 (2000).
- Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W., and Steitz, T.A., Crystal structure of *Thermus aquaticus* DNA polymerase, *Nature* **376**, 612–616 (1995).
- Kong, X.-P., Onrust, R., O'Donnell, M., and Kuriyan, J., Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp, *Cell* **69**, 425–437 (1992).
- Kool, E.T., Active site tightness and substrate fit in DNA replication, *Annu. Rev. Biochem.* **71**, 191–219 (2002); and Hydrogen-bonding, base stacking, and steric effects in DNA replication, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1–2 (2001).
- Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G., Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helices bound to single-stranded DNA and ADP, *Cell* **90**, 635–647 (1997).
- Kunkel, T.A. and Bebenek, K., DNA replication fidelity, *Annu. Rev. Biochem.* **69**, 497–529 (2000).
- Lamers, M.H., Georgescu, R.E., Lee, S.-G., O'Donnell, M., and Kuriyan, J., Crystal structure of the catalytic α subunit of *E. coli* replicative DNA polymerase III, *Cell* **126**, 881–892 (2006); and Bailey, S., Wing, R.A., and Steitz, T.A., The structure of *T. aquaticus* polymerase III is distinct from eukaryotic replicative DNA polymerases, *Cell* **126**, 893–904 (2006).
- Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J.K., Kim, H.-K., Kwon, S.-T., and Suh, S.W., Crystal structure of NAD⁺-dependent DNA ligase: modular architecture and functional implications, *EMBO J.* **19**, 1119–1129 (2000).
- Li, Y., Korolev, S., and Waksman, G., Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation, *EMBO J.* **17**, 7514–7525 (1998).
- Mott, M.L. and Berger, J.M., DNA replication initiation: mechanisms and regulation in bacteria, *Nature Rev. Microbiol.* **5**, 343–354 (2007).
- Mott, M.L., Erzberger, J.P., Coons, M.M., and Berger, J.M., Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators, *Cell* **134**, 623–634 (2008). [The X-ray structure of DnaC protein.]
- Mulcair, M.D., Schaeffer, P.M., Oakley, A.J., Cross, H.F., Neylon, C., Hill, T.M., and Dixon, N.E., A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*, *Cell* **125**, 1309–1319 (2006).
- Nandakumar, J., Nair, P.A., and Shuman, S., Last stop on the road to repair: Structure of *E. coli* DNA ligase bound to nicked DNA-adenylate, *Mol. Cell* **26**, 257–271 (2007); and Shuman, S., DNA ligase: progress and prospects, *J. Biol. Chem.* **284**, 17365–17369 (2009).
- O'Donnell, M., Replisome architecture and dynamics in *E. coli*, *J. Biol. Chem.* **281**, 10653–10656 (2006); and Johnson, A. and O'Donnell, M., Cellular DNA replicases: Components and dynamics at the replication fork, *Annu. Rev. Biochem.* **74**, 283–315 (2005).
- Patel, S.S. and Donmez, I., Mechanisms of helicases, *J. Biol. Chem.* **281**, 18265–18268 (2006); and Patel, S.S. and Picha, K.M., Structure and function of hexameric helicases, *Annu. Rev. Biochem.* **69**, 651–697 (2000).
- Raghunathan, S., Kozlov, A.G., Lohman, T.M., and Waksman, G., Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA, *Nature Struct. Biol.* **7**, 648–652 (2000).
- Rothwell, P.J. and Waksman, G., Structure and mechanism of DNA polymerases, *Adv. Prot. Chem.* **71**, 401–440 (2005).
- Simonetta, K.R., Kazmirska, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M., and Kuriyan, J., The mechanism of ATP-dependent primer-template recognition by a clamp loader complex, *Cell* **137**, 659–671 (2009).
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B., Structure and mechanism of helicases and nucleic acid translocases, *Annu. Rev. Biochem.* **76**, 23–50 (2007).
- Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B., Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides, *Cell* **101**, 589–600 (2000).
- Soultanas, P. and Wigley, D.B., Unwinding the 'Gordian knot' of helicase action, *Trends Biochem. Sci.* **26**, 47–54 (2001).
- Watson, J.D. and Crick, F.H.C., Genetical implications of the structure of deoxyribonucleic acid, *Nature* **171**, 964–967 (1953). [The paper in which semiconservative DNA replication was first postulated.]

Eukaryotic DNA Replication

- Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W., and Harley, C.B., Telomere length predicts replicative capacity of human fibroblasts, *Proc. Natl. Acad. Sci.* **89**, 10114–10118 (1992).
- Arezi, B. and Kuchta, R.D., Eukaryotic DNA primase, *Trends Biochem. Sci.* **25**, 572–576 (2000).
- Armanios, M., Syndromes of telomere shortening, *Annu. Rev. Genomics Hum. Genet.* **10**, 45–61 (2009).

- Autexier, C. and Lue, N.F., The structure and function of telomerase reverse transcriptase, *Annu. Rev. Biochem.* **75**, 493–517 (2006).
- Bell, S.P. and Dutta, A., DNA replication in eukaryotic cells, *Annu. Rev. Biochem.* **71**, 333–374 (2002).
- Blackburn, E.H., Switching and signaling at the telomere, *Cell* **106**, 661–673 (2001); and Telomere states and cell fates, *Nature* **408**, 53–56 (2000).
- Blackburn, E.H., Greider, C., and Szostak, J.W., Telomeres and telomerase: the path from maize, *Tetrahymena* and yeast to human cancer and aging, *Nature Med.* **12**, 1133–1138 (2006). [A scientific memoir.]
- Blow, J.J. and Dutta, A., Preventing re-replication of chromosomal DNA, *Nature Rev. Mol. Cell Biol.* **6**, 476–486 (2005).
- Bowman, G.D., O'Donnell, M., and Kuriyan, J., Structural analysis of a eukaryotic sliding clamp–clamp loader complex, *Nature* **429**, 724–730 (2004).
- Burger, P.M.J., Polymerase dynamics at the eukaryotic replication fork, *J. Biol. Chem.* **284**, 4041–4045 (2009). [Discusses the roles of pol δ, pol ε and FEN1.]
- Cech, T.R., Life at the end of the chromosome: Telomeres and telomerase, *Angew. Chem.* **39**, 34–43 (2000).
- Croy, J.E. and Wuttke, D.S., Themes in ssDNA recognition by telomere-end protection proteins, *Trends Biochem. Sci.* **31**, 516–525 (2006).
- de Lange, T., Lundblad, V., and Blackburn, E. (Eds.), *Telomeres* (2nd ed.), Cold Spring Harbor Laboratory Press (2006).
- DePamphilis, M.L., Replication origins in metazoan chromosomes: fact or fiction, *BioEssays* **21**, 5–16 (1999).
- Diffley, J.F.X., DNA replication: Building the perfect switch, *Curr. Biol.* **11**, R367–R370 (2001).
- Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark, A.D., Jr., Jacobo-Molina, A., Tantillo, C., Hughes, S.H., and Arnold, E., Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution, *J. Mol. Biol.* **284**, 1095–1111 (1998).
- Franklin, M.C., Wang, J., and Steitz, T.A., Structure of the replicating complex of a pol α family DNA polymerase, *Cell* **105**, 657–667 (2001).
- Ellenberger, T. and Tomkinson, A.E., Eukaryotic DNA ligases: structural and functional insights, *Annu. Rev. Biochem.* **77**, 313–338 (2008).
- Garg, P. and Burgers, P.M.J., DNA polymerases that propagate the eukaryotic replication fork, *Crit. Rev. Biochem. Mol. Biol.* **40**, 115–128 (2005).
- Gilbert, D.M., Making sense out of eukaryotic DNA replication origins, *Science* **294**, 96–100 (2001).
- Gillis, A.J., Schuller, A.P., and Skordalakes, E., Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT, *Nature* **455**, 633–637 (2008).
- Gilson, E. and Géli, V., How telomeres are replicated, *Nature Rev. Mol. Cell Biol.* **8**, 825–838 (2007).
- Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T., Mammalian telomeres end in a large duplex loop, *Cell* **97**, 503–514 (1999).
- Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W., and Weinberg, R.A., Creation of human tumour cells with defined genetic elements, *Nature* **400**, 464–468 (1999).
- Harrison, J.C. and Haber, J.E., Surviving the breakup: the DNA damage checkpoint, *Annu. Rev. Genet.* **40**, 209–235 (2006).
- Holt, I.J., Mitochondrial DNA replication and repair: all a flap, *Trends Biochem. Sci.* **34**, 358–365 (2009).
- Horvath, M.P. and Schultz, S.C., DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex, *J. Mol. Biol.* **310**, 367–377 (2001).
- Hübscher, U., Maga, G., and Spadari, S., Eukaryotic DNA polymerases, *Annu. Rev. Biochem.* **71**, 133–163 (2002).
- Johansson, E. and MacNeil, S.A., The eukaryotic replicative DNA polymerase takes shape, *Trends Biochem. Sci.* **35**, 339–347 (2010).
- Kelleher, C., Teixeira, M.T., Förstemann, K., and Lingner, J., Telomerase: Biochemical considerations for enzyme and substrate, *Trends Biochem. Sci.* **27**, 572–579 (2002).
- Kelly, T.J. and Brown, G.W., Regulation of chromosome replication, *Annu. Rev. Biochem.* **69**, 829–880 (2000).
- Lansdorp, P.M., Major cutbacks at chromosome ends, *Trends Biochem. Sci.* **30**, 388–395 (2005) [Reviews mechanisms of telomere erosion.]
- Liu, Y., Kao, H.-I., and Bambara, R.A., Flap endonuclease 1: A central component of DNA metabolism, *Annu. Rev. Biochem.* **73**, 589–615 (2004).
- Machida, Y.J., Hamlin, J.L., and Dutta, A., Right place, right time, and only once: Replication initiation in metazoans, *Cell* **123**, 13–24 (2005); and Machida, Y.J. and Dutta, A., Cellular checkpoint mechanisms monitoring proper initiation of DNA replication, *J. Biol. Chem.* **280**, 6253–6256 (2005).
- Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M., Eukaryotic chromosome DNA replication: Where, when, and how? *Annu. Rev. Biochem.* **79**, 89–130 (2010).
- McCulloch, S.D. and Kunkel, T.A., The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases, *Cell Res.* **18**, 148–161 (2008).
- McEachern, M.J., Krauskopf, A., and Blackburn, E.H., Telomeres and their control, *Annu. Rev. Genet.* **34**, 331–358 (2000).
- Moldovan, G.-L., Pfander, B., and Jentsch, S., PCNA, the maestro of the replication fork, *Cell* **129**, 665–679 (2007).
- Neidle, S. and Parkinson, G., Telomere maintenance as a target for anticancer drug discovery, *Nature Rev. Drug Discov.* **1**, 383–393 (2002); and The structure of telomeric DNA, *Curr. Opin. Struct. Biol.* **13**, 275 (2003).
- Osterhage, J.L. and Friedman, K.L., Chromosome end maintenance by telomerase, *J. Biol. Chem.* **284**, 16061–16065 (2009).
- O'Sullivan, R.J. and Karlseder, J., Telomeres: protecting chromosomes against genome instability, *Nature Rev. Mol. Cell Biol.* **11**, 171–181 (2010).
- Pursell, Z.F., Isoz, I., Lundström, E.-B., Johansson, E., and Kunkel, T.A., Yeast DNA polymerase ε participates in leading-strand replication, *Science* **317**, 127–130 (2007).
- Riethman, H., Human telomere structure and biology, *Annu. Rev. Genomics Hum. Genet.* **9**, 1–19 (2008).
- Schultze, P., Smith, F.W., and Feigon, J., Refined solution structure of the dimeric quadruplex formed from the *Oxytricha* telomeric oligonucleotide d(GGGGTTTTGGGG), *Structure* **2**, 221–233 (1994).
- Smogorzewska, A., and de Lange, T., Regulation of telomerase by telomeric proteins, *Annu. Rev. Biochem.* **73**, 177–208 (2004).
- Swan, M.K., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K., Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ, *Nature Struct. Mol. Biol.* **16**, 979–986 (2009).
- Tye, B.K. and Sawyer, S., The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts, *J. Biol. Chem.* **275**, 34833–34836 (2000); and Tye, B.K., MCM proteins in DNA replication, *Annu. Rev. Biochem.* **68**, 649–686 (1999).
- Urquidi, V., Tarin, D., and Goddison, S., Role of telomerase in cell senescence and oncogenesis, *Annu. Rev. Med.* **51**, 65–79 (2000).
- Verdun, R.E. and Karlseder, J., Replication and protection of telomeres, *Nature* **447**, 924–931 (2007).

Repair of DNA

- Ames, B.N., Identifying environmental chemicals causing mutations and cancer, *Science* **204**, 587–593 (1979).
- Beckman, K.B. and Ames, B.N., Oxidative decay of DNA, *J. Biol. Chem.* **272**, 19633–19636 (1997).
- Broyde, S., Wang, L., Rechkoblit, O., Geacintov, N.E., and Patel, D.J., Lesion processing: high-fidelity versus lesion-bypass DNA polymerases, *Trends Biochem. Sci.* **33**, 209–219 (2008).
- David, S.S., O'Shea, V.L., and Kundu, S., Base excision repair of oxidative DNA damage, *Nature* **447**, 941–950 (2007).
- Deaconescu, A.M., Savery, N., and Darst, S.A., The bacterial transcription repair coupling factor, *Curr. Opin. Struct. Biol.* **17**, 96–102 (2007).
- Friedberg, E.C., Wagner, R., and Radman, M., Specialized DNA polymerases, cellular survival, and the genesis of mutations, *Science* **296**, 1627–1630 (2002).
- Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T., *DNA Repair and Mutagenesis* (2nd ed.), ASM Press (2006).
- Garber, P.M., Vidanes, G., and Toczyński, D.P., Damage in transition, *Trends Biochem. Sci.* **30**, 63–66 (2005). [Describes how broken DNA can be repaired by NHEJ or recombination.]
- Goodman, M.F., Error-prone repair DNA polymerases in prokaryotes and eukaryotes, *Annu. Rev. Biochem.* **71**, 17–50 (2002).
- Harfe, B.D. and Jinks-Robertson, S., DNA mismatch repair and genetic instability, *Annu. Rev. Genet.* **34**, 359–399 (2000).
- Heller, R.C. and Marians, K.J., Replisome assembly and the direct restart of replication forks, *Nature Rev. Mol. Cell Biol.* **7**, 932–943 (2006).
- Hopfner, K.-P., Putnam, C.D., and Tainer, J.A., DNA double-strand break repair from head to tail, *Curr. Opin. Struct. Biol.* **12**, 115–122 (2002).
- Kunkel, T.A. and Erie, D.A., DNA mismatch repair, *Annu. Rev. Biochem.* **74**, 681–710 (2005).
- Lieber, M.R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway, *Annu. Rev. Biochem.* **79**, 181–211 (2010).
- Ling, H., Boudsocq, F., Woogate, R., and Yang, W., Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication, *Cell* **107**, 91–102 (2001).
- Mees, A., Klar, T., Gnau, P., Hennecke, U., Eker, A.P.M., Carell, T., and Essen, L.-O., Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair, *Science* **306**, 1789–1793 (2004).
- McCullough, A.K., Dodson, M.L., and Lloyd, R.S., Initiation of base excision repair: glycosylase mechanism and structures, *Annu. Rev. Biochem.* **68**, 255–285 (1999).
- Modrich, P., Mechanisms of eukaryotic mismatch repair, *J. Biol. Chem.* **281**, 30305–30309 (2006).
- Mol, C.D., Parikh, S.S., Putnam, C.D., Lo, T.P., and Tainer, J.A., DNA repair mechanism for the recognition and removal of damaged DNA bases, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 101–128 (1999).
- Moore, M.H., Gulbis, J.M., Dodson, E.J., Demple, B., and Moody, P.C.E., Crystal structure of a suicidal DNA repair protein: Ada O⁶-methylguanine-DNA methyltransferase from *E. coli*, *EMBO J.* **13**, 1495–1501 (1994).
- Myers, L.C., Verdine, G.L., and Wagner, G., Solution structure of the DNA methyl triester repair domain of *Escherichia coli* Ada, *Biochemistry* **32**, 14089–14094 (1993).
- Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E., and Tainer, J.A., Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA, *EMBO J.* **17**, 5214–5226 (1998).
- Pham, P., Rangarajan, S., Woodgate, R., and Goodman, M.F., Roles of DNA polymerases V and II in SOS-induced error-prone and error-free repair in *Escherichia coli*, *Proc. Natl. Acad. Sci.* **98**, 8350–8354 (2001); and Goodman, M.F., Coping with replication 'train wrecks' in *Escherichia coli* using Pol V, Pol II, and RecA proteins, *Trends Biochem. Sci.* **25**, 189–195 (2000).
- Prakash, S., Johnson, R.E., and Prakash, L., Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function, *Annu. Rev. Biochem.* **74**, 317–353 (2005).
- Sancar, A., Lindsey-Bolz, L.A., Ünsal-Kaçmaz, K., and Linn, S., Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints, *Annu. Rev. Biochem.* **73**, 39–85 (2005).
- Sarker, A.H., et al., Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIB: Insights for transcription-coupled repair and Cockayne syndrome, *Mol. Cell* **20**, 187–198 (2005).
- Sutton, M.D., Smith, B.T., Godoy, V.G., and Walker, G.C., The SOS response: recent insights into *umuDC*-dependent mutagenesis and DNA damage tolerance, *Annu. Rev. Genet.* **34**, 479–497 (2000).
- Tainer, J.A. and Friedberg, E.C. (Eds.), *Biological Implications from Structures of DNA Repair Proteins*, *Mutat. Res.* **460**, 139–335 (2000). [A series of authoritative reviews.]
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Chapters 28 and 32 contain discussions of xeroderma pigmentosum, Cockayne syndrome, and hereditary nonpolyposis colorectal cancer.]
- Walker, J.R., Corpina, R.A., and Goldberg, J., Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair, *Nature* **412**, 607–614 (2001).
- Wyman, C. and Kanaar, R., DNA double-strand break repair: All's well that ends well, *Annu. Rev. Genet.* **40**, 363–383 (2006).
- Yang, W. (Ed.), *DNA Replication and Repair*, *Adv. Prot. Chem.* **69** (2004).

Recombination and Mobile Genetic Elements

- Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K., Crystal structure of the Holliday junction DNA in complex with a single RuvA tetramer, *Proc. Natl. Acad. Sci.* **97**, 8257–8262 (2000).
- Biertümpfel, C., Yang, W., and Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction; and Hadden, J.M., Déclais, A.-C., Carr, S.B., Lilley, D.M., and Phillips, S.E.V., The structural basis of Holliday junction resolution by T7 endonuclease I, *Nature* **449**, 616–620 and 621–625 (2007).
- Changela, A., Perry, K., Taneja, B., and Mondragón, A., DNA manipulators: caught in the act, *Curr. Opin. Struct. Biol.* **13**, 15–22 (2003).
- Chen, Z., Yang, H., and Pavletich, N.P., Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures, *Nature* **453**, 489–494 (2008).
- Cox, M.M., Motoring along with the bacterial RecA protein, *Nature Rev. Mol. Cell Biol.* **8**, 127–138 (2007); and Regulation of bacterial RecA protein function, *Crit. Rev. Biochem. Mol. Biol.* **42**, 41–63 (2007).
- Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (Eds.), *Mobile DNA II*, ASM Press (2002). [A compendium of authoritative articles.]
- Grindley, N.G.F., Whiteson, K.L., and Rice, P.A., Mechanisms of site-specific recombination, *Annu. Rev. Biochem.* **75**, 567–605 (2006).
- Haber, J.E., Partners and pathways. Repairing a double-strand break, *Trends Genet.* **16**, 259–264 (2000).

- Ho, P.S. and Eichman, B.F., The crystal structures of Holliday junctions, *Curr. Opin. Struct. Biol.* **11**, 302–308 (2001).
- Kuzminov, A., Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ , *Microbiol. Mol. Biol. Rev.* **63**, 751–813 (1999).
- Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D.F., and Steitz, T.A., Structure of a synaptic $\gamma\delta$ resolvase tetramer covalently linked to two cleaved DNAs, *Science* **309**, 1210–1215 (2005); and Yang, W. and Steitz, T.A., Crystal structure of the site-specific recombinase $\gamma\delta$ resolvase complexed with a 34 bp cleavage site, *Cell* **82**, 193–207 (1995).
- Lusetti, S.L. and Cox, M.M., The bacterial RecA protein and the recombinational DNA repair of stalled replication forks, *Annu. Rev. Biochem.* **71**, 71–100 (2002).
- Marians, K.J., PriA-directed replication fork restart in *Escherichia coli*, *Trends Biochem. Sci.* **25**, 185–189 (2000).
- Reznikoff, W. S., Transposon Tn5, *Annu. Rev. Genet.* **42**, 269–286 (2008).
- Rice, P.A. and Baker, T.A., Comparative architecture of transposase and integrase complexes, *Nature Struct. Biol.* **8**, 302–307 (2001).
- Roe, S.M., Barlow, T., Brown, T., Oram, M., Keeley, A., Tsaneva, I.R., and Pearl, L.H., Crystal structure of an octameric RuvA–Holliday junction complex, *Mol. Cell* **2**, 361–372 (1998).
- Saikrishnan, K., Griffiths, S.P., Cook, N., Court, R., and Wigley, D.B., DNA binding to RecD: role of the 1B domain in SF1B helicase activity, *EMBO J.* **27**, 2222–2229 (2008).
- Simon, M., Zieg, J., Silverman, M., Mandel, G., and Doolittle, R., Phase variation: evolution of a controlling element, *Science* **209**, 1370–1374 (1980).
- Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B., Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks, *Nature* **432**, 187–193 (2004).
- Sung, P. and Klein, H., Mechanism of homologous recombination: mediators and helicases take on regulatory functions, *Nature Rev. Mol. Cell Biol.* **7**, 739–750 (2006).
- Van Duyne, G.D., A structural view of Cre–*loxP* site-specific recombination, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104 (2001).
- Yamada, K., Ariyoshi, M., and Morikawa, K., Three-dimensional structural views of branch migration and resolution in DNA homologous recombination, *Curr. Opin. Struct. Biol.* **14**, 130–137 (2004).
- Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K., Crystal structure of the Holliday junction migration motor protein RuvB from *Thermus thermophilus* HB8, *Proc. Natl. Acad. Sci.* **98**, 1442–1447 (2001).
- DNA Methylation and Trinucleotide Repeat Expansions**
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M., Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism; and Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S., Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, *Nature* **455**, 818–821 and 822–825 (2008).
- Bowater, R.P. and Wells, R.D., The intrinsically unstable life of DNA repeats associated with human hereditary disorders, *Prog. Nucleic Acid Res. Mol. Biol.* **66**, 159–202 (2001).
- Castel, A.L., Cleary, J.D., and Pearson, C.E., Repeat instability as the basis for human diseases and as a potential target for therapy, *Nature Rev. Mol. Cell Biol.* **11**, 165–170 (2010).
- Cummings, C.J. and Zoghbi, H.Y., Trinucleotide repeats: mechanisms and pathophysiology, *Annu. Rev. Genomics Hum. Genet.* **1**, 281–328 (2002).
- Goodman, J. and Watson, R.E., Altered DNA methylation: a secondary mechanism involved in carcinogenesis, *Annu. Rev. Pharmacol. Toxicol.* **42**, 501–525 (2002).
- Jones, P.A. and Baylin, S.B., The fundamental role of epigenetic events in cancer, *Nature Rev. Genet.* **3**, 415–428 (2002); and Jones, P.A. and Takai, D., The role of DNA methylation in mammalian epigenetics, *Science* **293**, 1068–1070 (2001).
- Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X., HhaI methyltransferase flips its target base out of the DNA helix, *Cell* **76**, 357–369 (1994).
- Klose, R.J. and Bird, A.P., Genomic DNA methylation: the mark and its mediators, *Trends Biochem. Sci.* **31**, 89–96 (2006).
- Mirkin, S.M., Expandable DNA repeats and human disease, *Nature* **447**, 932–940 (2007).
- O'Donnell, W.T. and Warren, S.T., A decade of molecular studies of fragile X syndrome, *Annu. Rev. Neurosci.* **25**, 315–338 (2002).
- Perutz, M.F. and Windle, A.H., Causes of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats, *Nature* **42**, 143–144 (2001); and Perutz, M.F., Glutamine repeats and neurodegenerative diseases: molecular aspects, *Trends Biochem. Sci.* **24**, 58–63 (1999).
- Reik, W., Dean, W., and Walter, J., Epigenetic reprogramming in mammalian development, *Science* **293**, 1089–1093 (2001).
- Rideout, W.M., III, Eggan, K., and Jaenisch, R., Nuclear cloning and epigenetic reprogramming of the genome, *Science* **293**, 1093–1098 (2001).
- Shubert, H.L., Blumenthal, R.M., and Cheng, X., Many paths to methyltransfer: a chronicle of convergence, *Trends Biochem. Sci.* **28**, 329–335 (2003).
- Szyf, M. and Detich, N., Regulation of the DNA methylation machinery and its role in cellular transformation, *Prog. Nucleic Acid Res. Mol. Biol.* **69**, 47–79 (2001).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Chaps. 64, 223, and 226 discuss fragile X syndrome, Huntington disease, and the spinocerebellar ataxias.]

PROBLEMS

- Explain how certain mutant varieties of Pol I can be nearly devoid of DNA polymerase activity but retain almost normal levels of $5' \rightarrow 3'$ exonuclease activity.
- Why haven't Pol I mutants been found that completely lack $5' \rightarrow 3'$ activity at all temperatures?
- Why aren't type I topoisomerases necessary in *E. coli* DNA replication?

4. The $3' \rightarrow 5'$ exonuclease activity of Pol I excises only unpaired 3'-terminal nucleotides from DNA, whereas this enzyme's pyrophosphorolysis activity removes only properly paired 3'-terminal nucleotides. Discuss the mechanistic significance of this phenomenon in terms of the polymerase reaction.

- You have isolated *E. coli* with temperature-sensitive mutations in the following genes. What are their phenotypes above their restrictive temperatures? Be specific. (a) *dnaB*, (b) *dnaE*, (c) *dnaG*, (d) *lig*, (e) *polA*, (f) *rep*, (g) *ssb*, and (h) *recA*.

6. About how many Okazaki fragments are synthesized in the replication of an *E. coli* chromosome?

***7.** What are the minimum and maximum number of replication forks that occur in a contiguous chromosome of an *E. coli* that is dividing every 25 min; every 80 min?

8. To put the *E. coli* replication system on a human scale, let us imagine that the 20-Å-diameter B-DNA was expanded to 1 m in diameter. If everything were proportionally expanded, then each DNA polymerase III holoenzyme would be about the size of a medium-sized truck. In such an expanded system: (a) How fast would each replisome be moving? (b) How far would each replisome travel during a complete replication cycle? (c) What would be the length of an Okazaki fragment? (d) What would be the average distance a replisome would travel between each error it made? Provide your answers in km/hr and km.

9. Why can't linear duplex DNAs, such as occur in bacteriophage T7, be fully replicated by only *E. coli*-encoded proteins?

***10.** What is the half-life of a particular purine base in the human genome assuming that it is subject only to spontaneous depurination? What fraction of the purine bases in a human genome will have depurinated in the course of a single generation (assume 25 years)? The DNAs of ~4000-year-old Egyptian mummies have been sequenced. Assuming that mummification did not slow the rate of DNA depurination, what fraction of the purine bases originally present in the mummy would still be intact today.

11. Why is the methylation of DNA to form O^6 -methylguanine mutagenic?

12. A replication fork encountering a single-strand lesion may either dissociate or leave a single-strand gap. The latter process is more likely to occur during lagging strand synthesis than during leading strand synthesis. Explain.

13. The *E. coli* genome contains 1009 Chi sequences. Do these sequences occur at random, and, if not, how much more or less frequently than random do they occur?

14. *Deinococcus radiodurans*, which the *Guinness Book of World Records* has dubbed the world's toughest bacterium, can tolerate doses of ionizing radiation ~3000-fold greater than those that are lethal to humans (it was first discovered growing in a can of ground meat that had been "sterilized" by radiation). It appears to have several strategies to repair radiation damage to its DNA (which large doses of ionizing radiation fragment to many pieces) including a particularly large number of genes encoding proteins involved in DNA repair and 4 to 10 copies per cell of its genome, which consists of two circular chromosomes and two circular plasmids. Yet, these strategies, alone, do not account for *D. radiodurans*' enormously high radiation resistance. However, in an addi-

tional strategy, it organizes its multiple identical dsDNA circles into stacks in which, it is thought, the identical genes in the neighboring circles are aligned side by side. How would this latter strategy help *D. radiodurans* efficiently repair its fragmented DNA?

15. CpG islands occur in eukaryotic genomes at about one-fifth their expected random frequency. Suggest an evolutionary (mutational) process that eliminates CpG islands.

16. Explain why the brief exposure of a cultured eukaryotic cell line to 5-azacytosine results in permanent phenotypic changes to these cells.

17. Explain why chi structures, such as that shown in Fig. 30-70b, have two pairs of equal length arms.

***18.** Single-stranded circular DNAs containing a transposon have a characteristic stem-and-double-loop structure such as that shown in Fig. 30-106. What is the physical basis of this structure?

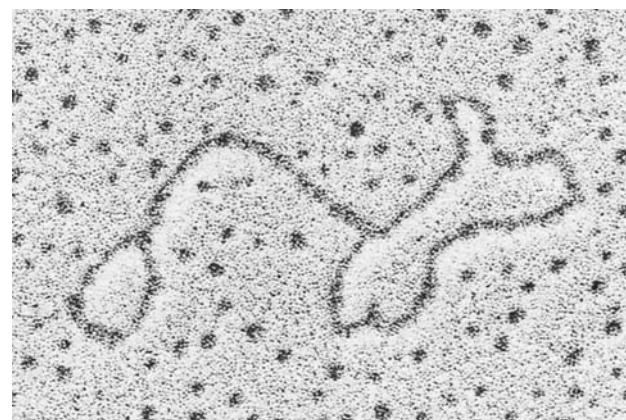
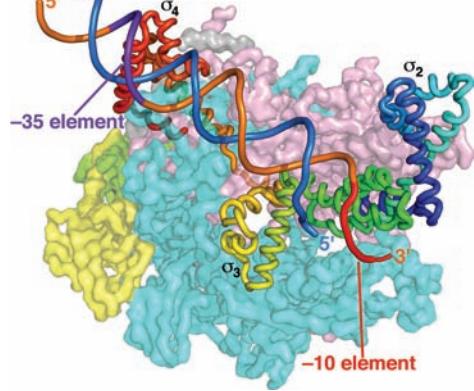


Figure 30-106 Electron micrograph of a single-stranded circular DNA containing a transposon. [Courtesy of Stanley N. Cohen, Stanford University School of Medicine.]

19. A composite transposon integrated in a circular plasmid occasionally transposes the DNA comprising the original plasmid rather than the transposon's central region. Explain how this is possible.

***20.** Cre recombinase has an additional function to that of circularizing the linear P1 dsDNA (Fig. 30-96). It is also required to resolve the circular dimers of P1 plasmids that result from their recombinational repair during replication, thereby permitting both daughter cells to receive a copy of the P1 plasmid. Using simple line diagrams, outline how these plasmids become dimerized and how Cre resolves them to circular monomers.



Transcription

CHAPTER 31

1 The Role of RNA in Protein Synthesis

- A. Enzyme Induction
- B. Messenger RNA

2 RNA Polymerase

- A. Template Binding
- B. Chain Initiation
- C. Chain Elongation
- D. Chain Termination
- E. Eukaryotic RNA Polymerases

3 Control of Transcription in Prokaryotes

- A. Promoters
- B. *lac* Repressor I: Binding
- C. Catabolite Repression: An Example of Gene Activation
- D. Sequence-Specific Protein–DNA Interactions
- E. *araBAD* Operon: Positive and Negative Control by the Same Protein
- F. *lac* Repressor II: Structure
- G. *trp* Operon: Attenuation
- H. Riboswitches Are Metabolite-Sensing RNAs
- I. Regulation of Ribosomal RNA Synthesis: The Stringent Response

4 Post-Transcriptional Processing

- A. Messenger RNA Processing
- B. Ribosomal RNA Processing
- C. Transfer RNA Processing

There are three major classes of RNA, all of which participate in protein synthesis: **ribosomal RNA (rRNA)**, **transfer RNA (tRNA)**, and **messenger RNA (mRNA)**. All of these RNAs are synthesized under the direction of DNA templates, a process known as **transcription**.

RNA's involvement in protein synthesis became evident in the late 1930s through investigations by Torbjörn Caspersson and Jean Brachet. Caspersson, using microscopic techniques, found that DNA is confined almost exclusively to the eukaryotic cell nucleus, whereas RNA occurs largely in the cytosol. Brachet, who had devised methods for fractionating cellular organelles, came to similar conclusions based on direct chemical analyses. He found, in addition, that the cytosolic RNA-containing particles are also protein rich. Both investigators noted that the concentration of these RNA–protein particles (which were later named ribosomes) is correlated with the rate at which a cell synthesizes protein, implying a relationship be-

tween RNA and protein synthesis. Indeed, Brachet even suggested that *the RNA–protein particles are the site of protein synthesis*.

Brachet's suggestion was shown to be valid when radioactively labeled amino acids became available in the 1950s. A short time after injection of a rat with a labeled amino acid, most of the label that had been incorporated in proteins was associated with ribosomes. This experiment also established that *protein synthesis is not immediately directed by DNA because, at least in eukaryotes, DNA and ribosomes are never in contact*.

In 1958, Francis Crick summarized the then dimly perceived relationships among DNA, RNA, and protein by what he called the **central dogma** of molecular biology: *DNA directs its own replication and its transcription to RNA which, in turn, directs its translation to proteins* (Fig. 5-21). The peculiar use of the word “dogma,” one definition of which is a religious doctrine that the true believer cannot doubt, stemmed from a misunderstanding. When Crick formulated the central dogma, he was under the impression that dogma meant “an idea for which there was no reasonable evidence.”

We begin this chapter by discussing experiments that led to the elucidation of mRNA's central role in protein synthesis. We then study the mechanism of transcription and its control in prokaryotes. Finally, in the last section, we consider post-transcriptional processing of RNA in both prokaryotes and eukaryotes. Translation is the subject of Chapter 32. Note that these subjects were outlined in Section 5-4. Here we shall delve into much greater detail.

1 THE ROLE OF RNA IN PROTEIN SYNTHESIS

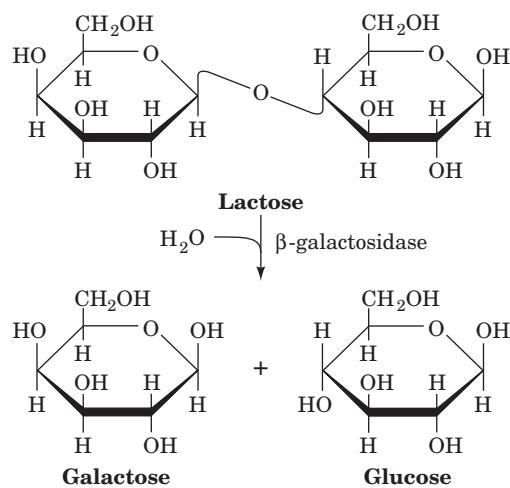
The idea that proteins are specified by mRNA and synthesized on ribosomes arose from the study of **enzyme induction**, a phenomenon in which bacteria vary the synthesis rates of specific enzymes in response to environmental changes. In this section, we discuss the classic experiments that explained the basis of enzyme induction and revealed the existence of mRNA. We shall see that *enzyme induction occurs as a consequence of the regulation of mRNA synthesis by proteins that specifically bind to the mRNA's DNA templates*.

A. Enzyme Induction

E. coli can synthesize an estimated ~4300 different polypeptides. There is, however, enormous variation in the amounts of these different polypeptides that are produced. For instance, the various ribosomal proteins may each be present in over 10,000 copies per cell, whereas certain regulatory proteins (see below) normally occur in <10 copies per cell. Many enzymes, particularly those involved in basic cellular “housekeeping” functions, are synthesized at a more or less constant rate; they are called **constitutive enzymes**. Other enzymes, termed **adaptive** or **inducible enzymes**, are synthesized at rates that vary with the cell’s circumstances.

a. Lactose-Metabolizing Enzymes Are Inducible

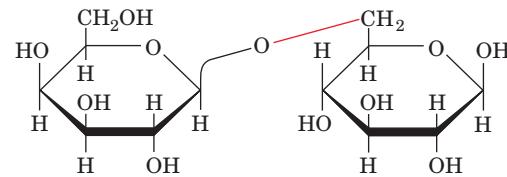
Bacteria, as has been recognized since 1900, adapt to their environments by producing enzymes that metabolize certain nutrients, for example, lactose, only when those substances are available. *E. coli* grown in the absence of lactose are initially unable to metabolize this disaccharide. To do so they require the presence of two proteins: **β -galactosidase**, which catalyzes the hydrolysis of lactose to its component monosaccharides,



and **galactoside permease** (also known as **lactose permease**; Section 20-4B), which transports lactose into the cell. *E. coli* grown in the absence of lactose contain only a few (<5) molecules of these proteins. Yet, a few minutes after lactose is introduced into their medium, *E. coli* increase the rate at which they synthesize these proteins by ~1000-fold (such that β -galactosidase can account for up to 10% of their soluble protein) and maintain this pace until lactose is no longer available. The synthesis rate then returns to its minuscule **basal level** (Fig. 31-1). *This ability to produce a series of proteins only when the substances they metabolize are present permits bacteria to adapt to their environment without the debilitating need to continuously synthesize large quantities of otherwise unnecessary substances.*

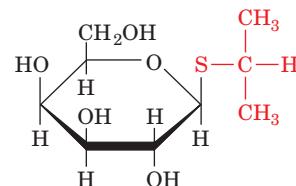
Lactose or one of its metabolic products must somehow trigger the synthesis of the above proteins. Such a sub-

stance is known as an **inducer**. The physiological inducer of the lactose system, the lactose isomer **1,6-allolactose**,



1,6-Allolactose

arises from lactose’s occasional transglycosylation by β -galactosidase. Most experimental studies of the lactose system use **isopropylthiogalactoside (IPTG)**,



Isopropylthiogalactoside (IPTG)

a potent inducer that structurally resembles allolactose but that is not degraded by β -galactosidase.

Lactose system inducers also stimulate the synthesis of **thiogalactoside transacetylase**, an enzyme that, *in vitro*, transfers an acetyl group from acetyl-CoA to the C6-OH group of a β -thiogalactoside such as IPTG. Since lactose fermentation proceeds normally in the absence of thiogalactoside transacetylase, however, this enzyme’s physiological role is unknown.

b. lac System Genes Form an Operon

The genes specifying wild-type β -galactosidase, galactoside permease, and thiogalactoside transacetylase are designated Z^+ , Y^+ , and A^+ , respectively. Genetic mapping of the defective mutants Z^- , Y^- , and A^- indicated that these

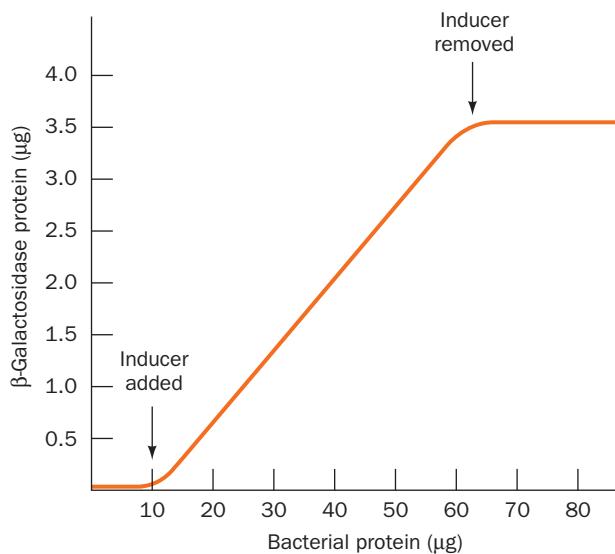


Figure 31-1 The induction kinetics of β -galactosidase in *E. coli*. [After Cohn, M., *Bacteriol. Rev.* **21**, 156 (1957).]

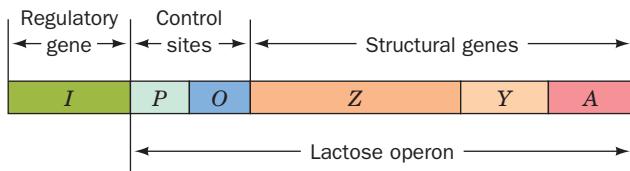


Figure 31-2 Genetic map of the *E. coli* lac operon. The map shows the genes encoding the proteins mediating lactose metabolism and the genetic sites that control their expression. The *Z*, *Y*, and *A* genes, respectively, encode β -galactosidase, galactoside permease, and thiogalactoside transacetylase.

lac structural genes (genes that specify polypeptides) are contiguously arranged on the *E. coli* chromosome (Fig. 31-2). These genes, together with the control elements *P* and *O*, form a genetic unit called an **operon**, specifically the **lac operon**. The nature of the control elements is discussed below. The role of operons in prokaryotic gene expression is examined in Section 31-3.

c. Bacteria Can Transmit Genes via Conjugation

An important clue as to how *E. coli* synthesizes protein was provided by a mutation that causes the proteins of the *lac* operon to be synthesized in large amounts in the absence of inducer. This so-called **constitutive mutation** occurs in a gene, designated *I*, that is distinct from although closely linked to the genes specifying the *lac* enzymes (Fig. 31-2). What is the nature of the *I* gene product? This riddle was solved in 1959 by Arthur Pardee, François Jacob, and Jacques Monod through an ingenious experiment that is known as the **PaJaMo experiment**. To understand this experiment, however, we must first consider **bacterial conjugation**.

Bacterial conjugation is a process, discovered in 1946 by Joshua Lederberg and Edward Tatum, through which some bacteria can transfer genetic information to others. The ability to conjugate (“mate”) is conferred on an otherwise indifferent bacterium by a plasmid named **F factor** (for fertility). Bacteria that possess an F factor (designated F⁺ or male) are covered by hairlike projections known as **F pili**. These bind to cell-surface receptors on bacteria that lack the F factor (F⁻ or female), which leads to the formation of



Figure 31-3 Bacterial conjugation. An electron micrograph in false color showing an F⁺ (left) and an F⁻ (right) *E. coli* engaged in sexual conjugation. [Dennis Kunkel/Phototake.]

a cytoplasmic bridge between these cells (Fig. 31-3). The F factor then replicates and, as the newly replicated single strand is formed, it passes through the cytoplasmic bridge to the F⁻ cell where the complementary strand is synthesized (Fig. 31-4). This converts the F⁻ cell to F⁺ so that the F factor is an infectious agent (a bacterial venereal disease?).

On very rare occasions, the F factor spontaneously integrates into the chromosome of the F⁺ cell. In the resulting

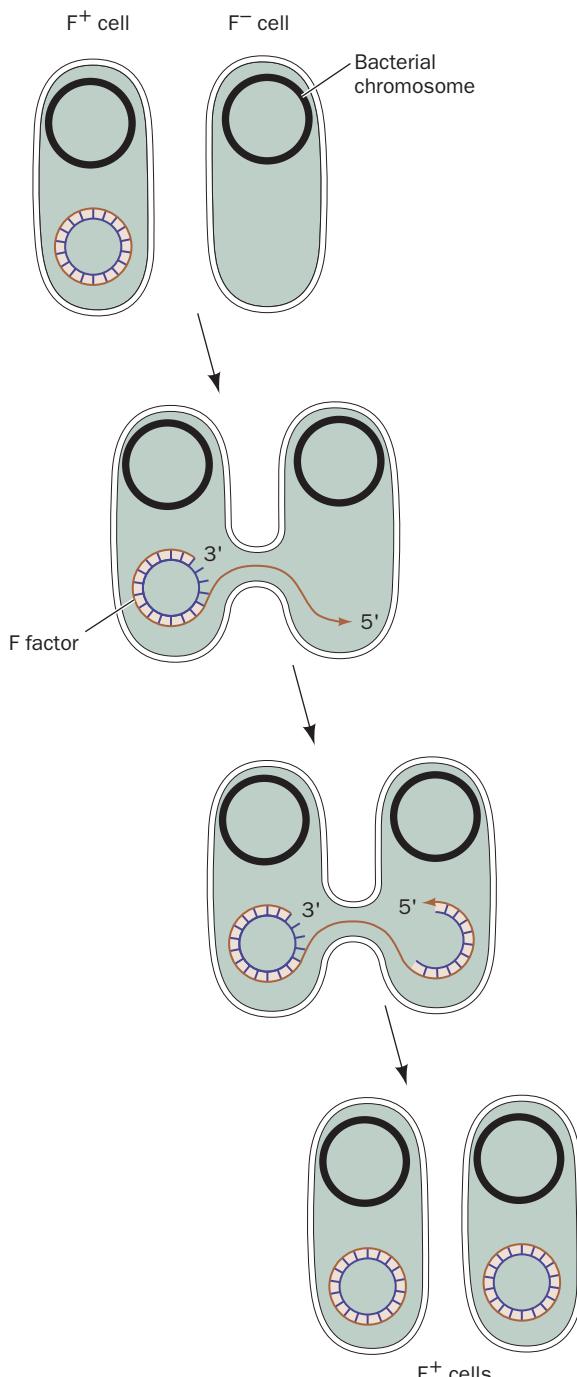


Figure 31-4 Diagram showing how an F⁻ cell acquires an F factor from an F⁺ cell. A single strand of the F factor is replicated, via the rolling circle mode (Section 30-3Bb), and is transferred to the F⁻ cell where its complementary strand is synthesized to form a new F factor.

Hfr (for high frequency of recombination) cells, the F factor behaves much as it does in the autonomous state. Its replication commences at a specific internal point in the F factor, and the replicated section passes through a cytoplasmic bridge to the F⁻ cell, where its complementary strand is synthesized. In this case, however, the replicated chromosome of the Hfr cell is also transmitted to the F⁻ cell (Fig. 31-5). *Bacterial genes are transferred from the Hfr cell to the F⁻ cell in fixed order*. This is because the F factor in a given Hfr strain is integrated into the bacterial chromosome at a specific site and because only a particular strand of the Hfr chromosomal DNA is replicated and transferred

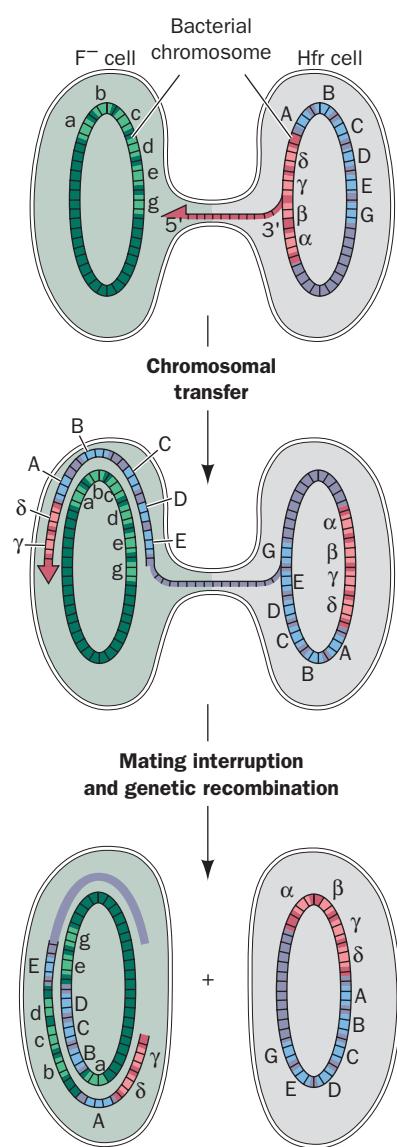


Figure 31-5 Transfer of the bacterial chromosome from an Hfr cell to an F⁻ cell and its subsequent recombination with the F⁻ chromosome. Here, Greek letters represent F factor genes, uppercase Roman letters represent bacterial genes from the Hfr cell, and lowercase Roman letters represent the corresponding alleles in the F⁻ cell. Since chromosomal transfer, which begins within the F factor, is rarely complete, the entire F factor is seldom transferred. Hence the recipient cell usually remains F⁻.

to the F⁻ cell. Usually, only part of the Hfr bacterial chromosome is transferred during sexual conjugation because the cytoplasmic bridge almost always breaks off sometime during the ~90 min required to complete the transfer process. In the resulting **merozygote** (a partially diploid bacterium), the chromosomal fragment, which lacks a complete F factor, neither transforms the F⁻ cell to Hfr nor is subsequently replicated. However, the transferred chromosomal fragment recombines with the chromosome of the F⁻ cell (Section 30-6A), thereby permanently endowing the F⁻ cell with some of the traits of the Hfr strain.

The integrated F factor in an Hfr cell occasionally undergoes spontaneous excision to yield an F⁺ cell. In rare instances, the F factor is aberrantly excised such that a portion of the adjacent bacterial chromosome is incorporated in the subsequently autonomously replicating F factor. Bacteria carrying such a so-called **F' factor** are permanently diploid for its bacterial genes.

d. lac Repressor Inhibits the Synthesis of lac Operon Proteins

In the PaJaMo experiment, Hfr bacteria of genotype I^+Z^+ were mated to an F⁻ strain of genotype I^-Z^- in the absence of inducer while the β -galactosidase activity of the culture was monitored (Fig. 31-6). At first, as expected,

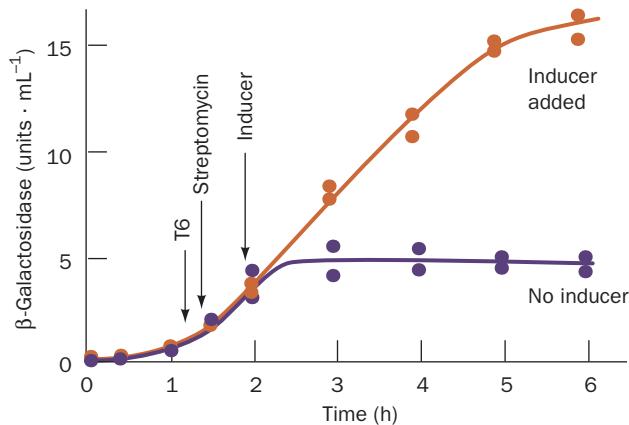


Figure 31-6 The PaJaMo experiment. This experiment demonstrated the existence of the *lac* repressor through the appearance of β -galactosidase in the transient merozygotes (partial diploids) formed by mating I^+Z^+ Hfr donors with I^-Z^- F⁻ recipients. The F⁻ strain was also resistant to both **bacteriophage T6** and **streptomycin**, whereas the Hfr strain was sensitive to these agents. Both types of cells were grown and mated in the absence of inducer. After sufficient time had passed for the transfer of the *lac* genes, the Hfr cells were selectively killed by the addition of T6 phage and streptomycin. In the absence of inducer (lower curve), β -galactosidase synthesis commenced at around the time at which the *lac* genes had entered the F⁻ cells but stopped after ~1 h. If inducer was added shortly after the Hfr donors had been killed (upper curve), enzyme synthesis continued unabated. This demonstrates that the cessation of β -galactosidase synthesis in uninduced cells is not due to the intrinsic loss of the ability to synthesize this enzyme but to the production of a repressor specified by the I^+ gene. [After Pardee, A.B., Jacob, F., and Monod, J., *J. Mol. Biol.* **1**, 173 (1959).]

there was no β -galactosidase activity because the Hfr donors lacked inducer and the F⁻ recipients were unable to produce active enzyme (only DNA passes through the cytoplasmic bridge connecting mating bacteria). About 1 h after conjugation began, however, when the I^+Z^+ genes had just entered the F⁻ cells, β -galactosidase synthesis began and only ceased after about another hour. The explanation for these observations is that the donated Z^+ gene, on entering the cytoplasm of the I^- cell, directs the synthesis of β -galactosidase in a constitutive manner. Only after the donated I^+ gene has had sufficient time to be expressed is it able to repress β -galactosidase synthesis. *The I^+ gene must therefore give rise to a diffusible product, the lac repressor, which inhibits the synthesis of β -galactosidase (and the other lac proteins).* Inducers such as IPTG temporarily inactivate lac repressor, whereas I^- cells constitutively synthesize lac enzymes because they lack a functional repressor. Lac repressor, as we shall see in Section 31-3B, is a protein.

B. Messenger RNA

The nature of the lac repressor's target molecule was deduced in 1961 through a penetrating genetic analysis by Jacob and Monod. A second type of constitutive mutation in the lactose system, designated O^c (for **operator constitutive**), which complementation analysis (Section 1-4Cc) has shown to be independent of the I gene, maps between the I and Z genes (Fig. 31-2). In the partially diploid F' strain $O^cZ^-/F\ O^+Z^+$, β -galactosidase activity is inducible by IPTG, whereas the strain $O^cZ^+/F\ O^+Z^-$ constitutively synthesizes this enzyme. *An O^+ gene can therefore only control the expression of a Z gene on the same chromosome.* The same is true with the Y^+ and A^+ genes.

Jacob and Monod's observations led them to conclude that the proteins are synthesized in a two-stage process:

1. The structural genes on DNA are transcribed onto complementary strands of **messenger RNA (mRNA)**.
2. The mRNAs transiently associate with ribosomes, which they direct in polypeptide synthesis.

This hypothesis explains the behavior of the lac system that we previously outlined in Section 5-4Ab (Fig. 5-25;  **See Guided Exploration 2: Regulation of gene expression by the lac repressor system.**) *In the absence of inducer, the lac repressor specifically binds to the O gene (the operator) so as to prevent the enzymatic transcription of mRNA. On binding inducer, the repressor dissociates from the operator, thereby permitting the transcription and subsequent translation of the lac enzymes.* The operator-repressor-inducer system thereby acts as a molecular switch so that the lac operator can only control the expression of lac enzymes on the same chromosome. The O^c mutants constitutively synthesize lac enzymes because they are unable to bind repressor. The **coordinate** (simultaneous) expression of all three lac enzymes under the control of a single operator site arises, as Jacob and Monod theorized, from the

transcription of the lac operon as a single **polycistronic mRNA** which directs the ribosomal synthesis of each of these proteins (the term **cistron** is a somewhat archaic synonym for gene). This transcriptional control mechanism is further discussed in Section 31-3. [DNA sequences that are on the same DNA molecule are said to be "in cis" (Latin: on this side of), whereas those on different DNA molecules are said to be "in trans" (Latin: across). Control sequences such as the O gene, which are only active on the same DNA molecule as the genes they control, are called **cis-acting elements**. Genes such as $lacI$, which specify the synthesis of diffusible products and can therefore be located on a different DNA molecule from the genes they control, are said to direct the synthesis of **trans-acting factors**.]

a. mRNAs Have Their Predicted Properties

The kinetics of enzyme induction, as indicated, for example, in Figs. 31-1 and 31-6, requires that the postulated mRNA be both rapidly synthesized and rapidly degraded. An RNA with such quick turnover had, in fact, been observed in T2-infected *E. coli*. Moreover, the base composition of this RNA fraction resembles that of the viral DNA rather than that of the bacterial RNA (keep in mind that base sequencing techniques would not be formulated for another ~ 15 years). Ribosomal RNA, which comprises up to 90% of a cell's RNA, turns over much more slowly than mRNA. Ribosomes are therefore not permanently committed to the synthesis of a particular protein (a once popular hypothesis). Rather, *ribosomes are nonspecific protein synthesizers that produce the polypeptide specified by the mRNA with which they are transiently associated.* A bacterium can therefore respond within a few minutes to changes in its environment.

Evidence favoring the Jacob and Monod model rapidly accumulated. Sydney Brenner, Jacob, and Matthew Meselson carried out experiments designed to characterize the RNA that *E. coli* synthesized after T4 phage infection. *E. coli* were grown in a medium containing ^{15}N and ^{13}C so as to label all cell constituents with these heavy isotopes. The cells were then infected with T4 phages and immediately transferred to an unlabeled medium (which contained only the light isotopes ^{14}N and ^{12}C) so that cell components synthesized before and after phage infection could be separated by equilibrium density gradient ultracentrifugation in CsCl solution (Section 6-5Bb). No "light" ribosomes were observed, which indicates, in agreement with the above-mentioned T2 phage results, that no new ribosomes are synthesized after phage infection.

The growth medium also contained either ^{32}P or ^{35}S so as to radioactively label the newly synthesized and presumably phage-specific RNA and protein, respectively. Much of the ^{32}P -labeled RNA was associated, as was postulated for mRNA, with the preexisting "heavy" ribosomes (Fig. 31-7). Likewise, the ^{35}S -labeled proteins were transiently associated with, and therefore synthesized by, these ribosomes.

Sol Spiegelman developed the RNA-DNA hybridization technique (Section 5-3Cb) in 1961 to characterize the

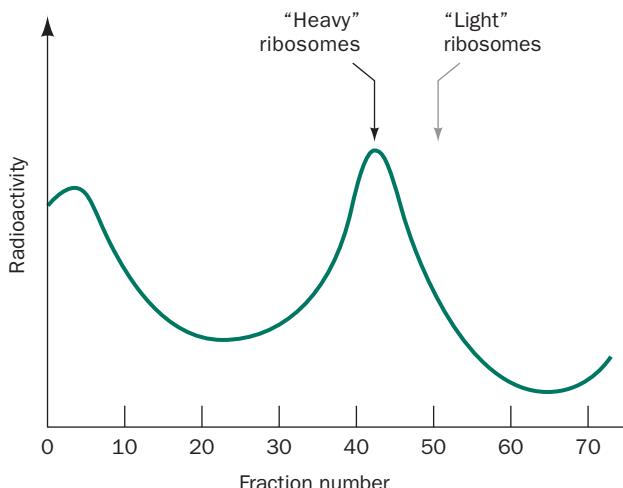


Figure 31-7 The distribution, in a CsCl density gradient, of ^{32}P -labeled RNA that had been synthesized by *E. coli* after T4 phage infection. Free RNA, being relatively dense, bands at the bottom of the centrifugation cell (left). Much of the RNA, however, is associated with the ^{15}N - and ^{13}C -labeled “heavy” ribosomes that had been synthesized before the phage infection. The predicted position of unlabeled “light” ribosomes, which are not synthesized by phage-infected cells, is also indicated. [After Brenner, S., Jacob, F., and Meselson, M., *Nature* **190**, 579 (1961).]

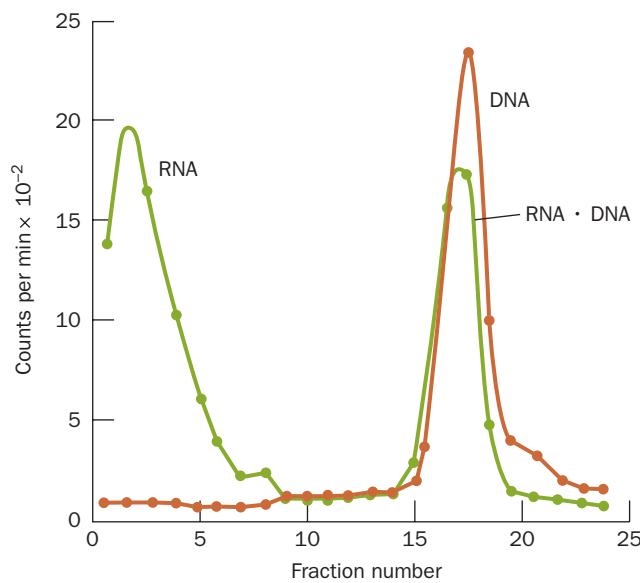


Figure 31-8 The hybridization of ^{32}P -labeled RNA produced by T2-infected *E. coli* with ^{3}H -labeled T2 DNA. On radioactive decay, ^{32}P and ^{3}H emit β particles (electrons) with characteristically different energies so that these isotopes can be independently detected. Although free RNA (left) in a CsCl density gradient is denser than DNA, much of the RNA bands with the DNA (right). This indicates that the two polynucleotides have hybridized and are therefore complementary in sequence. [After Hall, B.D. and Spiegelman, S., *Proc. Natl. Acad. Sci.* **47**, 141 (1961).]

RNA synthesized by T2-infected *E. coli*. He found that this phage-derived RNA hybridizes with T2 DNA (Fig. 31-8) but does not hybridize with DNAs from unrelated phage nor with the DNA from uninfected *E. coli*. This RNA must therefore be complementary to T2 DNA in agreement with Jacob and Monod’s prediction; that is, the phage-specific RNA is a messenger RNA. Hybridization studies have likewise shown that mRNAs from uninfected *E. coli* are complementary to portions of *E. coli* DNA. In fact, other RNAs, such as transfer RNA and ribosomal RNA, have corresponding complementary sequences on DNA from the same organism. Thus, *all cellular RNAs are transcribed from DNA templates*.

2 RNA POLYMERASE

RNA polymerase (RNAP), the enzyme responsible for the DNA-directed synthesis of RNA, was discovered independently in 1960 by Samuel Weiss and Jerard Hurwitz. The enzyme couples together the ribonucleoside triphosphates ATP, CTP, GTP, and UTP on DNA templates in a reaction that is driven by the release and subsequent hydrolysis of PP_i :



All cells contain RNAP. In bacteria, one species of this enzyme synthesizes all of the cell’s RNA except the RNA primers employed in DNA replication (Section 30-1D). Various bacteriophages encode RNAPs that synthesize only phage-specific RNAs. Eukaryotic cells contain four or five RNAPs that each synthesize a different class of RNA. In this section we first consider the properties of the bacterial RNAPs and then consider the eukaryotic enzymes.

E. coli RNAP’s so-called **holoenzyme** is an ~ 459 -kD protein with subunit composition $\alpha_2\beta\beta'\omega\sigma$ (Table 31-1) in which the β and β' subunits contain several colinearly arranged homologous segments. Once RNA synthesis has been initiated, however, the σ subunit (also called **σ factor** or σ^{70} since its molecular mass is 70 kD) dissociates from the **core enzyme**, $\alpha_2\beta\beta'\omega$, which carries out the actual polymerization process (see below).

Table 31-1 Components of *E. coli* RNA Polymerase Holoenzyme

Subunit	Number of Residues	Structural Gene
α	329	<i>rpoA</i>
β	1342	<i>rpoB</i>
β'	1407	<i>rpoC</i>
ω	91	<i>rpoZ</i>
σ^{70}	613	<i>rpsD</i>

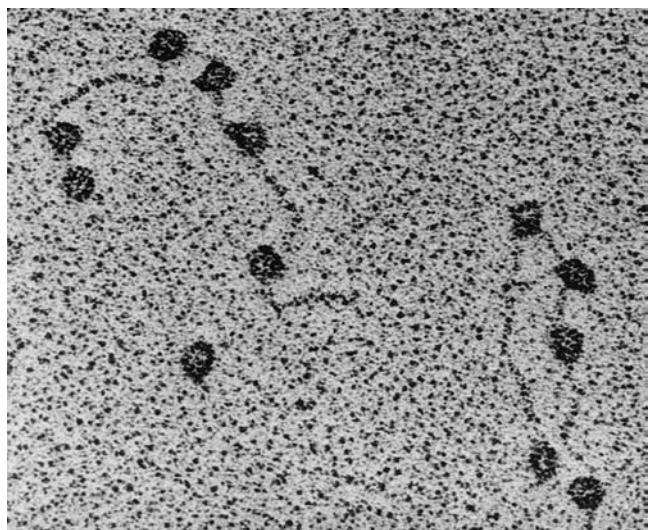


Figure 31-9 An electron micrograph of *E. coli* RNA polymerase (RNAP) holoenzyme attached to various promoter sites on bacteriophage T7 DNA. RNAP is one of the largest known soluble enzymes. [From Williams, R.C., *Proc. Natl. Acad. Sci.* **74**, 2313 (1977).]

Electron micrographs (Fig. 31-9) clearly indicate that RNAP, which has a characteristic large size, binds to DNA as a protomer. This large size is presumably a consequence of the holoenzyme's several complex functions including (1) template binding, (2) RNA chain initiation, (3) chain elongation, and (4) chain termination. We discuss these various functions below.

A. Template Binding

RNA synthesis is normally initiated only at specific sites on the DNA template. This was first demonstrated through hybridization studies of bacteriophage ϕ X174 DNA with the RNA produced by ϕ X174-infected *E. coli*. Bacteriophage ϕ X174 carries a single strand of DNA known as the (+) strand. On its injection into *E. coli*, the (+) strand directs the synthesis of the complementary (−) strand with which it combines to form a circular duplex DNA known as the replicative form (Section 30-3Ba). The RNA produced by ϕ X174-infected *E. coli* does not hybridize with DNA from intact phages but does so with the replicative form. Thus only the (−) strand of ϕ X174 DNA, the so-called **antisense strand**, is transcribed, that is, acts as a template; the (+) strand, the **sense strand** (or **coding strand**; so called because it has the same sequence as the transcribed RNA), does not do so. Similar studies indicate that in larger phages, such as T4 and λ , the two viral DNA strands are the antisense (template) strands for different sets of genes. The same is true of cellular organisms.

a. Holoenzyme Specifically Binds to Promoters

RNA polymerase binds to its initiation sites through base sequences known as **promoters** that are recognized

by the corresponding σ factor. The existence of promoters was originally suggested by mutations that enhance or diminish the transcription rates of certain genes, including those of the *lac* operon. *Genetic mapping of such mutations indicated that the promoter consists of an ~40-bp sequence that is located on the 5' side of the transcription start site.* [By convention, the sequence of template DNA is represented by its sense (nontemplate) strand so that it will have the same directionality as the transcribed RNA. A base pair in a promoter region is assigned a negative or positive number that indicates its position, upstream or downstream in the direction of RNAP travel, from the first nucleotide that is transcribed to RNA; this start site is +1 and there is no 0.] RNA, as we shall see, is synthesized in the 5' \rightarrow 3' direction (Section 31-2C). Consequently, the promoter lies on the “upstream” side of the RNA's starting nucleotide. Sequencing studies indicate that the *lac* promoter (*lacP*) overlaps the *lac* operator (Fig. 31-2).

The holoenzyme forms tight complexes with promoters (dissociation constant $K \approx 10^{-14} M$) and thereby protects the bound DNA segments from digestion by DNase I. The region from about −20 to +20 is protected against exhaustive DNase I degradation. The region extending upstream to about −60 is also protected but to a lesser extent, presumably because it binds holoenzyme less tightly.

Sequence determinations of the protected regions from numerous *E. coli* and phage genes have revealed the consensus sequence of *E. coli* promoters (Fig. 31-10). *Their most conserved sequence is a hexamer centered at about the −10 position, the so-called Pribnow box* (named after David Pribnow, who pointed out its existence in 1975). It has a consensus sequence of TATAAT in which the leading TA and final T are highly conserved. *Upstream sequences around position −35 also have a region of sequence similarity, TTGACA, which is most evident in efficient promoters.* The sequence of the segment between the −10 and the −35 sites is unimportant but its length is critical; it ranges from 16 to 19 bp in the great majority of promoters. The initiating (+1) nucleotide, which is nearly always A or G, is centered in a poorly conserved CAT or CGT sequence. Most promoter sequences vary considerably from the consensus sequence (Fig. 31-10). Nevertheless, a mutation in one of the partially conserved regions can greatly increase or decrease a promoter's initiation efficiency. In addition, Richard Gourse discovered that certain highly expressed genes contain an A + T-rich segment between positions −40 and −60, the **upstream promoter (UP) element**, which binds to the C-terminal domain of RNAP's α subunits. The UP element-containing genes include those encoding the ribosomal RNAs, the *rrn* genes (e.g., Fig. 31-10), which collectively account for 60% of the RNA synthesized by *E. coli*. *The rates at which genes are transcribed, which span a range of at least 1000, vary directly with the rate at which their promoters form stable initiation complexes with the holoenzyme.* Promoter mutations that increase or decrease the rate at which the associated gene is transcribed are known as **up mutations** and **down mutations**.

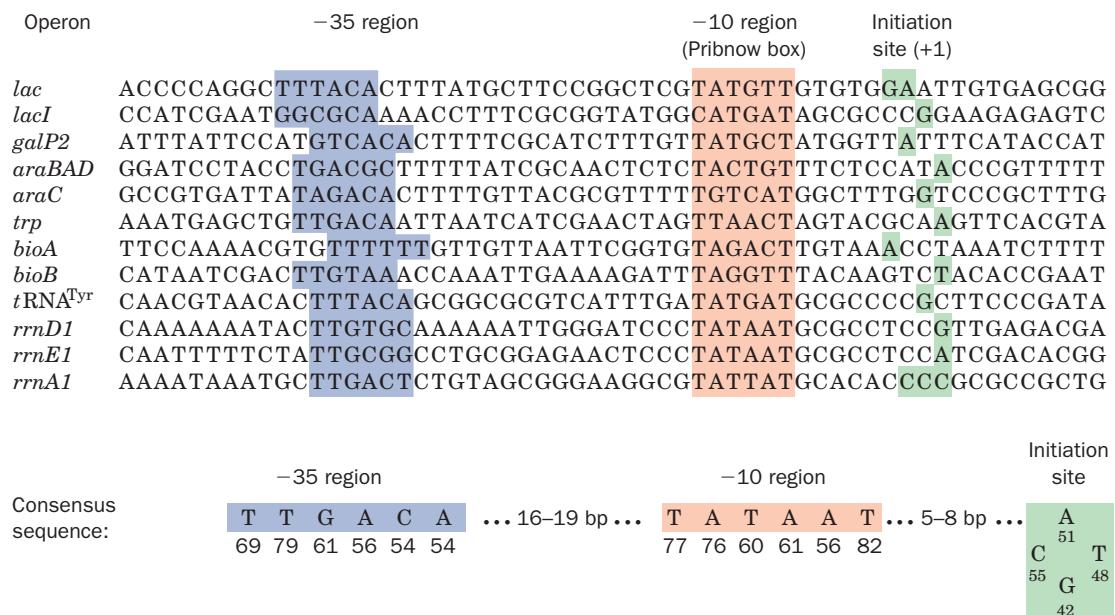


Figure 31-10 The sense (nontemplate) strand sequences of selected *E. coli* promoters. A 6-bp region centered around the –10 position (red shading) and a 6-bp sequence around the –35 region (blue shading) are both conserved. The transcription initiation sites (+1), which in most promoters occur at a single purine nucleotide, are shaded in green. The bottom row shows

the consensus sequence of 298 *E. coli* promoters with the number below each base indicating its percentage occurrence. The downstream portions of the *rrn* genes' UP elements can be seen. [After Rosenberg, M. and Court, D., *Annu. Rev. Genet.* **13**, 321–323 (1979). Consensus sequence from Lisser, S. and Margalit, H., *Nucleic Acids Res.* **21**, 1512 (1993).]

b. Initiation Requires the Formation of an Open Complex

The promoter regions in contact with the holoenzyme were identified by determining where the enzyme alters the susceptibility of the DNA to alkylation by agents such as dimethyl sulfate (DMS), a procedure named **DMS footprinting** (Section 34-3Bh). These experiments demonstrated that the holoenzyme contacts the promoter mainly around its –10 and –35 regions. These protected sites are both on the same side of the B-DNA double helix as the initiation site, which suggests that holoenzyme binds to only one face of the promoter.

DMS methylates G residues at N7, A residues at N1 and N3, and C residues at N3. Since N1 on A and N3 on C participate in base pairing interactions, however, they can only react with DMS in single-stranded DNA. This differential methylation of single- and double-stranded DNAs provides a sensitive test for DNA strand separation or “melting.” Such chemical footprinting studies indicate that the binding of holoenzyme “melts out” the promoter in a region of ~14 bp extending from the middle of the –10 region to just past the initiation site, thereby forming a so-called **transcription bubble**. The need to form this **open complex** explains why promoter efficiency tends to decrease with the number of G · C base pairs in the –10 region; this presumably increases the difficulty in opening the double helix as is required for chain initiation (recall that G · C pairs are more stable than A · T pairs).

Core enzyme, which does not specifically bind promoter (except when it has an UP element), tightly binds duplex

DNA (the complex's dissociation constant is $K \approx 5 \times 10^{-12} M$ and its half-life is ~60 min). Holoenzyme, in contrast, binds to nonpromoter DNA comparatively loosely ($K \approx 10^{-7} M$ and a half-life >1 s). Evidently, the σ subunit allows holoenzyme to move rapidly along a DNA strand in search of the σ subunit's corresponding promoter. Once transcription has been initiated and the σ subunit jettisoned, the tight binding of core enzyme to DNA apparently stabilizes the ternary enzyme–DNA–RNA complex.

B. Chain Initiation

The 5'-terminal base of prokaryotic RNAs is almost always a purine with A occurring more often than G. The initiating reaction of transcription is simply the coupling of two nucleoside triphosphates in the reaction



and hence, unlike DNA replication, does not require a primer. Bacterial RNAs therefore have 5'-triphosphate groups as was demonstrated by the incorporation of radioactive label into RNA when it was synthesized with [γ -³²P]ATP. Only the 5' terminus of the RNA can retain the label because the internal phosphodiester groups of RNA are derived from the α -phosphate groups of nucleoside triphosphates.

RNAP has a curious behavior: It frequently releases its newly synthesized RNA after only ~10 nt have been polymerized, a process known as **abortive initiation**. When RNAP initiates transcription, it keeps its grip on the pro-

motor (which is on the DNA's nontemplate/sense strand). Consequently, conformational tension builds up as the template/antisense strand is pulled through the RNAP's active site, a process called **scrunching** because the resulting increased size of the transcription bubble in the downstream direction must somehow be accommodated within the RNAP. In abortive initiation, the RNAP fails to escape the promoter and instead relieves the conformational tension by releasing the newly synthesized RNA fragment, thereby letting the transcription bubble relax to its normal size. The RNAP then reinitiates transcription from the +1 position. In successful initiation, the strain eventually provides sufficient energy to strip the promoter from the RNAP, which then commences the processive (continuous) transcription of the template. This process requires the dissociation of the σ factor from the core–DNA–RNA complex to form the elongation complex, although recent experiments indicate that this process often occurs stochastically (randomly) over several nucleotide additions. The σ factor can then join with another core to form a new initiation complex as was demonstrated by a burst of RNA synthesis on addition of core enzyme to a transcribing reaction mixture that initially contained only holoenzyme.

a. Bacterial RNAP Has a Highly Complex Structure

The X-ray structure of *E. coli* RNAP has not been determined. However, Seth Darst and Dmitry Vassylyev independently determined the X-ray structures of the closely similar *Thermus aquaticus* (Taq) and *Thermus thermophilus* (Tth) RNAP core enzymes and holoenzymes. The structure of the Tth core enzyme in complex with DNA and RNA, in agreement with EM studies of *E. coli* RNAP, has the overall shape of a crab claw whose two “pincers” are formed by the β and β' subunits (Fig. 31-11). The protein is ~ 150 Å long (parallel to the pincers), ~ 115 Å high, and ~ 110 Å deep, with the tunnel between the two pincers ~ 27 Å wide. The β and β' subunits extensively interact with one another, particularly at the base of the tunnel (also called the main channel) where an active site Mg^{2+} ion is located, which is also where their homologous segments converge. The β' subunit binds two Zn^{2+} ions, each via four Cys residues that are invariant in prokaryotes but not in eukaryotes. The outer surface of the RNAP is almost uniformly negatively charged, whereas those surfaces that interact with nucleic acids are positively charged.

The downstream dsDNA occupies the main channel, which directs the template strand to the active site. There it base-pairs with the incoming NTP (not present in this structure) at the so-called $i + 1$ site near the Mg^{2+} ion. The 3' end of the RNA forms a 9-bp hybrid helix with the 5' end of the template DNA strand and then exits the protein through a channel between the β and β' subunits (the RNA exit channel) in which it adopts a conformation similar to that of a single strand within an RNA double helix. Thus the structure resembles that of a post-translocated elongation complex, although the paths taken by the template and nontemplate DNA strands to rejoin at the end of the transcription bubble are unclear.

The X-ray structure of the Tth holoenzyme indicates that its σ subunit (σ^{70}) has three flexibly linked, largely α

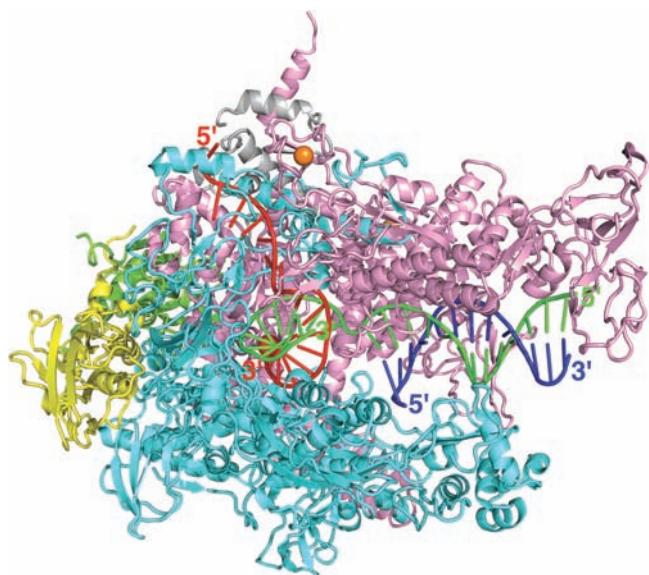


Figure 31-11 X-ray structure of Tth core RNAP in complex with a 23-nt template DNA, a 14-nt nontemplate DNA, and a 16-nt RNA. The protein is drawn in ribbon form with its two α subunits yellow and green, its β subunit cyan, its β' subunit pink, and its ω subunit gray. The bound Mg^{2+} and Zn^{2+} ions are represented by red and orange spheres, respectively. The DNA and RNA are shown in ladder form with template DNA green, nontemplate DNA blue, and RNA red. Note that residues 208 to 390 of the β' subunit, which extend from the tip of its pincer, are disordered and hence not visible, as are the 86-residue C-terminal domains of both α subunits. [Based on an X-ray structure by Dmitry Vassylyev, University of Alabama at Birmingham. PDBid 5O5I.]

helical domains, σ_2 , σ_3 , and σ_4 , that extend across the top of the holoenzyme (Fig. 31-12; σ_1 is not visible). The holoenzyme's pincers are ~ 10 Å farther apart than in the elongation complex. The binding cavities for the downstream dsDNA and the transcription bubble are partially occupied by the σ subunit's σ_4 domain and its σ_{3-4} linker. This partially accounts for the above-described nucleic acid scrunching that precedes the transition from the initiation complex to the elongation complex. Thus, the release of only these segments of the σ subunit from RNAP is compatible with the formation of an elongation complex, thereby accounting for the above-mentioned stochastic release of the σ subunit from a successfully initiated RNAP complex.

A low (6.5 Å) resolution X-ray structure of Taq holoenzyme in complex with a dsDNA segment containing the promoter's -10 and -35 elements reveals that the DNA lies across one face of the holoenzyme, completely outside of the main channel (Fig. 31-13). All sequence-specific contacts that the holoenzyme makes with the -10 and -35 elements as well as with the so-called extended -10 region just upstream of the -10 element are mediated by the σ subunit via conserved residues. This structure presumably resembles the so-called **closed complex** in which the DNA has not yet entered the main channel to form a transcription bubble. The mechanism through which this occurs is largely unknown.

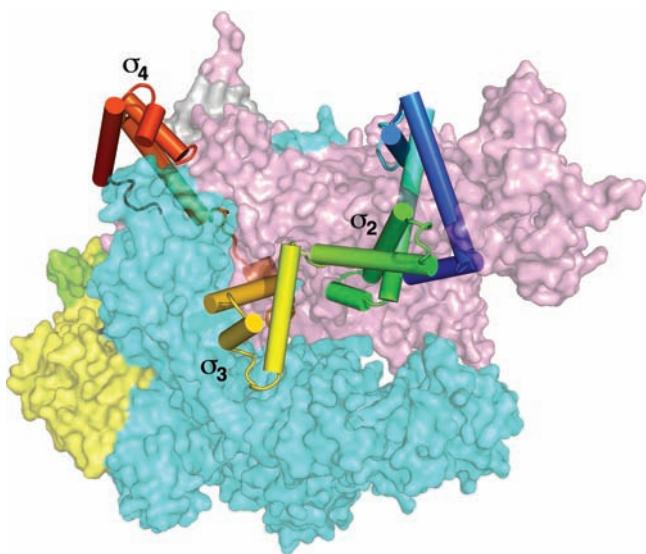


Figure 31-12 X-ray structure of Tth RNAP holoenzyme viewed similarly to Fig. 31-11. The subunits of the core enzyme are represented by their partially transparent molecular surface colored as in Fig. 31-11. The σ subunit is drawn with its α helices as cylinders and colored in rainbow order from its N-terminus (blue) to its C-terminus (red). [Based on an X-ray structure by Dmitry Vassylyev, University of Alabama at Birmingham. PDBid 1IW7.]

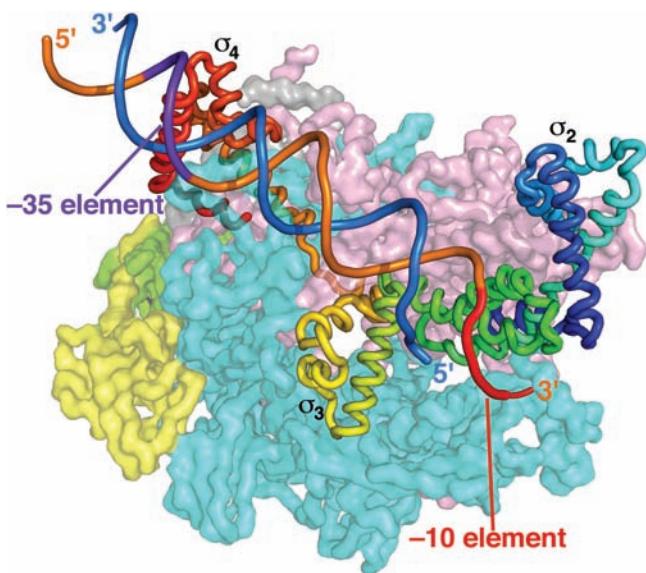
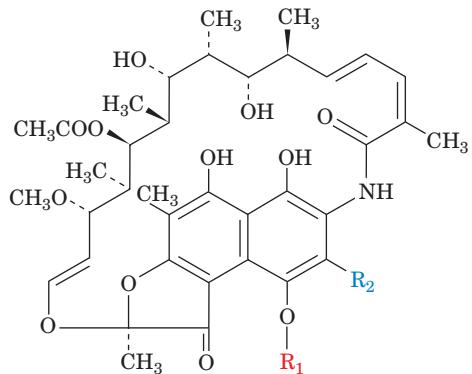


Figure 31-13 Low (6.5 Å) resolution X-ray structure of Taq RNAP holoenzyme in complex with a promoter-containing dsDNA viewed as in Fig. 31-12. The subunits of the core enzyme are represented by their partially transparent molecular surface colored as in Fig. 31-11 (which appears striated due to the structure's low resolution, which permits only the polypeptide backbones to be visualized). The σ subunit is drawn in worm form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA's sugar-phosphate backbone is drawn in cartoon form with the template DNA strand blue and the nontemplate DNA's -10 element red, its -35 element purple, and its remaining portions orange. [Based on an X-ray structure by Seth Darst, The Rockefeller University. PDBid 1L9Z.]

b. Rifamycins Inhibit Prokaryotic Transcription Initiation

Two related antibiotics, **rifamycin B**, which is produced by *Streptomyces mediterranei*, and its semisynthetic derivative **rifampicin**,



Rifamycin B $R_1 = \text{CH}_2\text{COO}^-$; $R_2 = \text{H}$

Rifampicin $R_1 = \text{H}$; $R_2 = \text{CH} = \text{N} - \text{N}(\text{CH}_3) - \text{CH}_3$

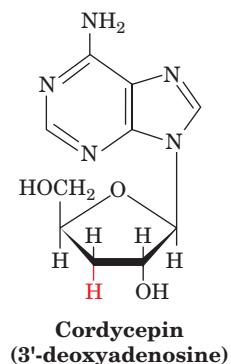
specifically inhibit transcription by prokaryotic, but not eukaryotic, RNAPs. This selectivity and their high potency (bacterial RNAP is 50% inhibited by $2 \times 10^{-8} \text{ M}$ rifampicin) made them medically useful bacteriocidal agents against gram-positive bacteria and tuberculosis. Indeed, few other antibiotics are effective against tuberculosis, which has reached epidemic levels in some parts of the world.

The finding that the β subunits of rifamycin-resistant mutants have altered electrophoretic mobilities first demonstrated that this subunit contains the rifamycin-binding site. Rifamycins inhibit neither the binding of RNAP to the promoter nor the formation of the first phosphodiester bond, but they prevent further chain elongation. The inactivated RNAP remains bound to the promoter, thereby blocking its initiation by uninhibited enzymes. Once RNA chain initiation has occurred, however, rifamycins have no effect on the subsequent elongation process. The rifamycins are therefore useful research tools because they permit the transcription process to be dissected into its initiation and its elongation phases.

The X-ray structure of Taq core enzyme in complex with rifampicin reveals how this antibiotic inhibits RNAP. Rifampicin binds with close complementary fit but little conformational change in a pocket in the β subunit that is located within the main channel, $\sim 12 \text{ \AA}$ distant from the active site Mg^{2+} ion. Model building indicates that the bound rifampicin would sterically interfere with the RNA transcript at positions -2 to -5 in the transcription bubble. Thus, as is observed, rifampicin would not interfere with the initiation of transcription but would mechanically block the extension of the RNA transcript. The residues lining the pocket in which rifampicin binds are highly conserved among prokaryotes but not in eukaryotes, thereby explaining why rifamycins inhibit only bacterial RNAPs.

C. Chain Elongation

The direction of RNA chain elongation; that is, whether it occurs by the addition of incoming nucleotides to the 3' end of the nascent (growing) RNA chain ($5' \rightarrow 3'$ growth; Fig. 31-14a) or by their addition to its 5' terminus ($3' \rightarrow 5'$ growth; Fig. 31-14b), was established by determining the rate at which the radioactive label from $[\gamma-^{32}\text{P}]$ GTP is incorporated into RNA. For $5' \rightarrow 3'$ elongation, the 5' γ -P is permanently labeled and, hence, the chain's level of radioactivity would not change on replacement of the labeled GTP with unlabeled GTP. However, for $3' \rightarrow 5'$ elongation, the 5' γ -P is replaced with the addition of every new nucleotide so that, on replacement of labeled with unlabeled GTP, the nascent RNA chains would lose their radioactivity. The former was observed. *Chain growth must therefore occur in the $5' \rightarrow 3'$ direction* (Fig. 31-14a), the same direction as DNA is synthesized. This conclusion is corroborated by the observation that the antibiotic **cordycepin**,



an adenosine analog that lacks a 3'-OH group, inhibits bacterial RNA synthesis. Its addition to the 3' end of RNA, as is expected for $5' \rightarrow 3'$ growth, prevents the RNA chain's

further elongation. Cordycepin would not have this effect if chain growth occurred in the opposite direction because it could not be appended to an RNA's 5' end.

a. Transcription Supercoils DNA

RNA chain elongation requires that the double-stranded DNA template be opened up at the point of RNA synthesis so that the template strand can be transcribed to its complementary RNA strand. In doing so, the RNA chain only transiently forms a short length of RNA–DNA hybrid duplex, as is indicated by the observation that transcription leaves the template duplex intact and yields single-stranded RNA. The unpaired transcription bubble of the DNA in the open initiation complex apparently travels along the DNA with the RNAP. There are two ways this might occur (Fig. 31-15):

1. If the RNAP followed the template strand in its helical path around the DNA, the DNA would build up little supercoiling because the DNA duplex would never be unwound by more than about a turn. However, the RNA transcript would wrap around the DNA, once per duplex turn. This model is implausible since it is unlikely that its DNA and RNA could be readily untangled: The RNA would not spontaneously unwind from the long and often circular DNA in any reasonable time, and no known topoisomerase can accelerate this process.

2. If the RNAP moves in a straight line while the DNA rotates, the RNA and DNA will not become entangled. Rather, the DNA's helical turns are pushed ahead of the advancing transcription bubble so as to more tightly wind the DNA ahead of the bubble (which promotes positive supercoiling), and the DNA behind the bubble becomes equivalently unwound (which promotes negative supercoiling, although note that the linking number of the entire DNA remains unchanged). This model is supported by the observa-

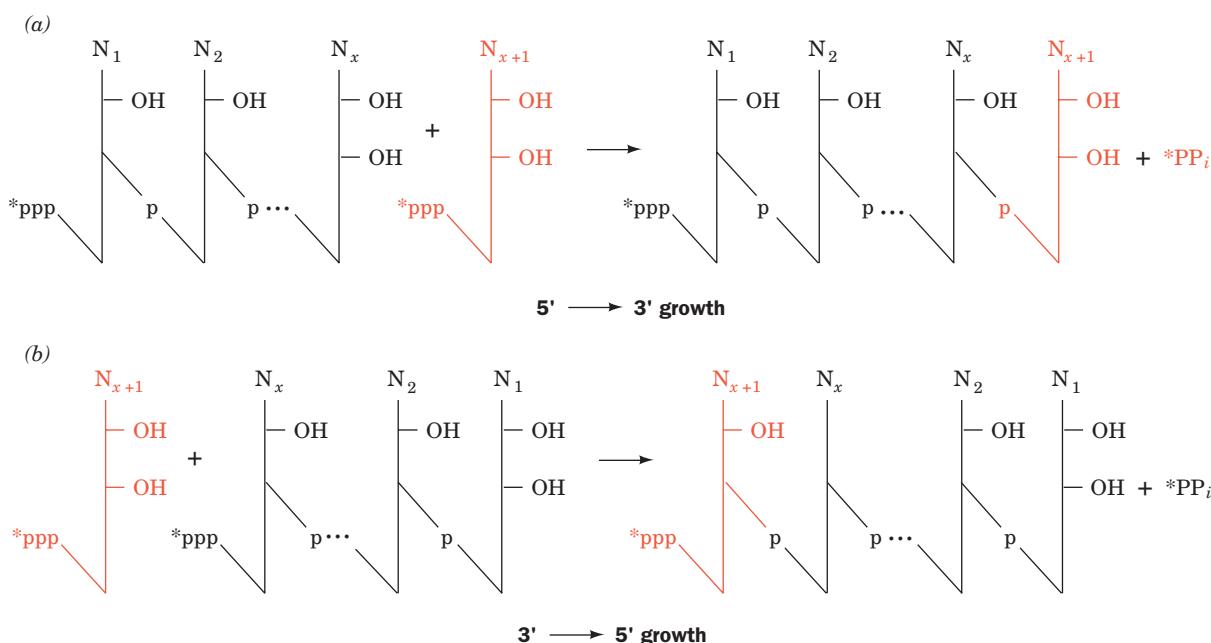


Figure 31-14 The two possible modes of RNA chain growth. Growth may occur (a) by the addition of nucleotides to the 3'

end and (b) by the addition of nucleotides to the 5' end. RNA polymerase catalyzes the former reaction.

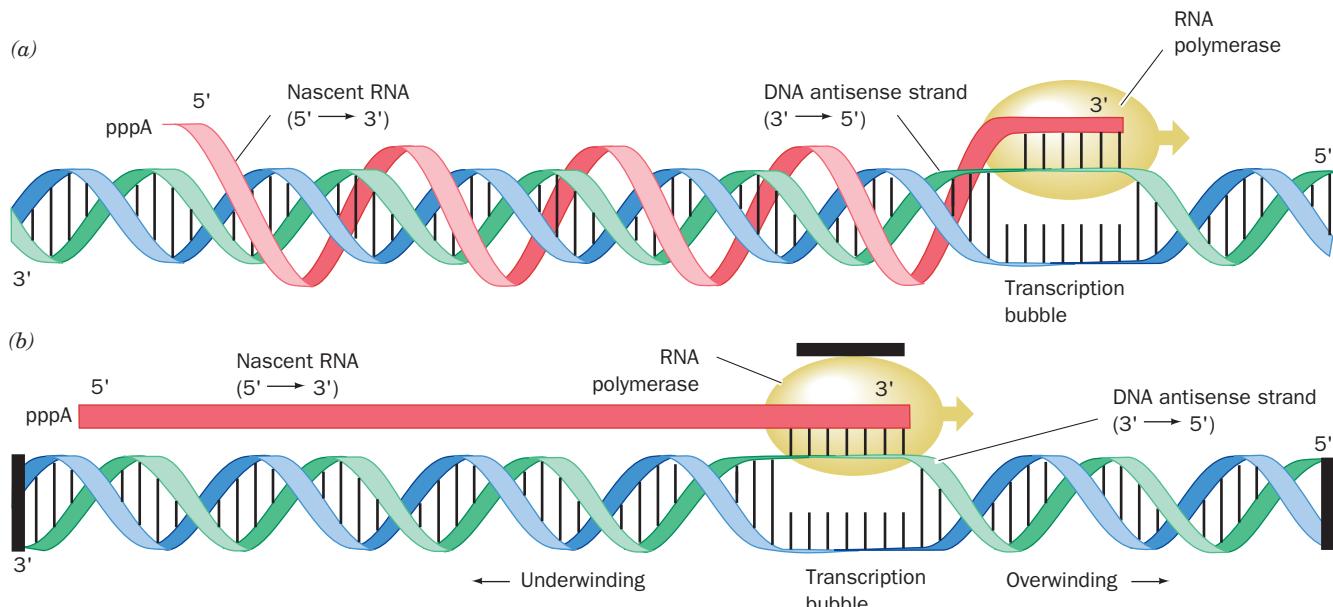


Figure 31-15 RNA chain elongation by RNA polymerase. In the region being transcribed, the DNA double helix is unwound by about a turn to permit the DNA's sense strand to form a short segment of DNA–RNA hybrid double helix with the RNA's 3' end. As the RNAP advances along the DNA template (here to the right), the DNA unwinds ahead of the RNA's growing 3' end and rewinds behind it, thereby stripping the newly synthesized RNA from the template (antisense) strand. (a) One way this might occur is by the RNAP following the path of the template strand about the DNA double helix, in which case the transcript would become wrapped about the DNA once per duplex turn.

(b) A second and more plausible possibility is that the RNA moves in a straight line while the DNA rotates beneath it. In this case the RNA would not wrap around the DNA but the DNA would become overwound ahead of the advancing transcription bubble and unwind behind it (consider the consequences of placing your finger between the twisted DNA strands in this model and pushing toward the right). The model presumes that the ends of the DNA, as well as the RNAP, are prevented from rotating by attachments within the cell (black bars). [After Futcher, B., *Trends Genet.* **4**, 271, 272 (1988).]

tions that the transcription of plasmids in *E. coli* causes their positive supercoiling in gyrase mutants (which cannot relax positive supercoils; Section 29-3Cd) and their negative supercoiling in topoisomerase I mutants (which cannot relax negative supercoils; Section 29-3Ca). In fact, by tethering RNAP to a glass surface and allowing it to transcribe DNA that had been fluorescently labeled at one end, Kazuhiko Kinoshita demonstrated, through fluorescence microscopy (using techniques similar to those showing that the F_1F_0 -ATPase is a rotary engine; Section 22-3Ce), that single DNA molecules rotated in the expected direction during transcription.

Inappropriate superhelicity in the DNA being transcribed halts transcription (Section 29-3C). Quite possibly the torsional tension in the DNA generated by negative superhe-

licity behind the transcription bubble is required to help drive the transcriptional process, whereas too much such tension prevents the opening and maintenance of the transcription bubble.

b. Transcription Occurs Processively and Rapidly

The *in vivo* rate of transcription is 20 to 70 nucleotides per second. Once an RNAP molecule has initiated transcription and moved away from the promoter, another RNAP can follow suit. The synthesis of RNAs that are needed in large quantities, ribosomal RNAs, for example, is initiated as often as is sterically possible, about once per second (Fig. 31-16). Processivity is accomplished without an obvious clamplike structure such as the sliding clamp of

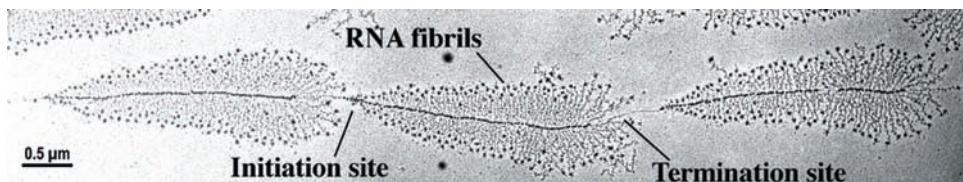


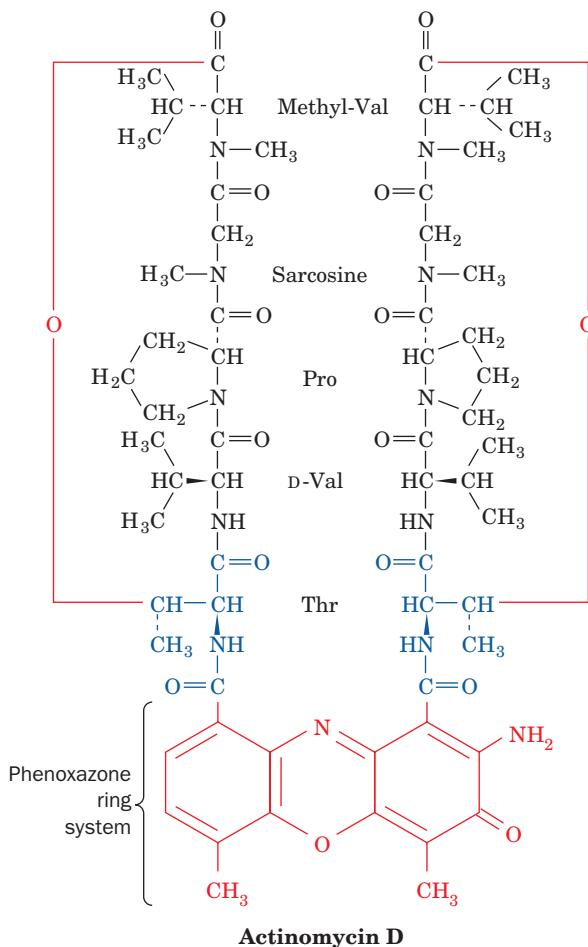
Figure 31-16 An electron micrograph of three contiguous ribosomal genes from oocytes of the salamander *Pleurodeles waltl* undergoing transcription. The “arrowhead” structures result from the increasing lengths of the nascent RNA chains as

the RNAP molecules synthesizing them move from the initiation site on the DNA to the termination site. [Courtesy of Ulrich Scheer, University of Würzburg, Germany.]

E. coli DNA polymerase III (Fig. 30-14). However, the RNAP itself apparently functions as a sliding clamp by binding tightly but flexibly to the DNA–RNA complex. In experiments in which the RNAP was immobilized and a magnetic bead was attached to the DNA, the bead was observed to undergo up to 180 rotations (representing nearly 2000 base pairs at 10.4 bp per turn) before the polymerase slipped.

c. Intercalating Agents Inhibit Both RNA and DNA Polymerases

Actinomycin D,



a useful antineoplastic (anticancer) agent produced by *Streptomyces antibioticus*, tightly binds to duplex DNA and, in doing so, strongly inhibits both transcription and DNA replication, presumably by interfering with the passage of RNA and DNA polymerases. The NMR structure of actinomycin D in complex with a duplex DNA composed of two strands of the self-complementary octamer d(GAAGCTTC) reveals that the DNA assumes a B-like conformation in which the actinomycin's **phenoxazone** ring system, as had previously been shown, is intercalated between the DNA's central G · C base pairs (Fig. 31-17). Consequently, the DNA helix is unwound by $\sim 30^\circ$ at the intercalation site and the central G · C base pairs are separated by ~ 7 Å. The DNA helix is severely distorted from the normal B-DNA conformation such that it is bent toward its major groove by $\sim 30^\circ$ and its minor groove is wide and

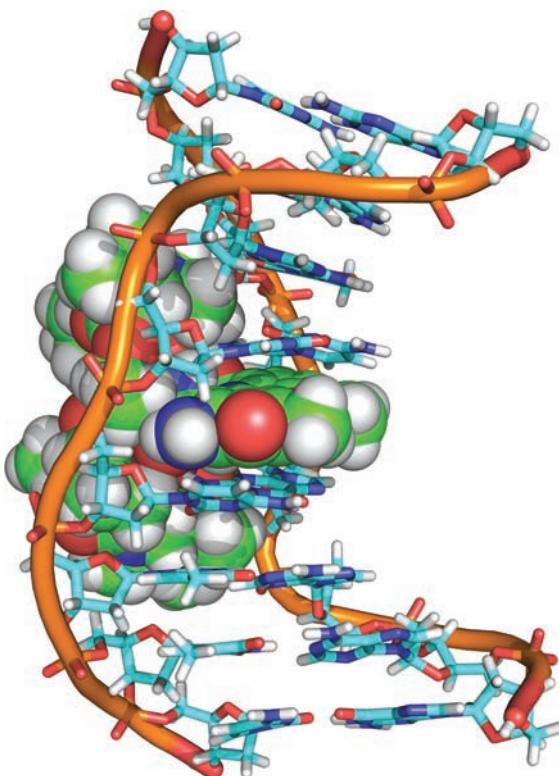


Figure 31-17 NMR structure of actinomycin D in complex with a dsDNA of self-complementary sequence d(GAAGCTTC). The actinomycin D is drawn in space-filling form and the DNA is drawn in stick form with successive P atoms on the same strand connected by orange rods, all colored according to atom type with actinomycin C green, DNA C cyan, H white, N blue, O red, and P orange. The complex is viewed toward the DNA's major groove. The actinomycin D's two cyclic depsipeptides are tightly wedged into the DNA's minor groove and the actinomycin D's phenoxazone ring system is intercalated between the DNA's central G · C base pairs. [Based on an NMR structure by Andrew Wang, University of Illinois. PDBid 1DSC.]

shallow in a manner resembling that of A-DNA. Actinomycin D's two chemically identical cyclic **depsipeptides** (having both peptide bonds and ester linkages) extend in opposite directions from the intercalation site along the minor groove of the DNA. The complex is stabilized through the formation of base-peptide and phenoxazone–sugar–phosphate backbone hydrogen bonds, as well as by hydrophobic interactions, in a way that explains the preference of actinomycin D to bind to DNA with its phenoxazone ring intercalated between the base pairs of a 5'-GC-3' sequence. Several other intercalation agents, including ethidium and acridine orange (Sections 6-6Ca and 29-3Ba), also inhibit nucleic acid synthesis, presumably by similar mechanisms.

D. Chain Termination

Electron micrographs such as Fig. 31-16 suggest that DNA contains specific sites at which transcription is terminated. In this section we discuss how transcription is terminated in bacteria. The eukaryotic process is discussed in Section 31-4Ab.

a. The RNA at Intrinsic Terminators Has an Oligo(U) Tract Preceded by a G + C-Rich Stem

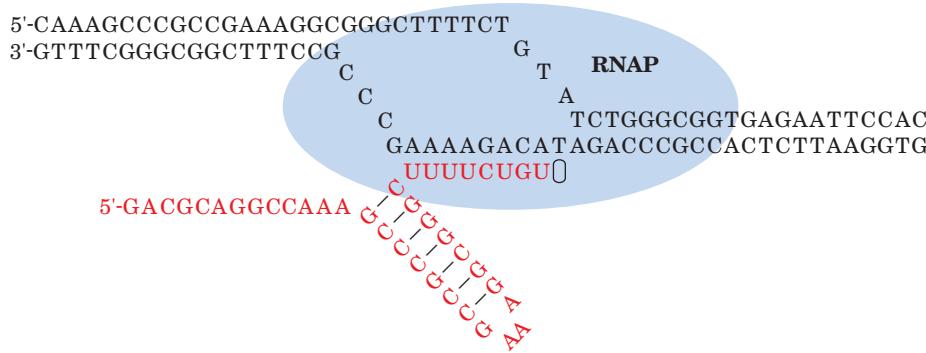
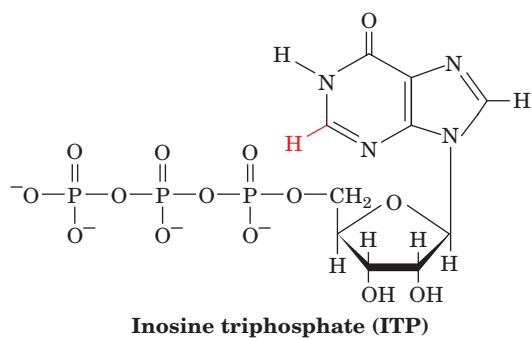
Around half the transcriptional termination sites in *E. coli* are **intrinsic** or **spontaneous terminators**, that is, they induce termination without assistance. The sequences of these terminators share two common features (Fig. 31-18):

1. A tract of 7 to 10 consecutive A · T's with the A's on the template strand, sometimes interrupted by one or more different base pairs. The transcribed RNA is terminated in or just past this sequence.

2. A G + C-rich segment with a palindromic (2-fold symmetric) sequence that is immediately upstream of the series of A · T's.

The RNA transcript of this region can therefore form a self-complementary “hairpin” structure that is terminated by several U residues (Fig. 31-18).

The stability of a terminator's G + C-rich hairpin and the weak base pairing of its oligo(U) tail to template DNA are important factors in ensuring proper chain termination. In fact, model studies have shown that oligo(dA · rU) forms a particularly unstable hybrid helix although oligo(dA · dT) forms a helix of normal stability. In fact, oligo(dA · rU) tracts as long as 8 or 9 bp are unstable at room temperature when not bound to RNAP. The formation of the G + C-rich hairpin causes RNAP to pause for several seconds at the termination site. Mutations in the termination site that decrease the strengths of these associations reduce the efficiency of chain termination (the fraction of transcripts that are terminated at that site) and often eliminate it. Termination efficiency is similarly diminished when *in vitro* transcription is carried out with GTP replaced by **inosine triphosphate (ITP)**:



I · C pairs are weaker than G · C pairs because the hypoxanthine base of I, which lacks the 2-amino group of G, can only make two hydrogen bonds to C, thereby decreasing the hairpin's stability.

Despite the foregoing, experiments by Michael Chamberlin in which segments of highly efficient terminators were swapped via recombinant DNA techniques indicate that the RNA terminator hairpin and U-rich 3' tail do not function independently of their corresponding DNA's upstream and downstream flanking regions. Indeed, terminators that lack a U-rich segment can be highly efficient when joined to the appropriate sequence immediately downstream from the termination site.

These and other observations have led to three not necessarily mutually exclusive models to explain how intrinsic terminators work:

1. The forward translocation model, in which hairpin formation pushes the RNAP forward without the concomitant elongation of the RNA transcript. This would shorten the RNA–DNA hybrid by as much as several base pairs, thereby destabilizing it.

2. The RNA pullout model, in which hairpin formation mechanically pulls the RNA out of the RNA–DNA hybrid.

3. The allosteric model, in which hairpin formation induces a conformational change in the RNAP that permits the upstream nontemplate DNA strand to displace the weakly bound oligo(U) tail from the template DNA strand.

In an effort to differentiate these models, Robert Landick and Steven Block used **optical traps** to exert a pulling force on one or the other end of the DNA that a single RNAP molecule was transcribing or on its RNA transcript. An optical trap consists of a highly focused laser beam that is typically generated by sending it through a microscope objective lens. The resulting strong electric field gradient across the constricted region of the beam attracts dielectric (insulating) particles such as submicrometer sized polystyrene beads to the center of the beam where the electric field is strongest. The force on the particle varies directly with its displacement from the center of the beam. By attaching a single macromolecule of interest to such a bead in an optical trap and fixing the macromolecule's other end or attaching it to a bead in a second optical trap,

Figure 31-18 An *E. coli* intrinsic terminator. Its transcription yields an RNA (red) with a self-complementary G + C-rich segment that forms a base-paired hairpin immediately followed by a sequence of 4 to 10 consecutive U's that base-pair with the template A's in the transcription bubble. The oval symbol represents the binding site for an incoming NTP. [After a drawing by Park, J.-S. and Roberts, J.W., Cornell University.]

a force can be exerted on the molecule by laterally displacing the beam as little as subnanometer distances. Such a device is known as an **optical tweezers**.

In the optical tweezers diagrammed in Fig. 31-19a, pulling apart the two optical traps would assist the RNAP in translocating along the DNA, whereas attaching the other end of the DNA to a bead would hinder this process. The application of either an assisting or hindering force to the DNA does not significantly affect the termination efficiencies of any of the three terminators shown in Fig. 31-19c. Evidently, forward translocation is not a general feature of intrinsic termination. However, the efficiency of the t500 terminator with a mutation in its hairpin varies with the force on the DNA, which indicates that forward translocation occurs with some terminators.

In the optical tweezers diagrammed in Fig. 31-19b, pulling on the RNA with sufficient force to disrupt the first 2 or 3 base pairs of the terminator hairpin reduces the ter-

mination efficiencies of all three terminators. If the force is greater than that required to fully unfold the hairpin, transcription efficiency is indistinguishable from that of the corresponding terminator containing only its oligo(U) tract. This suggests that the formation of the hairpin base pairs disrupts the adjacent RNA–DNA hybrid as predicted by the RNA pullout model.

Curiously, a force weaker than that required to disrupt hairpin base pairs increases termination efficiency. However, the presence of ssDNA complementary to the transcript eliminates this latter effect. Evidently, the RNA upstream of the terminator forms weakly base-paired secondary structures that compete with the formation of the terminator hairpin. In this way, the RNA sequence upstream of an intrinsic terminator modulates the efficiency of termination.

Since RNAP stabilizes the RNA–DNA hybrid, allosteric changes to RNAP by hairpin formation may also

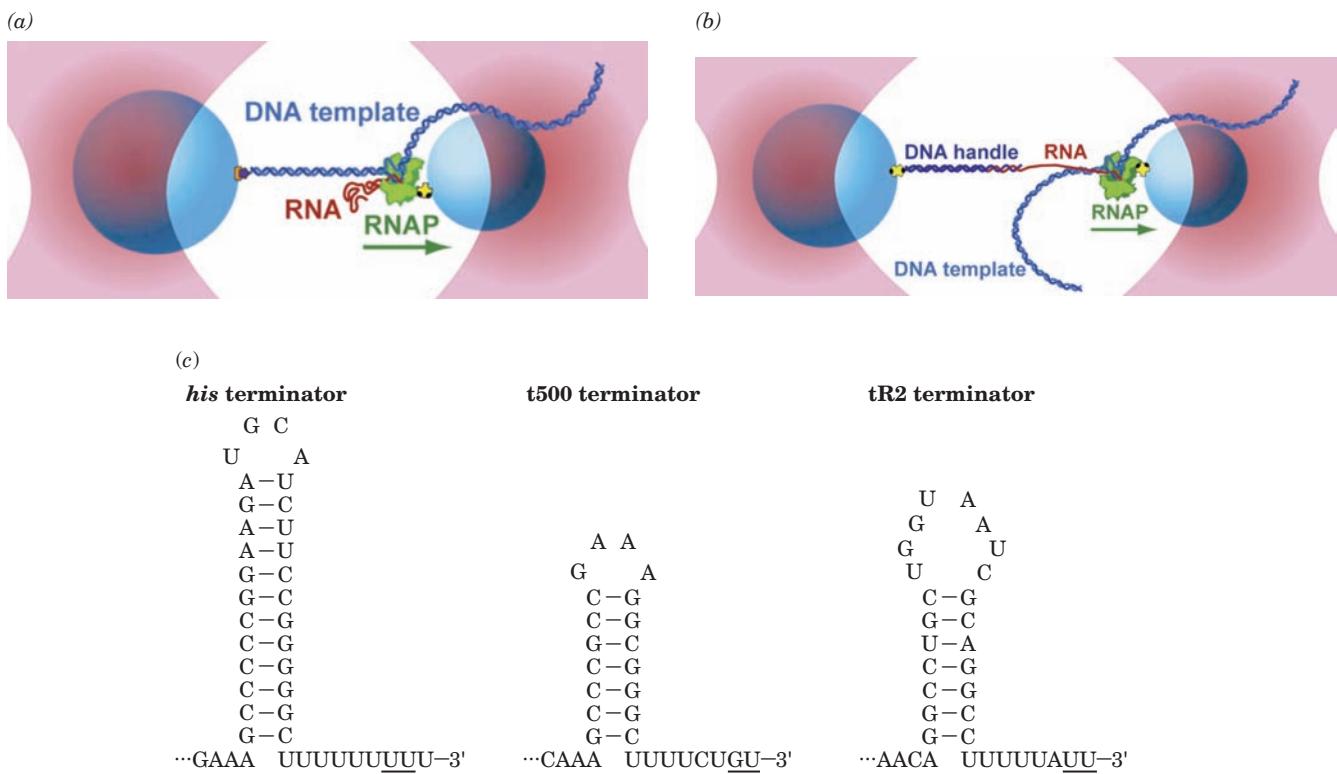


Figure 31-19 Apparatus for single molecule pulling assays on RNAP elongation complexes (not to scale). (a) A DNA-pulling assay. The RNAP (green; its direction of translocation is indicated by the green arrow) in an elongation complex was attached to an avidin-coated polystyrene bead (light blue) via a biotin-avidin linkage (yellow and black; Section 22-3Ce), and either the upstream end of the template DNA (dark blue) as shown or the downstream end was attached to a somewhat larger polystyrene bead via a digoxigenin–antidigoxigenin linkage [purple and orange; digoxigenin is a steroid related to digitalin (Fig. 20-21b) that has high antigenicity and antidigoxigenin is an antibody to which it specifically binds]. The RNA product of the complex (red) was untethered. The two beads were then placed in separate optical traps (pink) and the beads were pulled apart

while the elongation complex synthesized RNA. The horizontal displacement of a bead from the center of its optical trap is indicative of the pulling force on the bead. (b) An RNA-pulling assay. As in Part a but with the template DNA untethered and the RNA emerging from the RNAP attached to the second bead via a DNA handle (which had a 25-nt 3' overhang complementary to the 5' end of the RNA) tethered to the larger bead via a biotin–avidin linkage. (c) Structures of the three intrinsic terminators investigated in this study showing their hairpins and poly(U) tracts. The underlined bases are the transcript termination sites. The termination efficiencies of the his, t500, and tR2 terminators are normally 77%, 98%, and 46%, respectively. [Courtesy of Steven Block, Stanford University.]

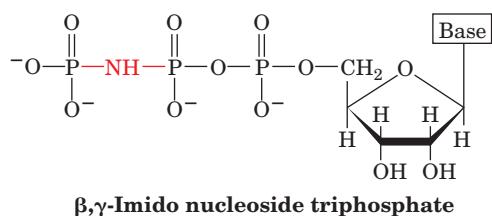
influence termination efficiency. Indeed, mutations in the β subunit of RNAP can both increase and decrease termination efficiency. However, the way in which hairpin formation induces allosteric changes to RNAP is as yet unknown.

b. Many Bacterial Terminators Require the Assistance of Rho Factor

Around half the termination sites in *E. coli* lack any obvious similarities and are unable to form strong hairpins; they require the participation of a protein known as **Rho factor** to terminate transcription. Rho factor was discovered through the observation that *in vivo* transcripts are often shorter than the corresponding *in vitro* transcripts. Rho factor, a RecA family hexameric helicase (Section 30-2Ca) of identical 419-residue subunits, enhances the termination efficiency of spontaneously terminating transcripts as well as inducing the termination of nonspontaneously terminating transcripts.

Several key observations have led to a model of Rho-dependent termination:

1. Rho unwinds RNA–DNA and RNA–RNA double helices by translocating along a single strand of RNA in its $5' \rightarrow 3'$ direction. This process is powered by the hydrolysis of NTPs to NDPs + P_i with little preference for the identity of the base. NTPase activity is required for Rho-dependent termination as is demonstrated by its *in vitro* inhibition when the NTPs are replaced by their β,γ -imido analogs,



substances that are RNAP substrates but cannot be hydrolyzed by Rho.

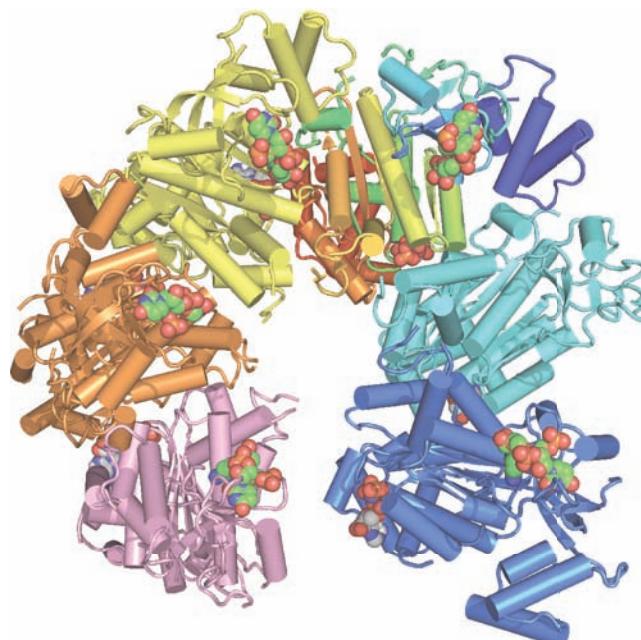
2. Genetic manipulations indicate that Rho-dependent termination requires the presence of a specific recognition sequence on the newly transcribed RNA upstream of the termination site. The recognition sequence must be on the nascent RNA rather than the DNA as is demonstrated by Rho's inability to terminate transcription in the presence of pancreatic RNase A. The essential features of this termination site have not been fully elucidated; the construction of synthetic termination sites indicates that it consists of 80 to 100 nucleotides that lack a stable secondary structure and contain multiple regions that are rich in C and poor in G.

These observations suggest that Rho attaches to nascent RNA at its recognition sequence [named ***rut*** (for Rho utilization), a C-rich segment of at least 40 nt] and then translocates along the RNA in the $5' \rightarrow 3'$ direction until it encounters an RNAP paused at the termination site (without the pause, Rho might not be able to overtake the RNA

polymerase). There, as Jeffrey Roberts has shown, Rho pushes the RNAP forward in a way that partially rewinds its dsDNA helix at the transcription bubble while unwinding the RNA–DNA hybrid helix (forward translocation), thus releasing the RNA. Rho-terminated transcripts have 3' ends that typically vary over a range of \sim 50 nucleotides. This suggests that Rho pries the RNA away from the template DNA rather than “pushing” an RNA release “button.” TCRF (alternatively, Mfd), which functions during transcription-coupled repair in *E. coli* to release a stalled RNAP from a damaged template by stripping away its bound RNA (Section 30-5Bb), is an ATP-powered DNA translocase that is thought to mechanically act on RNAP in much the same way as Rho.

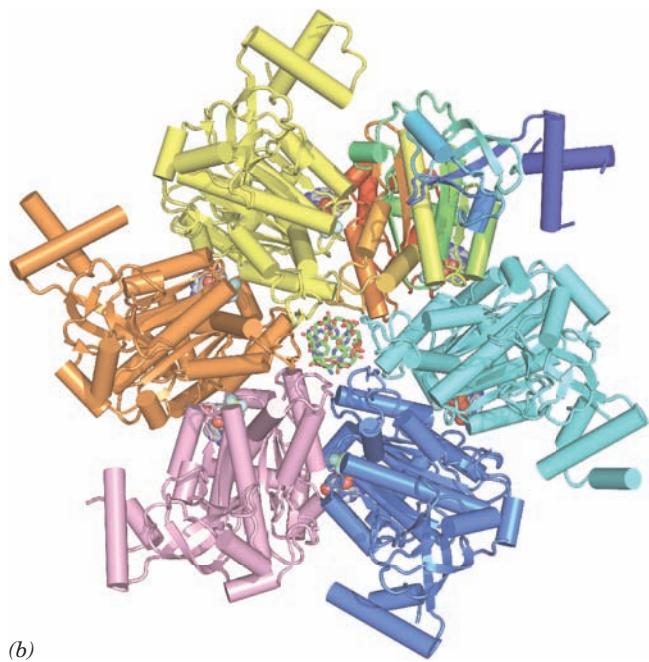
Each Rho subunit consists of two domains that can be separated by proteolysis: Its N-terminal domain binds single-stranded polynucleotides and its C-terminal domain, which is homologous to the α and β subunits of the F₁-ATPase (Section 22-3Cb), binds an NTP. The X-ray structure of Rho in complex with AMPPNP and an 8-nt RNA, r(UC)₄ (Fig. 31-20a), determined by James Berger, reveals that Rho forms a hexameric lock washer-shaped helix that is 120 Å in diameter with an \sim 30-Å-diameter central hole and whose first and sixth subunits are separated by a 12-Å gap and a rise of 45 Å along the helix axis. The RNAs, only a single UC unit of which is visible in each chain, bind along the top of the helix to the so-called primary RNA binding sites on the N-terminal domains, whereas AMPPNP binds to the C-terminal domains (which are further from the viewer in Fig. 31-20a than the N-terminal domains) at the interface between subunits. This X-ray structure represents an open state that has bound to *rut* site mRNA and is poised to bind additional mRNA upon its entry into the central cavity through the gap.

In the X-ray structure of Rho in complex with rU₁₂ and the ATP mimic ADP · BeF₃ (Fig. 31-20b), also determined by Berger, the helicase's six subunits have formed a closed ring in which each subunit has a different conformation. The RNA, only 6 nt of which are visible, assumes a right-handed helical conformation with its 5' end closest to the viewer in Fig. 31-20b and binds to Rho's N-terminal domains in the helicase's central channel, the so-called secondary RNA binding site. Protein loops that extend from the walls of the central channel to interact with the RNA are helically arranged like the steps of a right-handed spiral staircase such that they track the RNA's sugar–phosphate backbone, much like the central loops of E1 protein (an AAA+ family hexagonal helicase) tracks its centrally bound ssDNA (Section 30-2Ca). The different conformations of Rho's six subunits indicate that they sequentially undergo a series of six NTP-driven conformational changes and that these changes are allosterically coupled so that they progress around the hexamer in a wavelike manner. Since each of the foregoing loops maintains its grip on the same nucleotide during this process, the helicase translocates along its bound RNA, in much the same way that E1 protein translocates along its bound ssDNA. Why, then, do Rho and E1 protein move in opposite directions? Comparison of the structures of Rho and E1 protein



(a)

Figure 31-20 X-ray structures of Rho factor. (a) Rho in complex with r(UC)₄ (only one UC unit of which is visible) and AMPPNP. Each of the protein's six subunits are drawn in tube-and-arrow form in different colors with the upper right subunit colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The UC units and the AMPPNP are shown in space-filling form with UC C green, AMPPNP C gray, N blue, O red, and P orange. The hexamer has a lock washer-like shape with the blue subunit ~45 Å closer to the viewer than the pink



(b)

subunit. (b) Rho in complex with rU₁₂ (only 6 nt of which are visible) and ADP · BeF₃. The protein is drawn and colored as in Part a. The RNA and ADP · BeF₃ are shown in stick and space-filling form, respectively, with RNA C green, ADP C gray, N blue, O red, P orange, Be light green, and F light blue. Note that each of the Rho subunits has a different conformation. [Based on X-ray structures by James Berger, University of California at Berkeley. PDBid 1PVO and 3ICE.]

indicates that the relative order of the conformational states around the hexamer for Rho is opposite that for E1 protein. Evidently, the “firing order” of Rho’s NTPase sites around the hexamer is the reverse of that of E1 protein, thus accounting for their differing directions of translocation. Presumably, other RecA and AAA+ family hexagonal helicases otherwise have similar mechanisms.

E. Eukaryotic RNA Polymerases

Eukaryotic nuclei, as Robert Roeder and William Rutter discovered, contain three distinct types of RNAPs that differ in the RNAs they synthesize:

1. RNA polymerase I (RNAP I; also called Pol I and RNAP A), which is located in the nucleoli (dense granular bodies in the nuclei that contain the ribosomal genes; Section 31-4Bb), synthesizes precursors of most ribosomal RNAs (rRNAs).

2. RNA polymerase II (RNAP II; also called Pol II and RNAP B), which occurs in the nucleoplasm, synthesizes mRNA precursors.

3. RNA polymerase III (RNAP III; also called Pol III and RNAP C), which also occurs in the nucleoplasm, syn-

thesizes the precursors of 5S ribosomal RNA, the tRNAs, and a variety of other small nuclear and cytosolic RNAs.

Eukaryotic nuclear RNAPs have considerably greater subunit complexity than those of prokaryotes. These enzymes have molecular masses of up to 600 kD and, as is indicated in Table 31-2, each contains two nonidentical “large” (>120 kD) subunits comprising ~65% of its mass that are homologs of the prokaryotic RNAP β' and β subunits and up to 12 additional “small” (<50 kD) subunits, two of which are homologs of prokaryotic RNAP α , and one of which is a homolog of prokaryotic RNAP ω . Of these small subunits, five are identical in all three eukaryotic RNAPs and two others (the RNAP α homologs) are identical in RNAPs I and III. Two of the RNAP II subunits, Rbp4 and Rbp7, are not essential for activity and, in fact, are present in RNAP II in less than stoichiometric amounts. (Curiously, Rbp7 has a 102-residue segment that is 30% identical to a portion of σ^{70} , the predominant *E. coli* σ factor.) Thus 10 of the 12 RNAP II subunits are either identical or closely similar to subunits of RNAPs I and III (Table 31-2). Moreover, the sequences of these subunits are highly conserved (~50% identical) across species from yeast to humans (and to a lesser extent between eukaryotes and bac-

Table 31-2 RNA Polymerase Subunits^a

<i>S. cerevisiae</i> RNAP I (14 subunits)	<i>S. cerevisiae</i> RNAP II (12 subunits)	<i>S. cerevisiae</i> RNAP III (15 subunits)	<i>E. coli</i> RNAP Core (5 subunits)	Class ^b
Rpa1 (A190)	Rbp1 (B220)	Rpc1 (C160)	β'	Core
Rpa2 (A135)	Rbp2 (B150)	Rpc2 (C128)	β	Core
Rpc5 (AC40)	Rpb3 (B44.5)	Rpc5 (AC40)	α	Core
Rpc9 (AC19)	Rpb11 (B13.6)	Rpc9 (AC19)	α	Core
Rbp6 (ABC23)	Rbp6 (ABC23)	Rpb6 (ABC23)	ω	Core/common
Rpb5 (ABC27)	Rpb5 (ABC27)	Rpb5 (ABC27)		Common
Rpb8 (ABC14.4)	Rpb8 (ABC14.4)	Rpb8 (ABC14.4)		Common
Rbp10 (ABC10 β)	Rpb10 (ABC10 β)	Rpb10 (ABC10 β)		Common
Rpb12 (ABC10 α)	Rpb12 (ABC10 α)	Rpb12 (ABC10 α)		Common
Rpa9 (A12.2)	Rpb9 (B12.6)	Rpc12 (C11)		
Rpa8 (A14) ^c	Rpb4 (B32)	—		
Rpa4 (A43) ^c	Rpb7 (B16)	Rpc11 (C25)		
+2 others ^d		+4 others ^d		

^aHomologous subunits occupy the same row. In the alternative subunit names in parentheses, the letter(s) indicates the RNAPs in which the subunit is a component (A, B, and C for RNAPs I, II, and III) and the numbers indicate its approximate molecular mass in kilodaltons.

^bCore: sequence partially homologous in all RNAPs; common: shared by all eukaryotic RNAPs.

^cPotential homologs of Rpb4 and Rpb7.

^dRpa3 (A49) and Rpa5 (A34.5) in RNAP I and Rpc3 (C74), Rpc4 (C53), Rpc6 (C34), and Rpc8 (C31) in RNAP III.

Source: Mainly Cramer, P., *Curr. Opin. Struct. Biol.* **12**, 89 (2002).

teria). In fact, in all ten cases tested, a human RNAP II subunit could replace its counterpart in yeast without loss of cell viability.

Rpb1, the β' homolog in RNAP II, has an extraordinary C-terminal domain (**CTD**). In mammals, it contains 52 highly conserved repeats of the heptad PTSPSYS (26 repeats in yeast with other eukaryotes having intermediate values). Five of the seven residues in these particularly hydrophilic repeats bear hydroxyl groups and at least 50 of them, predominantly those on the second Ser residue in each heptad, are subject to reversible phosphorylation by **CTD kinases** and **CTD phosphatases**. RNAP II initiates transcription only when the CTD is unphosphorylated but commences elongation only after the CTD has been phosphorylated, which suggests that this process triggers the conversion of RNAP II's initiation complex to its elongation complex. Charge-charge repulsions between nearby phosphate groups probably cause a highly phosphorylated CTD to project as far as 500 Å from the globular portion of RNAP II. Indeed, as we shall see, the phosphorylated CTD provides the binding sites for numerous auxiliary factors that have essential roles in the transcription process.

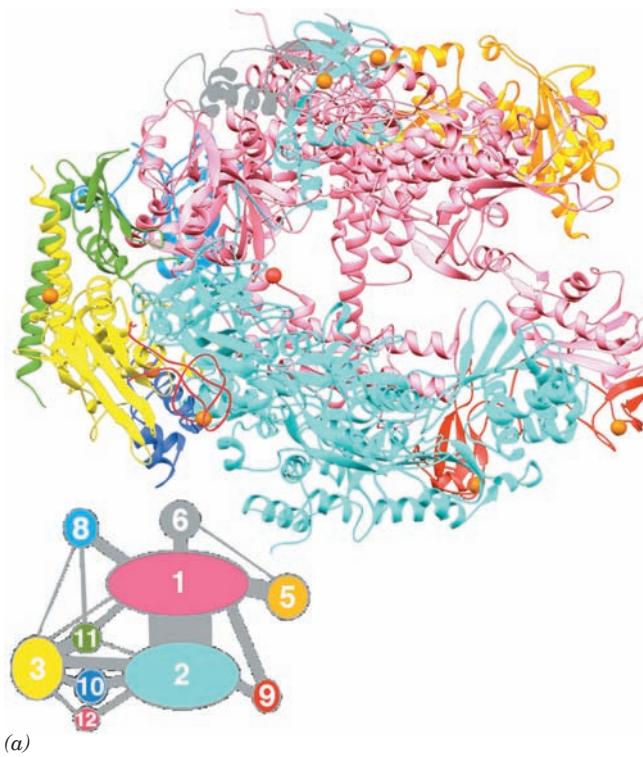
In contrast to the somewhat smaller prokaryotic RNAP holoenzymes, eukaryotic RNAPs do not independently bind their target DNAs. Rather, as we shall see in Section 34-3B, they are recruited to their target promoters through the mediation of complexes of transcription factors and their ancillary proteins that, in the case of RNAP II-transcribed genes, are so large and complicated that they collectively dwarf RNAP II.

In addition to the foregoing nuclear enzymes, eukaryotic cells contain separate mitochondrial and (in plants) chloroplast RNAPs. These small (~100 kD) single-subunit RNAPs, which resemble those encoded by certain bacteriophages, are much simpler than the nuclear RNAPs although they catalyze the same reaction.

a. X-Ray Structures of Yeast RNAP II Reveal a Transcribing Complex

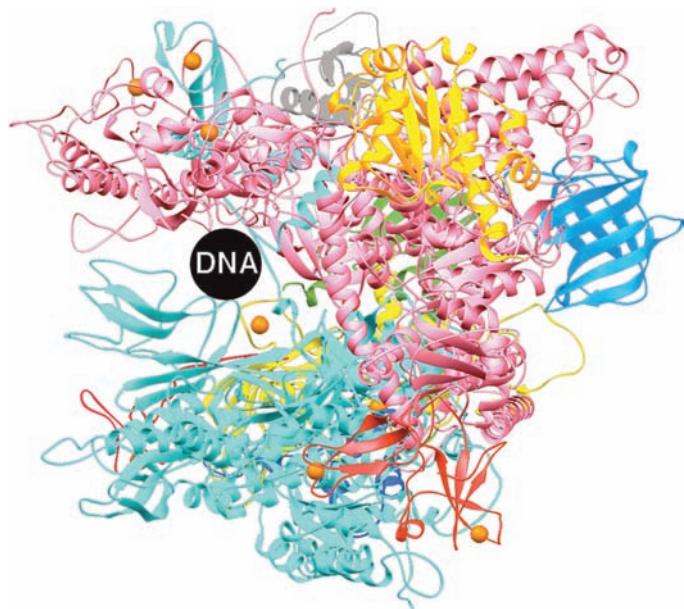
In a crystallographic tour de force, Roger Kornberg determined the X-ray structure of yeast (*S. cerevisiae*) RNAP II that lacks its nonessential Rpb4 and Rpb7 subunits (Fig. 31-21). This enzyme, as expected, resembles Tth RNAP (Fig. 31-11) in its overall crab claw-like shape and in the positions and core folds of their homologous subunits although, of course, RNAP II is somewhat larger than and has several subunits that have no counterpart in bacterial RNAPs. RNAP II binds two Mg²⁺ ions at its active site in the vicinity of five conserved acidic residues (although one of these Mg²⁺ ions appears to be weakly bound and hence is only faintly visible in the X-ray structure; it apparently accompanies the incoming NTP). This suggests that RNAPs catalyze RNA elongation via a two-metal ion mechanism similar to that employed by DNA polymerases (Section 30-2Af). As is the case with bacterial RNAPs, the surface of RNAP II is almost entirely negatively charged except for its main channel and the region about the active site, which are positively charged.

Although, as mentioned above, RNAP II does not normally initiate transcription by itself, Kornberg found that it



(a)

Figure 31-21 The X-ray structure of yeast RNAP II that lacks its Rpb4 and Rpb7 subunits. (a) The enzyme is oriented similarly to Tth RNAP in Fig. 31-11 and its subunits are colored as is indicated in the accompanying diagram, with the subunits homologous to those of Tth RNAP given the same colors. The strongly bound Mn^{2+} ion (physiologically Mg^{2+}) that marks the active site is shown as a red sphere and the enzyme's 8 bound Zn^{2+} ions are shown as orange spheres. The Rpb1 C-terminal



(b)

domain (CTD) is not visible due to disorder. In the accompanying diagram, the area of each numbered ellipsoid is proportional to the corresponding subunit's size and the width of each gray line connecting a pair of subunits is proportional to the surface area of their interface. (b) View of the enzyme from the right in Part a looking into its DNA-binding main channel. The black circle has the approximate diameter of B-DNA. [Based on an X-ray structure by Roger Kornberg, Stanford University. PDBid 1I50.]

will do so on a dsDNA bearing a 3' single-stranded tail at one end. Consequently, incubating yeast RNAP II with the DNA shown in Fig. 31-22a and all NTPs but UTP yielded the DNA–RNA hybrid helix diagrammed in Fig. 31-22a bound to RNAP II. The X-ray structure of this paused transcribing complex revealed, as expected, that the dsDNA had bound in the enzyme's main channel (Fig. 31-22b,c; transcription resumed on soaking the crystals in UTP, thereby demonstrating that the crystalline complex was active). In comparison with the X-ray structure of RNAP II alone, a massive (~50 kD) portion of Rpb1 and Rpb2 named the “clamp” has swung down over the DNA to trap it in the main channel, in large part accounting for the enzyme's essentially infinite processivity. The mainly rigid motion of the clamp is mediated by conformational changes at five so-called switch regions at the base of the clamp in which three of these switches, which are disordered in the structure of RNAP II alone, become ordered in the transcribing complex.

The DNA unwinds by three bases before entering the active site (which is contained on Rpb1). Past this point, however, a portion of Rpb2 dubbed the “wall” directs the

template strand out of the cleft in an ~90° turn. As a consequence, the template base at the active ($i + 1$) site points toward the floor of the cleft where it can be read out by the active site. This base is paired with the ribonucleotide at the 3' end of the RNA, which is positioned above a 12-Å-diameter pore at the end of a funnel to the protein exterior (also called the secondary channel) through which NTPs presumably gain access to the otherwise sealed off active site. The RNA–DNA hybrid helix adopts a nonstandard conformation intermediate between those of A- and B-DNAs, which is underwound relative to that in the X-ray structure of an RNA–DNA hybrid helix alone (Fig. 29-4). Nearly all contacts that the RNAP makes with the RNA and DNA are with their sugar–phosphate backbones; none are with the edges of their bases. The specificity of the enzyme for a ribonucleotide rather than a deoxyribonucleotide is attributed to the enzyme's recognition of both the incoming ribose sugar and the RNA–DNA hybrid helix. After about one turn of hybrid helix, a loop extending from the clamp called the “rudder” separates the RNA and template DNA strands, thereby permitting the DNA double helix to reform as it exits the enzyme (although the unpaired 5' tail of

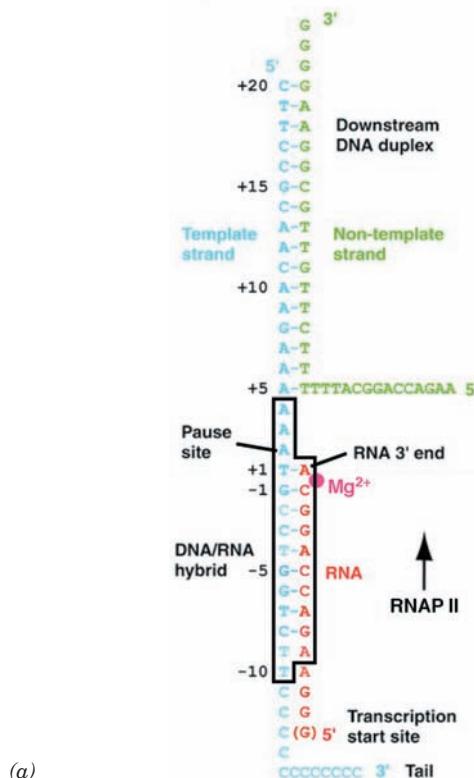
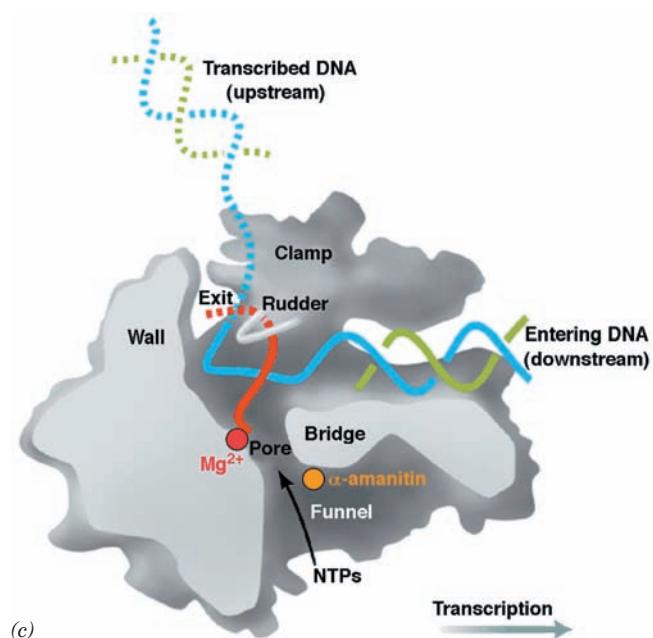
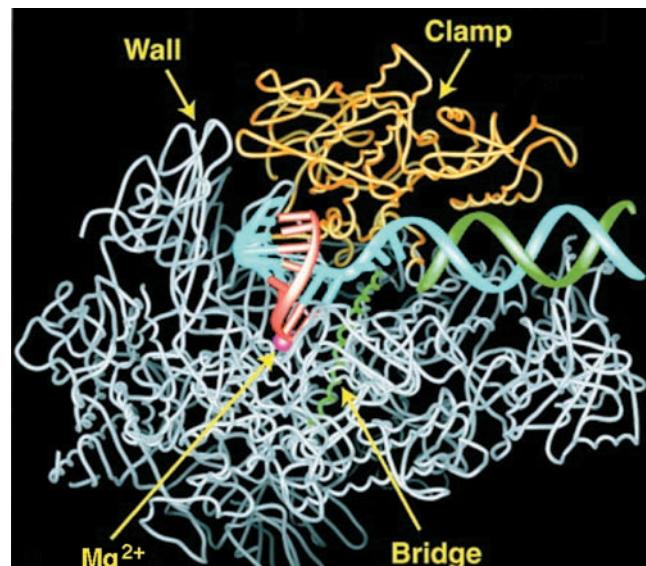


Figure 31-22 X-ray structure of an RNAP II elongation complex. (a) The RNA · DNA complex in the structure with the template DNA cyan, the nontemplate DNA green, and the newly synthesized RNA red. The magenta dot marked Mg^{2+} represents the strongly bound active site metal ion. The black box encloses those portions of the complex that are clearly visible in the structure; the double-stranded portion of the DNA marked “Downstream DNA duplex” is poorly ordered, and the remaining portions of the complex are disordered. (b) View of the transcribing complex from the bottom of Fig. 31-21a in which portions of Rpb2 that form the near side of the cleft have been removed to expose the bound RNA · DNA complex. The protein is represented by its backbone in which the clamp, which is closed over the downstream DNA duplex, is yellow, the bridge helix is green, and the remaining portions of the protein are gray. The DNA and RNA are colored as in Part a with their well-ordered portions drawn in ladder form and their less ordered portions drawn in backbone form. The active site Mg^{2+} ion is represented by a magenta sphere. (c) Cutaway schematic diagram of the transcribing complex in Part b in which the cut surfaces of the protein are light gray, its remaining surfaces are darker gray, and several of its functionally important structural features are labeled. The DNA, RNA, and active site Mg^{2+} ion



are colored as in Part a with portions of the DNA and RNA that are not visible in the X-ray structure represented by dashed lines. The α -amanitin binding site is marked by an orange circle. [Modified from diagrams by Roger Kornberg, Stanford University. PDBid 1I6H.]  See Interactive Exercise 37

the nontemplate strand and the 3' tail of the template strand are disordered in the X-ray structure).

How does RNAP translocate its bound RNA–DNA assembly in preparation for a new round of synthesis? The highly conserved helical segment of Rpb1, dubbed the “bridge” because it bridges the two pincers forming the enzyme’s cleft (Figs. 31-21 and 31-22), nonspecifically contacts

the template DNA base at the $i + 1$ position. Although this helix is straight in all X-ray structures of RNAP II yet determined, it is bent in that of Taq core RNAP. If the bridge helix, in fact, alternates between its straight and bent conformations, it would move by 3 to 4 Å. Kornberg has therefore speculated that translocation occurs through the bending of the bridge helix so as to push the paired nucleotides at posi-

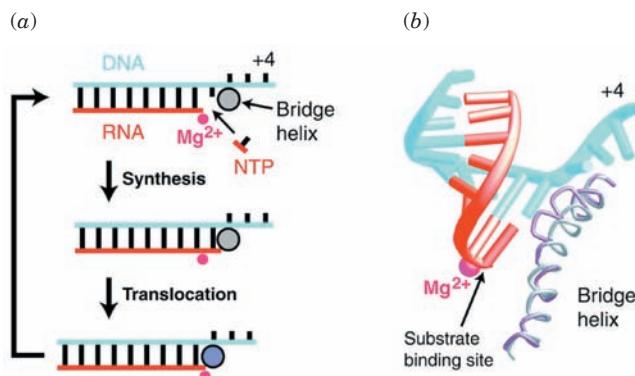


Figure 31-23 Proposed transcription cycle and translocation mechanism of RNAP. (a) The nucleotide addition cycle in which the enzyme active site is marked by its strongly bound Mg²⁺ ion (magenta). The translocation of the transcribing RNA · DNA complex is proposed to be motivated by a conformational change of the bridge helix from straight (gray circle) to bent (violet circle). The relaxation of the bridge helix back to its straight form would complete the cycle by yielding an empty NTP binding site at the active ($i + 1$) site. (b) The RNA · DNA complex in RNAP II viewed and colored as in Fig. 31-22b. The RNAP II bridge helix is gray and the superimposed (and bent) Taq polymerase bridge helix is violet. The side chains extending from the bent helix would sterically clash with the hybrid base pair at position $i + 1$. [Courtesy of Roger Kornberg, Stanford University.]

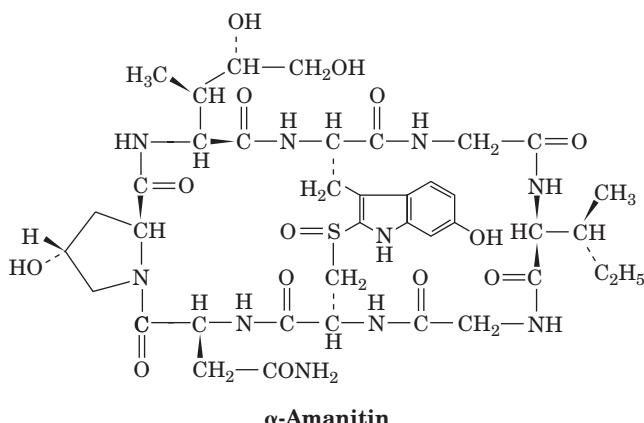
tion $i + 1$ to position $i - 1$ (Fig. 31-23). The recovery of the bridge helix to its straight conformation would then yield an empty site at position $i + 1$ for entry of the next NTP, thereby preparing the enzyme for a new round of nucleotide addition. The reversal of this process is presumably prevented by the binding of the next substrate NTP and hence this mechanism is that of a **Brownian ratchet** [in which otherwise random thermal (Brownian) back-and-forth fluctuations are converted to coherent forward motion by inhibiting the backward motion; Section 12-4Bg].

RNAP II selects its substrate ribonucleotide through a two-stage process. The incoming NTP gains access to the active site through the funnel and pore (secondary channel) diagrammed in Fig. 31-22c. There it first binds to the so-called E (for entry) site (Fig. 31-24), which exhibits no selectivity for the identity of its base. The NTP then pivots to enter the A (for addition) site, which only accepts an NTP that forms a Watson-Crick base pair with the template base in the $i + 1$ position. This process is mediated by the Rbp1 subunit's so-called trigger loop, which swings in beneath the correctly base-paired NTP in the A site to form an extensive hydrogen-bonded network involving both the NTP and other portions of the RNAP, interactions that acutely discriminate against dNTPs.

b. Amatoxins Specifically Inhibit RNA Polymerases II and III

The poisonous mushroom *Amanita phalloides* (death cap), which is responsible for the majority of fatal mushroom poisonings, contains several types of toxic substances,

including a series of unusual bicyclic octapeptides known as **amatoxins**. **α-Amanitin**,



α-Amanitin

which is representative of the amatoxins, forms a tight 1:1 complex with RNAP II ($K = 10^{-8} M$) and a looser one with RNAP III ($K = 10^{-6} M$). Its binding slows an RNAP's rate of RNA synthesis from several thousand to only a few nucleotides per minute. α-Amanitin is therefore a useful tool for mechanistic studies of these enzymes. RNAP I as well as mitochondrial, chloroplast, and bacterial RNAPs are insensitive to α-amanitin.

The X-ray structure of RNAP II in complex with α-amanitin, also determined by Kornberg, reveals that α-amanitin binds in the funnel beneath the protein's bridge helix (Fig. 31-22c), where it interacts with residues of the bridge helix and the trigger loop. The α-amanitin binding site is too far away from the enzyme active site to directly interfere with NTP entry or RNA synthesis, consistent with the observation that α-amanitin does not influence the affinity of RNAP II for NTPs, although it reduces its selectivity. Mutation of Rbp1 His 1085, an invariant member of the trigger loop, to Tyr mimics the effects of α-amanitin. Moreover, this mutation renders RNAP II highly resistant to α-amanitin, in agreement with an X-ray structure indicating that α-amanitin interacts with the side chain of His 1085 so as to lock the trigger loop in a previously unobserved conformation. Evidently, α-amanitin interferes with the conformational change of the trigger loop postulated to promote catalysis (Fig. 31-24), which further supports this mechanism.

Despite the amatoxins' high toxicity (5–6 mg, which occurs in ~40 g of fresh mushrooms, is sufficient to kill a human adult), they act slowly. Death, usually from liver dysfunction, occurs no earlier than several days after mushroom ingestion (and after recovery from the effects of other mushroom toxins). This, in part, reflects the slow turnover of eukaryotic mRNAs and proteins.

c. RNAPs Can Correct Their Mistakes

RNAPs cannot read through a damaged template strand and consequently stall at the damage site. Moreover, if a deoxynucleotide or a mispaired ribonucleotide is mistakenly incorporated into RNA, the DNA–RNA hybrid helix becomes distorted, which also causes the RNAP

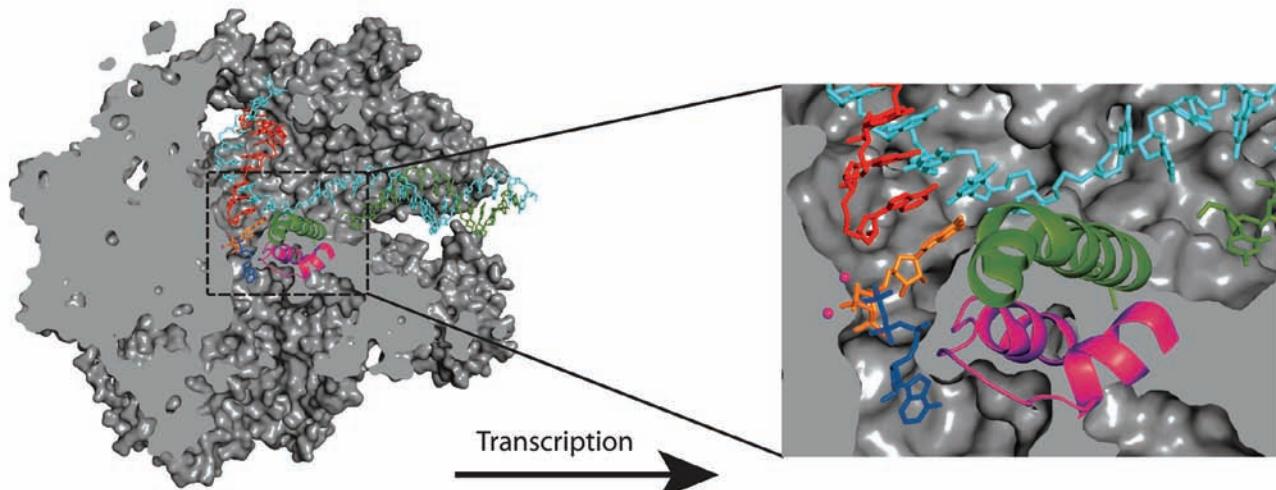


Figure 31-24 The A and E sites and the trigger loop in RNA polymerase II. A cutaway view of the transcribing complex viewed as in Fig. 31-22b. Its bound nucleic acids and nucleotides are colored differently with template DNA cyan, nontemplate DNA green, newly transcribed RNA red, GTP in the A site

orange, and ATP in the E site blue. The trigger loop is magenta, the bridge helix is green, and the two Mg^{2+} ions at the active site are represented by magenta spheres. The RNAP II surface is gray. [Courtesy of Dong Wang and Roger Kornberg, Stanford University. PDBid 2E2H.]

to stall. How, then, do RNAPs avoid accumulating at damaged or mispaired sites, which, if it occurred on an essential gene, would be lethal?

RNAPs do not monotonically move forward along the template DNA. Instead, they frequently backtrack such that the RNA's penultimate nucleotide, which was in the $i - 1$ position, has re-entered the $i + 1$ position and the 3'-nucleotide, now in the $i + 2$ position, enters the secondary channel where it binds in the so-called P (for proofreading) site. If the forward movement of the RNA is impeded by damage to the template or by mispairing, further backtracking becomes favored so that several more ribonucleotides enter the secondary channel. The backtracking of only one or a few nucleotides is reversible. Otherwise, transcription is arrested until the RNA is hydrolytically cleaved at the active site. In *E. coli*, this requires the assistance of the homologous proteins **GreA** and/or **GreB**, whereas with RNAP II, this function is carried out by the unrelated protein **TFIIS**. These proteins induce the RNAP active site to hydrolyze the phosphodiester bond between the ribonucleotides in the $i + 1$ and $i - 1$ positions (a reaction that is not the reverse of the polymerase reaction since this would be pyrophosphorolysis). In this way, RNAP can correct its mistakes and resume RNA synthesis. RNAP I and RNAP III also efficiently correct their mistakes.

Kornberg determined the X-ray structure of RNAP II in complex with TFIIS, a 28-nt template DNA, a 14-nt nontemplate DNA, and a 13-nt RNA that is complementary to 5' end of the template DNA except for the last two residues at the RNA's 3' end, which are mismatched (Fig. 31-25). The RNA-DNA hybrid has backtracked such that these latter residues occupy the $i + 1$ and $i + 2$ positions. The C-terminal domain of TFIIS is bound in the RNAP's funnel with one of its loops insinuated through the pore to interact with the RNAP's active site residues. There it presum-

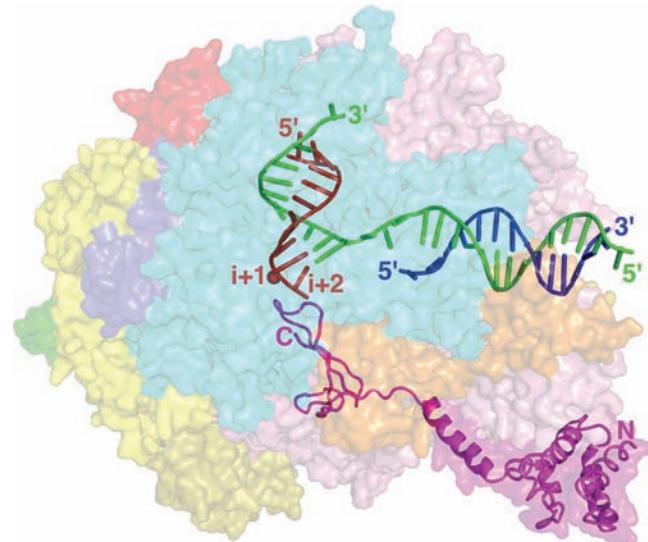


Figure 31-25 X-ray structure of backtracked RNA polymerase II in complex with DNA, RNA, and TFIIS. The RNAP, which is represented by its semitransparent surface diagram, is viewed as in Figs. 31-22b and 31-24 with its subunits colored as in Fig. 31-21a. The DNA and RNA are shown in ladder form with the RNA (13 nt) red, the template DNA (28 nt) green, and the nontemplate DNA (14 nt) blue. The base pairs at the $i + 1$ and $i + 2$ positions of the RNA · DNA hybrid helix are mismatched. An Mg^{2+} ion, which is represented by a red sphere near the $i + 1$ position, marks the RNAP's active site. TFIIS is drawn in cartoon form in magenta. Its C-terminal domain is inserted into the RNAP's funnel (Fig. 31-22c) with a loop occupying the pore, where it is in proximity to the RNAP's active site. [Based on an X-ray structure by Roger Kornberg, Stanford University. PDBid 3GTM.]

ably facilitates the hydrolytic reaction, perhaps by liganding the active site Mg^{2+} ion that normally accompanies the incoming NTP (Section 31-2Ea) and/or positioning a hydrolytic water molecule. The trigger loop in this structure is in the “open” conformation. Interestingly, the cryo-EM-based structure of *E. coli* RNAP in complex with GreB indicates that GreB likewise inserts an extended protein finger in the RNAPs active site via its secondary channel, even though the structures of GreB and TFIIS are unrelated.

Despite the foregoing, transcription is less accurate than DNA replication: RNAPs incorporate one incorrect base for every $\sim 10^4$ transcribed, whereas, for example, *E. coli* Pol I incorporates one incorrect base in $\sim 10^7$ (Section 30-2Ab). Cells can tolerate the former rate because most genes are repeatedly transcribed. In contrast, errors in DNA synthesis alter all the affected gene’s transcripts in the cell in which the error occurred and all of its progeny.

d. Mammalian RNA Polymerase I Has a Bipartite Promoter

Since, as we shall see in Section 31-4B, the numerous rRNA genes in a given eukaryotic cell have essentially identical sequences, its RNAP I only recognizes one promoter. Yet, in contrast to the case for RNAPs II and III, RNAP I promoters are species specific, that is, an RNAP I only recognizes its own promoter and those of closely related species. This is because only closely related species exhibit recognizable sequence identities near the transcriptional start sites of their rRNA genes. RNAP I promoters were therefore identified by determining how the transcription rate of an rRNA gene is affected by a series of increasingly longer deletions approaching its start site from either its upstream or its downstream sides. Such studies have indicated, for example, that mammalian RNAPs I require the presence of a so-called **core promoter element**, which spans positions -31 to $+6$ and hence overlaps the transcribed region. However, efficient transcription additionally requires an **upstream promoter element**, which is located between residues -187 and -107 . These elements, which are G + C-rich and $\sim 85\%$ identical, are bound by specific transcription factors which then recruit RNAP I to the transcription start site.

e. RNA Polymerase II Promoters Are Complex and Diverse

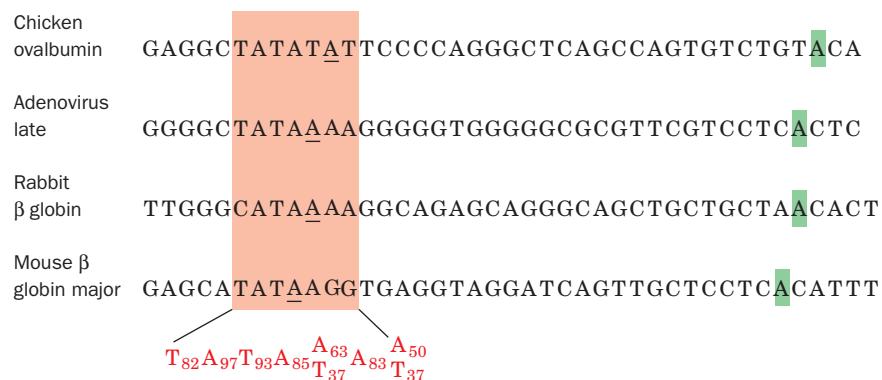
The promoters recognized by RNAP II are considerably longer and more diverse than those of prokaryotic genes but have not yet been fully described. The structural genes expressed in all tissues, the so-called **housekeeping genes**, which are thought to be constitutively transcribed, have one or more copies of the sequence GGGCGG or its complement (the **GC box**) located upstream from their transcription start sites. The analysis of deletion and point mutations in eukaryotic viruses such as SV40 indicates that GC boxes function analogously to prokaryotic promoters. On the other hand, structural genes that are selectively expressed in one or a few types of cells often lack these GC-rich sequences. Rather, *many contain a conserved AT-rich sequence located 25 to 30 bp upstream from their transcription start sites* (Fig. 31-26). Note that this so-called **TATA box** resembles the -10 region of prokaryotic promoters (TATAAT), although they differ in their locations relative to the transcription start site (-27 vs -10). The functions of these two promoter elements are not strictly analogous, however, since the deletion of the TATA box does not necessarily eliminate transcription. Rather, TATA box deletion or mutation generates heterogeneities in the transcriptional start site, thereby indicating that the TATA box participates in selecting this site.

The gene region extending between about -50 and -110 also contains promoter elements. For instance, many eukaryotic structural genes, including those encoding the various globins, have a conserved sequence of consensus CCAAT (the **CCAAT box**) located between about -70 and -90 whose alteration greatly reduces the gene’s transcription rate. Globin genes have, in addition, a conserved **CACCC box** upstream from the CCAAT box that has also been implicated in transcriptional initiation. Evidently, the promoter sequences upstream of the TATA box form the initial DNA-binding sites for RNA polymerase II and the other proteins involved in transcriptional initiation (see below).

f. Enhancers Are Transcriptional Activators That Can Have Variable Positions and Orientations

Perhaps the most surprising aspect of eukaryotic transcriptional control elements is that some of them need not

Figure 31-26 The promoter sequences of selected eukaryotic structural genes. The homologous segment, the TATA box, is shaded in red with the base at position -27 underlined and the initial nucleotide to be transcribed ($+1$) shaded in green. The bottom row indicates the consensus sequence of several such promoters with the subscripts indicating the percent occurrence of the corresponding base. [After Gannon, F., et al., *Nature* **278**, 433 (1978).]



have fixed positions and orientations relative to their corresponding transcribed sequences. For example, the SV40 genome, in which such elements were first discovered, contains two repeated sequences of 72 bp each that are located upstream from the promoter for early gene expression. Transcription is unaffected if one of these repeats is deleted but is nearly eliminated when both are absent. The analysis of a series of SV40 mutants containing only one of these repeats demonstrated that its ability to stimulate transcription from its corresponding promoter is all but independent of its position and orientation. Indeed, transcription is unimpaired when this segment is several thousand base pairs upstream or downstream from the transcription start site. Gene segments with such properties are named **enhancers** to indicate that they differ from promoters, with which they must be associated in order to trigger site-specific and strand-specific transcription initiation (although the characterization of numerous promoters and enhancers indicates that their functional properties are similar). Enhancers occur in both eukaryotic viruses and cellular genes.

Enhancers are required for the full activities of their cognate promoters. It was originally thought that enhancers somehow acted as entry points on DNA for RNAP II (perhaps by altering DNA's local conformation or through a lack of binding affinity for the histones that normally coat eukaryotic DNA; Section 34-1B). However, it is now clear that *enhancers are recognized by specific transcription factors that stimulate RNA polymerase II to bind to the corresponding but distant promoter.* This requires that the DNA between the enhancer and promoter loop around so that the transcription factor can simultaneously contact the enhancer and the RNAP II and/or its associated proteins at the promoter. Most cellular enhancers are associated with genes that are selectively expressed in specific tissues. It therefore seems, as we discuss in Section 34-3B, that *enhancers mediate much of the selective gene expression in eukaryotes.*

g. RNA Polymerase III Promoters Can Be Located Downstream from Their Transcription Start Sites

The promoters of genes transcribed by RNAP III can be located entirely within the genes' transcribed regions. Donald Brown established this through the construction of a series of deletion mutants of a *Xenopus borealis* 5S RNA gene. Deletions of base sequences that start from outside one or the other end of the transcribed portion of the 5S gene only prevent transcription if they extend into the segment between nucleotides +40 and +80. Indeed, a fragment of the 5S RNA gene consisting of only nucleotides 41 to 87, when cloned in a bacterial plasmid, is sufficient to direct specific initiation by RNAP III at an upstream site. This is because, as was subsequently demonstrated, the sequence contains the binding site for transcription factors that stimulate the upstream binding of RNAP III. Further studies have shown, however, that the promoters of other RNAP III-transcribed genes lie entirely upstream of their start sites. These upstream sites also bind transcription factors that recruit RNAP III.

3 CONTROL OF TRANSCRIPTION IN PROKARYOTES

Prokaryotes respond to sudden environmental changes, such as the influx of nutrients, by inducing the synthesis of the appropriate proteins. This process takes only minutes because transcription and translation in prokaryotes are closely coupled: *Ribosomes commence translation near the 5' end of a nascent mRNA soon after it is extruded from RNA polymerase* (Fig. 31-27). Moreover, *most prokaryotic mRNAs are enzymatically degraded within 1 to 3 min of their synthesis*, thereby eliminating the wasteful synthesis of unneeded proteins after a change in conditions (protein degradation is discussed in Section 32-6). In fact, the 5' ends of some mRNAs are degraded before their 3' ends have been synthesized.

In contrast, the induction of new proteins in eukaryotic cells frequently takes hours or days, in part because transcription takes place in the nucleus and the resulting mRNAs must be transported to the cytoplasm, where translation occurs. However, eukaryotic cells, particularly those of multicellular organisms, have relatively stable environments; major changes in their transcription patterns usually occur only during cell differentiation.

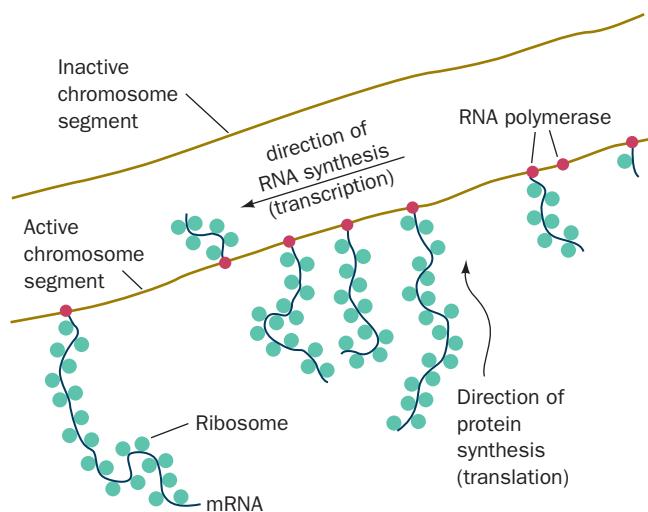
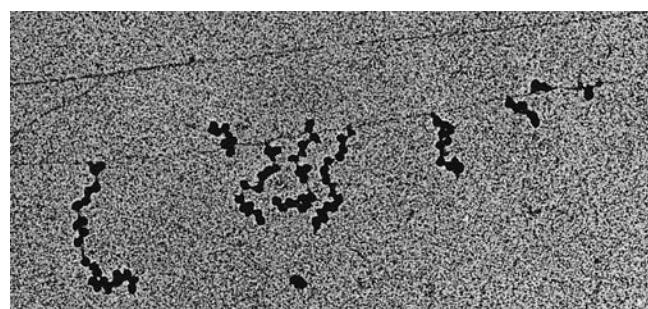


Figure 31-27 An electron micrograph and its interpretive drawing showing the simultaneous transcription and translation of an *E. coli* gene. RNA polymerase molecules are transcribing the DNA from right to left while ribosomes are translating the nascent RNAs (mostly from bottom to top). [Courtesy of Oscar L. Miller, Jr. and Barbara Hamkalo, University of Virginia.]

In this section we examine some of the ways in which prokaryotic gene expression is regulated through transcriptional control. Eukaryotes, being vastly more complex creatures than are prokaryotes, have a correspondingly more complicated transcriptional control system whose general outlines are beginning to come into focus. We therefore defer discussion of eukaryotic transcriptional control until Section 34-3B, where it can be considered in light of what we know about the structure and organization of the eukaryotic chromosome.

A. Promoters

In the presence of high concentrations of inducer, the *lac* operon (Section 31-1Ab) is rapidly transcribed. In contrast, the *lacI* gene is transcribed at such a low rate that a typical *E. coli* cell contains <10 molecules of the *lac* repressor. Yet, the *I* gene has no repressor. Rather, it has such an inefficient promoter (Fig. 31-10) that it is transcribed an average of about once per bacterial generation. *Genes that are transcribed at high rates have efficient promoters.* In general, the more efficient a promoter, the more closely its sequence resembles that of the corresponding consensus sequence.

a. Gene Expression Can Be Controlled by a Succession of σ Factors

The processes of development and differentiation involve the temporally ordered expression of sets of genes according to genetically specified programs. Phage infections are among the simplest examples of developmental processes. Typically, only a subset of the phage genome, often referred to as *early* genes, are expressed in the host immediately after phage infection. As time passes, *middle* genes start to be expressed, and the *early* genes as well as the bacterial genes are turned off. In the final stages of phage infection, the *middle* genes give way to the *late* genes. Of course some phage types express more than three sets of genes and some genes may be expressed in more than one stage of an infection.

One way in which families of genes are sequentially expressed is through “cascades” of σ factors. In the infection of *Bacillus subtilis* by bacteriophage SP01, for example, the *early* gene promoters are recognized by the bacterial RNAP holoenzyme. Among these *early* genes is gene 28, whose gene product is a new σ subunit, designated σ^{sp28} , that displaces the bacterial σ subunit from the core enzyme. The reconstituted holoenzyme recognizes only the phage *middle* gene promoters, which all have similar -35 and -10 regions but bear little resemblance to the corresponding regions of bacterial and phage *early* genes. The *early* genes therefore become inactive once their corresponding mRNAs have been degraded. The phage *middle* genes include genes 33 and 34, which together specify yet another σ factor, $\sigma^{sp33/34}$, which, in turn, permits the transcription of only *late* phage genes.

Most bacteria, including *E. coli* and *B. subtilis*, likewise have several different σ factors (*E. coli* has seven). These are not necessarily utilized in a sequential manner. Rather,

those that differ from the predominant or primary σ factor (σ^{70} in *E. coli*) control the transcription of coordinately expressed groups of special purpose genes, whose promoters are quite different from those recognized by the primary σ factor. For example, in *E. coli*, the alternative σ factor σ^{32} is the master regulator of the heat shock response (Section 9-2C), whereas σ^{54} directs the expression of proteins involved in nitrogen assimilation. Likewise, sporulation in *B. subtilis*, a process in which the bacterial cell is asymmetrically partitioned into two compartments, the **forespore** (which becomes the **spore**, a germline cell from which subsequent progeny arise) and the **mother cell** (which synthesizes the spore's protective cell wall and is eventually discarded), is governed by five σ factors in addition to that of the **vegetative** (nonsporulating) cell: one that is active before cell partition occurs, two that are sequentially active in the forespore, and two that are sequentially active in the mother cell. Cross-regulation of the compartmentalized σ factors permits the forespore and mother cell to tightly coordinate this differentiation process.

B. *lac* Repressor I: Binding

In 1966, Benno Müller-Hill and Walter Gilbert isolated *lac* repressor on the basis of its ability to bind ^{14}C -labeled IPTG (Section 31-1Aa) and demonstrated that it is a protein. This was an exceedingly difficult task because *lac* repressor comprises only $\sim 0.002\%$ of the protein in wild-type *E. coli*. Now, however, *lac* repressor is available in quantity via molecular cloning techniques (Section 5-5G).

a. *lac* Repressor Finds Its Operator by Sliding Along DNA

The *lac* repressor is a tetramer of identical 360-residue subunits, each of which binds one IPTG molecule with a dissociation constant of $K = 10^{-6} \text{ M}$. In the absence of inducer, the repressor tetramer nonspecifically binds duplex DNA with a dissociation constant of $K \approx 10^{-4} \text{ M}$. However, it specifically binds to the *lac* operator with far greater affinity: $K \approx 10^{-13} \text{ M}$. Limited proteolysis of *lac* repressor with trypsin reveals that each subunit consists of two functional domains: Its 58-residue N-terminal peptide binds DNA but not IPTG, whereas the remaining “core tetramer” binds only IPTG.

The observed rate constant for the binding of *lac* repressor to *lac* operator is $k_f \approx 10^{10} \text{ M}^{-1} \text{ s}^{-1}$. This “on” rate is much greater than that calculated for the diffusion-controlled process in solution: $k_f \approx 10^7 \text{ M}^{-1} \text{ s}^{-1}$ for molecules the size of *lac* repressor. Since it is impossible for a reaction to proceed faster than its diffusion-controlled rate, the *lac* repressor must not encounter operator from solution in a random three-dimensional search. Rather, *it appears that lac repressor finds operator by nonspecifically binding to DNA and diffusing along it in a far more efficient one-dimensional search.*

b. *lac* Operator Has a Nearly Palindromic Sequence

The availability of large quantities of *lac* repressor made it possible to characterize the *lac* operator. *E. coli* DNA

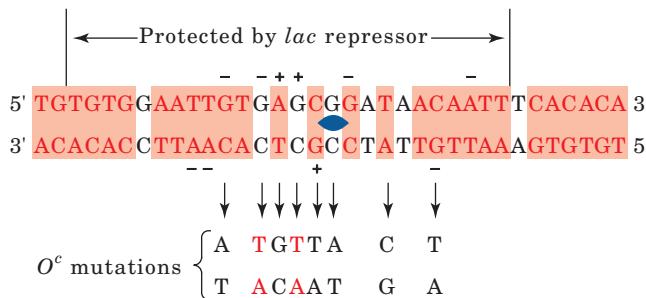


Figure 31-28 The base sequence of the *lac* operator. The symmetry related regions (red) comprise 28 of its 35 bp. A “+” denotes positions at which repressor binding enhances methylation by dimethyl sulfate (which methylates G at N7 and A at N3) and a “-” indicates where this footprinting reaction is inhibited. The bottom row indicates the positions and identities of different point mutations that prevent *lac* repressor binding (O^c mutants). Those in red increase the operator’s symmetry. [After Sobell, H.M., in Goldberger, R.F. (Ed.), *Biological Regulation and Development*, Vol. 1, p. 193, Plenum Press (1979).]

that had been sonicated to small fragments was mixed with *lac* repressor and passed through a nitrocellulose filter. Protein, with or without bound DNA, sticks to nitrocellulose, whereas duplex DNA, by itself, does not. The DNA was released from the filter-bound protein by washing it with IPTG solution, recombined with *lac* repressor, and the resulting complex treated with DNase I. The DNA fragment that *lac* repressor protects from nuclelease degradation consists of a run of 26 bp that is embedded in a nearly 2-fold symmetric sequence of 35 bp (Fig. 31-28, top). Such palindromic symmetry is a common feature of DNA segments that are specifically bound by proteins (recall, for example, that restriction endonuclease recognition sites are also palindromic; Section 5-5Aa).

Palindromic DNA sequences, as we have seen, bind to proteins that have matching 2-fold symmetry. However, methylation protection experiments on the *lac* repressor-

operator system do not fully support this model: There is an asymmetric pattern of differences between free and repressor-bound operator in the susceptibility of its bases to reaction with DMS (Fig. 31-28). Furthermore, point mutations that render it operator constitutive (O^c), and that invariably weaken the binding of repressor to operator, may increase as well as decrease the operator’s 2-fold symmetry (Fig. 31-28).

c. *lac* Repressor Prevents RNA Polymerase from Forming a Productive Initiation Complex

Operator occupies positions -7 through +28 of the *lac* operon relative to the transcription start site (Fig. 31-29). Nuclease protection studies, it will be recalled, indicate that, in the initiation complex, RNA polymerase tightly binds to the DNA between positions -20 and +20 (Section 31-2Aa). Thus, the *lac* operator and promoter sites overlap. It was therefore widely assumed for many years that *lac* repressor simply physically obstructs the binding of RNA polymerase to the *lac* promoter. However, the observation that *lac* repressor and RNA polymerase can simultaneously bind to the *lac* operon indicates that *lac* repressor must act by somehow interfering with the initiation process. Closer investigation of this phenomenon revealed that, in the presence of bound *lac* repressor, RNA polymerase holoenzyme still abortively synthesizes oligonucleotides, although they tend to be shorter than those made in the absence of repressor. Evidently, *lac* repressor acts by somehow increasing the already high kinetic barrier for RNA polymerase to generate the open complex and commence processive elongation.

We discuss the *lac* repressor structure and further aspects of *lac* operator organization in Section 31-3F.

C. Catabolite Repression: An Example of Gene Activation

Glucose is *E. coli*’s metabolite of choice; the availability of adequate amounts of glucose prevents the full expression of >100 genes that encode proteins involved in the fermentation

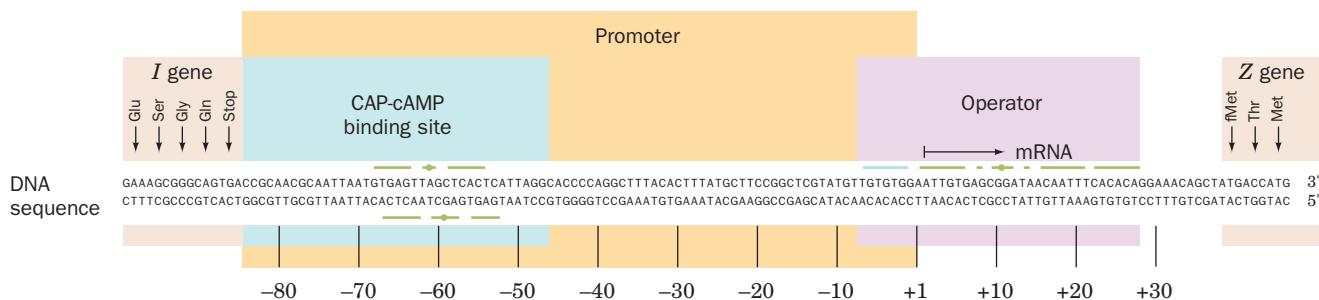


Figure 31-29 The nucleotide sequence of the *E. coli* lac promoter-operator region. The region extends from the C-terminal portion of *lacI* (left) to the N-terminal portion of *lacZ* (right). The palindromic sequences of the operator and

the CAP-binding site (Section 31-3C) are overscored or underscored. [After Dickson, R.C., Abelson, J., Barnes, W.M., and Reznikoff, W.A., *Science* **187**, 32 (1975).]

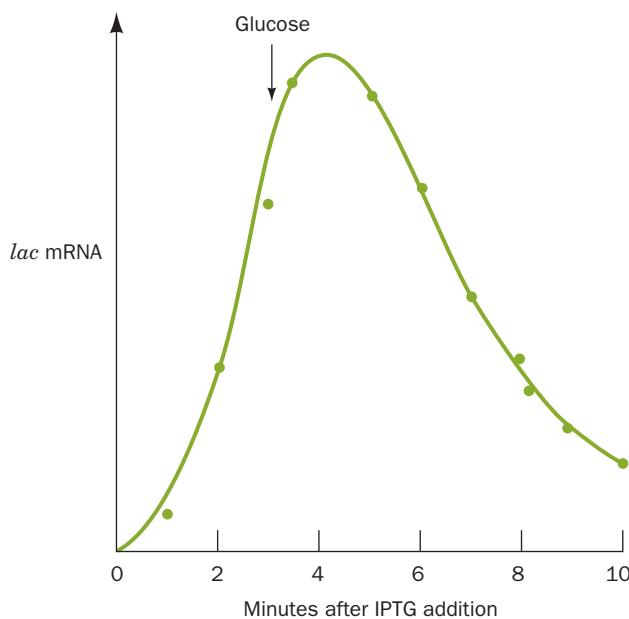


Figure 31-30 The kinetics of *lac* operon mRNA synthesis following its induction with IPTG, and of its degradation after glucose addition. *E. coli* were grown on a medium containing glycerol as their only carbon-energy source and ^3H -labeled uridine. IPTG was added to the medium at the beginning of the experiment to induce the synthesis of the *lac* enzymes. After 3 min, glucose was added to stop the synthesis. The amount of ^3H -labeled *lac* RNA was determined by hybridization with DNA containing the *lacZ* and *lacY* genes. [After Adesnik, M. and Levinthal, C., *Cold Spring Harbor Symp. Quant. Biol.* **35**, 457 (1970).]

of numerous other catabolites, including lactose (Fig. 31-30), arabinose, and galactose, even when these metabolites are present in high concentrations. This phenomenon, which is known as **catabolite repression**, prevents the wasteful duplication of energy-producing enzyme systems.

a. cAMP Signals the Lack of Glucose

The first indication of the mechanism of catabolite repression was the observation that, in *E. coli*, the level of cAMP, which was known to be a second messenger in animal cells (Section 18-3Cb), is greatly diminished in the presence of glucose. This observation led to the finding that the addition of cAMP to *E. coli* cultures overcomes catabolite repression by glucose. Recall that, in *E. coli*, adenylate cyclase is activated by the phosphorylated enzyme $\text{EI}^{\text{A}}\text{glc}$ (or possibly inactivated by dephospho- $\text{EI}^{\text{A}}\text{glc}$), which is dephosphorylated on the transport of glucose across the cell membrane (Section 20-3D). *The presence of glucose, therefore, normally lowers the cAMP level in *E. coli*.*

b. CAP-cAMP Complex Stimulates the Transcription of Catabolite Repressed Operons

Certain *E. coli* mutants, in which the absence of glucose does not relieve catabolite repression, are missing a cAMP-

binding protein that is synonymously named **catabolite gene activator protein (CAP)** and **cAMP receptor protein (CRP)**. CAP is a homodimer of 209-residue subunits that undergoes a large conformational change on binding cAMP. Its function was elucidated by Ira Pastan, who showed that *CAP-cAMP complex, but not CAP itself, binds to the lac operon (among others) and stimulates transcription from its otherwise low-efficiency promoter in the absence of lac repressor*. CAP is therefore a **positive regulator** (turns on transcription), in contrast to *lac* repressor, which is a **negative regulator** (turns off transcription).

The X-ray structure, by Thomas Steitz, of CAP-cAMP in complex with a palindromic 30-bp segment of duplex DNA whose sequence resembles that of the CAP binding sequence (Fig. 31-29) reveals that the DNA is bent by $\sim 90^\circ$ around the protein (Fig. 31-31a). The bend arises from two $\sim 45^\circ$ kinks in the DNA between the fifth and sixth bases out from the complex's 2-fold axis in both directions. This distortion results in the closing of the major groove and an enormous widening of the minor groove at each kink.

Why is the CAP-cAMP complex necessary to stimulate the transcription of its target operons? And how does it do so? The *lac* operon has a weak (low efficiency) promoter; its -10 and -35 sequences (TATGTT and TTTACA; Fig. 31-10) differ significantly from the corresponding consensus sequences of strong (high-efficiency) promoters (TATAAT and TTGACA; Fig. 31-10). Such weak promoters evidently require some sort of help for efficient transcriptional initiation.

Richard Ebright has shown that CAP interacts directly with RNAP via the C-terminal domain of its 85-residue α subunit (αCTD) in a way that stimulates RNAP to initiate transcription from a nearby promoter. The αCTD also binds dsDNA nonspecifically but does so with higher affinity at A + T-rich sites such as those of UP elements (Section 31-2Aa). It is flexibly linked to the rest of the α subunit and hence is not seen in the X-ray structure of Tth RNAP (Fig. 31-11) due to disorder.

Three classes of the over one hundred CAP-dependent promoters have been characterized:

1. Class I promoters, such as that of the *lac* operon, require only CAP-cAMP for transcriptional activation. The CAP binding site on the DNA can be located at various distances from the promoter provided that CAP and RNAP bind to the same face of the DNA helix. Thus, CAP-cAMP activates the transcription of the *lac* operon if its DNA binding site is centered near positions -62 (its wild-type position; Fig. 31-29), -72 , -83 , -93 , or -103 , all of which are one helical turn apart. For the latter sites, this requires that the DNA loop around to permit CAP-cAMP to contact the αCTD . Such looping is likely to be facilitated by the bending of the DNA around CAP-cAMP.

2. Class II promoters also require only CAP-cAMP for transcriptional activation. However, in class II promoters, the CAP binding site only occupies a fixed position that overlaps the RNAP binding site, apparently by replacing the promoter's -35 promoter region. CAP then interacts

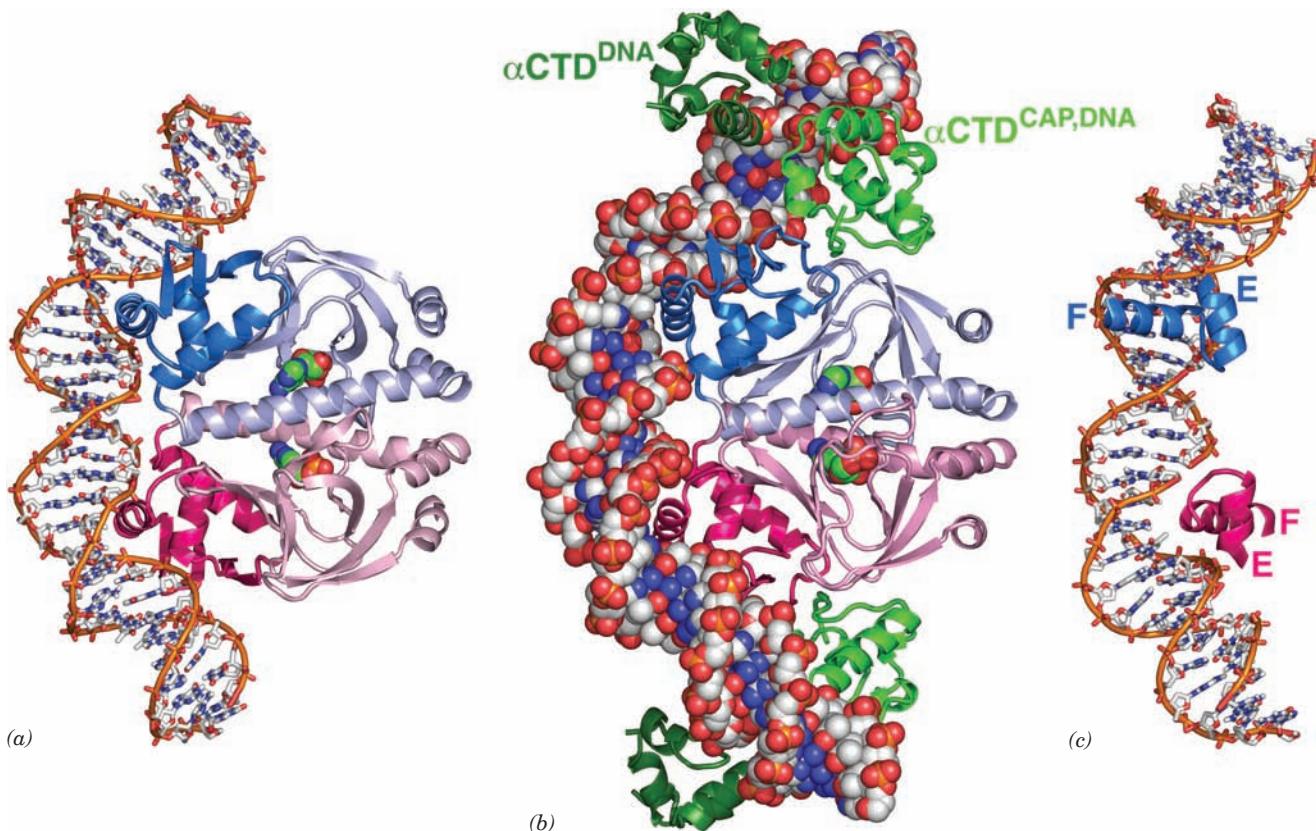


Figure 31-31 X-ray structures of CAP-cAMP-dsDNA complexes. The dsDNA and cAMP in these 2-fold symmetric complexes are colored according to atom type with DNA C white, cAMP green, N blue, O red, and P orange. (a) CAP-cAMP in complex with a palindromic 30-bp self-complementary DNA viewed with its 2-fold axis horizontal. The protein is drawn in ribbon form with its identical subunits pink and blue and with their C-terminal domains in darker shades. The DNA is shown in stick form with successive P atoms in the same strand connected by orange rods and with the cAMP drawn in space-filling form. (b) CAP-cAMP in complex with a 44-bp palindromic DNA and four α CTD subunits. The DNA, CAP, and cAMP are viewed as in

Part a with the DNA drawn in space-filling form. The α CTD subunits are drawn in ribbon form with the α CTD^{CAP,DNA} green and the α CTD^{DNA} dark green. (c) The same structure as in Part a showing the binding of the CAP dimer's two helix-turn-helix (HTH) motifs in successive major grooves of the DNA. The view is rotated 45° about the vertical axis relative to Part a. Note how CAP's F (recognition) helix is inserted into the DNA's major groove, as can also be seen in Parts a and b. [Parts a and c based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1CGP. Part b based on an X-ray structure by Helen Berman and Richard Ebright, Rutgers University. PDBid 1LB2.]

See Interactive Exercise 38

with RNAP via interactions with both the α CTD and the α subunit's N-terminal domain.

3. Class III promoters require multiple activators to maximally stimulate transcription. These may be two or more CAP-cAMP complexes or a CAP-cAMP complex acting in concert with promoter-specific activators as occurs in the *araBAD* operon (Section 31-3E).

The X-ray structure of CAP-cAMP in complex with the *E. coli* α CTD and a 44-bp palindromic DNA containing the 22-bp CAP-cAMP binding site and 5'-AAAAAA-3' at each end, determined by Helen Berman and Ebright, reveals how these components interact (Fig. 31-31b). The 2-fold symmetric CAP-cAMP- α CTD complex contains two differently located pairs of α CTDs. Each member of the pair designated α CTD^{CAP,DNA} binds to both CAP and to the DNA. CAP and

α CTD^{CAP,DNA} interact over a surprisingly small surface area involving only six residues on each protein that mutagenesis experiments had previously implicated. α CTD^{CAP,DNA} also interacts with the minor groove of a 6-bp segment of the DNA (5'-AAAAAG-3') centered 19 bp from the center of the DNA. Each member of the other pair of α CTDs, designated α CTD^{DNA}, interacts with the minor groove of an UP element-like sequence (5'-GAAAAA-3') that is fortuitously present in the DNA but it makes no contacts with other protein molecules. The common portions of the two CAP complexes pictured in Fig. 31-31a,b are closely superimposable, thereby indicating that the conformation of CAP and its interaction with DNA are not significantly altered by its association with the α CTD. Evidently, CAP-cAMP transcriptionally activates RNAP via a simple "adhesive" mechanism that facilitates and/or stabilizes its interaction with the

promoter DNA. The structures of α CTD^{CAPDNA} and α CTD^{CAP} and their interactions with DNA are nearly identical, thereby suggesting that they are representative of the interaction of an α CTD with an UP element.

D. Sequence-Specific Protein-DNA Interactions

Since genetic expression is controlled by proteins such as CAP and *lac* repressor, an important issue in the study of gene regulation is how these proteins recognize their target base sequences on DNA. Sequence-specific DNA-binding proteins generally do not disrupt the base pairs of the duplex DNA to which they bind. Consequently, these proteins can only discriminate among the four base pairs (A · T, T · A, G · C, and C · G) according to the functional groups of these base pairs that project into DNA's major and minor grooves. An inspection of Fig. 5-12 reveals that the groups exposed in the major groove have a greater variation in their types and arrangements than do those that are exposed in the minor groove. Indeed, the positions of the hydrogen bonding acceptors in the major groove vary with both the identity and orientation of the base pair, whereas in the minor groove they are largely sequence independent. Moreover, the \sim 5-Å-wide and \sim 8-Å-deep minor groove of canonical (ideal) B-DNA is too narrow to admit protein structural elements such as an α helix, whereas its \sim 12-Å-wide and \sim 8-Å-deep major groove can do so. Thus, in the absence of major conformational changes to B-DNA, it would be expected that proteins could more readily differentiate base sequences from its major groove than from its minor groove. We shall see below that this is, in fact, the case.

a. The Helix-Turn-Helix Motif Is a Common DNA Recognition Element in Prokaryotes

 See Guided Exploration 30: Transcription factor-DNA interactions The CAP dimer's two symmetrically disposed F helices protrude from the protein surface in such a way that they fit into successive major grooves of B-DNA (Fig. 31-31). CAP's E and F helices form a **helix-turn-helix (HTH) motif** (supersecondary structure) that conformationally resembles analogous HTH motifs in numerous other prokaryotic repressors of known X-ray and NMR structure, including the *lac* repressor, the *E. coli trp* repressor (Section 31-3G), and the **cI** repressors and **Cro proteins** from **bacteriophages λ and 434** (Section 33-3D). HTH motifs are \sim 20-residue polypeptide segments that form two α helices which cross at \sim 120° (Fig. 31-31c). They occur as components of domains that otherwise have widely varying structures, although all of them bind DNA. Note that HTH motifs are structurally stable only when they are components of larger proteins.

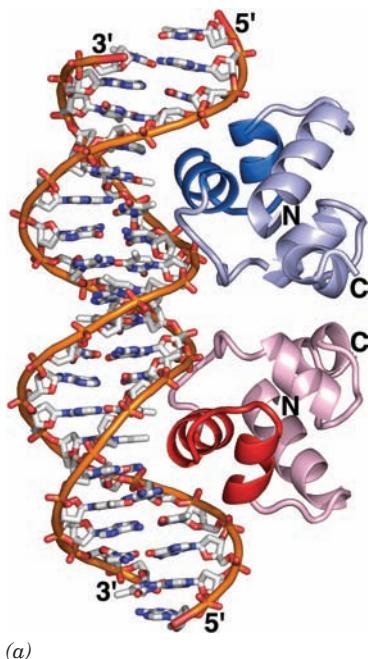
The X-ray and NMR structures of a number of protein-DNA complexes (see below) indicate that *DNA-binding proteins containing an HTH motif associate with their target base pairs mainly via the side chains extending from the second helix of the HTH motif, the so-called recognition helix* (helix F in CAP, E in *trp* repressor, and α 3 in the phage proteins). Indeed, replacing the outward-facing

residues of the 434 repressor's recognition helix with the corresponding residues of the related **bacteriophage P22** yields a hybrid repressor that binds to P22 operators but not to those of 434. Moreover, the HTH motifs in all these proteins have amino acid sequences that are similar to each other and to polypeptide segments in numerous other prokaryotic DNA-binding proteins, including *lac* repressor. Evidently, *these proteins are evolutionarily related and bind their target DNAs in a similar manner*.

How does the recognition helix recognize its target sequence? Since each base pair presents a different and presumably readily differentiated constellation of hydrogen bonding groups in DNA's major groove, it seemed likely that there would be a simple correspondence, analogous to Watson-Crick base pairing, between the amino acid residues of the recognition helix and the bases they contact in forming sequence-specific associations. The above X-ray structures, however, indicate this idea to be incorrect. Rather, base sequence recognition arises from complex structural interactions. For instance:

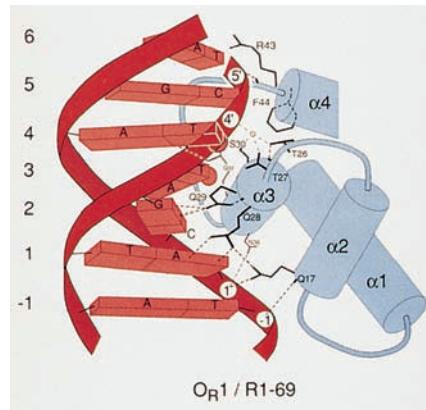
1. The X-ray structures of the 48% identical N-terminal domain of 434 repressor (residues 1–69) and the entire 71-residue 434 Cro protein in their complexes with the identical 20-bp target DNA (the expression of phage 434 is regulated through the differential binding of these proteins to the same DNA segments; Section 33-3Db) were both determined by Stephen Harrison. Both homodimeric proteins, as seen for CAP (Fig. 31-31), associate with the DNA in a 2-fold symmetric manner with their recognition helices bound in successive turns of the DNA's major groove (Figs. 31-32 and 31-33). In both complexes, the protein closely conforms to the DNA surface and interacts with its paired bases and sugar-phosphate chains through elaborate networks of hydrogen bonds, salt bridges, and van der Waals contacts. Nevertheless, the detailed geometries of these associations are significantly different. In the repressor-DNA complex (Fig. 31-32), the DNA bends around the protein in an arc of radius \sim 65 Å which compresses the minor groove by \sim 2.5 Å near its center (between the two protein monomers) and widens it by \sim 2.5 Å toward its ends. In contrast, the DNA in complex with Cro (Fig. 31-33), although also bent, is nearly straight at its center and has a less compressed minor groove (compare Figs. 31-32a and 31-33a). This explains why the simultaneous replacement of three residues in the repressor's recognition helix with those occurring in Cro does not cause the resulting hybrid protein to bind DNA with Cro-like affinity: *The different conformations of the DNA in the repressor and Cro complexes prevents any particular side chain from interacting identically with the DNA in the two complexes*.

2. Paul Sigler determined the X-ray structure of *E. coli trp* repressor in complex with a DNA containing an 18-bp palindrome (TGTACTAGTTAACTAGTAC, where the *trp* repressor's target sequence is underlined) that closely resembles the *trp* operator (Section 31-3G). The dimeric protein's recognition helices bind, as expected, in successive major grooves of the DNA, each in contact with an



(a)

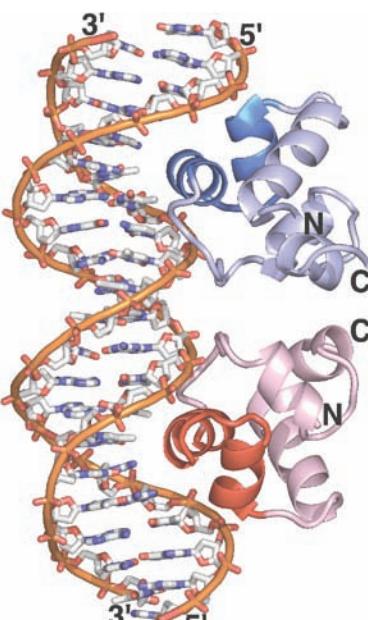
Figure 31-32 X-ray structure of the 69-residue N-terminal domain of 434 phage repressor in complex with a 20-bp dsDNA containing its target sequence. One strand of the DNA has the sequence d(TATACAAGAAAGTTTGTACT). (a) The complex viewed with the homodimeric protein's 2-fold axis horizontal. The protein is drawn in ribbon form with one of its two identical subunits blue and the other red and with their helix-turn-helix (HTH) motifs in darker shades. The DNA is drawn in stick form with C white, N blue, O red, and P orange, and with successive P atoms in the same chain connected by orange rods. (b) A



(b)

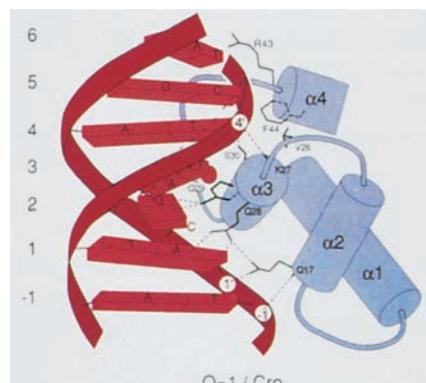
schematic drawing indicating how the HTH motif, which encompasses helices α_2 and α_3 , interacts with its target DNA. Short bars emanating from the polypeptide chain represent peptide NH groups, hydrogen bonds are represented by dashed lines, and DNA phosphates are represented by numbered circles. The small circle is a water molecule. [Part a based on an X-ray structure by and Part b courtesy of Aneel Aggarwal, John Anderson, and Stephen Harrison, Harvard University. PDBid 2OR1.] See Interactive Exercise 39 and Kinemage

Exercise 18-1



(a)

Figure 31-33 X-ray structure of the 71-residue 434 Cro protein in complex with the same 20-bp DNA shown in Fig. 31-32. Parts a and b correspond to those in Fig. 31-32. Note the close but not identical correspondence between the two structures and, in particular, the difference in the widths of the



(b)

minor groove between the two subunits in each structure. [Part a based on an X-ray structure by and Part b courtesy of Alfonso Mondragón, Cynthia Wolberger, and Stephen Harrison, Harvard University. PDBid 3CRO.]

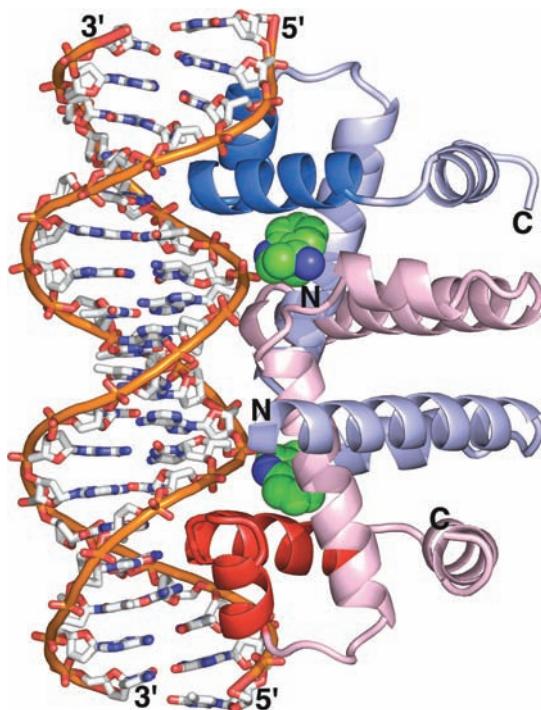


Figure 31-34 X-ray structure of an *E. coli* *trp* repressor-operator-tryptophan complex. The complex is viewed with its molecular 2-fold axis horizontal. The protein's two identical subunits are drawn in ribbon form colored pink and blue with their HTH motifs (helices D and E) more deeply colored. The 18-bp-containing self-complementary dsDNA is shown in stick form with C white, N blue, O red, P orange, and with successive P atoms in the same chain connected by orange rods. The *trp* repressor binds its operator only when L-tryptophan, drawn in space-filling form with C green, is simultaneously bound. Note that the protein's recognition helices (E) bind, as expected, in successive major grooves of the DNA but extend approximately perpendicular to the DNA duplex axis. In contrast, the recognition helices of 434 repressor and Cro proteins are nearly parallel to the major grooves of their bound DNAs (Figs. 31-32 and 31-33), whereas those of CAP assume an intermediate orientation (Fig. 31-31). [Based on an X-ray structure by Paul Sigler, Yale University. PDBid 1TRO.]

See Interactive Exercise 40

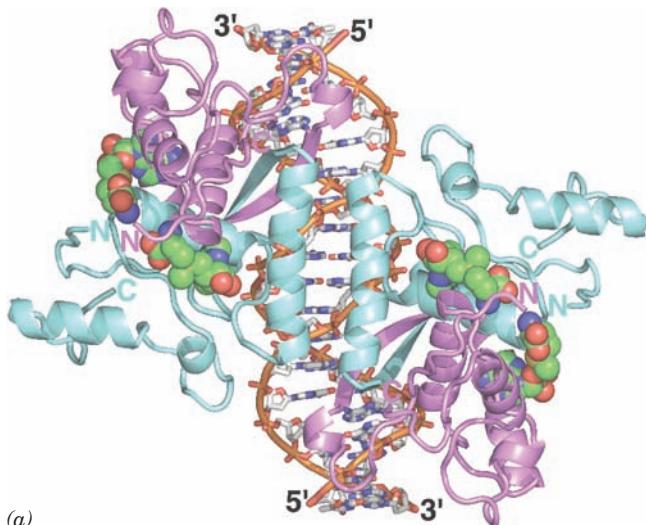
operator half-site (ACTAGT; Fig. 31-34). There are numerous hydrogen bonding contacts between the *trp* repressor and its bound DNA's nonesterified phosphate oxygens. Astoundingly, however, *there are no direct hydrogen bonds or nonpolar contacts that can explain the repressor's specificity for its operator*. Rather, *all but one of the side chain-base hydrogen bonding interactions are mediated by bridging water molecules* (the one direct interaction involves a base that can be mutated without greatly affecting repressor binding affinity). Such buried water molecules have therefore been described as "honorary" protein side chains. In addition, the operator contains several base pairs that are not in contact with the repressor but whose mutation nevertheless greatly decreases repressor binding affinity. This suggests that the operator assumes a sequence-specific conformation that makes favorable contacts with the repressor. Indeed, comparison of the X-ray structure of an uncomplexed 10-bp self-complementary DNA containing the *trp* operator's half-site (CCACTAGTGG) with that of the DNA in the *trp* repressor-operator complex reveals that the ACTAGT half-site assumes nearly identical idiosyncratic conformations and patterns of hydration in both structures. However, the B-DNA helix, which is straight in the DNA 10-mer, is bent by 15° toward the major groove in each operator half-site of the repressor-operator complex. Other DNA sequences could conceivably assume the repressor-bound operator's conformation but at too high an energy cost to form a stable complex with repressor (*trp* repressor's measured 10⁴-fold preference for its operator over other DNAs implies an ~23 kJ · mol⁻¹ difference in their binding free energies). This phenomenon, in which a protein senses the base

sequence of DNA through the DNA's backbone conformation and/or flexibility, is referred to as **indirect readout**. 434 repressor apparently also employs indirect readout: Replacing the central A · T base pair of the operator shown in Fig. 31-32 with G · C reduces repressor binding affinity by 50-fold even though 434 repressor does not contact this region of the DNA.

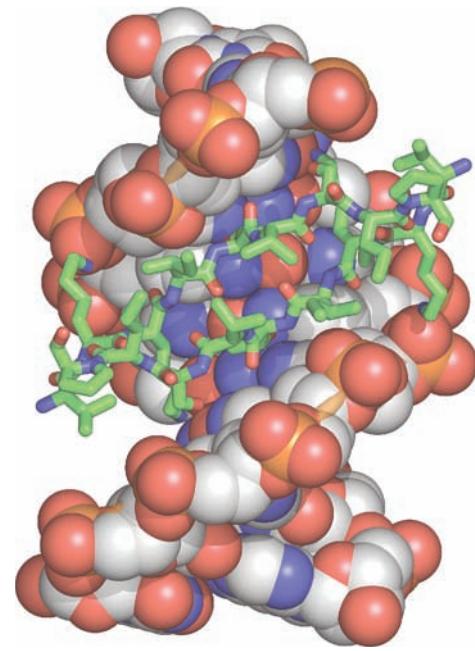
It therefore appears that *there are no simple rules governing how particular amino acid residues interact with bases. Rather, sequence specificity results from an ensemble of mutually favorable interactions between a protein and its target DNA*.

b. The *met* Repressor Contains a Two-Stranded Antiparallel β Sheet That Binds in Its Target DNA's Major Groove

The *E. coli* *met* repressor (**MetJ**), when complexed with S-adenosylmethionine (SAM; Fig. 26-18), represses the transcription of its own gene and those encoding enzymes involved in the synthesis of methionine (Fig. 26-60) and SAM. The X-ray structure of the *met* repressor-SAM-operator complex (Fig. 31-35), determined by Simon Phillips, reveals a symmetric dimer of intertwined homodimers that lacks an HTH motif. Rather, *met* repressor binds to its palindromic target DNA sequence through two symmetry-related pairs of symmetrical two-stranded antiparallel β sheets (called **β ribbons**) that are inserted in successive major grooves of the DNA. Each β ribbon makes sequence-specific contacts with its target DNA sequence via hydrogen bonding and, probably, indirect readout.



(a)



(b)

Figure 31-35 X-ray structure of the *E. coli* *met* repressor-SAM-operator complex. (a) The overall structure of the complex as viewed along its 2-fold axis of symmetry. The self-complementary 18-bp DNA is drawn in stick form, and SAM, which must be bound to the repressor for it to also bind DNA, is shown in space-filling form with the DNA C white, SAM C green, N blue, O red, P orange, and S yellow. The DNA binds four identical 104-residue repressor subunits. Pairs of subunits (light cyan and lavender) form symmetric dimers in which each subunit donates one strand of the 2-stranded antiparallel β ribbon that is inserted in the DNA's major groove

(upper left and lower right). Two such dimers pair across the complex's 2-fold axis via their antiparallel N-terminal helices, which contact one another over the DNA's minor groove. (b) Detailed view of the lower half of Part a showing the 2-stranded antiparallel β ribbon (residues 21-29) inserted into the DNA's major groove, as viewed along its local 2-fold axis (rotated relative to Part a by 50° about the vertical axis). The DNA is shown in space-filling form and the polypeptide chains are drawn in stick form with C green. [Based on an X-ray structure by Simon Phillips, University of Leeds, U.K. PDBid 1CMA.]

 See Interactive Exercise 41

Phillips first determined the X-ray structure of *met* repressor in the absence of DNA. Model building studies aimed at elucidating how *met* repressor binds to its palindromic target DNA assumed that the 2-fold rotation axes of both molecules would be coincident, as they are in all prokaryotic protein-DNA complexes of known structure. There were, consequently, two reasonable choices: (1) The protein could dock to the DNA with the above pairs of β ribbons entering successive major grooves; or (2) a symmetry-related pair of protruding α helices on the opposite face of the protein could do so in a manner resembling the way in which the recognition helices of HTH motifs interact with DNA. A variety of structural criteria suggested that the α helices make significantly better contacts with the DNA than do the β ribbons. Thus, the observation that it is, in fact, the β ribbons that bind to the DNA provides an important lesson: *The results of model building studies must be treated with utmost caution.* This is because our imprecise understanding of the energetics of intermolecular interactions (Sections 8-4 and 29-2) prevents us from reliably predicting how associating macromolecules conform to one another. In the case of the *met* repressor, unpredicted mutual structural accommodations of the protein and DNA yielded a significantly more extensive interface than

had been predicted by simply docking the uncomplexed Met repressor to canonical B-DNA.

The numerous prokaryotic transcriptional regulators of known structure either contain an HTH motif or pairs of β ribbons like the *met* repressor (although numerous prokaryotic DNA-binding proteins, including CAP, contain an elaboration of the HTH motif known as the **winged helix** motif in which two protein loops, one of which contacts the DNA's minor groove, flank the HTH recognition helix like the wings of a butterfly). Moreover, most of these proteins are homodimers that bind to palindromic or pseudopalindromic DNA target sequences. However, eukaryotic transcription factors, as we shall see in Section 34-3B, employ a much wider variety of structural motifs to bind their target DNAs, many of which lack symmetry.

E. *araBAD* Operon: Positive and Negative Control by the Same Protein

Humans neither metabolize nor intestinally absorb the plant sugar L-arabinose. Hence, the *E. coli* that normally inhabit the human gut are periodically presented with a banquet of this pentose. Three of the five *E. coli* enzymes that

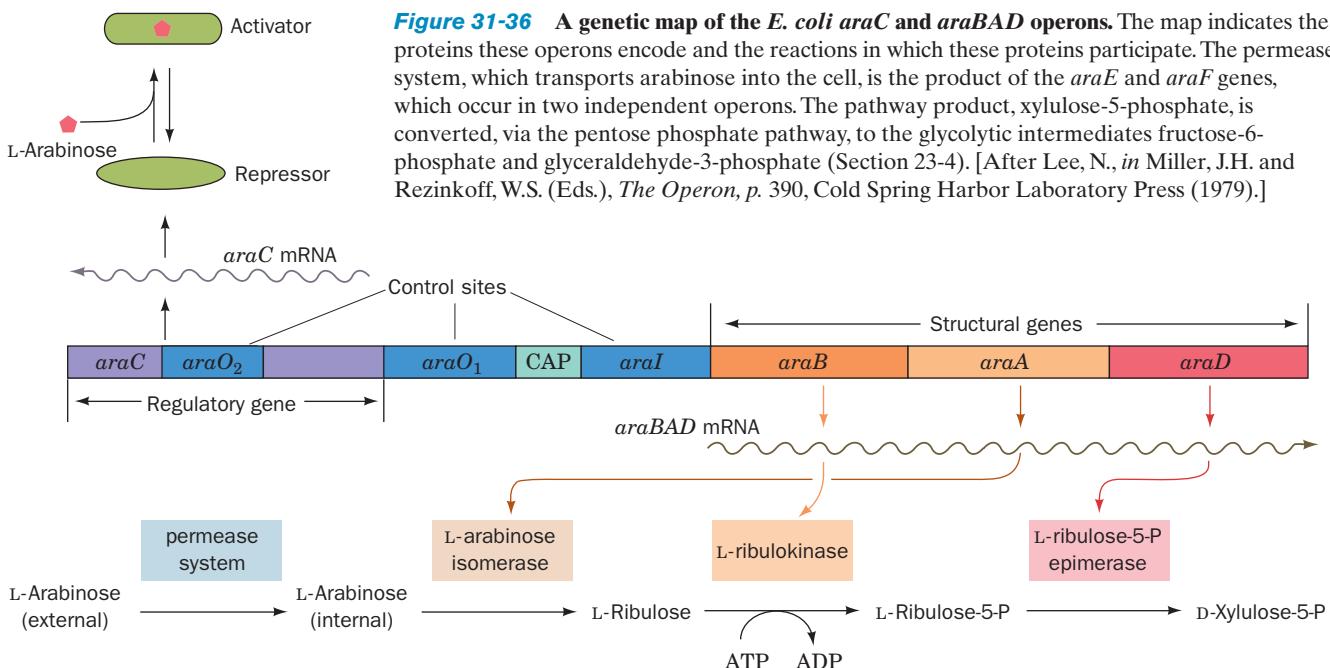


Figure 31-36 A genetic map of the *E. coli* *araC* and *araBAD* operons. The map indicates the proteins these operons encode and the reactions in which these proteins participate. The permease system, which transports arabinose into the cell, is the product of the *araE* and *araF* genes, which occur in two independent operons. The pathway product, xylulose-5-phosphate, is converted, via the pentose phosphate pathway, to the glycolytic intermediates fructose-6-phosphate and glyceraldehyde-3-phosphate (Section 23-4). [After Lee, N., in Miller, J.H. and Reznikoff, W.S. (Eds.), *The Operon*, p. 390, Cold Spring Harbor Laboratory Press (1979).]

metabolize arabinose are products of the catabolite repressible ***araBAD* operon** (Fig. 31-36).

The *araBAD* operon, as Robert Schleif has shown, contains, moving upstream from its transcriptional start site, the *araI*, *araO₁*, and *araO₂* control sites (Fig. 31-37a). The *araI* site (*I* for inducer) consists of two closely similar 17-bp half-sites, *araI₁* and *araI₂*, that are direct repeats separated by 4 bp and are oriented such that *araI₂*, which overlaps the -35 region of the *araBAD* promoter, is downstream of *araI₁*. Likewise, *araO₁* consists of two directly repeating half-sites, *O_{1L}* and *O_{1R}*. Intriguingly, however, *araO₂* consists of a single half-site that is located in a noncoding upstream region of the *araC* gene (see below), at position -270 relative to the *araBAD* start site.

The transcription of the *araBAD* operon is regulated by both CAP-cAMP and the arabinose-binding protein **AraC**. Each 292-subunit of the homodimeric AraC consists of an N-terminal, arabinose-binding, dimerization domain (residues 1–170) connected via a flexible linker to a C-terminal DNA-binding domain (residues 178–292). Regulation of the *araBAD* operon occurs as follows (Fig. 31-37):

1. In the absence of AraC, RNA polymerase initiates transcription of the *araC* gene in the direction away from its upstream neighbor, *araBAD*. The *araBAD* operon is expressed at a low basal level.

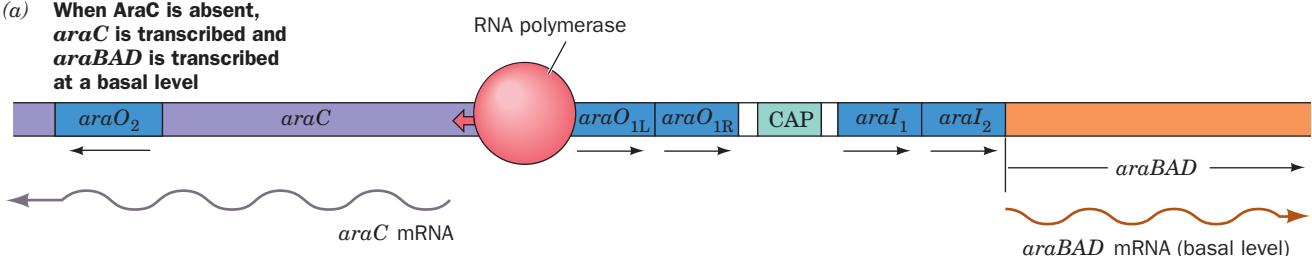
2. When AraC is present, but neither arabinose nor CAP-cAMP (high glucose), AraC binds to *araO₂* and *araI₁* via two HTH motifs in each of its subunits. The binding of AraC to *araI₁* prevents RNAP from initiating transcription of the *araBAD* operon (negative control). A series of deletion mutations indicate that the presence of *araO₂* is also required for the repression of *araBAD*. The remarkably large 210-bp separation between *araO₂* and *araI₁* therefore

strongly suggests that the DNA between them is looped such that a dimeric molecule of AraC protein simultaneously binds to both *araO₂* and *araI₁*. This is corroborated by the observation that the level of repression is greatly diminished by the insertion of 5 bp (half a turn) of DNA between these two sites, thereby transferring *araO₂* to the opposite face of the DNA helix relative to *araI₁* in the putative loop. Yet, the insertion of 11 bp (one turn) of DNA has no such effect. Moreover, looping does not readily occur unless the DNA is supercoiled, which presumably drives the looping process. The AraC dimer also binds to *araO₁*, the operator of the *araC* gene, so as to block the transcription of *araC* but only at high concentrations. Thus, it is likely that DNA looping itself represses the transcription of *araC*. In either case, the expression of *araC* is autoregulatory.

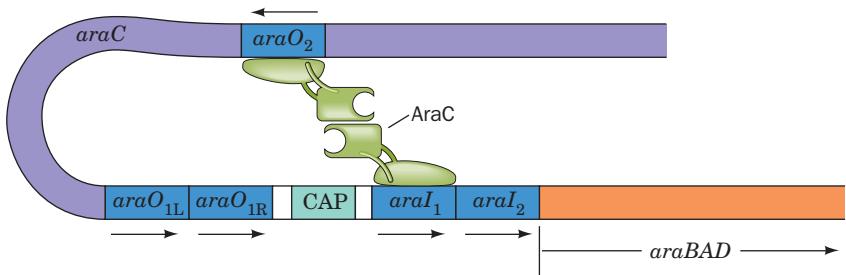
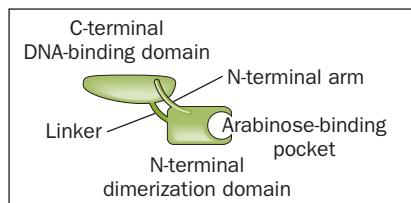
3. When arabinose is present, it allosterically induces the AraC subunit bound to *araO₂* to instead bind to *araI₂*. This activates RNAP to transcribe the *araBAD* genes (positive control). When the cAMP level is high (low glucose), CAP-cAMP, whose presence is required to achieve the maximum level of transcriptional activation, binds to a site between *araO₁* and *araI₁*, where it functions to help break the loop between *araO₂* and *araI₁* and hence to increase the affinity of AraC for *araI₂*. The orientation of *araO₁* with respect to *araC* is opposite to that of *araI* with respect to *araBAD*, and hence the binding of AraC–arabinose at *araO₁* blocks RNAP binding at the *araC* promoter, that is, it represses the expression of AraC.

If the *araI₂* subsite is mutated so as to increase AraC's affinity for it, arabinose is no longer required for transcriptional activation. This suggests that arabinose does not conformationally transform AraC to an activator but, rather, weakens its binding affinity for *araO₂*. If the *araI* site is

- (a) When AraC is absent, *araC* is transcribed and *araBAD* is transcribed at a basal level



- (b) When cAMP and L-arabinose are low, AraC represses *araBAD* transcription



- (c) When cAMP and L-arabinose are abundant, *araBAD* transcription is activated

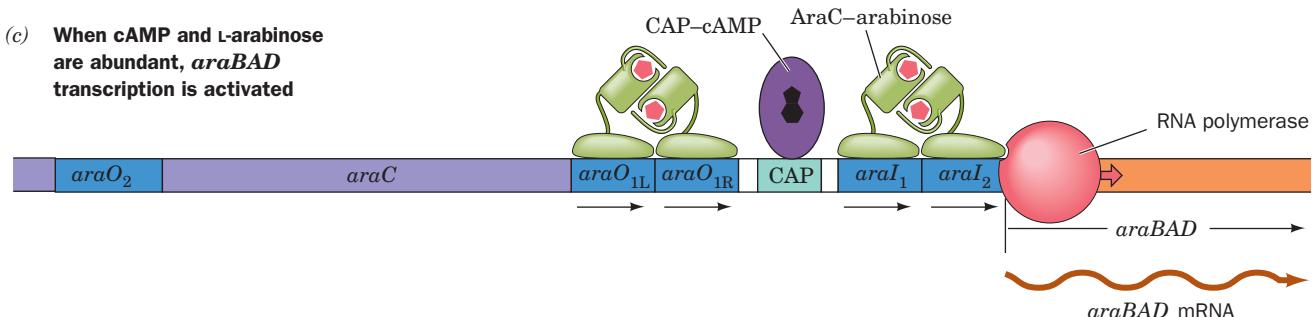


Figure 31-37 The mechanism of *araBAD* regulation. (a) In the absence of AraC, RNAP initiates the transcription of *araC*. *araBAD* is also expressed but at a low basal level. (b) When AraC is present, but not L-arabinose or cAMP, AraC links together *araO*₂ and *araI*₁ to form a DNA loop, thereby repressing both *araC* and *araBAD*. (c) When AraC and L-arabinose are

both present and cAMP is abundant, the resulting AraC–arabinose complex releases *araO*₂ and instead binds *araI*₂, thereby activating *araBAD* transcription. This process is facilitated by the binding of CAP–cAMP. *araC* is repressed by the binding of AraC–arabinose to *araO*₁.

turned around or if it is moved upstream so that *araI*₂ does not overlap the *araBAD* promoter, AraC cannot stimulate transcription. Evidently, AraC activates RNAP through specific and relatively inflexible protein–protein interactions.

The X-ray structures of the N-terminal domain of AraC (residues 2–178), in both the presence and the absence of arabinose, were determined by Schleif and Cynthia Wolberger. In the presence of arabinose, this domain consists of an 8-stranded β barrel followed by two antiparallel α helices (Fig. 31-38). Two such domains associate via an antiparallel coiled coil between each of their C-terminal helices to form the protein’s dimerization interface. An arabinose molecule binds in a pocket of each β barrel via a network of direct and water-mediated hydrogen bonds with side chains that line the pocket. Residues 7 to 18 of the N-terminal arm lie across the mouth of the sugar-binding pocket (residues 2–6 are disordered), thereby fully enclosing the arabinose. The structure of the N-terminal domain in the absence of arabinose is largely superimposable on that in the complex with arabinose, with the exception that

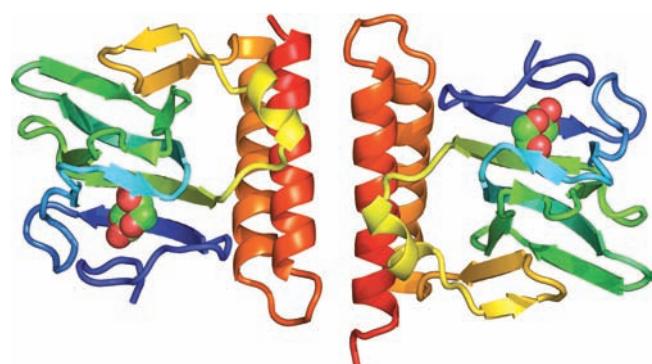


Figure 31-38 X-ray structure of *E. coli* AraC in complex with L-arabinose. The homodimeric protein is viewed along its 2-fold axis with each of its subunits colored in rainbow order from N-terminus (blue) to C-terminus (red). The arabinose is drawn in space-filling form with C green and O red. [Based on an X-ray structure by Robert Schleif and Cynthia Wolberger, Johns Hopkins University. PDBid 2ARC.]

the N-terminal arm is disordered, a not unexpected observation considering that it interacts with bound arabinose via a series of hydrogen bonds.

How does arabinose binding induce the AraC subunit bound at *araO*₂ to instead bind to *araI*₂? Several lines of evidence indicate that, in the absence of arabinose, AraC's N-terminal arm binds to its DNA-binding domain in a way that favors loop formation: (1) the deletion of the N-terminal arm beyond its sixth residue makes AraC act as if arabinose is present; (2) mutations to surface residues on the DNA-binding domain that presumably eliminate its binding of the N-terminal arm also constitutively activate AraC; and (3) mutations in the DNA-binding domain that weaken the binding of arabinose to the protein, presumably by strengthening the binding of the N-terminal arm, can be suppressed by a second mutation in the N-terminal arm or by the deletion of its five N-terminal residues. Evidently, *the binding of the N-terminal arms to the DNA-binding domains in the absence of arabinose rigidifies the AraC dimer such that it cannot simultaneously bind to the directly repeated *araI*₁ and *araI*₂ and hence induce the transcription of *araBAD**. This is corroborated by the observations that (1) joining two AraC DNA-binding domains by flexible polypeptide linkers yields proteins that behave like AraC in the presence of arabinose, and (2) a construct consisting of two double-stranded *araI*₁ half-sites flexibly connected by a 24-nt segment of ssDNA binds wild-type AraC with an affinity that is unaffected by arabinose.

F. *lac* Repressor II: Structure

Here we continue our discussions of the *lac* repressor, but now in terms of the concepts learned in Sections 31-3C–E.

a. Loop Formation Is Important in the Expression of the *lac* Operon

DNA loop formation, which is now known to occur in numerous bacterial and eukaryotic systems, apparently permits several regulatory proteins and/or regulatory sites on one protein to simultaneously influence transcription initiation by RNAP. In fact, *the lac repressor has three binding sites on the lac operon*: the primary operator (Fig. 31-28), now known as *O*₁, and two so-called pseudo-operators (previously thought to be nonfunctional evolutionary fossils), *O*₂ and *O*₃, which are located 401 bp downstream and 92 bp upstream of *O*₁ (within the *lacZ* gene and overlapping the CAP binding site, respectively). Müller-Hill determined the relative contributions of these various operators to the repression of the *lac* operon through the construction of a set of eight plasmids: Each contained the *lacZ* gene under the control of the natural *lac* promoter as well as the three *lac* operators (*O*₁, *O*₂, and *O*₃), which were either active or mutagenically inactive in all possible combinations. When all three operators are active, *lacZ* expression is repressed 1300-fold relative to when all three operators are inactive. The inactivation of only *O*₁ results in almost complete loss of repression whereas the inactivation of only *O*₂ or *O*₃ causes only a ~2-fold loss in repression. However, when *O*₂ and *O*₃ are both inactive, repression is decreased ~70-fold. These results suggest that efficient repression requires the formation of a DNA loop between *O*₁ and either *O*₂ or *O*₃. Indeed, such loop formation, and/or the cooperativity of repressor binding arising from it, appears to be a greater contributor to repression than repressor binding to *O*₁ alone, which provides only 19-fold repression.

b. The *lac* Repressor Is a Dimer of Dimers

Ponzy Lu and Mitchell Lewis determined the X-ray structures of the *lac* repressor alone, in its complex with IPTG, and in its complex with a 21-bp duplex DNA segment whose sequence is a palindrome of the left half of *O*₁ (Fig. 31-28). Each repressor subunit consists of five functional units (Fig. 31-39): (1) an N-terminal DNA-binding domain (residues 1–49) which is known as the “headpiece” because it is readily proteolytically cleaved away from the remaining still tetrameric “core” protein; (2) a hinge helix (residues 50–58) that also binds to the DNA; (3 and 4) a sugar-binding domain (residues 62–333) that is divided into an N-subdomain and a C-subdomain; and (5) a C-terminal tetramerization helix (residues 340–360).

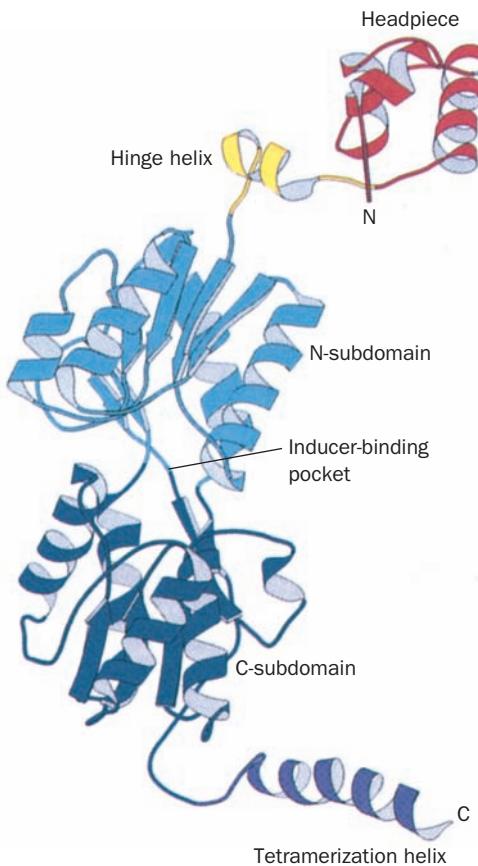


Figure 31-39 X-ray structure of the *lac* repressor subunit. The DNA-binding domain (the headpiece), which contains an HTH motif, is red, the DNA-binding hinge helix is yellow, the N-subdomain of the sugar-binding domain is light blue, its C-subdomain is dark blue, and the tetramerization helix is purple. [Courtesy of Ponzy Lu and Mitchell Lewis, University of Pennsylvania. PDBid 1LBI.]

The *lac* repressor has an unusual quaternary structure (Fig. 31-40a). Whereas nearly all homotetrameric nonmembrane proteins of known structure have D_2 symmetry (three mutually perpendicular 2-fold axes; Fig. 8-65b), *lac* repressor is a V-shaped protein that has only 2-fold symmetry. Each leg of the V consists of a locally symmetric dimer of closely associated repressor subunits. Two such dimers

associate rather tenuously, but with 2-fold symmetry, at the base (point) of the V to form a dimer of dimers.

In the structures of *lac* repressor alone and that of its IPTG complex, the DNA-binding domain is not visible, apparently because the hinge region that loosely tethers it to the rest of the protein is disordered. However, in the DNA complex, in which one DNA duplex binds to each of the two dimers forming the repressor tetramer, the DNA domain forms a compact globule containing three helices, the first two of which form a helix-turn-helix (HTH) motif. The two DNA-binding domains extending from each repressor dimer (at the top of each leg of the V) bind in successive major grooves of a DNA molecule via their HTH motifs, much as is seen, for example, in the complexes of 434 phage repressor and *trp* repressor with their target DNAs (Figs. 31-32 and 31-34). The binding of the *lac* repressor distorts the operator DNA such that it bends away from the DNA-binding domain with an ~ 60 Å radius of curvature due to an $\sim 45^\circ$ kink at the center of the operator that widens the DNA's minor groove to over 11 Å and reduces its depth to less than 1 Å. These distortions permit the now ordered hinge helix to bind in the minor groove so as to contact the identically bound hinge helix from the other subunit of the same dimer. NMR structures by Robert Kaptein and Rolf Boelens reveal that the DNA-binding domain, when cleaved from the repressor, binds to the *lac* operator without distorting the DNA, but that the DNA-binding domain together with the hinge helix forms a complex with the *lac* operator in which the hinge helix binds in the DNA's distorted minor groove (Fig. 31-40b) as in the X-ray structure. Thus, the binding of the two hinge helices to the *lac* operator appears necessary for DNA distortion. The two DNA duplexes that are bound to each repressor tetramer are ~ 25 Å apart and do not interact.

The sugar-binding domain consists of two topologically similar subdomains that are bridged by three polypeptide segments (Fig. 31-39). The two sugar-binding domains of a dimer make extensive contacts (Fig. 31-40a). IPTG binds to each sugar-binding domain between its subdomains. This does not significantly change the conformations of these subdomains, but it changes the angle between them. Although the hinge helix is not visible in the IPTG complex, model building indicates that, since the dimer's two hinge helices extend from its sugar-binding domains, this conformation change levers apart these hinge helices by 3.5 Å such that they and their attached HTH motifs can no longer simultaneously bind to their operator half-sites. Thus, inducer binding, which is allosteric within the dimer (has a positive homotropic effect; Section 10-4), greatly loosens the repressor's grip on the operator.

The C-terminal helices from each subunit, which are located on the opposite end of each subunit from the DNA-binding portion (at the point of the V), associate to form a bundle of four parallel helices that holds together the two repressor dimers, thereby forming the tetramer (Fig. 31-40a). The allosteric effects of inducer binding within each dimer are apparently not transmitted between dimers.

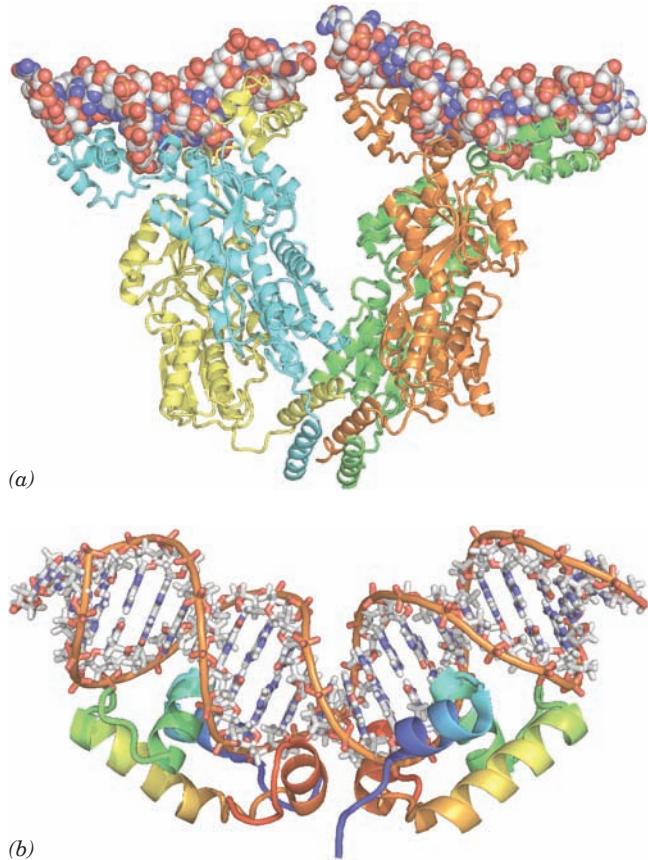


Figure 31-40 The structure of the *lac* repressor in complex with DNA. (a) The X-ray structure of the *lac* repressor tetramer bound to two 21-bp segments of symmetric *lac* operator DNA. The protein subunits are shown in ribbon form in yellow, cyan, green, and orange and the dsDNA segments are drawn in space-filling form with C white, N blue, O red, and P orange. [Courtesy of Ponzy Lu and Mitchell Lewis with coordinates generated by Benjamin Weider, University of Pennsylvania. PDBid 1L8G.] (b) The NMR structure of the 23-bp O_1 *lac* operator DNA in complex with two identical segments of the *lac* repressor consisting of its DNA-binding domain and its hinge helix. Each of the protein subunits is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA is represented in stick form with C white, N blue, O red, and P orange and with successive P atoms in the same chain connected by orange rods. The complex is viewed with its 2-fold axis vertical. Note that the protein dimer's two HTH motifs are inserted in successive major grooves at the periphery of the complex and that the insertion of the two centrally located hinge helices into the DNA's minor groove greatly widens and flattens the minor groove at this point and kinks the DNA in an upward bend. [Based on an NMR structure by Robert Kaptein and Rolf Boelens, Utrecht University, The Netherlands. PDBid 2KEI.]

Moreover, the *E. coli* purine repressor (**PurR**), which is homologous to the *lac* repressor but lacks its C-terminal helix, crystallizes as a dimer whose X-ray structure closely resembles that of the *lac* repressor dimer. What then is the function of *lac* repressor tetramerization?

Model building suggests that when the *lac* repressor tetramer simultaneously binds to both the O_1 and O_3 operators, the 93-bp DNA segment containing them forms a loop ~ 80 Å in diameter (Fig. 31-41). Furthermore, the CAP-cAMP binding site is exposed on the inner surface of the loop. Adding the CAP-cAMP at its proper position to this model reveals that the $\sim 90^\circ$ curvature which CAP-cAMP binding imposes on DNA (Fig. 31-31) has the correct direction and magnitude to stabilize the DNA loop, thereby stabilizing this putative CAP-cAMP-*lac* repressor-DNA complex. It may seem paradoxical that the binding of CAP-cAMP, a transcriptional activator, stabilizes the repressor-DNA complex. However, when both glucose and lactose are in short supply, it is important that the bacterium lower its basal rate of *lac* operon expression in order to conserve energy. The binding site (promoter) for RNAP is also located on the inner surface of the loop. Thus, the large size of the RNAP molecule would prevent

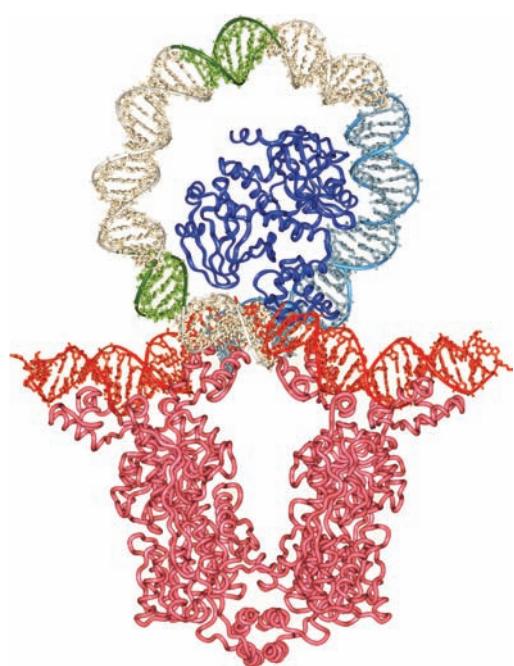


Figure 31-41 Model of the 93-bp DNA loop formed when *lac* repressor binds to O_1 and O_3 . The proteins are represented by their C_α backbones and the DNA is drawn in stick form with its sugar-phosphate backbones traced by helical ribbons. The model was constructed from the X-ray structure of the *lac* repressor (magenta) in complex with two 21-bp operator DNA segments (red) and the X-ray structure of CAP-cAMP (blue) in complex with its 30-bp target DNA (cyan; Fig. 31-28). The remainder of the DNA loop was generated by applying a smooth curvature to canonical B-DNA (white) with the -10 and -35 regions of the *lac* promoter highlighted in green. [Courtesy of Ponzy Lu and Mitchell Lewis, University of Pennsylvania.]

it from fully engaging the promoter in this looped complex, thereby maximizing repression.

c. Combining Genetic and Structural Studies of the *lac* Repressor Reveals Its Allosterically Important Residues

The phenotypes of 4042 point mutations of the *lac* repressor, which encompass nearly all of its 360 residues (making the *lac* repressor the most exhaustively mutationally characterized protein known) have been mapped onto its X-ray structure. Mutations with an I^- phenotype (*lac* repressors that fail to bind to the *lac* operator, so that β-galactosidase is constitutively synthesized) are located at the *lac* repressor's DNA-binding interface, at its dimer interface, or at internal residues of its inducer-binding core domain. Residues whose mutations result in the I^s phenotype (S for super-repressed; *lac* repressors that, in the presence of inducer, continue to repress the synthesis of β-galactosidase) appear to be of two types: (1) residues that are in direct contact with the inducer, whose alteration therefore interferes with inducer binding; and (2) residues at the dimer interface that are >8 Å from (not in direct contact with) the inducer-binding site. These latter mutations reveal which residues mediate the *lac* repressor's allosteric mechanism rather than directly binding the inducer or the DNA. Most of the allosterically important residues are located at the dimer interface and are members of the N-subdomain of the core domain, which links the inducer-binding sites to the operator DNA-binding sites. This is consistent with the observation that inducer binding causes a relative twist and translation of the N-subdomain, a movement which is propagated to the hinge helix and DNA-binding domain. This study demonstrates the power of combining genetic analysis with structural studies to elucidate structure-function relationships.

G. *trp* Operon: Attenuation

We now discuss a sophisticated transcriptional control mechanism named **attenuation** through which bacteria regulate the expression of certain operons involved in amino acid biosynthesis. This mechanism was discovered through the study of the *E. coli* *trp* operon (Fig. 31-42), which encodes five polypeptides comprising three enzymes that mediate the synthesis of tryptophan from chorismate (Section 26-5Bc). Charles Yanofsky established that the *trp* operon genes are coordinately expressed under the control of the *trp* repressor, a dimeric protein of identical 107-residue subunits that is the product of the *trpR* gene (which forms an independent operon). The *trp* repressor binds L-tryptophan, the pathway's end product, to form a complex that specifically binds to *trp* operator (*trpO*, Fig. 31-43) so as to reduce the rate of *trp* operon transcription 70-fold. The X-ray structure of the *trp* repressor-operator complex (Section 31-3Da) indicates that tryptophan binding allosterically orients *trp* repressor's two symmetry related helix-turn-helix "DNA reading heads" so that they can simultaneously bind to *trpO* (Fig. 31-34). Moreover, the bound tryptophan forms a hydrogen bond to a DNA phosphate group, thereby

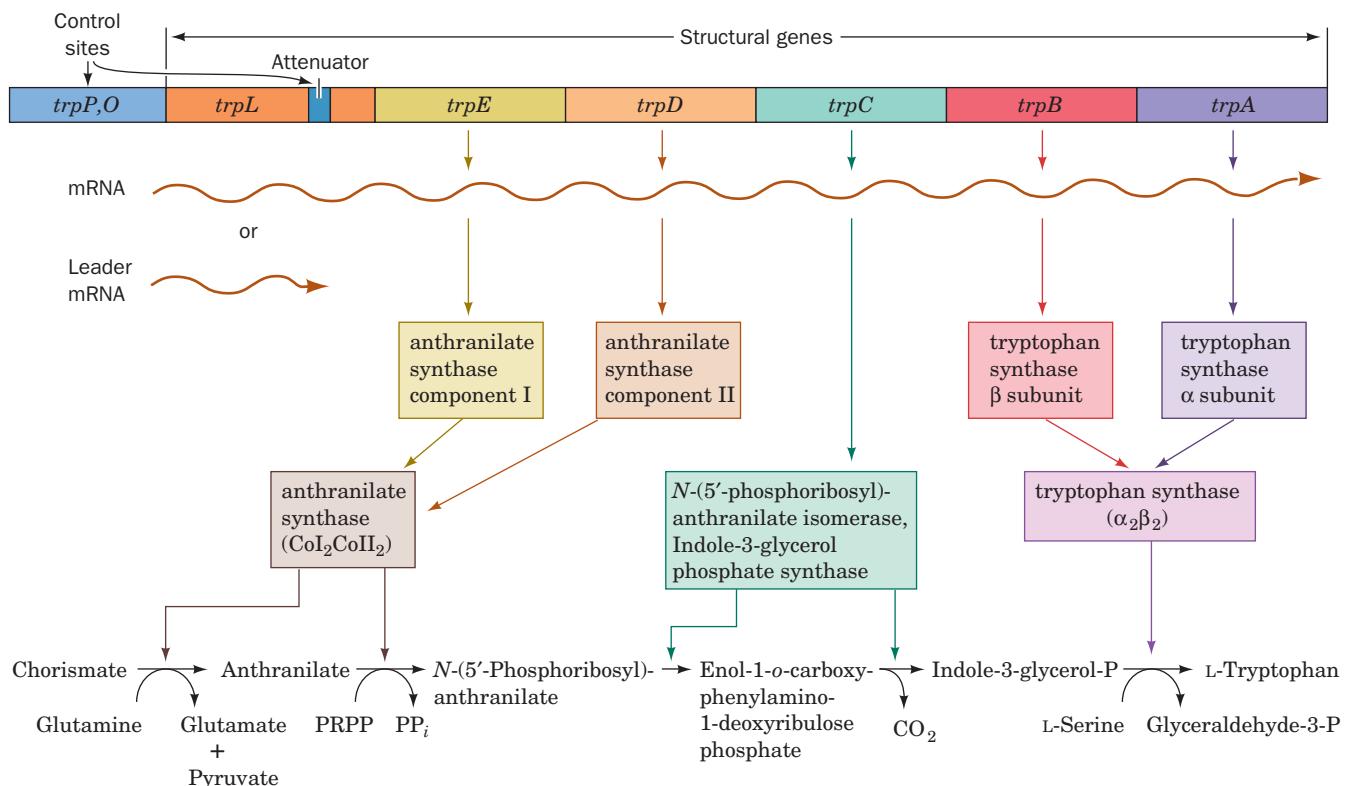


Figure 31-42 A genetic map of the *E. coli* *trp* operon indicating the enzymes it specifies and the reactions they catalyze. The gene product of *trpC* catalyzes two sequential

reactions in the synthesis of tryptophan. [After Yanofsky, C., *J. Am. Med. Assoc.* **218**, 1027 (1971).]

strengthening the repressor–operator association. Tryptophan therefore acts as a **corepressor**; its presence prevents what is then superfluous tryptophan biosynthesis (SAM similarly functions as a corepressor with the *met* repressor; Fig. 31-35a). The *trp* repressor also controls the synthesis of at least two other operons: the ***trpR* operon** and the ***aroH* operon** (which encodes one of three isozymes that catalyze the initial reaction of chorismate biosynthesis; Section 26-5Bc).

a. Tryptophan Biosynthesis Is Also Regulated by Attenuation

The *trp* repressor–operator system was at first thought to fully account for the regulation of tryptophan biosynthesis in *E. coli*. However, the discovery of *trp* deletion mu-

tantants located downstream from *trpO* that increase *trp* operon expression 6-fold indicated the existence of an additional transcriptional control element. Sequence analysis established that *trpE*, the *trp* operon's leading structural gene, is preceded by a 162-nucleotide **leader sequence** (*trpL*). Genetic analysis indicated that the new control element is located in *trpL*, ~30 to 60 nucleotides upstream of *trpE* (Fig. 31-42).

When tryptophan is scarce, the entire 6720-nucleotide polycistronic *trp* mRNA, including the *trpL* sequence, is synthesized. As the tryptophan concentration increases, the rate of *trp* transcription decreases as a result of the *trp* repressor–corepressor complex's consequent greater abundance. Of the *trp* mRNA that is transcribed, however, an increasing proportion consists of only a 140-nucleotide segment corresponding to the 5' end of *trpL*. The availability of tryptophan therefore results in the premature termination of *trp* operon transcription. The control element responsible for this effect is consequently termed an **attenuator**.

b. The *trp* Attenuator's Transcription Terminator Is Masked when Tryptophan Is Scarce

What is the mechanism of attenuation? The attenuator transcript contains four complementary segments that can form one of two sets of mutually exclusive base-paired

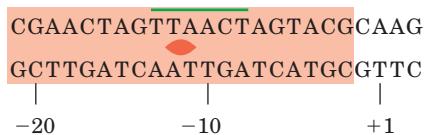


Figure 31-43 The base sequence of the *trp* operator. The nearly palindromic sequence is boxed and its *-10* region is underscored.

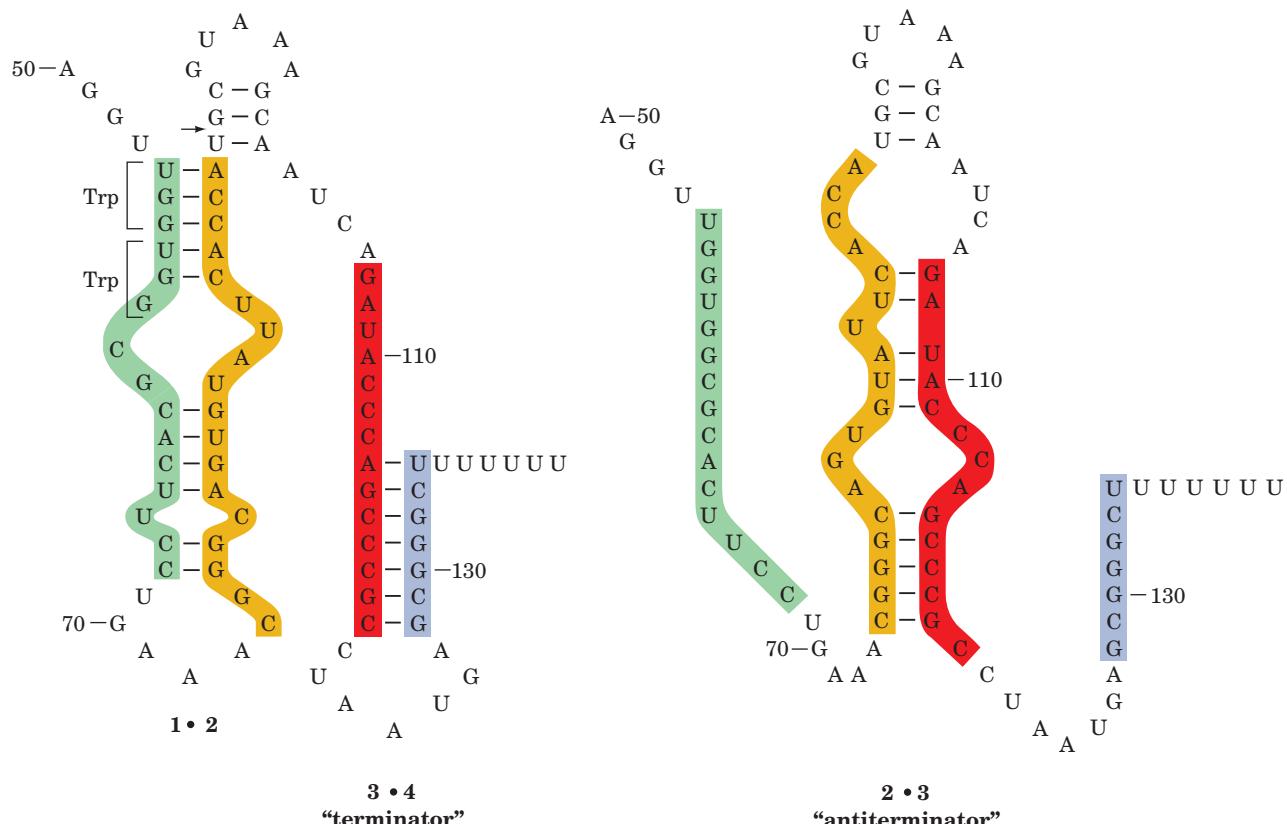


Figure 31-44 The alternative secondary structures of *trpL* mRNA. The formation of the base paired 2 · 3 (antiterminator) hairpin (right) precludes the formation of the 1 · 2 and 3 · 4 (terminator) hairpins (left) and vice versa. Attenuation results in the premature termination of transcription immediately after

nucleotide 140 when the 3 · 4 hairpin is present. The arrow indicates the mRNA site past which RNA polymerase pauses until approached by an active ribosome. [After Fisher, R.F. and Yanofsky, C., *J. Biol. Chem.* **258**, 8147 (1983).]

hairpins (Fig. 31-44). Segments 3 and 4 together with the succeeding residues comprise a normal intrinsic transcription terminator (Section 31-2Da): a G + C-rich sequence that can form a self-complementary hairpin structure followed by several sequential U's (compare with Fig. 31-18). Transcription rarely proceeds beyond this termination site unless tryptophan is in short supply.

A section of the leader sequence, which includes segment 1 of the attenuator, is translated to form a 14-residue polypeptide that contains two consecutive Trp residues (Fig. 31-44, left). The position of this particularly rare dipeptide segment (1.1% of the residues in *E. coli* proteins are Trp; Table 4-1) provided an important clue to the mechanism of attenuation. An additional essential aspect of this mechanism is that ribosomes commence the translation of a prokaryotic mRNA shortly after its 5' end has been synthesized.

The above considerations led Yanofsky to propose the following model of attenuation (Fig. 31-45). An RNA polymerase that has escaped repression initiates *trp* operon transcription. Soon after the ribosomal initiation site of the *trpL* gene has been transcribed, a ribosome attaches to it and begins translation of the leader peptide. When tryptophan is abundant, so that there is a plentiful supply of **tryptophanyl-tRNA^{Trp}** (the transfer RNA specific for Trp with

an attached Trp residue; Section 32-2C), the ribosome follows closely behind the transcribing RNA polymerase so as to sterically block the formation of the 2 · 3 hairpin. Indeed, RNA polymerase pauses past position 92 of the transcript and only continues transcription on the approach of a ribosome, thereby ensuring the proximity of these two entities at this critical position. The prevention of 2 · 3 hairpin formation permits the formation of the 3 · 4 hairpin, the transcription terminator pause site, which results in the termination of transcription (Fig. 31-45a). When tryptophan is scarce, however, the ribosome stalls at the tandem UGG codons (which specify Trp; Table 5-3) because of the lack of tryptophanyl-tRNA^{Trp}. As transcription continues, the newly synthesized segments 2 and 3 form a hairpin because the stalled ribosome prevents the otherwise competitive formation of the 1 · 2 hairpin (Fig. 31-45b). The formation of the transcriptional terminator's 3 · 4 hairpin is thereby pre-empted for sufficient time for RNA polymerase to transcribe through it and consequently through the remainder of the *trp* operon. The cell is thus provided with a regulatory mechanism that is responsive to the tryptophanyl-tRNA^{Trp} level, which, in turn, depends on the protein synthesis rate as well as on the tryptophan supply.

There is considerable evidence supporting this model of attenuation. The *trpL* transcript is resistant to limited

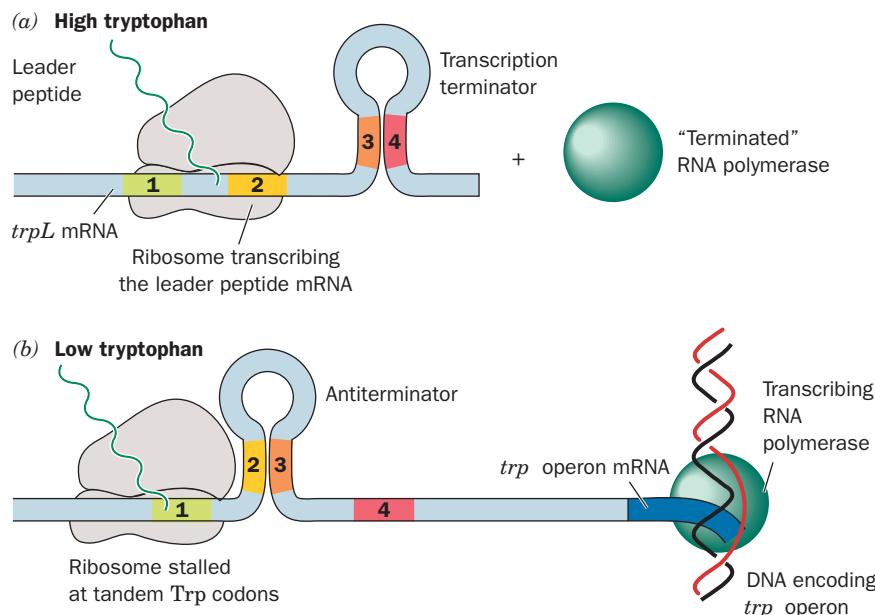


Figure 31-45 Attenuation in the *trp* operon. (a) When tryptophanyl-tRNA^{Trp} is abundant, the ribosome translates *trpL* mRNA. The presence of the ribosome on segment 2 prevents the formation of the base-paired 2 · 3 hairpin. The 3 · 4 hairpin, an essential component of the transcriptional terminator, can thereby form, thus aborting transcription. (b) When tryptophanyl-

tRNA^{Trp} is scarce, the ribosome stalls on the tandem Trp codons of segment 1. This situation permits the formation of the 2 · 3 hairpin which, in turn, precludes the formation of the 3 · 4 hairpin. RNA polymerase therefore transcribes through this unformed terminator and continues *trp* operon transcription.

RNase T1 digestion, indicating that it has extensive secondary structure. The significance of the tandem Trp codons in the *trpL* transcript is corroborated by their presence in *trp* leader regions of several other bacterial species. Moreover, the leader peptides of the five other amino acid-biosynthesizing operons known to be regulated by attenuation (most exclusively so) are all rich in their corresponding amino acid residues (Table 31-3). For example, the *E. coli his operon*, which specifies enzymes synthesizing histidine (Fig. 26-65), has seven tandem His residues in its leader peptide whereas the *ilv operon*, which specifies enzymes participating in isoleucine, leucine, and valine biosynthesis

(Fig. 26-61), has five Ile's, three Leu's, and six Val's in its leader peptide. Finally, the leader transcripts of these operons resemble that of the *trp* operon in their capacity to form two alternative secondary structures, one of which contains a trailing termination structure.

H. Riboswitches Are Metabolite-Sensing RNAs

We have just seen how the formation of secondary structure in a growing RNA transcript can regulate gene expression through attenuation. The conformational flexibility of mRNA also allows it to regulate genes by directly interacting

Table 31-3 Amino Acid Sequences of Some Leader Peptides in Operons Subject to Attenuation

Operon	Amino Acid Sequence ^a
<i>trp</i>	Met-Lys-Ala-Ile-Phe-Val-Leu-Lys-Gly-TRP-TRP-Arg-Thr-Ser
<i>pheA</i>	Met-Lys-His-Ile-Pro-PHE-PHE-PHE-Ala-PHE-PHE-PHE-Thr-PHE-Pro
<i>his</i>	Met-Thr-Arg-Val-Gln-Phe-Lys-HIS-HIS-HIS-HIS-HIS-HIS-Pro-Asp
<i>leu</i>	Met-Ser-His-Ile-Val-Arg-Phe-Thr-Gly-LEU-LEU-LEU-LEU-Asn-Ala-Phe-Ile-Val-Arg-Gly-Arg-Pro-Val-Gly-Gly-Ile-Gln-His
<i>thr</i>	Met-Lys-Arg-ILE-Ser-THR-THR-ILE-THR-THR-THR-ILE-THR-ILE-THR-THR-Gln-Asn-Gly-Ala-Gly
<i>ilv</i>	Met-Thr-Ala-LEU-LEU-Arg-VAL-ILE-Ser-LEU-VAL-VAL-ILE-Ser-VAL-VAL-VAL-ILE-ILE-ILE-Pro-Pro-Cys-Gly-Ala-Ala-Leu-Gly-Arg-Gly-Lys-Ala

^aResidues in uppercase are synthesized in the pathway catalyzed by the operon's gene products.

Source: Yanofsky, C., *Nature* **289**, 753 (1981).

with certain cellular metabolites, thereby eliminating the need for sensor proteins such as the *lac* repressor, CAP, and the *trp* repressor.

In *E. coli*, the biosynthesis of thiamine pyrophosphate (TPP; Section 17-3Ba) requires the action of several proteins whose levels vary according to the cell's need for TPP. In at least two of the relevant genes the untranslated regions at the 5' end of the mRNA include a highly conserved sequence called the ***thi* box**. The susceptibility of the *thi* box to chemical or enzymatic cleavage, as Ronald Breaker showed, differs in the presence and absence of TPP, suggesting that the RNA changes its secondary structure when TPP binds to it (the binding of a metabolite by RNA is not unprecedented; synthetic oligonucleotides known as **aptamers** bind specific molecules with high specificity and affinity; Section 7-6C). The TPP-sensing mRNA element has been dubbed a **riboswitch**.

The predicted secondary structure of the TPP-sensing riboswitch and its proposed mechanism are shown in Fig. 31-46a. In the absence of TPP, the mRNA assumes a con-

formation that allows a ribosome to begin translation. In the presence of TPP, an alternative secondary structure masks the sequence that identifies its translational initiation site to the ribosome (its so-called **Shine-Dalgarno sequence**; Section 32-3Cb) so that the ribosome cannot initiate the mRNA's translation. Thus, *the concentration of a metabolite can regulate the expression of genes required for its synthesis*. The X-ray structure of the 80-nt TPP-binding domain from the *E. coli* TPP-sensing riboswitch, determined by Breaker and Dinshaw Patel, reveals an intricately folded RNA that binds TPP in an extended conformation (Fig. 31-46b).

Over 20 classes of riboswitches have as yet been identified, including those that regulate the expression of enzymes involved in the metabolism of coenzyme B₁₂ (Fig. 25-21), riboflavin (Fig. 16-8), S-adenosylmethionine (SAM; Fig. 26-18), lysine, and adenine. In general, they consist of two components, an aptamer that binds an effector and a so-called **expression platform** that transduces effector binding to a change in gene expression. In some cases, the

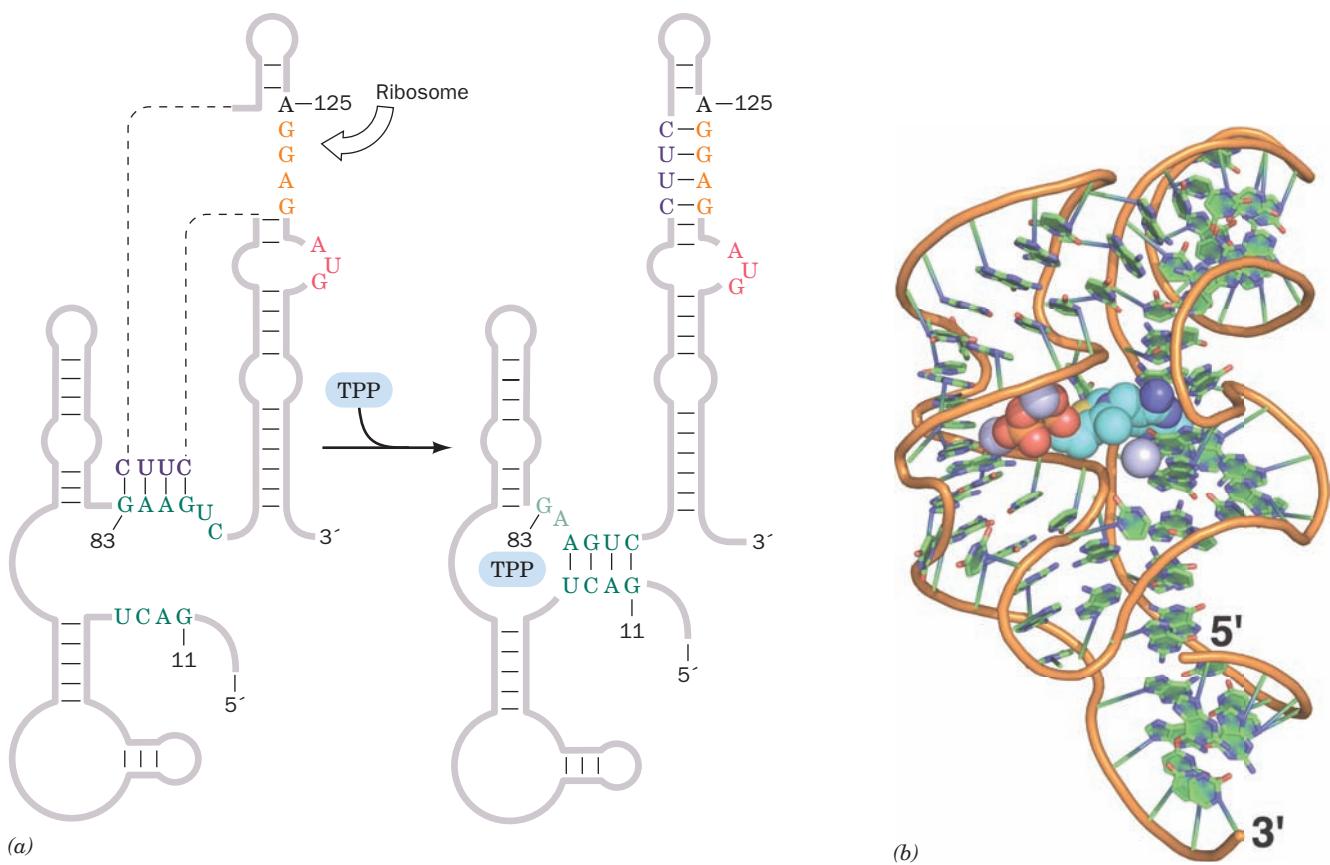


Figure 31-46 Structure of the TPP-sensing riboswitch from *E. coli*. (a) The predicted secondary structure of a 165-residue segment at the 5' end of the *thiM* gene is shown in the absence (left) and presence (right) of TPP. The TPP-binding conformation masks the Shine-Dalgarno sequence (orange) required by the ribosome to initiate translation at the AUG start codon (red) just downstream. [After Winkler, W., Nahvi, A., and Breaker, R.R., *Nature* **419**, 952 (2002).] (b) The X-ray structure of the riboswitch's 80-nt TPP-sensing domain. The RNA is drawn in

cartoon form with its sugar-phosphate backbone represented by an orange rod and its bases represented by paddles with C green, N blue, and O red. The TPP is drawn in space-filling form with C cyan, N blue, O red, and S yellow. Mg²⁺ ions are represented by lavender spheres. [Based on an X-ray structure by Ronald Breaker, Yale University, and Dinshaw Patel, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 2GDI.]

activated expression platform forms an intrinsic transcription termination site (Section 31-2Da) so that transcription beyond this site proceeds only when the effector is absent. In others, the activated expression platform cleaves itself, thereby inactivating the mRNA (the ability of RNA to act as an enzyme is discussed in Section 31-4Ae).

Riboswitches collectively regulate >2% of the genes in certain bacteria. Plants and fungi also contain riboswitches (although the above TTP-sensing riboswitch could not function in eukaryotes because eukaryotic ribosomes do not bind Shine–Dalgarno sequences; Section 32-3Cd). The fact that the interaction of riboswitches with their effectors does not require the participation of proteins suggests that they are relics of the RNA world (Section 1-5Ca) and hence among the oldest regulatory systems.

I. Regulation of Ribosomal RNA Synthesis: The Stringent Response

E. coli cells growing under optimal conditions divide every 20 min. Such cells contain up to 20,000 ribosomes and hence must synthesize ~10,000 ribosomes per cell division cycle. Yet RNAP can initiate the transcription of an rRNA gene no faster than about once per second. If *E. coli* contained only one copy of each of the three types of rRNA genes (those specifying the so-called 23S, 16S, and 5S rRNAs; Section 32-3A), fast-growing cells could synthesize no more than ~1200 ribosomes during their cell division cycle. However, the *E. coli* genome contains seven separately located rRNA operons, all of which contain one nearly identical copy of each type of rRNA gene. Moreover, rapidly growing cells contain multiple copies of their replicating chromosomes (Section 30-3Cb), thereby accounting for the observed rRNA synthesis rate.

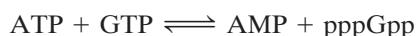
Cells have the remarkable ability to coordinate the rates at which their thousands of components are synthesized. For example, *E. coli* adjust their ribosome content to match the rate at which they can synthesize proteins under the prevailing growth conditions. The rate of rRNA synthesis is therefore proportional to the rate of protein synthesis. One mechanism by which this occurs is known as the **stringent response**: *A shortage of any species of amino acid–charged tRNA (usually a result of “stringent” or poor growth conditions) that limits the rate of protein synthesis triggers a sweeping metabolic readjustment.* A major facet of this change is an abrupt 10- to 20-fold reduction in the rate of rRNA and tRNA synthesis. This **stringent control**, moreover, depresses numerous metabolic processes (including DNA replication and the biosynthesis of carbohydrates, lipids, nucleotides, proteoglycans, and glycolytic intermediates) while stimulating others (such as amino acid biosynthesis). The cell is thereby prepared to withstand nutritional deprivation.

a. (p)ppGpp Mediates the Stringent Response

The stringent response is correlated with a rapid intracellular accumulation of two unusual nucleotides, ppGpp and pppGpp [known collectively as (p)ppGpp], and their prompt decay when amino acids become available. The ob-

servation that mutants, designated *relA*[−], which do not exhibit the stringent response (they are said to have **relaxed control**) lack (p)ppGpp suggests that these substances mediate the stringent response. This idea was corroborated by *in vitro* studies demonstrating, for example, that (p)ppGpp inhibits the transcription of rRNA genes but stimulates the transcription of the *trp* and *lac* operons as does the stringent response *in vivo*. Apparently, (p)ppGpp acts by somehow altering RNAP's promoter specificity at stringently controlled operons, a hypothesis that is supported by the isolation of RNAP mutants that exhibit reduced responses to (p)ppGpp. In addition, (p)ppGpp causes an increased frequency of pausing in RNAPs engaged in elongation, thereby reducing the rate of transcription.

The protein encoded by the wild-type *relA* gene, named **stringent factor (RelA)**, catalyzes the reaction



and, to a lesser extent,



However, several ribosomal proteins convert pppGpp to ppGpp so that ppGpp is the stringent response's usual effector. Stringent factor is only active in association with a ribosome that is actively engaged in translation. (p)ppGpp synthesis occurs when a ribosome binds its mRNA-specified but uncharged tRNA (lacking an appended amino acid residue). The binding of a specified and charged tRNA greatly reduces the rate of (p)ppGpp synthesis. *The ribosome apparently signals the shortage of an amino acid by stimulating the synthesis of (p)ppGpp which, acting as an intracellular messenger, influences the rates at which a great variety of operons are transcribed.*

(p)ppGpp degradation is catalyzed by the *spoT* gene product. The *spoT*[−] mutants show a normal increase in (p)ppGpp level on amino acid starvation but an abnormally slow decay of (p)ppGpp to basal levels when amino acids again become available. The *spoT*[−] mutants therefore exhibit a sluggish recovery from the stringent response. *The (p)ppGpp level is apparently regulated by the countervailing activities of stringent factor and the spoT gene product.*

4 POST-TRANSCRIPTIONAL PROCESSING

The immediate products of transcription, the **primary transcripts**, are not necessarily functional entities. In order to acquire biological activity, many of them must be specifically altered in several ways: (1) by the exo- and endonucleolytic removal of polynucleotide segments; (2) by appending nucleotide sequences to their 3' and 5' ends; and (3) by the modification of specific nucleosides. The three major classes of RNAs, mRNA, rRNA, and tRNA, are altered in different ways in prokaryotes and in eukaryotes. In this section we shall outline these **post-transcriptional modification** processes.

A. Messenger RNA Processing

In prokaryotes, most primary mRNA transcripts function in translation without further modification. Indeed, as we have seen, ribosomes in prokaryotes usually commence translation on nascent mRNAs. In eukaryotes, however, mRNAs are synthesized in the cell nucleus, whereas translation occurs in the cytosol. Eukaryotic mRNA transcripts can therefore undergo extensive post-transcriptional processing while still in the nucleus.

a. Eukaryotic mRNAs Are Capped

Eukaryotic mRNAs have a peculiar enzymatically appended **cap structure** consisting of a **7-methylguanosine (m⁷G)** residue joined to the transcript's initial (5') nucleoside via a 5'-5' triphosphate bridge (Fig. 31-47). This m⁷ cap, which is added to the growing transcript before it is ~30 nucleotides long, defines the eukaryotic translational start site (Section 32-3Cd). A cap may be O²'-methylated at the transcript's leading nucleoside (**cap-1**, the predominant cap in multicellular organisms), at its first two nucleosides (**cap-2**), or at neither of these positions (**cap-0**, the predominant cap in unicellular eukaryotes). If the leading nucleoside is adenosine (it is usually a purine), it may also be N⁶-methylated.

Capping involves several enzymatic reactions: (1) the removal of the leading phosphate group from the mRNA's 5' terminal triphosphate group by an **RNA triphosphatase**; (2) the guanylation of the mRNA by **capping enzyme**, which requires GTP and yields the 5'-5' triphosphate bridge and PP_i; (3) the methylation of guanine by **guanine-7-methyltransferase** in which the methyl group is supplied by S-adenosylmethionine (SAM); and possibly (4) the O2' methylation of the mRNA's first and perhaps its second nucleotide by a SAM-requiring **2'-O-methyltransferase**. Both the capping enzyme and the guanine-7-methyltransferase bind to RNAP II's phosphorylated CTD (Section 31-2E). Hence it is likely that capping marks the completion of RNAP II's switch from transcription initiation to elongation.

b. Eukaryotic mRNAs Have Poly(A) Tails

Eukaryotic mRNAs, unlike those of prokaryotes, are invariably monocistronic. Moreover, in contrast to the case in bacteria (Section 31-2D), no eukaryotic transcriptional termination sequence has been identified (but see below). In fact, the eukaryotic termination process is imprecise; that is, the primary transcripts of a given structural gene have heterogeneous 3' sequences. Nevertheless, mature eukaryotic mRNAs have well-defined 3' ends; *almost all of them in mammals have 3'-poly(A) tails of ~250 nucleotides (~80 in yeast)*. The poly(A) tails are enzymatically appended to the primary transcripts in two reactions that are mediated by a 500- to 1000-kD complex that consists of at least six proteins:

1. A transcript is cleaved to yield a free 3'-OH group at a specific site that is 15 to 25 nucleotides past an AAUAAA sequence and within 50 nucleotides before a

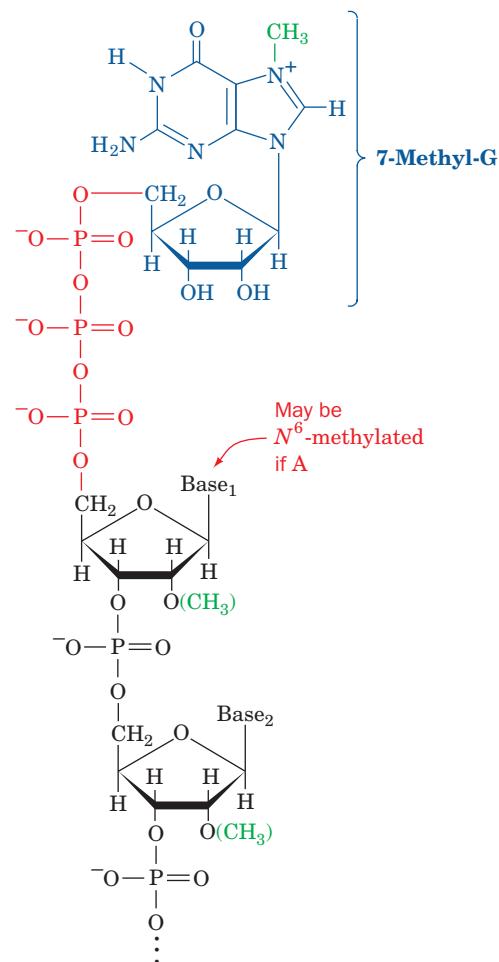


Figure 31-47 The structure of the 5' cap of eukaryotic mRNAs. It is known as cap-0, cap-1, or cap-2, respectively, if it has no further modifications, if the leading nucleoside of the transcript is O²'-methylated, or if its first two nucleosides are O²'-methylated.

U-rich or G + U-rich sequence. The AAUAAA sequence is highly conserved in higher eukaryotes (but not yeast) in which its mutation abolishes cleavage and polyadenylation. The precision of the cleavage reaction has apparently eliminated the need for accurate transcriptional termination. Nevertheless, the identity of the endonuclease that cleaves the RNA is uncertain although **cleavage factors I** and **II (CFI and CFI)** are required for this process.

2. The poly(A) tail is subsequently generated from ATP through the stepwise action of **poly(A) polymerase (PAP)**. This enzyme, which by itself only weakly binds RNA, is recruited by **cleavage and polyadenylation specificity factor (CPSF)** on this heterotetramer's recognition of the AAUAAA sequence, which it does with almost no tolerance for sequence variation. The downstream G + U-rich element is recognized by the heterotrimeric **cleavage stimulation factor (CstF)**, which increases the affinity with which CPSF binds the AAUAAA sequence. However,

once the poly(A) tail has grown to \sim 10 residues, the AAUAAA sequence is no longer required for further chain elongation. This suggests that CPSF becomes disengaged from its recognition site in a manner reminiscent of the way σ factor is released from the transcriptional initiation site once the elongation of prokaryotic mRNA is under way (Section 31-2B). The final length of the poly(A) tail is controlled by **poly(A)-binding protein II (PAB II)**, multiple copies of which bind to successive segments of poly(A). PAB II also increases the processivity of PAP.

Both CPSF and CstF bind to the phosphorylated RNAP II CTD (Section 31-2E); deleting the CTD inhibits polyadenylation. Evidently, the CTD couples polyadenylation to transcription. The mechanism that controls the length of a poly(A) tail is unclear.

PAP is a template-independent RNA polymerase that elongates an mRNA primer with a free 3'-OH group. The X-ray structure of the 530-residue D154A mutant form of yeast PAP (D154 is a catalytically essential active site residue) in complex with A₅ and ATP, determined by Andrew Bohm, reveals that this monomeric protein consists of three domains that form a prominent U-shaped cleft (Fig. 31-48). Hence it has the handlike domain arrangement of template-directed polymerases (Section 30-2Ad). Indeed, PAP's N-terminal domain, which contains the enzyme's active site, is homologous to the palm domain of

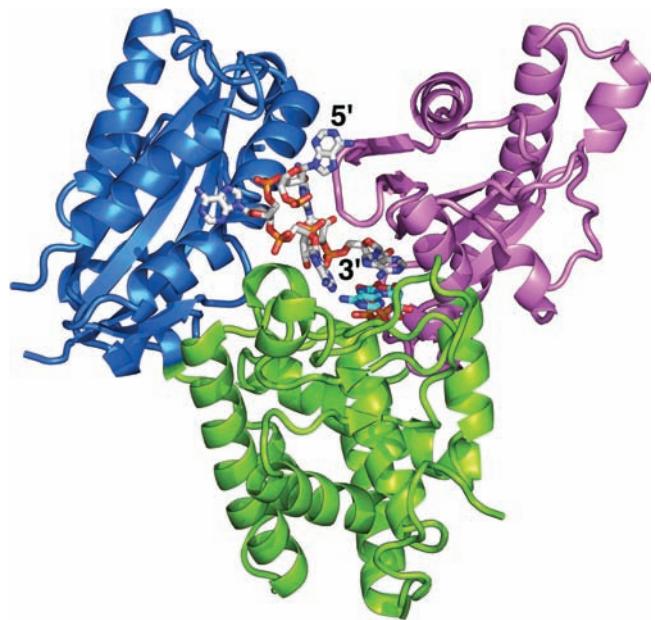


Figure 31-48 X-ray structure of the D154A mutant of yeast poly(A) polymerase (PAP) in complex with A₅ and ATP. PAP is drawn in ribbon form with its N-terminal domain lavender, its central domain yellow-green, and its C-terminal domain light blue. The A₅ and ATP are drawn in stick form with A₅ C white, ATP C cyan, N blue, O red, and P orange. [Based on an X-ray structure by Andrew Bohm, Tufts University School of Medicine. PDBid 2Q66.]

DNA polymerase β , although it forms the thumb side of the cleft rather than its base. PAP's central domain, which forms the base of the cleft, is functionally but not structurally analogous to the fingers domain of template-directed polymerases in that it interacts with the β and γ phosphates of the incoming ATP. However, the C-terminal domain shows no resemblance to a fingers domain. Rather, it is topologically similar to the **RNA-recognition motif [RRM]**; also known as the **RNA-binding domain (RBD)**] that occurs in >200 different RNA-binding proteins (see below). The A₅ binds in the cleft in an extended conformation such that, in contrast to the nucleic acids bound to template-dependent polymerases, its bases are not in contact. However, the 3'-terminal base of the A₅ stacks on the base of the ATP.

In comparison to the X-ray structure of yeast PAP in complex with 3'-dATP, also determined by Bohm, the N-terminal domain in the PAP-A₅-ATP structure has undergone an \sim 23° hingelike motion toward the C-terminal domain to form a closed conformation resembling that observed in the structure of Taq DNA polymerase I in complex with its substrates (Section 30-2Ae). This motion forms several new interactions, both direct and water-mediated, between PAP and its ATP substrate that helps differentiate adenine from other bases. In contrast, in template-dependent polymerases, the incoming base only makes sequence-specific contacts with the template base (Section 30-2Ae).

In vitro studies indicate that a poly(A) tail is not required for mRNA translation. Rather, the observations that an mRNA's poly(A) tail shortens as it ages in the cytosol and that unadenylated mRNAs have abbreviated cytosolic lifetimes suggest that poly(A) tails have a protective role. In fact, the only mature mRNAs that lack poly(A) tails, those of histones (which, with few exceptions, lack the AAUAAA cleavage–polyadenylation signal), have lifetimes of <30 min in the cytosol, whereas most other mRNAs last hours or days. The poly(A) tails are specifically complexed in the cytosol by **poly(A) binding protein (PABP**; not related to PAB II), which organizes poly(A)-bearing mRNAs into ribonucleoprotein particles. PABP is thought to protect mRNA from degradation as is suggested, for example, by the observation that the addition of PABP to a cell-free system containing mRNA and mRNA-degrading nucleases greatly reduces the rate at which the mRNAs are degraded and the rate at which their poly(A) tails are shortened.

All known PABPs contain four tandem and highly conserved RNA-recognition motifs (RRMs) followed by a less conserved Pro-rich C-terminal segment of variable length. A variety of evidence suggests that PABP's first two RRMs support most of the biochemical functions of full-length PABP. The X-ray structure of the first two RRMs of human PABP (RRM1/2; the N-terminal 190 residues of this 636-residue protein) in complex with A₁₁, determined by Stephen Burley, reveals that RRM1/2 forms a continuous trough-shaped surface in which the poly(A) binds in an extended conformation via interactions with conserved

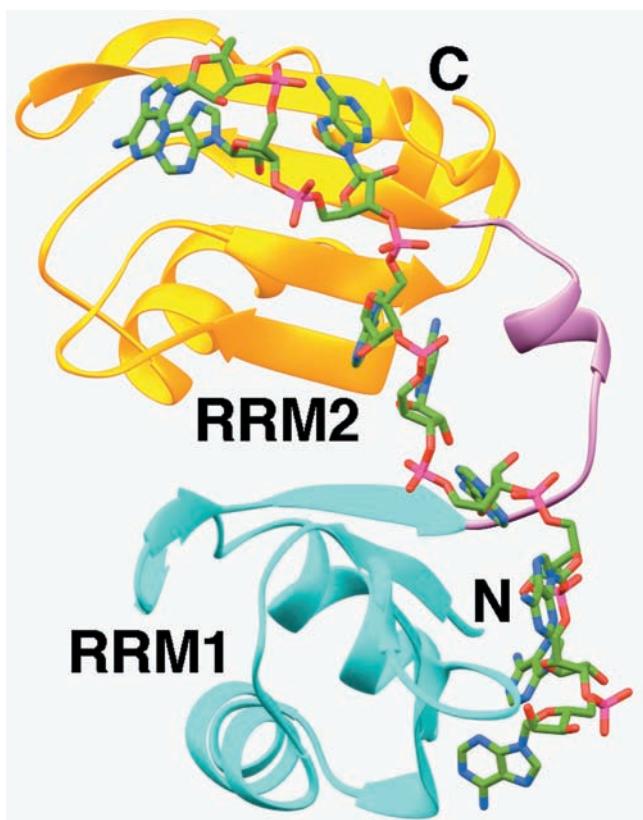


Figure 31-49 X-ray structure of the N-terminal two RNA-recognition motifs (RRMs) of human PABP in complex with A₁₁. RRM1 is cyan, RRM2 is gold, and their linking segment is lavender. The poly(A), only nine of whose nucleotides are observed, is drawn in stick form with C green, N blue, O red, and P magenta. [Based on an X-ray structure by Stephen Burley, The Rockefeller University. PDBid 1CVJ.]

residues (Fig. 31-49). Each RRM, as also seen in the structures of a variety of other RNA-binding proteins, consists of a compact globule made of a 4-stranded antiparallel sheet that forms the RNA-binding surface backed by two helices.

The cleavage of a transcript past its AAUAAA sequence does not, in itself, terminate transcription. However, in yeast, the protein **Rtt103**, which binds to the phosphorylated CTD of RNAP II, recognizes the AAUAAA sequence and recruits the 5' → 3' exonuclease known as **Rat1** (Xrn2 in humans). Then, in what is termed the **torpedo model**, the highly processive Rat1/Xrn2 loads onto the newly liberated 5' end of the still nascent RNA and rapidly degrades it until it intercepts the RNAP and induces it to terminate RNA synthesis. It has been hypothesized that this occurs in much the same way as Rho factor terminates bacterial transcription (Section 31-2Db). This frees the RNAP to initiate a new round of transcription.

c. Eukaryotic Genes Consist of Alternating Expressed and Unexpressed Sequences

The most striking difference between eukaryotic and prokaryotic structural genes is that the coding sequences of

most eukaryotic genes are interspersed with unexpressed regions. Early investigations of eukaryotic structural gene transcription found, quite surprisingly, that primary transcripts are highly heterogeneous in length (from ~2000 to well over 20,000 nucleotides) and are much larger than was expected from the known sizes of eukaryotic proteins. Rapid labeling experiments demonstrated that little of this so-called **heterogeneous nuclear RNA (hnRNA)** is ever transported to the cytosol; most of it is quickly turned over (degraded) in the nucleus. Yet, the hnRNA's 5' caps and 3' tails eventually appear in cytosolic mRNAs. *The straightforward explanation of these observations, that pre-mRNAs are processed by the excision of internal sequences, seemed so bizarre that it came as a great surprise in 1977 when Phillip Sharp and Richard Roberts independently demonstrated that this is actually the case.* In fact, mammalian pre-mRNAs typically contain eight noncoding **intervening sequences (introns)** whose aggregate length averages 4 to 10 times that of their flanking **expressed sequences (exons)**. This situation is graphically illustrated in Fig. 31-50, which is an electron micrograph of chicken **ovalbumin** mRNA hybridized to the antisense strand of the ovalbumin gene (ovalbumin is the major protein component of egg white).

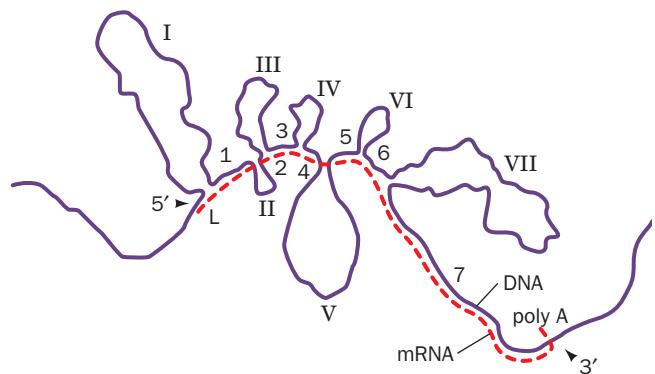
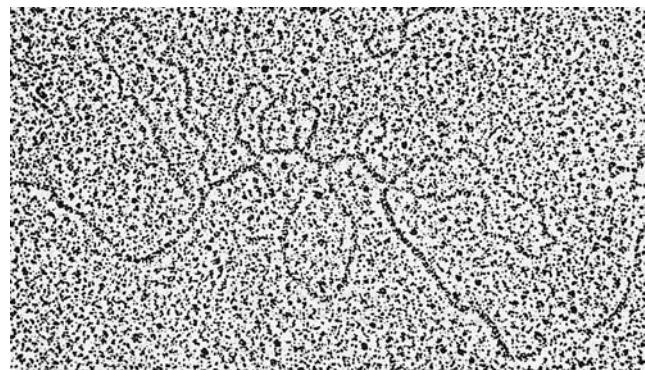


Figure 31-50 An electron micrograph and its interpretive drawing of a hybrid between the antisense strand of the chicken ovalbumin gene and its corresponding mRNA. The complementary segments of the DNA (purple line in the drawing) and mRNA (red dashed line) have annealed to reveal the exon positions (L, 1–7). The looped-out segments (I–VII), which have no complementary sequences in the mRNA, are the introns. [From Chambon, P., *Sci. Am.* **244**(5), 61 (1981).]

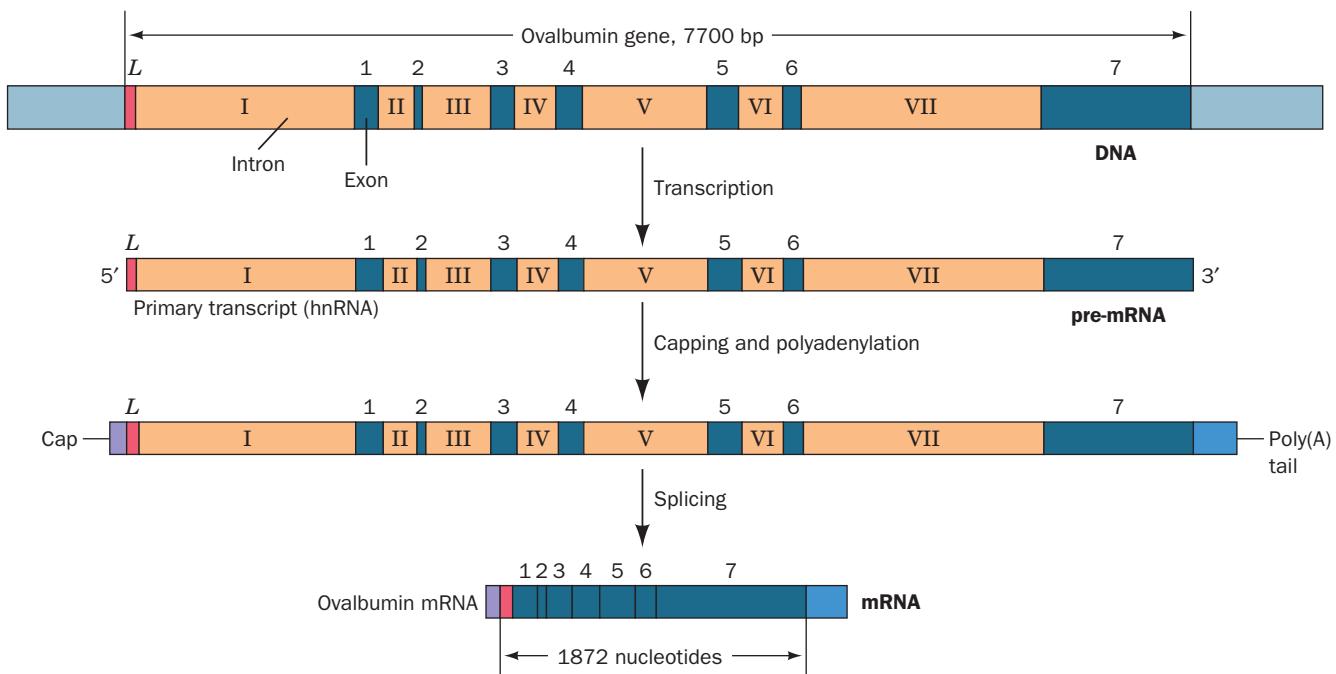


Figure 31-51 The sequence of steps in the production of mature eukaryotic mRNA as shown for the chicken ovalbumin gene. Following transcription, the primary transcript is capped

and polyadenylated. The introns are then excised and the exons spliced together to form the mature mRNA. However, splicing may also occur cotranscriptionally.

Exons have lengths that range up to 17,106 nt (in the gene encoding the 34,350-residue muscle protein **titin**, the largest known single-chain protein; Section 35-3Ae) but with most <300 nt (and averaging 150 nt in humans). Introns, in contrast, are usually much longer, with lengths averaging ~3500 nt and as high as ~800,000 nt [in the gene encoding the muscle protein **dystrophin** (Section 35-3Ae), whose length is ~2400 kb and hence is the largest human gene] with no obvious periodicity. Moreover, the corresponding introns from genes in two vertebrate species can vary extensively in both length and sequence so as to bear little resemblance to one another. The number of introns in a gene averages 7.8 in the human genome and varies from none to 364 (with the latter number occurring in the gene encoding titin).

The formation of eukaryotic mRNA begins with the transcription of an entire structural gene, including its introns, to form pre-mRNA (Fig. 31-51). Then, following capping, the introns are excised and their flanking exons are

connected, a process called **gene splicing** or just **splicing**, that often occurs cotranscriptionally. *The most striking aspect of gene splicing is its precision; if one nucleotide too few or too many were excised, the resulting mRNA could not be translated properly (Section 32-1B). Moreover, exons are never shuffled; their order in the mature mRNA is exactly the same as that in the gene from which it is transcribed.*

d. Exons Are Spliced in a Two-Stage Reaction

Sequence comparisons of exon–intron junctions from a diverse group of eukaryotes indicate that they have a high degree of homology (Fig. 31-52), including, as Richard Breathnach and Pierre Chambon first pointed out, *an invariant GU at the intron's 5' boundary and an invariant AG at its 3' boundary. These sequences are necessary and sufficient to define a splice junction*: Mutations that alter the sequences interfere with splicing, whereas mutations that change a nonjunction to a consensus-like sequence can generate a new splice junction.

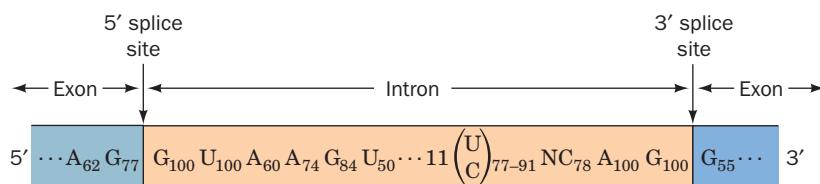


Figure 31-52 The consensus sequence at the exon–intron junctions of vertebrate pre-mRNAs. The subscripts indicate the percentage of pre-mRNAs in which the specified base(s) occurs. Note that the 3' splice site is preceded by a tract of 11

predominantly pyrimidine nucleotides. [Based on data from Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S.S., and Sharp, P.A., *Annu. Rev. Biochem.* **55**, 1123 (1986).]

Investigations of both cell-free and *in vivo* splicing systems by Argiris Efstratiadis, Tom Maniatis, Michael Rosbash, and Sharp established that intron excision occurs via two transesterification reactions that are remarkably similar from yeast to humans (Fig. 31-53):

1. The formation of a 2',5'-phosphodiester bond between an intron adenosine residue and its 5'-terminal phosphate group with the concomitant liberation of the 5' exon's 3'-OH group. *The intron thereby assumes a novel lariat structure.* The adenosine residue at the lariat branch has been identified in yeast as the last A in the highly conserved sequence UACUAAC and in vertebrates as the A in the equivalent but more permissive sequence YNCURAY [where R represents purines (A or G), Y represents pyrimidines (C or U), and N represents any nucleotide]. In yeast and vertebrates, the branch point A occurs ~50 and 18 to 40 residues upstream of the associated 3' splice site, respectively. In yeast, which have relatively few introns, mutations that change this branch point A residue abolish splicing at that site. However, in higher eukaryotes, the mutation or deletion of a branch site often activates a so-called **cryptic branch site** that is also near the 3' splice site. Evidently, the branch site functions to identify the nearest 3' splice site as a target for linkage to the 5' splice site.

2. The now free 3'-OH group of the 5' exon forms a phosphodiester bond with the 5'-terminal phosphate of the 3' exon yielding the spliced product and releasing the intron lariat with a free 3'-OH group. The intron lariat is then

debranched (linearized) and, *in vivo*, is rapidly degraded. Mutations that alter the conserved AG at the 3' splice site block this second step, although they do not interfere with lariat formation.

Note that the splicing process proceeds without free energy input; its transesterification reactions preserve the free energy of each cleaved phosphodiester bond through the concomitant formation of a new one.

The sequences required for splicing are the short consensus sequences at the 3' and 5' splice sites and at the branch site. Nevertheless, these sequences are poorly conserved. However, other short sequence elements within exons and introns that are known as **exonic and intronic splicing enhancers (ESEs and ISEs)** and **silencers (ESSs and ISSs)** also play important roles in splice site selection although their characteristics are poorly understood (even highly sophisticated computer programs are only ~50% successful in predicting actual splice sites over apparently equally good candidates that are not). In contrast, large portions of most introns can be deleted without impeding splicing.

e. Some Eukaryotic Genes Are Self-Splicing

It is now recognized that there are eight distinct types of introns, seven of which occur in eukaryotes (Table 31-4). **Group I introns** occur in the nuclei, mitochondria, and chloroplasts of diverse eukaryotes (but not vertebrates), and even in some bacteria. Thomas Cech's study of how

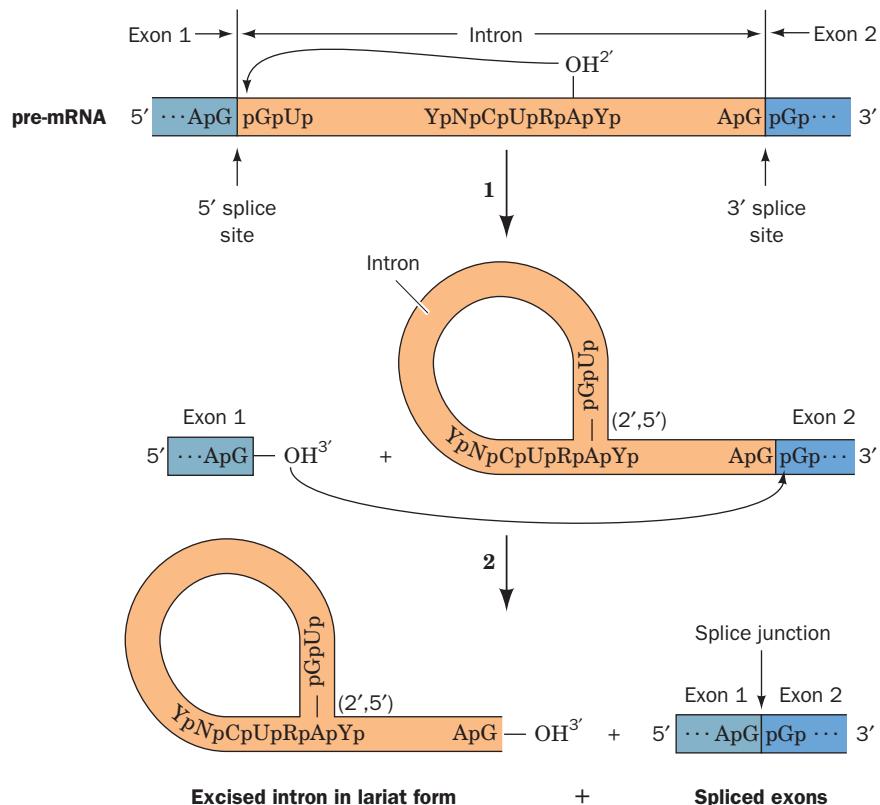


Figure 31-53 The sequence of transesterification reactions that splice together the exons of eukaryotic pre-mRNAs. The exons and introns are drawn in blue and orange, and R and Y represent purine and pyrimidine residues. **(1)** The 2'-OH group of a specific intron A residue nucleophilically attacks the 5'-phosphate at the 5' intron boundary to yield an unusual 2',5'-phosphodiester bond and thus form a lariat structure. **(2)** The liberated 3'-OH group forms a 3',5'-phosphodiester bond with the 5' terminal residue of the 3' exon, thereby splicing the two exons together and releasing the intron in lariat form with a free 3'-OH.

Table 31-4 Types of Introns

Intron Type	Where Found
GU-AG introns	Eukaryotic nuclear pre-mRNA
AU-AC introns	Eukaryotic nuclear pre-mRNA
Group I	Eukaryotic nuclear pre-mRNA, organelle RNAs, a few bacterial RNAs
Group II	Organelle RNAs, a few prokaryotic RNAs
Group III	Organelle RNAs
Twintrons (composites of two and/or more group II or III introns)	Organelle RNAs
Pre-tRNA introns	Eukaryotic nuclear pre-tRNAs
Archaeal introns	Various RNAs

Source: Brown, T.A., *Genomes* (3rd ed.), Garland Science, p. 355 (2007).

group I introns are spliced in the ciliated protozoan *Tetrahymena thermophila* led to an astonishing discovery: *RNA can act as an enzyme*. When the isolated pre-rRNA of this organism is incubated with guanosine or a free guanine nucleotide (GMP, GDP, or GTP), but in the absence of protein, its single 421-nucleotide intron excises itself and splices together its flanking exons; that is, this pre-rRNA is self-splicing. The three-step reaction sequence of this process (Fig. 31-54) resembles that of mRNA splicing:

1. The 3'-OH group of the guanosine forms a phosphodiester bond with the intron's 5' end, liberating the 5' exon.
2. The 3'-terminal OH group of the newly liberated 5' exon forms a phosphodiester bond with the 5'-terminal phosphate of the 3' exon, thereby splicing together the two exons and releasing the intron.
3. The 3'-terminal OH group of the intron forms a phosphodiester bond with the phosphate of the nucleotide 15 residues from the intron's 5' end, yielding the 5'-terminal fragment with the remainder of the intron in cyclic form.

This self-splicing process consists of a series of transesterifications and therefore does not require free energy input. Cech further established the enzymatic properties of the

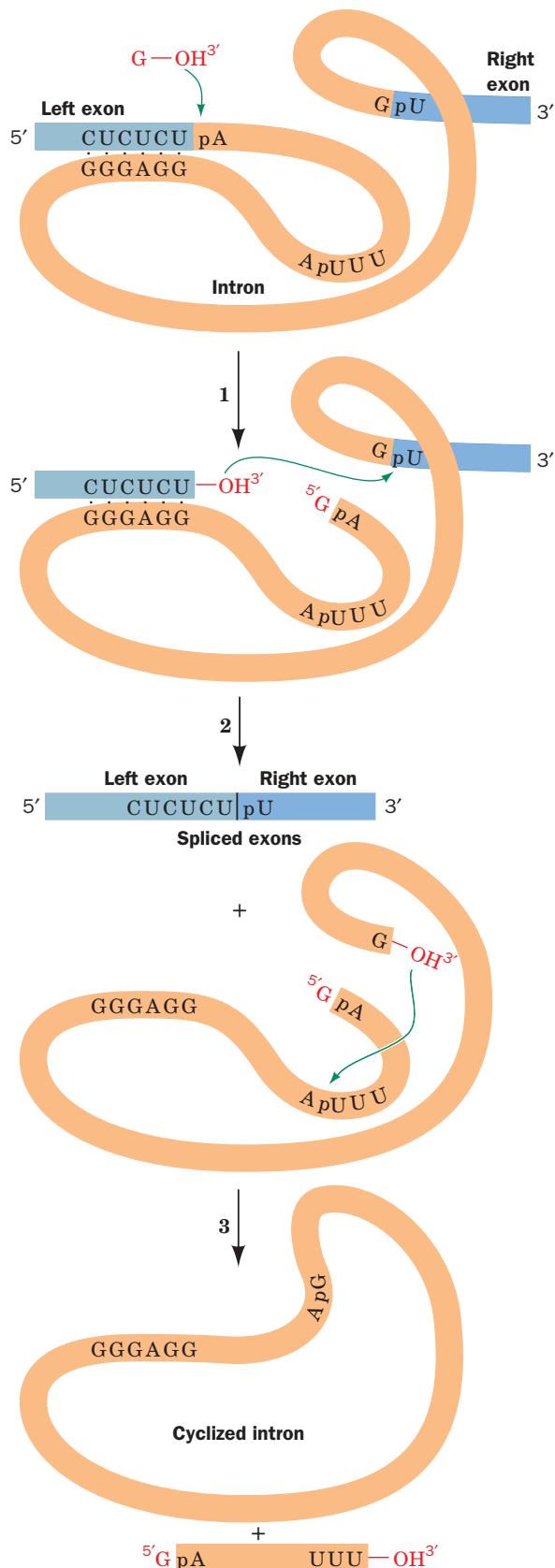


Figure 31-54 The sequence of reactions in the self-splicing of *Tetrahymena* group I intron. (1) The 3'-OH group of a guanine nucleotide attacks the intron's 5'-terminal phosphate so as to form a phosphodiester bond and release the 5' exon. (2) The newly generated 3'-OH group of the 5' exon attacks the 5'-terminal phosphate of the 3' exon, thereby splicing the two exons and releasing the intron. (3) The 3'-OH group of the intron attacks the phosphate of the nucleotide that is 15 residues from the 5' end so as to cyclize the intron and release its 5'-terminal fragment. Throughout this process, the RNA maintains a folded, internally hydrogen bonded conformation that permits the precise excision of the intron.

Tetrahymena intron, which stem from its three-dimensional structure, by demonstrating that it catalyzes the *in vitro* cleavage of poly(C) with an enhancement factor of 10^{10} over the rate of spontaneous hydrolysis. Indeed, this RNA catalyst even exhibits Michaelis–Menten kinetics ($K_M = 42 \mu\text{M}$ and $k_{\text{cat}} = 0.033 \text{ s}^{-1}$ for C₅). Such RNA enzymes have been named **ribozymes**.

Although the idea that an RNA can have enzymatic properties may seem unorthodox, *there is no fundamental reason why an RNA, or any other macromolecule, cannot have catalytic activity* (recall that it was likewise once generally accepted that nucleic acids lack the complexity to carry hereditary information; Section 5-2). Of course, in order to be an efficient catalyst, a macromolecule must be able to assume a stable structure but, as we shall see below and in Sections 32-2B and 32-3Ae, RNAs, including tRNAs and rRNAs, can do so. In fact, the *Tetrahymena* intron undergoes a series of well-defined conformational changes during its reaction sequence. [Synthetic ssDNAs are also known to have catalytic properties although such **deoxyribozymes** are unknown in biology.]

The **group II introns**, which occur in the mitochondria of fungi and plants and comprise the majority of the introns in chloroplasts, are also self-splicing. They generally employ an internal A residue as their initial attacking nucleophile (instead of an external G) to form a lariat intermediate, a process that resembles the splicing of nuclear pre-mRNAs (Fig. 31-53). We shall see below that nuclear pre-mRNA splicing is mediated by complex ribonucleoprotein particles known as **spliceosomes**. The chemical similarities of the pre-mRNA and group II intron splicing reactions therefore suggest that *spliceosomes are ribozymal systems whose RNA components have evolved from primordial self-splicing RNAs and that their protein components serve mainly to fine-tune ribozymal structure and function*. Similarly, the RNA components of ribosomes, which are two-thirds RNA and one-third protein, clearly have a catalytic function in addition to the structural and recognition roles traditionally attributed to them (Section 32-3). Thus, the observations that nucleic acids but not proteins can direct their own synthesis, that cells contain batteries of protein-based enzymes for manipulating DNA but relatively few for processing RNA, and that many coenzymes are ribonucleotides (e.g., ATP, NAD⁺, and CoA), led to the hypothesis that *RNAs were the original biological catalysts in precellular times (the RNA world) and that the chemically more versatile proteins were relative latecomers in macromolecular evolution* (Section 1-5Ca).

f. The X-Ray Structures of a Group I Ribozyme

Group I introns are the most abundant self-splicing introns, with >2000 such sequences known. The sequence of the 413-nt *Tetrahymena* group I intron, together with phylogenetic comparisons, indicates that it contains nine double helical segments that are designated P1 through P9 (Fig. 31-55a; P for base-paired segment). Such analysis further indicates that the conserved catalytic core of group I introns consists of sets of coaxially stacked helices inter-

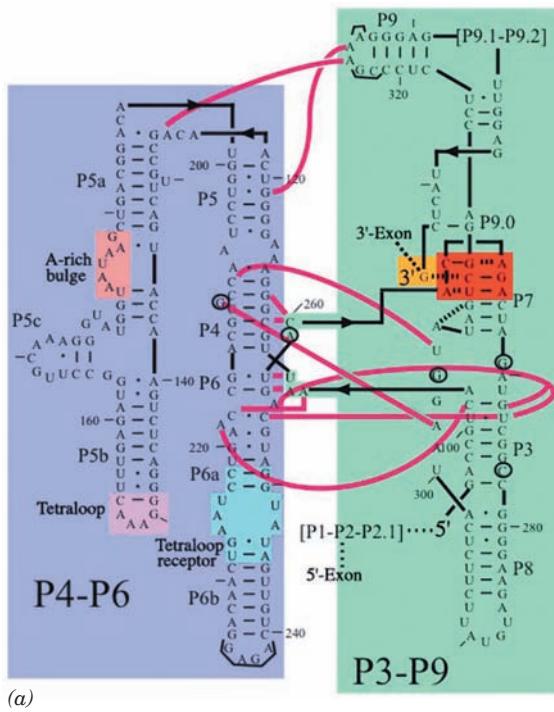
spersed with internal loops that are organized into two domains, the P4-P5-P6 domain (also called P4-P6) and the P3-P7-P8-P9 domain (also called P3-P9).

Cech designed a 247-nt RNA (Fig. 31-55a) that encompasses both the P4-P6 and P3-P9 domains of the *Tetrahymena* group I intron (it lacks the P1-P2 domain and the attached exons), with the addition of a 3' G (ωG), which functions as an internal guanosine nucleophile. This RNA is catalytically active; it binds the P1-P2 domain via tertiary interactions and, with the assistance of ωG , cleaves P1 in a manner similar to the intact intron.

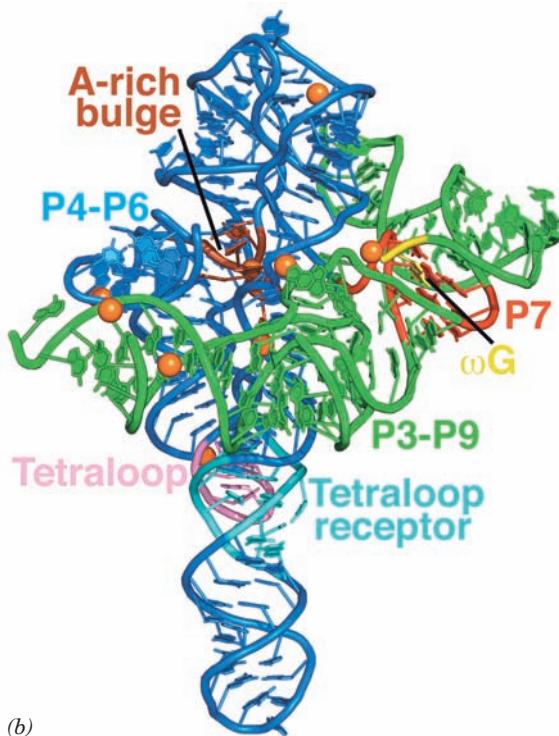
The X-ray structure of this RNA (Fig. 31-55b,c) reveals that it is largely composed of three coaxially stacked sets of A-RNA-like helices with P4-P6 consisting of two pseudo-continuous and straight parallel helices connected by a sharp bend and P3-P9 consisting of a curved helix that wraps around one side of P4-P6 through extensive interdomain interactions that form the ribozyme's active site. Of particular note are its so-called A-rich bulge, a 7-nt sequence about halfway along the short arm of the U-shaped P4-P6, and the 6-nt sequence at the tip of the short arm of the U, whose central GAAA assumes a characteristic conformation known as a **tetraloop**. In both of these substructures, the bases are splayed outward so as to stack on each other and to associate in the minor groove of specific segments of the long arm of the U via hydrogen bonding interactions involving ribose residues as well as bases. In many such interactions, the close packing of phosphate groups is mediated by hydrated Mg²⁺ ions. Throughout this structure, the defining characteristic of RNA, its 2'-OH group, is both a donor and an acceptor of hydrogen bonds to phosphates, bases, and other 2'-OH groups. Interestingly, although this overall fold is highly conserved among group I introns, their sequences are poorly conserved with the exception of a few crucial active site residues.

In the initial reaction catalyzed by group I ribozymes, the 3'-OH group of ωG nucleophilically attacks the phosphate group linking the 5' exon to the ribozyme (Fig. 31-54). But how is only this 3'-OH group activated as a nucleophile? The binding site for the ωG substrate is composed of four coplanar base triples with the ωG –G264–C311 triple sandwiched by three other base triples (Fig. 31-56). Consequently, the base of ωG is stacked between those of A261 and C262 (Fig. 31-56a), which stabilizes the binding of ωG to this site through base stacking.

Divalent metal ions, usually Mg²⁺, are often required for both the structural stability and the catalytic activity of ribozymes. Unfortunately, the relatively low (3.8 Å) resolution of the ribozyme X-ray structure precluded the direct observation of Mg²⁺ ions (which have the same number of electrons as water molecules). However, there is good evidence that the heavy metal ions (Eu and Ir) used to solve the X-ray structure occupied many of the same positions in the ribozyme as do Mg²⁺ ions and moreover, several of these sites were observed to contain Mg²⁺ ions in other group I introns whose X-ray structures are known. In particular, an Mg²⁺ ion is in contact with the 2'-OH group of ωG as well as being liganded by the phosphate groups of

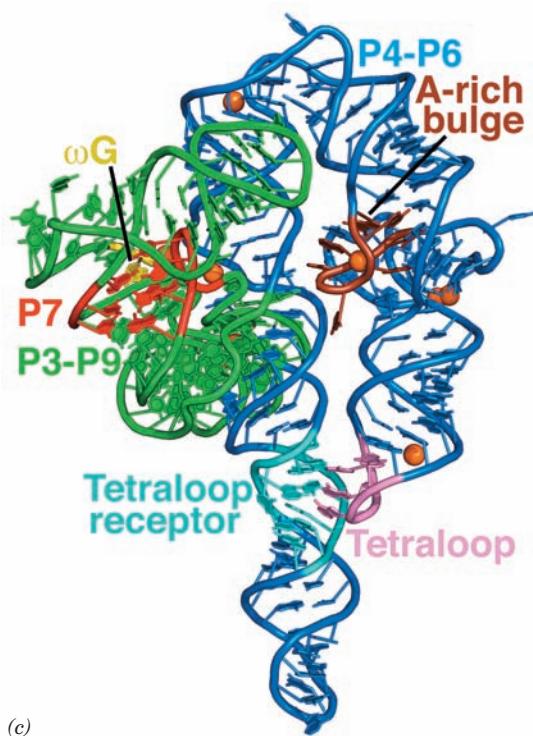


(a)



(b)

Figure 31-55 The group I intron from *Tetrahymena thermophila*. (a) The secondary structure of the 414-nt ribozyme. Its P4-P6 and P3-P9 domains are shaded in blue and green, respectively, with the catalytically active 3' ω G residue shaded in yellow, the base triples of the P7 domain shaded in red, and the A-rich bulge, the tetraloop, and the tetraloop receptor on the P4-P6 domain shaded in brown, pink, and cyan, respectively. Watson-Crick and non-Watson-Crick base pairing interactions are represented by short horizontal lines and small filled circles, whereas interdomain interactions are indicated by magenta lines. Every tenth residue is marked by an outwardly pointing dash. The positions of five residues that have been mutated to stabilize the ribozyme structure are circled and those of the seven mutations that facilitated crystallization are bracketed (this mutant form retains its catalytic activity). (b) The X-ray structure of the ribozyme, drawn in cartoon form with its bases shown as paddles, and colored as in Part a. The inferred positions of Mg^{2+} ions are represented by orange spheres. (c) As in Part b but rotated 140° about the vertical axis to better show the A-rich bulge and the interaction between the tetraloop and the tetraloop receptor. [Part a modified from a drawing by and Parts b and c based on an X-ray structure by Thomas Cech, University of Colorado. PDBid 1X8W.] 



(c)

three surrounding nucleotides. This both orients the ribose group of ω G and nucleophilically activates its 3'-OH group. Biochemical studies indicate that a second Mg^{2+} ion, which accompanies the phosphate group of the RNA

substrate, also participates in the catalytic reaction. Note that two Mg^{2+} ions similarly participate in the phosphoryl-transfer reactions catalyzed by protein enzymes such as DNA polymerase (Section 30-2Af).

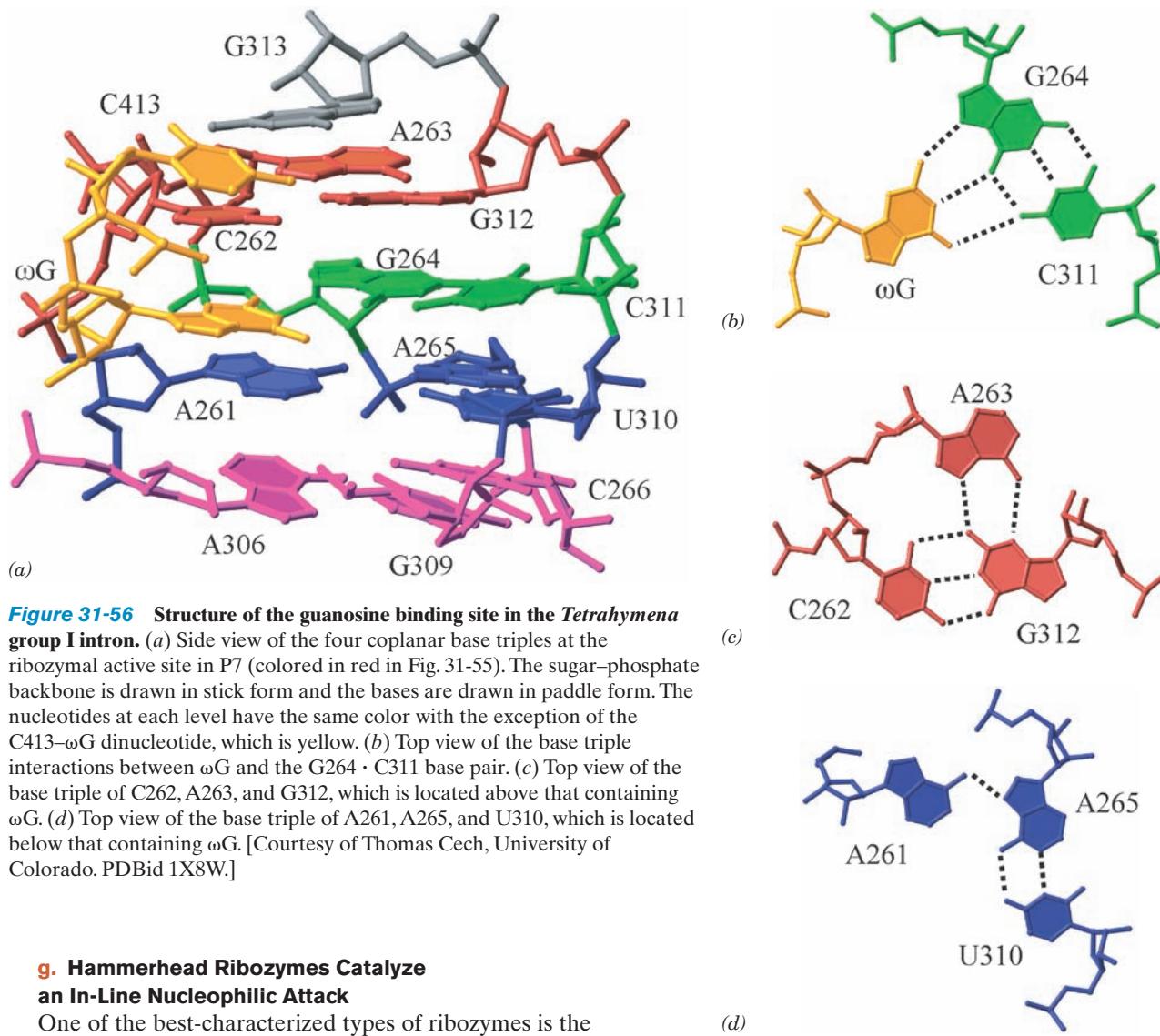


Figure 31-56 Structure of the guanosine binding site in the *Tetrahymena* group I intron. (a) Side view of the four coplanar base triples at the ribozymal active site in P7 (colored in red in Fig. 31-55). The sugar-phosphate backbone is drawn in stick form and the bases are drawn in paddle form. The nucleotides at each level have the same color with the exception of the C413- ω G dinucleotide, which is yellow. (b) Top view of the base triple interactions between ω G and the G264 · C311 base pair. (c) Top view of the base triple of C262, A263, and G312, which is located above that containing ω G. (d) Top view of the base triple of A261, A265, and U310, which is located below that containing ω G. [Courtesy of Thomas Cech, University of Colorado. PDBid 1X8W.]

g. Hammerhead Ribozymes Catalyze an In-Line Nucleophilic Attack

One of the best-characterized types of ribozymes is the **hammerhead ribozyme**, so called because of the superficial resemblance of its secondary structure, as it was originally laid out, to a hammer. This minimally ~40-nt RNA participates in the replication of certain viruslike RNAs that infect plants and also occurs in schistosomes (species of parasitic flatworms). The hammerhead ribozyme catalyzes the site-specific cleavage of one of its own phosphodiester bonds with an ~10⁷-fold rate enhancement. However, it is not a true catalyst because it cannot return to its original state.

The secondary structure of the 63-nt hammerhead ribozyme from *Schistosoma mansoni* has three duplex stems and an active site core of two nonhelical segments (Fig. 31-57a). This ribozyme cleaves itself between its C-17 and C-1.1 nucleotides to yield a cyclic 2',3'-phosphodiester on C-17 with inversion of configuration about the P atom, together with a free 5'-OH on C-1.1, much like the intermediate product in the RNA hydrolysis reaction catalyzed by RNase A (Section 15-1Ab). This suggests that the reaction proceeds via an “in-line” mechanism such as that diagrammed in Fig. 16-6b with the transition state forming a trigonal bipyramidal intermediate in which the attacking

nucleophile, the 2'-OH group (Y in Fig. 16-6b), and the leaving group, which forms the free 5'-OH group (X in Fig. 16-6b), occupy the axial positions.

The X-ray structure of the *S. mansoni* hammerhead ribozyme, determined by William Scott, reveals that its stem II, stem III and the terminal end of stem I are coaxially stacked. The remainder of stem I curves around in a manner that more closely resembles the handle of a suitcase than that of a hammer to form a junction with stems II and III that contains the ribozyme’s active site core (Fig. 31-57b). The nucleotides in the helical stems mainly form normal Watson-Crick base pairs, whereas the nucleotides of the active site core participate in non-Watson-Crick base pairs. This explains the observations that most helical positions can be occupied by any Watson-Crick base pair but that few core bases can be changed without significantly reducing ribozymal activity.

The bases of the active site core participate in a hydrogen bonded network (Fig. 31-57c). This helps position C-17 such that its O2' atom is properly oriented for an in-line

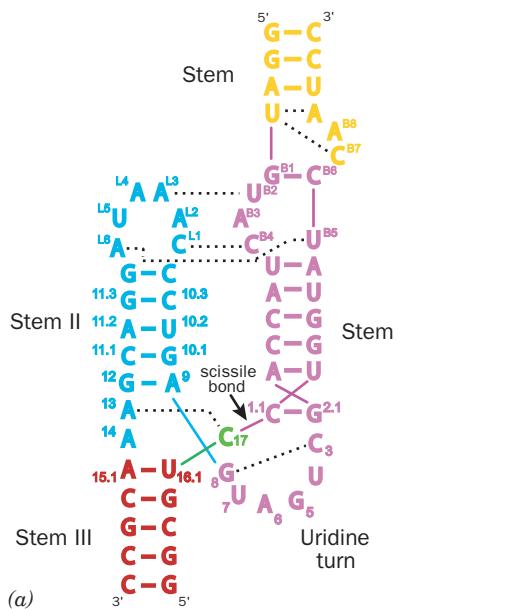
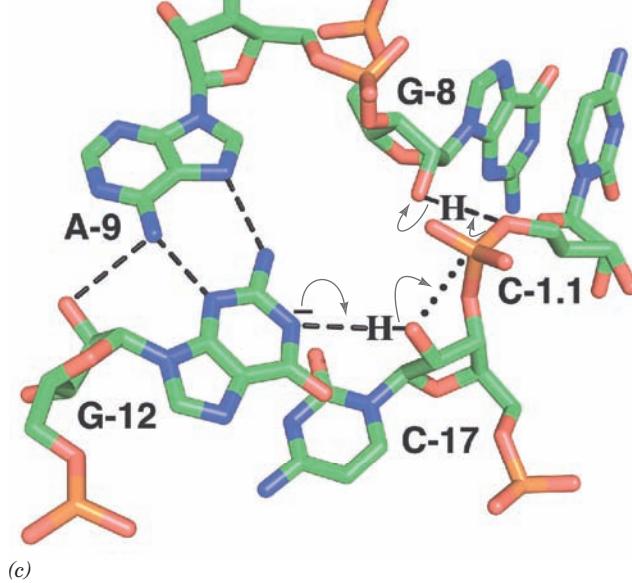
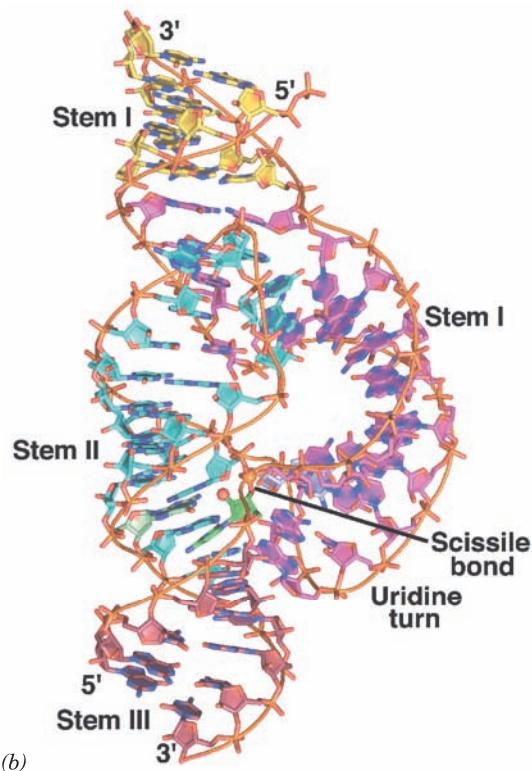


Figure 31-57 Structure of the *Schistosoma mansoni* hammerhead ribozyme. (a) The sequence and schematic structural organization of the ribozyme colored to match the X-ray structure drawn in Part b. Base pairs and tertiary interactions are represented by dashes and dotted lines, respectively. Nucleotides are labeled according to the universal numbering system for hammerhead ribozymes. (b) X-ray structure of the ribozyme drawn in paddle form with C atoms the same color as in Part a except that those of G-12 are light green and those of C-1.1 are light blue, N blue, O red, and P orange. Adjacent P atoms in the same strand are connected by thin orange rods. The P atom of the scissile phosphate group (bridging C-17 and C-1.1) and the nucleophile (O_{2'} of C-17) are represented by small spheres. (c) The ribozyme's active site residues, which are drawn in stick form with C green, O red, N blue, and P orange. Hydrogen bonds are represented by dashed black lines. The proposed reaction mechanism is indicated by the curved arrows with the dotted black line marking the in-line trajectory taken by atom O_{2'} of C-17 in nucleophilically attacking the P atom of the scissile phosphate group. [Part a courtesy of and Parts b and c based on an X-ray structure by William Scott, University of California at Santa Cruz. PDBid 2GOZ.] 



nucleophilic attack on the P atom that links atom O_{3'} of C-17 to atom O_{5'} of C-1.1. The N1 of the invariant G-12 when deprotonated and the 2'-OH of the invariant G-8 appear to be properly positioned to respectively act as base and acid catalysts in this reaction (Fig. 31-57c), which strongly suggests that the reaction occurs via a concerted acid–base catalyzed mechanism. This reaction mechanism does not involve the participation of metal ions and none are observed in the ribozyme's catalytic core. However, in solution, the presence of divalent metal ions provides an ~50-fold enhancement rate over the presence of only monovalent metal ions. Perhaps divalent metal ions stabilize the negative charge on the trigonal bipyramidal intermediate (Fig. 16-6b) and/or they may help position and orient reactive groups.

h. Splicing of Pre-mRNAs Is Mediated by snRNPs in the Spliceosome

How are the splice junctions of pre-mRNAs recognized and how are the two exons to be joined brought together in the splicing process? Part of the answer to this question was established by Joan Steitz going on the assumption that one nucleic acid is best recognized by another. The eukaryotic nucleus, as has been known since the 1960s, contains numerous copies of several highly conserved 60- to 300-nucleotide RNAs called **small nuclear RNAs**

(snRNAs), which form protein complexes termed **small nuclear ribonucleoproteins (snRNPs; pronounced “snurps”)**. Steitz recognized that the 5' end of one of these snRNAs, **U1-snRNA** (so called because it is a member of a U-rich subfamily of snRNAs), is partially complementary to the consensus sequence of the 5' splice site. The consequent hypothesis, that *U1-snRNA recognizes the 5' splice site*, was

corroborated by the observations that splicing is inhibited by the selective destruction of the U1-snRNA sequences that are complementary to the 5' splice site or by the presence of anti-U1-snRNP antibodies (produced by patients suffering from **systemic lupus erythematosus**, an often fatal autoimmune disease). Three other snRNPs are also implicated in splicing: **U2-snRNP**, **U4–U6-snRNP** (in which the **U4-** and **U6-snRNAs** associate via base pairing), and **U5-snRNP**.

*Splicing takes place in an as yet poorly characterized ~2700 kD particle dubbed the **spliceosome*** (Fig. 31-58). The spliceosome brings together a pre-mRNA, the foregoing four snRNPs, and a variety of pre-mRNA binding proteins. Note that the spliceosome, which consists of 5 RNAs and ~150 polypeptides, is comparable in size and complexity to the ribosome (which in *E. coli* consists of 3 RNAs and 52 polypeptides with an aggregate mass of ~2500 kD; Section 32-3A).

In addition to its size and complexity, the spliceosome is a highly dynamic entity, with its various components associating and dissociating during specific stages of the splicing reaction (Fig. 31-59), while undergoing a variety of ATP-driven conformational changes. For example, to carry out the first transesterification reaction yielding the lariat structure (Fig. 31-53), the spliceosome undergoes a complex series of rearrangements that are schematically diagrammed in Fig. 31-60. Similarly extensive rearrangements are required to carry out the second transesterification reaction and to recycle the spliceosome for subsequent splicing reactions.

Although spliceosomal transesterification reactions were initially assumed to be mediated by protein catalysts, their chemical resemblance to the reactions carried out by the self-splicing group II introns suggests, as is noted above, that it is really the snRNAs that catalyze the splicing of pre-mRNAs (pre-mRNA introns have such varied sequences outside of their splice and branch sites that they are unlikely to play an active role in splicing). In fact, James Manley has shown that, in the absence of protein, segments of human U2- and U6-snRNAs catalyze an Mg^{2+} -dependent reaction in an intron branch site sequence-containing RNA that resembles splicing's first transesterification reaction.

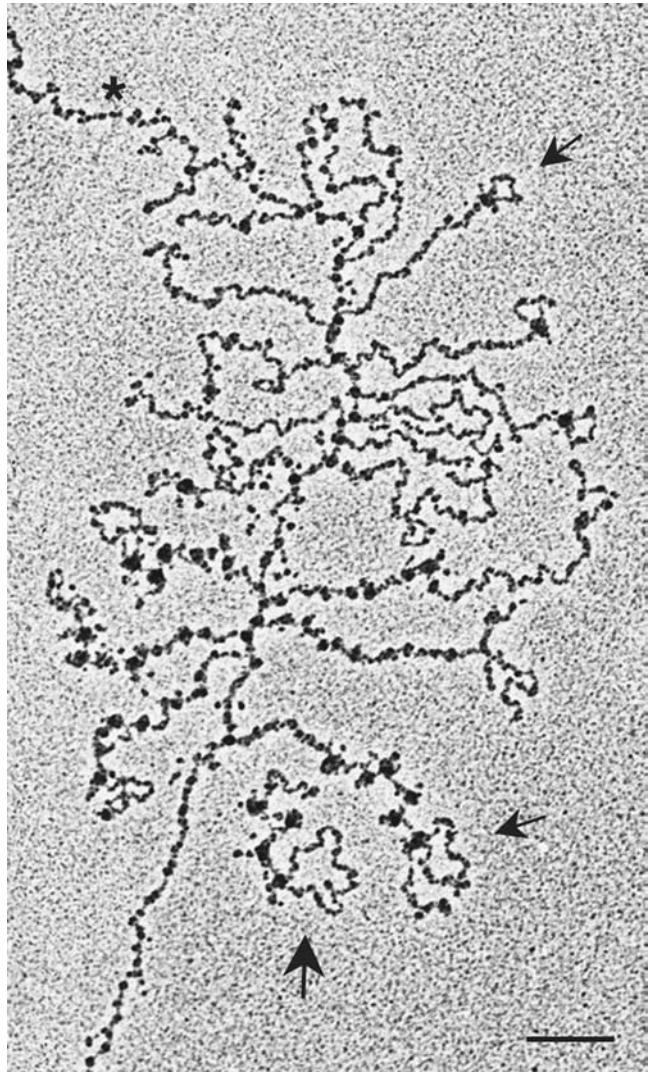


Figure 31-58 An electron micrograph of spliceosomes in action. A *Drosophila* gene that is ~6 kb long enters from the upper left of the micrograph and exits at the lower left. Transcription initiates near the point marked by an asterisk. The growing RNA chains appear as fibrils of increasing lengths that emanate from the DNA. The transcripts are undergoing cotranslational splicing as revealed by the progressive formation and loss of intron loops near the 5' ends of the RNA transcripts (arrows). The beads at the base of each intron loop as well as elsewhere on the transcripts are the spliceosomes. The large arrow points to a transcript near the 3' end of the gene that is no longer attached to the DNA template and hence appears to have recently been terminated and released. The bar is 200 nm long. [Courtesy of Ann Beyer and Yvonne Osheim, University of Virginia.]

i. Splicing Also Requires the Participation of Splicing Factors

Around 170 different proteins known as **splicing-associated factors** that are extrinsic to spliceosomes also participate in splicing, with individual assembly intermediates (e.g., complexes A, B, and C in Fig. 31-59) each associated with ~125 such proteins. Among them are **branch point-binding protein [BBP; also known as splicing factor 1 (SF1)]** and **U2-snRNP auxiliary factor (U2AF)**, which cooperate to select the intron's branch point. U2AF binds to the polypyrimidine tract upstream of the 3' splice site (Fig. 31-52), whereas BBP recognizes the nearby branch point sequence (Figs. 31-53 and 31-60). The NMR structure of the 131-residue RNA-binding segment of the 638-residue BBP in complex with an 11-nt RNA containing a branch

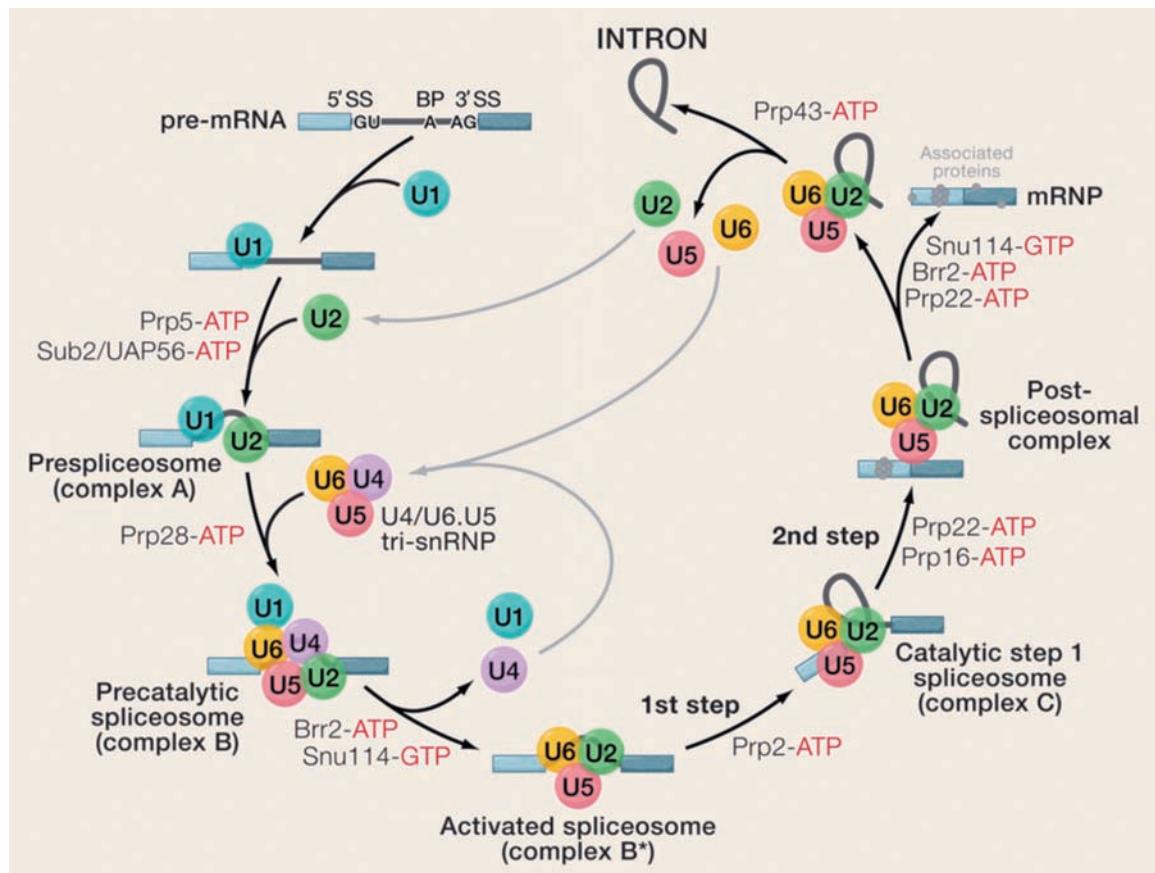


Figure 31-59 The spliceosomal assembly/disassembly cycle.

The sequential actions of the spliceosomal snRNPs (colored circles), but not the non-snRNP proteins, are diagrammed in the process of excising an intron from a pre-RNA containing two exons (blue). Here 5'SS, BP, and 3'SS stand for the pre-mRNA's 5' splice site, its branch point, and its 3' splice site, respectively.

Eight conserved DExD/H box-containing **RNA-dependent RNA ATPases/helicases** as well as the GTPase **Snu114** act in specific steps of the splicing cycle to motivate RNA–RNA rearrangements and RNP remodeling reactions. [Courtesy of Reinhard Lührmann, Max-Planck-Institut für biophysikalische Chemie, Göttingen, Germany.]

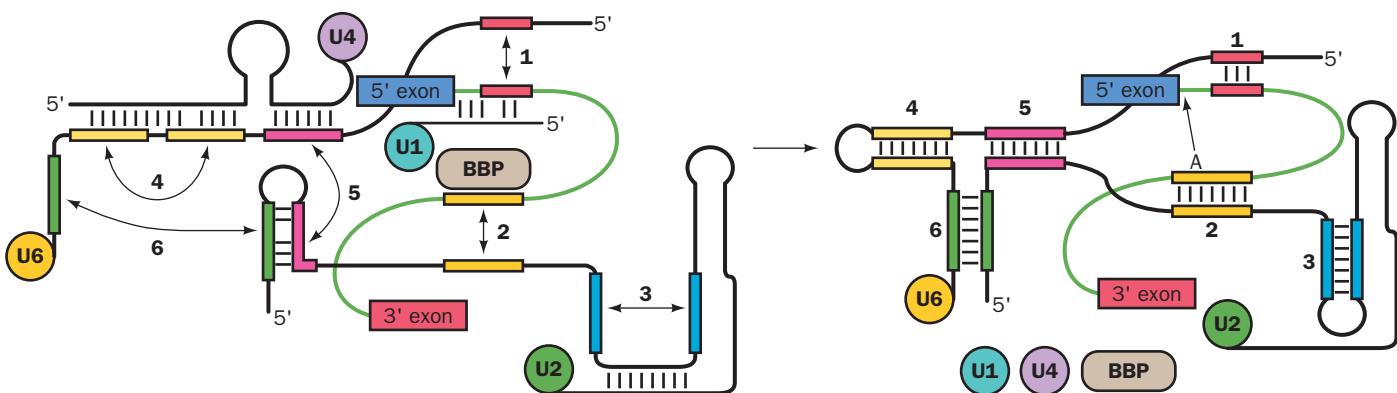


Figure 31-60 Schematic diagram of six rearrangements that the spliceosome undergoes in mediating the first transesterification reaction in pre-mRNA splicing. The RNA is color coded to indicate segments that become base-paired. The black and green lines represent snRNA and pre-mRNA, and BBP stands for branch point-binding protein. U5, which participates in the second transesterification reaction, has been omitted for clarity. (1) Exchange of U1 for U6 in base pairing to the intron's 5' splice site. (2) Exchange of BBP for U2 in binding to the intron's branch site. (3) Intramolecular rearrangement in

U2. (4) Disruption of a base-paired stem between U4 and U6 to form a stem-loop in U6. (5) Disruption of a second stem between U4 and U6 to form a stem between U2 and U6. (6) Disruption of a stem-loop in U2 to form a second stem between U2 and U6. The order of these rearrangements is unclear. The transesterification reaction is represented by the arrow from the A in the yellow segment of the pre-mRNA (right panel) to the 3' end of the 5' exon. [Adapted from Staley, J.P. and Guthrie, C., *Cell* **92**, 315 (1998).]

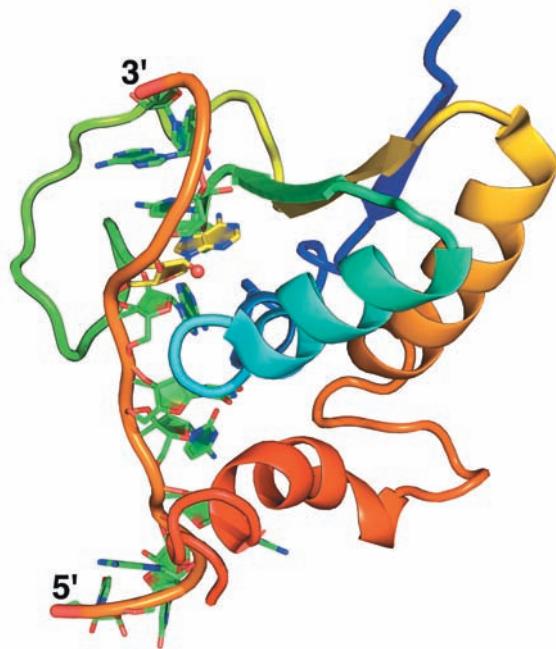


Figure 31-61 The NMR structure of the RNA-binding portion of human branch point-binding protein (BBP) in complex with its target RNA. The 11-nt RNA contains the sequence 5'-UAUACUA**A**CAA-3' in which the branch site sequence for both yeast and vertebrates is underlined and the branch point A is in bold. The protein is drawn in cartoon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The RNA is drawn in paddle form with C green except for the C atoms of the branch point A, which are yellow, N blue, and O red and with successive P atoms connected by an orange rod. The branch point O2' is represented by a small red sphere. [Based on an NMR structure by Michael Sattler, European Molecular Biology Laboratory, Heidelberg, Germany. PDBid 1K1G.]

point sequence, determined by Michael Sattler, reveals that the RNA assumes an extended conformation and is largely buried in a groove that is lined with both aliphatic and basic residues (Fig. 31-61). The branch point adenosine, whose mutation abolishes BBP binding, is deeply buried and binds to BBP via hydrogen bonds that mimic Watson–Crick base pairing with uracil.

Other splicing factors include **SR proteins** and several members of the **heterogeneous nuclear ribonucleoprotein (hnRNP)** family. SR proteins each have one or more RRM (RNA recognition motifs) near their N-terminus and a distinctive C-terminal **RS domain** that contains numerous Ser-Arg (SR) repeats and which participates in protein–protein interactions. SR proteins, when appropriately phosphorylated on their RS domains, specifically bind to their corresponding exonic splicing enhancers (ESEs) via their RRM and thereby recruit the splicing machinery to the flanking 5' and 3' splice sites. The hnRNP proteins, which are highly abundant RNA-binding proteins, lack RS domains and hence cannot recruit the splicing machinery. Instead, they bind to their corresponding ESSs and ISSs (exonic and intronic splicing silencers) so as to block the binding of the splicing machinery at the flanking splice sites.

A simplistic interpretation of Fig. 31-53 suggests that any 5' splice site could be joined with any following 3' splice site, thereby eliminating all the intervening exons together with the introns joining them. However, such **exon skipping** does not normally take place (but see below). Rather, all of a pre-mRNA's introns are individually excised in what appears to be a largely fixed order that more or less proceeds in the 5' → 3' direction. This occurs, at least in part, because splicing takes place cotranscriptionally (Fig. 31-58). Thus, as a newly synthesized exon emerges

from an RNAP II, it is bound by splicing factors that are also bound to the RNAP II's highly phosphorylated C-terminal domain (CTD; Section 31-2E). This tethers the exon and its associated spliceosome to the CTD so as to ensure that splicing occurs when the next exon emerges from the RNAP II.

j. Spliceosomal Structures

All four snRNPs involved in pre-mRNA splicing contain the same so-called **snRNP core protein**, which consists of seven **Sm proteins** (so called because they react with autoantibodies of the Sm serotype from patients with systemic lupus erythematosus), which are named **B, D1, D2, D3, E, F, and G proteins**. Each of these Sm proteins contains two conserved segments, Sm1 and Sm2, that are separated by a linker of variable length. The seven Sm proteins collectively bind to a conserved RNA sequence, the **Sm RNA motif**, which occurs in U1-, U2-, U4-, and U5-snRNAs and which has the single-stranded sequence AAUU-UGUG. However, in the absence of a U-snRNA, the Sm proteins form three stable complexes D1–D2, D3–B, and E–F–G. None of these complexes alone bind U-snRNA. However, the D1–D2 and E–F–G complexes form a stable subcore snRNP with U-snRNA, to which D3–B binds to form the complete **Sm core domain**.

The X-ray structures of the D3–B and D1–D2 heterodimers, determined by Reinhard Lührmann and Kiyoshi Nagai, reveals that these four proteins share a common fold which consists of an N-terminal helix followed by a 5-stranded antiparallel β sheet that is strongly bent so as to form a hydrophobic core (Fig. 31-62a). The subunits of both dimers associate in a similar manner with the β 5 strands of D3 and D1 binding to the β 4 strands of B and D2, respectively, so as to join their β

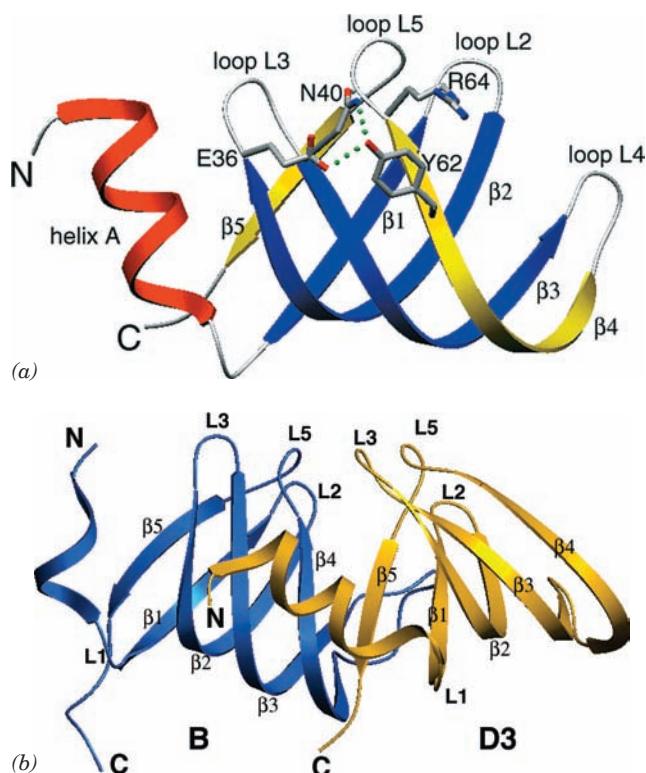


Figure 31-62 X-ray structures of Sm proteins. (a) The structure of D3 protein. The N-terminal helix and the β strands of its Sm1 domain are red and blue and the β strands of its Sm2 domain are yellow. The B, D1, and D2 Sm proteins have similar structures with their L4 loops and N-terminal segments, including helix A, comprising their most variable portions. Several highly conserved residues are shown in stick form (with C gray, N blue, and O red), and a conserved hydrogen bonding network is represented by green dotted lines. (b) The D3-B dimer with D3 gold and B blue. The β 5 strand of D3 associates with the β 4 strand of B to form a continuous antiparallel β sheet. Note that their corresponding loops extend in similar directions. [Courtesy of Kiyoshi Nagai, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1D3B.]

sheets (Fig. 31-62b). This, together with biochemical and mutagenic experiments, indicates that the Sm proteins form a closed heteroheptameric ring whose subunits are arranged in the order –B–D3–G–E–F–D2–D1–. This model is corroborated by the X-ray structure of an Sm-like protein from the hyperthermophilic archeon *Pyrobaculum aerophilum*, determined by David Eisenberg, that forms a homoheptameric ring that is structurally similar to the heteroheptameric model. This structure also supports the hypothesis that the seven eukaryotic Sm proteins arose through a series of duplications of an archaeal Sm-like protein gene.

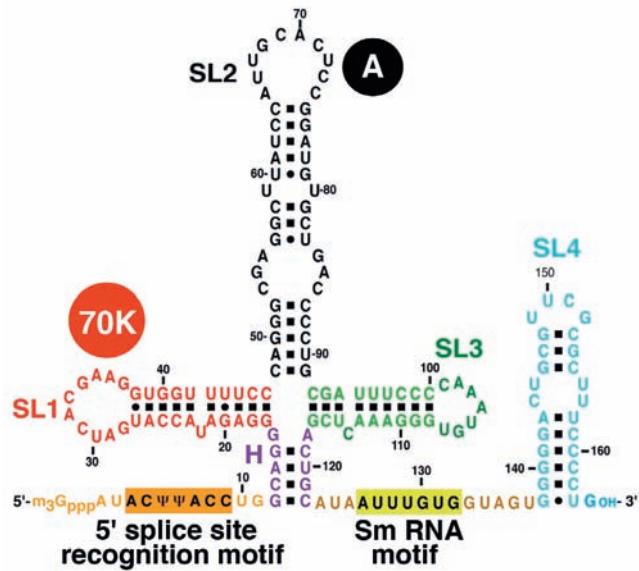
Mammalian U1-snRNP consists of U1-snRNA and ten proteins, the seven Sm proteins that are common to all U-snRNPs as well as three that are specific to U1-snRNP: **U1-70K**, **U1-A**, and **U1-C** (437, 282, and 159 residues, respectively, in humans). The predicted secondary structure of the 165-nt U1-snRNA contains five double helical stems, four of which come together at a 4-way junction (Fig. 31-63a). U1-70K and U1-A bind directly to RNA stem-loops 1 and 2 (SL1 and SL2), respectively, whereas U1-C is bound by other proteins.

Nagai determined the X-ray structure of human U1 snRNP at 5.5 Å resolution. At this low resolution the major and minor grooves of the double-stranded RNA stems are visible. Moreover, the helices and sheets of the proteins are apparent, which allows the placement of protein folds of known structure. SL2 of the U1-snRNA had been altered and shortened to promote crystallization. This eliminated the binding site for U1-A. However, U1 snRNP in which U1-A is depleted is active in a splicing assay.

SL1 and SL2 are stacked coaxially as are SL3 and helix H, and these two stacked helices cross at an angle of $\sim 90^\circ$. The Sm proteins form the predicted heteroheptameric ring, which is ~ 70 Å in diameter (Fig. 31-63b). The Sm RNA motif together with SL4 are threaded through the ring's funnel-shaped central hole such the Sm RNA motif interacts with the Sm proteins (Fig. 31-63c). U1-C associates with the Sm ring via an interaction between its zinc finger domain and the D3 subunit (Fig. 31-63c). An RRM in the C-terminal segment of U1-70K interacts with the loop of SL1 and this protein's N-terminal segment is draped across the external face of the Sm ring.

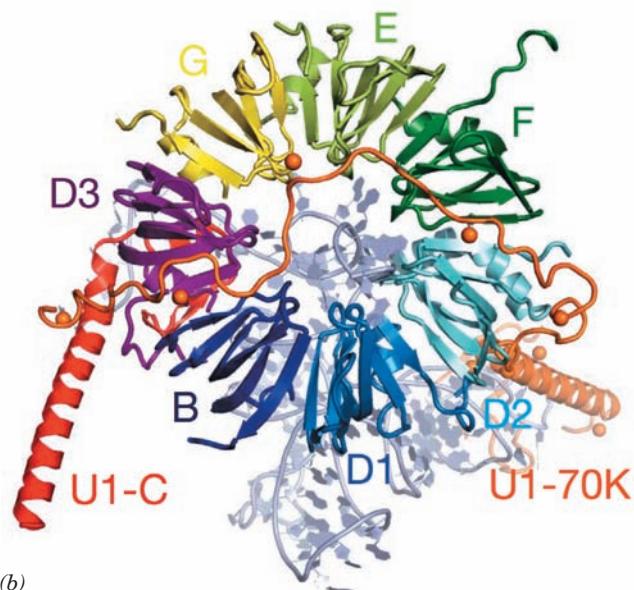
Fortunately, the 5' ends of two neighboring U1-snRNAs in the crystal form a double helical segment. Since, in the spliceosome, the 5' end of U1-snRNA base-pairs with the 5' splice site of the pre-RNA (Fig. 31-60), this interaction provides a model of how the spliceosome recognizes the 5' splice site. The zinc finger domain of U1-C interacts with this double helix (left side of Fig. 31-63c) and presumably stabilizes it. This is consistent with the observation that mutants of U1-C in this region cannot initiate spliceosome formation (the first reaction in Fig. 31-59).

Difficulties in obtaining homogeneous preparations of spliceosomal subassemblies and their poor stabilities have hindered their crystallization and limited the resolution of their cryo-EM-based images. Nevertheless, the cryo-EM-based structures of a variety of spliceosomal components have been reported. For example, the ~ 40 -Å resolution structure of human complex B after it has released U1-snRNP but before it has released U4-snRNP (Fig. 31-59) was determined by Lührmann and Holger

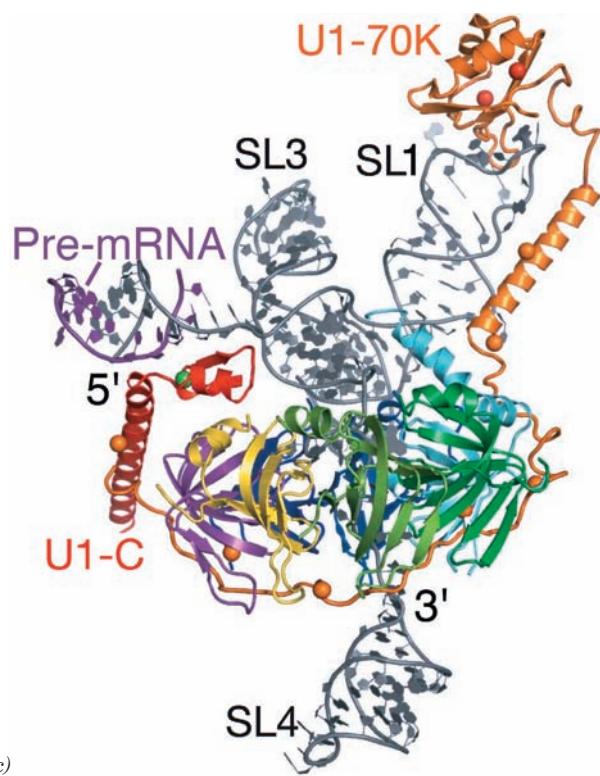


(a)

Figure 31-63 X-ray structure of human U1-snRNP. (a) The predicted secondary structure of U1-snRNA with the RNA segments at which the proteins U1-70K and U1-A bind indicated. Ψ is the symbol for pseudouridine residues (Section 30-5Be). (b) View of the X-ray structure along the axis of the Sm ring. The proteins are drawn in ribbon form in different colors and the RNA is shown in paddle form in gray. SL4 has been deleted for clarity. The yellow spheres represent the selenium atoms in selenoMet residues (Met with its S atom replaced by Se), which were mutagenically inserted into the proteins to aid in solving the structure. Their known positions in the polypeptide chains helped trace the polypeptides' paths. (c) The structure as viewed from above in Part a. The Zn^{2+} ion bound to the zinc finger motif of U1-C is represented by a green sphere. The 5' terminal portion of the U1-snRNA forms a double helical segment with the 5' end of a neighboring U1-snRNA, which is drawn in paddle form in purple. [Part a modified from a drawing by, and Parts b and c courtesy of, Kiyoshi Nagai, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 3CW1.]



(b)



(c)

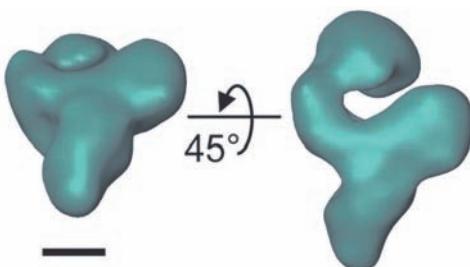


Figure 31-64 Cryo-EM-based structure of the human spliceosome at its $B\Delta U1$ stage of assembly/disassembly at 40 Å resolution. The bar represents 100 Å. [Courtesy of Holger Stark, Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany.]

Stark (Fig. 31-64). This $370 \times 270 \times 170 \text{ \AA}$ particle, which is named $B\Delta U1$, has a roughly triangular body connected to a head domain. Other spliceosome components are equally irregular.

k. The Significance of Gene Splicing

The analysis of the large body of known DNA sequences reveals that introns are rare in prokaryotic structural genes, uncommon in lower eukaryotes such as yeast (which has a total of 239 introns in its ~ 6600 genes and, with two exceptions, only one intron per polypeptide), and abundant in higher eukaryotes (the only known vertebrate structural genes lacking introns are those encoding histones and the antiviral proteins known as interferons). Pre-mRNA introns, as we have seen, can be quite long and many genes contain large numbers of them. Consequently, unexpressed sequences constitute $\sim 80\%$ of a typical vertebrate structural gene and $\sim 99\%$ of a few of them.

The argument that introns are only molecular parasites (**junk DNA**) seems untenable since it would then be difficult to rationalize why the evolution of complex splicing machinery offered any selective advantage over the elimination of the split genes. What then is the function of gene splicing? Although, since its discovery, the significance of gene splicing has been often vehemently debated, two important roles for it have emerged: (1) It is an agent for rapid protein evolution; and (2) through **alternative splicing**, it permits a single gene to encode several (sometimes many) proteins that may have significantly different functions. In the following paragraphs, we discuss these aspects of gene splicing.

I. Many Eukaryotic Proteins Consist of Modules That Also Occur in Other Proteins

The 839-residue LDL receptor is a plasma membrane protein that functions to bind low-density lipoprotein (LDL) to coated pits for transport into the cell via endocytosis (Section 12-5Bc). LDL receptor's 45-kb gene contains 18 exons, most of which encode specific functional domains of the protein. Moreover, 13 of these exons specify polypeptide segments that are homologous to segments in other proteins:

1. Five exons encode a 7-fold repeat of a 40-residue sequence that occurs once in **complement C9** (an immune system protein; Section 35-2F).
2. Three exons each encode a 40-residue repeat similar to that occurring four times in **epidermal growth factor**

(EGF; Section 19-3A) and once each in three blood clotting system proteins: **factor IX**, **factor X**, and **protein C** (Section 35-1).

3. Five exons encode a 400-residue sequence that is 33% identical with a polypeptide segment that is shared only with EGF.

Evidently, the LDL receptor gene is modularly constructed from exons that also encode portions of other proteins. Numerous other eukaryotic proteins are similarly constituted including, as we have seen, many of the proteins involved in signal transduction (e.g., those containing SH2 and SH3 domains; Section 19-3C). Moreover, many exons encode complete domains that frequently have independent functions. *It therefore appears that the genes encoding these modular proteins arose by the stepwise collection of exons that were assembled by (aberrant) recombination between their neighboring introns.*

m. Alternative Splicing Greatly Increases the Number of Proteins Encoded by Eukaryotic Genomes

The expression of numerous cellular genes is modulated by the selection of alternative splice sites. Thus, certain exons in one type of cell may be introns in another. For example, a single rat gene encodes seven tissue-specific isoforms (splice variants) of the muscle protein **α -tropomyosin** (Section 35-3Ca) through the selection of alternative splice sites (Fig. 31-65).

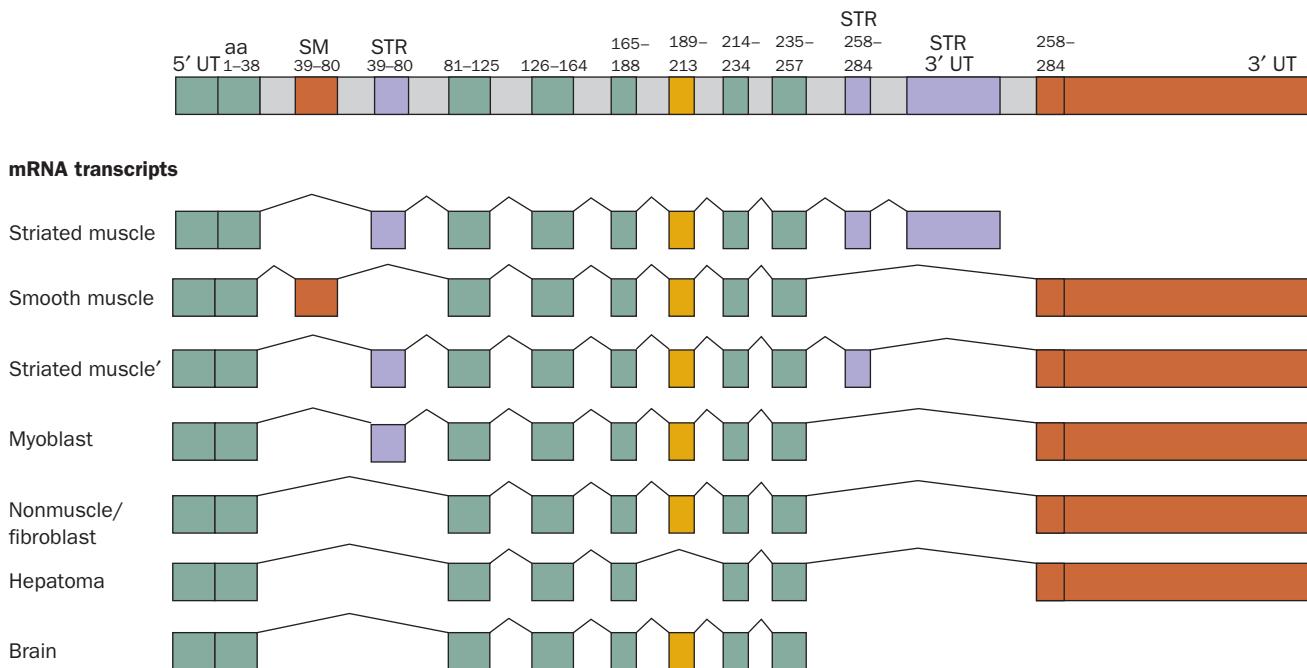


Figure 31-65 The organization of the rat α -tropomyosin gene and the seven alternative splicing pathways that give rise to cell-specific α -tropomyosin isoforms. The thin kinked lines indicate the positions occupied by the introns before they are spliced out to form the mature mRNAs. Tissue-specific exons are indicated together with the amino acid (aa) residues they encode: “constitutive” exons (those expressed in all tissues) are

green, those expressed only in smooth muscle (SM) are brown, those expressed only in striated muscle (STR) are purple, and those variably expressed are yellow. Note that the smooth and striated muscle exons encoding amino acid residues 39 to 80 are mutually exclusive; likewise, there are alternative 3'-untranslated (UT) exons. [After Breitbart, R.E., Andreadis, A., and Nadal-Ginard, B., *Annu. Rev. Biochem.* **56**, 481 (1987).]

Alternative splicing occurs in all metazoa and is especially prevalent in vertebrates. In fact, microarray-based comparisons of the cDNAs obtained from various tissues indicate that ~95% of human structural genes are subject to at least one alternative splicing event. This rationalizes the discrepancy between the ~23,000 genes identified in the human genome (Section 7-2Bc) and earlier estimates that it contains 50,000 to 140,000 structural genes.

The variation in mRNA sequence can take several different forms: Exons can be retained in an mRNA or they can be skipped; introns may be excised or retained; and the positions of 5' and 3' splice sites can be shifted to make exons shorter or longer. Alterations in the transcriptional start site and/or the polyadenylation site can further contribute to the diversity of the mRNAs that are transcribed from a single gene. In a particularly striking example, the *Drosophila* protein **Dscam** (for *Down syndrome cell-adhesion molecule*), which functions in neuronal development, is encoded by 24 exons of which there are 12 mutually exclusive variants of exon 4, 48 of exon 6, 33 of exon 9, and 2 of exon 17 (which are therefore known as **cassette exons**) for a total of 38,016 possible isoforms of this protein (compared to ~14,000 identified genes in the *Drosophila* genome). Although it is unknown if all possible Dscam isoforms are produced, experimental evidence suggests that the *Dscam* gene expresses many thousands of them. [Dscam is a membrane-anchored cell-surface protein of the immunoglobulin superfamily. The specific isoform expressed in a given neuron binds to itself but rarely to other isoforms. This permits the neuron to distinguish its own processes (axons and dendrites) from those of other neurons and thereby plays an essential role in neural patterning. However, the precise identity of a given isoform appears to be unimportant.] Clearly, the number of genes in an organism's genome does not by itself provide an adequate assessment of its protein diversity. Indeed, it has been estimated that, on average, each human structural gene encodes three different proteins.

The types of changes that alternative splicing confers on expressed proteins spans the entire spectrum of protein properties and functions. Entire functional domains or even single amino acid residues may be inserted into or deleted from a protein, and the insertion of a stop codon may truncate the expressed polypeptide. Splice variations may, for example, control whether a protein is soluble or membrane bound, whether it is phosphorylated by a specific kinase, the subcellular location to which it is targeted, whether an enzyme binds a particular allosteric effector, and the affinity with which a receptor binds a ligand. Changes in an mRNA, particularly in its noncoding regions, may also influence the rate at which it is transcribed and its susceptibility to degradation. Since the selection of alternative splice sites is both tissue- and developmental stage-specific, splice site choice must be tightly regulated in both space and time. In fact, it is estimated that from ~15% to 50% of human genetic diseases are caused by point mutations that result in pre-mRNA splicing defects. Some of these mutations delete functional splice sites, thereby activating nearby pre-existing **cryptic splice sites**. Others generate

new splice sites that are used instead of the normal ones and yet others are in the genes encoding components of the splicing machinery. In addition, tumor progression is correlated with changes in levels of proteins implicated in alternative splice site selection.

How are alternative splice sites selected? Well-understood examples of such processes occur in the pathway responsible for sex determination in *Drosophila*, two of which we discuss here:

1. Exon 2 of *transformer (tra)* pre-mRNA contains two alternative 3' splice sites (which succeed the excised intron), with the proximal (close to exon 1) site used in males and the distal (far) site used in females (Fig. 31-66a). The region between these two sites contains a Stop codon (UAG). In males, the splicing factor U2AF binds to the proximal 3' splice site to yield an mRNA containing this premature stop codon, which thereby directs the synthesis of truncated and hence nonfunctional **TRA** protein. In females, however, the proximal 3' splice site is bound by the female-specific **SXL** protein, the product of the *sex-lethal (sxl)* gene (which is only expressed in females), so as to block the binding of U2AF, which then binds to the distal 3' splice site, thereby excising the UAG and inducing the expression of functional TRA protein (U2AF and TRA both contain RS domains but not RRM so that neither is an SR protein).

2. In *doublesex (dsx)* pre-mRNA, the first three exons are constitutively spliced in both males and females. However, the branch site immediately upstream of exon 4 has a suboptimal pyrimidine tract to which U2AF does not bind (Fig. 31-66b). Hence in males, exon 4 is not included in *dsx* mRNA, leading to the synthesis of male-specific **DSX-M** protein that functions as a repressor of female-specific genes. However, in females, TRA protein promotes the cooperative binding of the SR protein **RBP1** and the SR-like protein **TRA2** [the product of the *transformer 2 (tra-2)* gene] to six copies of an exonic splice enhancer (ESE) within exon 4. This heterotrimeric complex recruits the splicing machinery to the upstream 3' splice site of exon 4, leading to its inclusion in *dsx* mRNA. The resulting female-specific **DSX-F** protein is a repressor of male-specific genes.

Thus, the synthesis of functional TRA protein involves the repression of a splice site, whereas the synthesis of female-specific DSX-F protein involves the activation of a splice site. Similar mechanisms of alternative splice site selection have been identified in vertebrates.

In general, the decision as to whether an alternative exon is kept or eliminated is determined by the activities and concentrations of its various regulators, many of which are SR proteins and hnRNPs. Hence the tissue-specific expression of these regulators and the phosphorylation state of the SR proteins are important contributors to the complex regulation of mRNA splicing. Moreover, extensive analysis of the sequences of numerous alternative splice sites has revealed the existence of a "splicing code" that uses combinations of over 200 RNA features that are

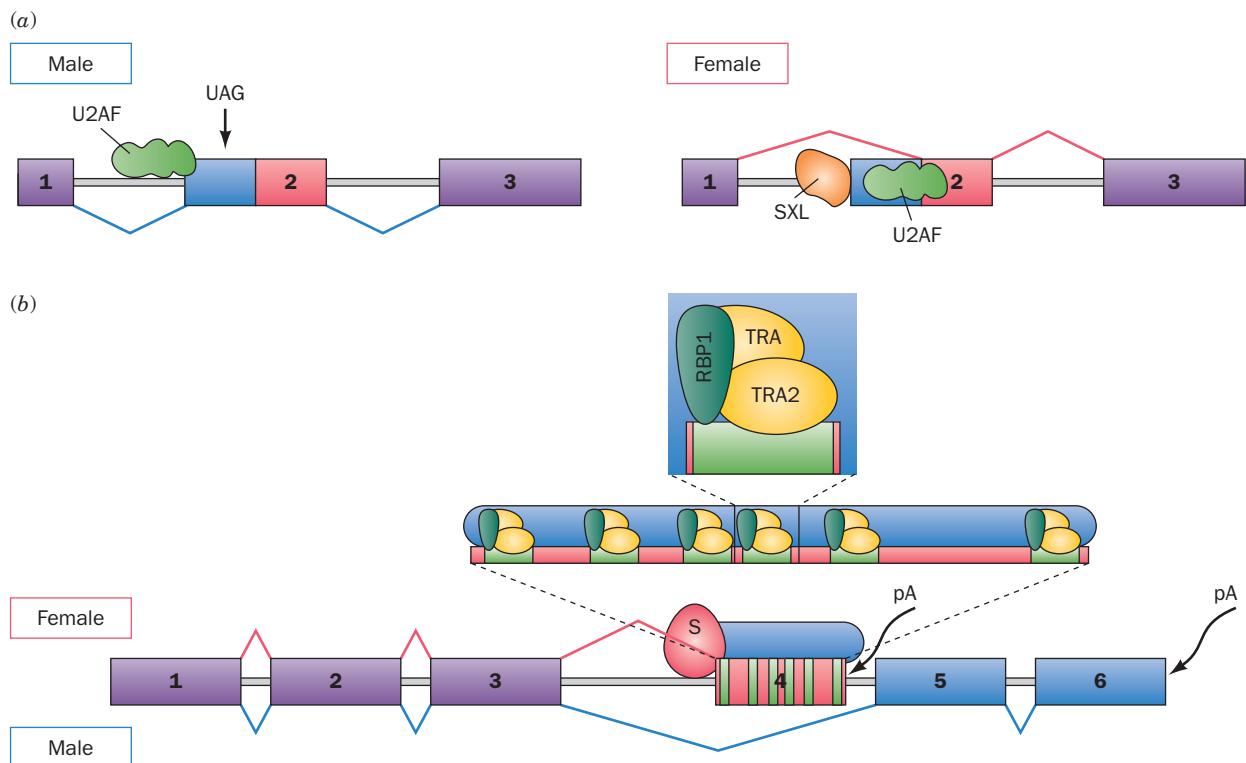


Figure 31-66 Mechanisms of alternative splice site selection in the *Drosophila* sex-determination pathway as described in the text. In all panels, exons are represented by colored rectangles and introns are shown as pale gray lines. (a) Alternative splicing in *tra* pre-mRNA. UAG is a Stop codon. (b) Alternative splicing in *dsx* pre-mRNA. The six ESEs (exonic splice enhancers) in

present in both introns and exons and which are recognized by the foregoing regulators.

Riboswitches (Section 31-3H) have been implicated in controlling alternative splicing in eukaryotes. For example, as Breaker has shown, in the *NMT1* gene of the bread mold *Neurospora crassa* [which expresses an enzyme that participates in the metabolism of TPP (thiamine pyrophosphate)], a TPP-sensing riboswitch is contained in an intron that is located upstream of the mRNA's normal AUG translational start codon. This intron contains one 3' splice site and two 5' splice sites with two AUG start codons between them. When the concentration of TPP is low so that it does not bind to the riboswitch, one strand of the riboswitch's P4-P5 segment base-pairs with the second (downstream) splice site so as to inactivate it (Fig. 31-67a). The spliceosome then efficiently excises the entire intron yielding an mRNA that is readily translated (I-3; Fig. 31-67b). However, when the concentration of TPP is high so that it binds to the riboswitch, the riboswitch assumes a conformation that activates the second splice site but occludes the branch point A (Fig. 31-67c). Consequently, the spliceosome inefficiently excises only the downstream portion of the intron. The two upstream AUG codons, which are present in both the unspliced mRNA (I-1) and in the spliced mRNA containing only the upstream portion of the intron (I-2), compete for ribosomes with mRNA's normal AUG

exon 4 are indicated by green rectangles and S represents the splicing machinery. In females, polyadenylation (pA) of *dsx* mRNA occurs downstream of exon 4, whereas in males, it occurs downstream of exon 6. [After a drawing by Maniatis, T. and Tasic, B., *Nature* **418**, 236 (2002).]

start codon and thereby repress the mRNA's translation (Fig. 31-67c).

n. AU-AC Introns Are Excised by a Novel Spliceosome

A small fraction of introns (~0.3%) have AU rather than GU at their 5' ends and AC rather than AG at their 3' ends, but are nevertheless excised via a lariat structure to an internal intron A. These so-called **AU-AC introns** (alternatively, **AT-AC introns** after their DNA sequences), which occur in organisms as diverse as *Drosophila*, plants, and humans, are excised by a novel so-called **AU-AC spliceosome** (alternatively, an **AT-AC spliceosome**) that has one snRNP, U5, in common with the major (GU-AG) spliceosome, and three others, **U11**, **U12**, and **U4atac-U6atac**, which are distinct from but structurally and functionally analogous to U1, U2, and U4-U6. Curiously, all genes known to contain AU-AC introns also contain multiple major class introns. Moreover, AU-AC introns are not conserved in either length or position in their host genes. Thus, the functional and evolutionary significance of the AU-AC spliceosome and introns is obscure.

o. Trans-Splicing

The types of splicing we have so far considered occur within single RNA molecules and hence are known as

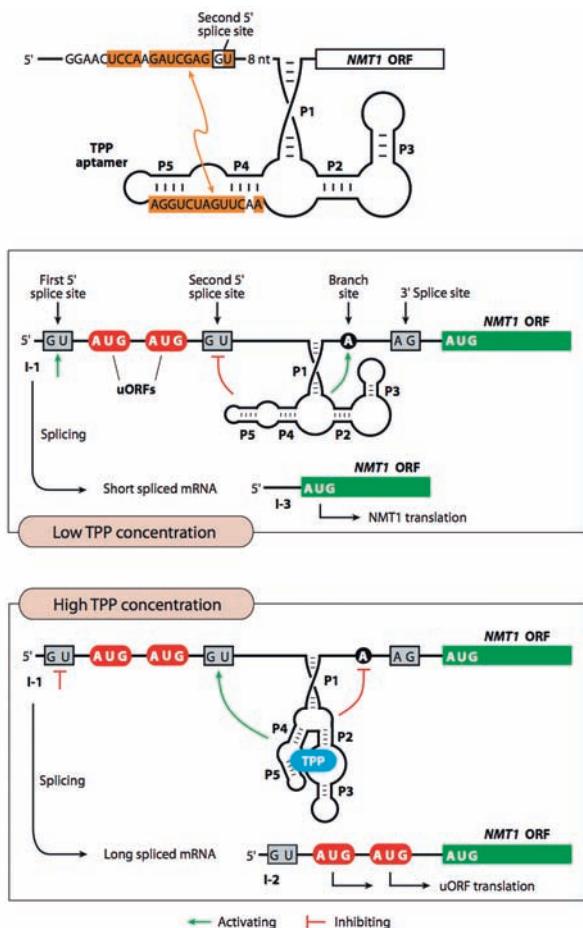


Figure 31-67 Control of translation by the *N. crassa*

TPP-sensing riboswitch through alternative splicing. (a) The predicted secondary structure of the TPP aptamer, which resides in the 5' untranslated region of the *NMT1* mRNA. The sequence of one strand of the P4-P5 stem is complementary (orange shading) to a segment that overlaps the second 5' splice site. (b) At low TPP concentrations, the aptamer inhibits (red tee) splicing from the second 5' splice site while activating (green arrow) the branch point A so that the spliceosome excises the RNA between the first 5' splice site and the 3' splice site, yielding an ORF (open reading frame; I-3) that is normally translated from its AUG start codon. (c) At high TPP concentrations, the binding of TPP to the aptamer activates the second 5' splice site but occludes the branch point A. The spliceosome therefore inefficiently excises the intron's downstream portion yielding an mRNA (I-2) that contains two uORFs (upstream ORFs) that compete with the translation of the primary ORF. The unspliced mRNA (I-1) also contains the two uORFs and hence is likewise inefficiently translated. [Courtesy of Ronald Breaker, Yale University.]

cis-splicing. The chemistry of the spliceosomal cis-splicing reaction, however, is the same as would occur if the two exons to be joined initially resided on two different RNA molecules, a process called **trans-splicing**. This, in fact, occurs in trypanosomes (kinetoplastid protozoa; the cause of African sleeping sickness). Trypanosomal mRNAs all have the same 35-nt noncoding leader sequence, although this leader sequence is not present in the corresponding genes. Rather, this sequence is part of a so-called **spliced leader (SL) RNA** that is transcribed from an independent gene. The 5' splice site that succeeds the SL RNA leader sequence, and the branch site and 3' splice site that precede the exon sequence have the same consensus sequences as occur in the RNAs spliced by the major spliceosome. Consequently, the SL RNA leader and the pre-mRNA are joined in a trans-splicing reaction that resembles the spliceosomal cis-splicing reaction (Fig. 31-53) with the exception that the product of the first transesterification reaction is necessarily Y-shaped rather than lariat-shaped (Fig. 31-68). Trypanosomes, whose pre-mRNAs lack introns, nevertheless have U2- and U4–U6-snRNPs but lack U1- and U5-snRNPs. However, the SL RNA, which is predicted to fold into three stem-loops and a single-stranded Sm RNA-like motif as does U1-snRNA (Fig. 31-63a), apparently carries out the functions of U1-snRNA in the trans-splicing reaction.

Trans-splicing has been shown to occur in nematodes (roundworms; e.g., *C. elegans*) and flatworms. These organisms also carry out cis-splicing and, indeed, perform both types of splicing on the same pre-mRNA. There are also several reports that trans-splicing occurs in higher eukaryotes such as *Drosophila* and vertebrates, but if it does occur, it does so in only a few pre-mRNAs and at a very low level.

p. mRNA Is Methylated at Certain Adenylate Residues

During or shortly after the synthesis of vertebrate pre-mRNAs, ~0.1% of their A residues are methylated at their N6 atoms. These m⁶A's tend to occur in the sequence RRM⁶ACX, where X is rarely G. Although the functional significance of these methylated A's is unknown, it should be noted that a large fraction of them are components of the corresponding mature mRNAs.

q. RNA Can Be Edited by the Insertion or Deletion of Specific Nucleotides

Certain mRNAs from a variety of eukaryotic organisms have been found to differ from their corresponding genes in several unexpected ways, including C → U and U → C changes, the insertion or deletion of U residues, and the insertion of multiple G or C residues. The most extreme examples of this phenomenon, which occur in the mitochon-

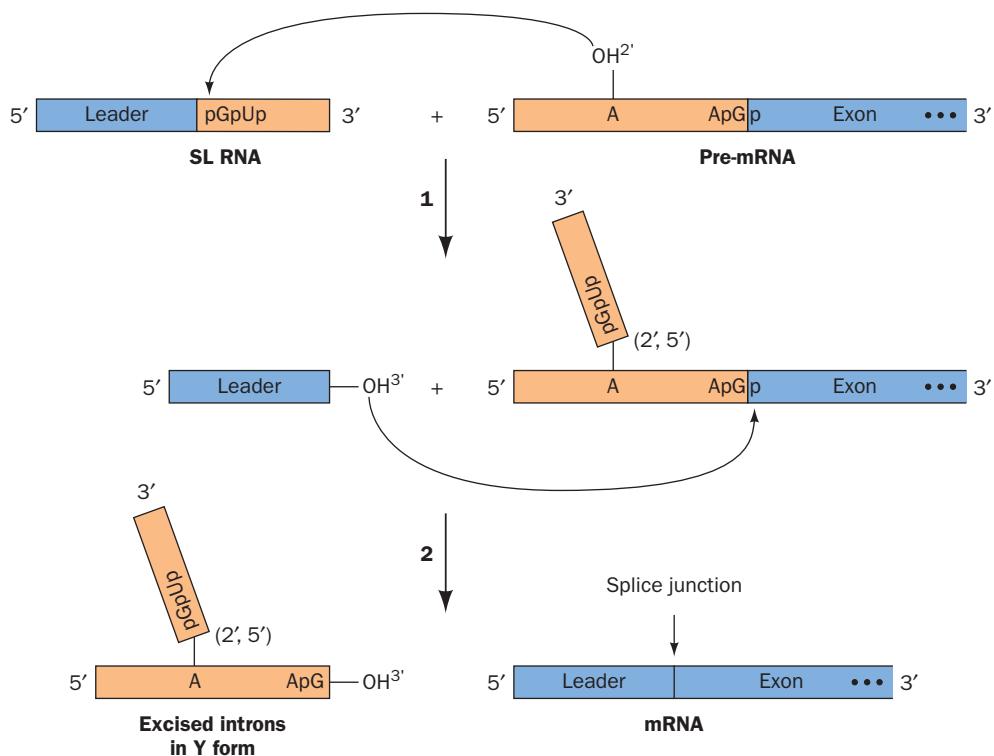


Figure 31-68 The sequence of transesterification reaction that occurs in trans-splicing. The chemistry is closely similar to that of pre-mRNA cis-splicing (Fig. 31-53).

dria of trypanosomes (whose DNA encodes only 20 genes), involve the addition and removal of up to hundreds of U's to and from 12 otherwise untranslatable mRNAs. The process whereby a transcript is altered in this manner is called **RNA editing** because it originally seemed that the required enzymatic reactions occurred without the direction of a nucleic acid template and hence violated the central dogma of molecular biology (Fig. 5-21). Eventually, however, a new class of trypanosomal mitochondrial transcripts called **guide RNAs (gRNAs)** was identified. gRNAs, which consist of 40 to 80 nucleotides, have 3' oligo(U) tails, an internal segment that is precisely complementary

to the edited portion of the pre-edited mRNA (if G · U pairs, which are common in RNAs, are taken to be complementary), and a 10- to 15-nt so-called anchor sequence near the 5' end that is largely complementary in the Watson-Crick sense to a segment of the mRNA that is not edited.

An unedited transcript presumably associates with the corresponding gRNA via its anchor sequence (Fig. 31-69). Then, in a process mediated by the appropriate enzymatic machinery in an ~20S RNP named the **editosome**, the gRNA's internal segment is used as a template to “correct” the transcript, thereby yielding the edited mRNA. Inser-

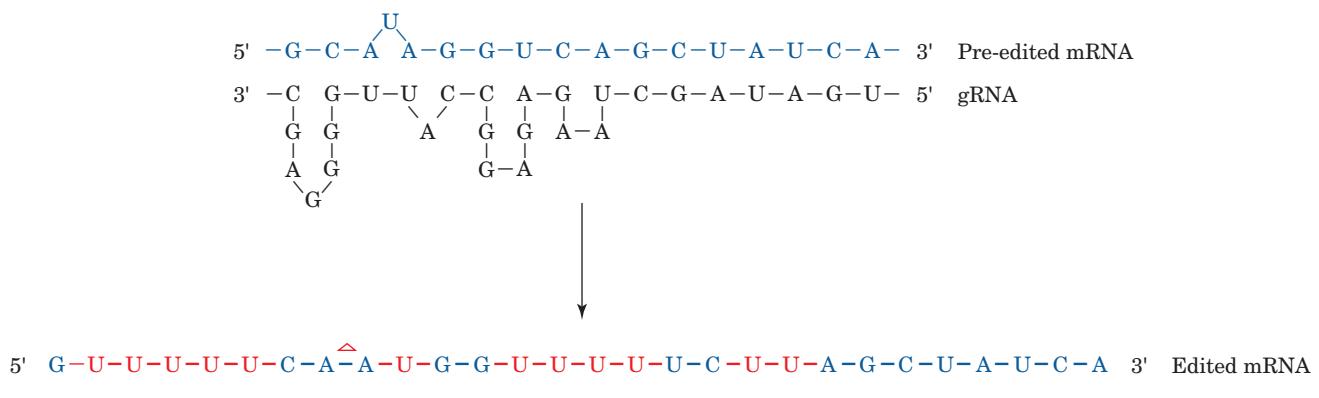


Figure 31-69 A schematic diagram indicating how gRNAs direct the editing of trypanosomal pre-edited mRNAs. The red U's in the edited mRNA are insertions and the triangle (Δ) marks a deletion. Several gRNAs may be necessary to direct the

editing of consecutive segments of a pre-edited mRNA. [After Bass, B.L., in Gesteland, R.F. and Atkins, J.F. (Eds.), *The RNA World*, p. 387, Cold Spring Harbor Laboratory Press (1993).]

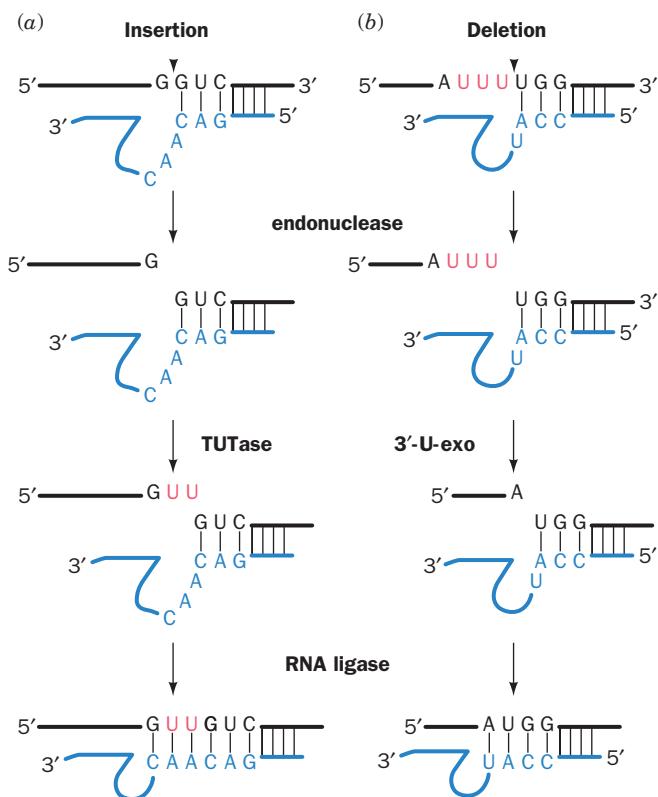


Figure 31-70 Trypanosomal RNA editing pathways. The RNAs being edited (black) are shown base-paired to the gRNAs (blue) with the U's that are (a) inserted by TUTase or (b) deleted by 3'-U-exo drawn in red. The arrowheads indicate the positions that are cleaved by the endonuclease. [After Madison-Antenucci, S., Grams, J., and Hajduk, S.L., *Cell* **108**, 435 (2002).]

tion editing requires at least three enzymatic activities that, somewhat surprisingly, are encoded by nuclear genes (Fig. 31-70a): (1) an endonuclease at a mismatch between the gRNA and the pre-edited mRNA to cleave the pre-edited mRNA on the 5' side of the insertion point; (2) **terminal uridyltransferase (TUTase)** to insert the new U(s); and (3) an **RNA ligase** to reseal the RNA. Deletion requires similar enzymatic apparatus with the exceptions that the endonuclease cleaves the RNA being edited on the 3' side of the U(s) to be deleted and TUTase is replaced by **3'-U-exonuclease (3'-U-exo)**, which excises the U(s) at the deletion site (Fig. 31-70b). A single gRNA mediates the editing of a block of 1 to 10 sites. Thus, the genetic information specifying an edited mRNA is derived from two or more genes. The functional advantage of this complicated process, either presently or more likely in some ancestral organism, is obscure.

r. RNA Can Be Edited by Base Deamination

Humans express two forms of **apolipoprotein B (apoB)**: **apoB-48**, which is made only in the small intestine and functions in chylomicrons to transport triacylglycerols from the intestine to the liver and peripheral tissues; and **apoB-100**, which is made only in the liver and functions in VLDL, IDL, and LDL to transport cholesterol from the

liver to the peripheral tissues (Sections 12-5A and 12-5B). ApoB-100 is an enormous 4536-residue protein, whereas apoB-48 consists of apoB-100's N-terminal 2152 residues and therefore lacks the C-terminal domain of apoB-100 that mediates LDL receptor binding.

Despite their differences, both apoB-48 and apoB-100 are expressed from the same gene. How does this occur? Comparison of the mRNAs encoding the two proteins indicates that they differ by a single C → U change: The codon for Gln 2153 (CAA) in apoB-100 mRNA is, in apoB-48 mRNA, a UAA Stop codon. The activity that catalyzes this conversion is a protein: It is destroyed by proteases and protein-specific reagents but not by nucleases. When apoB mRNA is synthesized with $[\alpha-^{32}\text{P}]$ CTP, *in vitro* editing yields a $[\text{P}^{32}]$ UMP residue solely at the editing site. Evidently, the editing activity is a site-specific **cytidine deaminase**. This type of RNA editing differs in character from that in trypanosomal mitochondria, which inserts and deletes multiple U's into mRNAs under the direction of gRNAs. ApoB mRNA editing therefore falls into a different class of RNA editing that is called **substitutional editing**.

The several other known examples of pre-mRNA substitutional editing all occur on pre-mRNAs that encode ion channels and G protein-coupled receptors in nerve tissue. Among them is vertebrate brain **glutamate receptor** pre-mRNA, which undergoes an A → I deamination [where I is inosine (guanosine lacking its 2-amino group), which the translational apparatus reads as G] that transforms a Gln codon (CAG) to that of a functionally important Arg (CIG; normally CGG). The vertebrate enzymes that catalyze such A → I RNA editing of pre-mRNAs, **ADAR1** (1200 residues), **ADAR2** (729 residues), and **ADAR3** (739 residues; ADAR for adenosine deaminases acting on RNA), have the curious requirement that their target A residues must be members of RNA double helices that are formed between the editing site and a complementary sequence that is usually located in a downstream intron (Fig. 31-71). Hence, ADAR-mediated editing must precede splicing.

Substitutional editing may contribute to protein diversity. For example, *Drosophila cacophony* pre-mRNA that encodes a voltage-gated Ca^{2+} channel subunit contains 10 different substitutional editing sites and hence has the potential of generating 1000 different isoforms in the absence of alternative splicing.

Substitutional editing can also generate alternative splice sites. For example, rat ADAR2 edits its own pre-mRNA by converting an intronic AA dinucleotide to AI, which mimics the AG normally found at 3' splice sites (Fig. 31-53). The consequent new splice site adds 47 nucleotides near the 5' end of the ADAR2 mRNA so as to generate a new translational initiation site. The resulting ADAR2 isozyme is catalytically active but is produced in smaller amounts than that from unedited transcripts, perhaps due to a less efficient translational initiation site. Thus, rat ADAR2 appears to regulate its own rate of expression.

ADAR1 contains an N-terminal Z-DNA-binding domain, Zab, that is composed of two subdomains, $Z\alpha$ and $Z\beta$. We have seen that in the X-ray structure of $Z\alpha$ in complex with Z-DNA (Fig. 29-3), $Z\alpha$ binds Z-DNA via

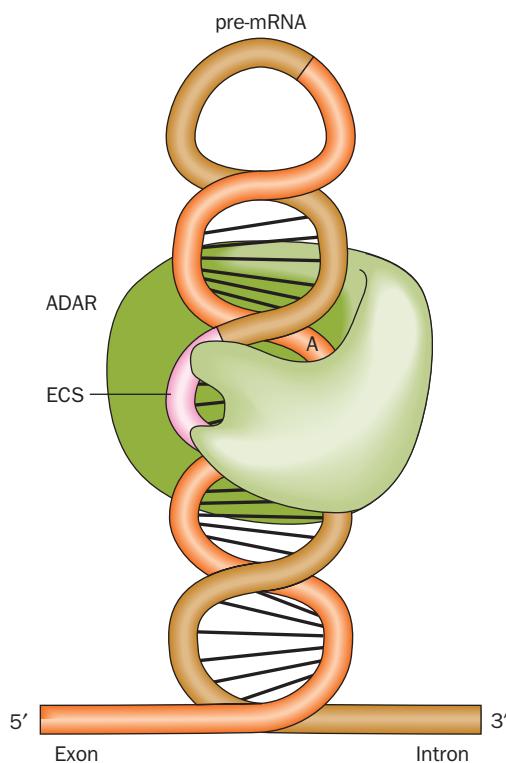


Figure 31-71 The recognition of ADAR editing sites. Both ADAR1 and ADAR2 bind to a 9- to 15-bp double-stranded RNA that is formed between the editing site (orange) on a pre-mRNA exon and a so-called **editing site complementary sequence** (ECS; pink) that is often located in a downstream intron (brown). A represents the adenosine that the ADAR (green) converts to inosine. [After Keegan, L.P., Gallo, A., and O'Connell, M.A., *Nature Rev. Genet.* **2**, 869 (2001).]

sequence-independent complementary surfaces (Section 29-1Bb). What is the function of Zab? Alexander Rich has proposed that since the negative supercoiling of the DNA immediately behind actively transcribing RNAP (Section 31-2Ca) stimulates the transient formation of Z-DNA (recall that Z-DNA has a left-handed helix), Zab targets ADAR1 to genes that are undergoing transcription. This would facilitate rapid A → I editing, which must take place before the next splicing reaction occurs.

s. RNA Interference Degrades mRNAs

Since the 1990s it has become increasingly clear that noncoding RNAs have important roles in controlling gene expression. One of the first indications of this phenomenon occurred in Richard Jorgensen's attempt to genetically engineer more vividly purple petunias by introducing extra copies of the gene that directs the synthesis of the purple pigment. Surprisingly, the resulting transgenic plants had variegated and often entirely white flowers. Apparently, the purple-making genes somehow switched each other off. Similarly, it is well known that **antisense RNA** (RNA that is complementary to at least a portion of an mRNA) prevents the translation of the corresponding mRNA because ribosomes cannot translate double-stranded RNA.

Yet, injecting **sense RNA** (RNA with the same sequence as an mRNA) into the nematode *Caenorhabditis elegans* also blocks protein production. Since the added RNA somehow interferes with gene expression, this phenomenon is known as **RNA interference (RNAi)**. RNAi is now known to occur in all eukaryotes investigated except baker's yeast.

The mechanism of RNAi began to come to light in 1998 when Andrew Fire and Craig Mello showed that double-stranded RNA (**dsRNA**) was substantially more effective in causing RNAi in *C. elegans* than were either of its component strands alone. RNAi is induced by only a few molecules of dsRNA per affected cell, suggesting that RNAi is a catalytic rather than a stoichiometric effect. Further investigations, in large part in *Drosophila*, have led to the elucidation of the following pathway mediating RNAi (Fig. 31-72):

1. The trigger dsRNA, as Phillip Zamore discovered, is chopped up into ~21- to 25-nt-long double-stranded fragments known as **small interfering RNAs (siRNAs)**, each of

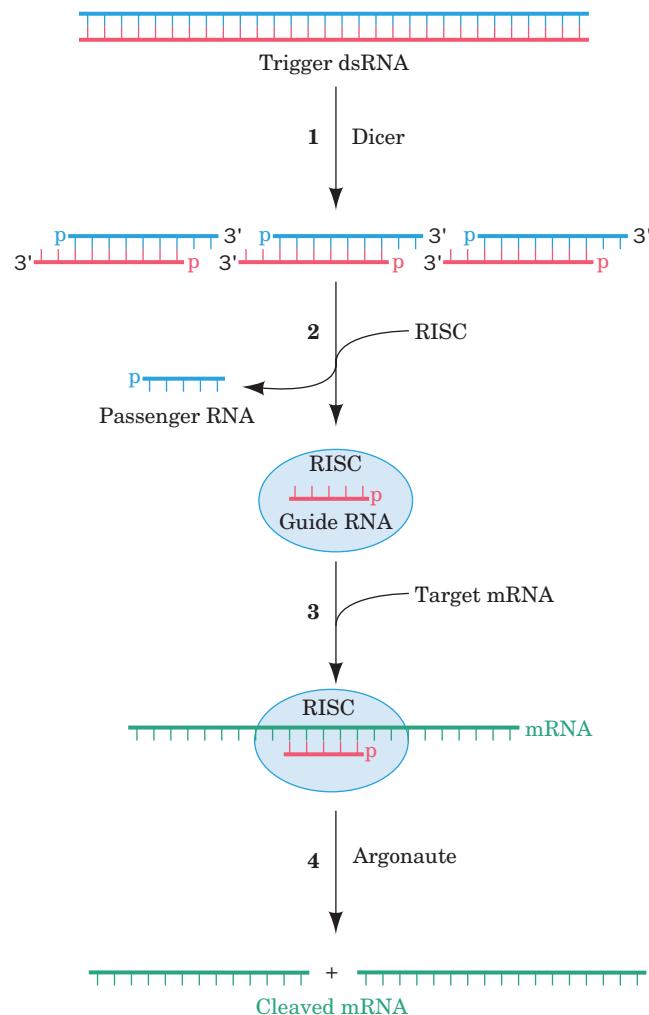


Figure 31-72 A mechanism for RNA interference (RNAi). See the text for details. ATP is required for Dicer-catalyzed cleavage of RNA and for RISC-associated helicase unwinding of double-stranded RNA. Depending on the species, the mRNA may not be completely degraded.

whose strands has a 2-nt overhang at its 3' end and a 5' phosphate. This reaction is mediated by an ATP-dependent RNase named **Dicer**, a homodimer of ~1900-residue subunits in animals that is a member of the **RNase III** family of double-strand-specific RNA endonucleases.

2. An siRNA is transferred to a 250- to 500-kD multi-subunit complex known as **RNA-induced silencing complex (RISC)**. RISC has at least four protein components, one of which is an ATP-dependent RNA helicase that separates the two strands of the siRNA. The strand whose 5' end has the lower free energy of binding, the **guide RNA**, is bound by the RISC whereas its complementary strand, the **passenger RNA**, is cleaved and discarded. In some species, but apparently not in humans, the original siRNA signal is amplified by the action of an **RNA-dependent RNA polymerase (RdRP)**.

3. The **guide RNA** recruits the RISC complex to an mRNA with the complementary sequence.

4. An RNase III component of RISC known as **Argonaute (AGO; also called Slicer)** cleaves the mRNA opposite the bound guide RNA. The cleaved mRNA is then further degraded by cellular nucleases, thereby preventing its translation.

The X-ray structure of Dicer from the parasitic protozoan *Giardia intestinalis*, determined by Jennifer Doudna, reveals that its shape resembles that of a hatchet, with its two RNase III domains forming the blade and its PAZ domain (named for three proteins in which it is contained, PIWI, Argonaute, and Zwille) forming the base of its handle (Fig. 31-73; Dicers from higher eukaryotes additionally contain an N-terminal DExD/H box helicase domain and a C-terminal dsRNA-binding domain). The two RNase III domains form an internal heterodimer that resembles the homodimeric structure of bacterial RNase III. Four conserved acidic residues in each RNase III domain bind two Mg^{2+} ions and hence are postulated to cleave an RNA strand via a two-metal-ion mechanism (Section 30-2Af). The two RNase III active sites are 17.5 Å apart, the width of dsRNA's major groove, and thus appear positioned to cleave the two strands of a bound dsRNA. The PAZ domain specifically binds dsRNA ends that have a 3' two-nt overhang. The distance between this binding site and its closest RNase III domain active site is 65 Å, the length of a 25-bp dsRNA. This explains how Dicer cleaves an ~25-bp segment from the end of dsRNA.

Argonaute proteins consist of four domains: an N-terminal (N), a PAZ, a middle (Mid), and a PIWI (for *P*-element induced *wimpy testis*) domain. The X-ray structures of several bacterial Argonaute proteins reveal that they have a bilobal architecture with the N and PAZ domains forming one lobe and the Mid and PIWI domains forming the other. The PIWI domain has an RNase H fold (RNase H cleaves the RNA strand of an RNA · DNA hybrid helix), which strongly suggests that it mediates Argonaute's "slicer" activity (bacterial Argonautes preferentially bind **guide DNA** over RNA). The X-ray structure of *T. thermophilus* Argonaute in ternary complex with a 21-nt

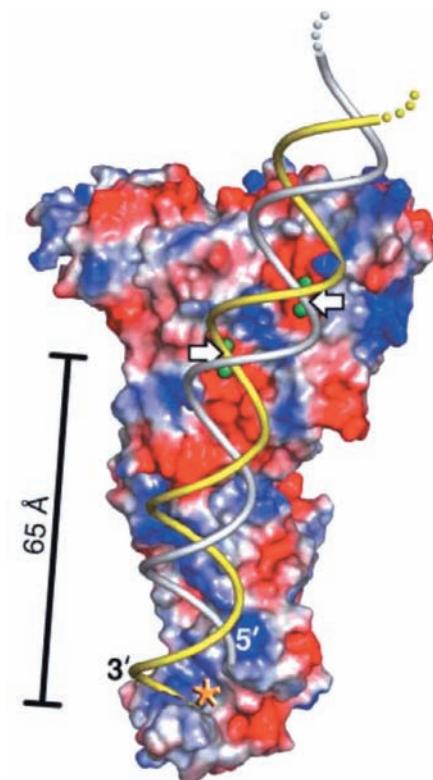


Figure 31-73 X-ray structure of Dicer from *G. intestinalis*.

The protein is represented by its molecular surface colored according to its surface charge with red negative, blue positive, and white neutral. Bound Mg^{2+} ions, which are represented by green spheres, mark the active site of each of the protein's two RNase III domains. A dsRNA has been modeled into the structure with its 3' overhang entering the PAZ domain's binding pocket (asterisk). The white arrows point to the dsRNA's scissile phosphate groups. Note that much of the surface to which the anionic dsRNA is presumably bound is positively charged (the calculated surface charge does not take into account the bound Mg^{2+} ions). [Courtesy of Jennifer Doudna, University of California at Berkeley. PDBid 2FFL.]

guide DNA and a 19-nt target RNA (Fig. 31-74), determined by Patel, reveals that the DNA · RNA hybrid helix binds in the cleft between Argonaute's two lobes with the phosphate group bridging RNA nucleotides 10 and 11 positioned for cleavage at the PIWI active site. The comparison of this structure with those of similar complexes that lack the target RNA or in which the target RNA has 12 or 15 nucleotides indicates that the guide DNA (and presumably the guide RNA in eukaryotes) initially binds to Argonaute with its 3' end in the PAZ binding pocket, but as the hybrid helix lengthens beyond one turn, this 3' end is released, which facilitates further winding of the hybrid helix.

t. RNAi Defends against Viral Infection and Regulates Gene Expression

What is the physiological function of RNAi? Since many eukaryotic viruses store and replicate their genomes as RNA (Chapter 33), it seems likely that RNAi arose as a defense against viral infections. Indeed, many plant viruses contain genes that suppress various steps of RNAi and which are essential for pathogenesis. RNAi has also been

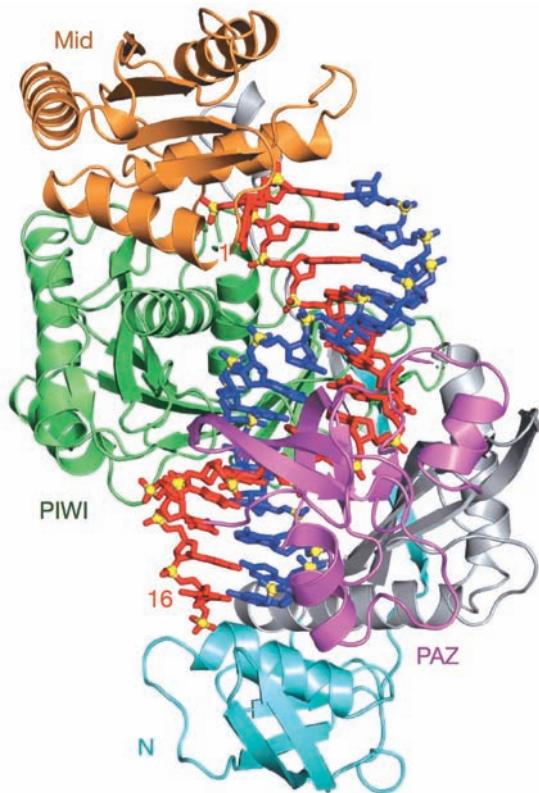


Figure 31-74 X-ray structure of *T. thermophilus* Argonaute in ternary complex with a 21-nt guide DNA and a 19-nt target RNA. The protein is shown in ribbon form with its N, PAZ, Mid, and PIWI domains cyan, magenta, orange, and green, respectively, and with the linkers connecting these domains gray. The guide DNA and target RNA are drawn in stick form in red and blue with their P atoms yellow. Only DNA nucleotides 1 to 16 and RNA nucleotides 2 to 16 are visible. [Courtesy of Dinshaw Patel, Memorial-Sloan Kettering Cancer Center, New York, New York. PDBid 3HK2.]

shown to inhibit the intragenomic spread of retrotransposons (Section 30-6Bh).

A wide variety of eukaryotes, including plants, nematodes, flies, fish, and mammals, use RNAi to control gene expression. Certain mRNAs expressed by these organisms contain ~70-nt, imperfectly base-paired, stem-loop structures that are excised by a 1374-residue RNase III named **Drosha** (Fig. 31-75). The stem-loops are exported from the nucleus to the cytosol where they are cleaved by Dicer to liberate ~22-bp dsRNAs known as **microRNAs (miRNAs)**; so called to differentiate these endogenous RNAs from exogenous siRNAs. The transcripts from which miRNAs are derived are known as **pri-miRNAs** (pri for *primary*), whereas the stem-loops are called a **pre-miRNAs** (pre for *precursor*). Pre-miRNAs can be located within both the introns and, less commonly, the exons of a pri-miRNA. The miRNAs bind to **RISC** in which they function to identify the tens to hundreds of mRNAs containing segments that are partially complementary to the miRNA.

The RISC-bound miRNA binds to its targets site, which is usually in the 3' untranslated region (3'UTR) of an

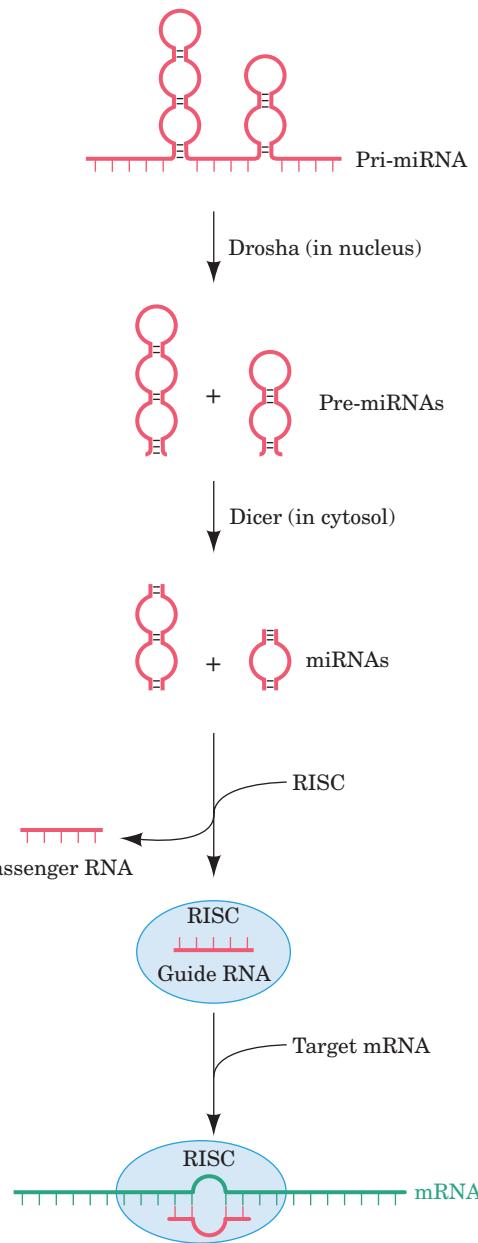


Figure 31-75 The generation of miRNAs from pri-miRNAs and their RISC-mediated binding to target mRNAs. See the text for details.

mRNA. A lack of perfect complementarity prevents Argonaute from cleaving the mRNA (Argonaute's PIWI domain catalyzes slicing only if there is perfect complementarity to the miRNA's so-called **seed sequence**, which consists of nucleotides 2–8 from its 5' end), and in fact, some species of Argonaute lack the catalytic residues to do so. Instead, miRNA-mediated silencing is thought to occur through the removal of its target mRNA's poly(A) tail or its m⁷G cap, which leads to the mRNA's degradation (Section 31-4Av), and/or the RISC-mediated repression of the target mRNA's translation by interfering with ribosomal initiation (Section 32-3Cd) and sequestering or degrading the mRNA in cytoplasmic granules known as **P bodies** (P for *processing*).

In 1993, Victor Ambros discovered the first known miRNA, which is encoded by the *lin-4* gene of *C. elegans* (Fig. 31-76a). The *lin-4* gene was known to control the timing of larval development, although it was thought that it encoded a protein that repressed the expression of the *lin-14* gene. In fact, the *lin-4* miRNA is complementary to seven sites on the 3'UTR of the *lin-14* gene, which had previously been shown to mediate the repression of *lin-14* by the *lin-4* gene product. A puzzling observation at the time was that this regulation greatly reduces the amount of LIN-14 protein produced without altering the level of *lin-14* mRNA. These findings were eventually followed by the discovery that the *C. elegans* *let-7* gene encodes what is now known to be an miRNA (Fig. 31-76b) that controls the transition from larval to adult stages of development. Subsequently, *let-7* homologs were identified in the *Drosophila* and human genomes and *let-7* RNA was detected in these organisms as well as in numerous other animals.

Both the *lin-4* and *let-7* miRNAs were discovered by genetic analyses. However most of the nearly 10,000 miRNAs in plants and animals that are now known, including those in Fig. 31-76c, were identified through bioinformatic approaches (Section 7-4). The known miRNAs are catalogued in the miRBase database (<http://www.mirbase.org/>). Nearly all miRNAs are conserved among closely related animals (e.g., mice and humans) and many are more broadly conserved throughout animal lineages (e.g., over one-third of the 174 *C. elegans* miRNAs have homologs in humans). *The significance of miRNAs is indicated by the fact that humans express over 720 miRNAs that participate in regulating ~30% of their protein-coding genes.*

RNAi has become the method of choice for “knocking out” specific genes in plants and invertebrates. For example, in *C. elegans*, RNAi has been used to systematically inactivate over 16,000 of its ~19,000 protein-coding genes in an attempt to assign a function to each gene. *C. elegans* is particularly amenable to the RNAi approach, since these worms eat *E. coli* cells, and it is relatively easy to genetically engineer the bacterial cells to express double-stranded RNA that becomes part of the worms’ diet. One limitation of the RNAi method is that it permits only the effects of gene inactivation—not gene activation—to be examined.

RNAi is of similar use in mammalian systems, even though mammals lack the mechanisms that amplify silencing in plants and nonvertebrates so that the effects of RNAi in mammals are transient. The exquisite specificity of RNAi may make it possible to prevent viral infections and to silence disease-causing mutant genes such as oncogenes. In fact, experiments have demonstrated that it is possible to use RNAi to block the liver’s inflammatory response to a hepatitis virus, at least in mice, and to prevent HIV replication in cultured human cells. One challenge for the future is to devise protocols for more specific and longer-lasting gene silencing that would make it possible to prevent viral infections or to block the effects of disease-causing mutant genes.

u. Mature Eukaryotic mRNAs Are Actively Transported from the Nucleus to the Cytoplasm

The translation of prokaryotic mRNAs is often initiated, as we have seen in Fig. 31-27, before their synthesis is

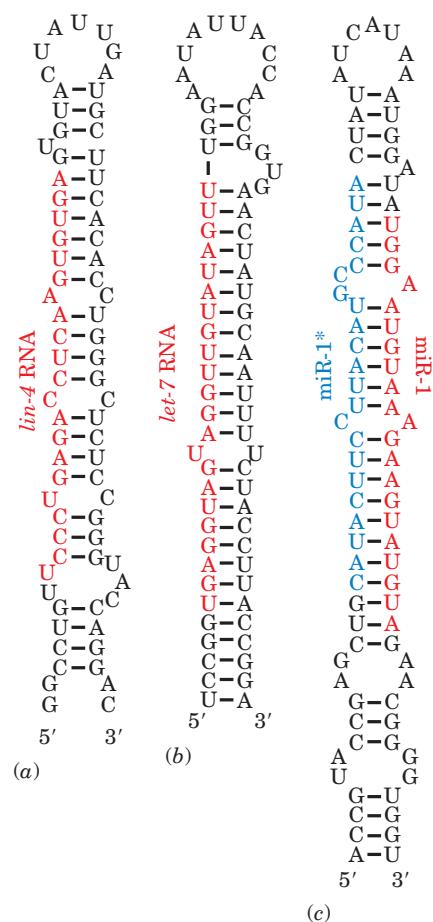


Figure 31-76 The predicted stem-loops of some pre-miRNAs. The miRNAs contained in these pre-mRNAs, all of which are from *C. elegans*, are red. (a) *lin-4*, (b) *let-7*, and (c) *miR-1* and *miR-1** (in blue), which are largely complementary to each other.

complete. This cannot occur in eukaryotes because the transcription and post-transcriptional processing of eukaryotic mRNAs occurs in the nucleus but their translation takes place in the cytosol. Consequently, mature mRNAs must be transported from the nucleus to the cytoplasm. This is a highly selective process because mature mRNAs comprise only a small fraction of the RNAs present in the nucleus, the remainder being pre-mRNAs, excised introns (which are usually much larger than the exons from which they were liberated), rRNAs, tRNAs, snRNAs, and a variety of RNAs that participate in the processing of rRNAs and tRNAs (Sections 31-4B and 31-4C). Indeed, only ~5% of the RNA that is synthesized ever leaves the nucleus.

How are mature mRNAs recognized and transported? As we have seen, throughout their residency in the nucleus, pre-mRNAs are continually associated with numerous proteins, including those that participate in synthesizing their m⁷G caps and poly(A) tails, and in splicing out their introns. In addition, the **exon junction complex (EJC)**, which consists of four core proteins and several transiently associating proteins, is deposited onto mRNA during the splicing process at a site that is 20 to 24 nt upstream of the splice junction without regard to its sequence. The population of proteins bound to an mRNA changes as the mRNA is processed but some of the proteins, including SR proteins, hnRNPs (Section 31-4Ai), and EJCs remain associ-

ated with mature mRNAs in the nucleus. However, it appears that it is its entire collection of bound proteins rather than any individual protein that serves to identify an mRNA to the nuclear export machinery.

The eukaryotic nucleus (Fig. 1-5) is a double membrane-enveloped organelle that in animals is penetrated by an average of ~ 3000 pores. These are formed by **nuclear pore complexes (NPCs)**, which are massive ($\sim 120,000$ kD), 8-fold symmetric assemblies of ~ 30 different proteins known as **nucleoporins**. NPCs, which have inner diameters of ~ 90 Å (although this may be expandable to as much as 260 Å), allow the free diffusion of molecules of up to 50 kD, but most macromolecules, including mRNAs in their complexes with proteins, require an active transport process to pass through an NPC. Some of the proteins associated with mature mRNAs bear nuclear export signals that are recognized by a protein receptor that in yeast is named **Dbp5**. This 482-residue DExD/H box protein (Dbp5 stands for *DExD/H box protein 5*) is an ATP-driven RNA helicase that also binds to the NPC. This permits Dbp5 to pull the mRNA out into the cytosol while simultaneously stripping away many of its bound proteins. These proteins are later recycled by returning them to the nucleus through the NPCs.

v. mRNA Degradation Is Elaborately Controlled

The synthesis and maturation of mRNAs, as we have seen, are subject to multiple controls. The same is true of their degradation. Indeed, the range of mRNA stability in eukaryotic cells, measured in half-lives, varies from a few minutes to many hours or days. The mRNA molecules themselves contain elements that dictate their decay rates. These elements include the 3' poly(A) tail and the 5' m⁷G cap, which protect against exonucleases, as well as sequences that are located within the coding region.

A major route for mRNA degradation begins with the progressive removal of its poly(A) tail, a process catalyzed by **deadenylases** that are located throughout the cytosol. When the residual poly(A) tail is less than ~ 10 nt long and hence no longer capable of binding poly(A) binding protein (Section 31-4Ab), the mRNA becomes a substrate for a **decapping enzyme**, which hydrolytically excises the mRNA's m⁷G cap. This is possible because the translational initiation factor **eIF4G** interacts with both poly(A) binding protein and cap binding protein (Section 32-3Cd), thereby circularizing the mRNA so that events at its 3' end can be coupled to events at its 5' end. The decapped and deadenylated mRNA is then degraded by exonucleases, mainly the 1706-residue 5' \rightarrow 3' exonuclease **Xrn1** and the 3' \rightarrow 5' exonuclease complex named the **exosome**. A decapping enzyme, 5' \rightarrow 3' exonucleases, and accessory proteins form P bodies (Section 31-4At) that function to either degrade mRNA or store it in an inactive form.

Proteins that bind to **AU-rich elements (AREs)** in the 3' untranslated region of mRNAs also appear to increase or decrease the rate of mRNA degradation, although their exact action is poorly understood. RNA secondary structure and RNA-binding proteins, which may be susceptible to modification by cellular signaling pathways, are thought to play a role in regulating mRNA stability.

The eukaryotic core exosome consists of single copies of nine different subunits. Its X-ray structure (Fig. 31-77), de-

termined by Christopher Lima, reveals that six of these subunits, **Rpr41** (Rpr for *rRNA* processing; the exosome was discovered as an activity that processed the 3' ends of rRNAs), **Rrp42**, **Mtr3**, **Rrp43**, **Rrp46**, and **Rrp45**, form a six-membered ring with the remaining three subunits, **Rrp4**, **Csl4**, and **Rrp40**, bound to the same face of this ring. These subunits are arranged such that the core exosome contains an ~ 9 -Å-wide central channel that allows the entrance of only single-stranded RNAs.

The archeal exosome appears to be a simpler version of the eukaryotic core exosome. Its six-membered ring consists of only two types of subunits, Rrp41 and Rrp42, that alternate around the ring, with three copies of Rrp4 bound to the same face of the ring. Only Rrp41 contains an active site although Rrp42 is required for activity. Not surprisingly, eukaryotic Rrp4, Mtr3, and Rrp46 are homologs of archeal Rrp41, eukaryotic Rrp42, Rrp43, and Rrp45 are homologs of archeal Rrp42, and eukaryotic Rrp4, Csl4, and Rrp40 are homologs of archeal Rrp4. Nevertheless, despite the fact that each of its core subunits are essential for viability, eukaryotic core exosomes, from yeast to humans, are catalytically inactive due to changes in active site residues relative to their archeal homologs. However, the core exosome associates with two 3'-exonucleases, **Rrp6** and **Rrp44**, whose catalytically inactive mutants are individually viable in yeast but lethal in combination. Moreover, the core exosome interacts with numerous mostly multisubunit cofactors that carry out a variety of RNA processing activities in both the nucleus and the cytosol. Thus, the eukaryotic core exosome appears to be a structural platform upon which many RNA processing enzymes can be mounted.

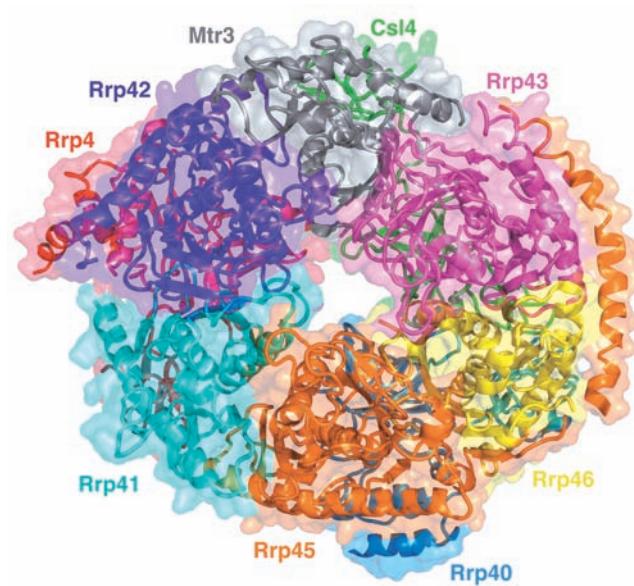


Figure 31-77 X-ray structure of the human core exosome. The protein complex is drawn in ribbon form embedded in its semitransparent molecular surface, with each of its nine different subunits separately colored. The view is toward the face of the six-membered ring of subunits opposite that to which the three other subunits bind. [Based on an X-ray structure by Christopher Lima, Sloan-Kettering Institute, New York, New York. PDBID 2NN6.]

B. Ribosomal RNA Processing

The seven *E. coli* rRNA operons all contain one (nearly identical) copy of each of the three types of rRNA genes (Section 32-3A). Their polycistronic primary transcripts, which are ~5500 nucleotides in length, contain 16S rRNA at their 5' ends followed by the transcripts for 1 or 2 tRNAs, 23S rRNA, 5S rRNA, and, in some rRNA operons, 1 or 2 more tRNAs at the 3' end (Fig. 31-78). The steps in processing these primary transcripts to mature rRNAs were elucidated with the aid of mutants defective in one or more of the processing enzymes.

The initial processing, which yields products known as **pre-rRNAs**, commences while the primary transcript is still being synthesized. It consists of specific endonucleolytic cleavages by **RNase III**, **RNase P**, **RNase E**, and **RNase F** at the sites indicated in Fig. 31-78. The base sequence of the primary transcript suggests the existence of several base-paired stems. The RNase III cleavages occur in a stem consisting of complementary sequences flanking the 5' and 3' ends of the 23S segment (Fig. 31-79) as well as that of the 16S segment. Presumably, certain features of these stems constitute the RNase III recognition site.

The 5' and 3' ends of the pre-rRNAs are trimmed away in secondary processing steps (Fig. 31-78) through the action of **RNases D**, **M16**, **M23**, and **M5** to produce the mature rRNAs. These final cleavages only occur after the pre-rRNAs become associated with ribosomal proteins.

a. Ribosomal RNAs Are Methylated

During ribosomal assembly, the 16S and 23S rRNAs are methylated at a total of 24 specific nucleosides. The methylation reactions, which employ S-adenosylmethionine (Section 26-3Ea) as a methyl donor, yield N^6,N^6 -dimethyladenine and O^2 -methylribose residues. O^2 -methyl groups may protect adjacent phosphodiester bonds from degradation by intracellular RNases (the mechanism of RNase hydrolysis involves utilization of the free 2'-OH group of ribose to eliminate the

substituent on the 3'-phosphoryl group via the formation of a 2',3'-cyclic phosphate intermediate; Figs. 5-3 and 15-3). However, the function of base methylation is unknown.

b. Eukaryotic rRNA Processing Is Guided by snoRNAs

The eukaryotic genome typically has several hundred tandemly repeated copies of rRNA genes that are contained in small, dark-staining nuclear bodies known as **nucleoli** (the site of rRNA transcription and processing and ribosomal subunit assembly; Fig. 1-5; note that nucleoli are not membrane enveloped). The primary rRNA transcript is an ~7500-nucleotide 45S RNA that contains, starting from its 5' end, the 18S, 5.8S, and 28S rRNAs separated by spacer sequences (Fig. 31-80). In the first stage of its processing, 45S RNA is specifically methylated at numerous sites (106 in humans) that occur mostly in its rRNA sequences. About 80% of these modifications yield O^2 -methylribose residues and the remainder form methylated bases such as N^6,N^6 -dimethyladenine and 2-methylguanine. In addition, many pre-rRNA U's (95 in humans) are converted to pseudouridines (Ψ 's) (Section 30-5Be), which may contribute to the rRNA's tertiary stability through hydrogen bonding involving its newly acquired ring NH group. The subsequent cleavage and trimming of the 45S RNA superficially resembles that of prokaryotic rRNAs. In fact, enzymes exhibiting RNase III- and RNase P-like activities occur in eukaryotes. The 5S eukaryotic rRNA is separately processed in a manner resembling that of tRNA (Section 31-4C).

The methylation sites in eukaryotic rRNAs occur exclusively within conserved domains that are therefore likely to participate in fundamental ribosomal processes. Indeed, the methylation sites generally occur in invariant sequences among yeast and vertebrates although the methylations themselves are not always conserved. These methylation sites do not appear to have a consensus structure

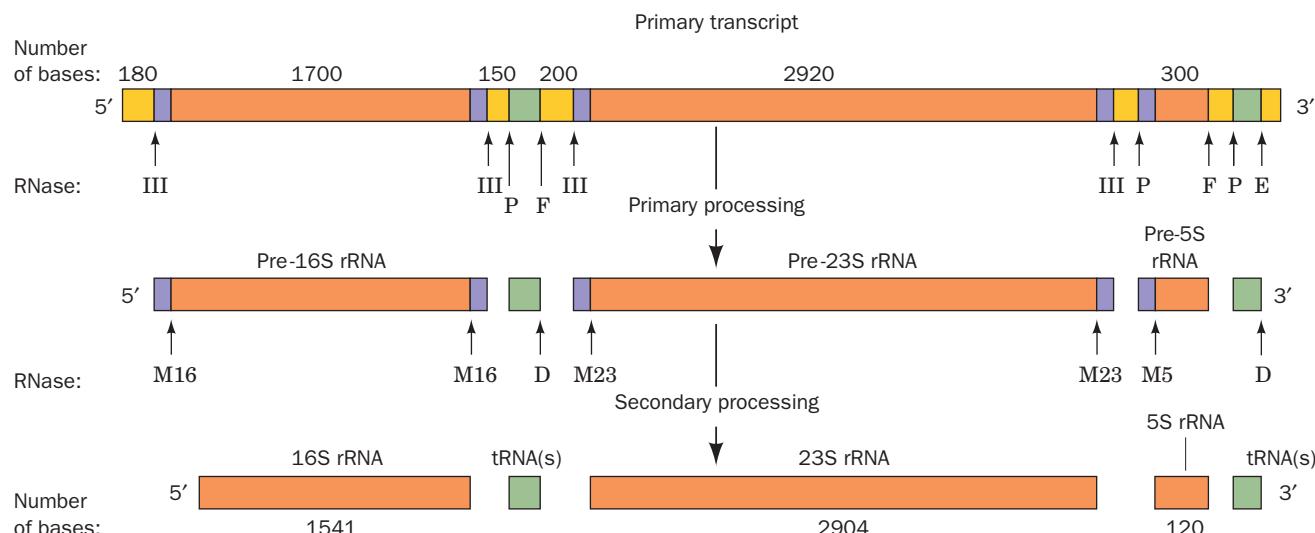


Figure 31-78 The post-transcriptional processing of *E. coli* rRNA. The transcriptional map is shown approximately to scale. The labeled arrows indicate the positions of the various nucleolytic cuts and the nucleases that generate them. [After

Apiron, D., Ghora, B.K., Plantz, G., Misra, T.K., and Gegenheimer, P., in Söll, D., Abelson, J.N., and Schimmel P.R. (Eds.), *Transfer RNA: Biological Aspects*, p. 148, Cold Spring Harbor Laboratory Press (1980).]

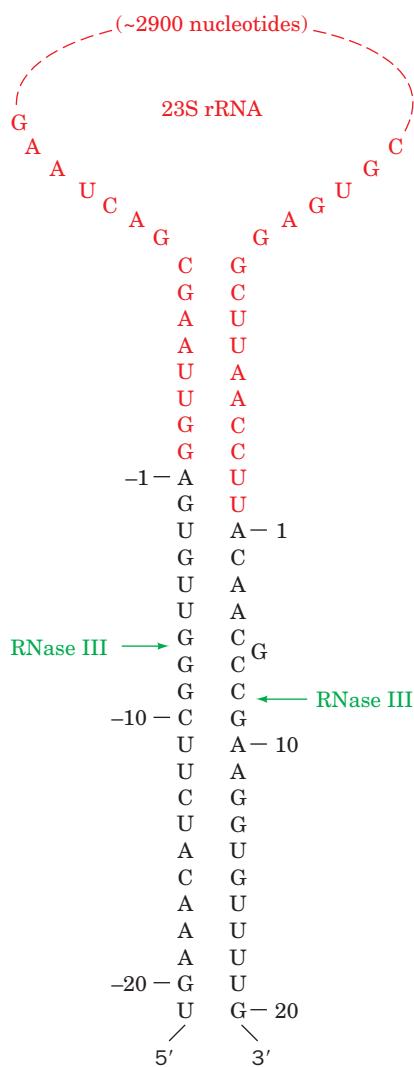


Figure 31-79 The stem-and-giant-loop secondary structure in the 23S region of the *E. coli* primary rRNA transcript. The RNase III cleavage sites are indicated. [After Young, R.R., Bram, R.J., and Steitz, J.A., *in* Söll, D., Abelson, J.N., and Schimmel, P.R. (Eds.), *Transfer RNA: Biological Aspects*, p. 102, Cold Spring Harbor Laboratory Press (1980).]

The snoRNA nucleotide that pairs with the nucleotide to be O₂'-methylated always precedes box D by exactly 5 nt. Evidently, each of these so-called **box C/D snoRNAs** act to guide the methylation of a single site. In fact, in those cases in which two adjacent ribose residues are methylated, two box C/D snoRNAs with overlapping sequences occur. The methylation is mediated by a complex of at least four nucleolar proteins, including **fibrillarin** (~325 residues; so called because it is located in the dense fibrillar region of the nucleolus), the likely methyltransferase, which together with a box C/D snoRNA form **snoRNPs**. The conversion of specific rRNA U's to Ψ's is similarly mediated by a different subgroup of snoRNAs, the **box H/ACA snoRNAs**, so called because they contain the sequence motifs ACANN at the snoRNA's 3' end and box H (ANANNA) at its 5' end, so as to flank a sequence that partially base-pairs to the pre-rRNA segment containing the U to be converted to Ψ. Archaea also modify their rRNAs via RNA-guided methylations and U to Ψ conversions but, interestingly, the analogous reactions in eubacteria are mediated by protein enzymes that lack RNA.

C. Transfer RNA Processing

tRNAs, as we discuss in Section 32-2A, consist of ~80 nucleotides that assume a secondary structure with four base-paired stems known as the **cloverleaf structure** (Fig. 31-81). All tRNAs have a large fraction of modified bases (whose structures and functions are discussed in Section 32-2Aa) and each has the 3'-terminal sequence —CCA to which the

that might be recognized by a single methyltransferase. How, then, are these methylation sites targeted?

An important clue as to how the methylation sites on rRNA are selected came from the observation that pre-rRNA interacts with the members of a large family of **small nucleolar RNAs (snoRNAs)** (~100 in yeast and ~200 in mammals). The snoRNAs, whose lengths vary from 70 to 100 nt, contain segments of 10 to 21 nt that are precisely complementary to segments of the mature rRNAs that contain the O₂'-methylation sites. These snoRNA sequences are located between the conserved sequence motifs known as box C (RUGAUGA) and box D (CUGA), which are respectively located on the 5' and 3' sides of the complementary segments. In intron-rich organisms such as vertebrates, most snoRNAs are encoded by the introns of structural genes so that not all excised introns are discarded.

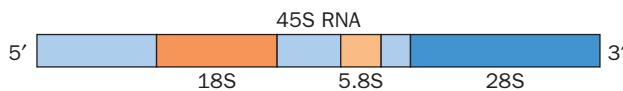


Figure 31-80 The organization of the 45S primary transcript of eukaryotic rRNA.

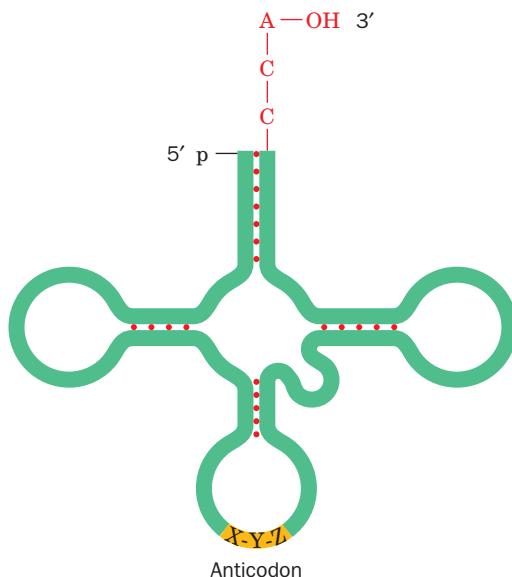


Figure 31-81 A schematic diagram of the tRNA cloverleaf secondary structure. Each dot indicates a base pair in the hydrogen bonded stems. The position of the anticodon triplet and the 3'-terminal —CCA are indicated.

corresponding amino acid is appended in the amino acid-charged tRNA. The **anticodon** (which is complementary to the codon specifying the tRNA's corresponding amino acid) occurs in the loop of the cloverleaf structure opposite the stem containing the terminal nucleotides.

The *E. coli* chromosome contains ~60 tRNA genes. Some of them are components of rRNA operons (Section 31-4B); the others are distributed, often in clusters, throughout the chromosome. The primary tRNA transcripts, which contain from one to as many as four or five identical tRNA copies, have extra nucleotides at the 3' and 5' ends of each tRNA sequence. The excision and trimming of these tRNA sequences resemble those for *E. coli* rRNAs (Section 31-4B) in that both processes employ some of the same nucleases.

a. RNase P Is a Ribozyme

RNase P, which generates the 5' ends of tRNAs (Fig. 31-78), is a particularly interesting enzyme because it has, in *E. coli*, a 377-nucleotide RNA component (~125 kD vs 14 kD for its 119-residue protein subunit) that is essential for its enzymatic activity. The enzyme's RNA was, quite understandably, first proposed to function in recognizing the sub-

strate RNA through base pairing and to thereby guide the protein subunit, which was presumed to be the actual nuclease, to the cleavage site. However, Sidney Altman demonstrated that *the RNA component of RNase P is, in fact, the enzyme's catalytic subunit* by showing that protein-free RNase P RNA catalyzes the cleavage of substrate RNA at high salt concentrations. RNase P protein, which is basic, evidently functions at physiological salt concentrations to electrostatically reduce the repulsions between the polyanionic ribozyme and substrate RNAs. The argument that trace quantities of RNase P protein are really responsible for the RNase P reaction was disposed of by showing that catalytic activity is exhibited by RNase P RNA that has been transcribed in a cell-free system. RNase P activity occurs in eukaryotes (nuclei, mitochondria, and chloroplasts) as well as in prokaryotes although eukaryotic nuclear RNase P's have 9 or 10 protein subunits, none of which are related to the bacterial protein. Indeed, RNase P

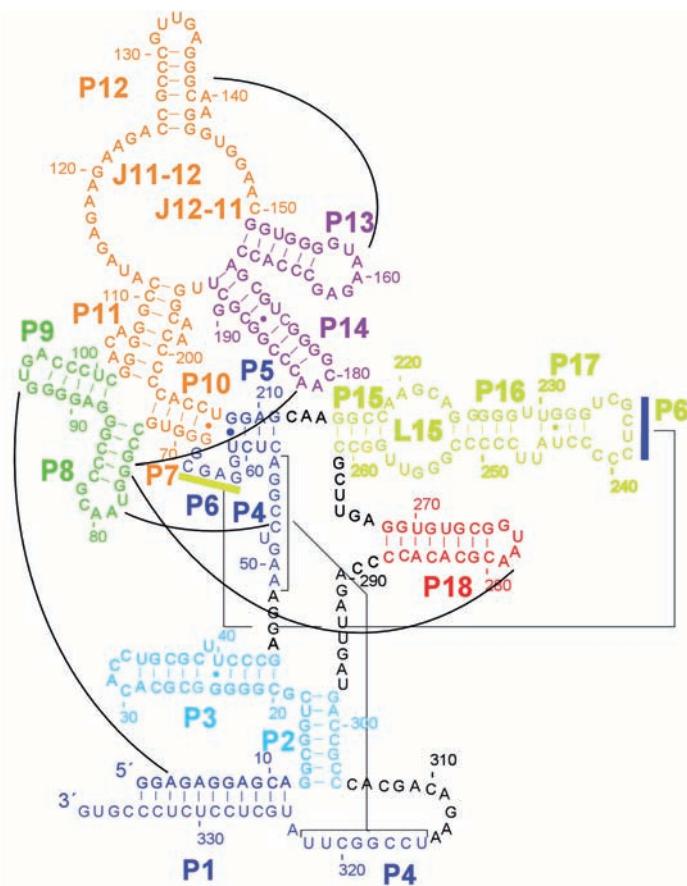
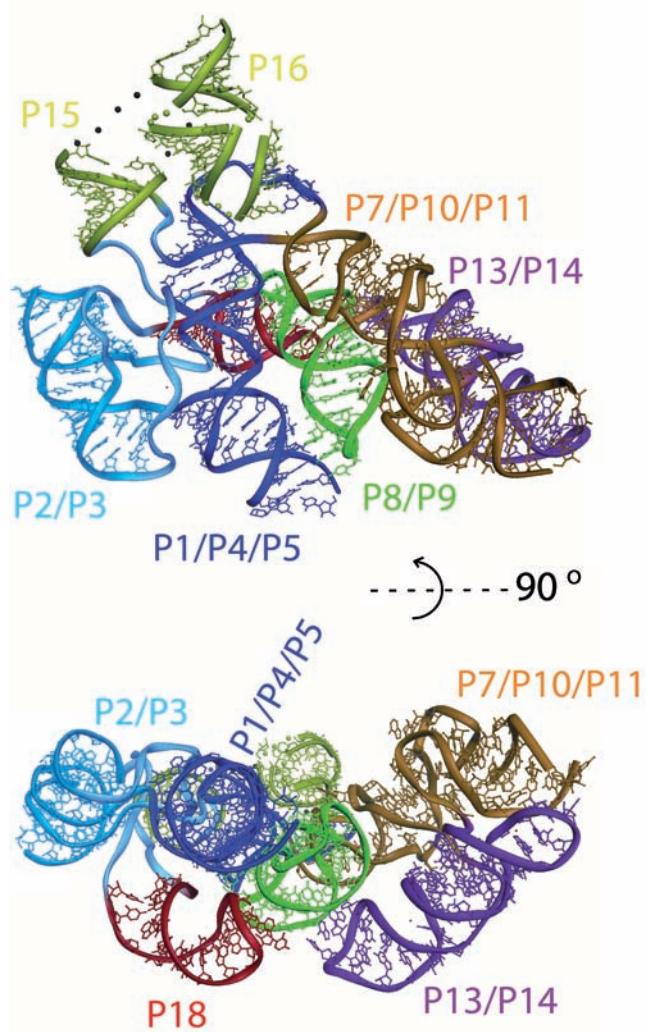


Figure 31-82 Structure of the RNA component of *T. maritima* RNase P. (a) Its sequence and secondary structure. The various segments (P for paired region, J for joining region, and L for loop) are shown in different colors. The black lines indicate major interactions that are observed in the X-ray structure, dashes indicate Watson-Crick base pairs, and small



filled circles represent non-Watson-Crick base pairs. (b) Its X-ray structure, which is colored as in Part a. Of its 338 nucleotides, 309 are visible. The lower view is related to the upper view by a 90° rotation about the horizontal axis. [Courtesy of Alfonso Mondragón, Northwestern University. PDBid 2A2E.]

mediates one of the two ribozymal activities that occur in all cellular life, the other being associated with ribosomes (Section 32-3Dg).

The X-ray structures of the RNA components of RNase P from *Thermotoga maritima* (338 nt) and *Bacillus stearothermophilus* (417 nt), which were independently determined by Alfonso Mondragón and Norman Pace, reveal that these ribozymes consist mainly of stacked helical stems with overall compact structures typical of protein enzymes (Fig. 31-82). Biochemical studies and modeling indicate that the RNase active site lies in a cleft between the P2/P3 region (cyan in Fig. 31-82) and the P1/P4/P5 region (dark blue in Fig. 31-82). That portion of the structure encompassing P1 through P4, P9 through P11, J11-12 and J12-11 is present in all known RNase P's and hence is known as the universal minimum consensus structure. This structure was presumably present in the primordial RNase P.

b. Many Eukaryotic Pre-tRNAs Contain Introns

Eukaryotic genomes contain from several hundred to several thousand tRNA genes. Many eukaryotic primary

tRNA transcripts, for example, that of yeast tRNA^{Tyr} (Fig. 31-83), contain a small intron adjacent to their anticodons as well as extra nucleotides at their 5' and 3' ends. Note that this intron is unlikely to disrupt the tRNA's cloverleaf structure.

c. The —CCA Ends of Eukaryotic tRNAs Are Post-Transcriptionally Appended

Eukaryotic tRNA transcripts lack the obligatory —CCA sequence at their 3' ends. This is appended to the immature tRNAs by the enzyme **CCA-adding polymerase**, which sequentially adds two C's and an A to tRNA using CTP and ATP as substrates. This enzyme also occurs in prokaryotes, although, at least in *E. coli*, the tRNA genes all encode a —CCA terminus. The *E. coli* CCA-adding polymerase is therefore likely to function in the repair of degraded tRNAs.

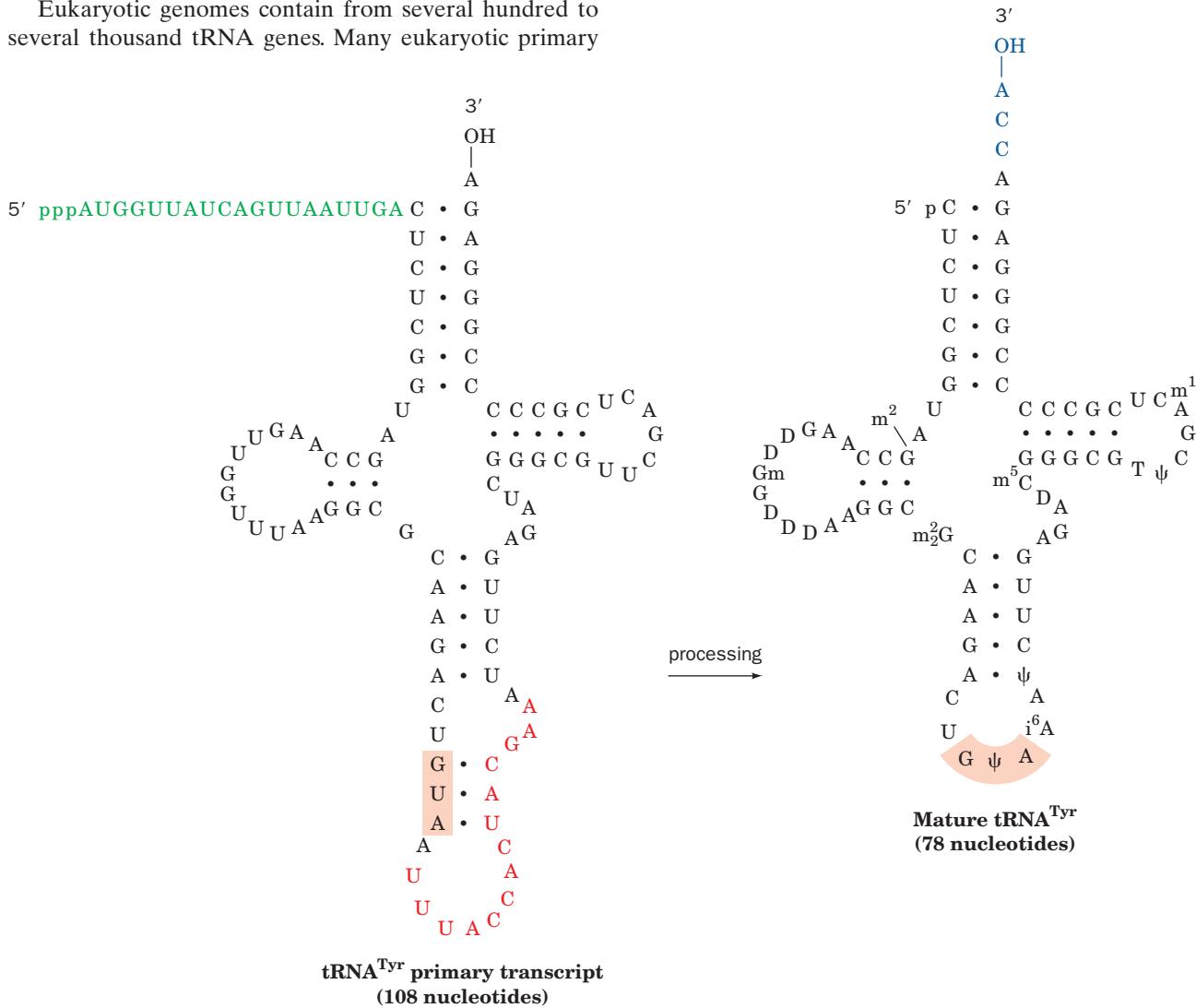


Figure 31-83 The post-transcriptional processing of yeast tRNA^{Tyr}. A 14-nucleotide intervening sequence (red) and a 19-nucleotide 5'-terminal sequence (green) are excised from the primary transcript, a —CCA (blue) is appended to the 3' end,

and several of the bases are modified (their symbols are defined in Fig. 32-13) to form the mature tRNA. The anticodon is shaded. [After DeRobertis, E.M. and Olsen, M.V., *Nature* **278**, 142 (1989).]

CHAPTER SUMMARY

1 The Role of RNA in Protein Synthesis The central dogma of molecular biology states that “DNA makes RNA makes protein” (although RNA can also “make” DNA). There is, however, enormous variation among the rates at which the various proteins are made. Certain enzymes, such as those of the *lac* operon, are synthesized only when the substances whose metabolism they catalyze are present. The *lac* operon consists of the control sequences *lacP* and *lacO* followed by the tandemly arranged genes for β -galactosidase (*lacZ*), galactoside permease (*lacY*), and thiogalactoside transacetylase (*lacA*). In the absence of inducer, physiologically allolactose, the *lac* repressor, the product of the *lacI* gene, binds to operator (*lacO*) so as to prevent the transcription of the *lac* operon by RNA polymerase. The binding of inducer causes the repressor to release the operator that allows the *lac* structural genes to be transcribed onto a single polycistronic mRNA. The mRNAs transiently associate with ribosomes so as to direct them to synthesize their encoded polypeptides.

2 RNA Polymerase The holoenzymes of bacterial RNA polymerases (RNAPs) have the subunit structure $\alpha_2\beta\beta'\omega\sigma$. They initiate transcription on the antisense (template) strand of a gene at a position designated by its promoter. In *E. coli* the most conserved region of the promoter is centered at about the -10 position and has the consensus sequence TATAAT. The -35 region is also conserved in efficient promoters. DMS footprinting studies indicate that the RNAP holoenzyme forms an open initiation complex with the promoter in which the template and nontemplate DNA strands are separated to form an ~ 14 -nt transcription bubble. RNAPs have the shape of a crab claw. In the elongation complex, the template strand in the transcription bubble passes through a tunnel in the core enzyme to the active site, where it pairs with incoming ribonucleotides. The RNA product exits the enzyme through a channel between its β and β' subunits. In the closed complex of the bacterial RNAP holoenzyme, the σ subunit, which extends along the “top” of the holoenzyme, makes all the sequence-specific contacts with the promoter. After the initiation of RNA synthesis, the σ subunit dissociates from the core enzyme, which then autonomously catalyzes chain elongation in the $5' \rightarrow 3'$ direction. RNA synthesis is terminated by a segment of the transcript that forms a G + C-rich hairpin with an oligo(U) tail that spontaneously dissociates from the DNA. Termination sites that lack these sequences require the assistance of Rho factor for proper chain termination.

In the nuclei of eukaryotic cells, RNAPs I, II, and III, respectively, synthesize rRNA precursors, mRNA precursors, and tRNAs + 5S RNA. The structure of yeast RNAP II resembles that of bacterial RNAPs but is somewhat larger and has more subunits. The structure of its transcribing complex reveals a one-turn segment of RNA · DNA hybrid helix at the active site, which is in contact with the solvent via a pore leading into a funnel through which NTPs presumably pass. RNAPs can hydrolytically correct their mistakes with the aid of TFIIS in eukaryotes and GreA and GreB in bacteria. The minimal RNA polymerase I promoter extends between nucleotides -31 and $+6$. Many RNA polymerase II promoters contain a conserved TATAAAA sequence, the TATA box, located around position -27 . Enhancers are transcriptional activators that can have variable positions and orientations rela-

tive to the transcription start site. RNA polymerase III promoters are located within the transcribed regions of their gene between positions $+40$ and $+80$.

3 Control of Transcription in Prokaryotes Prokaryotes can respond rapidly to environmental changes, in part because the translation of mRNAs commences during their transcription and because most mRNAs are degraded within 1 to 3 min of their synthesis. The expression of specific sets of genes is controlled, in most bacteria and some bacteriophages, by σ factors. The *lac* repressor is a tetrameric protein of identical subunits that, in the absence of inducer, nonspecifically binds to duplex DNA but binds much more tightly to *lac* promoter. The promoter sequence that *lac* repressor protects from nucleic acid digestion has nearly palindromic symmetry. Yet, methylation protection and mutational studies indicate that repressor is not symmetrically bound to promoter. *lac* repressor prevents RNA polymerase from properly initiating transcription at the *lac* promoter.

The presence of glucose represses the transcription of operons specifying certain catabolic enzymes through the mediation of cAMP. On binding cAMP, which accumulates only in the absence of glucose, catabolite gene activator protein (CAP) binds at or immediately upstream of the promoters of these operons, including the *lac* operon, thereby activating their transcription through the binding to the C-terminal domain of the associated RNAP's α subunit (α CTD). CAP's two symmetry equivalent DNA-binding domains each bind in the major groove of their target DNA via a helix-turn-helix (HTH) motif that also occurs in numerous prokaryotic repressors. The binding between these repressors and their target DNAs is mediated by mutually favorable associations between these macromolecules rather than any specific interactions between particular base pairs and amino acid side chains analogous to Watson-Crick base pairing. Sequence-specific interactions between the *met* repressor and its target DNA occur through a 2-fold symmetric antiparallel β ribbon that this protein inserts into the DNA's major groove. *araBAD* transcription is controlled by the levels of L-arabinose and CAP-cAMP through a remarkable complex of the control protein AraC to two binding sites, *araO*₂ and *araI*₁, that forms an inhibitory DNA loop. On binding L-arabinose and when CAP-cAMP is adjacently bound, AraC releases *araO*₂ and instead binds *araI*₂, thereby releasing the loop and activating RNA polymerase to transcribe the *araBAD* operon. The expression of the *lac* operon is also in part controlled by DNA loop formation. The *lac* repressor is a dimer of homodimers, one of which binds to the operator *lacO*₁ and the other to *lacO*₂ or *lacO*₃ to form a DNA loop that may interfere with RNAP binding to the *lac* promoter. The binding of an inducer such as IPTG to a *lac* repressor dimer core domain alters the angle between its two attached DNA-binding domains such that they cannot simultaneously bind to the *lac* operator, thereby weakening the repressor's grip on the DNA.

The expression of the *E. coli* *trp* operon is regulated by both attenuation and repression. On binding tryptophan, its corepressor, *trp* repressor binds to the *trp* operator, thereby blocking *trp* operon transcription. When tryptophan is available, much of the *trp* transcript that has escaped repression is prematurely terminated in the *trpL* sequence because its

transcript contains a segment that forms a normal intrinsic terminator. When tryptophanyl-tRNA^{Trp} is scarce, ribosomes stall at the transcript's two tandem Trp codons. This permits the newly synthesized RNA to form a base-paired stem and loop that prevents the formation of the terminator structure. Several other operons are similarly regulated by attenuation. Riboswitches are mRNA components that regulate gene expression by specifically binding metabolites. The stringent response is another mechanism by which *E. coli* match the rate of transcription to charged tRNA availability. When a specified charged tRNA is scarce, stringent factor on active ribosomes synthesizes ppGpp, which inhibits the transcription of rRNA and some mRNAs while stimulating the transcription of other mRNAs.

4 Post-Transcriptional Processing Most prokaryotic mRNA transcripts require no additional processing. However, eukaryotic mRNAs have an enzymatically appended 5' cap and, in most cases, an enzymatically generated poly(A) tail. Moreover, the introns of eukaryotic mRNA primary transcripts (hnRNAs) are precisely excised via lariat intermediates and their flanking exons are spliced together. Group I and group II introns are self-splicing, that is, their RNAs function as ribozymes (RNA enzymes). Ribozymes, such as the *Tetrahymena* pre-rRNA and hammerhead ribozymes, have complex structures containing several base-paired stems. Pre-mRNAs are spliced by large and complex particles named spliceosomes that consist of four different small nuclear ribonucleoproteins (snRNPs) and which are assisted by the participation of a variety of protein splicing factors. Many eukaryotic proteins consist of modules that also occur in other proteins and hence appear to have evolved via the stepwise collection of exons through recombination events. The alternative splicing of pre-mRNAs greatly increases the variety of proteins expressed by eukaryotic genomes. Certain mRNAs are subject to RNA editing, either by the replacement, inser-

tion, or deletion of specific bases in a process that is directed by guide RNAs (gRNAs), or by substitutional editing mediated by cytidine deaminases or adenosine deaminases.

In RNA interference (RNAi), dsRNA is cleaved by the endoribonuclease Dicer to small interfering RNAs (siRNAs) that guide the hydrolytic cleavage of the complementary mRNAs by the Argonaute component of the RNA-induced silencing complex (RISC), thereby preventing the mRNAs' transcription. MicroRNAs (miRNAs) are generated through the excision of imperfectly base paired stem-loop structures from pri-miRNAs through the actions of Drosha and Dicer. The guide RNA strand of miRNAs binds to RISC and directs it to partially complementary sequences on 3' untranslated regions (3'UTRs) of its target mRNAs, thus inhibiting the expression of the mRNAs and providing a major although only recently recognized mechanism for the regulation of gene expression. Mature eukaryotic mRNAs are selectively and actively transported from the nucleus to the cytosol via the nuclear pore complex. The degradation of mRNAs, which is elaborately controlled, is mediated in part by exosomes.

The primary transcript of *E. coli* rRNAs contains all three rRNAs together with some tRNAs. These are excised and trimmed by specific endonucleases and exonucleases. The eukaryotic 18S, 5.8S, and 28S rRNAs are similarly transcribed as a 45S precursor, which is processed in a manner resembling that of *E. coli* rRNAs. Eukaryotic rRNAs are modified by the methylation of specific nucleosides, as are prokaryotic rRNA, and by the conversion of certain U's to pseudouridines (Ψ's). These processes are guided by small nucleolar RNAs (snoRNAs). Prokaryotic tRNAs are excised from their primary transcripts and trimmed in much the same way as are rRNAs. In RNase P, one of the enzymes mediating this process, the catalytic subunit is an RNA. Eukaryotic tRNA transcripts also require the excision of a short intron and the enzymatic addition of a 3'-terminal —CCA to form the mature tRNA.

REFERENCES

General

- Brown, T.A., *Genomes* (3rd ed.), Chapter 12, Wiley-Liss (2007).
 Gesteland, R.F., Cech, T.R., and Atkins, J.F. (Eds.), *The RNA World* (3rd ed.), Cold Spring Harbor Laboratory (2006). [A series of authoritative articles on the nature of the prebiotic "RNA world" as revealed by the RNA "relics" in modern organisms.]
 Hodgson, D.A. and Thomas, C.M. (Eds.), *Signals, Switches, Regulons and Cascades: Control of Bacterial Gene Expression*, Cambridge University Press (2002).
 Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R., *Molecular Biology of the Gene* (6th ed.), Chapters 12, 13, 16, and 18, Cold Spring Harbor Laboratory Press (2008).

The Genetic Role of RNA

- Brachet, J., Reminiscences about nucleic acid cytochemistry and biochemistry, *Trends Biochem. Sci.* **12**, 244–246 (1987).
 Brenner, S., Jacob, F., and Meselson, M., An unstable intermediate carrying information from genes to ribosomes for protein synthesis, *Nature* **190**, 576–581 (1960). [The experimental verification of mRNA's existence.]
 Crick, F., Central dogma of molecular biology, *Nature* **227**, 561–563 (1970).

- Hall, B.D. and Spiegelman, S., Sequence complementarity of T2-DNA and T2-specific RNA, *Proc. Natl. Acad. Sci.* **47**, 137–146 (1964). [The first use of RNA–DNA hybridization.]
 Jacob, F. and Monod, J., Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318–356 (1961). [The classic paper postulating the existence of mRNA and operons and explaining how the transcription of operons is regulated.]
 Pardee, A.B., Jacob, F., and Monod, J., The genetic control and cytoplasmic expression of "inducibility" in the synthesis of β-galactosidase by *E. coli*, *J. Mol. Biol.* **1**, 165–178 (1959). [The Pajam experiment.]
 Thieffry, D., Forty years under the central dogma, *Trends Biochem. Sci.* **23**, 312–316 (1998).

RNA Polymerase and mRNA

- Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A., Structural mechanism for rifampicin inhibition of bacterial RNA polymerase, *Cell* **104**, 901–912 (2001).
 Cramer, P., et al., Structure of eukaryotic RNA polymerases, *Annu. Rev. Biophys.* **37**, 337–352 (2008).
 Cramer, P. and Arnold, A., Proteins: how RNA polymerases work, *Curr. Opin. Struct. Biol.* **19**, 680–682 (2009). [The editorial]

- overview of a series of authoritative reviews on numerous aspects of RNA polymerases. See, in particular, the reviews beginning on pages 691, 701, 708, and 732.]
- Cramer, P., Bushnell, D.A., and Kornberg, R.D., Structural basis of transcription: RNA polymerase at 2.8 Å resolution, *Science* **292**, 1863–1876 (2001); and Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D., Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution, *Science* **292**, 1876–1882 (2001).
- Dahmus, M.E., Reversible phosphorylation of the C-terminal domain of RNA polymerase II, *J. Biol. Chem.* **271**, 19009–19012 (1996).
- Darst, S.A., Bacterial RNA polymerase, *Curr. Opin. Struct. Biol.* **11**, 155–162 (2001).
- Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L., Identification of an UP element consensus sequence for bacterial promoters, *Proc. Natl. Acad. Sci.* **95**, 9761–9766 (1998).
- Gilmour, D.S. and Fan, R., Derailing the locomotive: transcription termination, *J. Biol. Chem.* **283**, 661–664 (2008).
- Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., and Kinoshita, K., Jr., Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase, *Nature* **409**, 113–115 (2001).
- Kadonaga, K.T., The RNA polymerase II core promoter, *Annu. Rev. Biochem.* **72**, 449–479 (2003).
- Kaplan, C.D., Larsson, K.-M., and Kornberg, R.D., The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by α-amanitin, *Mol. Cell* **30**, 547–556 (2008).
- Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M., Applied force reveals mechanistic and energetic details of transcription termination, *Cell* **132**, 971–982 (2008); and Brueckner, F. and Cramer, P., Structural basis for transcription inhibition by α-amanitin and implications for RNA polymerase II translocation, *Nature Struct. Mol. Biol.* **15**, 811–818 (2008).
- Lian, C., Robinson, H., and Wang, A.H.-J., Structure of actinomycin D bound with (GAAGCTTC)₂ and (GATGCTTC)₂ and its binding to the (CAG)_n(CTG)_n triplet sequence as determined by NMR analysis, *J. Am. Chem. Soc.* **118**, 8791–8801 (1996).
- Mooney, R.A., Darst, S.A., and Landick, R., Sigma and RNA polymerase: An on-again, off-again relationship? *Mol. Cell.* **20**, 335–345 (2006).
- Murikami, K.S., Masuda, S., and Darst, S., Structural basis of transcription initiation: RNA polymerase at 4 Å resolution, *Science* **296**, 1280–1284 (2002); Murikami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S., Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex, *Science* **296**, 1285–1290 (2002); and Murakami, K.S. and Darst, S.A., Bacterial RNA polymerases: the whole story, *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
- Nudler, E., RNA polymerase active center: the molecular engine of transcription, *Annu. Rev. Biochem.* **78**, 335–361 (2009).
- Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R., Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching; and Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H., Initial transcription by RNA polymerase proceeds through DNA-scrunching mechanism, *Science* **314**, 1139–1143 and 1144–1150 (2006).
- Reynolds, R. and Chamberlin, M.J., Parameters affecting transcription termination by *Escherichia coli* RNA. II. Construction and analysis of hybrid terminators, *J. Mol. Biol.* **224**, 53–63 (1992).
- Roberts, J.W., Shankar, S., and Filter, J.J., RNA polymerase elongation factors, *Annu. Rev. Microbiol.* **62**, 211–233 (2008); and Park, J.-S. and Roberts, J.W., Role of DNA bubble rewinding in enzymatic transcription termination, *Proc. Natl. Acad. Sci.* **103**, 4870–4875 (2006).
- Shilatifard, A., Conway, R.C., and Conway, J.W., The RNA polymerase II elongation complex, *Annu. Rev. Biochem.* **72**, 693–715 (2003).
- Skordalakes, E. and Berger, J.M., The structure of Rho transcription terminator: Mechanism of mRNA recognition and helicase loading, *Cell* **114**, 135–146 (2003).
- Svejstrup, J.Q., Contending with transcriptional arrest during RNAPII transcript elongation, *Trends Biochem. Sci.* **32**, 165–171 (2007).
- Thompson, N.D. and Berger, J.M., Running in reverse: The structural basis for translocation polarity in hexameric helicases, *Cell* **139**, 523–534 (2009).
- Vassylyev, D.G., Sekine, S.-I., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S., Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution, *Nature* **417**, 712–718 (2002).
- Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I., Structural basis for transcription elongation by bacterial RNA polymerase; and Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R., Structural basis for substrate loading in bacterial RNA polymerase, *Nature* **448**, 157–162 and 163–168 (2007).
- Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D., Structural basis of transcription: Backtracked RNA polymerase II at 3.4 Å resolution, *Science* **324**, 1203–1206 (2009); and Sydow, J.F., Brueckner, F., Cheung, A.C.M., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P., Structural basis of transcription: Mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA, *Mol. Cell* **34**, 710–721 (2009).
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D., Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis, *Cell* **127**, 941–954 (2006).
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A., Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution, *Cell* **98**, 811–824 (1999).

Control of Transcription in Prokaryotes

- Anderson, J.E., Ptashne, M., and Harrison, S.C., The structure of the repressor-operator complex of bacteriophage 434, *Nature* **326**, 846–852 (1987).
- Gilbert, W. and Müller-Hill, B., Isolation of the lac repressor, *Proc. Natl. Acad. Sci.* **56**, 1891–1898 (1966).
- Harmor, T., Wu, M., and Schleif, R., The role of rigidity in DNA looping-unlooping by AraC, *Proc. Natl. Acad. Sci.* **98**, 427–431 (2001).
- Huffman, J.L. and Brennan, R.G., Prokaryotic transcription regulators: more than just the helix-turn-helix motif, *Curr. Opin. Struct. Biol.* **12**, 98–106 (2002).
- Kolter, R. and Yanofsky, C., Attenuation in amino acid biosynthetic operons, *Annu. Rev. Genet.* **16**, 113–134 (1982).
- Lamond, A.I. and Travers, A.A., Stringent control of bacterial transcription, *Cell* **41**, 6–8 (1985).
- Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M., and Ebright, R.H., Catabolite activator protein: DNA binding and transcription activation, *Curr. Opin. Struct. Biol.* **14**, 10–20 (2004); and Bennoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M., and Ebright, R.H., Structural basis of transcription activation:

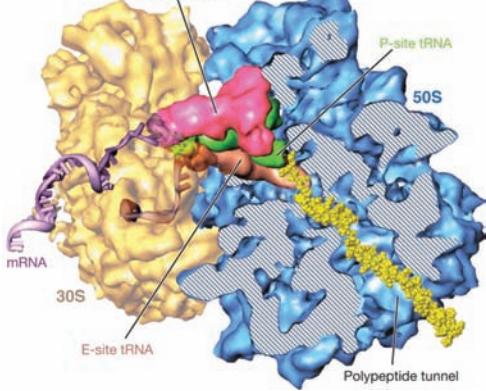
- The CAP- α CTD-DNA complex, *Science* **297**, 1562–1566 (2002).
- Lewis, M., The lac repressor, *C. R. Biologies* **328**, 521–548 (2005).
- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P., Crystal structure of the lactose operon repressor and its complexes with DNA and inducer, *Science* **271**, 1247–1254 (1996).
- Mondragón, A. and Harrison, S.C., The phage 434 Cro/O_R1 complex at 2.5 Å resolution, *J. Mol. Biol.* **219**, 321–334 (1991); and Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S.C., Structure of phage 434 Cro/DNA complex, *Nature* **335**, 789–795 (1988).
- Montange, R.K. and Batey, R.T., Riboswitches: emerging themes in RNA structure and function, *Annu. Rev. Biophys.* **387**, 117–133 (2007).
- Mooney, R.A., Darst, S.A., and Landick, R., Sigma and RNA polymerase: an on-again, off-again relationship, *Mol. Cell* **20**, 335–345 (2005).
- Oehler, S., Eismann, E.R., Krämer, H., and Müller-Hill, B., The three operators of the lac operon cooperate in repression, *EMBO J.* **9**, 973–979 (1990).
- Pace, H.C., Kercher, M.A., Lu, P., Markiewicz, P., Miller, J.H., Chang, G., and Lewis, M., Lac repressor genetic map in real space, *Trends Biochem. Sci.* **22**, 334–339 (1997).
- Reeder, T. and Schleif, R., AraC protein can activate transcription from only one position and when pointed in only one direction, *J. Mol. Biol.* **231**, 205–218 (1993).
- Romanuka, J., Folkers, G.E., Biris, N., Tishchenko, E., Wienk, H., Bonvin, A.M.J.J., Kaptein, R., and Boelens, R., Specificity and affinity of lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein–DNA complexes, *J. Mol. Biol.* **390**, 478–489 (2009).
- Roth, A. and Breaker, R.R., The structural and functional diversity of metabolite-binding riboswitches, *Annu. Rev. Biochem.* **78**, 305–334 (2009).
- Schleif, R., AraC protein: a love–hate relationship, *BioEssays* **25**, 274–282 (2003); and Regulation of the L-arabinose operon of *Escherichia coli*, *Trends Genet.* **16**, 559–565 (2000).
- Schultz, S.C., Shields, G.C., and Steitz, T.A., Crystal structure of a CAP-DNA complex: The DNA is bent by 90°, *Science* **253**, 1001–1007 (1991).
- Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J., Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch, *Nature* **441**, 1167–1171 (2006).
- Shakked, Z., Guzikevich-Guerstein, G., Frolov, F., Rabinovich, D., Joachimiak, A., and Sigler, P.B., Determinants of repressor/operator recognition from the structure of the trp operator binding site, *Nature* **368**, 469–473 (1994).
- Soisson, S.M., MacDougall-Shackleton, B., Schleif, R., and Wolberger, C., Structural basis for ligand-regulated oligomerization of AraC, *Science* **276**, 421–425 (1997). [The X-ray structure of AraC alone and in complex with arabinose.]
- Somers, W.S. and Phillips, S.E.V., Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands, *Nature* **359**, 387–393 (1992).
- Yanofsky, C., Transcription attenuation, *J. Biol. Chem.* **263**, 609–612 (1988); and Attenuation in the control of expression of bacterial operons, *Nature* **289**, 751–758 (1981).
- Post-transcriptional Processing**
- Bachellerie, J.-P. and Cavaillé, J., Guiding ribose methylation of rRNA, *Trends Biochem. Sci.* **22**, 257–261 (1997).
- Balbo, P.B. and Bohm, A., Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis, *Structure* **15**, 1117–2232 (2007).
- Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencoe, B.J., and Frey, B.J., Deciphering the splicing code, *Nature* **465**, 53–59 (2010).
- Bartel, D.P., MicroRNAs: target recognition and regulatory function, *Cell* **136**, 215–233 (2009); and MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**, 281–297 (2004).
- Bass, B.L., RNA editing by adenosine deaminases that act on RNA, *Annu. Rev. Biochem.* **71**, 817–846 (2002).
- Blencowe, B.J., Alternative splicing: new insights from global analyses, *Cell* **126**, 37–47 (2006).
- Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Lührmann, R., and Stark H., Three-dimensional structure of pre-catalytic human spliceosomal complex B, *Nature Struct. Biol.* **11**, 463–468 (2004).
- Brantl, S., Antisense regulation and RNA interference, *Biochim. Biophys. Acta* **1575**, 15–25 (2002).
- Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R., Control of alternative RNA splicing and gene expression by eukaryotic riboswitches, *Nature* **447**, 497–500 (2007).
- Chen, M. and Manley, J.L., Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches, *Nature Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
- Cole, C.N. and Scarcelli, J.J., Transport of messenger RNA from the nucleus to the cytoplasm, *Curr. Opin. Cell. Biol.* **18**, 299–306 (2006).
- Davis, R.E., Spliced leader RNA trans-splicing in metazoa, *Parasitology Today* **12**, 33–40 (1996).
- Decatur, W.A. and Fournier, M.J., RNA-guided nucleotide modification of ribosomal and other RNAs, *J. Biol. Chem.* **278**, 695–698 (2003).
- Evans, D., Marquez, S.M., and Pace, N.R., RNase P: interface of the RNA and protein worlds, *Trends Biochem. Sci.* **31**, 333–341 (2006).
- Fedor, M.J., Comparative enzymology and structural biology of RNA self-cleavage, *Annu. Rev. Biophys.* **38**, 271–299 (2009).
- Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N., Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
- Gannan, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P., Organization and sequences of the 5' end of a cloned complete ovalbumin gene, *Nature* **278**, 428–434 (1979).
- Garneau, N.L., Wilusz, J., and Wilusz, C.J., The highways and byways of mRNA decay, *Nature Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
- Gerber, A.P. and Keller, W., RNA editing by base deamination: more enzymes, more targets, new mysteries, *Trends Biochem. Sci.* **26**, 376–384 (2001).
- Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G., RNA-binding proteins and post-transcriptional gene regulation, *FEBS Lett.* **582**, 1977–1986 (2008).
- Gott, J.M. and Emeson, R.B., Functions and mechanisms of RNA editing, *Annu. Rev. Genet.* **34**, 499–531 (2000).
- Gu, M. and Lima, C.D., Processing the message: structural insights into capping and decapping mRNA, *15*, 99–106 (2005).
- Guo, F., Gooding, A.R., and Cech, T.R., Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site, *Mol. Cell* **16**, 351–362 (2004).
- Jinek, M. and Doudna, J.A., A three-dimensional view of the molecular machinery of RNA interference, *Nature* **457**, 405–412 (2009).
- Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., and Nagai, K., Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs, *Cell* **96**, 375–387 (1999).

- Kawamata, T. and Tomari, Y., Making RISC, *Trends Biochem. Sci.* **35**, 368–376 (2010).
- Keegan, L.P., Gallo, A., and O’Connell, M.A., The many roles of an RNA editor, *Nature Rev. Genet.* **2**, 869–878 (2001).
- Krummel, D.A.P., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K., Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution, *Nature* **458**, 475–480 (2009).
- Li, Y. and Breaker, R.R., Deoxyribozymes: new players in an ancient game of biocatalysis, *Curr. Opin. Struct. Biol.* **9**, 315–323 (1999).
- Liu, Q., Greimann, J.C. and Lima, C.D., Reconstitution, activities, and structure of the eukaryotic RNA exosome, *Cell* **127**, 1223–1237 (2006); and Errata, *Cell* **131**, 188–190 (2007).
- Maas, S., Rich, A., and Nishikura, K., A-to-I RNA editing: Recent news and residual mysteries, *J. Biol. Chem.* **278**, 1391–1394 (2003); and Blanc, V. and Davidson, N.O., C-to-U RNA editing: Mechanisms leading to genetic diversity, *J. Biol. Chem.* **278**, 1395–1398 (2003).
- MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A., Structural basis for double-stranded RNA processing by Dicer, *Science* **311**, 195–198 (2006).
- Madison-Antenucci, S., Grams, J., and Hajduk, S.L., Editing machines: the complexities of trypanosome editing, *Cell* **108**, 435–438 (2002).
- Maniatis, T. and Tasic, B., Alternative pre-mRNA splicing and proteome expansion in metazoans, *Nature* **418**, 236–243 (2002).
- Martick, M. and Scott, W.G., Tertiary contacts distant from the active site prime a ribozyme for catalysis, *Cell* **126**, 309–320 (2006). [The X-ray structure of *S. mansoni* hammerhead ribozyme.]
- Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D.S., The crystal structure of a heptameric archaeal Sm protein: Implication for the eukaryotic snRNP core, *Proc. Natl. Acad. Sci.* **98**, 5532–5537 (2001).
- Nishikura, K., Functions and regulation of RNA editing by ADAR deaminases, *Annu. Rev. Biochem.* **79**, 321–349 (2010); and Hundley, H.A. and Bass, B.L., ADAR editing in double-stranded UTRs and other noncoding sequences, *Trends Biochem. Sci.* **35**, 377–383 (2010).
- Pratt, A.J. and MacRae, I.J., The RNA-induced silencing complex: a versatile gene-silencing machine, *J. Biol. Chem.* **284**, 17897–17901 (2009).
- Proudfoot, N.J., Furger, A., and Dye, M.J., Integrating mRNA processing with transcription, *Cell* **108**, 501–512 (2002).
- Schmid, M. and Jensen, T.H., The exosome: a multipurpose RNA-decay machine, *Trends Biochem. Sci.* **33**, 501–510 (2008).
- Serganov, A. and Patel, D.J., Ribozymes, riboswitches and beyond: regulation of gene expression without proteins, *Nature Rev. Genet.* **8**, 776–790 (2007).
- Sharp, P.A., Split genes and RNA splicing, *Cell* **77**, 805–815 (1994).
- Smith, H.C. (Ed.), *RNA and DNA Editing*, Wiley (2008).
- Stahley, M.R. and Strobel, S.A., RNA splicing: group I crystal structures reveal the basis of splice site selection and metal ion catalysis, *Curr. Opin. Struct. Biol.* **16**, 319–326 (2006).
- Stark, H. and Lührmann, R., Cryo-electron microscopy of spliceosomal components, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 435–457 (2006).
- Stuart, K.D., Schnaufer, A., Ernst, N.E., and Panigrahi, A.K., Complex management: RNA editing in trypanosomes, *Trends Biochem. Sci.* **30**, 97–105 (2005).
- Tanaka Hall, T.M., Poly(A) tail synthesis and regulation: recent structural insights, *Curr. Opin. Struct. Biol.* **12**, 82–88 (2002).
- Tarn, W.-Y. and Steitz, J.A., Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge, *Trends Biochem. Sci.* **22**, 132–137 (1997).
- Toor, N., Keating, K.S., and Pyle, A.M., Structural insights into RNA splicing, *Curr. Opin. Struct. Biol.* **19**, 260–266 (2009).
- Torres-Larios, A., Swinger, K.K., Krasilnikov, A.S., Pan, T., and Mondragón, A., Crystal structure of the RNA component of bacterial ribonuclease P, *Nature* **437**, 584–587 (2005); and Kazantsev, A.V., Krivenko, A.A., Harrington, D.J., Holbrook, S.R., Adams, P.D., and Pace, N.R., Crystal structure of a bacterial ribonuclease P RNA, *Proc. Natl. Acad. Sci.* **102**, 13392–13397 (2005).
- Vicens, Q. and Cech, T.A., Atomic level architecture of group I introns revealed, *Trends Biochem. Sci.* **31**, 41–51 (2006).
- Wahl, M.C., Will, C.L., and Lührmann, R., The spliceosome: design principles in a dynamic RNP machine, *Cell* **136**, 701–718 (2009). [A detailed review.]
- Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J., Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes, *Nature* **461**, 754–762 (2009). [X-ray structures of Argonaute–DNA · RNA complexes.]
- Weinstein, L.B. and Steitz, J.A., Guided tours: from precursor to snoRNA to functional snoRNP, *Curr. Opin. Cell Biol.* **11**, 378–384 (1999).
- Xiao, S., Scott, F., Fierke, C.A., and Enelke, D.R., Eukaryotic ribonuclease P: A plurality of ribonucleoprotein enzymes, *Annu. Rev. Biochem.* **71**, 165–189 (2002).

PROBLEMS

- Indicate the phenotypes of the following *E. coli* lac partial diploids in terms of inducibility and active enzymes synthesized.
 - $I^-P^+O^+Z^+Y^-/I^+P^-O^+Z^+Y^+$
 - $I^-P^+O^cZ^+Y^-/I^+P^+O^+Z^-Y^+$
 - $I^-P^+O^cZ^+Y^+/I^-P^+O^+Z^+Y^+$
 - $I^+P^-O^cZ^+Y^+/I^-P^+O^cZ^-Y^-$
 - Superrepressed** mutants, I^s , encode lac repressors that bind operator but do not respond to the presence of inducer. Indicate the phenotypes of the following genotypes in terms of inducibility and enzyme production.
 - $I^sO^+Z^+$
 - $I^sO^cZ^+$
 - $I^+O^+Z^+/I^sO^+Z^+$
 - Why do $lacZ^-$ *E. coli* fail to show galactoside permease activity after the addition of lactose in the absence of glucose? Why do lac Y^- mutants lack β -galactosidase activity under the same conditions?
 - What is the experimental advantage of using IPTG instead of 1,6-*allo*lactose as an inducer of the lac operon?
 - Describe the probable genetic defect that abolishes the sensitivity of the lac operon to the absence of glucose when other metabolic operons continue to be sensitive to the absence of glucose.
 - Indicate the -10 region, the -35 region, and the initiating nucleotide on the sense strand of the *E. coli* tRNA^{Tyr} promoter shown below.
- 5' CAACGTAACACTTACAGCGCGCGTCATTGATATGATGCGCCCGCTTCCCGATA 3'
3' GTTGCATTGTGAAATGTCGCCCGCAGTAACCTACTACCGGGGGCAAGGGCTAT 5'
- Why are *E. coli* that are diploid for rifamycin resistance and rifamycin sensitivity (rif^R/rif^S) sensitive to rifamycin?

- 8.** Why does promoter efficiency tend to decrease with the number of G · C base pairs in the -10 region of a prokaryotic gene?
- 9.** A eukaryotic ribosome contains 4 different rRNA molecules and ~ 82 different proteins. Why does a cell contain many more copies of the rRNA genes than the ribosomal protein genes?
- 10.** What is the probability that the 4026-nucleotide DNA sequence coding for the β subunit of *E. coli* RNA polymerase will be transcribed with the correct base sequence? Perform the calculations for the probabilities of 0.0001, 0.001, and 0.01 that each base is incorrectly transcribed.
- 11.** If an enhancer is placed on one plasmid and its corresponding promoter is placed on a second plasmid that is catenated (linked) with the first, initiation is almost as efficient as when the enhancer and promoter are on the same plasmid. However, initiation does not occur when the two plasmids are unlinked. Explain.
- 12.** What is the probability that the symmetry of the *lac* operator is merely accidental?
- 13.** Why does the inhibition of DNA gyrase in *E. coli* inhibit the expression of catabolite-sensitive operons?
- 14.** Describe the transcription of the *trp* operon in the absence of active ribosomes and tryptophan.
- 15.** Why can't eukaryotic transcription be regulated by attenuation?
- 16.** Predict the effect of deleting the leader peptide sequence on regulation of the *trp* operon.
- 17.** Charles Yanofsky and his associates have synthesized a 15-nucleotide RNA that is complementary to segment 1 of *trpL* mRNA (but only partially complementary to segment 3). What is its effect on the *in vitro* transcription of the *trp* operon? What is its effect if the *trpL* gene contains a mutation in segment 2 that destabilizes the 2 · 3 stem and loop?
- 18.** Why are *relA*⁻ mutants defective in the *in vivo* transcription of *his* and *trp* operons?
- 19.** Why aren't primary rRNA transcripts present in wild-type *E. coli*?
- 20.** Why can't hammerhead ribozymes catalyze the cleavage of ssDNA?
- 21.** Explain why the active site of poly(A) polymerase is much narrower than that of DNA and RNA polymerases.
- 22.** Would you expect spliceosome-catalyzed intron removal to be reversible in a highly purified *in vitro* system and *in vivo*? Explain.
- 23.** Introns in eukaryotic protein-coding genes may be quite large, but almost none are smaller than about 65 bp. What is the reason for this minimum intron size?
- 24.** Infection with certain viruses inhibits snRNA processing in eukaryotic cells. Explain why this favors the expression of viral genes in the host cell.
- 25.** Explain why RNAi would be a less efficient mechanism for regulating the expression of specific genes if Dicer hydrolyzed double-stranded RNA every 11 bp rather than every 22 bp.



Translation

CHAPTER 32

1 The Genetic Code

- A. Chemical Mutagenesis
- B. Codons Are Triplets
- C. Deciphering the Genetic Code
- D. The Nature of the Code

2 Transfer RNA and Its Aminoacylation

- A. Primary and Secondary Structures of tRNA
- B. Tertiary Structure of tRNA
- C. Aminoacyl-tRNA Synthetases
- D. Codon–Anticodon Interactions
- E. Nonsense Suppression

3 Ribosomes and Polypeptide Synthesis

- A. Ribosome Structure
- B. Polypeptide Synthesis: An Overview
- C. Chain Initiation
- D. Chain Elongation
- E. Translational Accuracy
- F. Chain Termination
- G. Protein Synthesis Inhibitors: Antibiotics

4 Control of Eukaryotic Translation

- A. Regulation of eIF2
- B. Regulation of eIF4E
- C. mRNA Masking and Cytoplasmic Polyadenylation
- D. Antisense Oligonucleotides

5 Post-Translational Modification

- A. Proteolytic Cleavage
- B. Covalent Modification
- C. Protein Splicing: Inteins and Exteins

6 Protein Degradation

- A. Degradation Specificity
- B. Degradation Mechanisms

translation process. Following this, we consider the structure and functions of **ribosomes**, the complex molecular machines that catalyze peptide bond formation between the mRNA-specified amino acids. Peptide bond formation, however, does not necessarily yield a functional protein; many polypeptides must first be post-translationally modified as we discuss in the subsequent section. Finally, we study how cells degrade proteins, a process that must balance protein synthesis.

1 THE GENETIC CODE

How does DNA encode genetic information? According to the one gene–one polypeptide hypothesis, the genetic message dictates the amino acid sequences of proteins. Since the base sequence of DNA is the only variable element in this otherwise monotonously repeating polymer, the amino acid sequence of a protein must somehow be specified by the base sequence of the corresponding segment of DNA.

A DNA base sequence might specify an amino acid sequence in many conceivable ways. With only 4 bases to code for 20 amino acids, a group of several bases, termed a **codon**, is necessary to specify a single amino acid. A triplet code, that is, one with 3 bases per codon, is minimally required since there are $4^3 = 64$ different triplets of bases, whereas there can be only $4^2 = 16$ different doublets, which is insufficient to specify all the amino acids. In a triplet code, as many as 44 codons might not code for amino acids. On the other hand, many amino acids could be specified by more than one codon. Such a code, in a term borrowed from mathematics, is said to be **degenerate**.

Another mystery was, how does the polypeptide synthesizing apparatus group DNA's continuous sequence of bases into codons? For example, the code might be overlapping; that is, in the sequence

ABCDEFGHIJ...

ABC might code for one amino acid, BCD for a second, CDE for a third, and so on. Alternatively, the code might be nonoverlapping, so that ABC specifies one amino acid, DEF a second, GHI a third, and so on. The code might also contain internal “punctuation” such as in the nonoverlapping triplet code

ABC,DEF,GHI,...

In this chapter we consider **translation**, the mRNA-directed biosynthesis of polypeptides. Although peptide bond formation is a relatively simple chemical reaction, the complexity of the translational process, which involves the coordinated participation of over 100 macromolecules, is mandated by the need to link 20 different amino acid residues accurately in the order specified by a particular mRNA. Note that we previewed this process in Section 5-4B.

We begin by discussing the **genetic code**, the correspondence between nucleic acid sequences and polypeptide sequences. Next, we examine the structures and properties of **tRNAs**, the amino acid–bearing entities that mediate the

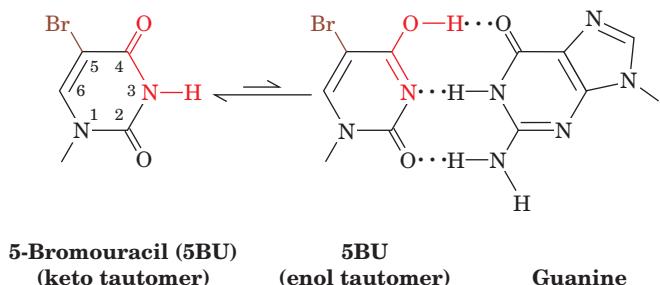


Figure 32-1 5-Bromouracil. Its keto form (left) is its most common tautomer. However, it frequently assumes the enol form (right), which base pairs with guanine.

in which the commas represent particular bases or base sequences. A related question is, how does the genetic code specify the beginning and the end of a polypeptide chain?

The genetic code is, in fact, a nonoverlapping, comma-free, degenerate, triplet code. How this was determined and how the genetic code dictionary was elucidated are the subjects of this section.

A. Chemical Mutagenesis

The triplet character of the genetic code, as we shall see below, was established through the use of **chemical mutagens**, substances that chemically induce mutations. We therefore precede our study of the genetic code with a discussion of these substances. There are two major classes of mutations:

1. Point mutations, in which one base pair replaces another. These are subclassified as

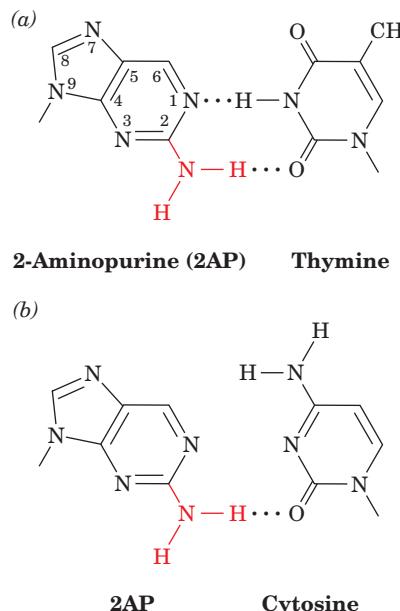


Figure 32-2 Base pairing by the adenine analog 2-aminopurine. It normally base pairs with thymine (a) but occasionally also does so with cytosine (b).

- (a) **Transitions**, in which one purine (or pyrimidine) is replaced by another.
- (b) **Transversions**, in which a purine is replaced by a pyrimidine or vice versa.

2. Insertion/deletion mutations, in which one or more nucleotide pairs are inserted in or deleted from DNA.

A mutation in any of these three categories may be reversed by a subsequent mutation of the same but not another category.

a. Point Mutations Are Generated by Altered Bases

Point mutations can result from the treatment of an organism with base analogs or with substances that chemically alter bases. For example, the base analog **5-bromouracil (5BU)** sterically resembles thymine (5-methyluracil) but, through the influence of its electronegative Br atom, frequently assumes a tautomeric form that base pairs with guanine instead of adenine (Fig. 32-1). Consequently, when 5BU is incorporated into DNA in place of thymine, as it usually is, it occasionally induces an $A \cdot T \rightarrow G \cdot C$ transition in subsequent rounds of DNA replication. Occasionally, 5BU is also incorporated into DNA in place of cytosine, which instead generates a $G \cdot C \rightarrow A \cdot T$ transition.

The adenine analog **2-aminopurine (2AP)**, normally base pairs with thymine (Fig. 32-2a) but occasionally forms an undistorted but singly hydrogen bonded base pair with cytosine (Fig. 32-2b). Thus 2AP generates $A \cdot T \rightarrow G \cdot C$ transitions.

In aqueous solutions, **nitrous acid** (HNO_2) oxidatively deaminates aromatic primary amines so that it converts cytosine to uracil (Fig. 32-3a) and adenine to the guanine-like **hypoxanthine** (which forms two of guanine's three hydrogen

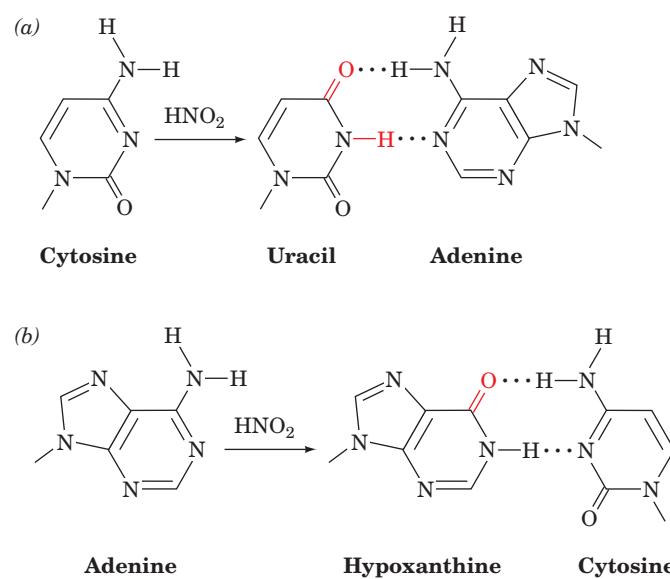


Figure 32-3 Oxidative deamination by nitrous acid. (a) Cytosine is converted to uracil, which base pairs with adenine. (b) Adenine is converted to hypoxanthine, a guanine derivative (it lacks guanine's 2-amino group) that base pairs with cytosine.

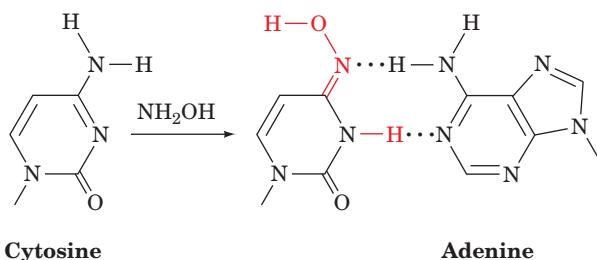
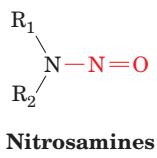


Figure 32-4 Reaction with hydroxylamine converts cytosine to a derivative that base pairs with adenine.

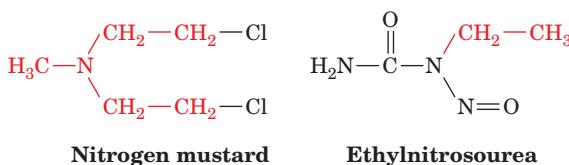
bonds with cytosine; Fig. 32-3b). Hence, treatment of DNA with nitrous acid, or compounds such as **nitrosamines**



that react to form nitrous acid, results in both $\text{A} \cdot \text{T} \rightarrow \text{G} \cdot \text{C}$ and $\text{G} \cdot \text{C} \rightarrow \text{A} \cdot \text{T}$ transitions. **Hydroxylamine** (NH_2OH) also induces $\text{G} \cdot \text{C} \rightarrow \text{A} \cdot \text{T}$ transitions by specifically reacting with cytosine to convert it to a compound that base pairs with adenine (Fig. 32-4).

Nitrite, the conjugate base of nitrous acid, has long been used as a preservative of prepared meats such as frankfurters. However, the observation that many mutagens are also carcinogens (Section 30-5Fa) suggests that the consumption of nitrite-containing meat is harmful to humans. Proponents of nitrite preservation nevertheless argue that to stop it would result in far more fatalities. This is because lack of such treatment would greatly increase the incidence of **botulism**, an often fatal form of food poisoning caused by the ingestion of protein neurotoxins secreted by the anaerobic bacterium *Clostridium botulinum* (Section 12-4Dd).

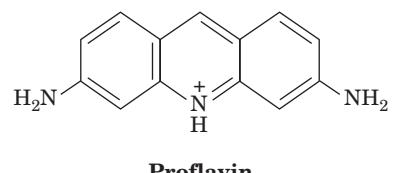
The use of alkylating agents such as dimethyl sulfate, **nitrogen mustard**, and **ethylnitrosourea**



often generates transversions. The alkylation of the N7 position of a purine nucleotide causes its subsequent depurination. The resulting gap in the sequence is filled in by an error-prone repair system (Section 30-5D). Transversions arise when the missing purine is replaced by a pyrimidine. The repair of DNA that has been damaged by UV radiation may also generate transversions.

b. Insertion/Deletion Mutations Are Generated by Intercalating Agents

Insertion/deletion mutations (also called indels), may arise from the treatment of DNA with intercalating agents such as acridine orange (Section 6-6Ca) or proflavin.



The distance between two consecutive base pairs is doubled by the intercalation of such a molecule between them. The replication of such a distorted DNA occasionally results in the insertion or deletion of one or more nucleotides in the newly synthesized polynucleotide. (Insertions and deletions of large DNA segments generally arise from aberrant crossover events; Section 34-2De.)

B. Codons Are Triplets

In 1961, Francis Crick and Sydney Brenner, through genetic investigations into the previously unknown character of proflavin-induced mutations, determined the triplet character of the genetic code. In bacteriophage T4, a particular proflavin-induced mutation, designated *FC0*, maps in the *rIIB* cistron (Section 1-4Eb). The growth of this mutant phage on a permissive host (*E. coli* B) resulted in the occasional spontaneous appearance of phenotypically wild-type phages as was demonstrated by their ability to grow on a restrictive host [*E. coli* K12(λ)]; recall that *rIIB* mutants form characteristically large plaques on *E. coli* B but cannot lyse *E. coli* K12(λ); Section 1-4Eb]. Yet, these doubly mutated phages are not genotypically wild type; the simultaneous infection of a permissive host by one of them and true wild-type phage yielded recombinant progeny that have either the *FC0* mutation or a new mutation designated *FC1*. Thus the phenotypically wild-type phage is a double mutant that actually contains both *FC0* and *FC1*. *These two genes are therefore suppressors of one another; that is, they cancel each other's mutant properties.* Furthermore, since they map together in the *rIIB* cistron, they are mutual **intragenic suppressors** (suppressors in the same gene).

The treatment of *FC1* in a manner identical to that described for *FC0* provided similar results: the appearance of a new mutant, *FC2*, that is an intragenic suppressor of *FC1*. By proceeding in this iterative manner, Crick and Brenner collected a series of different *rIIB* mutants, *FC3*, *FC4*, *FC5*, etc., in which each mutant *FC(n)* is an intragenic suppressor of its predecessor, *FC(n - 1)*. Recombination studies showed, moreover, that odd-numbered mutations are intragenic suppressors of even-numbered mutations, but neither pairs of different odd-numbered mutations nor pairs of different even-numbered mutations suppress each other. However, recombinants containing three odd-numbered mutations or three even-numbered mutations all are phenotypically wild type.

Crick and Brenner accounted for these observations by the following set of hypotheses:

1. The proflavin-induced mutation *FC0* is either an insertion or a deletion of one nucleotide pair from the *rIIB* cistron. If it is a deletion then *FC1* is an insertion, *FC2* is a deletion, and so on, and vice versa.

2. The code is read in a sequential manner starting from a fixed point in the gene. The insertion or deletion of a nucleotide shifts the **frame** (grouping) in which succeeding nucleotides are read as codons (insertions or deletions of nucleotides are therefore also known as **frameshift mutations**). Thus the code has no internal punctuation that indicates the reading frame; that is, *the code is comma free*.

3. The code is a triplet code.

4. All or nearly all of the 64 triplet codons code for an amino acid; that is, *the code is degenerate*.

These principles are illustrated by the following analogy. Consider a sentence (gene) in which the words (codons) each consist of three letters (bases).

THE BIG RED FOX ATE THE EGG

(Here the spaces separating the words have no physical significance; they are only present to indicate the reading frame.) The deletion of the fourth letter, which shifts the reading frame, changes the sentence to

THE IGR EDF OXA TET HEE GG

so that all words past the point of deletion are unintelligible (specify the wrong amino acids). An insertion of any letter, however, say an X in the ninth position,

THE IGR EDX FOX ATE THE EGG

restores the original reading frame. Consequently, only the words between the two changes (mutations) are altered. As in this example, such a sentence might still be intelligible (the gene could still specify a functional protein), particularly if the changes are close together. Two deletions or two insertions, no matter how close together, would not suppress each other but just shift the reading frame. However, three insertions, say X, Y, and Z in the fifth, eighth, and twelfth positions, respectively, would change the sentence to

THE BXI GYR EDZ FOX ATE THE EGG

which, after the third insertion, restores the original reading frame. The same would be true of three deletions. As before, if all three changes were close together, the sentence might still retain much of its meaning.

Crick and Brenner did not unambiguously demonstrate that the genetic code is a triplet code because they had no proof that their insertions and deletions involved only single nucleotides. Strictly speaking, they showed that a codon consists of $3r$ nucleotides where r is the number of nucleotides in an insertion or deletion. Although it was generally assumed at the time that $r = 1$, proof of this assertion had to await the elucidation of the genetic code (Section 32-1C).

C. Deciphering the Genetic Code

The genetic code could, in principle, be determined by simply comparing the base sequence of an mRNA with the amino acid sequence of the polypeptide it specifies. In the 1960s, however, techniques for isolating and sequencing

mRNAs had not yet been developed. The elucidation of the genetic code dictionary therefore proved to be a difficult task.

a. UUU Specifies Phe

The major breakthrough in deciphering the genetic code came in 1961 when Marshall Nirenberg and Heinrich Matthaei established that UUU is the codon specifying Phe. They did so by demonstrating that the addition of poly(U) to a cell-free protein synthesizing system stimulates only the synthesis of poly(Phe). The cell-free protein synthesizing system was prepared by gently breaking open *E. coli* cells by grinding them with powdered alumina and centrifuging the resulting cell sap to remove the cell walls and membranes. This extract contained DNA, mRNA, ribosomes, enzymes, and other cell constituents necessary for protein synthesis. When fortified with ATP, GTP, and amino acids, the system synthesized small amounts of proteins. This was demonstrated by the incubation of the system with ^{14}C -labeled amino acids followed by the precipitation of its proteins by the addition of trichloroacetic acid. The precipitate proved to be radioactive.

A cell-free protein synthesizing system, of course, produces proteins specified by the cell's DNA. On addition of DNase, however, protein synthesis stops within a few minutes because the system can no longer synthesize mRNA, whereas the mRNA originally present is rapidly degraded. Nirenberg found that crude mRNA-containing fractions from other organisms were highly active in stimulating protein synthesis in a DNase-treated protein synthesizing system. This system is likewise responsive to synthetic mRNAs.

The synthetic mRNAs that Nirenberg used in subsequent experiments were synthesized by the *Azotobacter vinelandii* enzyme **polynucleotide phosphorylase**. This enzyme, which was discovered by Severo Ochoa and Marianne Grunberg-Manago, links together nucleotides in the reaction



In contrast to RNA polymerase, however, polynucleotide phosphorylase does not utilize a template. Rather, it randomly links together the available NDPs so that the base composition of the product RNA reflects that of the reactant NDP mixture.

Nirenberg and Matthaei demonstrated that poly(U) stimulates the synthesis of poly(Phe) by incubating poly(U) and a mixture of 1 radioactive and 19 unlabeled amino acids in a DNase-treated protein synthesizing system. Significant radioactivity appeared in the protein precipitate only when phenylalanine was labeled. *UUU must therefore be the codon specifying Phe*. In similar experiments using poly(A) and poly(C), it was found that poly(Lys) and poly(Pro), respectively, were synthesized. Thus *AAA specifies Lys and CCC specifies Pro*. [Poly(G) cannot function as a synthetic mRNA because, even under denaturing conditions, it aggregates to form a four-stranded helix (Section 30-4De). An mRNA must be single stranded to direct its translation; Section 32-2D.]

Table 32-1 Amino Acid Incorporation Stimulated by a Random Copolymer of U and G in Mole Ratio 0.76 : 0.24

Codon	Probability of Occurrence	Relative Incidence ^a	Amino Acid	Relative Amount of Amino Acid Incorporated
UUU	0.44	100	Phe	100
UUG	0.14	32	Leu	36
UGU	0.14	32	Cys	35
GUU	0.14	32	Val	37
UGG	0.04	9	Trp	14
GUG	0.04	9		
GGU	0.04	9	Gly	12
GGG	0.01	2		

^aRelative incidence is defined here as $100 \times$ probability of occurrence/0.44.

Source: Matthaei, J.H., Jones, O.W., Martin, R.G., and Nirenberg, M., *Proc. Natl. Acad. Sci.* **48**, 666 (1962).

Nirenberg and Ochoa independently employed ribonucleotide copolymers to further elucidate the genetic code. For example, in a poly(UG) composed of 76% U and 24% G, the probability of a given triplet being UUU is $0.76 \times 0.76 \times 0.76 = 0.44$. Likewise, the probability of a particular triplet consisting of 2U's and 1G, that is, UUG, UGU, or GUU, is $0.76 \times 0.76 \times 0.24 = 0.14$. The use of this poly(UG) as an mRNA therefore indicated the base compositions, but not the sequences, of the codons specifying several amino acids (Table 32-1). Through the use of copolymers containing two, three, and four bases, the base compositions of codons specifying each of the 20 amino acids were inferred. Moreover, *these experiments demonstrated that the genetic code is degenerate since, for example, poly(UA), poly(UC), and poly(UG) all direct the incorporation of Leu into a polypeptide.*

b. The Genetic Code Was Elucidated through Triplet Binding Assays and the Use of Polyribonucleotides with Known Sequences

In the absence of GTP, which is necessary for protein synthesis, trinucleotides but not dinucleotides are almost as effective as mRNAs in promoting the ribosomal binding of specific tRNAs. This phenomenon, which Nirenberg and Philip Leder discovered in 1964, permitted the various codons to be identified by a simple binding assay. Ribosomes, together with their bound tRNAs, are retained by a nitrocellulose filter but free tRNA is not. The bound tRNA was identified by using charged tRNA mixtures in which only one of the pendent amino acid residues was radioactively labeled. For instance, it was found, as expected, that UUU stimulates the ribosomal binding of only Phe tRNA. Likewise, UUG, UGU, and GUU stimulate the binding of Leu, Cys, and Val tRNAs, respectively. Hence

UUG, UGU, and GUU must be codons that specify Leu, Cys, and Val, respectively. In this way, the amino acids specified by some 50 codons were identified. For the remaining codons, the binding assay was either negative (no tRNA bound) or ambiguous.

The genetic code dictionary was completed and previous results confirmed through H. Gobind Khorana's synthesis of polyribonucleotides with specified repeating sequences (Section 7-6A). In a cell-free protein synthesizing system, UCUCUCUCU... for example, is read

UCU CUC UCU CUC UCU C...

so that it specifies a polypeptide chain of two alternating amino acid residues. In fact, it was observed that this mRNA stimulated the production of

Ser—Leu—Ser—Leu—Ser—Leu—...

which indicates that either UCU or CUC specifies Ser and the other specifies Leu. This information, together with the tRNA-binding data, permitted the conclusion that UCU codes for Ser and CUC codes for Leu. These data also prove that codons consist of an odd number of nucleotides, thereby relieving any residual suspicions that codons consist of six rather than three nucleotides.

Alternating sequences of three nucleotides, such as poly(UAC), specify three different homopolypeptides because ribosomes may initiate polypeptide synthesis on these synthetic mRNAs in any of the three possible reading frames (Fig. 32-5). Analyses of the polypeptides specified by various alternating sequences of two and three nucleotides confirmed the identity of many codons and filled out missing portions of the genetic code.

c. mRNAs Are Read in the 5' → 3' Direction

The use of repeating tetranucleotides indicated the reading direction of the code and identified the chain

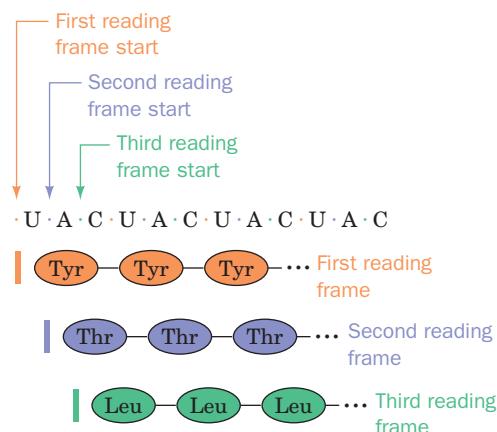


Figure 32-5 The three potential reading frames of an mRNA. Each reading frame would yield a different polypeptide.

termination codons. Poly(UAUC) specifies, as expected, a polypeptide with a tetrapeptide repeat:

5' UAU CUA UCU AUC UAU CUA ... 3'
Tyr — Leu — Ser — Ile — Tyr — Leu — ...

The amino acid sequence of this polypeptide indicates that the mRNA's 5' end corresponds to the polypeptide's N-terminus; that is, *mRNA is read in the 5' → 3' direction*.

d. UAG, UAA, and UGA Are Stop Codons

In contrast to the above results, poly(AUAG) yields only dipeptides and tripeptides. This is because *UAG is a signal to the ribosome to terminate protein synthesis*:

AUA GAU AGA UAG AUA GAU ...
Ile — Asp — Arg Stop Ile — Asp — ...

Likewise, poly(GUAA) yields dipeptides and tripeptides because UAA is also a chain termination signal:

GUA AGU AAG UAA GUA AGU ...
Val — Ser — Lys Stop Val — Ser — ...

UGA is a third stop signal. These **Stop codons**, whose existence was first inferred from genetic experiments, are known, somewhat inappropriately, as **nonsense codons** because they are the only codons that do not specify amino acids. UAG, UAA, and UGA are sometimes referred to as **amber**, **ochre**, and **opal** codons. [They were so named as the result of a laboratory joke: The German word for amber is Bernstein, the name of an individual who helped discover *amber* mutations (mutations that change some other codon to UAG); *ochre* and *opal* are puns on *amber*.]

e. AUG and GUG Are Chain Initiation Codons

The codons AUG, and less frequently GUG, form part of the chain initiation sequence (Section 32-3Ca). However, they also specify the amino acid residues Met and Val, respectively, at internal positions of polypeptide chains. (Nirenberg and Matthaei's discovery that UUU specifies Phe was only possible because ribosomes indiscriminately initiate polypeptide synthesis on an mRNA when the Mg²⁺ concentration is unphysiologically high as it was, serendipitously, in their experiments.)

D. The Nature of the Code

The genetic code dictionary, as elucidated by the above methods, is presented in Table 32-2 as well as in Table 5-3. Examination of the table indicates that the genetic code has several remarkable features:

1. The code is highly degenerate. Three amino acids, Arg, Leu, and Ser, are each specified by six codons, and

Table 32-2 The “Standard” Genetic Code^a

First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	UUU Phe	UCU	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG	UAG Stop	UGG Trp	G
C	CUU	CCU	CAU His	CGU	U
	CUC	CCC	CAC	CGC	C
	CUA Leu	CCA	CAA Gln	CGA Arg	A
	CUG	CCG	CAG	CGG	G
A	AUU	ACU	AAU Asn	AGU Ser	U
	AUC Ile	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met ^b	ACG	AAG	AGG	G
G	GUU	GCU	GAU Asp	GGU	U
	GUC	GCC	GAC	GGC	C
	GUU Val	GCA	GAA Glu	GGA Gly	A
	GUG	GCG	GAG	GGG	G

^aNonpolar amino acid residues are tan, basic residues are blue, acidic residues are red, and nonpolar uncharged residues are purple.

^bAUG forms part of the initiation signal as well as coding for internal Met residues.

most of the rest are specified by either four, three, or two codons. Only Met and Trp, two of the least common amino acids in proteins (Table 4-1), are represented by a single codon. Codons that specify the same amino acid are termed **synonyms**.

2. The arrangement of the code table is nonrandom. Most synonyms occupy the same box in Table 32-2; that is, they differ only in their third nucleotide. The only exceptions are Arg, Leu, and Ser, which, having six codons each, must occupy more than one box. XYU and XYG always specify the same amino acid; XYA and XYG do so in all but two cases. Moreover, changes in the first codon position tend to specify similar (if not the same) amino acids, whereas codons with second position pyrimidines encode mostly hydrophobic amino acids (tan in Table 32-2), and those with second position purines encode mostly polar amino acids (blue, red, and purple in Table 32-2). Apparently *the code evolved so as to minimize the deleterious effects of mutations*.

Many of the mutations causing amino acid substitutions in a protein can be rationalized, according to the genetic code, as single point mutations. *As a consequence of the genetic code's degeneracy, however, many point mutations at a third codon position are phenotypically silent; that is, the*

mutated codon specifies the same amino acid as the wild type. Degeneracy may account for as much as 33% of the 25 to 75% range in the G + C content among the DNAs of different organisms (Section 5-1Ba). The frequent occurrence of Arg, Ala, Gly, and Pro also tends to give a high G + C content, whereas Asn, Ile, Lys, Met, Phe, and Tyr contribute to a low G + C content.

a. Some Phage DNA Segments Contain Overlapping Genes in Different Reading Frames

Since any nucleotide sequence may have three reading frames, it is possible, at least in principle, for a polynucleotide to encode two or even three different polypeptides. This idea was never seriously entertained, however, because it seemed that the constraints on even two overlapping genes in different reading frames would be too great for them to evolve so that both could specify sensible proteins. It therefore came as a great surprise, in 1976, when Frederick Sanger reported that the 5386-nucleotide DNA of bacteriophage ϕ X174 (which, at the time, was the largest DNA to have been sequenced) contains two genes that are completely contained within larger genes of different reading frames (Fig. 32-6). Moreover, the end of the overlapping D and E genes contains the control sequence for the ribosomal initiation of the J gene so that this short DNA segment performs triple duty. Bacteria

also exhibit such coding economy; the ribosomal initiation sequence of one gene in a polycistronic mRNA often overlaps the end of the preceding gene. Nevertheless, completely overlapping genes have only been found in small single-stranded DNA phages, which presumably must make maximal use of the little DNA that they can pack inside their capsids.

b. The “Standard” Genetic Code Is Widespread but Not Universal

For many years it was thought that the “standard” genetic code (that given in Table 32-2) is universal. This assumption was, in part, based on the observations that one kind of organism (e.g., *E. coli*) can accurately translate the genes from quite different organisms (e.g., humans). This phenomenon is, in fact, the basis of genetic engineering. Once the “standard” genetic code had been established, presumably during the time of prebiotic evolution (Section 1-5B), any mutation that would alter the way the code is translated would result in numerous, often deleterious, protein sequence changes. Undoubtedly there is strong selection against such mutations.

Despite the foregoing, DNA sequencing studies in 1981 revealed that *the genetic codes of certain mitochondria (mitochondria contain their own genes and protein synthesizing systems but produce only a few mitochondrial proteins; Section 12-4E) are variants of the “standard” genetic code* (Table 32-3). For example, in mammalian mitochondria, UUA, as well as the standard AUG, is a Met/initiation codon; UGA specifies Trp rather than “Stop”; and AGA and AGG are “Stop” rather than Arg. Note that all mitochondrial genetic codes except those of plants simplify the “standard” code by increasing its degeneracy. For example, in the mammalian mitochondrial code, each amino acid is specified by at least two codons that differ only in their third nucleotide. Apparently the constraints preventing alterations of the genetic code are eased by the small sizes of mitochondrial genomes. More recent

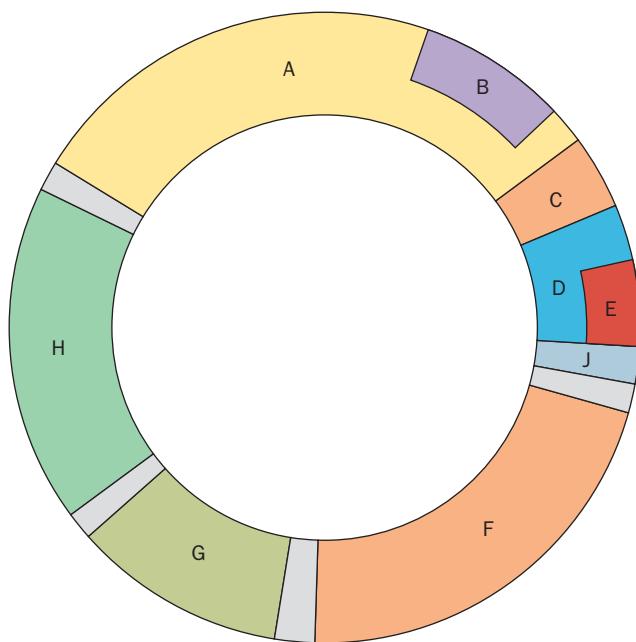


Figure 32-6 Genetic map of bacteriophage ϕ X174 as determined by DNA sequence analysis. Genes are labeled A, B, C, etc. Note that gene B is wholly contained within gene A and gene E is wholly contained within gene D. These pairs of genes are read in different reading frames and therefore specify unrelated proteins. The unlabeled regions correspond to untranslated control sequences.

Table 32-3 Mitochondrial Deviations from the “Standard” Genetic Code

Mitochondrion	UGA	AUA	CUN ^a	AG _G ^A	CGG
Mammalian	Trp	Met ^b			Stop
Baker's yeast	Trp	Met ^b	Thr		
<i>Neurospora crassa</i>	Trp				
<i>Drosophila</i>	Trp	Met ^b			Ser ^c
Protozoan	Trp				
Plant					Trp
“Standard” code	Stop	Ile	Leu	Arg	Arg

^aN represents any of the four nucleotides.

^bAlso acts as part of an initiation signal.

^cAGA only; no AGG codons occur in *Drosophila* mitochondrial DNA.

Source: Mainly Breitenberger, C.A. and RajBhandary, U.L., *Trends Biochem. Sci.* **10**, 481 (1985).

studies, however, have revealed that in ciliated protozoa, the codons UAA and UAG specify Gln rather than "Stop." Perhaps UAA and UAG were sufficiently rare codons in a primordial ciliate (which molecular phylogenetic studies indicate branched off very early in eukaryotic evolution) to permit the code change without unacceptable deleterious effects. At any rate, the "standard" genetic code, although very widely utilized, is not universal. Indeed, as we shall see in Section 32-2D, under the proper context in mRNA, certain codons can specify "nonstandard" amino acids.

2 TRANSFER RNA AND ITS AMINOACYLATION

The establishment of the genetic function of DNA led to the realization that cells somehow "translate" the language of base sequences into the language of polypeptides. Yet, nucleic acids originally appeared unable to bind specific amino acids [more recently RNA aptamers for specific amino acids have been generated; aptamers are nucleic acids that have been selected for their ability to bind specific ligands (Section 7-6C)]. In 1955, Crick, in what became known as the **adaptor hypothesis**, postulated that translation occurs through the mediation of "adaptor" molecules. Each adaptor was postulated to carry a specific enzymatically appended amino acid and to recognize the corresponding codon (Fig. 32-7). Crick suggested that these adaptors contain RNA because codon recognition could then occur by complementary base pairing. At about this time, Paul Zamecnik and Mahlon Hoagland discovered that in the course of protein synthesis, ^{14}C -labeled amino acids became transiently bound to a low molecular mass fraction of RNA. Further investigations indicated that these RNAs, which at first were called "soluble RNA" or "sRNA" but are now known as **transfer RNA (tRNA)**, are, in fact, Crick's putative adaptor molecules.

A. Primary and Secondary Structures of tRNA

 **See Guided Exploration 26: The structure of tRNA** In 1965, after a 7-year effort, Robert Holley reported the first known base sequence of a biologically significant nucleic acid, that of yeast alanine tRNA (tRNA^{Ala} ; Fig. 32-8). To do so Holley had to overcome several major obstacles:

1. All organisms contain many species of tRNAs (usually at least one for each of the 20 amino acids) which, because of their nearly identical properties (see below), are not easily separated. Preparative techniques had to be developed to provide the gram or so of pure yeast tRNA^{Ala} Holley required for its sequence determination.

2. Holley had to invent the methods that were initially used to sequence RNA (Section 7-2).

3. Ten of the 76 bases of yeast tRNA^{Ala} are modified (see below). Their structural formulas had to be elucidated

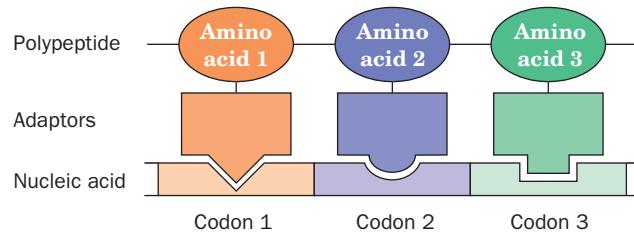


Figure 32-7 The adaptor hypothesis. It postulates that the genetic code is read by molecules that recognize a particular codon and carry the corresponding amino acid.

although they were never available in more than milligram quantities.

Since 1965, the techniques for tRNA purification and sequencing have vastly improved. A tRNA may now be sequenced in a few hours' time with only $\sim 1 \mu\text{g}$ of material. Presently, the base sequences of many thousands of tRNAs from nearly 800 organisms are known, most from genomic sequences (they are compiled at the Genomic

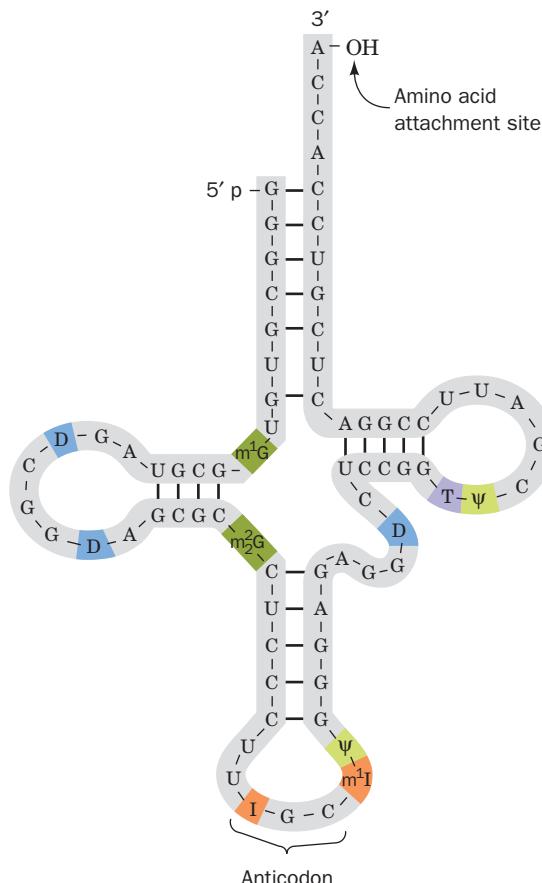


Figure 32-8 Base sequence of yeast tRNA^{Ala} drawn in the cloverleaf form. The symbols for the modified nucleosides (color) are explained in Fig. 32-10.

tRNA Database, <http://gtrnadb.ucsc.edu/>). They vary in length from 54 to 100 nucleotides (18–28 kD) although most have ~76 nucleotides.

Almost all known tRNAs, as Holley first recognized, may be schematically arranged in the so-called cloverleaf secondary structure (Fig. 32-9). Starting from the 5' end, they have the following common features:

1. A 5'-terminal phosphate group.
2. A 7-bp stem that includes the 5'-terminal nucleotide and that may contain non-Watson–Crick base pairs such as G · U. This assembly is known as the **acceptor** or **amino acid stem** because the amino acid residue carried by the tRNA is appended to its 3'-terminal OH group (Section 32-2C).
3. A 3- or 4-bp stem ending in a loop that frequently contains the modified base **dihydrouridine** (**D**; see below). This stem and loop are therefore collectively termed the **D arm**.
4. A 5-bp stem ending in a loop that contains the **anticodon**, the triplet of bases that is complementary to the codon specifying the tRNA. These features are known as the **anticodon arm**.
5. A 5-bp stem ending in a loop that usually contains the sequence T ψ C (where ψ is the symbol for **pseudouridine**; see below). This assembly is called the **T ψ C** or **T arm**.
6. All tRNAs terminate in the sequence CCA with a free 3'-OH group. The CCA may be genetically specified or enzymatically appended to immature tRNA (Section 31-4Cc).
7. There are 15 invariant positions (always have the same base) and 8 **semi-invariant** positions (only a purine or only a pyrimidine) that occur mostly in the loop regions. These regions also contain **correlated invariants**, that is, pairs of nonstem nucleotides that are base paired in all tRNAs. The purine on the 3' side of the anticodon is invariably modified. The structural significance of these features is examined below.

The site of greatest variability among the known tRNAs occurs in the so-called **variable arm**. It has from 3 to 21 nucleotides and may have a stem consisting of up to 7 bp. The D loop also varies in length from 5 to 7 nucleotides.

a. Transfer RNAs Have Numerous Modified Bases

One of the most striking characteristics of tRNAs is their large proportion, up to 25%, of post-translationally modified or hypermodified bases. Nearly 80 such bases, found at >60 different tRNA positions, have been characterized. A few of them, together with their standard abbreviations, are indicated in Fig. 32-10. Hypermodified nucleosides, such as i⁶A, are usually adjacent to the anticodon's 3' nucleotide when it is A or U. Their low polarities probably strengthen the otherwise relatively weak pairing associations of these bases with the codon, thereby increasing translational fidelity. Conversely, certain methylations block base pairing and hence prevent inappropriate structures from forming. Some of these modifications form important recognition elements for the enzyme that attaches the correct amino acid to a tRNA (Section 32-2Cb). However, none of them are essential for maintaining a tRNA's structural integrity (see below) or for its proper binding to the ribosome. Nevertheless, mutant bacteria unable to form certain modified bases compete poorly against the corresponding normal bacteria.

B. Tertiary Structure of tRNA

 **See Guided Exploration 26: The structure of tRNA** The earliest physicochemical investigations of tRNA indicated that it has a well-defined conformation. Yet, despite numerous hydrodynamic, spectroscopic, and chemical cross-linking studies, its three-dimensional structure remained an enigma until 1974. In that year, the 2.5-Å resolution X-ray crystal structure of yeast tRNA^{Phe} was separately elucidated by Alexander Rich in collaboration with Sung-Hou Kim and, in a different crystal form, by Aaron Klug. *The molecule assumes an L-shaped conformation in which one leg of the L is formed by the acceptor and T stems folded into a continuous A-RNA-like double helix (Section 29-1Ba) and the other leg is similarly composed of the D and*

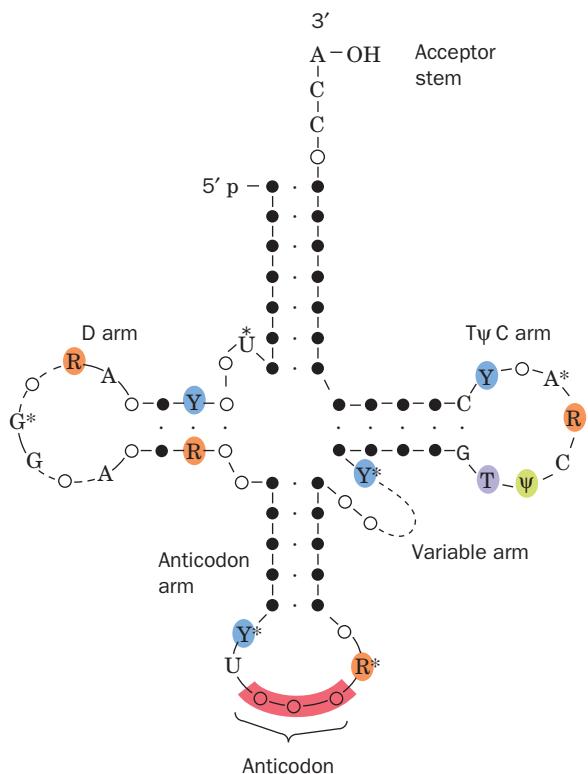
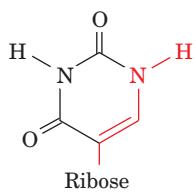
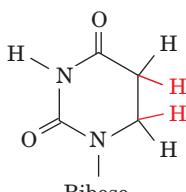
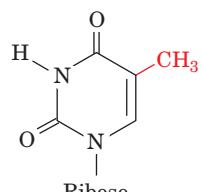


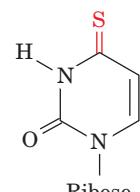
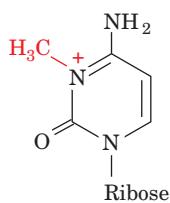
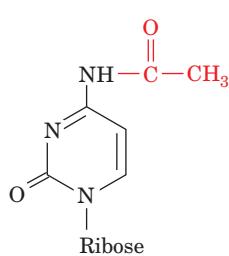
Figure 32-9 Cloverleaf secondary structure of tRNA. Filled circles connected by dots represent Watson–Crick base pairs, and open circles in the double-helical regions indicate bases involved in non-Watson–Crick base pairing. Invariant positions are indicated: R and Y represent invariant purines and pyrimidines and ψ signifies pseudouridine. The starred nucleosides are often modified. The dashed regions in the D and variable arms contain different numbers of nucleotides in the various tRNAs.

Uracil derivativesPseudouridine (ψ)

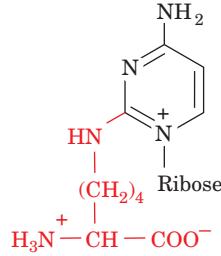
Dihydrouridine (D)



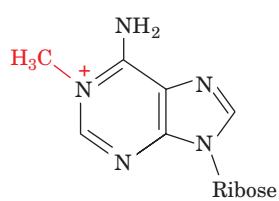
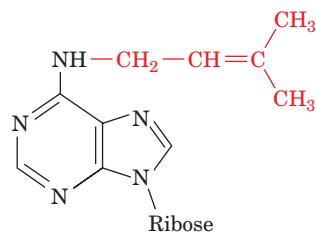
Ribothymidine (T)

4-Thiouridine (s^4U)**Cytosine derivatives**3-Methylcytidine (m^3C)

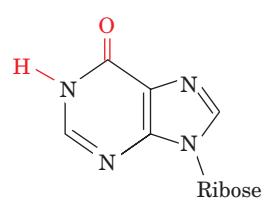
N⁴-Acetylcystidine (ac⁴C)



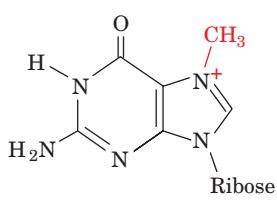
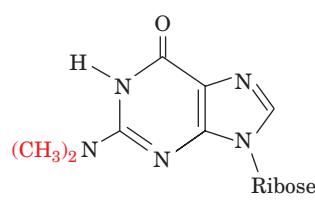
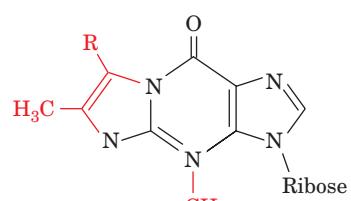
Lysidine (L)

Adenine derivatives1-Methyladenosine (m^1A)

N⁶-Isopentenyladenosine (i⁶A)



Inosine (I)

Guanine derivativesN⁷-Methylguanosine (m^7G)N²,N²-Dimethylguanosine (m^2G)

Wyosine (Wyo)

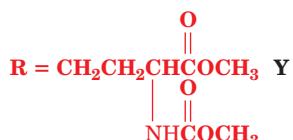


Figure 32-10 A selection of the modified nucleosides that occur in tRNAs together with their standard abbreviations. Note that although inosine chemically resembles guanosine, it is

biochemically derived from adenosine. Nucleosides may also be methylated at their ribose 2' positions to form residues symbolized, for instance, by Cm, Gm, and Um.

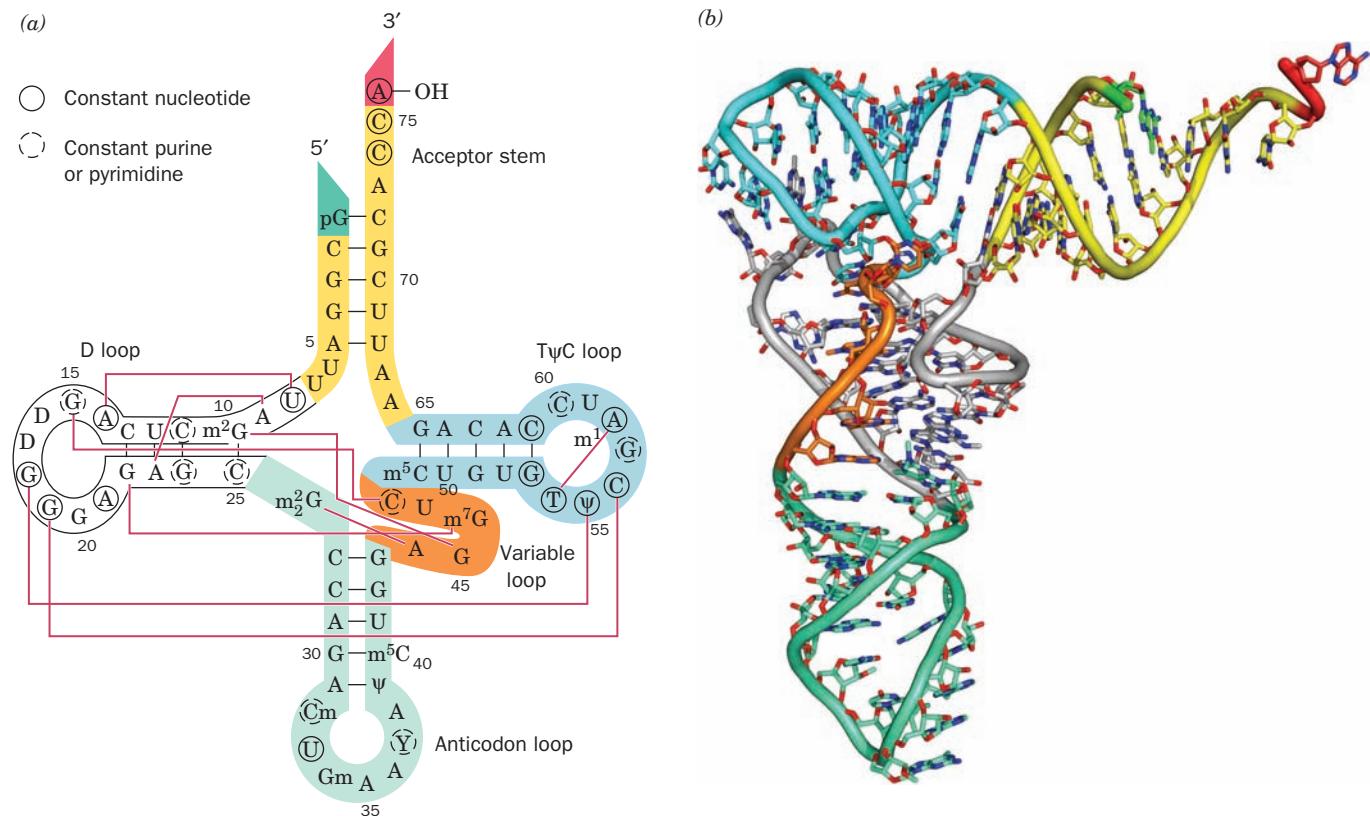


Figure 32-11 Structure of yeast tRNA^{Phe}. (a) The base sequence drawn in cloverleaf form. Tertiary base pairing interactions are represented by thin red lines connecting the participating bases. Bases that are invariant or semi-invariant in all tRNAs are circled by solid and dashed lines, respectively. The 5' terminus is colored bright green, the acceptor stem is yellow, the D arm is white, the anticodon arm is light green, and the variable

arm is orange, the T_ψC arm is cyan, and the 3' terminus is red. (b) The X-ray structure drawn to show how its base paired stems form an L-shaped molecule. The tRNA is drawn in stick form with C atoms colored as in Part a, N blue, and O red. Adjacent P atoms are connected by rods colored as in Part a. [Based on an X-ray structure by Sung-Hou Kim, PDBid 6TRNA.] 

anticodon stems (Fig. 32-11). Each leg of the L is ~ 60 Å long and the anticodon and amino acid acceptor sites are at opposite ends of the molecule, some 76 Å apart. The narrow 20- to 25-Å width of native tRNA is essential to its biological function: During protein synthesis, three RNA molecules must simultaneously bind in close proximity at adjacent codons on mRNA (Section 32-3Ae).

a. tRNA's Complex Tertiary Structure Is Maintained by Hydrogen Bonding and Stacking Interactions

The structural complexity of yeast tRNA^{Phe} is reminiscent of that of a protein. Although only 42 of its 76 bases occur in double helical stems, 71 of them participate in stacking associations (Fig. 32-12). The structure also contains 9 base pairing interactions that cross-link its tertiary structure (Figs. 32-11a and 32-12). Remarkably, all but one of these tertiary interactions, which appear to be the mainstays of the molecular structure, are non-Watson-Crick associations. Moreover, most of the bases involved in these interactions are either invariant or semi-invariant, which strongly suggests that all tRNAs have similar conformations (see below). The structure is also stabilized by several

unusual hydrogen bonds between bases and either phosphate groups or the 2'-OH groups of ribose residues.

The compact structure of yeast tRNA^{Phe} results from its large number of intramolecular associations, which renders most of its bases inaccessible to solvent. The most notable exceptions to this are the anticodon bases and those of the amino acid-bearing—CCA terminus. Both of these groupings must be accessible in order to carry out their biological functions.

The observation that the molecular structures of yeast tRNA^{Phe} in two different crystal forms are essentially identical lends much credence to the supposition that its crystal structure closely resembles its solution structure. Transfer RNAs other than yeast tRNA^{Phe} have, unfortunately, been notoriously difficult to crystallize. As yet, the X-ray structures of only three other uncomplexed tRNAs have been reported (although the X-ray structures of numerous tRNAs in complex with the enzymes that append their corresponding amino acids and with ribosomes have been elucidated; Sections 32-2C and 32-3D). The major structural differences among them result from an apparent flexibility in the anticodon loop and the—CCA terminus as well as from a

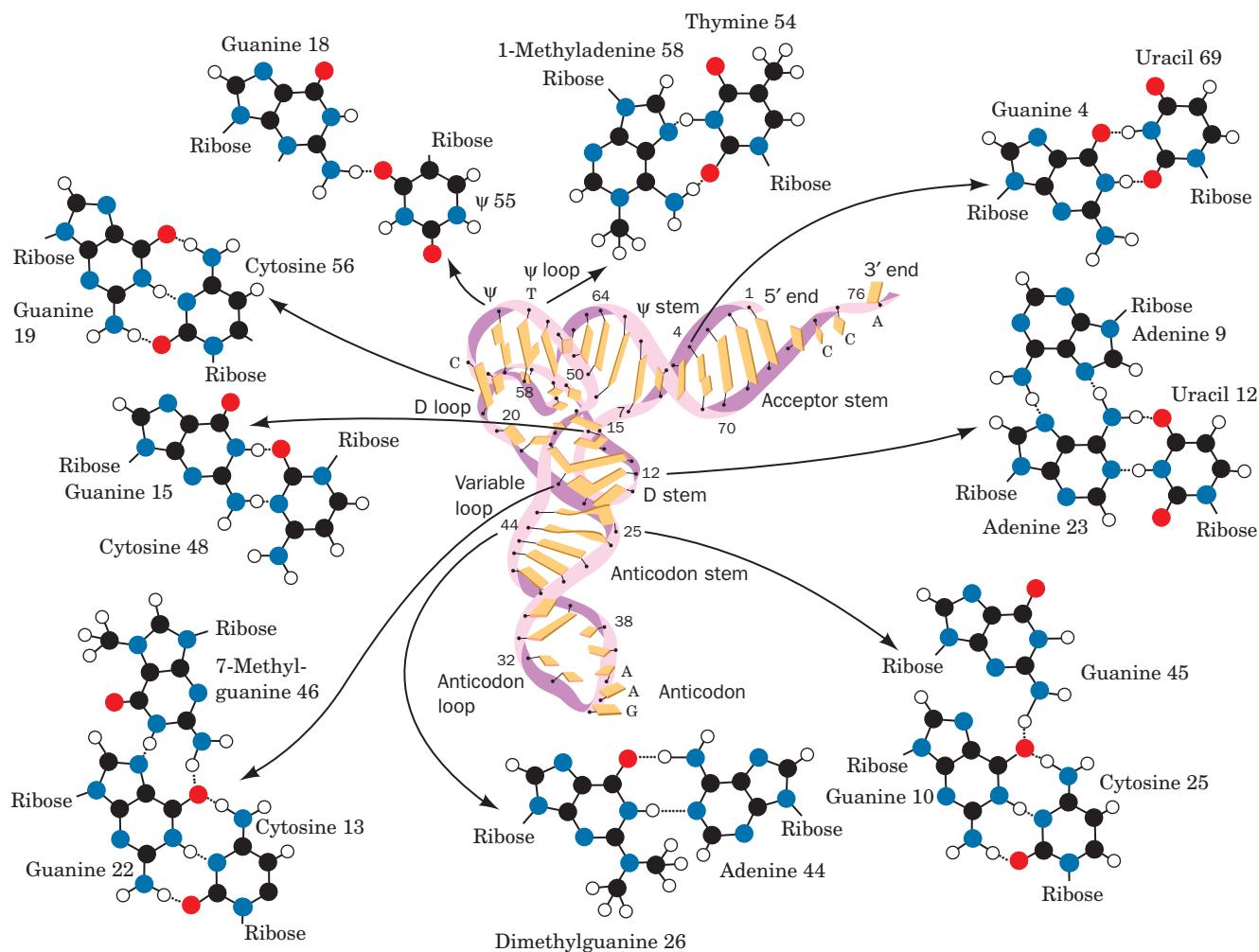


Figure 32-12 Tertiary base pairing interactions in yeast

tRNA^{Phe}. Note that all but one of these nine interactions involve non-Watson-Crick pairs and that they are all located near the corner of the L. [After Kim, S.H., in Schimmel, P.R., Söll, D., and

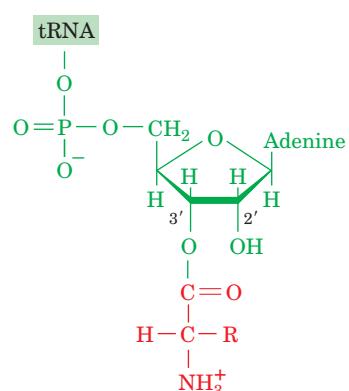
Abelson, J.N. (Eds.), *Transfer RNA: Structure, Properties and Recognition*, p. 87, Cold Spring Harbor Laboratory Press (1979). Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]

See Kinemage Exercise 19-3

hingelike mobility between the two legs of the L that gives, for instance, yeast tRNA^{Asp} a boomerang-like shape. Such observations are in accord with the expectation that all tRNAs fit into the same ribosomal cavities.

C. Aminoacyl-tRNA Synthetases

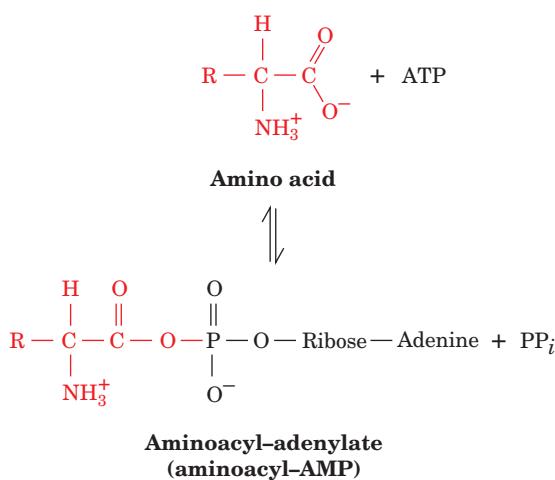
See Guided Exploration 27: The structures of aminoacyl-tRNA synthetases and their interactions with tRNAs Accurate translation requires two equally important recognition steps: (1) the choice of the correct amino acid for covalent attachment to a tRNA; and (2) the selection of the amino acid-charged tRNA specified by mRNA. The first of these steps, which is catalyzed by amino acid-specific enzymes known as **aminoacyl-tRNA synthetases (aaRSs)**, appends an amino acid to the 3'-terminal ribose residue of its cognate tRNA to form an aminoacyl-tRNA (Fig. 32-13). This otherwise unfavorable process is driven by the hydrolysis of ATP in two sequential reactions that are catalyzed by a single enzyme.



Aminoacyl-tRNA

Figure 32-13 An aminoacyl-tRNA. The amino acid residue is esterified to the tRNA's 3'-terminal nucleoside at either its 3'-OH group, as shown here, or its 2'-OH group.

1. The amino acid is first “activated” by its reaction with ATP to form an **aminoacyl-adenylate**



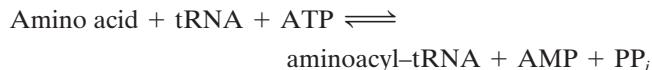
which, with all but three aaRSs, can occur in the absence of tRNA. Indeed, this intermediate may be isolated although it normally remains tightly bound to the enzyme.

2. This mixed anhydride then reacts with tRNA to form the aminoacyl-tRNA



Some aaRSs exclusively append an amino acid to the terminal 2'-OH group of their cognate tRNAs, and others do so at the 3'-OH group. This selectivity was established with the use of chemically modified tRNAs that lack either the 2'- or 3'-OH group of their 3'-terminal ribose residue. The use of these derivatives was necessary because, in solution, the aminoacyl group rapidly equilibrates between the 2' and 3' positions.

The overall aminoacylation reaction is



These reaction steps are readily reversible because the free energies of hydrolysis of the bonds formed in both the aminoacyl-adenylate and the aminoacyl-tRNA are comparable to that of ATP hydrolysis. The overall reaction is driven to completion by the inorganic pyrophosphatase-catalyzed hydrolysis of the PP_i generated in the first reaction step. Amino acid activation therefore chemically resembles fatty acid activation (Section 25-2A); the major difference between these two processes, which were both elucidated by Paul Berg, is that tRNA is the acyl acceptor in amino acid activation, whereas CoA performs this function in fatty acid activation.

a. There Are Two Classes of Aminoacyl-tRNA Synthetases

Most cells have one aaRS for each of the 20 amino acids. The similarity of the reactions catalyzed by these enzymes and the structural resemblance of all tRNAs suggests that

all aaRSs evolved from a common ancestor and should therefore be structurally related. This is not the case. In fact, the aaRSs form a diverse group of enzymes. The over 1000 such enzymes that have been characterized each have one of four different types of subunit structures, α , α_2 (the predominant forms), α_4 , and $\alpha_2\beta_2$, with known subunit sizes ranging from ~ 300 to ~ 1200 residues. Moreover, there is little sequence similarity among synthetases specific for different amino acids. Quite possibly, aminoacyl-tRNA synthetases arose very early in evolution, before the development of the modern protein synthesis apparatus other than tRNAs.

Detailed sequence and structural comparisons of aminoacyl-tRNA synthetases by Dino Moras indicate that these enzymes form two unrelated families, termed **Class I** and **Class II aaRSs**, that each have the same 10 members in nearly all organisms (Table 32-4). The Class I enzymes, although of largely dissimilar sequences, share two homologous polypeptide segments, not present in other proteins, that have the consensus sequences His-Ile-Gly-His (HIGH) and Lys-Met-Ser-Lys-Ser (KMSKS). The X-ray structures of Class I enzymes indicate that both of these segments are components of a dinucleotide-binding fold (Rossmann fold, which is also possessed by many NAD^+ - and ATP-binding proteins; Section 8-3Bi) in which they

Table 32-4 Characteristics of Bacterial Aminoacyl-tRNA Synthetases

Amino Acid	Quaternary Structure	Number of Residues
Class I		
Arg	α	577
Cys	α	461
Gln	α	553
Glu	α	471
Ile	α	939
Leu	α	860
Met	α, α_2	676
Trp	α_2	325
Tyr	α_2	424
Val	α	951
Class II		
Ala	α, α_4	875
Asn	α_2	467
Asp	α_2	590
Gly	$\alpha_2\beta_2$	303/689
His	α_2	424
Lys	α_2	505
Pro	α_2	572
Phe	$\alpha_2\beta_2, \alpha$	327/795
Ser	α_2	430
Thr	α_2	642

Source: Mainly Carter, C.W., Jr., *Annu. Rev. Biochem.* **62**, 715 (1993).

participate in ATP binding and are implicated in catalysis. The Class II synthetases lack the foregoing sequences but have three other sequences in common. Their X-ray structures reveal that these sequences occur in a so-called signature motif, a fold found only in Class II enzymes that consists of a 7-stranded antiparallel β sheet with three flanking helices, which forms the core of their catalytic domains.

Many Class I aaRSs require anticodon recognition to aminoacylate their cognate tRNAs. In contrast, several Class II enzymes, including **AlaRS** and **SerRS**, do not interact with their bound tRNA's anticodon. Indeed, several class II aaRSs accurately aminoacylate "microhelices" derived from only the acceptor stems of their cognate tRNAs. Another difference between Class I and Class II synthetases is that all Class I enzymes aminoacylate their bound tRNA's 3'-terminal 2'-OH group, whereas Class II enzymes, with the exception of **PheRS**, all charge the 3'-OH group. The amino acids for which the Class I synthetases are specific tend to be larger and more hydrophobic than those used by Class II synthetases. Finally, as Table 32-4 indicates, Class I aaRSs are mainly monomers, whereas most Class II aaRSs are homodimers.

LysRS has been classified as a Class II aaRS. However, a search of the genome sequences of *Methanococcus janaschii* and *Methanobacterium thermoautotrophicum* failed to reveal the presence of such a LysRS. This led to the discovery that the LysRSs expressed by these archaeabacteria are Class I rather than Class II enzymes. This raises the interesting question of how Class I LysRS evolved.

Prokaryotic aaRSs occur as individual protein molecules. However, in many higher eukaryotes (e.g., *Drosophila* and mammals), 9 aaRSs, some of each class, associate to form a multienzyme particle in which the glutamyl and prolyl synthetase functions are fused into a single polypeptide named **GluProRS**. The advantages of these systems are unknown.

b. The Structural Features Recognized by Aminoacyl-tRNA Synthetases May Be Quite Simple

As we shall see in Section 32-2D, ribosomes select aminoacyl-tRNAs only via codon–anticodon interactions, not according to the identities of their aminoacyl groups. *Accurate translation therefore requires not only that each tRNA be aminoacylated by its cognate aaRS but that it not be aminoacylated by any of its 19 noncognate aaRSs*. Moreover, since most cells express only one aaRS for each amino acid, each aaRS must aminoacylate all of the several, if not many, **isoaccepting tRNAs** (different tRNAs that are specific for the same amino acid) in each cell.

Considerable effort has therefore been expended, notably by LaDonne Schulman, Paul Schimmel, Olke Uhlenbeck, and John Abelson, in elucidating how aaRSs manage this feat, despite the close structural similarities of nearly all tRNAs. The experimental methods employed involved the use of specific tRNA fragments, mutationally altered tRNAs, chemical cross-linking agents, computerized sequence comparisons, and X-ray crystallography. The most common synthetase contact sites on tRNA occur on the inner (concave) face of the L. Other than that, there appears to be little regularity in how the various tRNAs are recognized by their cognate synthetases. Indeed, as we shall see, some aaRSs recognize only their cognate tRNA's acceptor stem, whereas others also interact with its anticodon region. Additional tRNA regions may also be recognized.

Genetic manipulations by Schimmel revealed that the tRNA features recognized by at least one type of aaRS are surprisingly simple. Numerous sequence alterations of *E. coli* tRNA^{Ala} do not appreciably affect its capacity to be aminoacylated with alanine. Yet, most base substitutions in the G3 · U70 base pair located in the tRNA's acceptor stem (Fig. 32-14a) greatly diminish this reaction. Moreover, the introduction of a G · U base pair into the analogous

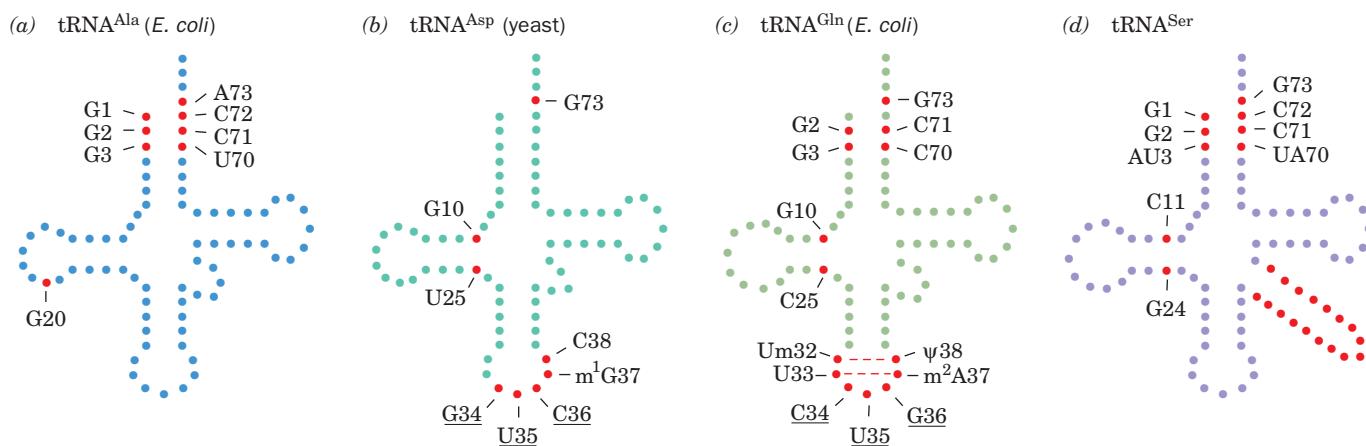


Figure 32-14 Major identity elements in four tRNAs. Each base in the tRNA is represented by a filled circle. Red circles indicate positions that have been shown to be identity elements for the recognition of the tRNA by its cognate aminoacyl-tRNA synthetase. The anticodon bases that are identity elements are

underlined. In each case, additional identity elements may yet be discovered. The base at position 73, which is an identity element in all four tRNAs shown here, is known as the **discriminator base**.

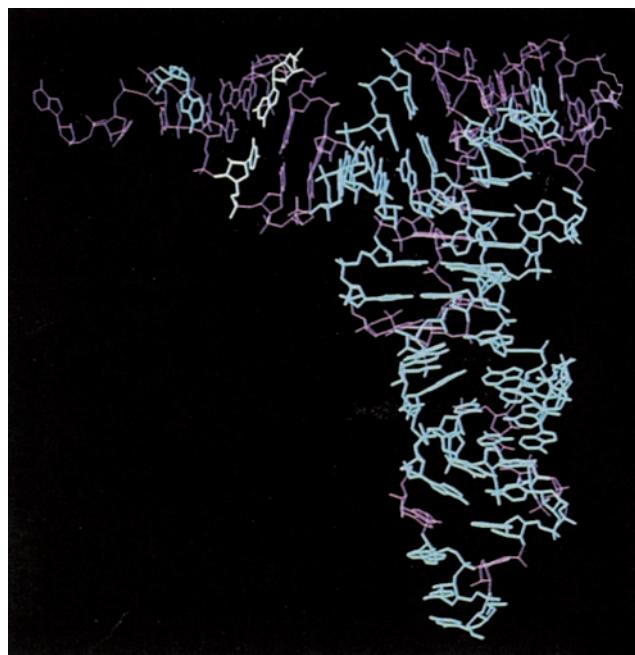


Figure 32-15 Three-dimensional model of *E. coli* tRNA^{Ala}. This model is based on the X-ray structure of yeast tRNA^{Phe} (Fig. 32-11b) in which the nucleotides that are different in *E. coli* tRNA^{Cys} are highlighted in cyan and the G3 · U70 base pair is highlighted in ivory. [Courtesy of Ya-Ming Hou, MIT.]

position of tRNA^{Cys} and tRNA^{Phe} causes them to be aminoacylated with alanine even though there are few other sequence identities between these mutant tRNAs and tRNA^{Ala} (e.g., Fig. 32-15). In fact, *E. coli* AlaRS even efficiently aminoacylates a 24-nt “microhelix” derived from only the G3 · U70-containing acceptor stem of *E. coli* tRNA^{Ala}. Since the only *E. coli* tRNAs that normally have a G3 · U70 base pair are the tRNA^{Ala}, and this base pair is also present in the tRNA^{Ala} from many organisms including yeast (Fig. 32-8), the foregoing observations strongly suggest that *the G3 · U70 base pair is a major feature recognized by AlaRSs*. These enzymes presumably recognize the distorted shape of the G · U base pair (Fig. 32-12), a hypothesis corroborated by the observation that base changes at G3 · U70 which least affect the acceptor identity of tRNA^{Ala} yield base pairs that structurally resemble G · U.

The elements of three other tRNAs, which are recognized by their cognate tRNA synthetases, are indicated in Fig. 32-14. As with tRNA^{Ala}, these identity elements appear to comprise only a few bases. Note that the anticodon forms an identity element in two of these tRNAs. In another example of an anticodon identifier, the *E. coli* tRNA^{Ile} specific for the codon AUA has the anticodon LAU, where L is **lysidine**, a modified cytosine whose 2-keto group is replaced by the amino acid lysine (Fig. 32-10). The L in this context pairs with A rather than G, a rare instance of base modification altering base pairing specificity. The replacement of this L with unmodified C, as expected, yields a tRNA that recognizes the Met codon AUG (codons bind anticodons in an antiparallel fashion). Surprisingly, how-

ever, this altered tRNA^{Ile} is also a much better substrate for **MetRS** than it is for **IleRS**. Thus, both the codon and the amino acid specificity of this tRNA are changed by a single post-transcriptional modification. The *N*¹-methylation of G37 in yeast tRNA^{Asp} (Fig. 32-14b) provides another example of a base modification forming an identity element. In the absence of this *N*¹-methyl group, tRNA^{Asp} is recognized by **ArgRS**, largely via its C36 and G37, whereas ArgRS normally recognizes only tRNA^{Arg}, mainly via its C35 and U36.

The available experimental evidence has largely located the various tRNA identifiers in the acceptor stem and the anticodon loop (Fig. 32-16). The X-ray structures of several aaRS · tRNA complexes, which we consider next, have structurally rationalized some of these observations.

c. The X-Ray Structure of GlnRS · tRNA^{Gln}, a Class I Complex

The X-ray structures of all 20 different amino acid-specific aaRSs from numerous organisms have been determined, many of which are in complex with ATP, their cognate amino acids, or their analogs. These structures reveal that the active sites of these enzymes bind the ATP and target amino acid in optimal positions for in-line nucleophilic displacement (Section 16-2B) during amino acid

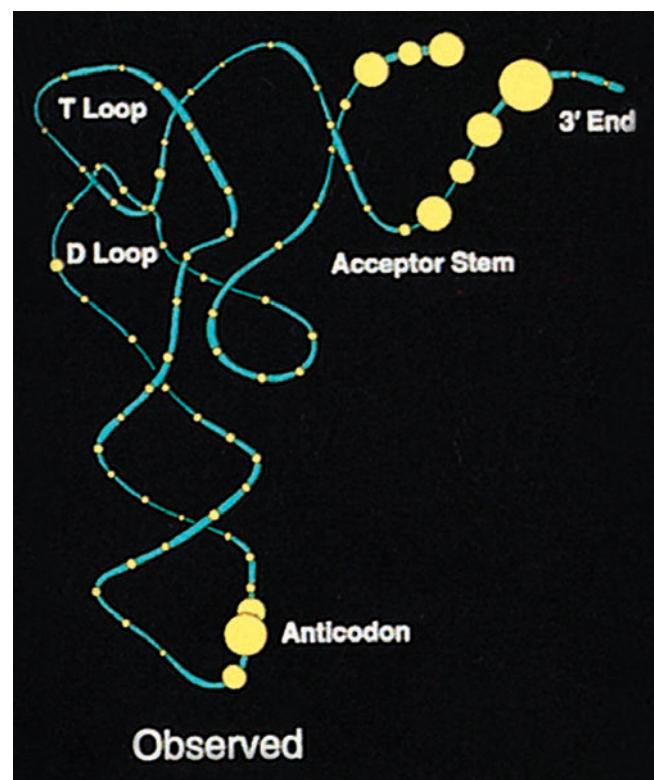


Figure 32-16 Experimentally observed identity elements of tRNAs. The tRNA backbone is cyan and each of its nucleotides is represented by a yellow circle whose diameter is proportional to the fraction of the 20 tRNA acceptor types for which the nucleotide is an observed determinant. [Courtesy of William McClain, University of Wisconsin.]

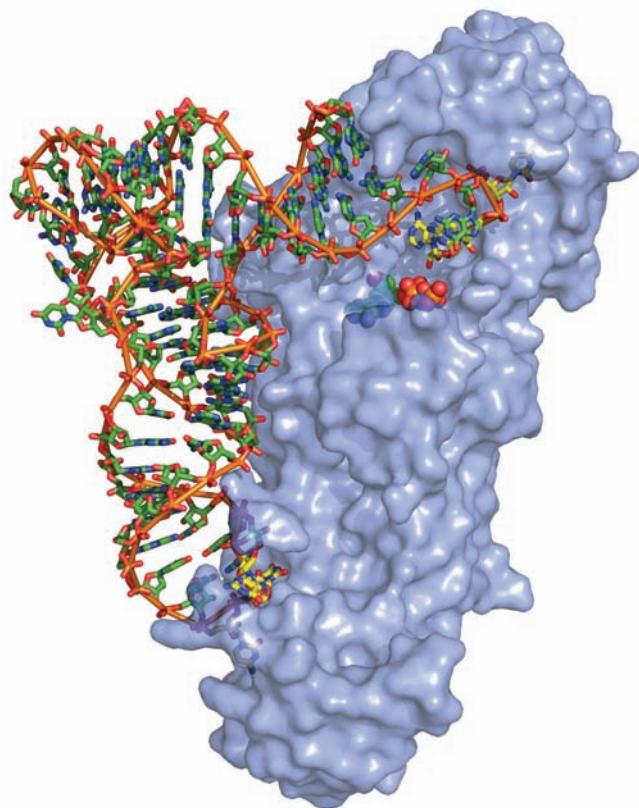
activation and that the specificity of an aaRS for its target amino acid is determined by idiosyncratic contacts with the side chain of the amino acid.

The X-ray structures of 16 different aaRSs in their complexes with their cognate tRNAs (all but those of Gly, Ala, Lys, and His) have so far been reported. The first of them to be elucidated, that of *E. coli* **GlnRS**, a Class I synthetase, in its complex with **tRNA^{Gln}** and ATP (Fig. 32-17), was determined by Thomas Steitz. The tRNA^{Gln} assumes an L-shaped conformation that resembles those of tRNAs of known structures (e.g., Fig. 32-11b). GlnRS, a 553-residue monomeric protein that consists of four domains arranged to form an elongated molecule, interacts with the tRNA along the entire inside face of the L such that the anticodon is bound near one end of the protein and the acceptor stem is bound near its other end.

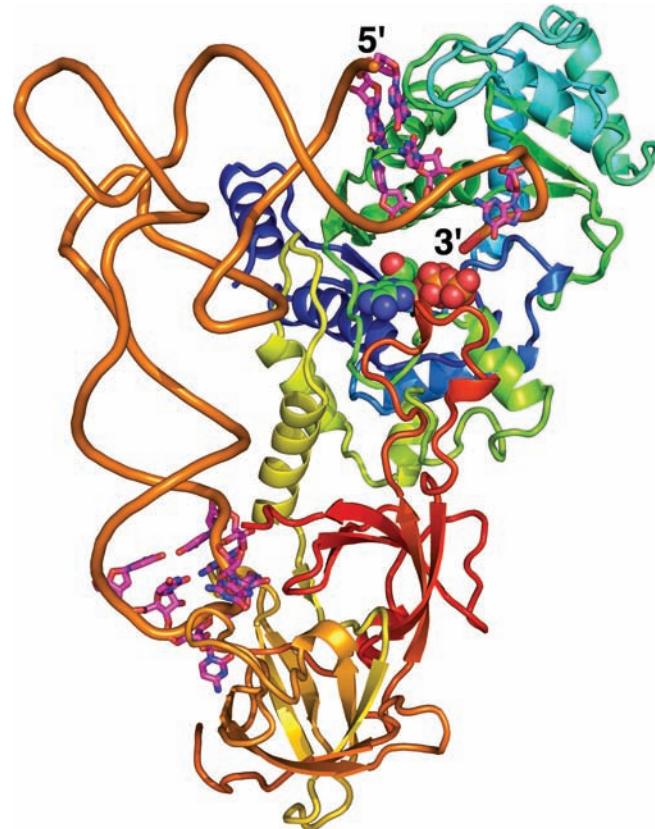
Genetic and biochemical data indicate that the identity elements of tRNA^{Gln} are largely clustered in its anticodon loop and acceptor stem (Fig. 32-14c). The anticodon loop of tRNA^{Gln} is extended by two novel non-Watson–Crick base

pairs (2'-*O*-methyl-U32 · ψ38 and U33 · m²A37), thereby causing the bases of the anticodon to unstack and splay outward in different directions so as to bind in separate recognition pockets of GlnRS. These structural features suggest that GlnRS uses all seven bases of the anticodon loop to discriminate among tRNAs. Indeed, changes to any one of the bases of residues C34 through ψ38 yield tRNAs with decreases in k_{cat}/K_M for aminoacylation by GlnRS by factors ranging from 70 to 28,000.

The GCCA at the 3' end of the tRNA^{Gln} makes a hairpin turn toward the inside of the L rather than continuing helically onward (as does the ACCA at the 3' end in the X-ray structure of tRNA^{Phe}; Fig. 32-11b). This conformation change is facilitated by the insinuation of a Leu side chain between the 5' and 3' ends of the tRNA so as to disrupt the first base pair of the acceptor stem (U1 · A72). The GlnRS reaction is therefore relatively insensitive to base changes in these latter two positions except when base pairing is strengthened by their conversion to G1 · C72. The GCCA end of the tRNA^{Gln} plunges deeply into a protein pocket



(a)



(b)

Figure 32-17 X-ray structure of *E. coli* GlnRS · tRNA^{Gln} · ATP. (a) The tRNA is drawn in stick form colored as the ATP but with the C atoms of the anticodon (UCG) and the 3'-CCA end yellow. An orange rod links its successive P atoms. The ATP bound in the protein's active site is drawn in space-filling form with C green, N blue, O red, and P orange. The protein is represented by a semitransparent light blue surface diagram that reveals the buried portions of the tRNA and ATP. Note that both the 3' end of the tRNA (top right) and its anticodon bases

(bottom) are inserted into deep pockets in the protein. (b) The complex viewed as in Part a. The tRNA's sugar-phosphate backbone is represented by an orange worm and the bases forming its identity elements (Fig. 32-14c) are drawn in stick form colored according to atom type with C magenta, N blue, and O red. The ATP is drawn as in Part a. The protein is shown in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). [Based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1GTR.]  See Kinemage Exercise 20

that also binds the enzyme's ATP and glutamine substrates. Three protein "fingers" are inserted into the minor groove of the acceptor stem to make sequence-specific interactions with base pairs G2 · C71 and G3 · C70 [recall that double helical RNA has an A-DNA-like structure (Section 29-1Bc) whose wide minor groove readily admits protein but whose major groove is normally too narrow to do so].

The GlnRS domain that binds glutamine, ATP, and the GCCA end of tRNA^{Gln}, the so-called catalytic domain, contains, as we previously discussed, a dinucleotide-binding fold. Much of this domain is nearly superimposable with and thus evolutionarily related to the corresponding domains of other Class I aaRSs.

d. The X-Ray Structure of AspRS · tRNA^{Asp}, a Class II Complex

Yeast AspRS, a Class II synthetase, is an α_2 dimer of 557-residue subunits. Its X-ray structure in complex with tRNA^{Asp}, determined by Moras, reveals that the protein symmetrically binds two tRNA molecules (Fig. 32-18). Like GlnRS, AspRS principally contacts its bound tRNA both at the end of its acceptor stem and in its anticodon region. The contacts in these two enzymes are, nevertheless, quite different in character (Fig. 32-19): Although both tRNAs approach their cognate synthetases along the inside of their L shapes, tRNA^{Gln} does so toward the direction of the minor groove of its acceptor stem, whereas tRNA^{Asp} does so toward the direction of its major groove. The GCCA at the 3' end of tRNA^{Asp} thereby continues its helical track as it plunges into AspRS's catalytic site, whereas, as we saw, the GCCA end of tRNA^{Gln} bends backward into a hairpin

turn that opens up the first base pair (U1 · A72) of its acceptor stem. Although the deep major groove of an A-RNA helix is normally too narrow to admit groups larger than water molecules (Section 29-1Bc), the major groove at the end of the acceptor stem in AspRS · tRNA^{Asp} is sufficiently widened for its base pairs to interact with a protein loop.

The anticodon arm of tRNA^{Asp} is bent by as much as 20 Å toward the inside of the L relative to that in the X-ray structure of uncomplexed tRNA^{Asp} and its anticodon bases are unstacked. The hinge point for this bend is a G30 · U40 base pair in the anticodon stem which, in nearly all other species of tRNA, is a Watson–Crick base pair. The anticodon bases of tRNA^{Gln} are also unstacked in contacting GlnRS but with a backbone conformation that differs from that in tRNA^{Asp}. Evidently, the conformation of a tRNA in complex with its cognate synthetase appears to be dictated more by its interactions with the protein (induced fit) than by its sequence.

Structural analyses of complexes of AspRS · tRNA^{Asp} with ATP and aspartic acid, and of GlnRS · tRNA^{Gln} with ATP, have permitted models of the aminoacyl-AMP complexes of these enzymes to be independently formulated. Comparison of these models reveals that the 3'-terminal A residues of tRNA^{Gln} and tRNA^{Asp} (to which the aminoacyl groups are appended; Fig. 32-13) are positioned on opposite sides of the enzyme-bound aminoacyl-AMP intermediate (Fig. 32-20). The 3'-terminal ribose residues are puckered C2'-endo for tRNA^{Asp} and C3'-endo for tRNA^{Gln} (see Fig. 29-8) such that the 2'-hydroxyl group of tRNA^{Gln} (Class I) is stereochemically positioned to attack the

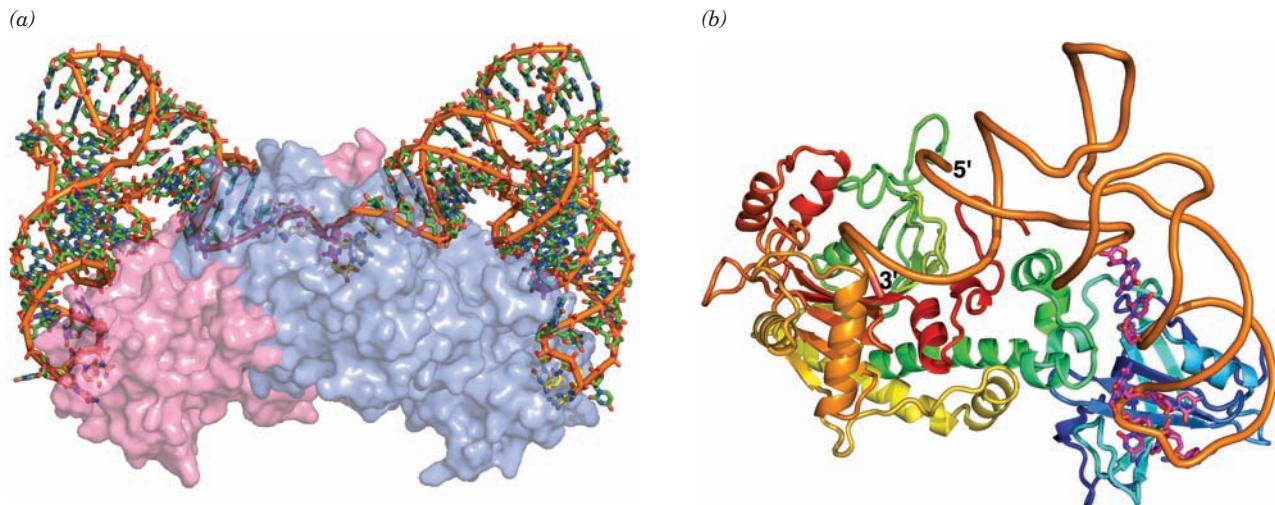


Figure 32-18 X-ray structure of yeast AspRS · tRNA^{Asp} · ATP. (a) The homodimeric enzyme with its two symmetrically bound tRNAs is viewed with its 2-fold axis approximately vertical. The tRNAs are drawn in skeletal form colored according to atom type with the C atoms of the anticodon (GUC) and the 3'-CCA end yellow, the remaining C atoms green, N blue, O red, and P orange. An orange rod connects its successive P atoms. The two protein subunits are represented by semitransparent pink and light blue surface diagrams that reveal the buried portions of

the tRNAs. (b) A ribbon diagram of the AspRS · tRNA^{Asp} protomer. The tRNA's sugar–phosphate backbone is represented by an orange worm and the bases forming its identity elements (Fig. 32-14b) are drawn in stick form colored according to atom type with C magenta, N blue, and O red. The protein is shown in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). [Based on an X-ray structure by Dino Moras, CNRS/INSERM/ULP, Illkirch Cédex, France. PDBID 1ASY.]

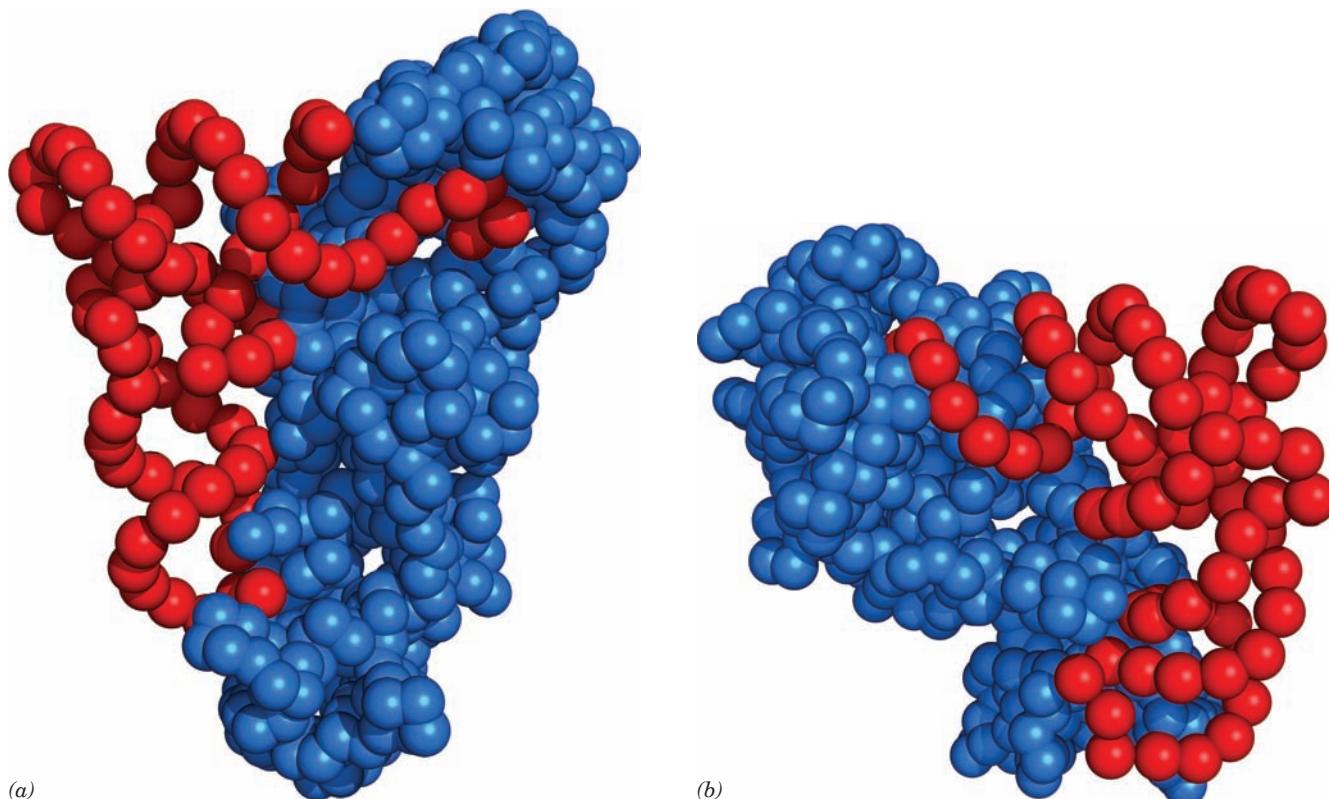


Figure 32-19 Comparison of the modes by which GlnRS and AspRS bind their cognate tRNAs. The proteins and tRNAs are represented by blue and red spheres centered on their C_{α} and P atom positions. Note how GlnRS (a), a Class I synthetase, binds tRNA^{Gln} from the minor groove side of its acceptor stem so as to bend its 3' end into a hairpin conformation. In contrast, AspRS

(b), a Class II synthetase, binds tRNA^{Asp} from the major groove side of its acceptor stem so that its 3' end continues its helical path on entering the active site. [After drawings by Dino Moras, CNRS/INSERM/ULP, Illkirch Cédex, France. PDBBids 1GTR abd 1ASY.]

aminoacyl-AMP's carboxyl group, whereas for tRNA^{Asp} (Class II), only the 3'-hydroxyl group is situated to do so. This clearly explains the different aminoacylation specificities of the Class I and Class II aaRSs.

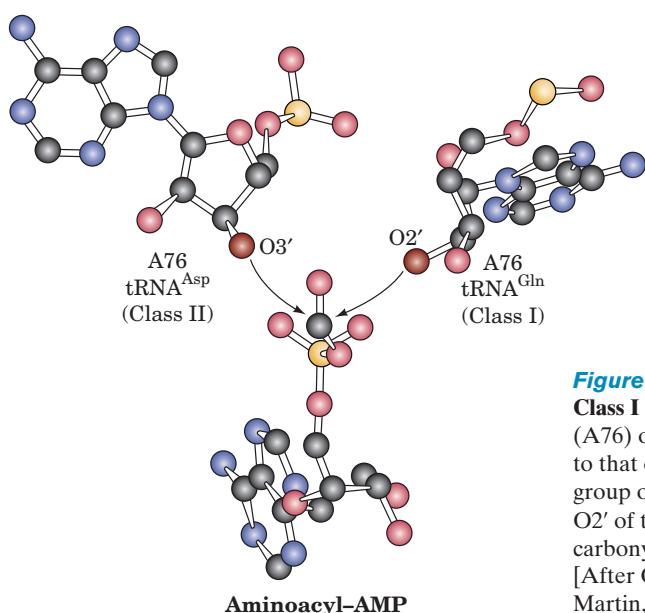


Figure 32-20 Comparison of the stereochemistries of aminoacylation by Class I and Class II aaRSs. The positions of the 3' terminal adenosine residues (A76) of AspRS (Class II, left) and GlnRS (Class I, right) are drawn relative to that of the enzyme-bound aminoacyl-AMP (below; only the carbonyl group of its aminoacyl residue is shown). Note how only O3' of tRNA^{Asp} and O2' of tRNA^{Gln} are suitably positioned to attack the aminoacyl residue's carbonyl group and thereby transfer the aminoacyl residue to the tRNA. [After Cavarelli, J., Eriani, G., Rees, B., Ruff, M., Boeglin, M., Mitschler, A., Martin, F., Gangloff, J., Thierry, J.-C., and Moras, D., *EMBO J.* **13**, 335 (1994).]

e. Proofreading Enhances the Fidelity of Amino Acid Attachment to tRNA

The charging of a tRNA with its cognate amino acid is a remarkably accurate process: aaRSs display an overall error rate of about 1 in 10,000. We have seen that aaRSs bind only their cognate tRNAs through an intricate series of specific contacts. But how do they discriminate among the various amino acids, some of which are quite similar?

Experimental measurements indicate, for example, that IleRS transfers as many as 40,000 isoleucines to tRNA^{Ile} for every valine it so transfers. Yet, as Linus Pauling first pointed out, *there are insufficient structural differences between Val and Ile to permit such a high degree of discrimination in the direct generation of aminoacyl-tRNAs*. The X-ray structure of *Thermus thermophilus* IleRS, a monomeric

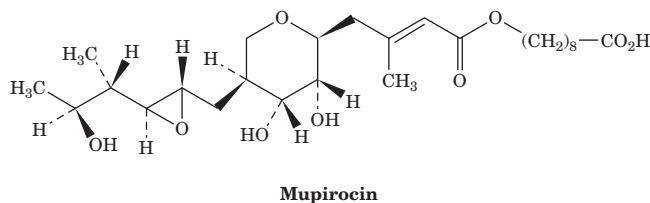
Class I aaRS, in complex with isoleucine, determined by Shigeyuki Yokoyama and Schimmel, indicates that isoleucine fits snugly into its binding site in the enzyme's Rossmann fold domain and hence that this binding site would sterically exclude leucine as well as larger amino acids. However, valine, which differs from isoleucine by only the lack of a single methylene group, fits into this isoleucine-binding site. The binding free energy of a methylene group is estimated to be $\sim 12 \text{ kJ} \cdot \text{mol}^{-1}$. Equation [3.17] indicates that the ratio f of the equilibrium constants, K_1 and K_2 , with which two substances bind to a given binding site is given by

$$f = \frac{K_1}{K_2} = \frac{e^{-\Delta G_1^{\circ\prime}/RT}}{e^{-\Delta G_2^{\circ\prime}/RT}} = e^{-\Delta \Delta G^{\circ\prime}/RT} \quad [32.1]$$

where $\Delta \Delta G^{\circ\prime} = \Delta G_1^{\circ\prime} - \Delta G_2^{\circ\prime}$ is the difference between the free energies of binding of the two substances. It is therefore estimated that isoleucyl-tRNA synthetase could discriminate between isoleucine and valine by no more than a factor of ~ 100 .

Berg resolved this apparent paradox by demonstrating that, in the presence of tRNA^{Ile}, IleRS catalyzes the nearly quantitative hydrolysis of valyl-adenylate to valine + AMP rather than forming Val-tRNA^{Ile}. Moreover, the few Val-tRNA^{Ile} molecules that do form are hydrolyzed to valine + tRNA^{Ile}. Thus, IleRS subjects both aminoacyl-adenylate and aminoacyl-tRNA^{Ile} to a **proofreading or editing step that occurs at a separate catalytic site**. This site binds Val residues but excludes the larger Ile residues. *The enzyme's overall selectivity is therefore the product of the selectivities of its synthesis and proofreading steps, thereby accounting for the high fidelity of aminoacylation. Note that in this so-called double-sieve mechanism, editing occurs at the expense of ATP hydrolysis, the thermodynamic price of high fidelity (increased order).*

The X-ray structure of *Staphylococcus aureus* IleRS in complex with tRNA^{Ile} and the clinically useful antibiotic **mupirocin**



(a product of *Pseudomonas fluorescens* that acts by specifically binding to bacterial IleRS so as to inhibit bacterial protein synthesis), determined by Steitz, suggests how IleRS carries out its editing process. The X-ray structure (Fig. 32-21) reveals that this complex resembles the GlnRS · tRNA^{Gln} · ATP complex (Fig. 32-17) but with IleRS having an additional editing domain (also called CP1 for connective peptide 1) inserted in its Rossmann fold domain. The two 3' terminal residues of the tRNA^{Ile}, C75 and A76, are disordered but, when modeled so as to continue the acceptor stem's stacked A-form helix, extend into a cleft in the editing domain that has been implicated as its hydrolytic site (Fig. 32-22a, left). Thus, this IleRS complex ap-

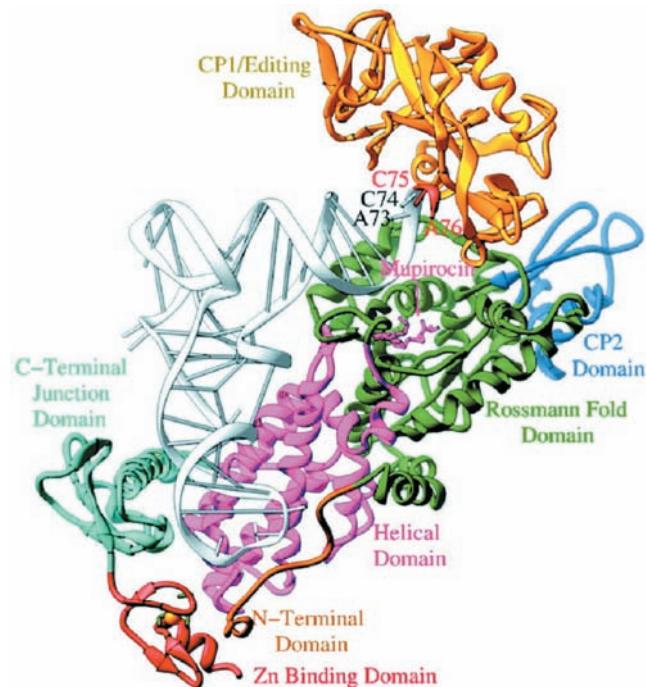
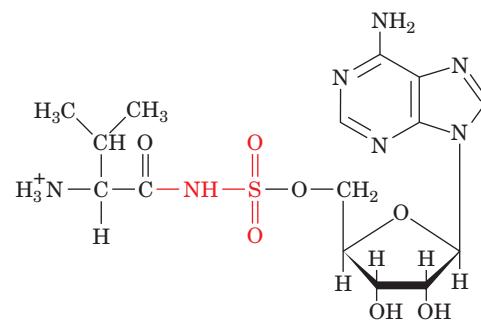


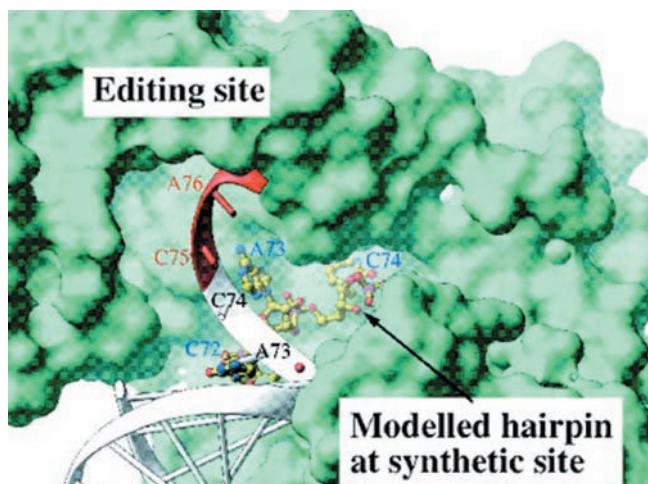
Figure 32-21 X-ray structure of *S. aureus* isoleucyl-tRNA synthetase in complex with tRNA^{Ile} and mupirocin. The tRNA is white, the protein is colored by domain, and the mupirocin is shown in stick form in pink. [Courtesy of Thomas Steitz, Yale University. PDBid 1QU2.]

pears to resemble an “editing complex” instead of a “transfer complex” as seen in the GlnRS structure. However, a transfer complex would form if the 3' ending segment of the tRNA^{Ile} assumes a hairpin conformation (Fig. 32-22a, right) similar to that in the GlnRS structure (Figs. 32-17b and 32-19a; recall that IleRS and GlnRS are both Class I aaRSs). Steitz has therefore postulated that the aminoacyl group is shuttled between the IleRS's aminoacylation site and its editing site by such a conformational change (Fig. 32-22b). This process functionally resembles the way in which DNA polymerase I edits its newly synthesized strand (Section 30-2Ag), which Steitz also elucidated.

ValRS is a monomeric Class I aaRS that resembles IleRS. The X-ray structure of the complex of *T. thermophilus* ValRS, tRNA^{Val}, and the nonhydrolyzable **valyl-adenylate** analog **5'-O-[N-(L-valyl)sulfamoyl]adenosine (Val-AMS)**,

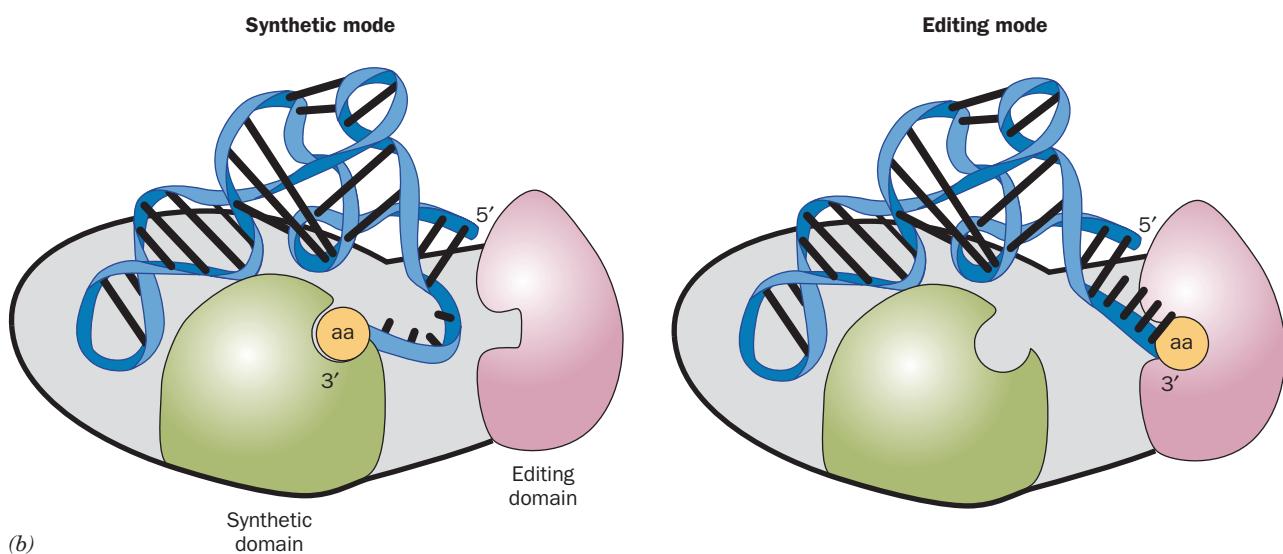


5'-O-[N-(L-Valyl)sulfamoyl]adenosine (Val-AMS)

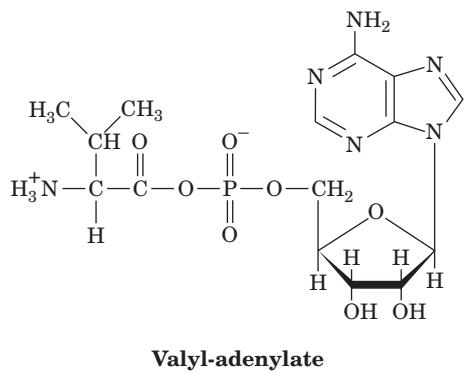


(a)

Figure 32-22 Comparison of the putative aminoacylation and editing modes of IleRS · tRNA^{Ile}. (a) The superposition of tRNA^{Ile} in these two binding modes on the solvent-accessible surface of IleRS (green). The acceptor strand of tRNA^{Ile} in the editing mode observed in the X-ray structure of IleRS · tRNA^{Ile} · mupirocin (Fig. 32-21) is drawn in ribbon form in white with the modeled positions of C75 and A76 in red. This places the tRNA's 3' end in the editing site. In contrast, the three 3' terminal residues of tRNA^{Ile}, as positioned through homology modeling based on the X-ray structure of GlnRS · tRNA^{Gln} · ATP (Fig. 32-17) and drawn in ball-and-stick form with C yellow, N blue, O red, and P magenta, places the tRNA's 3' end in the synthetic (aminoacylation) site, 34 Å distant from its position in the editing site. Note that there is a cleft running between the editing and synthetic sites and that the 3' end of the tRNA continues its A-form helical path in the editing mode but assumes a hairpin conformation in the synthetic mode. (b) A cartoon comparing the positions of the 3' end of tRNA^{Ile} in its complex with IleRS in its synthetic mode (left) and in its editing mode (right). [Part a courtesy of and Part b based on a drawing by Thomas Steitz, Yale University.]



(b)



determined by Yokoyama, reveals that the Val-AMS is bound in the aminoacylation pocket in the Rossmann fold domain, which accommodates the isosteric Val and Thr moieties but sterically excludes Ile. Modeling studies based on the IleRS · tRNA^{Ile} · mupirocin structure indicate that

the Thr side chain would fit into the ValRS editing pocket with its side chain hydroxyl group hydrogen bonded to the side chain of Asp 279 of ValRS, which protrudes into the pocket in contrast to the corresponding Asp 328 of IleRS, which does not. Consequently, a Val side chain would be excluded from the ValRS editing pocket because it cannot form such a hydrogen bond, thereby explaining why this editing pocket hydrolyzes **threonyl-adenylate** and Thr-tRNA^{Val} but not the corresponding Val derivatives. The ValRS · tRNA^{Val} structure also indicates that ValRS and tRNA^{Val} together form a tunnel connecting the ValRS's aminoacylation pocket with its editing pocket. Improperly formed threonyl-adenylate is proposed to be channeled through this tunnel for hydrolysis in the editing pocket, thereby explaining why tRNA^{Val} must be bound to ValRS for this pretransfer editing reaction to occur. Valyl-adenylate is presumably channeled through the similar IleRS · tRNA^{Ile} complex for its hydrolysis.

ThrRS, a Class II homodimer, has the opposite problem of ValRS: It must synthesize **Thr-tRNA^{Thr}** but not **Val-tRNA^{Thr}**. The X-ray structure of *E. coli* ThrRS that lacks its N-terminal domain but remains catalytically active in a complex with either threonine or the threonyl-adenylate analog **Thr-AMS**, determined by Moras, reveals that ThrRS's aminoacylation pocket contains a Zn^{2+} ion that is coordinated by the side chain hydroxyl and amino groups of the threonyl group as well as by three protein side chains. The isosteric valine could not coordinate the Zn^{2+} ion in this way and hence does not undergo adenylylation by ThrRS. However, what prevents ThrRS from synthesizing **Ser-tRNA^{Thr}**? In fact, the truncated ThrRS synthesizes **Ser-tRNA^{Thr}** at more than half the rate it synthesizes **Thr-tRNA^{Thr}**, thereby indicating that the N-terminal domain of wild-type ThrRS contains the enzyme's editing site. Mutational analysis of ThrRS has localized this editing site to a cleft in the N-terminal domain of wild-type ThrRS, whose X-ray structure in complex with tRNA^{Thr} was also determined by Moras. In this latter structure, the tRNA's 3' end follows a regular helical path similar to that seen in the X-ray structure of AspRS \cdot tRNA^{Asp} \cdot ATP (Fig. 32-18) so as to enter the aminoacylation site. However, if the 3' end of the bound tRNA^{Thr} assumed a hairpin conformation similar to that seen in X-ray structure of tRNA^{Gln} in complex with the Class I enzyme GlnRS and ATP (Fig. 32-17), its covalently linked aminoacyl group would enter the editing site. This indicates an intriguing "mirror symmetry" (Fig. 32-23): In Class I aaRSs that mediate a double-sieve editing mechanism, the 3' end of the bound cognate tRNA assumes a hairpin conformation when it enters the aminoacylation site and a helical conformation when it enters the editing site, whereas the converse holds for Class II aaRSs. Finally, ThrRS does not appear to mediate pretransfer editing (does not hydrolyze **seryl-adenylate**), and, in fact, the ThrRS \cdot tRNA^{Thr} complex lacks a channel connecting its aminoacylation and editing sites such as is seen in the ValRS \cdot tRNA^{Val} complex.

Synthetases that have adequate selectivity for their corresponding amino acid lack editing functions. Thus, for example, the TyrRS aminoacylation site discriminates between tyrosine and phenylalanine through hydrogen bonding with the tyrosine —OH group. The cell's other amino acids, standard as well as nonstandard, have even less resemblance to tyrosine, which rationalizes why TyrRS lacks an editing site.

f. **Gln-tRNA^{Gln}** May Be Formed via an Alternative Pathway

Although it was long believed that each of the 20 standard amino acids is covalently linked to a tRNA by its corresponding aaRS, it is now clear that gram-positive bacteria, archaeabacteria, cyanobacteria, mitochondria, and chloroplasts all lack GlnRS. Rather glutamate is linked to tRNA^{Gln} by the same GluRS that synthesizes **Glu-tRNA^{Glu}**. The resulting **Glu-tRNA^{Gln}** is then transamidated to Gln-tRNA^{Gln} by the enzyme **Glu-tRNA^{Gln} amidotransferase (Glu-AdT)** in an ATP-requiring reaction in which glutamine is the

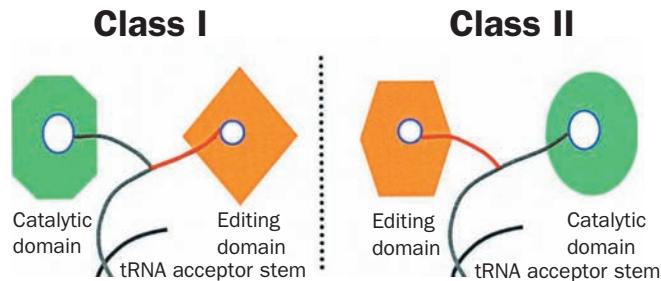


Figure 32-23 Schematic diagram of the aminoacylation and editing mechanisms of Class I and Class II aaRSs emphasizing the "mirror symmetry" of their overall mechanisms. With Class I aaRSs (left; e.g., IleRS), the 3' end of the bound tRNA's acceptor stem assumes a hairpin conformation in the synthetic mode and a helical conformation in the editing mode, whereas the converse occurs with Class II aaRSs (right; e.g., ThrRS). [Courtesy of Dino Moras, CNRS/INSERM/ULP, Illkirch Cedex, France.]

amide donor. Some microorganisms use a similar transamidation pathway for the synthesis of **Asn-tRNA^{Asn}** from **Asp-tRNA^{Asn}**.

The overall reaction catalyzed by Glu-AdT occurs in three stages (Fig. 32-24): (1) Glutamine is hydrolyzed to glutamate and the resulting NH₃ sequestered; (2) ATP reacts with the Glu side chain of Glu-tRNA^{Gln} to yield an activated acylphosphate intermediate and ADP; and (3) the acylphosphate intermediate reacts with the NH₃ to yield Gln-tRNA^{Gln} + P_i. Glu-AdT from *Bacillus subtilis*, which was characterized by Dieter Söll, is a heterotrimeric protein, none of whose subunits exhibit significant sequence similarity to GlnRS. The genes encoding these subunits, *gatA*, *gatB*, and *gatC*, form a single operon whose disruption is lethal, thereby demonstrating that *B. subtilis* has no alternative pathway for Gln-tRNA^{Gln} production. The **GatA** subunit of Glu-AdT appears to catalyze the activation of the side chain carboxyl of glutamic acid via a reaction resembling that catalyzed by carbamoyl phosphate synthetase (Section 26-2A). Nevertheless, GatA exhibits no sequence similarity with other known glutamine amidotransferases (members of the triad or Ntn families; Section 26-5Aa). The **GatB** subunit may be used to select the correct tRNA substrate. The role of the **GatC** subunit is unclear, although the observation that its presence is necessary for the expression of GatA in *E. coli* suggests that it participates in the modification, folding, and/or stabilization of GatA.

Since Glu is not misincorporated into *B. subtilis* proteins in place of Gln, the Glu-tRNA^{Gln} product of the above aminoacylation reaction must not be transported to the ribosome in the same way as Gln-tRNA^{Gln}. It is likely that this occurs because, as has been shown in chloroplasts, **EF-Tu**, the elongation factor that binds and transports most bacterial aminoacyl-tRNAs to the ribosome in a GTP-dependent process (Section 32-3D), does not bind Glu-tRNA^{Gln}. It is unclear why two independent routes have evolved for the synthesis of Gln-tRNA^{Gln}.

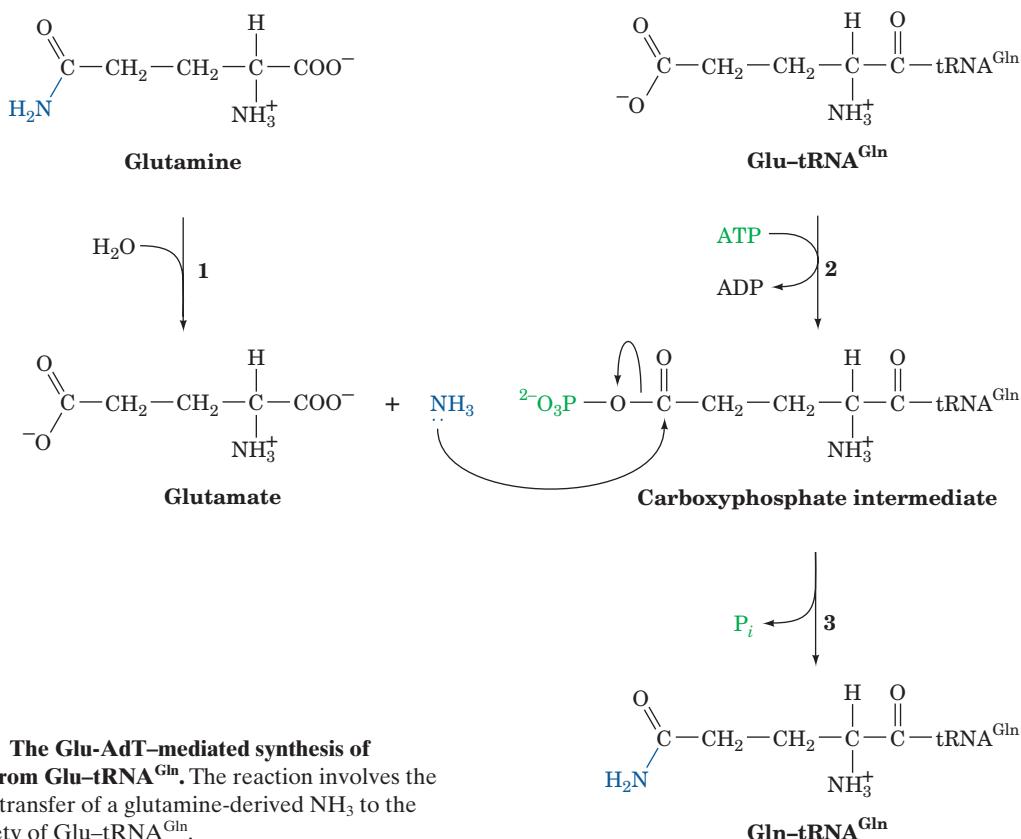


Figure 32-24 The Glu-AdT-mediated synthesis of Gln-tRNA^{Gln} from Glu-tRNA^{Gln}. The reaction involves the ATP-activated transfer of a glutamine-derived NH₃ to the glutamate moiety of Glu-tRNA^{Gln}.

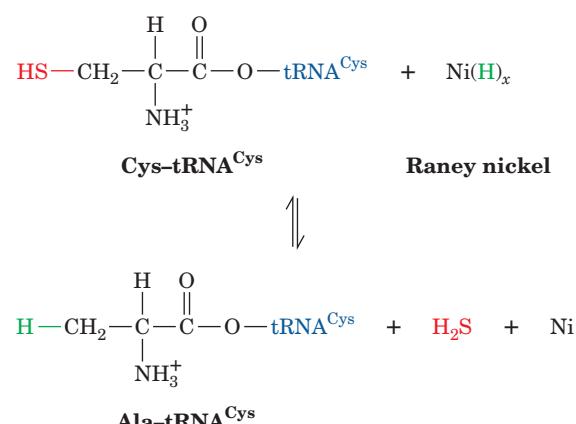
g. Some Archaeabacteria Lack a Separate CysRS

The genomes of certain archaeabacteria such as *M. janaschii* lack an identifiable gene for CysRS. This is because the enzyme responsible for synthesizing Pro-tRNA^{Pro} in these organisms also synthesizes Cys-tRNA^{Cys}. Interestingly, this enzyme, which is named ProCysRS, does not synthesize Pro-tRNA^{Cys} or Cys-tRNA^{Pro}. Although ProCysRS synthesizes **cysteinyl-adenylate** only in the presence of tRNA^{Cys}, it synthesizes **prolyl-adenylate** in the absence of tRNA^{Pro}. The binding of tRNA^{Cys} to ProCysRS blocks the activation of proline so that only cysteine can be activated. Conversely, the activation of proline facilitates the binding of tRNA^{Pro} while preventing the binding of tRNA^{Cys}. However, the mechanism through which ProCysRS carries out these mutually exclusive syntheses is unknown. In any case, it appears that some organisms can get by with as few as 17 different aaRSs; they may lack GlnRS, AspRS, and a separate CysRS.

D. Codon–Anticodon Interactions

In protein synthesis, the proper tRNA is selected only through codon–anticodon interactions; the aminoacyl group does not participate in this process. This phenomenon was demonstrated as follows. Cys-tRNA^{Cys}, in which the Cys

residue was ¹⁴C labeled, was reductively desulfurized with Raney nickel so as to convert the Cys residue to Ala:

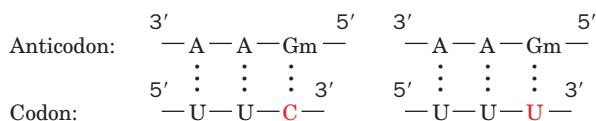


The resulting ¹⁴C-labeled hybrid, Ala-tRNA^{Cys}, was added to a cell-free protein synthesizing system extracted from rabbit reticulocytes. The product hemoglobin α chain's only radioactive tryptic peptide was the one that normally contains the subunit's only Cys. No radioactivity was found in the peptides that normally contain Ala but no Cys.

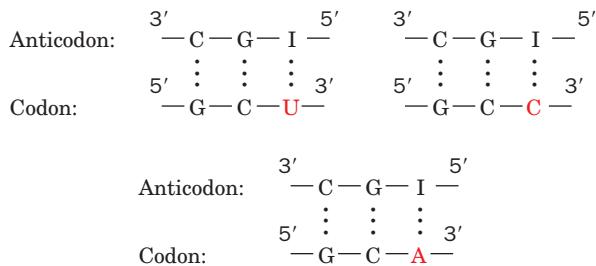
Evidently, only the anticodons of aminoacyl-tRNAs participate in codon recognition.

a. Genetic Code Degeneracy Is Largely Due to Variable Third Position Codon–Anticodon Interactions

One might naively guess that each of the 61 codons specifying an amino acid would be read by a different tRNA. Yet, even though most cells contain several groups of isoaccepting tRNAs, many tRNAs bind to two or three of the codons specifying their cognate amino acids. For example, yeast tRNA^{Phe}, which has the anticodon GmAA, recognizes the codons UUC and UUU (remember that the anticodon pairs with the codon in an antiparallel fashion),



and yeast tRNA^{Ala}, which has the anticodon IGC, recognizes the codons GCU, GCC, and GCA.



It therefore seems that non-Watson–Crick base pairing can occur at the third codon–anticodon position (the anticodon's first position is defined as its 3' nucleotide), the site of most codon degeneracy (Table 32-2). Note also that the third (5') anticodon position commonly contains a modified base such as Gm or I.

b. The Wobble Hypothesis Structurally Accounts for Codon Degeneracy

By combining structural insight with logical deduction, Crick proposed, in what he named the **wobble hypothesis**, how a tRNA can recognize several degenerate codons. He assumed that the first two codon–anticodon pairings have normal Watson–Crick geometry. The structural constraints that this places on the third codon–anticodon pairing ensure that its conformation does not drastically differ from that of a Watson–Crick pair. Crick then proposed that there could be a small amount of play or “wobble” in the third codon position which allows limited conformational adjustments in its pairing geometry. This permits the formation of several non-Watson–Crick pairs such as U · G and I · A (Fig. 32-25a). The allowed “wobble” pairings are indicated in Fig. 32-25b. Then, by analyzing the known pattern of codon–anticodon pairing, Crick deduced the most plausible sets of pairing combinations in the third codon–anticodon position (Table 32-5). Thus, an anticodon

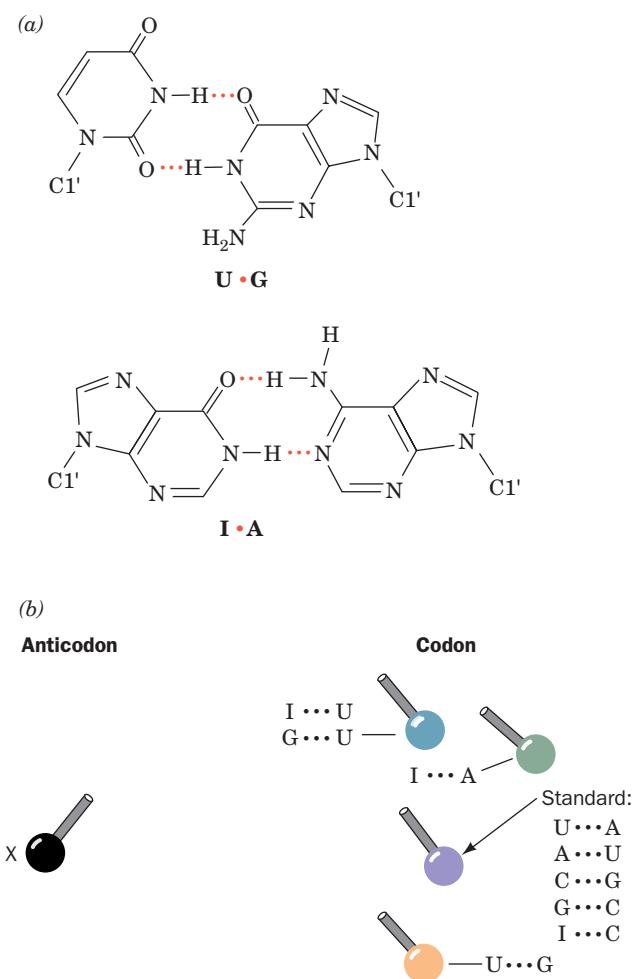


Figure 32-25 Wobble pairing. (a) U · G and I · A wobble pairs. Both have been observed in X-ray structures. (b) The geometry of wobble pairing. The spheres and their attached bonds represent the positions of ribose C1' atoms with their accompanying glycosidic bonds. X (left) designates the nucleoside at the 5' end of the anticodon (tRNA). The positions on the right are those of the 3' nucleoside of the codon (mRNA) in the indicated wobble pairings. [After Crick, F.H.C., *J. Mol. Biol.* **19**, 552 (1966).]

with C or A in its third position can only pair with its Watson–Crick complementary codon. If U, G, or I occupies the third anticodon position, two, two, or three codons are recognized, respectively.

Table 32-5 Allowed Wobble Pairing Combinations in the Third Codon–Anticodon Position

5'-Anticodon Base	3'-Codon Base
C	G
A	U
U	A or G
G	U or C
I	U, C, or A

No prokaryotic or eukaryotic cytoplasmic tRNA is known to participate in a nonwobble pairing combination. There is, however, no known instance of such a tRNA with an A in its third anticodon position, which suggests that the consequent A · U pair is not permitted. The structural basis of wobble pairing is poorly understood, although it is clear that it is influenced by base modifications.

A consideration of the various wobble pairings indicates that at least 31 tRNAs are required to translate all 61 coding triplets of the genetic code (there are 32 tRNAs in the minimal set because translational initiation requires a separate tRNA; Section 32-3Ca). Most cells have >32 tRNAs, some of which have identical anticodons. In fact, mammalian cells have >150 tRNAs. Nevertheless, *all isoaccepting tRNAs in a cell are recognized by a single aminoacyl-tRNA synthetase*.

c. Some Mitochondrial tRNAs Have More Permissive Wobble Pairings than Other tRNAs

The codon recognition properties of mitochondrial tRNAs must reflect the fact that mitochondrial genetic codes are variants of the “standard” genetic code (Table 32-3). For instance, the human mitochondrial genome, which consists of only 16,569 bp, encodes 22 tRNAs (together with 2 ribosomal RNAs and 13 proteins). Fourteen of these tRNAs each read one of the synonymous pairs of codons indicated in Tables 32-2 and 32-3 (MNX, where X is either C or U or else A or G) according to normal G · U wobble rules: The tRNAs have either a G or a modified U in their third anticodon position that, respectively, permits them to pair with codons having X = C or U or else X = A or G. The remaining 8 tRNAs, which, contrary to wobble rules, each recognize one of the groups of four synonymous codons (MNY, where Y = A, C, G, or U), all have anticodons with a U in their third position. Either this U can somehow pair with any of the four bases or these tRNAs read only the first two codon positions and ignore the third. Thus, not surprisingly, many mitochondrial tRNAs have unusual structures in which, for example, the GTψCRA sequence (Fig. 32-9) is missing, or, in the most bizarre case, a tRNA^{Ser} lacks the entire D arm.

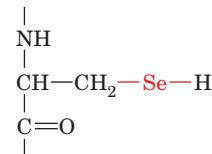
d. Frequently Used Codons Are Complementary to the Most Abundant tRNA Species

The analysis of the base sequences of several highly expressed structural genes of *S. cerevisiae* has revealed a remarkable bias in their codon usage. Only 25 of the 61 coding triplets are commonly used. *The preferred codons are those that are most nearly complementary, in the Watson-Crick sense, to the anticodons in the most abundant species in each set of isoaccepting tRNAs.* Furthermore, codons that bind anticodons with two consecutive G · C pairs or three A · U pairs are avoided so that the preferred codon-anticodon complexes all have approximately the same binding free energies. A similar phenomenon occurs in *E. coli*, although several of its 22 preferred codons differ from those in yeast. The degree with which the preferred codons occur in a given gene is strongly correlated, in both organisms, with the gene’s level of expression (the measured rates of aminoacyl-tRNA selection in *E. coli* span a

25-fold range). This, it has been proposed, permits the mRNAs of proteins that are required in high abundance to be rapidly and smoothly translated.

e. Selenocysteine and Pyrrolysine Are Carried by Specific tRNAs

Although it is widely stated, even in this text, that proteins are synthesized from the 20 “standard” amino acids, that is, those specified by the “standard” genetic code, some organisms, as Theresa Stadtman discovered, use a twenty-first amino acid, **selenocysteine (Sec; alternatively SeCys)**, in synthesizing a few of their proteins:

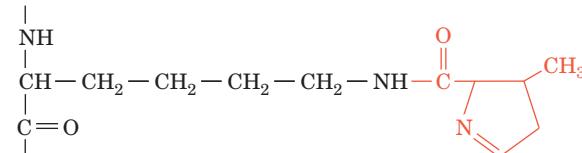


The selenocysteine (Sec) residue

Selenium, a biologically essential trace element, is a component of several enzymes in both prokaryotes and eukaryotes. These include thioredoxin reductase (Section 28-3Ae) and the **thyroid hormone deiodinases** (which participate in thyroid hormone synthesis; Section 19-1D) in mammals and three forms of **formate dehydrogenases** in *E. coli*, all of which contain Sec residues. The Sec residues are ribosomally incorporated into these proteins by a unique tRNA, **tRNA^{Sec}**, bearing a UCA anticodon that is specified by a particular (in the mRNA) UGA codon (normally the *opal* Stop codon). The Sec-tRNA^{Sec} is synthesized by the aminoacylation of tRNA^{Sec} with L-serine by the same SerRS that charges tRNA^{Ser}, followed by the enzymatic selenylation of the resulting Ser residue.

How does the ribosomal system differentiate a Sec-specifying UGA codon from a normal opal Stop codon? As we saw to be the case with Glu-tRNA^{Gln} (Section 32-2Cf), EF-Tu, the elongation factor that conducts most aminoacyl-tRNAs to the ribosome in a GTP-dependent process, does not bind Sec-tRNA^{Sec}. Instead it is bound by a specific elongation factor named **SELB**, a homolog of EF-Tu, that in its complex with GTP is recruited to a ribosomally bound mRNA stem-loop structure in the selenoprotein coding region on the 3' side of the UGA codon specifying Sec.

Certain methanogenic (methane-producing) archaea express the enzyme **methylamine methyltransferase**, which contains the amino acid residue **pyrrolysine (Pyl)**, a Lys with its ε-nitrogen in amide linkage to a pyrrolidine group:



The pyrrolysine (Pyl) residue

Unlike post-translationally modified Lys residues, such as 5-hydroxylysine (Hyl; Section 8-2B) and ε-N-acetyllysine

(Section 4-3B), Pyl is directly incorporated into proteins during translation. Pyl is specified by the codon UAG (normally the *amber* Stop codon). Pyl is carried to the ribosome by **tRNA^{Pyl}**, which contains a CUA anticodon and differs from typical tRNAs in having a D loop with five rather than eight residues, an anticodon stem with six rather than five base pairs, and a T ψ C loop that lacks the sequence T ψ C. A specific aminoacyl-tRNA synthetase, **PylRS**, that differs from known LysRSs, charges tRNA^{Pyl} with pyrrolysine in an ATP-dependent reaction, the first known example in nature of the direct aminoacylation of a tRNA with a “nonstandard” amino acid. Unlike the case for Ser-tRNA^{Sec}, Pyl-tRNA^{Pyl} is delivered to the ribosome by EF-Tu. This suggests that the mRNA contains a signal that causes UAG to be read as a Pyl codon rather than as a Stop codon. A conserved stem-loop structure located on the 3' side of UAG codons specifying Pyl may comprise this signal. Alternatively, a Pyl-tRNA^{Pyl} may occasionally read a UAG codon and is therefore a type of nonsense suppressor (see below).

E. Nonsense Suppression

Nonsense mutations are usually lethal when they prematurely terminate the synthesis of an essential protein. An organism with such a mutation may nevertheless be “rescued” by a second mutation on another part of the genome. For many years after their discovery, the existence of such **intergenic suppressors** was quite puzzling. It is now known, however, that they usually arise from mutations in a tRNA gene that causes the tRNA to recognize a nonsense codon. Such a **nonsense suppressor** tRNA appends its amino acid (which is the same as that carried by the corresponding wild-type tRNA) to a growing polypeptide in response to the recognized Stop codon, thereby preventing chain termination. For example, the *E. coli* amber suppressor known as **su3** is a tRNA^{Tyr} whose anticodon has mutated from the wild-type GUA (which reads the Tyr codons UAU and UAC) to CUA (which recognizes the amber Stop codon UAG). An **su3⁺** *E. coli* with an otherwise lethal *amber* mutation in a gene coding for an essential protein would be viable if the replacement of the wild-type amino acid residue by Tyr does not inactivate the protein.

There are several well-characterized examples of *amber* (UAG), *ochre* (UAA), and *opal* (UGA) suppressors in *E. coli* (Table 32-6). Most of them, as expected, have mutated anticodons. **UGA-1** tRNA, however, differs from the wild type only by a G → A mutation in its D stem, which changes a G · U pair to a stronger A · U pair. This mutation apparently alters the conformation of the tRNA's CCA anticodon so that it can form an unusual wobble pairing with UGA as well as with its normal codon, UGG. Nonsense suppressors also occur in yeast.

a. Suppressor tRNAs Are Mutants of Minor tRNAs

How do cells tolerate a mutation that both eliminates a normal tRNA and prevents the termination of polypeptide synthesis? They survive because the mutated tRNA is usually a minor member of a set of isoaccepting tRNAs and be-

Table 32-6 Some *E. coli* Nonsense Suppressors

Name	Codon Suppressed	Amino Acid Inserted
<i>su1</i>	UAG	Ser
<i>su2</i>	UAG	Gln
<i>su3</i>	UAG	Tyr
<i>su4</i>	UAA, UAG	Tyr
<i>su5</i>	UAA, UAG	Lys
<i>su6</i>	UAA	Leu
<i>su7</i>	UAA	Gln
UGA-1	UGA	Trp
UGA-2	UGA	Trp

Source: Körner, A.M., Feinstein, S.I., and Altman, S., in Altman, S. (Ed.), *Transfer RNA*, p. 109, MIT Press (1978).

cause nonsense suppressor tRNAs must compete for Stop codons with the protein factors that mediate the termination of polypeptide synthesis (Section 32-3F). Consequently, the rate of suppressor-mediated synthesis of active proteins with either UAG or UGA nonsense mutations rarely exceeds 50% of the wild-type rate, whereas mutants with UAA, the most common termination codon, have suppression efficiencies of <5%. Many mRNAs, moreover, have two tandem Stop codons so that even if their first Stop codon were suppressed, termination could occur at the second. Nevertheless, many suppressor-rescued mutants grow relatively slowly because they cannot make an otherwise prematurely terminated protein as efficiently as do wild-type cells.

Other types of suppressor tRNAs are also known. **Misense suppressors** act similarly to nonsense suppressors but substitute one amino acid in place of another. **Frameshift suppressors** have eight nucleotides in their anticodon loops rather than the normal seven. They read a four-base codon beyond a base insertion thereby restoring the wild-type reading frame.

3 RIBOSOMES AND POLYPEPTIDE SYNTHESIS

Ribosomes were first seen in cellular homogenates by dark-field microscopy in the late 1930s by Albert Claude who referred to them as “microsomes.” It was not until the mid-1950s, however, that George Palade observed them in cells by electron microscopy, thereby disposing of the contention that they were merely artifacts of cell disruption. The name ribosome derives from the fact that these particles in *E. coli* consist of approximately two-thirds RNA and one-third protein. (**Microsomes** are now defined as the artifactual vesicles formed by the endoplasmic reticulum on cell disruption. They are easily isolated by differential centrifugation and are rich in ribosomes.) The correlation between the amount of RNA in a cell and the rate at which it synthesizes protein led to the suspicion that ribosomes are the site of protein synthesis. This hypothesis was confirmed in 1955 by Paul Zamecnik, who demonstrated that ^{14}C -labeled amino acids are transiently associated with

Table 32-7 Components of *E. coli* Ribosomes

	Ribosome	Small Subunit	Large Subunit
Sedimentation coefficient	70S	30S	50S
Mass (kD)	2520	930	1590
RNA			
Major		16S, 1542 nucleotides	23S, 2904 nucleotides
Minor			5S, 120 nucleotides
RNA mass (kD)	1664	560	1104
Proportion of mass	66%	60%	70%
Proteins		21 polypeptides	31 polypeptides
Protein mass (kD)	857	370	487
Proportion of mass	34%	40%	30%

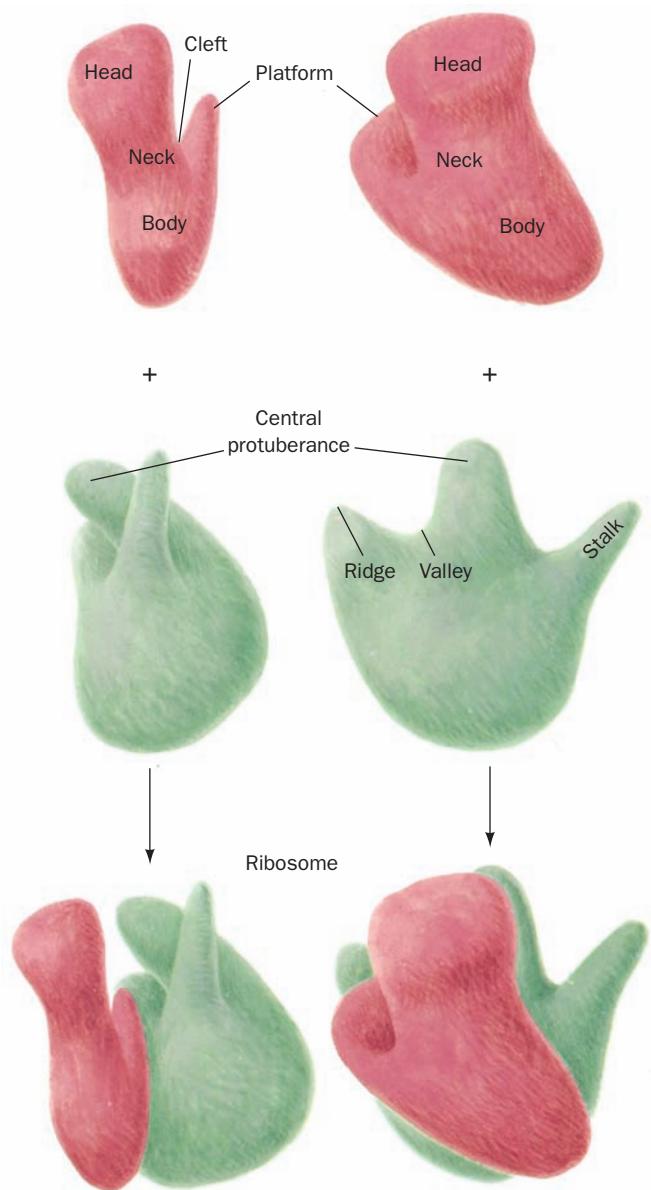
ribosomes before they appear in free proteins. Further research showed that ribosomal polypeptide synthesis has three distinct phases: (1) chain initiation, (2) chain elongation, and (3) chain termination.

In this section we examine the structure of the ribosome and then outline the ribosomal mechanism of polypeptide synthesis. In doing so we shall compare the properties of ribosomes from prokaryotes with those of eukaryotes.

A. Ribosome Structure

The *E. coli* ribosome, which has a particle mass of $\sim 2.5 \times 10^6$ D and a sedimentation coefficient of 70S, is a spheroidal particle that is ~ 250 Å across in its largest dimension. It may be dissociated, as James Watson discovered, into two unequal subunits (Table 32-7). The small (30S) subunit consists of a 16S rRNA molecule and 21 different polypeptides, whereas the large (50S) subunit contains a 5S and a 23S rRNA together with 31 different polypeptides. The up to 20,000 ribosomes in an *E. coli* cell account for $\sim 80\%$ of its RNA content and $\sim 10\%$ of its protein.

Structural studies of the ribosome through electron microscopy began soon after its discovery. Three-dimensional (3D) structures of the ribosome and its subunits at low (~ 50 Å) resolution first became available in the 1970s through image reconstruction techniques, pioneered by Klug, in which electron micrographs of a single particle or ordered sheets of particles taken from several directions are combined to yield its 3D image. The small subunit is a roughly mitten-shaped particle, whereas the large subunit is spheroidal with three protuberances on one side (Fig. 32-26).

**Figure 32-26** A low resolution model of the *E. coli* ribosome.

The small subunit (top; red) associates with the large subunit (middle; green) to form the intact ribosome (bottom). Two perpendicular views of each particle are provided. These models are based on transmission electron micrographs of negatively stained particles (in which the particle being imaged is embedded in electron-absorbing heavy metal salts, thereby providing contrast between the relatively electron-transparent particle and the background).

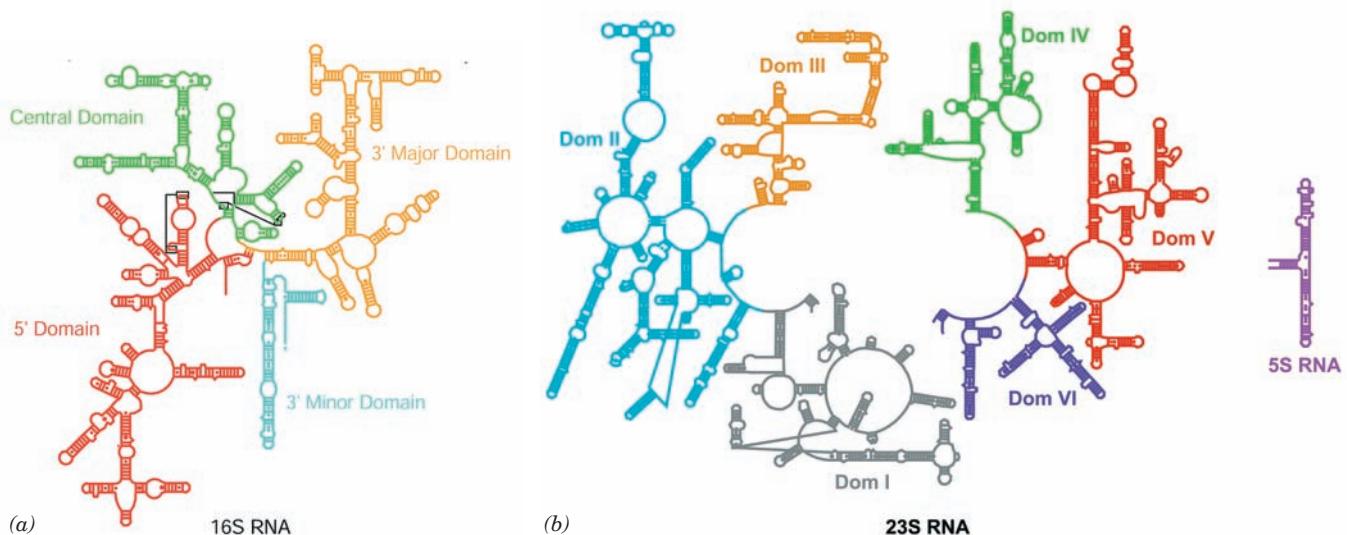


Figure 32-27 Secondary structures of the *E. coli* ribosomal RNAs. (a) 16S RNA and (b) 23S and 5S RNAs. The rRNAs are colored by domain with short lines spanning a stem representing Watson–Crick base pairs, small dots representing G · U base pairs, and large dots representing other non-Watson–Crick base

pairs. Note the flowerlike series of stems and loops forming each domain. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K., and Peter Moore, Yale University. Adapted from diagrams in <http://www.rna.ccb.utexas.edu/>.]

a. Ribosomal RNAs Have Complicated Secondary Structures

The *E. coli* 16S rRNA, which was sequenced by Harry Noller, consists of 1542 nucleotides. A computerized search of this sequence for stable double helical segments yielded many plausible but often mutually exclusive secondary structures. However, the comparison of the sequences of 16S rRNAs from several prokaryotes, under the assumption that their structures have been evolutionarily conserved, led to the flowerlike secondary structure for 16S rRNA seen in Fig. 32-27a. This four-domain structure is 54% base paired. Its double helical stems tend to be short (<8 bp) and many of them are imperfect. Intriguingly, electron micrographs of the 16S rRNA resemble those of the complete 30S subunit, thereby suggesting that the 30S subunit's overall shape is largely determined by the 16S rRNA. The large ribosomal subunit's 5S and 23S rRNAs, which consist of 120 and 2904 nucleotides, respectively, have also been sequenced. As with the 16S rRNA, they have extensive secondary structures (Fig. 32-27b).

b. Ribosomal Proteins Have Been Partially Characterized

Ribosomal proteins are difficult to separate because most of them are insoluble in ordinary buffers. By convention, ribosomal proteins from the small and large subunits are designated with the prefixes S and L, respectively, followed by a number indicating their position, from upper left to lower right, on a two-dimensional gel electrophoretogram (roughly in order of decreasing molecular mass; Fig. 32-28). Only protein S20/L26 appears to be common to both subunits. One of the large subunit proteins is partially acetylated at its N-terminus so that it gives rise to two elec-

trophoretic spots (L7/L12). Four copies of this protein, a dimer of dimers, are present in the large subunit. Moreover, these four copies of L7/L12 aggregate with L10 to form a stable complex that was initially thought to be a unique protein, “L8.” All the other ribosomal proteins occur in only one copy per subunit.

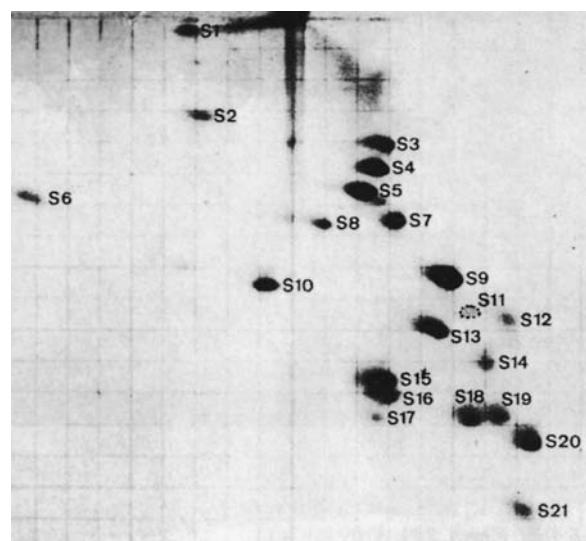


Figure 32-28 Two-dimensional gel electrophoretogram of *E. coli* small ribosomal subunit proteins. First dimension (vertical): 8% acrylamide, pH 8.6; second dimension (horizontal): 18% acrylamide, pH 4.6. [From Kaltschmidt, E. and Wittmann, H.G., *Proc. Natl. Acad. Sci.* **67**, 1277 (1970).]

The amino acid sequences of all 52 *E. coli* ribosomal proteins were elucidated, mainly by Heinz-Günter Wittmann and Brigitte Wittmann-Liebold. They range in size from 46 residues for L34 to 557 residues for S1. Most of these proteins, which exhibit little sequence similarity with one another, are rich in the basic amino acids Lys and Arg and contain few aromatic residues as is expected for proteins that are closely associated with polyanionic RNA molecules.

The X-ray and NMR structures of around half of the ribosomal proteins or their fragments have been independently determined. These proteins have a wide variety of structural motifs although most of their folds occur in other proteins of known structure. Around one-third of these ribosomal proteins contain the **RNA-recognition motif (RRM; Section 31-4Ab)**, which occurs in >200 RNA-binding proteins including rho factor (the transcriptional termination protein, which contains four such motifs; Section 31-2Db), poly(A) polymerase, poly(A)-binding protein (PABP), several proteins involved in gene splicing (Section 31-4A), and the translational initiation factor **EIF4B** (Section 32-3Cd). All of these proteins presumably evolved from an ancient RNA-binding protein.

c. Ribosomal Subunits Are Self-Assembling

Ribosomal subunits form, under proper conditions, from mixtures of their numerous macromolecular components. *Ribosomal subunits are therefore self-assembling entities*. Masayasu Nomura determined the order in which this occurs through partial reconstitution experiments. If one macromolecular component is left out of an otherwise

self-assembling mixture of proteins and RNA, the other components that fail to bind to the resulting partially assembled subunit must somehow interact with the omitted component. Through the analysis of a series of such partial reconstitution experiments, Nomura constructed an assembly map of the small (30S) subunit (Fig. 32-29). This map indicates that the initial steps in small subunit assembly are the independent binding to naked 16S rRNA of six so-called primary (1°) binding proteins (S4, S7, S8, S15, S17, and S20). The resulting assembly intermediates provide the molecular scaffolding for binding secondary (2°) binding proteins, which after a significant conformational change, form the attachment sites for tertiary (3°) binding proteins. An analogous assembly map for the large subunit was elucidated by Knud Nierhaus. The observation that similar assembly intermediates occur *in vivo* and *in vitro* suggests that *in vivo* and *in vitro* assembly processes are much alike.

In the cell, the 16S RNA folds in an ordered manner such that each domain is folded before the next domain is transcribed. The assembly of the small ribosomal subunit is then facilitated by a variety of **assembly factors**, proteins that bind to immature complexes but not to mature subunits. Many assembly factors associate with segments of the 16S RNA that change conformation during the latter stages of assembly. Presumably, the assembly of the large ribosomal subunit follows a similar course.

d. The Atomic Structure of the Prokaryotic Ribosome Has Been Long in Coming

The elucidation of the ribosome's atomic structure was a tortuous affair extending over four decades in which slow incremental improvements were occasionally punctuated by significant technical gains. The process began in the 1960s with shadowy transmission electron micrographs that provided only rough 2D shapes. This was followed in the 1970s by image reconstruction techniques that generated 3D models although still at low resolution (Fig. 32-26). Later in the 1970s, the sites of many of the ribosome's proteins were determined by James Lake and Georg Stöffler through **immune electron microscopy**, a technique in which antibodies raised against a particular ribosomal protein are used to mark its position in electron micrographs of the antibody complexed to a ribosomal subunit. These results were improved and extended in the 1980s by neutron scattering experiments conducted by Donald Engelman and Peter Moore on the 30S subunit, which indicated the distances between the centers of mass of its component proteins and hence their three-dimensional distribution. These structural studies were supplemented by a variety of chemical cross-linking and fluorescence transfer studies that demonstrated the proximity of various ribosomal components.

The molecular structure of the prokaryotic ribosome began to come into focus in the mid-1990s through the development of **cryo-electron microscopy (cryo-EM)**. In this technique, the sample is cooled to near liquid N₂ temperatures (-196°C) so rapidly (in a few milliseconds) that the water in the sample does not have time to crystallize but,

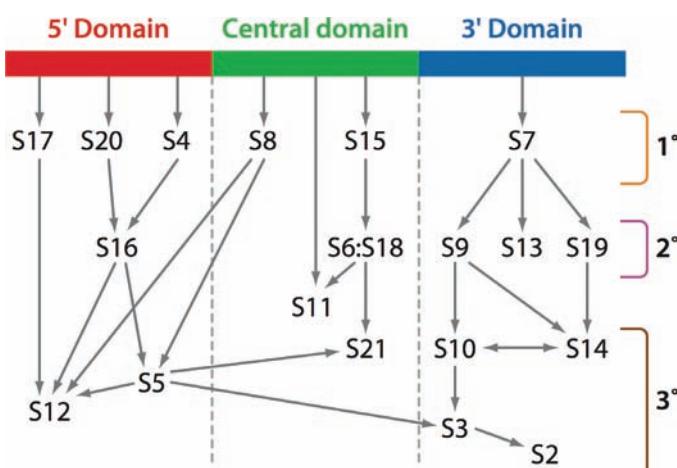
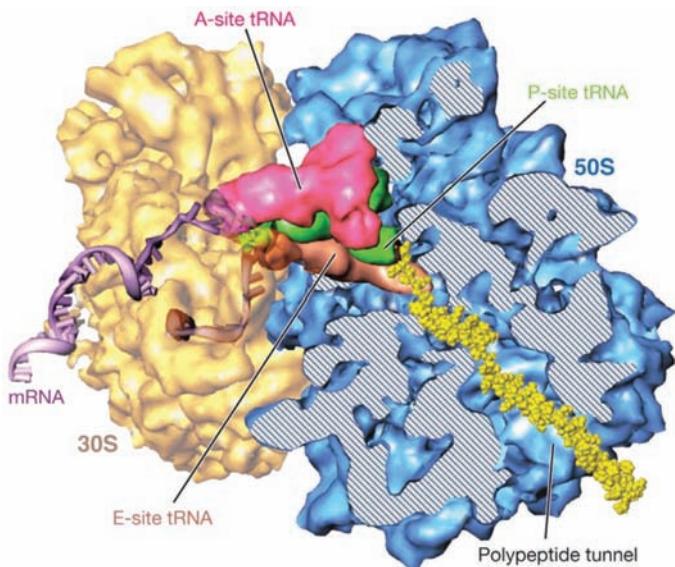


Figure 32-29 Assembly map of the *E. coli* small subunit. The map is organized according to the domains of the 16S rRNA (Fig. 32-27) with arrows indicating the facilitation of binding. For example, the arrow from the 5' domain of the 16S rRNA to S20 indicates that S20 binds directly to the 16S rRNA in the absence of other proteins and is therefore a primary (1°) binding protein; the arrow from S20 to S16 indicates that S20 facilitates the binding of S16, which is therefore a secondary (2°) binding protein; and the arrows from S16 to S5 and S12 indicate that S5 and S12 are tertiary (3°) binding proteins. [Courtesy of James Williamson, The Scripps Research Institute, La Jolla, California.]

Figure 32-30 Cryoelectron microscopy-based image of the *E. coli* ribosome. The 30S subunit (yellow) is on the left and the 50S subunit (blue) is on the right. The tRNAs that occupy the A, P, and E sites (Section 32-3B) are colored magenta, green, and brown. A portion of the 50S subunit has been cut away to reveal the polypeptide exit tunnel. A segment of mRNA (5' end brown and 3' end lavender) and the nascent polypeptide chain (yellow) have been modeled into the structure. [Courtesy of Joachim Frank, State University of New York at Albany.]



rather, assumes a vitreous (glasslike) state. Consequently, the sample remains hydrated and hence retains its native shape to a greater extent than in conventional electron microscopy (in which the sample is vacuum dried). Studies, carried out in large part by Joachim Frank, revealed the positions where tRNAs and mRNA as well as various soluble protein factors bind to the ribosome (Fig. 32-30). The highest resolution achieved by cryo-EM of ribosomes has gradually improved over the years to ~ 8 Å.

Ribosomal subunits were first crystallized by Ada Yonath in 1980 although they diffracted X-rays poorly. Over the course of several years, however, the quality of these crystals were incrementally improved until, in 1991, Yonath reported crystals of the 50S subunit that diffracted X-rays to 3-Å resolution. It was not until later in the 1990s, however, that technology was up to the task of determining the X-ray structures of these gargantuan molecular complexes. In 2000, the *annus mirabilis* (miracle year) of ribosomology, Moore and Steitz reported the X-ray structure of the 50S ribosomal subunit of the halophilic (salt-loving) bacterium *Haloarcula marismortui* at atomic (2.4-Å) resolution and Venki Ramakrishnan and Yonath independently reported the X-ray structure of the 30S subunit of *T. thermophilus* at ~ 3 -Å resolution. In 2001, Noller reported the 5.5-Å resolution structure of the entire *T. thermophilus* ribosome, which was gradually improved to 2.8 Å. In addition, the structures of the *E. coli* ribosome, by Jamie Cate, and the large subunit from *Deinococcus radiodurans*, by Yonath, have been determined. In the following paragraphs we discuss the properties of these groundbreaking structures. We consider their functional implications starting in Section 32-3C.

e. Ribosomal Architecture

Several generalizations can be made about ribosomal architecture based on the structures of the 30S and 50S subunits:

1. Both the 16S and 23S rRNAs are assemblies of coaxially stacked helical elements connected by loops, most of which are irregular extensions of helices (Fig. 32-31). These structures, which are in close accord with previous secondary structure predictions (Fig. 32-27), are stabilized by interactions between helices such as minor groove to minor groove packing, which also occur in the structures of the group I intron and RNase P (Sections 31-4Af and 31-4Ca; recall that A-form RNA has a very shallow minor groove); the insertion of a phosphate ridge into a minor groove; and adenines that are distant in sequence but often highly conserved that are inserted into minor grooves. The overall shapes of these subunits are relatively flat rather than globular with most regions having a thickness of two or three helical diameters [in contrast, other large (>100 nt) RNAs of known structure, such as the group I intron, RNase P, and the RNA from the signal recognition particle (Section 12-4Bb), are only one helix thick]. Although the determination of the structures of the 30S and 50S ribosomal subunits increased the amount of RNA structure that was then known at atomic resolution by ~ 10 -fold, nearly all of the secondary structural motifs seen in the ribosome also occur in smaller RNA structures. This suggests that the repertoire of RNA secondary structural motifs is limited.

2. Each of the 16S RNA's four domains, which extend out from a central junction (Fig. 32-27a), forms a morphologically distinct portion of the 30S subunit (Fig. 32-31a): The 5' domain forms most of the body (Fig. 32-26), the central domain forms the platform, the 3' major domain forms the entire head, and the 3' minor domain, which consists of just two helices, is located at the interface between the 30S and 50S subunits. In contrast, the 23S RNA's six domains (Fig. 32-27b) are intricately intertwined in the 50S subunit (Fig. 32-31b). Since the ribosomal proteins are embedded in the RNA (see below), this suggests that the domains of the 30S subunit can move relative to one

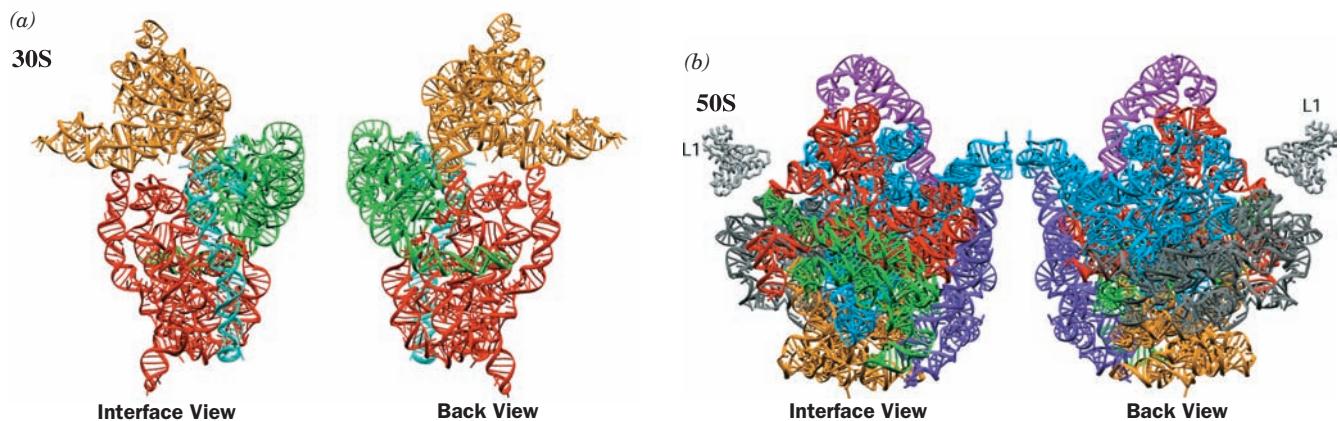


Figure 32-31 Tertiary structures of the ribosomal RNAs. (a) The 16S rRNA of *T. thermophilus*. (b) The 23S rRNA of *H. marismortui*. The rRNAs are colored according to domain as in Fig. 32-27. The interface view of a ribosomal subunit (*left*) is toward its surface that associates with the other subunit in the whole ribosome and the back view (*right*) is from the opposite (solvent-exposed) side. Note that the secondary structure

another during protein synthesis, whereas the 50S subunit appears to be rigid.

3. The distribution of the proteins in the two ribosomal subunits is not uniform (Fig. 32-32). The vast majority of the ribosomal proteins are located on the back and sides of their subunits. In contrast, the face of each subunit that forms the interface between the two subunits, particularly those regions that bind the tRNAs and mRNA (see below), is largely devoid of proteins.

4. Most ribosomal proteins consist of a globular domain, which is, for the most part, located on a subunit sur-

face (Fig. 32-32), and a long segment that is largely devoid of secondary structure and unusually rich in basic residues that infiltrates between the RNA helices into the subunit interior (Fig. 32-33). Indeed, a few ribosomal proteins lack a globular domain altogether (e.g., L39e in Fig. 32-33b). Ribosomal proteins make far fewer base-specific interactions than do other known RNA-binding proteins. They tend to interact with the RNA through salt bridges between their positively charged side chains and the RNAs' negatively charged phosphate oxygen atoms, thereby neutralizing the repulsive charge–charge interactions between nearby RNA segments. This is consistent with the hypothesis that

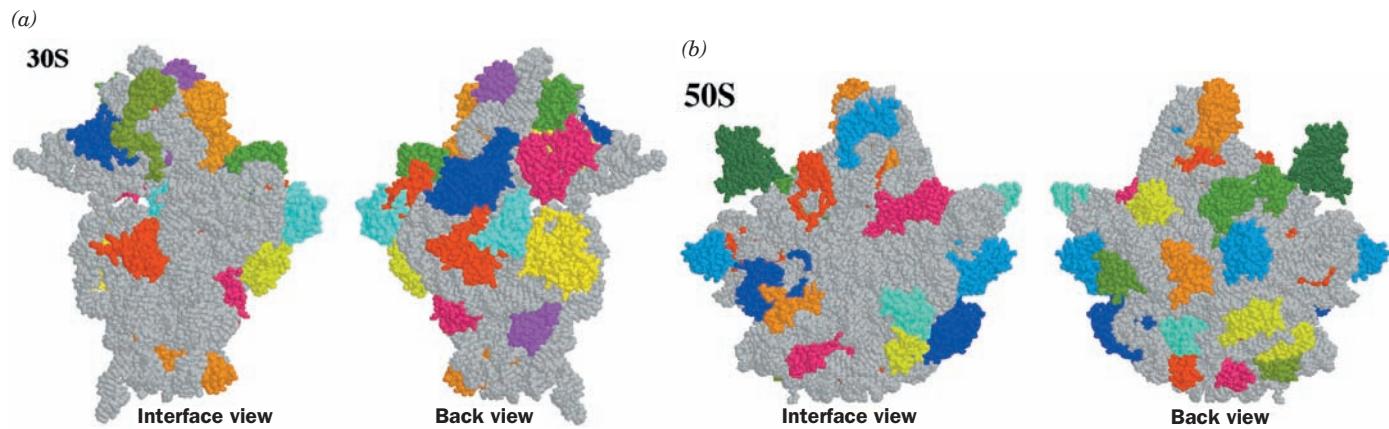


Figure 32-32 Distribution of protein and RNA in the ribosomal subunits. (a) The 30S subunit of *T. thermophilus*. (b) The 50S subunit of *H. marismortui*. The subunits are drawn in space-filling form with their RNAs gray and their proteins in various colors. Note that the interface side of each subunit is largely free of protein, particularly in its regions that interact

with mRNA and tRNAs. [Part a based on an X-ray structure by Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. Part b based on an X-ray structure by Peter Moore and Thomas Steitz, Yale University. PDBIDs 1J5E and 1JJ2.]

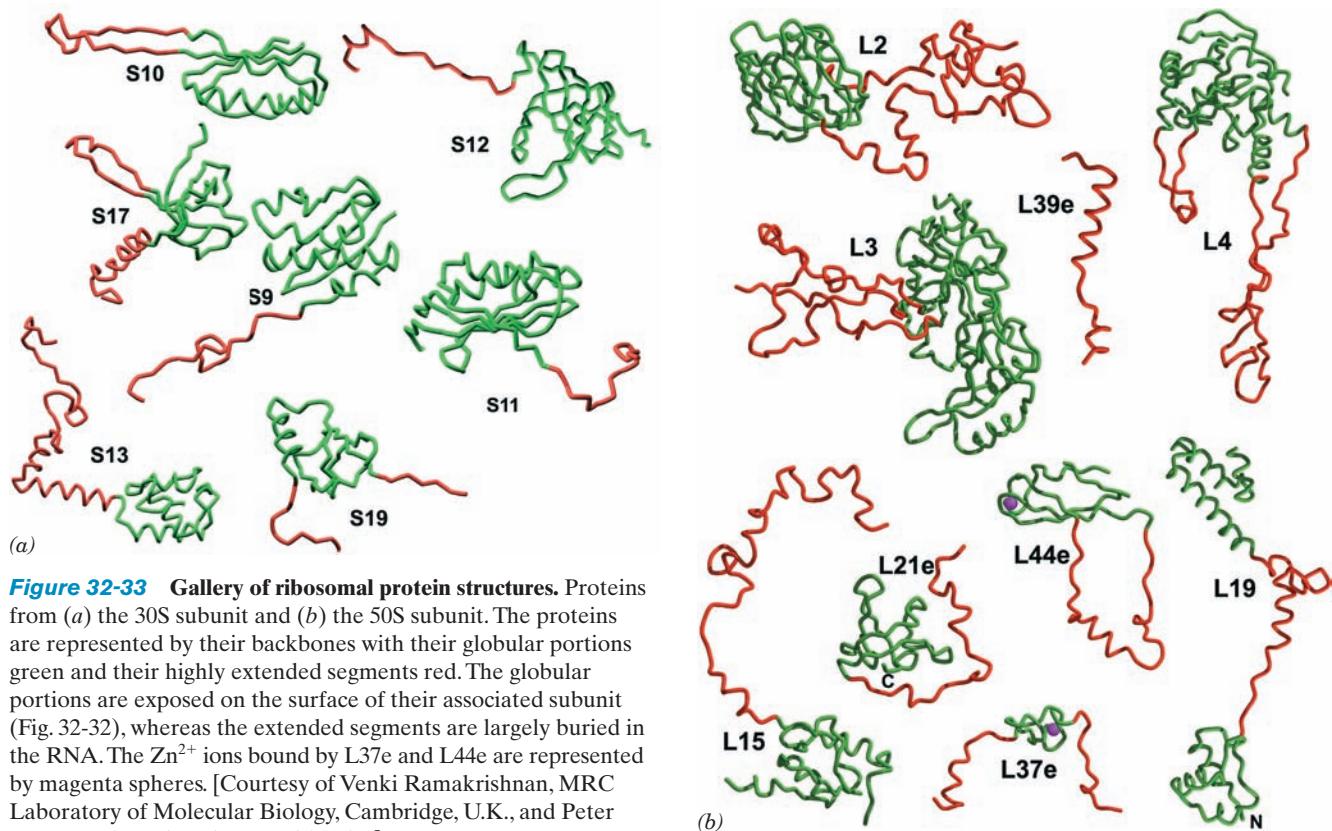


Figure 32-33 Gallery of ribosomal protein structures. Proteins from (a) the 30S subunit and (b) the 50S subunit. The proteins are represented by their backbones with their globular portions green and their highly extended segments red. The globular portions are exposed on the surface of their associated subunit (Fig. 32-32), whereas the extended segments are largely buried in the RNA. The Zn²⁺ ions bound by L37e and L44e are represented by magenta spheres. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K., and Peter Moore, Yale University. PDBid 1J5E.]

the primordial ribosome consisted entirely of RNA (the RNA world) and that the proteins that were eventually acquired stabilized its structure and fine-tuned its function.

The X-ray structure of the entire *T. thermophilus* ribosome in complex with three tRNAs and a 11-nt mRNA segment was determined by Ramakrishnan (Fig. 32-34; ribosomes, as we shall see in Section 32-3B, have three functionally distinct tRNA-binding sites known as the A, P, and E sites). The structures of the 30S and 50S subunits in this enormous molecular machine closely resemble those of the isolated subunits although there are several regions at the subunit interface that exhibit significant conformational shifts, which suggests that these changes occur as a consequence of subunit association. In addition, several disordered portions of the isolated *H. marismortui* 50S subunit are ordered in the intact *T. thermophilus* ribosome, although this may be a consequence of the latter's greater thermal stability.

The ribosome binds all three tRNAs in a similar manner with their anticodon stem-loops bound to the 30S subunit and their remaining portions, the D stem, elbow, and acceptor stem, bound to the 50S subunit. These interactions, which mainly consist of RNA–RNA contacts, are made to the tRNAs' universally conserved segments, thereby permitting the ribosome to bind different species of tRNAs in the similar ways.

The small and large ribosomal subunits contact each other at 12 positions via RNA–RNA, protein–protein, and

RNA–protein bridges. These intersubunit bridges have a distinct distribution: The RNA–RNA bridges are centrally located adjacent to the three bound tRNAs, whereas the protein–protein and RNA–protein bridges are peripherally located away from the ribosome's functional sites. The RNA–RNA contacts consist mainly of minor groove–minor groove interactions although major groove, loop, and backbone contacts also occur. In the RNA–protein bridges, the proteins contact nearly all types of RNA features including major groove, minor groove, backbone, and loop elements.

We discuss the path of the mRNA and how it interacts with the tRNAs in Section 32-3D. There we shall see that *the large subunit is mainly involved in mediating biochemical tasks such as catalyzing the reactions of polypeptide elongation, whereas the small subunit is the major actor in ribosomal recognition processes such as mRNA and tRNA binding* (although, as we have seen, the large subunit also participates in tRNA binding). We shall also see that *rRNA has the major functional role in ribosomal processes* (recall that RNA has demonstrated catalytic properties; Sections 31-4Ae and 31-4Ca).

f. Eukaryotic Ribosomes Are Larger and More Complex than Prokaryotic Ribosomes

Although eukaryotic and prokaryotic ribosomes resemble one another in both structure and function, they differ in nearly all details. Eukaryotic ribosomes have particle masses in the range 3.9 to 4.5×10^6 D and have a nominal

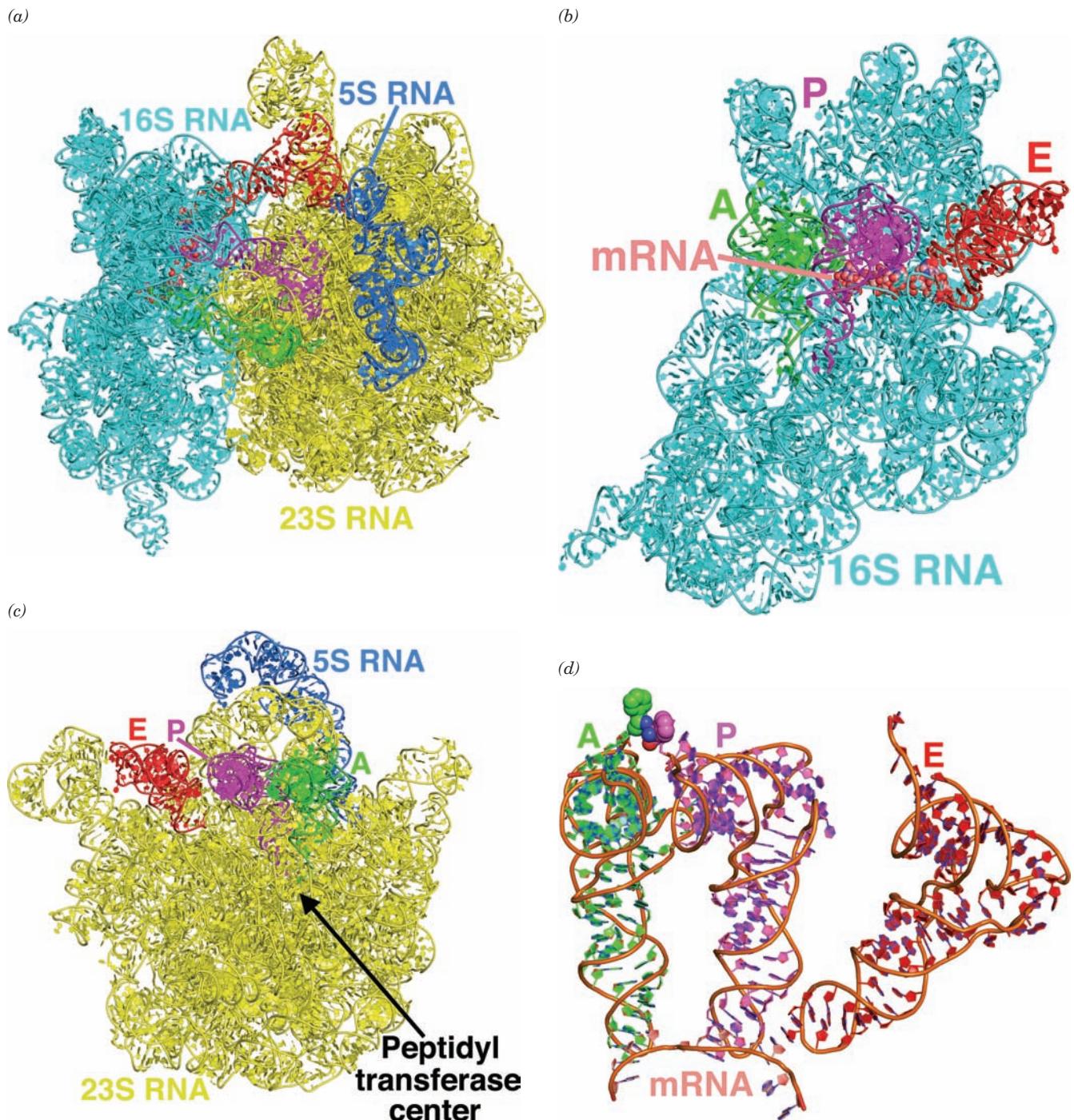


Figure 32-34 X-ray structure of the *T. thermophilus* ribosome in complex with tRNA and mRNA at 3.5 Å resolution. The E site binds tRNA^{Phe} and the A and P sites bind Phe-tRNA^{Phe} (in which the O atom forming an ester linkage from Phe to O3' of the 3' terminal A has been replaced by an NH group to prevent the hydrolysis of the Phe-tRNA linkage). (a) The RNA components of the ribosomal complex (its proteins are omitted for clarity) drawn in cartoon form except for the 11-residue mRNA, which is shown in space-filling form. The 16S RNA is cyan, the 23S RNA is yellow, the 5S RNA is blue, the tRNAs in the A, P, and E sites are green, magenta, and red, and the mRNA, which is largely occluded by the 16S RNA, is colored according to atom type with C pink, N blue, O red, and P orange. (b) The 16S RNA in interface view with its bound tRNAs and mRNA, all

represented as in Part a. (c) The 23S RNA in interface view (rotated 180° about the vertical direction relative to Part b) with its bound tRNAs all represented as in Part a. (d) The interactions of the tRNAs with the mRNA. This assembly is drawn in cartoon form with A-site C green, P-site C magenta, E-site C red, mRNA C pink, N blue, and O red and with successive P atoms connected by orange rods. The Phe residues appended to the A- and P-site tRNAs are represented in space-filling form. Note the close approach of these Phe residues and that the tRNAs in the A and P sites, but not that in the E site, form based paired codon-anticodon interactions with the mRNA. [Based on an X-ray structure by Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2WDK and 2WDL.]  See Interactive Exercise 44.

Table 32-8 Components of Rat Liver Cytoplasmic Ribosomes

	Ribosome	Small Subunit	Large Subunit
Sedimentation coefficient	80S	40S	60S
Mass (kD)	4220	1400	2820
RNA			
Major		18S, 1874 nucleotides	28S, 4718 nucleotides
Minor			5.8S, 160 nucleotides
			5S, 120 nucleotides
RNA mass (kD)	2520	700	1820
Proportion of mass	60%	50%	65%
Proteins		33 polypeptides	49 polypeptides
Protein mass (kD)	1700	700	1000
Proportion of mass	40%	50%	35%

sedimentation coefficient of 80S. They dissociate into two unequal subunits that have compositions that are distinctly different from those of prokaryotes (Table 32-8; compare with Table 32-7). The small (**40S**) subunit of the rat liver cytoplasmic ribosome, which together with the yeast ribosome is the most well-characterized eukaryotic ribosome, consists of 33 unique polypeptides and an **18S rRNA**. Its large (**60S**) subunit contains 49 different polypeptides and three rRNAs of **28S**, **5.8S**, and **5S**. The additional complexity of the eukaryotic ribosome relative to its prokaryotic counterpart is presumably due to the eukaryotic ribosome's additional functions: Its mechanism of translational initiation is more complex (Section 32-3Cd); it must be transported from the nucleus, where it is formed, to the cy-

toplasm, where translation occurs; and the machinery with which it participates in the secretory pathway is more complicated (Section 12-4B).

Sequence comparisons of the corresponding rRNAs from various species indicates that evolution has conserved their secondary structures rather than their base sequences (Figs. 32-27a and 32-35). For example, a G · C in a base paired stem of *E. coli* 16S rRNA has been replaced by an A · U in the analogous stem of yeast 18S rRNA. The **5.8S rRNA**, which occurs in the large eukaryotic subunit in base paired complex with the **28S rRNA**, is homologous in sequence to the 5' end of prokaryotic 23S rRNA. Apparently 5.8S RNA arose through mutations that altered rRNA's post-transcriptional processing, producing a fourth rRNA.

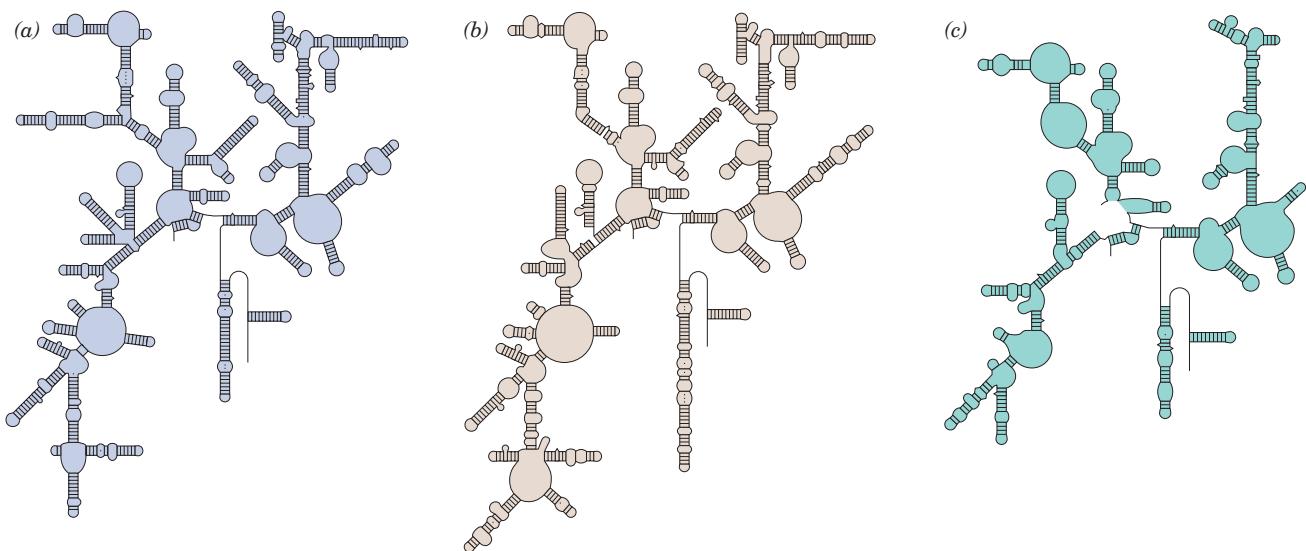


Figure 32-35 Predicted secondary structures of evolutionarily distant 16S-like rRNAs. (a) Archaeabacteria (*Halobacterium volcanii*), (b) eukaryotes (*S. cerevisiae*), and (c) mammalian mitochondria (bovine). Compare them with Fig. 32-27a, the secondary structure of 16S RNA from eubacteria (*E. coli*). Note the close similarities of these assemblies; they differ mostly by

insertions and deletions of stem-and-loop structures. The 23S-like rRNAs from a variety of species likewise have similar secondary structures. [After Gutell, R.R., Weiser, B., Woese, C.R., and Noller, H.F., *Prog. Nucleic Acid Res. Mol. Biol.* **32**, 183 (1985).]

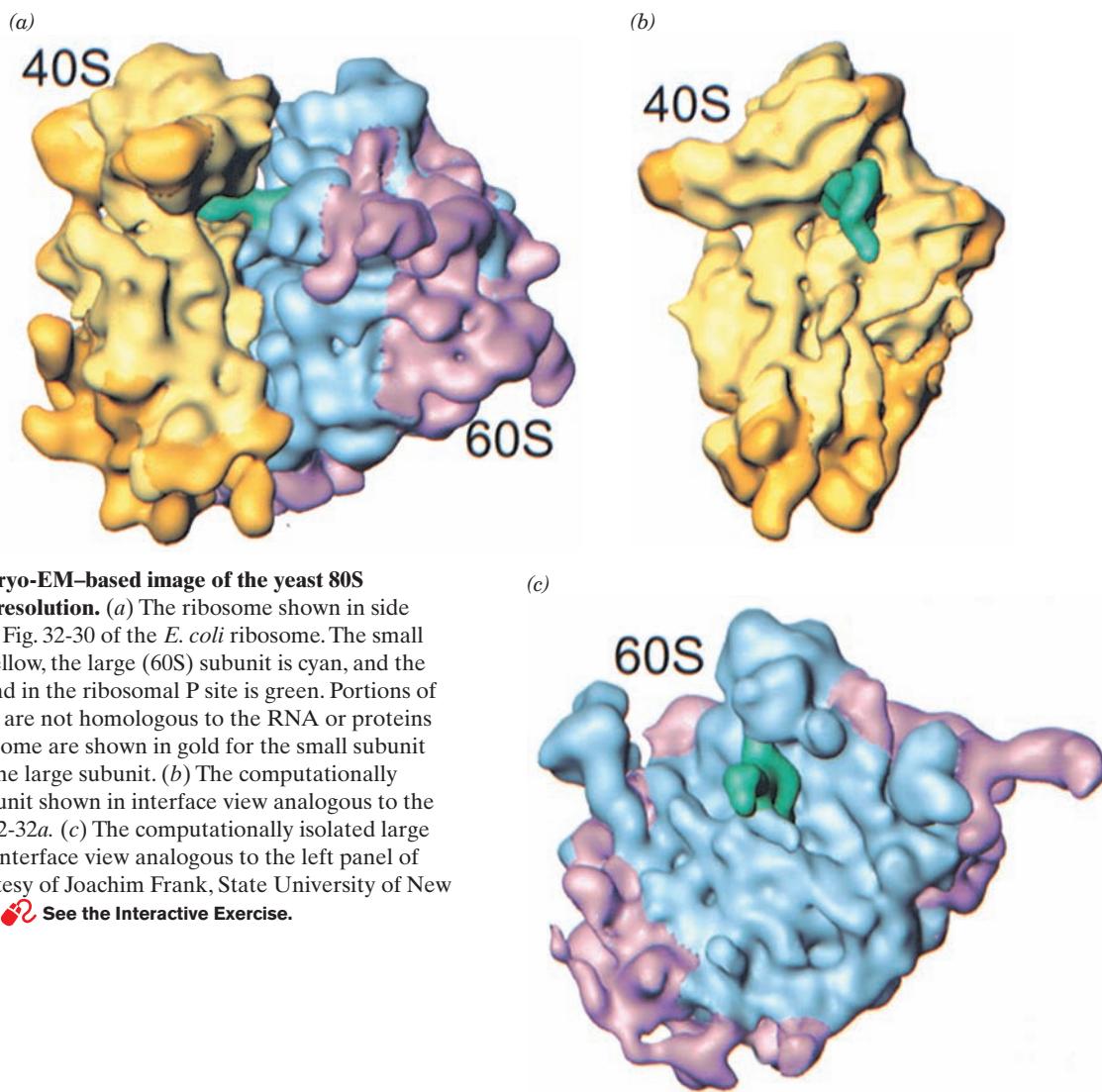


Figure 32-36 Cryo-EM-based image of the yeast 80S ribosome at 15 Å resolution. (a) The ribosome shown in side view analogous to Fig. 32-30 of the *E. coli* ribosome. The small (40S) subunit is yellow, the large (60S) subunit is cyan, and the tRNA that is bound in the ribosomal P site is green. Portions of this ribosome that are not homologous to the RNA or proteins of the *E. coli* ribosome are shown in gold for the small subunit and magenta for the large subunit. (b) The computationally isolated small subunit shown in interface view analogous to the left panel of Fig. 32-32a. (c) The computationally isolated large subunit shown in interface view analogous to the left panel of Fig. 32-32b. [Courtesy of Joachim Frank, State University of New York at Albany.] 

The cryo-EM-based image of the yeast 80S ribosome (Fig. 32-36), determined at 15 Å resolution by Andrej Sali, Günter Blobel, and Frank, reveals that there is a high degree of structural conservation between eukaryotic and prokaryotic ribosomes. Although the yeast 40S subunit (which consists of a 1798-nt 18S rRNA and 32 proteins) contains an additional 256 nt of RNA and 11 proteins relative to the *E. coli* 30S subunit (Table 32-8; 15 of the *E. coli* proteins are homologous to those of yeast), both exhibit a similar division into head, neck, body, and platform (Fig. 32-36b vs Figs. 32-32a and 32-34b). Many of the differences between these two small ribosomal subunits are accounted for by the 40S subunit's additional RNA and proteins, although their homologous portions exhibit several distinct conformational differences. Similarly, the yeast 60S subunit (Fig. 32-35c; which consists of an aggregate of 3671 nt and 45 proteins) structurally resembles the considerably smaller (Table 32-7) prokaryotic 50S subunit (Fig. 32-32b). The yeast ribosome exhibits 16 intersubunit bridges, 12 of which match the 12 that were observed in the X-ray structure of the *T. thermophilus* ribosome, a remarkable evolu-

tionary conservation that indicates the importance of these bridges. Moreover, the tRNA that occupies the P site of the yeast ribosome has a conformation that more closely resembles that of the P-site tRNA in the *T. thermophilus* ribosome than that of free tRNA^{Phe}.

B. Polypeptide Synthesis: An Overview

Before we commence our detailed discussion of polypeptide synthesis, it will be helpful to outline some of its major features.

a. Polypeptide Synthesis Proceeds from N-Terminus to C-Terminus

The direction of ribosomal polypeptide synthesis was established, in 1961 by Howard Dintzis, through radioactive labeling experiments. He exposed reticulocytes that were actively synthesizing hemoglobin to ³H-labeled leucine for times less than that required to make an entire polypeptide. The extent to which the tryptic peptides from the soluble (completed) hemoglobin molecules were labeled increased

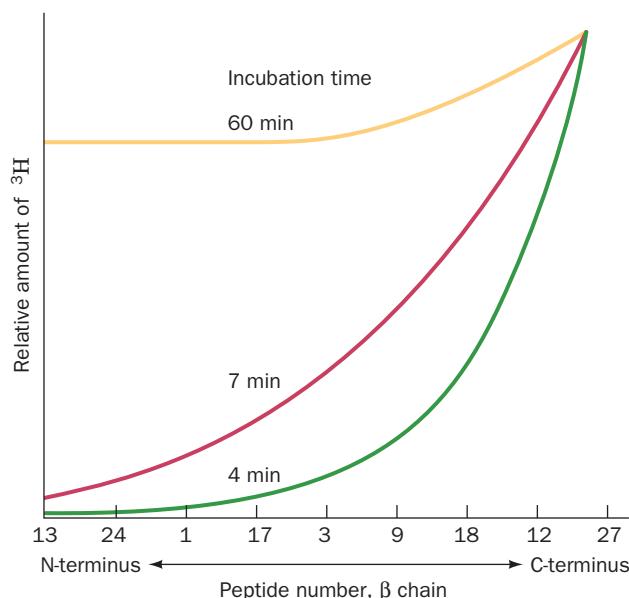


Figure 32-37 Demonstration that polypeptide synthesis proceeds from the N-terminus to the C-terminus. Rabbit reticulocytes were incubated with [^3H]leucine. The curves show the distribution of [^3H]Leu among the tryptic peptides from the β subunit of soluble rabbit hemoglobin after the indicated incubation times. The numbers on the horizontal axis are peptide identifiers arranged from N-terminus to C-terminus. [After Dintzis, H.M., *Proc. Natl. Acad. Sci.* **47**, 255 (1961).]

with their proximity to the C-terminus (Fig. 32-37). Incoming amino acids must therefore be appended to a growing polypeptide's C-terminus; that is, *polypeptide synthesis proceeds from N-terminus to C-terminus*.

b. Ribosomes Read mRNA in the 5' → 3' Direction

The direction in which the ribosome reads mRNAs was determined through the use of a cell-free protein synthesizing system in which the mRNA was poly(A) with a 3'-terminal C.



Such a system synthesizes a poly(Lys) that has a C-terminal Asn.



This, together with the knowledge that AAA and AAC code for Lys and Asn and the polarity of polypeptide synthesis, indicates that *the ribosome reads mRNA in the 5' → 3' direction*. Since mRNA is synthesized in the 5' → 3' direction, this accounts for the observation that, in prokaryotes, ribosomes initiate translation on nascent mRNAs (Section 31-3).

c. Active Translation Occurs on Polyribosomes

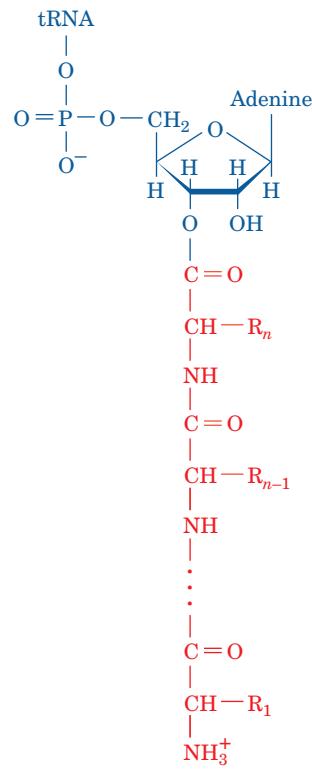
Electron micrographs, as Rich discovered, reveal that ribosomes engaged in protein synthesis are tandemly arranged on mRNAs like beads on a string (Fig. 31-27) to

form assemblies known as **polyribosomes (polysomes)**. Polysomes arise because once an active ribosome has cleared its initiation site on an mRNA, a second ribosome can initiate at that site.

Cryo-EM-based image reconstruction of *E. coli* polysomes, by Ulrich Hartl and Wolfgang Baumeister, has revealed that adjacent ribosomes are densely packed along the mRNA and have staggered or pseudohelical arrangements (Fig. 32-38). The mRNA is located along the inside of these assemblies, the tRNA entrance sites are accessible, and the polypeptide exit tunnel on each ribosome faces the cytosol. Modeling suggests that such an arrangement maximizes the distance between the nascent polypeptide chains exiting neighboring ribosomes. This reduces the ability of these incompletely folded polypeptides to nonspecifically aggregate and hence increases the yield of natively folded protein.

d. Chain Elongation Occurs by the Linkage of the Growing Polypeptide to the Incoming tRNA's Amino Acid Residue

During polypeptide synthesis, amino acid residues are sequentially added to the C-terminus of the nascent, ribosomally bound polypeptide chain. If the growing polypeptide is released from the ribosome by treatment with high salt concentrations, its C-terminal residue is invariably esterified to a tRNA molecule as a **peptidyl-tRNA**:



The nascent polypeptide must therefore grow by being transferred from the peptidyl-tRNA to the incoming aminoacyl-tRNA to form a peptidyl-tRNA with one more

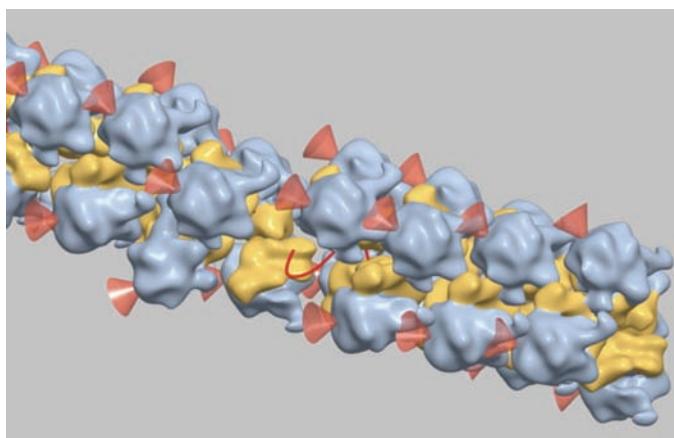


Figure 32-38 Cryo-EM-based image of an *E. coli* polysome.

The mRNA (which is mostly occluded) is represented by a red line, the small ribosomal subunits are yellow, the large subunits are blue-gray, and the red cones point to the polypeptide exit tunnel on each large subunit. The assembly shown has a pseudo-helical arrangement of ribosomes in which the center-to-center distance between adjacent ribosomes averages ~ 230 Å. Polysomes with somewhat different although equally densely packed arrangements of ribosomes have also been observed. [Courtesy of Ulrich Hartl and Wolfgang Baumeister, Max Planck Institute of Biochemistry, Martinsreid, Germany.]

residue (Fig. 32-39). Apparently, the ribosome has at least two tRNA-binding sites: the so-called **peptidyl** or **P site**, which binds the peptidyl-tRNA, and the **aminoacyl** or **A site**, which binds the incoming aminoacyl-tRNA (Fig. 32-39). Consequently, after the formation of a peptide bond, the newly deacylated P-site tRNA must be released and replaced by the newly formed peptidyl-tRNA from the A site, thereby permitting a new round of peptide bond for-

mation. The finding by Knud Nierhaus that each ribosome can bind up to three deacylated tRNAs but only two aminoacyl-tRNAs indicates, however, that the ribosome has a third tRNA-binding site, the **exit** or **E site**, which transiently binds the outgoing deacylated tRNA. All three sites, as we have seen (Fig. 32-34), extend over both ribosomal subunits.

The details of the chain elongation process are discussed in Section 32-3D. Chain initiation and chain termination, which are special processes, are examined in Sections 32-3C and 32-3F, respectively. In all of these sections we shall first consider the process of interest in *E. coli* and then compare it with the analogous eukaryotic activity.

C. Chain Initiation

a. fMet Is the N-Terminal Residue of Prokaryotic Polypeptides

The first indication that the initiation of translation requires a special codon, since identified as AUG (and, in prokaryotes, occasionally GUG), was the observation that almost half of the *E. coli* proteins begin with the otherwise uncommon amino acid Met. This was followed by the discovery of a peculiar form of Met-tRNA^{Met} in which the Met residue is *N*-formylated:

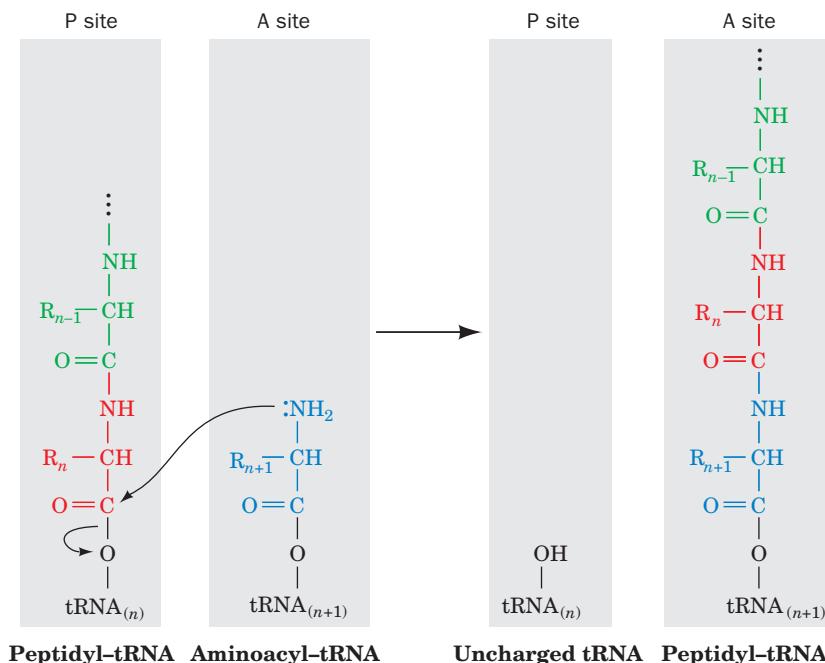
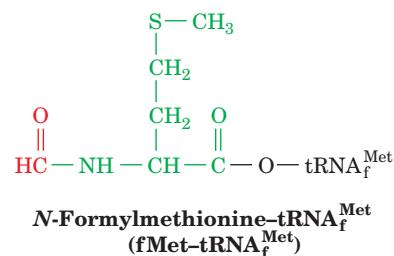


Figure 32-39 Ribosomal peptidyl transferase reaction forming a peptide bond. The ribosome catalyzes the nucleophilic attack of the amino group of the aminoacyl-tRNA in the A site on the peptidyl-tRNA ester in the P site, thereby forming a new peptide bond and transferring the nascent polypeptide to the A-site tRNA, while displacing the P-site tRNA.

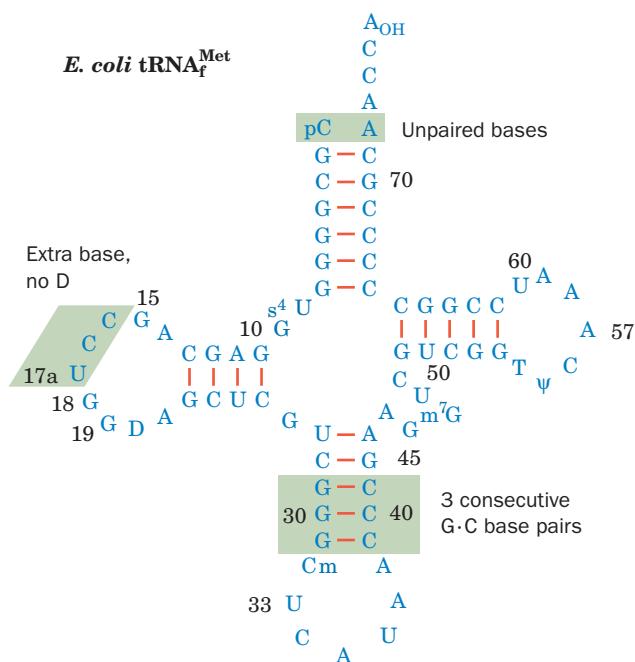


Figure 32-40 Nucleotide sequence of *E. coli* tRNA^{Met} shown in cloverleaf form. The shaded boxes indicate the significant differences between this initiator tRNA and noninitiator tRNAs such as yeast tRNA^{Ala} (Fig. 32-8). [After Woo, N.M., Roe, B.A., and Rich, A., *Nature* **286**, 346 (1980).]

The **N-formylmethionine** residue (**fMet**) already has an amide bond and can therefore only be the N-terminal residue of a polypeptide. In fact, polypeptides synthesized in an *E. coli*-derived cell-free protein synthesizing system always have a leading fMet residue. *fMet must therefore be E. coli's initiating residue.*

The tRNA that recognizes the initiation codon, tRNA^{Met} (Fig. 32-40), differs from the tRNA that carries internal Met residues, tRNA^{Met}, although they both recog-

nize the same codon. In *E. coli*, uncharged (deacylated) tRNA^{Met} is first aminoacylated with methionine by the same MetRS that charges tRNA^{Met}. The resulting **Met-tRNA^{Met}** is specifically *N*-formylated to yield **fMet-tRNA^{Met}** in an enzymatic reaction that employs *N*¹⁰-formyltetrahydrofolate (Section 26-4D) as its formyl donor. The formylation enzyme does not recognize Met-tRNA^{Met}. The X-ray structures of *E. coli* tRNA^{Met} and yeast tRNA^{Phe} (Fig. 32-11b) are largely similar but differ conformationally in their acceptor stems and anticodon loops. Perhaps these structural differences permit tRNA^{Met} to be distinguished from tRNA^{Met} in the reactions of chain initiation and elongation (see Section 32-3D).

E. coli proteins are post-translationally modified by a **deformylase**, which hydrolytically deforms the fMet residue, and, in many proteins, by the subsequent removal of the resulting N-terminal Met. This processing usually occurs on the nascent polypeptide, which accounts for the observation that mature *E. coli* proteins all lack fMet.

b. Base Pairing between mRNA and the 16S rRNA Helps Select the Translational Initiation Site

AUG codes for internal Met residues as well as the initiating Met residue of a polypeptide. Moreover, mRNAs usually contain many AUGs (and GUGs) in different reading frames. Clearly, a *translational initiation site must be specified by more than just an initiation codon*. This occurs in two ways: (1) the masking of AUGs that are not initiation codons by mRNA secondary structure; and (2) interactions between the mRNA and the 16S rRNA that select the initiating AUG as we now discuss.

The 16S rRNA contains a pyrimidine-rich sequence at its 3' end. This sequence, as John Shine and Lynn Dalgarno pointed out in 1974, is partially complementary to a purine-rich tract of 3 to 10 nucleotides, the **Shine-Dalgarno sequence**, that is centered ~10 nucleotides upstream from the start codon of nearly all known prokaryotic mRNAs (Fig. 32-41). *Base pairing interactions between an mRNA's*

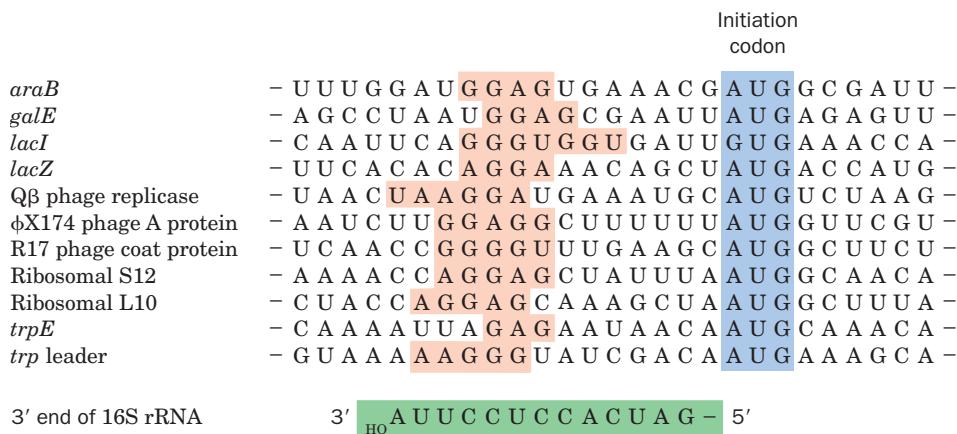


Figure 32-41 Some translational initiation sequences recognized by *E. coli* ribosomes. The mRNAs are aligned according to their initiation codons (blue shading). Their Shine-Dalgarno sequences (red shading) are complementary, counting G · U pairs, to a portion of the 16S rRNA's 3' end

(green shading; below). [After Steitz, J.A., in Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L., and Nomura, M. (Eds.), *Ribosomes. Structure, Function and Genetics*, pp. 481–482, University Park Press (1979).]

Shine–Dalgarno sequence and the 16S rRNA apparently permit the ribosome to select the proper initiation codon. Thus ribosomes with mutationally altered anti-Shine–Dalgarno sequences often have greatly reduced ability to recognize natural mRNAs, although they efficiently translate mRNAs whose Shine–Dalgarno sequences have been made complementary to the altered anti-Shine–Dalgarno sequences. Moreover, treatment of ribosomes with the bactericidal protein **colicin E3** (produced by *E. coli* strains carrying the E3 plasmid), which specifically cleaves a 49-nucleotide fragment from the 3' terminus of 16S rRNA, yields ribosomes that cannot initiate new polypeptide synthesis but can complete the synthesis of a previously initiated chain.

The X-ray structure of the 70S ribosome reveals, in agreement with Fig. 32-30, that a 27-nt segment of the mRNA is wrapped in a groove that encircles the neck of the 30S subunit (Fig. 32-42). The mRNA codons in the A and P sites are exposed on the interface side of the 30S subunit (as in Fig. 32-34b), whereas its 5' and 3' ends are bound in tunnels composed of RNA and protein. The mRNA's

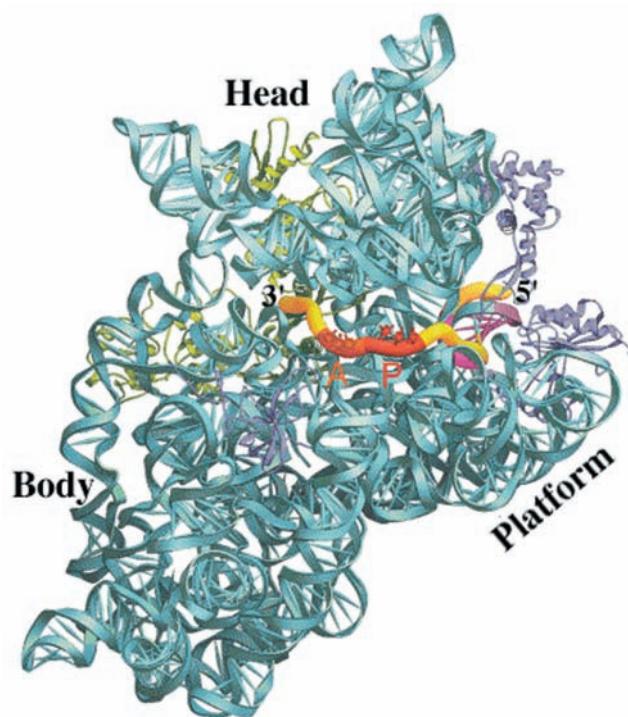


Figure 32-42 Path of mRNA through the ribosomal 30S subunit as viewed from its interface side. The 16S rRNA is cyan, and the 27-nt mRNA is represented in worm form with its A- and P-site codons orange and red, the Shine–Dalgarno helix (which includes a segment of 16S RNA) magenta, and its remaining segments yellow. The S3, S4, and S5 proteins are green, the S7, S11, and S12 proteins are purple, and the remaining ribosomal proteins have been omitted for clarity. The S3, S4, and S5 proteins, which in part form the tunnel through which the mRNA enters the ribosome, may function as a helicase to remove secondary structure from the mRNA that would otherwise interfere with tRNA binding. [Courtesy of Gloria Culver, Iowa State University. Based on an X-ray structure by Harry Noller, University of California at Santa Cruz. PDBid 1JGO.]

Shine–Dalgarno sequence, which is located near its 5' end, is base paired, as expected, with the 16S RNA's anti-Shine–Dalgarno sequence, which is situated close to the E site. The resulting double helical segment is accommodated in a cleft formed by both RNA and protein elements of the 16S subunit's head, neck, and platform (Fig. 32-26).

c. Prokaryotic Initiation Is a Three-Stage Process That Requires the Participation of Soluble Protein Initiation Factors

See Guided Exploration 28: Translational initiation Intact ribosomes do not directly bind mRNA so as to initiate polypeptide synthesis. Rather, *initiation is a complex process in which the two ribosomal subunits and fMet-tRNA_f^{Met} assemble on a properly aligned mRNA to form a complex that is competent to commence chain elongation. This assembly process also requires the participation of protein initiation factors that are not permanently associated with the ribosome.* Initiation in *E. coli* involves three initiation factors designated **IF-1**, **IF-2**, and **IF-3** (Table 32-9). Their existence was discovered when it was found that washing small ribosomal subunits with 1M ammonium chloride solution,

Table 32-9 The Soluble Protein Factors of *E. coli* Protein Synthesis

Factor	Number of Residues ^a	Function
Initiation Factors		
IF-1	71	Assists IF-3 binding
IF-2	890	Binds initiator tRNA and GTP
IF-3	180	Releases mRNA and tRNA from recycled 30S subunit and aids new mRNA binding
Elongation Factors		
EF-Tu	393	Binds aminoacyl-tRNA and GTP
EF-Ts	282	Displaces GDP from EF-Tu
EF-G	703	Promotes translocation through GTP binding and hydrolysis
Release Factors		
RF-1	360	Recognizes UAA and UAG Stop codons
RF-2	365	Recognizes UAA and UGA Stop codons
RF-3	528	Stimulates RF-1/RF-2 release via GTP hydrolysis
RRF	185	Together with EF-G, induces ribosomal dissociation to small and large subunits

^aAll *E. coli* translational factors are monomeric proteins.

which removes the initiation factors but not the “permanent” ribosomal proteins, prevents initiation.

The initiation sequence in *E. coli* ribosomes has three stages (Fig. 32-43):

1. On completing a cycle of polypeptide synthesis, the 30S and 50S subunits are separated (Section 32-3Fa). IF-3 then binds to the 30S subunit so as to prevent the reassociation of the 50S subunit. The X-ray structure of the 30S subunit in complex with the C-terminal domain of IF-3 (which by itself prevents the association of the 30S and 50S subunits), determined by Yonath and François Franceschi, indi-

cates that IF-3 binds to the upper end of the platform (Fig. 32-26) on its solvent (back) side. Hence IF-3 does not function by physically blocking the binding of the 50S subunit.

2. mRNA and IF-2 in a ternary complex with GTP and fMet-tRNA_f^{Met} that is accompanied by IF-1 subsequently bind to the 30S subunit in either order. Hence, fMet-tRNA_f^{Met} recognition must not be mediated by a codon–anticodon interaction. This interaction, nevertheless, helps bind fMet-tRNA_f^{Met} to the ribosome. IF-1 binds in the A site where it may function to prevent the inappropriate or premature binding of a tRNA. IF-3 also functions in this stage of the initiation process: it destabilizes the binding of tRNAs that lack the three G · C pairs in the anticodon stem of RNA_f^{Met} (Fig. 32-40) and helps discriminate between matched and mismatched codon–anticodon interactions.

3. Last, in a process that is preceded by IF-1 and IF-3 release, the 50S subunit joins the 30S initiation complex in a manner that stimulates IF-2 to hydrolyze its bound GTP to GDP + P_i. This irreversible reaction conformationally rearranges the 30S subunit and releases IF-2 for participation in further initiation reactions.

IF-2 is a member of the superfamily of regulatory GTPases such as Ras and hence is a **G protein** (Section 19-2A). The 30S initiation complex therefore functions as its **GAP** (GTPase-activating protein; Section 19-2Ca).

Initiation results in the formation of an fMet-tRNA_f^{Met} · mRNA · ribosome complex in which the fMet-tRNA_f^{Met} occupies the ribosome's P site while its A site is poised to accept an incoming aminoacyl-tRNA (an arrangement similar to that at the conclusion of a round of elongation; Section 32-3D). In fact, tRNA_f^{Met} is the only tRNA that directly enters the P site. All other tRNAs must do so via the A site during chain elongation (Section 32-3D). This arrangement was established through the use of the antibiotic puromycin as is discussed in Section 32-3Df.

d. Eukaryotic Initiation Is Far More Complicated than That of Prokaryotes

Although translational initiation in eukaryotes superficially resembles that in prokaryotes, it is, in fact, a far more complicated process. Whereas prokaryotic initiation only requires the assistance of three monomeric initiation factors, that in eukaryotes involves the participation of at least 12 initiation factors (designated eIF_n; “e” for eukaryotic) that consist of at least 26 polypeptide chains. Eukaryotic initiation occurs as follows (Fig. 32-44):

1. The process begins with the binding of **eIF3** (which in mammals consists of 13 different subunits) and **eIF1A** (a monomer and homolog of bacterial IF-1) to the 40S subunit in the inactive 80S ribosome (which had terminated elongation in its previous elongation cycle) so that it releases the 60S subunit.

2. The ternary complex of **eIF2** (a heterotrimer), GTP, and **Met-tRNA_i^{Met}** binds to the 40S ribosomal subunit accompanied by **eIF1** (a monomer) to form the so-called **43S**

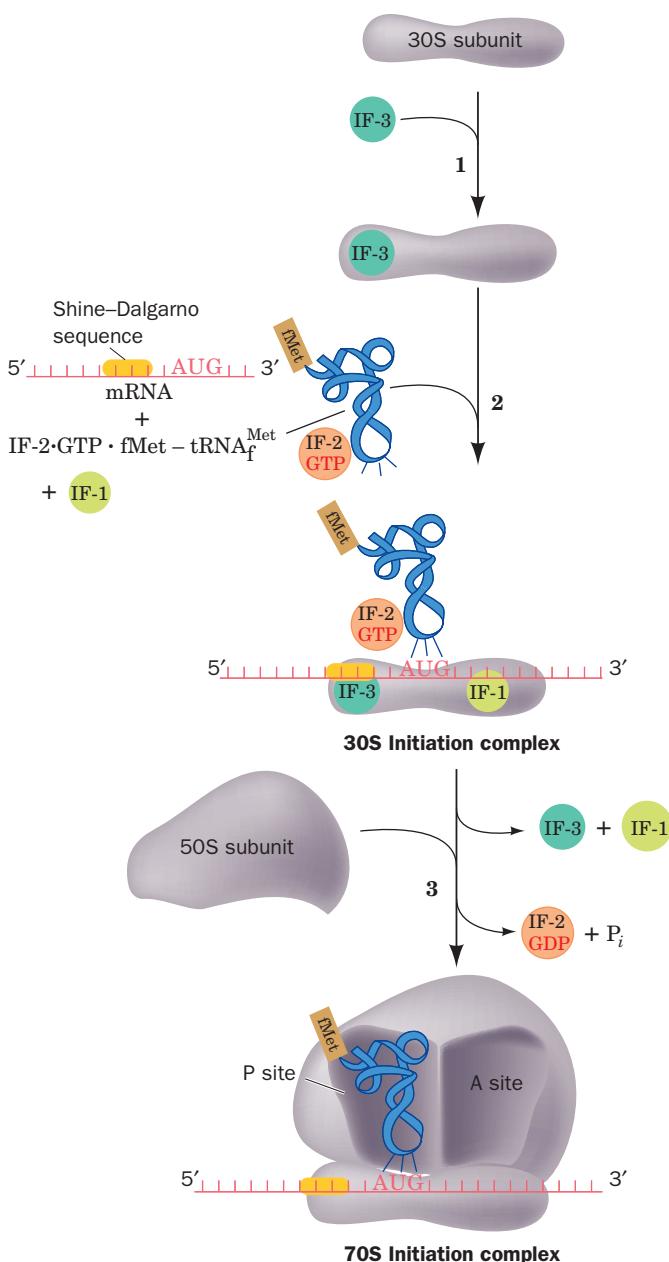


Figure 32-43 Translational initiation pathway in *E. coli*. The E site, which is unoccupied during this process, has been omitted for clarity.

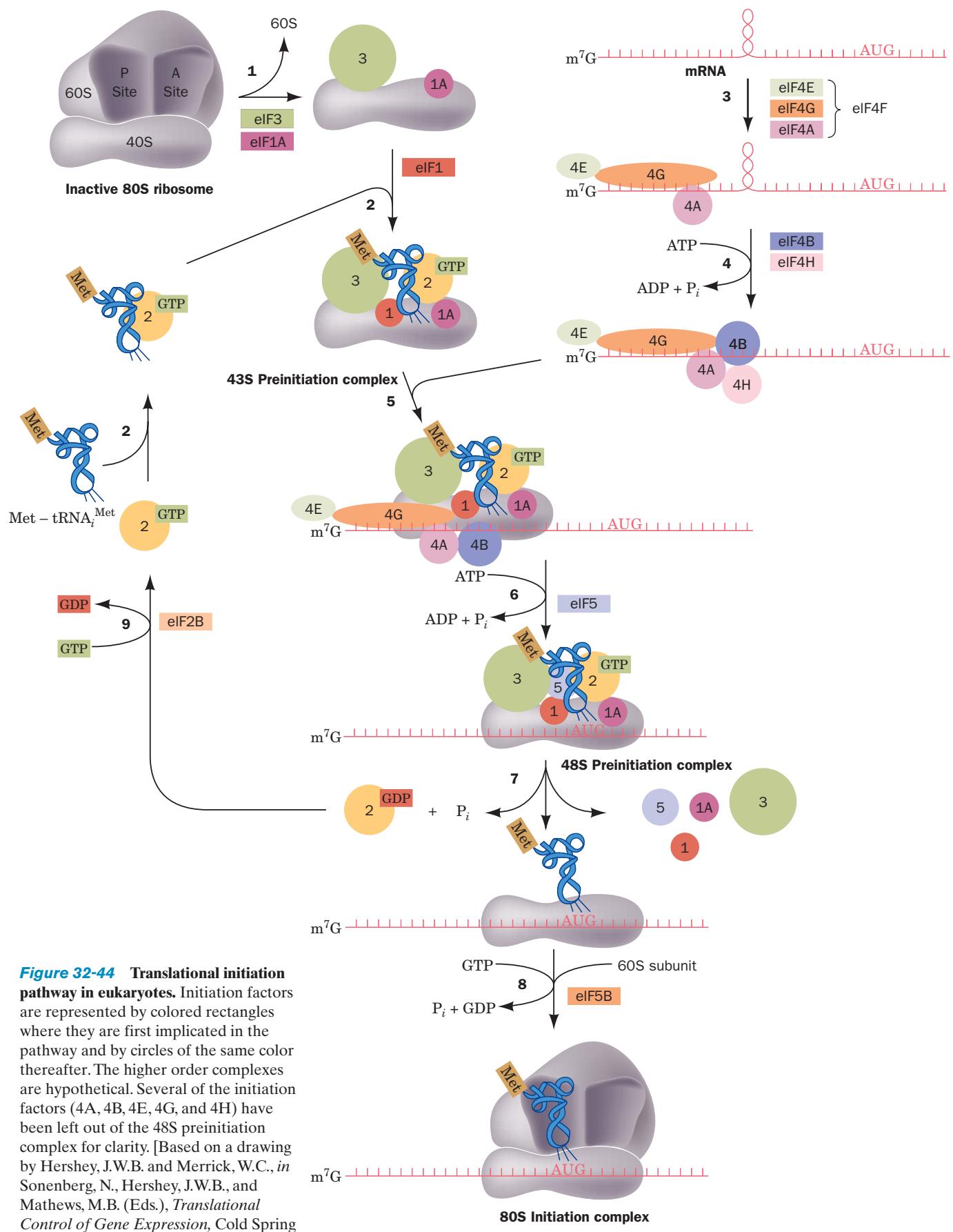


Figure 32-44 Translational initiation pathway in eukaryotes. Initiation factors are represented by colored rectangles where they are first implicated in the pathway and by circles of the same color thereafter. The higher order complexes are hypothetical. Several of the initiation factors (4A, 4B, 4E, 4G, and 4H) have been left out of the 48S preinitiation complex for clarity. [Based on a drawing by Hershey, J.W.B. and Merrick, W.C., in Sonenberg, N., Hershey, J.W.B., and Mathews, M.B. (Eds.), *Translational Control of Gene Expression*, Cold Spring Harbor Laboratory Press (2000).]

preinitiation complex. Here the subscript “i” on tRNA_i^{Met} distinguishes this eukaryotic initiator tRNA, whose appended Met residue is never *N*-formylated, from that of prokaryotes; both species are, nevertheless, readily interchangeable *in vitro*.

3. Eukaryotic mRNAs lack the complementary sequences to bind to the 18S rRNA in the Shine–Dalgarno manner. Rather, they have an entirely different mechanism for recognizing the mRNA’s initiating AUG codon. *Eukaryotic mRNAs, nearly all of which have an m⁷G cap and a poly(A) tail (Section 31-4Ab), are invariably monocistronic and almost always initiate translation at their leading AUG.* This AUG, which occurs at the end of a 5'-untranslated region of 50 to 70 nt, is embedded in the consensus sequence GCCRCCAUGG, with changes in the purine (R) 3 nt before the AUG and the G immediately following it reducing translational efficiency by ~10-fold each and with other changes having much smaller effects. In addition, secondary structure (stem-loops) in the mRNA upstream of the initiation site may affect initiation efficiency.

The recognition of the initiation site begins by the binding of **EIF4F** to the m⁷G cap. eIF4F is a heterotrimeric complex of **EIF4E**, **EIF4G**, and **EIF4A** (all monomers), in

which **EIF4E (cap-binding protein)** recognizes the mRNA’s m⁷G cap and eIF4G serves as a scaffold to join eIF4E with eIF4A. Both the X-ray and NMR structures of eIF4E in complex with **m⁷GDP**, determined by Nahum Sonenberg and Stephen Burley and by Sonenberg and Gerhard Wagner, reveal that the protein binds the m⁷G base by intercalating it between two highly conserved Trp residues (Fig. 32-45a) in a region that is adjacent to a positively charged cleft that forms the putative mRNA-binding site (Fig. 32-45b). The m⁷G base is specifically recognized by hydrogen bonding to protein side chains in a manner reminiscent of G · C base pairing. eIF4G also binds poly(A)-binding protein (PABP; Section 31-4Ab) bound to the mRNA’s poly(A) tail, thereby circularizing the mRNA (not shown in Fig. 32-44). Although this explains the synergism between an mRNA’s m⁷A cap and its poly(A) tail in stimulating translational initiation, the function of this circle is unclear. However, an attractive hypothesis is that it enables a ribosome that has finished translating the mRNA to reinitiate translation without having to disassemble and then reassemble. Another possibility is that it prevents the translation of incomplete (broken) mRNAs. This circularization, as we have seen in Section 31-4Av, also protects the mRNA from degradation

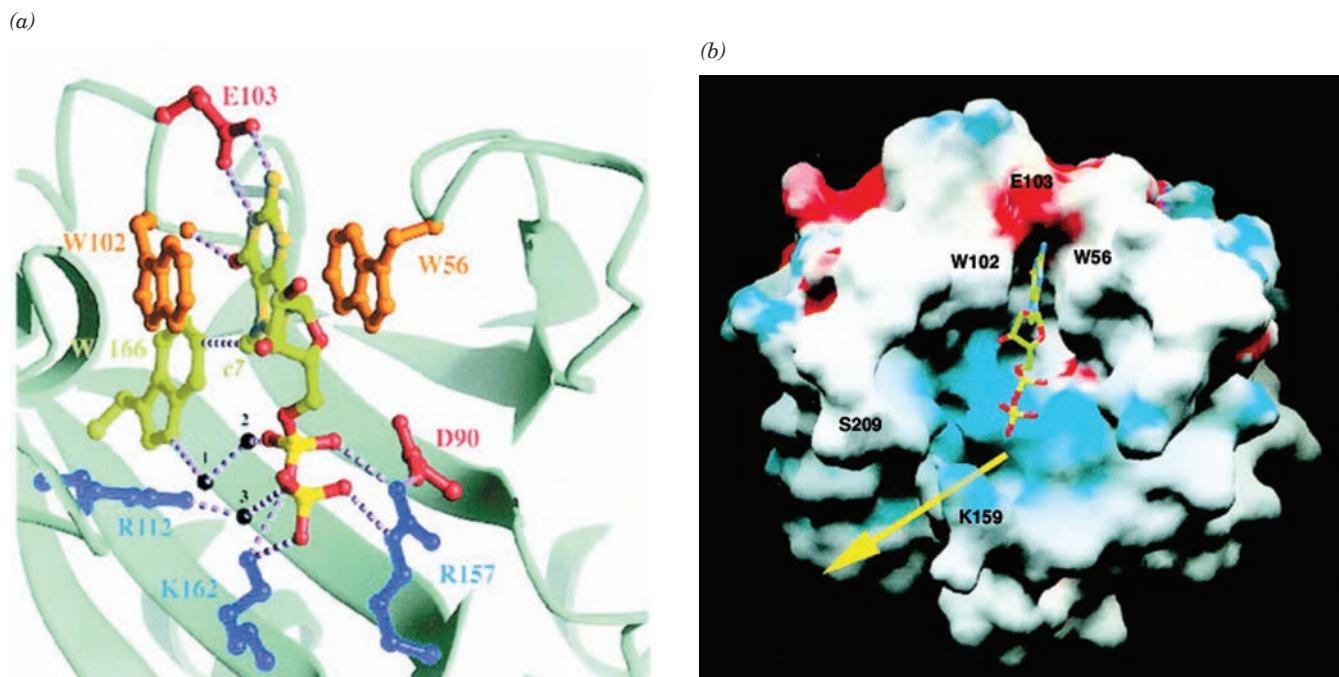


Figure 32-45 X-ray structure of murine eIF4E in complex with the m⁷G cap analog m⁷GDP. (a) The m⁷GDP-binding site with the m⁷GDP and the side chains that bind it drawn in ball-and-stick form with the atoms of the m⁷GDP colored according to type (C green, N dark yellow, O red, and P bright yellow) and the protein side chains with which the m⁷GDP interacts drawn in various colors. Hydrogen bonds, salt bridges, and van der Waals interactions are represented by dashed lines and bridging water molecules are drawn as black spheres. The m⁷G base is intercalated between the indole rings of Trp 56 and Trp 102, where it specifically interacts with protein side chains through

hydrogen bonds and van der Waals interactions. The GDP’s phosphate groups interact directly and indirectly with three basic side chains. (b) The solvent-accessible surface of eIF4E colored according to its electrostatic potential (red negative, blue positive, and white neutral) and viewed approximately as in Part a. The m⁷GDP is drawn in ball-and-stick form colored as in Part a. The mRNA presumably binds in the positively charged cleft (yellow arrow) that is adjacent to the m⁷G binding site and which passes between Lys 159 and Ser 209. [Courtesy of Nahum Sonenberg, McGill University, Montréal, Québec, Canada. PDBid 1EJ1.]

by preventing the action of decapping enzyme until the mRNA's poly(A) tail has been shortened to the point that it can no longer bind PABP.

4. eIF4B (an RRM-containing homodimer) and **eIF4H** (a monomer) join the eIF4F-mRNA complex where they stimulate the RNA helicase activity of eIF4A to unwind the mRNA's helical segments in an ATP-dependent process. This presumably also strips away the proteins that are bound to the mRNA (Section 31-4Au). eIF4A is the prototype of the so-called **DEAD-box family** of proteins (also known as DExD/H family proteins; Section 31-4Au), which is named after one of the sequence motifs shared by the diverse members of this family, all of which have NTPase activity.

5. The eIF4F-mRNA-eIF4B-eIF4H complex joins the 43S preinitiation complex through a protein-protein interaction between eIF4G and the 40S subunit-bound eIF3. This differs substantially from the corresponding prokaryotic process (Fig. 32-43) in which the mRNA is bound to the 30S ribosomal subunit via associations between RNA molecules (involving the Shine-Dalgarno sequence and the codon-anticodon interaction).

6. eIF5 (a monomer) joins the growing assembly. The 43S preinitiation complex then translocates along the mRNA, an ATP-dependent process called **scanning**, until it encounters the mRNA's AUG initiation codon, which is optimally in the sequence GCC(A/G)CCAUGG. This yields the **48S preinitiation complex**. The recognition of the AUG occurs mainly through base pairing with the CUA anticodon on the bound Met-tRNA_i^{Met}, as was demonstrated by the observation that mutating this anticodon results in the recognition of the new cognate codon instead of AUG. This explains why the initiator tRNA must bind to the small subunit before the mRNA.

7. The formation of the 48S preinitiation complex induces eIF2 to hydrolyze its bound GTP to GDP + P_i, which results in the release of all the initiation factors, thereby leaving the Met-tRNA_i^{Met} in the small subunit's P site. The hydrolysis reaction is stimulated by eIF5, acting as a GAP (Section 19-2Ca).

8. The 60S subunit then joins the mRNA-bound Met-tRNA_i^{Met}-40S subunit complex in a GTPase reaction mediated by **eIF5B** (a monomer and homolog of bacterial IF-2), thereby yielding the 80S ribosomal initiation complex. Thus eukaryotic translation initiation consumes two GTPs versus one for prokaryotic initiation (Fig. 32-43).

9. What remains is to recycle the eIF2 · GDP complex by exchanging its GDP for GTP. This reaction is mediated by **eIF2B** (a heteropentamer), which therefore functions as eIF2's **GEF** (guanine nucleotide exchange factor; Section 19-2Ca).

Many eukaryotic initiation factors are subject to phosphorylation/dephosphorylation and are therefore likely to par-

ticipate in the control of eukaryotic translation, a subject we discuss in Section 32-4.

Although the initiation sites on most eukaryotic mRNAs are identified by the above-described scanning mechanism, a few mRNAs have an **internal ribosome entry site (IRES)** to which the 40S subunit can directly bind in a process reminiscent of prokaryotic initiation. However, little is yet known about the mechanism of IRES-based initiation. Indeed, IRESs lack clearly identifiable consensus sequences.

D. Chain Elongation

See Guided Exploration 29: **Translational elongation** Ribosomes elongate polypeptide chains in a three-stage reaction cycle that adds amino acid residues to a growing polypeptide's C-terminus (Fig. 32-46):

1. Decoding, in which the ribosome selects and binds an aminoacyl-tRNA, whose anticodon is complementary to the mRNA codon in the A site.

2. Transpeptidation, in which the peptidyl group on the P-site tRNA is transferred to the aminoacyl group in the A site through the formation of a peptide bond (Fig. 32-39).

3. Translocation, in which A-site and P-site tRNAs are respectively transferred to the P site and E site accompanied by their bound mRNA; that is, the mRNA, together with its base paired tRNAs, is ratcheted through the ribosome by one codon.

Translational elongation, which occurs at a rate of 10 to 20 residues/s, involves the participation of several nonribosomal proteins known as **elongation factors** (Table 32-9). We describe these processes in the following paragraphs.

a. Decoding

In the decoding stage of the *E. coli* elongation cycle, a binary complex of GTP with the elongation factor **EF-Tu** (also called **EF1A**) combines with an aminoacyl-tRNA. The resulting ternary complex binds to the ribosome, and, in a reaction that hydrolyzes the GTP to GDP + P_i, the aminoacyl-tRNA is bound in a codon-anticodon complex to the ribosomal A site and EF-Tu · GDP + P_i is released. In the remainder of this stage, the bound GDP is replaced by GTP in a reaction mediated by the elongation factor **EF-Ts** (also called **EF1B**). EF-Tu, as are several other GTP-binding ribosomal factors, is a G-protein, and hence the ribosome functions as its GAP and EF-Ts is its GEF.

Aminoacyl-tRNAs can bind to the ribosomal A site without the mediation of EF-Tu but at a rate too slow to support cell growth. The importance of EF-Tu is indicated by the fact that it is the most abundant *E. coli* protein; it is present in ~100,000 copies per cell (>5% of the cell's protein), which is approximately the number of tRNA molecules in the cell. Consequently, *the cell's entire complement of aminoacyl-tRNAs is essentially sequestered by EF-Tu*.

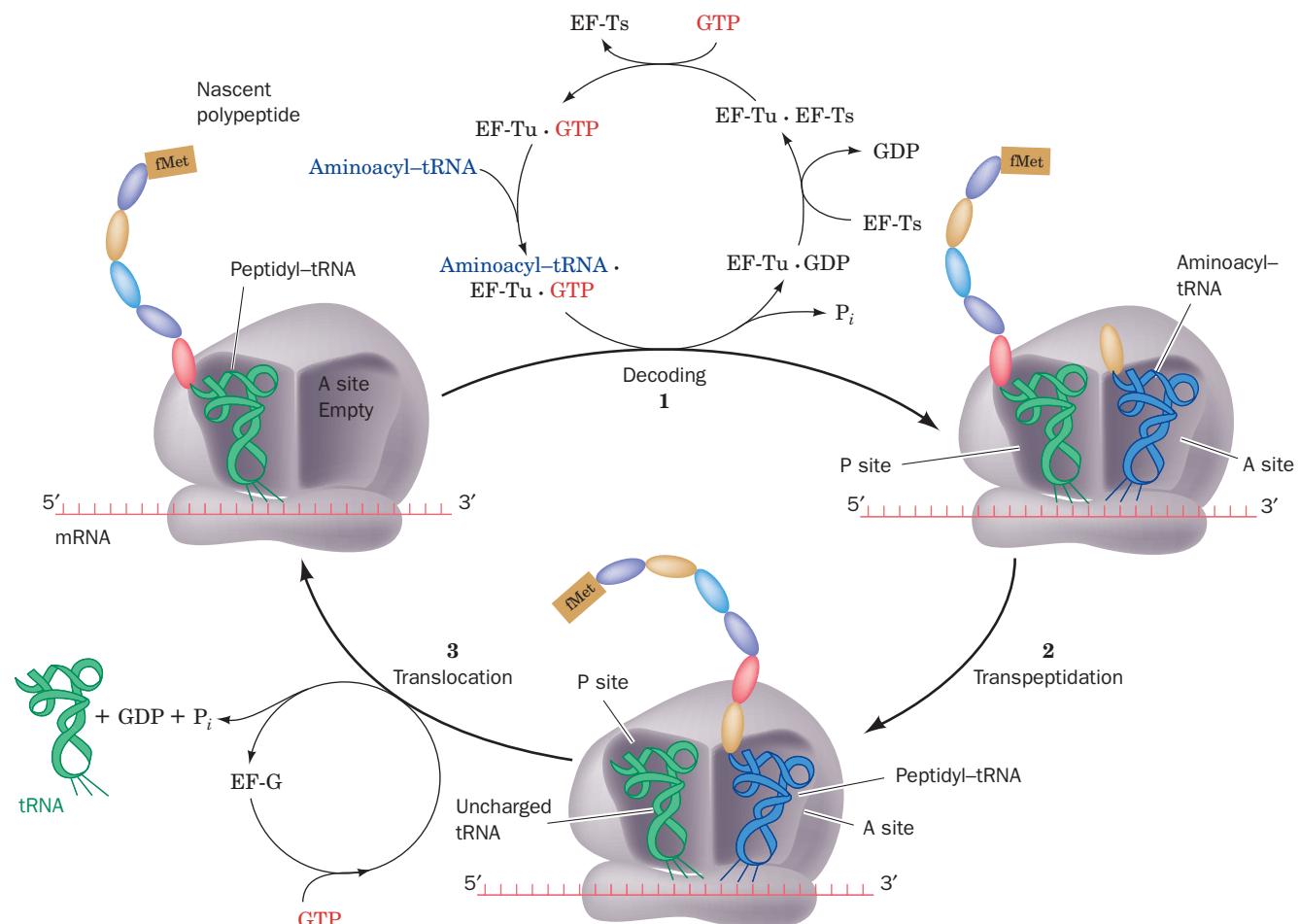


Figure 32-46 Elongation cycle in *E. coli* ribosomes. The E site, to which discharged tRNAs are transferred before being released to solution, is not shown. Eukaryotic elongation follows

a similar cycle but EF-Tu and EF-Ts are replaced by a single multisubunit protein, eEF1, and EF-G is replaced by eEF2.

b. EF-Tu Is Sterically Prevented from Binding Initiator tRNA

The X-ray structure of the Phe-tRNA^{Phe} · EF-Tu · GMPPNP ternary complex (GMPPNP is a nonhydrolyzable GTP analog; Section 19-3Ch), determined by Brian Clark and Jens Nyborg, reveals that these two macromolecules associate to form a corkscrew-shaped complex in which the EF-Tu and the tRNA's acceptor stem form a knoblike handle and the tRNA's anticodon helix forms the screw (Fig. 32-47). The conformation of the tRNA^{Phe} closely resembles that of the uncomplexed molecule (Fig. 32-11b). The EF-Tu folds into three distinct domains that are connected by flexible peptides, rather like beads on a string. The N-terminal domain 1, which binds guanine nucleotides and catalyzes GTP hydrolysis, structurally resembles other known G-proteins.

The two macromolecules associate rather tenuously via three major regions: (1) the CCA—Phe segment at the 3' end of the Phe-tRNA^{Phe} binds in a cleft between EF-Tu's domains 1 and 2 (the blue and green mainly helical domain

and the yellow β sheet domain in Fig. 32-47) that ends in a pocket large enough to accommodate all amino acid residues; (2) the 5'-phosphate of the tRNA binds in a depression at the junction of EF-Tu's three domains; and (3) one side of the tRNA's TΨC stem contacts the exposed main chain and side chains of EF-Tu's C-terminal domain 3 (the orange and red β barrel-containing domain in Fig. 32-47). The tight association of the aminoacyl group with EF-Tu appears to greatly increase the affinity of EF-Tu for the otherwise loosely bound uncharged elongator tRNAs, which explains why EF-Tu does not bind initiator tRNAs.

EF-Tu binds neither formylated aminoacyl-tRNAs nor unformylated Met-tRNA_f^{Met}, which is why the initiator tRNA never reads internal AUG or GUG codons. The first base pair of tRNA_f^{Met} is mismatched (C · A; Fig. 32-40) and hence this initiator tRNA has a 3' overhang of 5 nt vs 4 nt for an elongator tRNA. It seems likely that this mismatch, together with the formyl group, prevents fMet-tRNA_f^{Met} from binding to EF-Tu. Indeed, EF-Tu binds to *E. coli* tRNA_f^{Met} whose 5'-terminal C residue has been deaminated

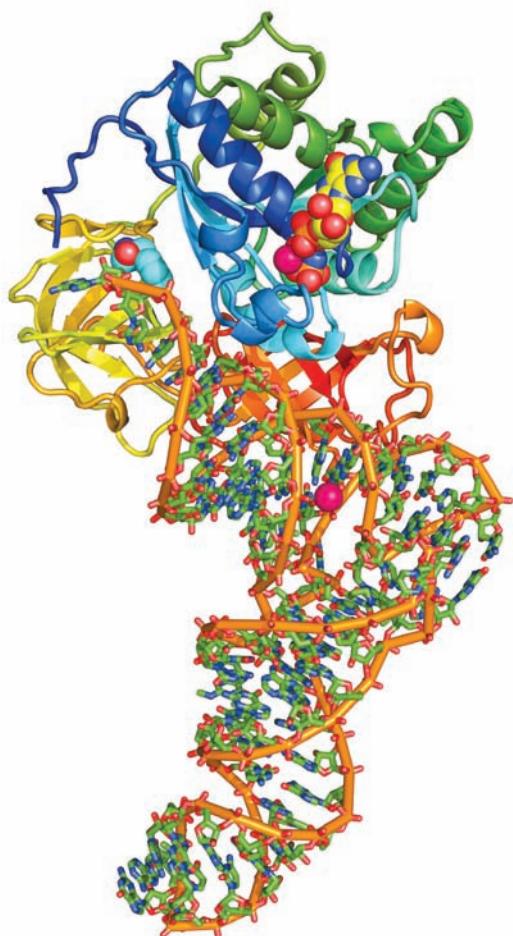


Figure 32-47 X-ray structure of the ternary complex of yeast Phe-tRNA^{Phe}, *Thermus aquaticus* EF-Tu, and GMPPNP. The EF-Tu is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The tRNA is shown in stick form colored according to atom type with C green, N blue, O red, and P orange and with orange rods linking successive P atoms. The tRNA's appended aminoacyl-Phe residue and the GMPPNP that is bound to the EF-Tu are drawn in space-filling form with C atoms cyan and yellow, respectively. Two bound Mg²⁺ ions are represented by magenta spheres. [Based on an X-ray structure by Jens Nyborg, University of Aarhus, Århus, Denmark. PDBid 1TTT.]

by bisulfite treatment (Section 30-7b), which reestablishes the “missing” base pair as U · A. Similarly, Sec-tRNA^{Sec}, which is also not bound by EF-Tu (but rather by SELB; Section 32-2De), has 8 bp in its acceptor stem vs 7 bp in those of other elongator tRNAs. However, initiator tRNAs from several sources have fully base paired acceptor stems, and the U1 · A72 base pair of tRNA^{Gln} is opened up on binding to GlnRS (Section 32-2Cc).

c. EF-Tu Undergoes a Major Conformational Change on Hydrolyzing GTP

Morten Kjeldgaard and Nyborg determined the X-ray structures of *T. aquaticus* EF-Tu (405 residues) in com-

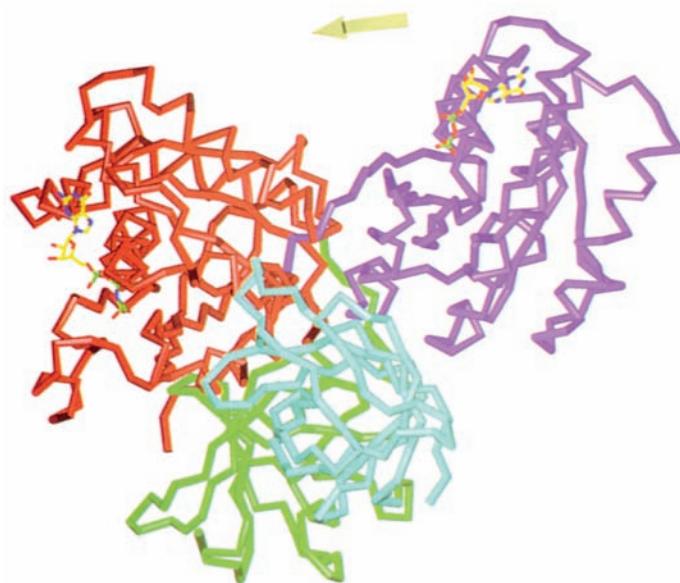


Figure 32-48 Comparison of the X-ray structures of EF-Tu in its complexes with GDP and GMPPNP. The protein is represented by its C_α backbone with domain 1, its GTP-binding domain, purple in the GDP complex and red in the GMPPNP complex. Domain 2 and domain 3, which have the same orientation in both complexes, are green and cyan. The bound GDP and GMPPNP are shown in stick form with C yellow, N blue, O red, and P green. [Courtesy of Morten Kjeldgaard and Jens Nyborg, University of Aarhus, Århus, Denmark. PDBid 1EFT.] See Interactive Exercise 45

plex with GMPPNP and the 70% identical *E. coli* EF-Tu (393 residues) in complex with GDP (Fig. 32-48). The conformation of EF-Tu in its complex with only GMPPNP closely resembles that in its ternary complex with Phe-tRNA^{Phe} and GMPPNP (Fig. 32-47). However, comparison of the GMPPNP and GDP complexes indicates that, on hydrolyzing its bound GTP, EF-Tu undergoes a major structural reorganization. Its greatest local conformational changes occur in the Switch I and Switch II regions of domain 1, which in all G-proteins signal the state of the bound nucleotide to interacting partners (Section 19-2Cb; here domains 2 and 3): Switch I converts from a β hairpin to a short α helix and the α helix of Switch II shifts toward the C-terminus by 4 residues. As a consequence, this latter helix reorients by 42°, which results in domain 1 rigidly changing its orientation with respect to domains 2 and 3 by a dramatic 91° rotation. The tRNA binding site is thereby eliminated.

d. EF-Ts Disrupts the Binding of GDP to EF-Tu

EF-Tu has a 100-fold higher affinity for GDP than GTP. Hence, replacement of the EF-Tu-bound GDP by GTP must be facilitated by the interaction of EF-Tu with EF-Ts (Fig. 32-46, top). The X-ray structure of the EF-Tu · EF-Ts complex, determined by Stephen Cusack and Reuben

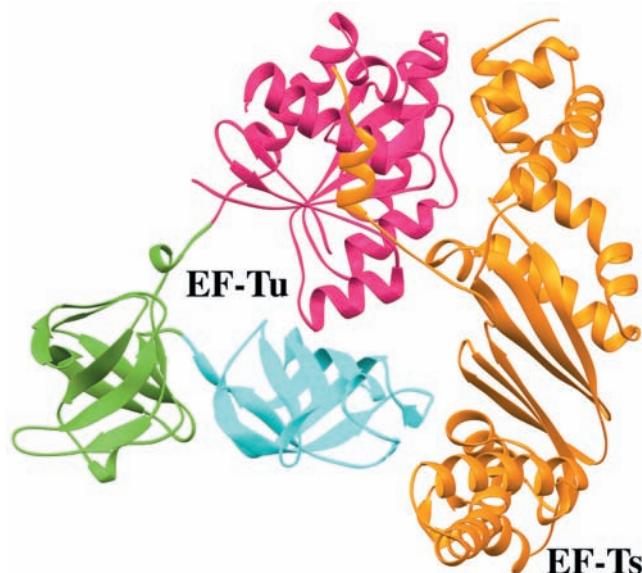


Figure 32-49 X-ray structure of the *E. coli* EF-Tu · EF-Ts complex. Domains 1, 2, and 3 of EF-Tu are magenta, green, and cyan, respectively, and EF-Ts is orange. [Based on an X-ray structure by Stephen Cusack and Reuben Leberman, EMBL, Grenoble Cedex, France. PDBid 1EFU.]

Leberman, reveals that the EF-Tu has a conformation resembling that of its GDP complex (Fig. 32-49) but with its domains 2 and 3 swung away from domain 1 by $\sim 18^\circ$. EF-Ts is an elongated molecule that binds along the right side of EF-Tu as shown in Fig. 32-49, where it contacts EF-Tu's domains 1 and 3. The intrusive interactions of EF-Ts side chains with the GDP binding pocket on EF-Tu domain 1 disrupts the Mg^{2+} ion binding site. This reduces the affinity of EF-Tu for GDP, thereby facilitating its exchange for GTP (after EF-Ts has dissociated), which has a 10-fold higher concentration in the cell than does GDP (the GEF-containing segment of Sos similarly interferes with Mg^{2+} binding and hence guanine nucleotide binding by Ras; Sec-

tion 19-3Cf). EF-Tu's subsequent binding of a charged elongator tRNA increases its affinity for GTP.

e. Transpeptidation

In the transpeptidation stage of the elongation cycle (Fig. 32-46), the peptide bond is formed through the nucleophilic displacement of the P-site tRNA by the amino group of the 3'-linked aminoacyl-tRNA in the A site (Fig. 32-39). The nascent polypeptide chain is thereby lengthened at its C-terminus by one residue and transferred to the A-site tRNA. The reaction occurs without the need of activating cofactors such as ATP because the ester linkage between the nascent polypeptide and the P-site tRNA is a “high-energy” bond. The **peptidyl transferase** center that catalyzes peptide bond formation is located entirely on the large subunit as is demonstrated by the observation that in high concentrations of organic solvents such as ethanol, the large subunit alone catalyzes peptide bond formation. The organic solvent apparently distorts the large subunit in a way that mimics the effect of small subunit binding.

f. Puromycin Is an Aminoacyl-tRNA Analog

The ribosomal elongation cycle was originally characterized through the use of the antibiotic **puromycin** (Fig. 32-50). This product of *Streptomyces alboniger*, which resembles the 3' end of Tyr-tRNA, causes the premature termination of polypeptide chain synthesis. Puromycin, in competition with the mRNA-specified aminoacyl-tRNA but without the need of elongation factors, binds to the ribosomal A site which, in turn, catalyzes a normal transpeptidation reaction to form peptidyl-puromycin. Yet, the ribosome cannot catalyze the transpeptidation reaction in the next elongation cycle because puromycin's “amino acid residue” is linked to its “tRNA” via an amide rather than an ester bond. Polypeptide synthesis is therefore aborted and the peptidyl-puromycin is released.

In the absence of the elongation factor EF-G (see below), an active ribosome cannot bind puromycin because its A site is at least partially occupied by a peptidyl-tRNA. A newly initiated ribosome, however, violates this rule; it

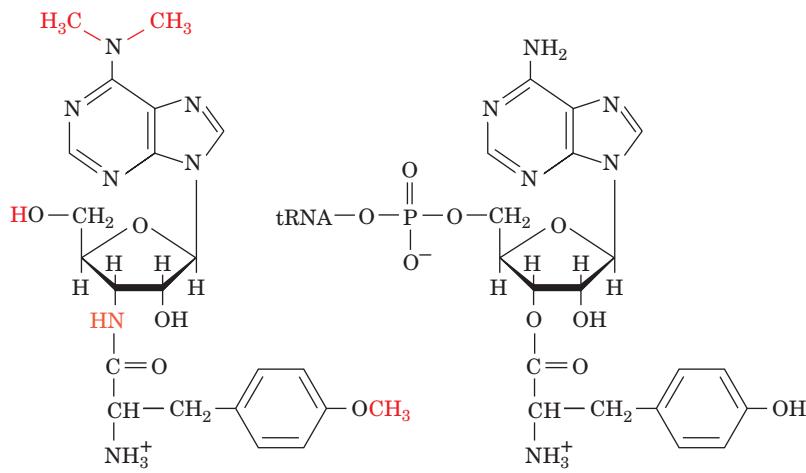


Figure 32-50 Puromycin. This antibiotic (left) resembles the 3'-terminus of tyrosyl-tRNA (right).

catalyzes fMet–puromycin formation. These observations demonstrated the functional existence of the ribosomal P and A sites and established that fMet-tRNA_f^{Met} binds directly to the P site, whereas other aminoacyl-tRNAs must first enter the A site.

g. Transpeptidation: The Ribosome Is a Ribozyme

What is the nature of the peptidyl transferase center, that is, does it consist of RNA, protein, or both? Since all proteins, including those associated with ribosomes, are ribosomally synthesized, the primordial ribosome must have preceded the primordial proteins and hence consisted entirely of RNA. Despite this (in hindsight) obvious evolutionary argument, the idea that rRNA functions catalytically was not seriously entertained until after it had been discovered that RNA can, in fact, act as a catalyst (Section 31-4Ae). Several other observations further indicate that the ribosome is a ribozyme:

1. The absence from the 50S subunit of any one of its proteins but L2, L3, and L4 does not abolish its peptidyl transferase function.
2. rRNAs are more highly conserved throughout evolution than are ribosomal proteins.
3. Most mutations that confer resistance to antibiotics that inhibit protein synthesis occur in genes encoding rRNAs rather than ribosomal proteins.

Nevertheless, the unambiguous demonstration that rRNA functions catalytically in polypeptide synthesis proved to be

surprisingly elusive. Noller succeeded in showing that the *T. thermophilus* large ribosomal subunit from which ~95% of the protein had been removed by treatment with SDS and **proteinase K** followed by phenol extraction (which denatures proteins; Section 6-6A) maintained >80% of its peptidyl transferase activity in a model reaction. Moreover, this activity was abolished by RNase treatment. However, since the remaining protein was due to several intact ribosomal proteins (which are presumably sequestered within the 23S RNA), it could be argued that these proteins are essential for ribosomal catalytic function, a reasonable expectation in light of the >3.5 billion years over which ribosomal proteins and RNAs have coevolved.

Steitz and Moore unequivocally determined the nature of the peptidyl transferase center through its identification in the X-ray structure of the 50S subunit. Peptide bond formation presumably resembles the reverse of peptide bond hydrolysis such as that catalyzed by serine proteases (Section 15-3C). The ribosomal reaction's tetrahedral intermediate (Fig. 32-51a) is mimicked by a compound synthesized by Michael Yarus that consists of the trinucleotide CCdA linked to puromycin via a phosphoramidite group (Fig. 32-51b). This compound, which is named **CCdA-p-Puro**, binds tightly to the ribosome so as to inhibit its peptidyl transferase activity. The X-ray structure of the 50S subunit in complex with CCdA-p-Puro reveals that the inhibitor binds to domain V of the 23S RNA (Fig. 32-27b) at the entrance to the ~100-Å-long polypeptide exit tunnel that runs through to the back of the subunit (Figs. 32-30 and 32-34c). There, the inhibitor is completely enveloped in RNA

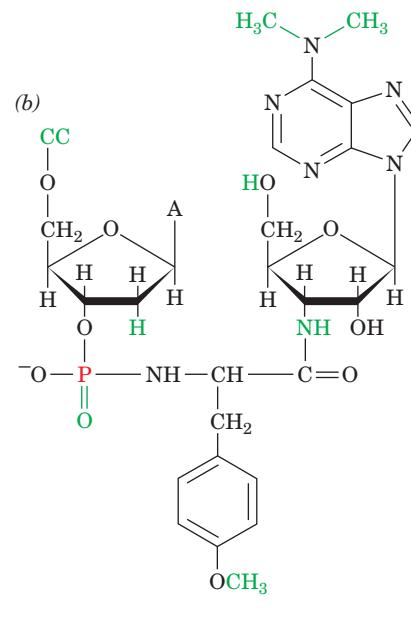
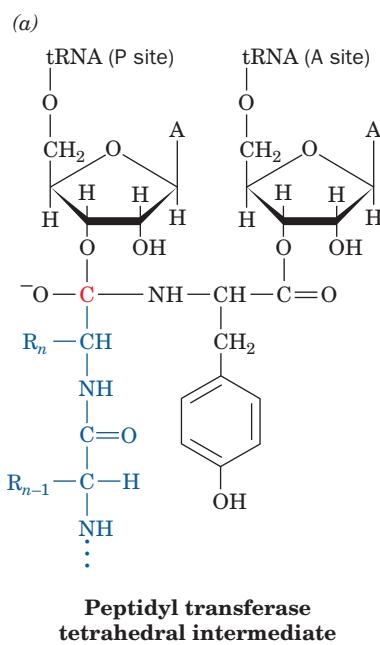


Figure 32-51 Ribosomal tetrahedral intermediate and its analog. (a) The chemical structure of the tetrahedral intermediate (red C) in ribosomally mediated peptide bond formation in which the A-site aminoacyl residue is Tyr. (b) CCdA-p-Puro, the

transition state analog of the tetrahedral intermediate in Part a produced by linking the 3'-OH group of CCdA to the amino group of puromycin's O-methyltyrosine residue via a phosphoryl group.

with no protein side chain approaching closer than $\sim 18 \text{ \AA}$ to the inhibitor's phosphoramidite group and with the nearest Mg^{2+} ion 8.5 \AA away—both too far away to participate in catalysis. Moreover, all the nucleotides that contact the CCdA-p-Puro are $>95\%$ conserved among all three kingdoms of life. Clearly, the ribosomal transpeptidase reaction is catalyzed by RNA.

Despite the foregoing, the X-ray structure of the *T. thermophilus* ribosome in complex with tRNA and mRNA (Fig. 32-34) reveals that the N-terminal tail of L27 interacts with the phosphate group of residue C75 of the A-site tRNA. Moreover, L16 interacts with the elbow region of the A-site tRNA via interactions between conserved Arg residues and phosphate and 2'-OH groups in the elbow region of the tRNA. These observations confirm previous mutational studies indicating that L16 and L27 help stabilize the binding of the A-site tRNA. However, some organisms lack L27, which indicates that its binding to A-site tRNA is not part of an evolutionarily conserved mechanism.

h. Peptide Bond Formation Does Not Occur via Acid–Base Catalysis

The ribosomal peptidyl transferase reaction occurs $\sim 10^7$ -fold faster than the uncatalyzed reaction. How does the ribosome catalyze this reaction? Peptide bond formation is naively expected to proceed via the nucleophilic attack of the amino group on the carbonyl group of an ester to form a tetrahedral intermediate that collapses to an amide and an alcohol (Fig. 32-39). However, in the physiological pH range, the attacking amino group is predominantly in its ammonium form (RNH_3^+), and hence lacks the lone pair necessary to undertake a nucleophilic attack. This suggests that the peptidyl transferase reaction is catalyzed in part by a general base that abstracts a proton from the ammonium group to generate the required free amino group (RNH_2).

Inspection of the peptidyl transferase center in *H. marismortui* reveals that the only basic group within 5 \AA of the inferred position of the attacking amino group is atom N3 of the invariant rRNA base A2486 (A2451 in *E. coli*). It is $\sim 3 \text{ \AA}$ from and hence hydrogen bonded to the attacking amino group (Fig. 32-52). This further suggests that the protonated A2486-N3 electrostatically stabilizes the oxyanion of the tetrahedral reaction intermediate and then donates the proton to the leaving group of the P-site tRNA to yield a 3'-OH group (general acid catalysis). However, in order for A2486-N3 to act as a general base in abstracting the proton from an ammonium group (whose pK is ~ 10), it must have a pK of at least 7 (recall that proton transfers between hydrogen-bonded groups occur at physiologically significant rates only when the pK of the proton donor is no more than 2 or 3 pH units greater than that of the proton acceptor; Section 15-3Dd). Yet, the pK of N3 in AMP is <3.5 . Moreover, several lines of evidence indicate that A2486 does not function as an acid–base catalyst including (1) the model displayed in Fig. 32-52 indicates that the tetrahedral intermediate's oxyanion would point away from and hence could not be stabilized by protonated

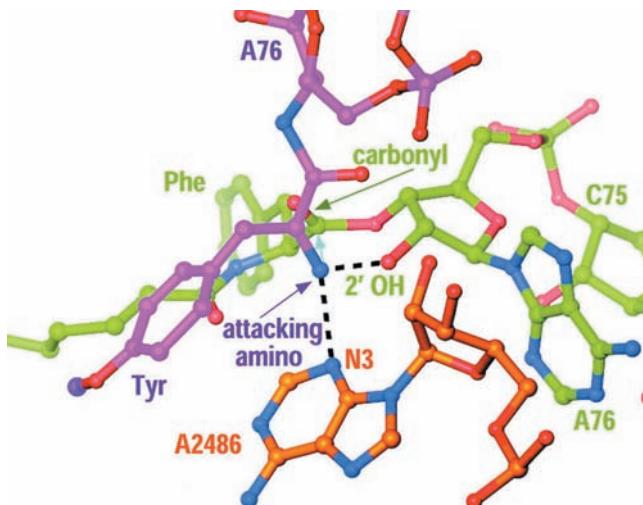


Figure 32-52 Model of the substrate complex of the 50S ribosomal subunit. Atoms are colored according to type with the A-site substrate C and P purple, P-site substrate C and P green, 23S rRNA C and P orange, N blue, and O red. The attacking amino group of the A-site aminoacyl residue is held in position for nucleophilic attack (cyan arrow) on the carbonyl C of the P-site aminoacyl ester through hydrogen bonds (dashed black lines) to A2486-N3 and the 2'-O of the P-site A76. [Courtesy of Peter Moore and Thomas Steitz, Yale University.]

A2486-N3; and (2) mutagenic replacement of A2486 by any other base does not greatly alter the ribosome's catalytic rate.

i. The Ribosome Is an Entropy Trap

If the peptidyl transfer reaction does not involve acid–base catalysis, what is the origin of the ribosome's catalytic power? Marina Rodnina and Richard Wolfenden noted that the uncatalyzed reaction of esters with amines to form amides occurs quite facilely in aqueous solution. They therefore measured the rates of both uncatalyzed peptide bond formation by model compounds and peptidyl transfer by the ribosome at several different temperatures. This provided values of $\Delta\Delta H_{\text{cat}}^\ddagger$ and $\Delta\Delta S_{\text{cat}}^\ddagger$, the reaction's change in the enthalpy and entropy of activation by the ribosome relative to the uncatalyzed reaction. Here $\Delta\Delta H_{\text{cat}}^\ddagger - T \Delta\Delta S_{\text{cat}}^\ddagger = \Delta\Delta G_{\text{cat}}^\ddagger = \Delta G_{\text{cat}}^\ddagger(\text{uncat}) - \Delta G^\ddagger(\text{cat})$, where $\Delta G_{\text{cat}}^\ddagger$ is the change in the reaction's free energy of activation by the ribosome, $\Delta G^\ddagger(\text{cat})$, relative to that of the uncatalyzed reaction, $\Delta G^\ddagger(\text{uncat})$ (Section 14-1C). The measured value of $\Delta\Delta H_{\text{cat}}^\ddagger$ is $-19 \text{ kJ} \cdot \text{mol}^{-1}$, a quantity that would be positive, not negative, if the ribosomal reaction had a significant component of chemical catalysis such as acid–base catalysis and/or the formation of new hydrogen bonds. In contrast, the value of $T \Delta\Delta S_{\text{cat}}^\ddagger$ is $52 \text{ kJ} \cdot \text{mol}^{-1}$, which indicates that the Michaelis complex in the ribosomal reaction is significantly more ordered relative to the transition state than is the uncatalyzed reaction. This value of $T \Delta\Delta S_{\text{cat}}^\ddagger$ largely accounts for the observed $\sim 10^7$ -fold rate enhancement of the ribosomal reaction relative to

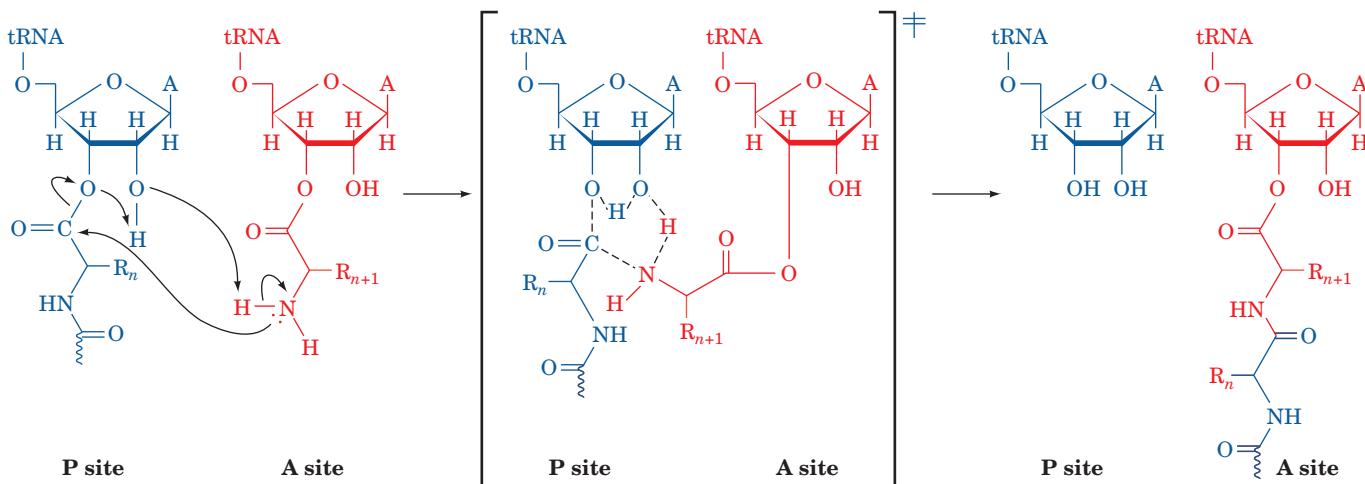


Figure 32-53 The mechanism of ribosome-catalyzed peptidyl transfer. The nucleophilic attack of the α -amino group of the aminoacyl-tRNA (red) on the carbonyl C of the peptidyl-tRNA (blue) occurs in concert with a proton shuttle that involves the O_{3'} and 2'-OH of the P-site A76 together with the α -amino

group of the aminoacyl-tRNA. The reaction proceeds through a transition state (center; enclosed by square brackets) that contains a six-membered ring of partially bonded atoms and which collapses to the reaction products drawn on the right.

the uncatalyzed reaction (the rate enhancement by the ribosome is given by $e^{\Delta\Delta G_{\text{cat}}^*/RT}$; Section 14-1Cd). Evidently, the ribosome enhances the rate of peptide bond formation by properly positioning and orienting its substrates and/or excluding water from the preorganized electrostatic environment of the active site (a form of reactant ordering) rather than by chemical catalysis.

The X-ray structures of the large ribosomal subunit in complex with aminoacyl-tRNA and peptidyl-tRNA indicate that the ribosome uses an induced fit mechanism, as occurs in enzymes such as hexokinase (Section 17-2Aa). Conformational changes in the 23S rRNA, presumably triggered by proper binding of the aminoacyl-tRNA in the A site, orient the ester group of the peptidyl-tRNA for nucleophilic attack. The hydrogen bond between the 2'-OH of the P-site A76 and the attacking amino group (Fig. 32-52) is crucial in doing so. In fact, replacing this 2'-OH group with H or F reduces the reaction rate by at least 10⁶. This suggests that the peptidyl transferase reaction occurs via the substrate-assisted proton shuttle mechanism diagrammed in Fig. 32-53. Although the peptidyl transferase reaction is relatively sluggish compared to the reactions catalyzed by many protein enzymes, it is sufficiently fast to keep up with the other ribosomal processes (which collectively link together 10–20 residues/s). Apparently, the ribosome's peptidyl transferase function is a molecular fossil from the RNA world.

In the absence of a tRNA in the A site, the ester bond linking the peptidyl group to the P-site tRNA is shielded by U2585 of the 23S RNA from nucleophilic attack by water, which would otherwise release the peptidyl group from the ribosome.

j. Translocation: The Ribosome Moves to the Next Codon

In the translocation stage of the elongation cycle, the now uncharged P-site tRNA (at first tRNA_f^{Met} but subsequently an elongator tRNA) is transferred to the E site (not shown in Fig. 32-46), its former occupant having been previously expelled (see below). Simultaneously, the peptidyl-tRNA in the A site, together with its bound mRNA, is moved to the P site. This prepares the ribosome for the next elongation cycle. The maintenance of the peptidyl-tRNA's codon–anticodon association is no longer necessary for amino acid specification. Rather, it acts as a place-keeper that permits the ribosome to precisely step off the three nucleotides along the mRNA required to preserve the reading frame. Indeed, the observation that frameshift suppressor tRNAs induce a four-nucleotide translocation (Section 32-2Ea) indicates that mRNA movement is directly coupled to tRNA movement. An Mg²⁺-stabilized kink in the mRNA between the A and P codons apparently helps prevent slippage.

k. EF-G Structurally Mimics the EF-Ts · tRNA Complex

The translocation process requires the participation of elongation factor **EF-G** (also called **EF2**), which binds to the ribosome together with GTP and is only released on hydrolysis of the GTP to GDP + P_i (Fig. 32-46). EF-G release is a prerequisite for beginning the next elongation cycle because EF-G and EF-Tu bind to the same site of the ribosome and hence their binding is mutually exclusive.

The X-ray structure of *T. thermophilus* EF-G · GMPPNP, determined by Anders Liljas and Derek Logan, reveals a tadpole-shaped monomeric protein that consists of five

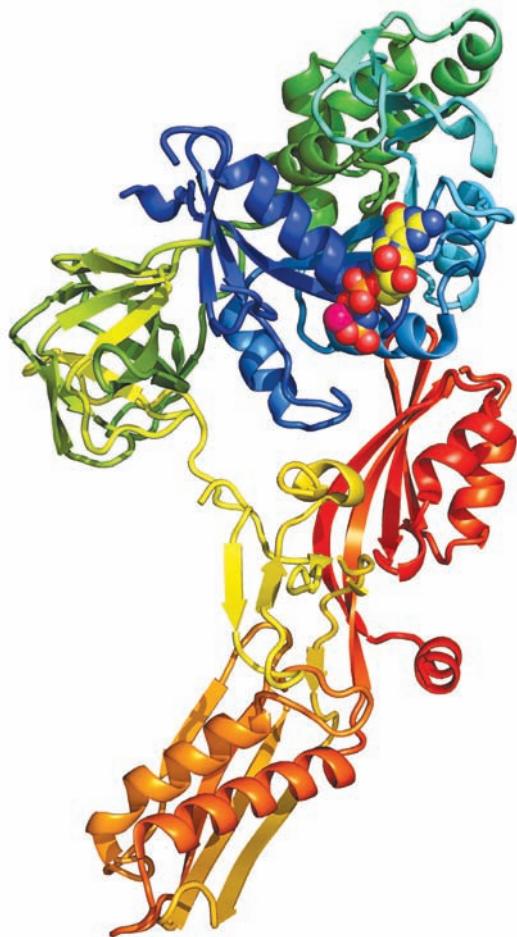


Figure 32-54 X-ray structure of EF-G from *T. thermophilus* in complex with GMPPNP. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The GMPPNP is drawn in space-filling form colored according to atom type with C yellow, N blue, O red, and P orange. An Mg^{2+} ion that is bound to the GMPPNP is represented by a magenta sphere. Portions of the structure are not visible. Note the remarkable resemblance in shape between this structure and that of Phe-tRNA^{Phe} · EF-Tu · GMPPNP (Fig. 32-47). [Based on an X-ray structure by Anders Liljas and Derek Logan, Lund University, Lund, Sweden. PDBid 2BV3.]

domains (Fig. 32-54). The first two domains closely resemble those in EF-Tu · GMPPNP rather than those in EF-Tu · GDP (Fig. 32-48). This, it is argued, is because the two elongation factors have reciprocal functions with EF-Tu · GTP facilitating the conversion of the ribosome from its post- to its pre-translocational state and EF-G · GTP promoting the reverse transition. This idea is supported by the intriguing observation that the Phe-tRNA^{Phe} · EF-Tu · GMPPNP and EF-G · GMPPNP complexes are almost identical in appearance: EF-G's three C-terminal domains (yellow through red in Fig. 32-54), which have no counterparts in EF-Tu, closely resemble the EF-Tu-bound tRNA in shape, a remarkable case of **macromolecular mimicry**. Indeed, EF-G occupies the same ribosomal site as does EF-Tu (Fig. 32-55).

EF-G is unusual among G-proteins in that it has no corresponding GEF. However, its N-terminal guanine nu-

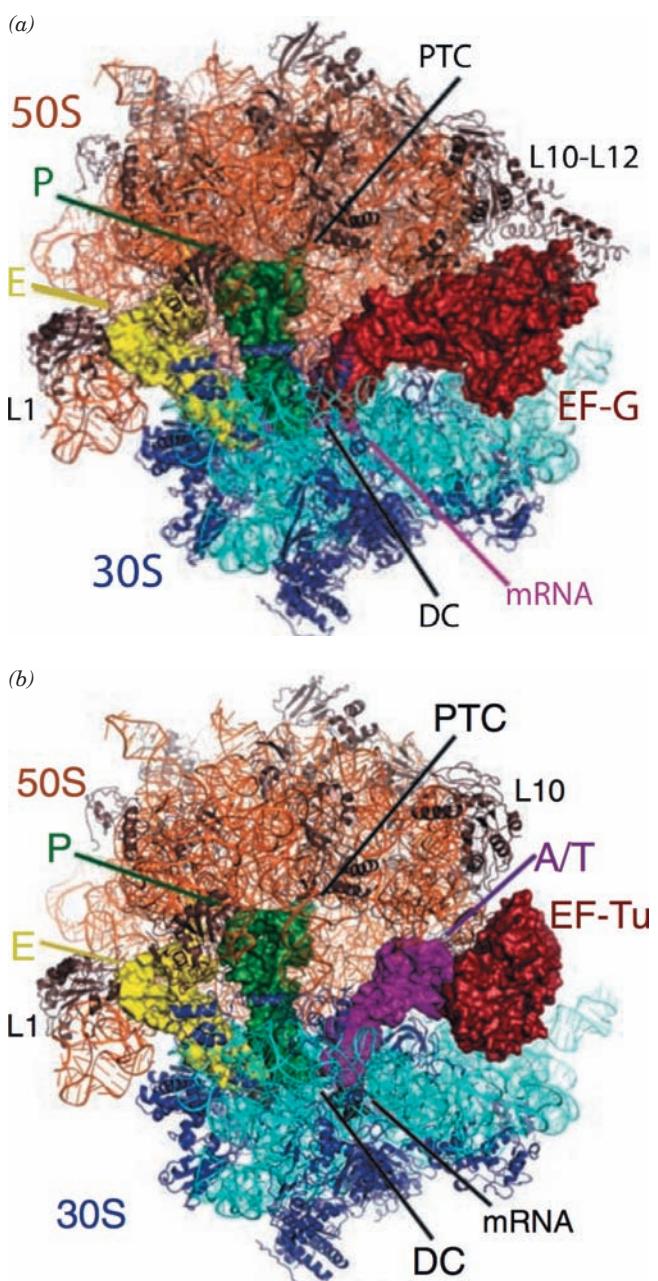


Figure 32-55 X-ray structure of the *T. thermophilus* ribosome in its complexes with (a) EF-G and (b) EF-Tu. The ribosome is drawn in ribbon form with its 23S RNA orange, its 50S subunit proteins brown, its 16S RNA cyan, and its 30S proteins blue. Its bound tRNAs, mRNA, EF-G, and EF-Tu are represented by their surface diagrams with E-site tRNA yellow, P-site tRNA green, A-site tRNA magenta, mRNA black, and both EF-G and EF-Tu red. In Part b, the A-site tRNA bound to EF-Tu is in the A/T conformation (see below). The positions of the decoding center (DC) and the peptidyl transferase center (PTC) are indicated. [Courtesy of Martin Schmeing and Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2WRI, 2WRJ, 2WRN, and 2WRO.]

cleotide-binding domain contains a unique α helical insert (green in Fig. 32-54) that contacts the domain's conserved core at sites analogous to those in EF-Tu that interact with EF-Ts. This suggests that this subdomain acts as an internal GEF.

I. Translocation Occurs via Intermediate States

Chemical footprinting studies (Section 31-2Ab) by Noller revealed that certain bases in the 16S rRNA are protected by tRNAs bound in the ribosomal A and P sites and that certain bases in the 23S rRNA are protected by tRNAs in the A, P, and E sites. Almost all of these protected bases are absolutely conserved in evolution and many of them have been implicated in ribosomal function through biochemical or genetic studies.

Variations in chemical footprinting patterns during the elongation cycle together with the more recently determined X-ray and cryo-EM structures indicate that the translocation of tRNA occurs in several discrete steps (Fig. 32-56):

1. Let us start with the ribosome in its **post-translocational state**: a deacylated tRNA bound to the E subsites of both the 30S and 50S subunits (the E/E binding state), a peptidyl-tRNA bound in the P subsites of both subunits

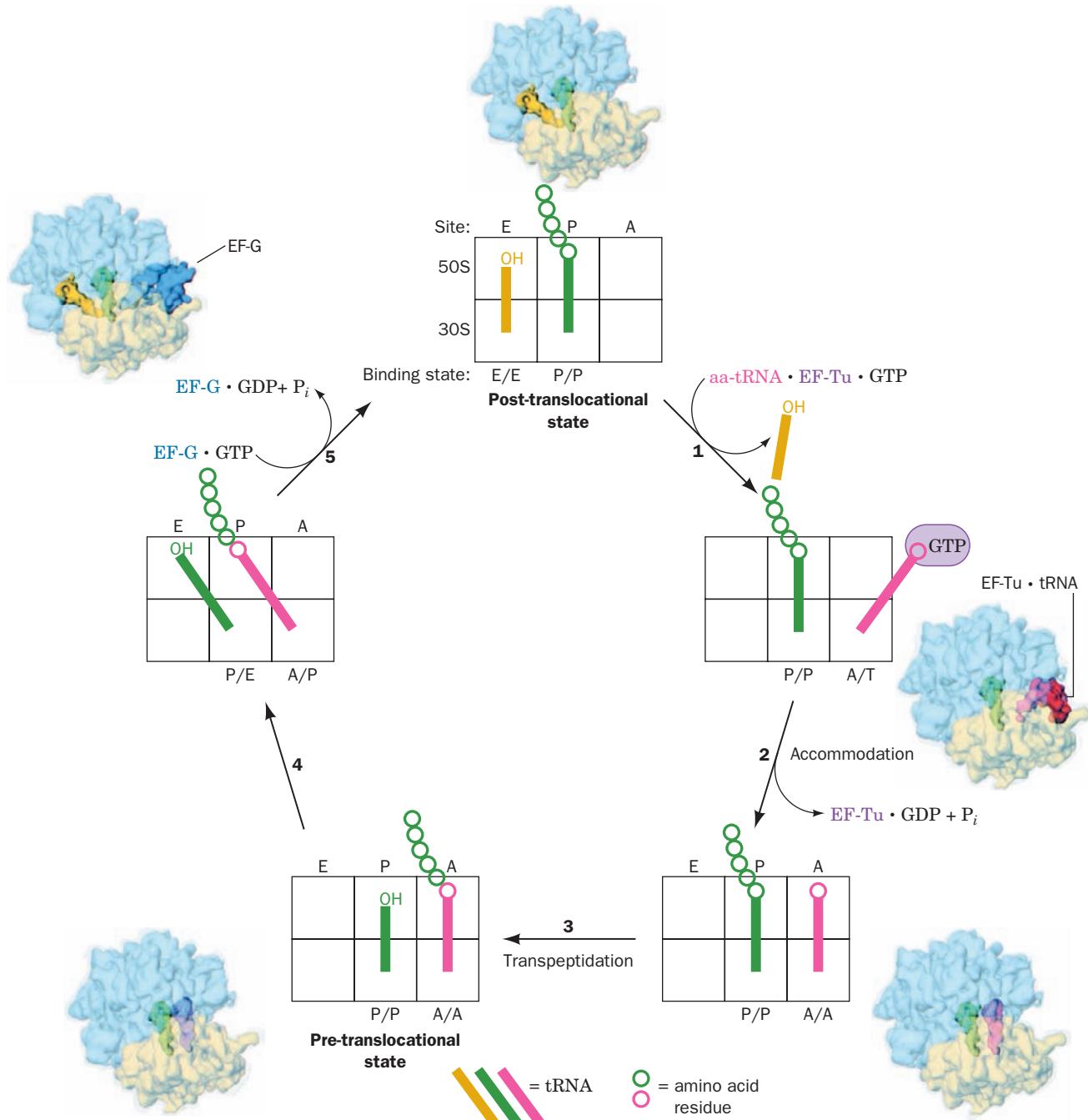


Figure 32-56 Ribosomal binding states in the elongation cycle.

Note how this scheme elaborates the classic elongation cycle diagrammed in Fig. 32-46. The drawings are accompanied by 17-Å-resolution cryo-EM-based images of the *E. coli* 70S ribosome in the corresponding binding states in which the 30S subunit is transparent yellow, the 50S subunit is transparent blue,

and the tRNAs and elongation factors are colored as in the drawing they accompany. [Cryo-EM images courtesy of Knud Nierhaus, Max-Planck-Institut für Molekulare Genetik, Berlin, Germany, and Joachim Frank, Wadsworth Center, State University of New York at Albany.]

(the P/P state), and the A site empty. An aminoacyl-tRNA (aa-tRNA) in ternary complex with EF-Tu and GTP binds to the A site accompanied by the release of the E-site tRNA (but see below). This yields a complex in which the incoming aa-tRNA is bound in the 30S subunit's A subsite via a codon–anticodon interaction (recall that the mRNA is bound to the 30S subunit) but with the EF-Tu preventing the entry of the tRNA's aminoacyl end into the 50S subunit's A subsite, an arrangement termed the A/T state (T for EF-Tu).

2. EF-Tu hydrolyzes its bound GTP to GDP + P_i and is released from the ribosome. This permits the aa-tRNA to fully bind to the A site (the A/A state), a process called **accommodation** in which the 3' end of the tRNA swings around by nearly 70 Å.

3. The peptidyl transferase reaction occurs, yielding the **pre-translocational state**.

4. The acceptor end of the new peptidyl-tRNA shifts from the A subsite of the 50S subunit to its P-subsite, while the tRNA's anticodon end remains associated with the A subsite of the 30S subunit (yielding the A/P hybrid binding state). The acceptor end of the newly deacylated tRNA simultaneously moves from the P subsite to the E subsite of the 50S subunit while its anticodon end remains associated with the P subsite of the 30S subunit (the P/E state).

5. The ribosomal binding of the EF-G · GTP complex and the subsequent GTP hydrolysis impel the anticodon ends of these tRNAs, together with their bound mRNA, to move relative to the small ribosomal subunit such that the peptidyl-tRNA assumes the P/P state and the deacylated tRNA assumes the E/E state (the post-translocational state), thereby completing the elongation cycle.

The binding of tRNA to the A and E sites, as Nierhaus has shown, exhibits negative allosteric cooperativity. In the pre-translocational state, the E site binds the newly deacylated tRNA with high affinity (the E site is sterically unable to bind an aminoacyl-tRNA), whereas the empty A site has low affinity for aminoacyl-tRNA. However, in the post-translocational state, the ribosome has undergone a conformational change that converts the A site to a high-affinity state and the E site to a low-affinity state, which consequently releases the deacylated tRNA when aa-tRNA · EF-Tu · GTP binds to the A site. Thus, the E site is not simply a passive holding site for spent tRNAs but performs an essential function in the translation process. The GTP hydrolysis by the elongation factors EF-Tu and EF-G as well as the peptidyl transferase reaction apparently function to reduce the activation barriers between these conformational states. The unidirectional A → P → E flow of tRNAs through the ribosome is thereby facilitated.

Certain aspects of the foregoing mechanism are not fully resolved. For example, X-ray studies of the 70S ribosome in complex with three tRNAs (e.g., Fig. 32-34) suggest that the E-site tRNA is not released from the ribosome until Step 2 of Fig. 32-56. However, Nierhaus and Frank argue that these complexes were crystallized in the presence of unphysiologically high tRNA concentrations.

Whatever the case, it is clear that the changes in binding states result in large-scale tRNA movements, in some instances >50 Å. Moreover, cryo-EM studies indicate that on binding EF-G · GDP(CH₂)P (like GMPPNP but with a CH₂ group rather than an NH group bridging its β and γ phosphates), the 30S subunit rotates with respect to the 50S subunit by 6° clockwise when viewed from the 30S subunit's solvent side, which results in a maximum displacement of ~19 Å at the periphery of the ribosome. This rotation is accompanied by many smaller conformational changes in both subunits, particularly in the regions about the entrance and exit to the mRNA channel. Clearly, we are far from fully understanding how the ribosome works at the molecular level.

m. The Eukaryotic Elongation Cycle Resembles That of Prokaryotes

The eukaryotic elongation cycle closely resembles that of prokaryotes. In eukaryotes, the functions of EF-Tu and EF-Ts are respectively assumed by the eukaryotic elongation factors **eEF1A** and **eEF1B**, with yeast eEF1B consisting of two subunits: **eEF1B α** , which catalyzes nucleotide exchange, and **eEF1B γ** , which has unknown function (in higher eukaryotes, eEF1B contains a third subunit, **eEF1B β** , that possesses a nucleotide exchange activity similar to that of eEF1B α). Likewise, **eEF2** functions in a manner analogous to EF-G. However, the corresponding eukaryotic and prokaryotic elongation factors are not interchangeable.

The X-ray structure of yeast eEF1A · eEF1B α (Fig. 32-57), determined by Kjeldgaard and Nyborg, reveals that eEF1A structurally resembles the homologous EF-Tu (Fig. 32-49), whereas eEF1B α exhibits no resemblance to EF-Ts, either in sequence or in structure. Nevertheless, eEF1B α functionally interacts with eEF1A much as EF-Ts interact with EF-Tu: Both GEFs associate with their corresponding G-protein so as to disrupt the Mg²⁺ binding site associated with its bound guanine nucleotide.

E. Translational Accuracy

The genetic code is normally expressed with remarkable fidelity. We have already seen that transcription and tRNA aminoacylation both proceed with high accuracy (Sections 31-2Ec and 32-2Ce). The accuracy of ribosomal mRNA decoding was estimated from the rate of misincorporation of [³⁵S]Cys into highly purified **flagellin**, an *E. coli* protein (Section 35-3I) that normally lacks Cys. These measurements indicated that the mistranslation rate is ~10⁻⁴ errors per codon. This rate is greatly increased in the presence of **streptomycin**, an antibiotic that increases the rate of ribosomal misreading (Section 32-3Ga). From the types of reading errors that streptomycin is known to induce, it was deduced that the mistranslation arose almost entirely from the confusion of the Arg codons CGU and CGC for the Cys codons UGU and UGC. The above error rate is therefore largely caused by mistakes in ribosomal decoding.

An aminoacyl-tRNA is selected by the ribosome only according to its anticodon. Yet the binding energy loss

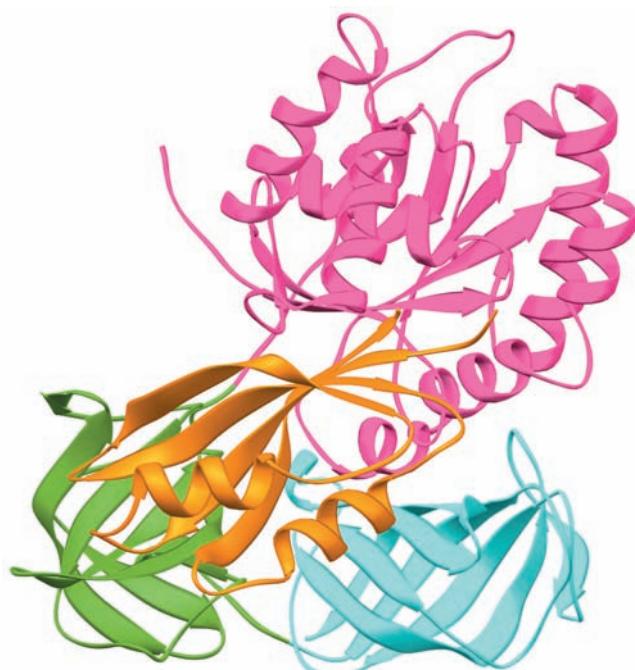


Figure 32-57 X-ray structure of yeast eEF1A · eEF1B α . Domains 1, 2, and 3 of eEF1A are magenta, green, and cyan, respectively, and eEF1B α is orange. The complex is oriented so as to emphasize the structural resemblance between eEF1A and the similarly colored EF-Tu in its complex with EF-Ts (Fig. 32-49). Note the lack of resemblance between eEF1B α and EF-Ts. [Based on an X-ray structure by Morten Kjeldgaard and Jens Nyborg, University of Aarhus, Århus, Denmark. PDBid 1F60.]

arising from a single base mismatch in a codon–anticodon interaction is estimated to be $\sim 12 \text{ kJ} \cdot \text{mol}^{-1}$, which, according to Eq. [32.1], cannot account for a ribosomal decoding accuracy of less than $\sim 10^{-2}$ errors per codon. Moreover, the base pairing interaction between the UUU codon for Phe and the GAA anticodon of tRNA^{Phe} would be naively expected to be less stable than the incorrect pairing be-

tween the UGC codon for Ser and the GCG anticodon of tRNA^{Arg}. This is because both interactions have one G · U base pair and the former correct interaction's remaining two A · U base pairs are weaker than the latter incorrect interaction's remaining two G · C base pairs. Evidently, the ribosome has some sort of proofreading mechanism that increases its overall decoding accuracy.

a. The Ribosome Monitors the Formation of a Correct Codon–Anticodon Complex

As we have seen (Figs. 32-55b and 32-56), the aminoacyl-tRNA · EF-Tu · GTP ternary complex initially binds to the ribosome with the tRNA in the A/T binding state. The tRNA only assumes the fully bound A/A state (accommodation) after the GTP has been hydrolyzed and the EF-Tu · GDP complex has been released from the ribosome. These two states presumably permit the ribosome to double-check (proofread) the codon–anticodon complex that the mRNA makes with the incoming tRNA.

The X-ray structure of the *T. thermophilus* 30S subunit in complex with a U₆ hexanucleotide mRNA and a 17-nt RNA consisting of the tRNA^{Phe} anticodon stem-loop (Fig. 32-11, although its nucleotides are unmodified), determined by Ramakrishnan, revealed how an mRNA-specified tRNA initially binds to the ribosome. The codon–anticodon association is stabilized by its interactions with three universally conserved ribosomal bases, A1492, A1493, and G530 (Fig. 32-58):

1. The first codon–anticodon base pair, that between mRNA U1 and tRNA A36, is stabilized by the binding of the rRNA A1493 base in the base pair's minor groove (Fig. 32-58a).

2. The second codon–anticodon base pair, that between U2 and A35, is bolstered by A1492 and G530, which both bind in this base pair's minor groove (Fig. 32-58b).

3. The third codon–anticodon base pair (the wobble pair; Section 32-2Db), that between U3 and G34, is reinforced through minor groove binding by G530 (Fig. 32-58c).

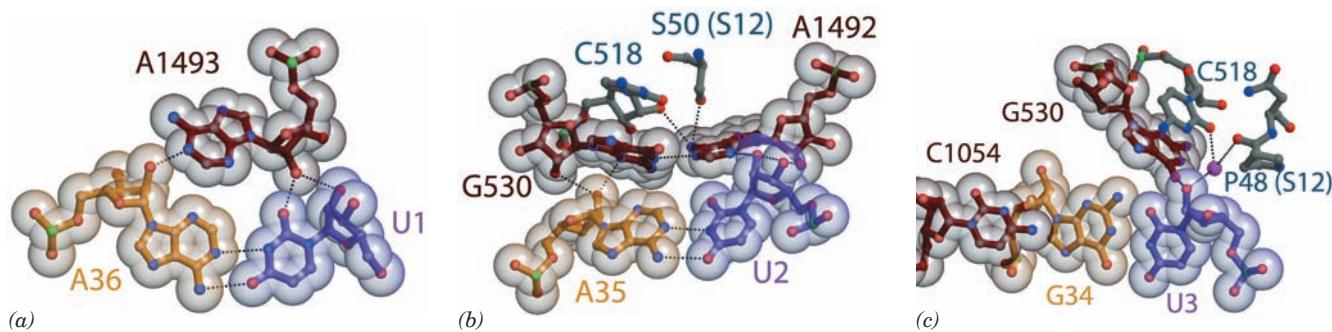


Figure 32-58 Codon–anticodon interactions in the ribosome.

The (a) first, (b) second, and (c) third codon–anticodon base pairs as seen in the X-ray structure of the *T. thermophilus* 30S subunit in complex with U₆ (a model mRNA) and the 17-nt anticodon stem-loop of tRNA^{Phe} (whose anticodon is GAA). The structures are drawn in ball-and-stick form embedded in

their semitransparent van der Waals surfaces. Codons are purple, anticodons are yellow, and rRNA is brown or gray with non-C atoms colored according to type (N blue, O red, and P green). Protein C atoms are gray and Mg²⁺ ions are represented by magenta spheres. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1IBM.]

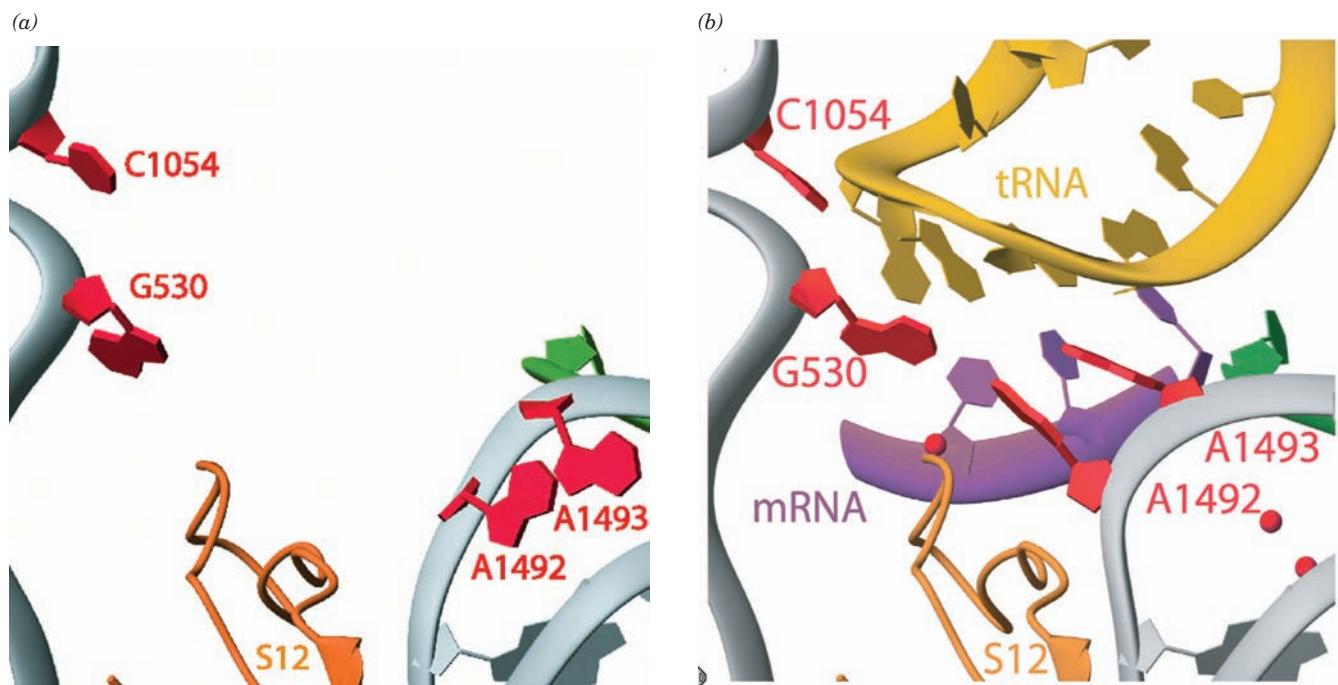


Figure 32-59 Ribosomal decoding site. The X-ray structures of *T. thermophilus* 30S subunit (a) alone and (b) in its complex with U₆ and the 17-nt anticodon stem-loop of tRNA^{Phe}. The RNAs are drawn as ribbons with their nucleotides in paddle form with tRNA gold, A-site mRNA purple, rRNA gray, and

nucleotides that undergo conformational changes red. Protein S12 is tan and Mg²⁺ ions are represented by red spheres. Compare Part b with Fig. 32-58. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids (a) 1FJF and (b) 1IBM.]

This latter interaction appears to be less stringent than those in the first and second codon–anticodon positions, which is consistent with the need for the third codon–anticodon pairing to tolerate non-Watson–Crick base pairs (Section 32-2D).

Comparison of this structure with that of the 30S subunit alone reveals that the foregoing rRNA nucleotides undergo conformational changes on the formation of a codon–anticodon complex (Fig. 32-59). In the absence of tRNA, the bases of A1492 and A1493 stack in the interior of an RNA loop but flip out of this loop to form the codon–anticodon complex, whereas the G530 base switches from the syn to the anti conformation (Section 29-2Aa). These interactions enable the ribosome to monitor whether an incoming tRNA is cognate to the codon in the A site; a non-Watson–Crick base pair could not bind these ribosomal bases in the same way. Indeed, any mutation of A1492 or A1493 is lethal because pyrimidines in these positions could not reach far enough to interact with the codon–anticodon complex or G530 and because a G in either position would be unable to form the required hydrogen bonds and its N2 would be subjected to steric collisions. An incorrect codon–anticodon provides insufficient free energy to bind the tRNA to the ribosome and it therefore dissociates from it, still in its ternary complex with EF-Tu and GTP.

b. GTP Hydrolysis by EF-Tu Is a Thermodynamic Prerequisite to Ribosomal Proofreading

A proofreading step must be entirely independent of the initial selection step. Only then can the overall probability of error be equal to the product of the probabilities of error of the individual selection steps. We have seen that DNA polymerases and aminoacyl-tRNA synthetases maintain the independence of their two selection steps by carrying them out at separate active sites (Sections 30-2Ac and 32-2Ce). Yet the ribosome only recognizes the incoming aminoacyl-tRNA according to its anticodon’s complementarity to the codon in the A site. Consequently, the ribosome must somehow examine this codon–anticodon interaction in two separate ways.

The formation of a correct codon–anticodon complex induces EF-Tu to hydrolyze its bound GTP, although how this occurs is unclear (note that EF-Tu’s GTPase domain is bound in the 50S subunit which, together with the observation that GTP hydrolysis requires an intact tRNA, suggests that the hydrolysis signal is at least in part transmitted through the tRNA). The resulting conformational change in EF-Tu (Fig. 32-48) swings its bound tRNA into the A/A state (accommodation), a process that moves the 3' end of the tRNA by nearly 70 Å. This, it is hypothesized, subjects the codon–anticodon interaction to a strain that only a correct pairing can withstand. The codon–anticodon interaction is thereby subjected to a second screening that only

permits a cognate aminoacyl-tRNA to enter the peptidyl transferase center. The irreversible GTPase reaction must precede this proofreading step because otherwise the dissociation of a noncognate tRNA (the release of its anticodon from the codon) would simply be the reverse of the initial binding step, that is, it would be part of the initial selection step rather than proofreading. *GTP hydrolysis therefore provides the second context necessary for proofreading; it is the entropic price the system must pay for accurate tRNA selection.*

F. Chain Termination

Polypeptide synthesis under the direction of synthetic mRNAs such as poly(U) terminates with a peptidyl-tRNA in association with the ribosome. However, *the translation of natural mRNAs, which contain the Stop codons UAA, UGA, or UAG, results in the release of free polypeptides.*

Accurate termination is essential, not only because it prevents the wasteful synthesis of nonfunctional polypeptides, but also because prematurely terminated polypeptides may be toxic.

a. Prokaryotic Termination

In *E. coli*, chain termination has several stages (Fig. 32-60):

1. The termination codons, the only codons that normally have no corresponding tRNAs, are recognized by class I **release factors** (Table 32-9): **RF-1** recognizes UAA and UAG, whereas the 39% identical **RF-2** recognizes UAA and UGA. Swapping a conserved PXT tripeptide in RF-1 with a conserved SPF tripeptide in RF-2 interchanges their Stop codon specificities, which suggests that these tripeptides mimic anticodons.

2. *On binding to their corresponding Stop codon, RF-1 and RF-2 induce the transfer of the peptidyl group from*

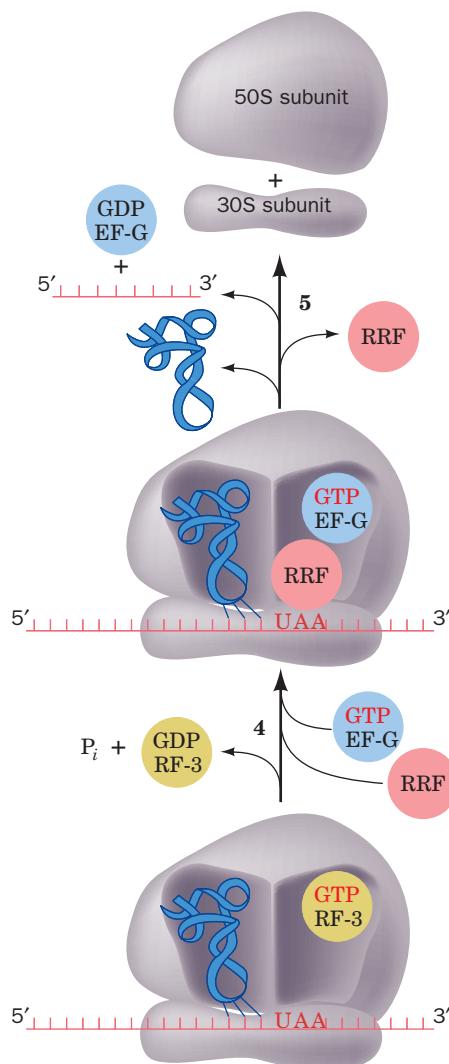
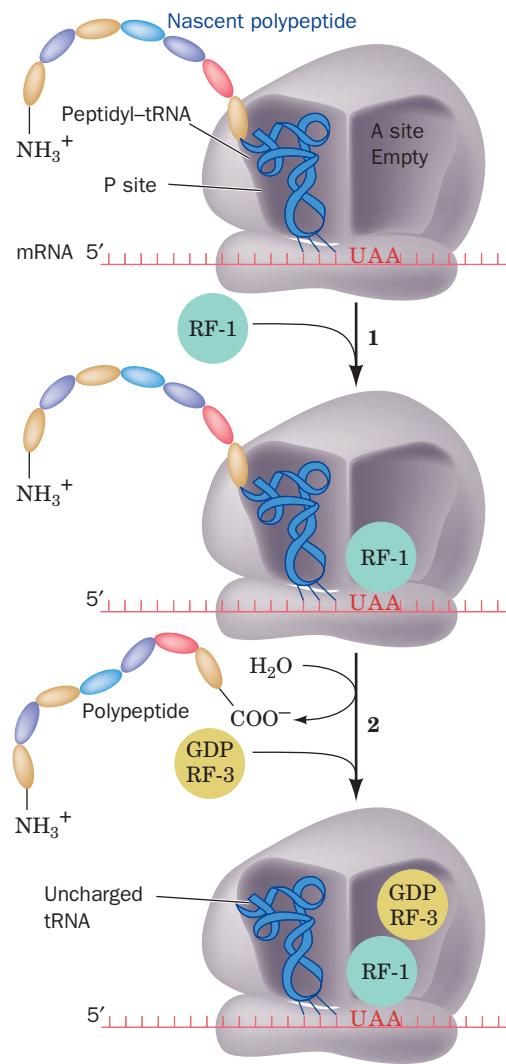


Figure 32-60 Termination pathway in *E. coli* ribosomes. RF-1 recognizes the Stop codons UAA and UAG, whereas RF-2 (not shown) recognizes UAA and UGA. Eukaryotic termination

follows an analogous pathway but requires only a single class I release factor, eRF1, that recognizes all three Stop codons.

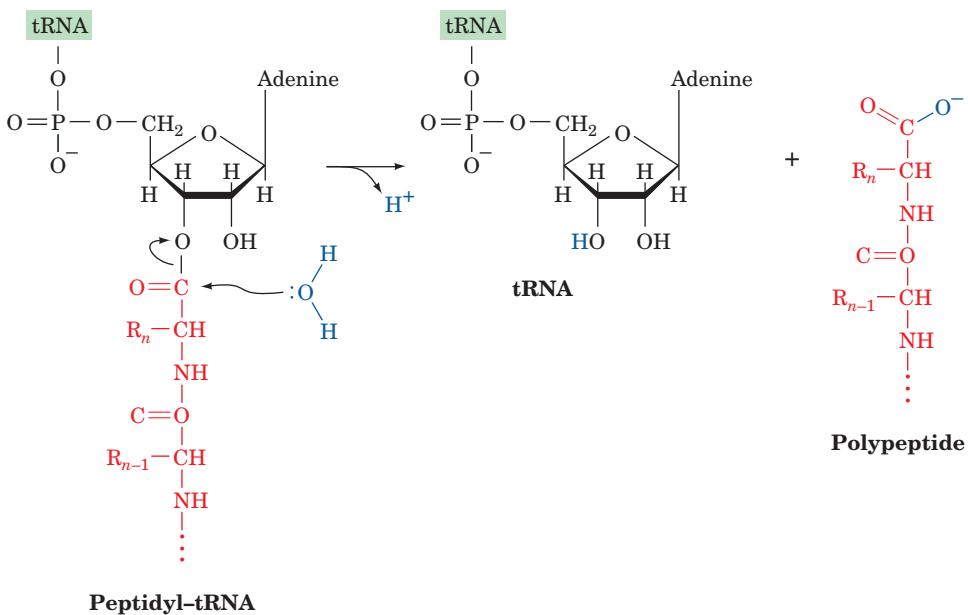


Figure 32-61 Ribosome-catalyzed hydrolysis of peptidyl-tRNA to form a polypeptide and free tRNA.

tRNA to water rather than to an aminoacyl-tRNA, thereby releasing the completed polypeptide (Fig. 32-61). This occurs with an error rate of 10^{-5} without proofreading. The class I release factors act at the ribosomal A site as is indicated by the observations that they compete with suppressor tRNAs for termination codons and that they cannot bind to the ribosome simultaneously with EF-G. A GGG tripeptide that is universally conserved in all class I release factors is implicated in catalyzing the hydrolysis of the peptidyl-tRNA ester linkage (see below).

3. Once the newly synthesized polypeptide has been released from the ribosome, the class II release factor **RF-3**, in its complex with GDP, binds to the ribosome at the same site as do EF-Tu and EF-G. In fact, the X-ray structure of RF-3 · GDP resembles that of EF-Tu · GMPPNP (Fig. 32-47). Free RF-3 has a greater affinity for GDP than GTP but on binding to the ribosome–RF-1/2 complex, it exchanges its bound GDP for GTP. The resulting change in the conformation of RF-3, as seen in cryo-EM studies, causes it to bind more tightly to the ribosome and expel the RF-1/2. RF-3 is not required for cell viability although it is necessary for maximum growth rate; RF-3 only accelerates the dissociation of RF-1/2 from the ribosome by ~5-fold.

4. The interaction of RF-3 · GTP with the ribosome stimulates it to hydrolyze its bound GTP, much as occurs with EF-Tu · GTP and EF-Tu · GTP. The resulting RF-3 · GDP then dissociates from the ribosome. Subsequently, **ribosomal recycling factor (RRF)** binds in the ribosomal A site followed by EF-G · GTP. RRF, which was discovered by Akira Kaji, is essential for cell viability.

5. EF-G hydrolyzes its bound GTP, which causes RRF to be translocated to the P site and the tRNAs previously in the P and E sites (the latter not shown in Fig. 32-60) to be released. Finally, the small and large ribosomal subunits

separate, a process that is facilitated by the binding of IF-3 (Section 32-3Cc), and RRF, EF-G · GDP, and mRNA are released. The ribosomal subunits can then participate in a new round of initiation (Fig. 32-43).

b. Eukaryotic Termination

Chain termination in eukaryotes resembles that in prokaryotes, but it has only one class I release factor, **eRF1**, that recognizes all three Stop codons. It is unrelated in sequence to RF-1 and RF-2. However, the eukaryotic class II release factor, **eRF3**, resembles RF-3 in both sequence and function. Nevertheless, eRF3 is essential for eukaryotic cell viability.

c. The Ribosome Binds RF-1 and RF-2 in a Conformation That Catalyzes Peptide Release

The X-ray structures of the *T. thermophilus* ribosome with RF-1, an mRNA containing a UAA Stop codon, and deacylated tRNAs in its P and E sites was determined by Noller (Fig. 32-62a), and the closely similar structures containing the tRNAs, RF-2, and an mRNA with a UAA or a UGA Stop codon, were respectively determined by Noller and Ramakrishnan. These are all product complexes since they lack peptidyl groups on their P-site tRNAs. The structurally similar RF-1 (Fig. 32-62b) and RF-2 each consist of four domains with domains 2 and 4 occupying the ribosome's decoding center (DC), where they contact the mRNA's Stop codon, and with domain 3 occupying the peptidyl transferase center (PTC), where it interacts with the ribose residue of the P-site tRNA's A76 to which a peptidyl group would be linked in a substrate complex. The deletion of domain 1 does not affect peptide release activity but is required for the RF-3-facilitated dissociation of RF-1/2 from the ribosome (see below).

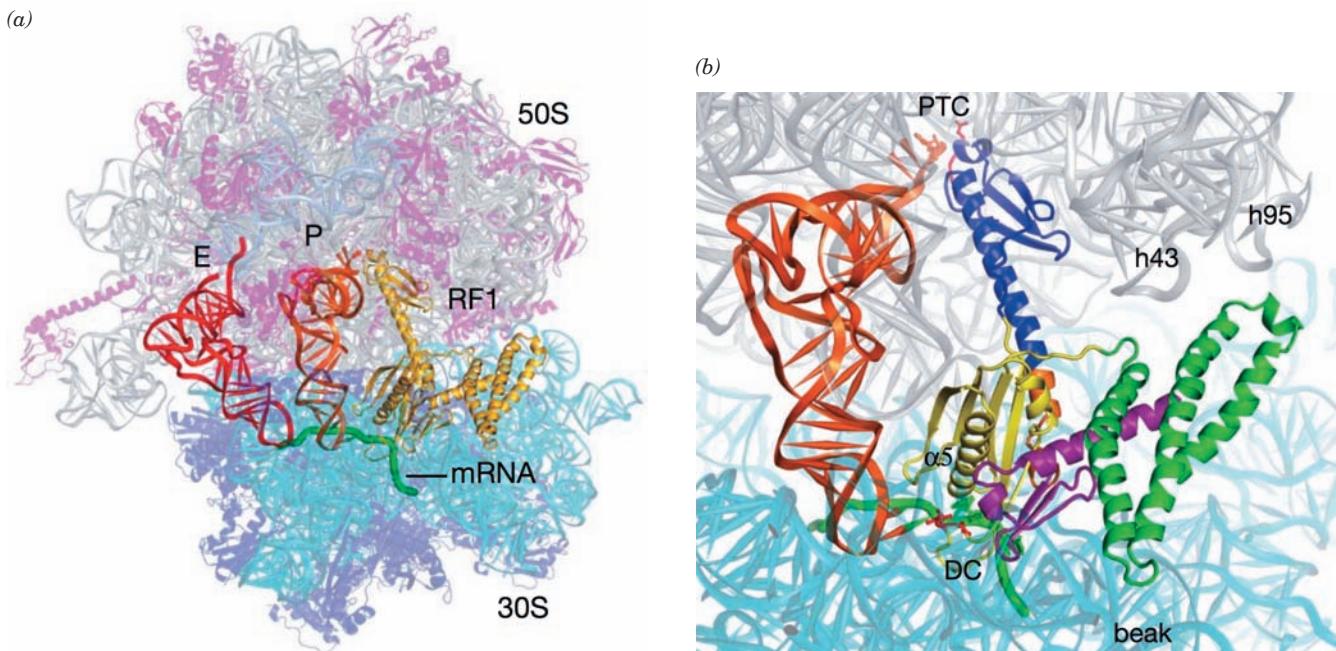


Figure 32-62 X-ray structure of the *T. thermophilus* ribosome in complex with RF-1, a UAA Stop codon-containing mRNA, and deacylated tRNAs in its P and E sites. (a) The overall structure with proteins shown in ribbon form and all RNAs shown in ladder form but the mRNA, which is drawn in worm form in green. The ribosome is semitransparent with its 23S RNA gray, its 5S RNA light blue, its 16S RNA cyan, the 50S subunit's proteins magenta, and the 30S subunit's proteins purple. The tRNAs occupying the P and E sites are orange and red, and the RF-1, which in part occupies the ribosomal A site, is yellow. (b) Close-up of the interactions between the P-site tRNA, mRNA,

and RF-1. The ribosome, tRNA, and mRNA are drawn as in Part a and the RF-1 is colored with its domains 1, 2, 3, and 4 green, yellow, blue, and magenta, respectively. The so-called switch-loop, which connects domains 3 and 4 and undergoes a major conformational rearrangement between the free and ribosome-bound RF-1, is orange. The PVT tripeptide implicated in Stop codon recognition and the GGQ tripeptide implicated in catalyzing the ester hydrolysis reaction are drawn in stick form in red. [Courtesy of Harry Noller, University of California at Santa Cruz. PDBIDs 3D5A and 3D5B.]

The binding of RF-1 or RF-2 in the DC causes A530 and A1492 of the 16S RNA to flip out from their resting state (Fig. 32-59a) as occurs with the binding of a tRNA to its cognate codon (Fig. 32-59b). However, A1493 does not flip out because in doing so it would clash with domain 2 of either release factor. Instead, it stacks on A1913 of the 23S RNA. The Stop codons are recognized by hydrogen bonding and van der Waals interactions with the similarly located PXT and SPF tripeptides on domain 2 of RF-1 and RF-2. However, the observation that mutations in RF-2 distant from its SQF motif result in altered specificity suggests that Stop codon recognition arises from a subtle balance of binding energy and conformational changes as we have seen to be the case for codon recognition by tRNA (Section 32-3Ea).

In the PTC, the GGQ tripeptide on domain 3 of both RF-1 and RF-2 contacts the 3'-terminal ribose residue (A76) of the deacylated P-site tRNA. Both Gly residues adopt backbone conformations that are forbidden for other amino acid residues, which accounts for the observations that the mutation of either residue results in an up to 10^4 -fold reduction in the rate of peptide release. The main chain NH group of the Gln residue is hydrogen bonded to the 3'-OH group of A76, which, it is hypothesized, positions it to also hydrogen bond to and thereby stabilize the transi-

sition state oxyanion in the hydrolysis reaction. In agreement with this hypothesis, the mutation of the Gln residue to Pro, which lacks a main chain NH group, abolishes the hydrolysis reaction. The side chain of the Gln residue is pointed away from the ribose (top of Fig. 32-62b) where, it is proposed, it helps position a water molecule for an in-line nucleophilic attack on the scissile ester bond. In addition, the observation that peptide release is nearly abolished by the removal of the 2'-OH from the P-site tRNA's 3' terminal residue suggests that peptide release utilizes a substrate-assisted proton shuttle mechanism similar to that of peptide bond formation (Section 32-3Di). Finally, the binding of RF-1/2 shifts U2585 so as to expose the otherwise protected scissile ester bond to nucleophilic attack (Section 32-3Di). Nevertheless, the formulation of a definitive mechanism for peptide release must await the X-ray structure of a ribosome in complex with both a release factor and a peptidyl-tRNA, that is, a substrate complex.

How does the binding of a release factor to a Stop codon in the DC induce the $\sim 75\text{-}\text{\AA}$ distant PTC to hydrolyze the scissile ester bond? In the X-ray structures of RF-1 or RF-2 alone, their PXT/SPF and GGQ motifs are only $\sim 23\text{ \AA}$ apart due to a change in conformation of their switch-loop segments (Fig. 32-62b) relative to that in the ribosomal complexes. The conformation of the

switch-loop observed in the ribosomal complexes is only possible when a Stop codon is recognized. This is because the flipping out of A1493, which only occurs when a tRNA binds its cognate codon in the DC (Fig. 32-59), or the failure of A1913 to stack on A1493, alters the binding pocket for the ribosomally bound switch-loop. Noller has therefore proposed that the binding of a Stop codon by its corresponding release factor and the consequent rearrangement of both its switch-loop and the DC cooperatively permit the GGQ motif to bind to the PTC in a way that catalyzes peptide release.

d. RRF Binds in the Ribosomal A Site

The X-ray structure of *T. thermophilus* ribosomal recycling factor (RRF), determined by Yoshikazu Nakamura, reveals it to be a two-domain structure that resembles tRNA in its overall shape (Fig. 32-63). The comparison of this structure with those of several other bacterial RRFs indicates that the two linkers connecting the RRF domains are flexible such that domain II can rotate about the axis of the three-helix bundle forming domain I.

The X-ray structure of the *T. thermophilus* ribosome in complex with RRF in its A site, the anticodon stem-loop (ASL) of tRNA^{Phe} in its P site, tRNA_f^{Met} in its E site, and an mRNA with a UAG Stop codon in the A site was determined by Ramakrishnan (Fig. 32-64). Domain I of the RRF spans the A and P sites of the 50S ribosome, a position in which the tip of its domain I would clash with a tRNA in the P site (and which rationalizes why a ribosomal complex of RRF and a tRNA in the P site has not been crystallized). This suggests that RRF binding forces a tRNA bound in the P site into the P/E hybrid binding state (Section 32-3D). Previous structural studies suggested that RRF binding induced changes in the bridges connecting the small and large subunits (Section 32-3Ae). However, no such changes are observed in the above structure. Perhaps they occur in the P/E hybrid state.

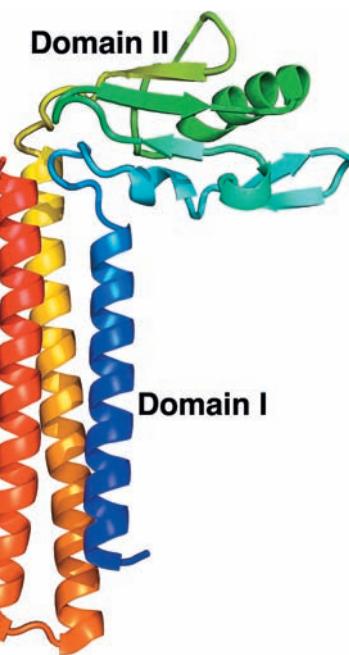


Figure 32-63 X-ray structure of *T. thermophilus* RRF. This monomeric protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). [Based on an X-ray structure by Yoshikazu Nakamura, The University of Tokyo, Japan. PDBid 1EH1.]

e. GTP Hydrolysis Speeds Up Ribosomal Processes

What is the role of the GTP hydrolysis reactions mediated by the various ribosomally associated G proteins (IF-2, EF-Tu, EF-G, and RF-3 in bacteria)? Translation occurs in the absence of GTP, albeit slowly, so that the free energy of the peptidyl transferase reaction is sufficient to drive the entire translational process. Moreover, none of the GTP

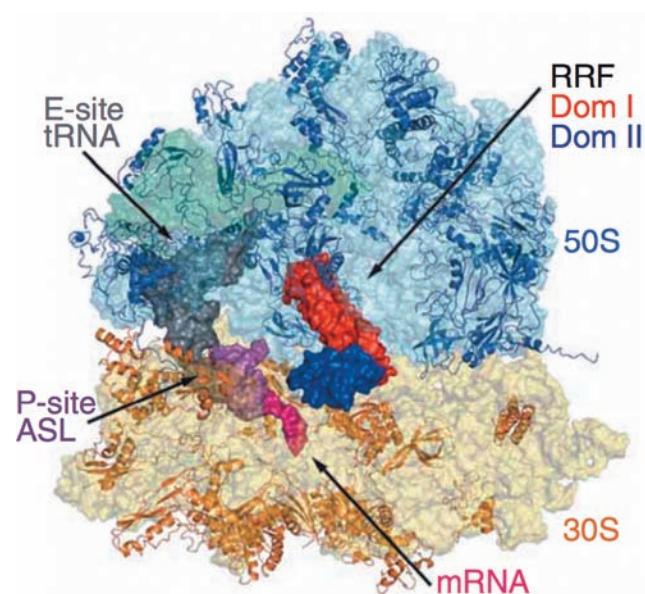


Figure 32-64 X-ray structure of the *T. thermophilus* ribosome in complex with RRF in its A site, the anticodon stem-loop (ASL) of tRNA^{Phe} in its P site, tRNA_f^{Met} in its E site, and an mRNA with a UAG Stop codon in the A site. The ribosomal RNAs are shown as semitransparent surface diagrams with 23S RNA light blue, 5S RNA blue-green, and 16S RNA yellow. The ribosomal proteins are drawn in ribbon form with 50S subunit proteins blue and 30S subunit proteins tan. The RRF, tRNAs, and mRNA are represented by their surface diagrams with domains I and II of RRF red and blue, the mRNA magenta, the P-site ASL purple, and the E-site tRNA_f^{Met} gray. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2V46 and 2V47.]

hydrolysis reactions yields a “high-energy” covalent intermediate as does, say, ATP hydrolysis in numerous biosynthetic reactions. Instead, the ribosomal binding of a G-protein induces it to hydrolyze its bound GTP to GDP resulting a conformational change that causes the ribosome to carry out a particular process (fMet-tRNA_f^{Met} binding for IF-2, accommodation for EF-Tu, translocation for EF-G, and RF-1/2 release for RF-3) and release the resulting G-protein · GDP complex. *The high rate and irreversibility of the GTP hydrolysis reaction ensures that the various complex ribosomal processes to which it is coupled, initiation, elongation, and termination, will themselves be fast and irreversible.* In essence, G-protein · GTP complexes act as Maxwell’s demons to trap the ribosome in functionally productive conformations. Hence, as we saw to be the case for ribosomal proofreading (Section 32-3Eb), the ribosome utilizes the free energy of GTP hydrolysis to gain a more ordered (lower entropy) state rather than a higher energy state as often occurs in ATP-dependent processes.

G. Protein Synthesis Inhibitors: Antibiotics

Antibiotics are bacterially, fungally, or synthetically produced substances that inhibit the growth of microorganisms. Antibiotics are known to inhibit a variety of essential biological processes, including DNA replication (e.g., ciprofloxacin; Section 29-3Cd), transcription (e.g., rifamycin B; Section 31-2Bb), and bacterial cell wall synthesis (e.g., penicillin; Section 11-3Bb). However, *the majority of known antibiotics, including a great variety of medically useful substances, block translation.* This situation is presumably a consequence of the translational machinery’s enormous complexity, which makes it vulnerable to disruption in many ways. Antibiotics have also been useful in an-

alyzing ribosomal mechanisms because, as we have seen for puromycin (Section 32-3Df), the blockade of a specific function often permits its biochemical dissection into its component steps. Table 32-10 and Fig. 32-65 present several medically significant and/or biochemically useful translational inhibitors. We study the mechanisms of a few of the best characterized of them below.

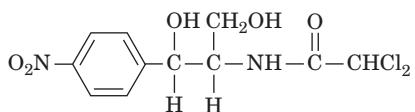
a. Streptomycin

Streptomycin, which was discovered in 1944 by Selman Waksman, is a medically important member of a family of antibiotics known as **aminoglycosides** that inhibit prokaryotic ribosomes in a variety of ways. At low concentrations, streptomycin induces the ribosome to characteristically misread mRNA: One pyrimidine may be mistaken for the other in the first and second codon positions and either pyrimidine may be mistaken for adenine in the first position. This inhibits the growth of susceptible cells but does not kill them. At higher concentrations, however, streptomycin prevents proper chain initiation and thereby causes cell death.

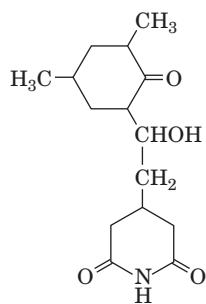
Certain streptomycin-resistant mutants (*str*^R) have ribosomes with an altered protein S12 compared with streptomycin-sensitive bacteria (*str*^S). Intriguingly, a change in base C912 of 16S rRNA (which lies in its central domain; Fig. 32-27a) also confers streptomycin resistance. (Some mutant bacteria are not only resistant to streptomycin but dependent on it; they require it for growth.) In partial diploid bacteria that are heterozygous for streptomycin resistance (*str*^R/*str*^S), streptomycin sensitivity is dominant. This puzzling observation is explained by the finding that, in the presence of streptomycin, *str*^S ribosomes remain bound to initiation sites, thereby excluding *str*^R ribosomes from these sites. Moreover, the mRNAs in these blocked complexes are degraded after a few

Table 32-10 Some Ribosomal Inhibitors

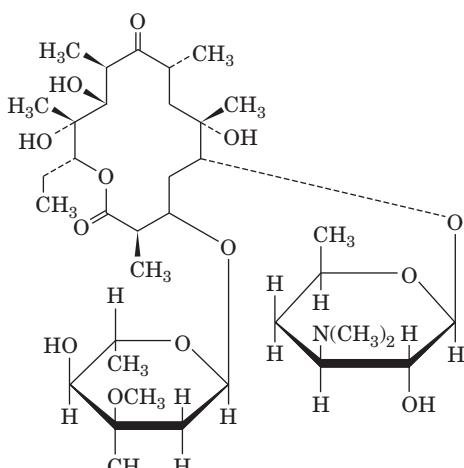
Inhibitor	Action
Chloramphenicol	Inhibits peptidyl transferase on the prokaryotic large subunit
Cycloheximide	Inhibits peptidyl transferase on the eukaryotic large subunit
Erythromycin	Inhibits translocation by the prokaryotic large subunit
Fusidic acid	Inhibits elongation in prokaryotes by binding to EF-G · GDP in a way that prevents its dissociation from the large subunit
Paromomycin	Increases the ribosomal error rate
Puromycin	An aminoacyl-tRNA analog that causes premature chain termination in prokaryotes and eukaryotes
Streptomycin	Causes mRNA misreading and inhibits chain initiation in prokaryotes
Tetracycline	Inhibits the binding of aminoacyl-tRNAs to the prokaryotic small subunit
Diphtheria toxin	Catalytically inactivates eEF2 by ADP-ribosylation
Ricin/abrin/α-sarcin	Ricin and abrin are poisonous plant glycosidases that catalytically inactivate the eukaryotic large subunit by hydrolytically depurinating a specific highly conserved A residue of the 28S RNA, which is located on the so-called sarcin–ricin loop that forms a critical part of the ribosomal factor–binding center; α-sarcin is a fungal protein that cleaves a specific phosphodiester bond in the sarcin–ricin loop



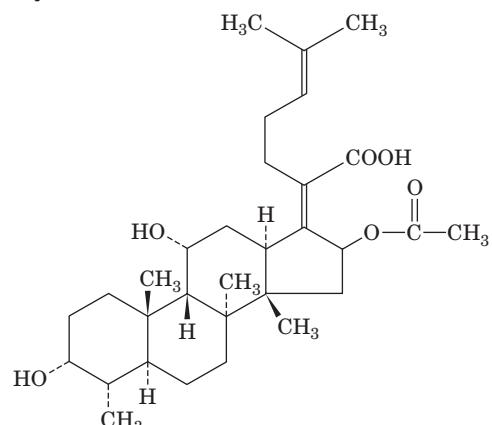
Chloramphenicol



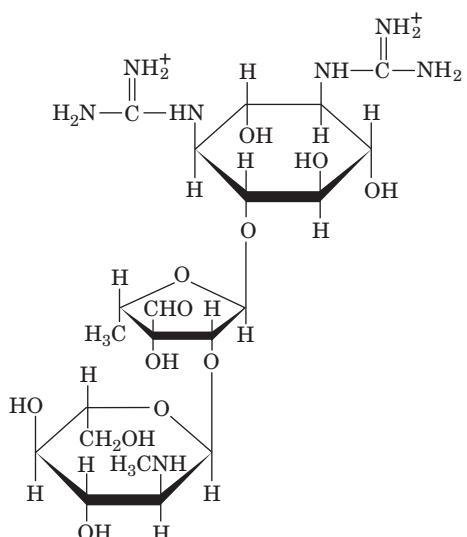
Cycloheximide



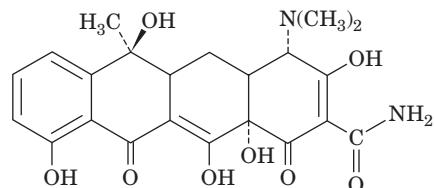
Erythromycin



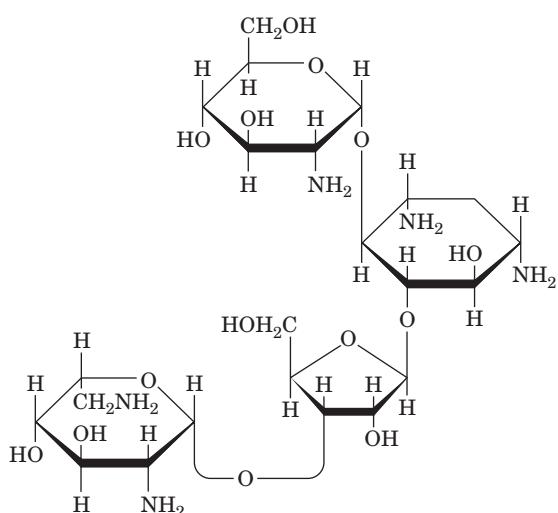
Fusidic acid



Streptomycin



Tetracycline



Paromomycin

Figure 32-65 Selection of antibiotics that act as translational inhibitors.

minutes, which allows the *str*^S ribosomes to bind to newly synthesized mRNAs as well.

b. Chloramphenicol

Chloramphenicol, the first of the “broad-spectrum” antibiotics, inhibits the peptidyl transferase activity on the large subunit of prokaryotic ribosomes. However, its clinical uses are limited to only severe infections because of its toxic side effects, which are caused, at least in part, by the chloramphenicol sensitivity of mitochondrial ribosomes. The 23S RNA is implicated in chloramphenicol binding by the observation that some of its mutants are chloramphenicol resistant. Indeed, X-ray studies indicate that chloramphenicol binds in the large subunit’s polypeptide exit tunnel in the vicinity of the A site. This explains why chloramphenicol competes for binding with the 3' end of aminoacyl-tRNAs as well as with puromycin (whose ribosomal binding site overlaps that of chloramphenicol) but not with peptidyl-tRNAs. These observations suggest that chloramphenicol inhibits peptidyl transfer by interfering with the interactions of ribosomes with A site-bound aminoacyl-tRNAs.

c. Paromomycin

Paromomycin, a clinically useful aminoglycoside antibiotic, increases the ribosomal error rate. The X-ray structure of the 30S subunit in complex with paromomycin (Fig. 32-66) reveals that it binds to the interior of the RNA loop in which the bases of A1492 and A1493 are normally stacked (Fig. 32-59a). This causes these bases to flip out of the loop and assume a conformation resembling that in the codon–anticodon–30S subunit complex (Fig. 32-59b). Indeed, this codon–anticodon–30S subunit complex is not significantly disturbed by the binding of paromomycin. As we have seen in Section 32-3Ea, the 30S subunit employs A1492 and A1493 to ascertain whether the first two codon–anticodon base pairs are Watson–Crick base pairs, that is, whether the incoming tRNA is cognate to the codon in the A site. Noncognate tRNAs normally have insufficient codon–anticodon binding energy to flip A1492 and A1493 out of the loop and consequently are rejected by the ribosome. However, the binding of paromomycin to the 30S subunit pays the energetic price of these base flips. This facilitates the ribosomal acceptance (stabilizes the binding) of near-cognate aminoacyl-tRNAs and hence the erroneous incorporation of their amino acid residues into the polypeptide being synthesized.

d. Tetracycline

Tetracycline and its derivatives are broad-spectrum antibiotics that bind to the small subunit of prokaryotic ribosomes, where they inhibit aminoacyl-tRNA binding. An X-ray structure of tetracycline in complex with the 30S subunit reveals that tetracycline mainly binds in a crevice comprised of only the 3' major domain of 16S RNA (Fig. 32-27a) and which is located in the neck of the 30S subunit just above its A site. This permits the initial screening of the aminoacyl-tRNA to proceed but physically blocks its accommodation into the peptidyl transferase (A/A) site after

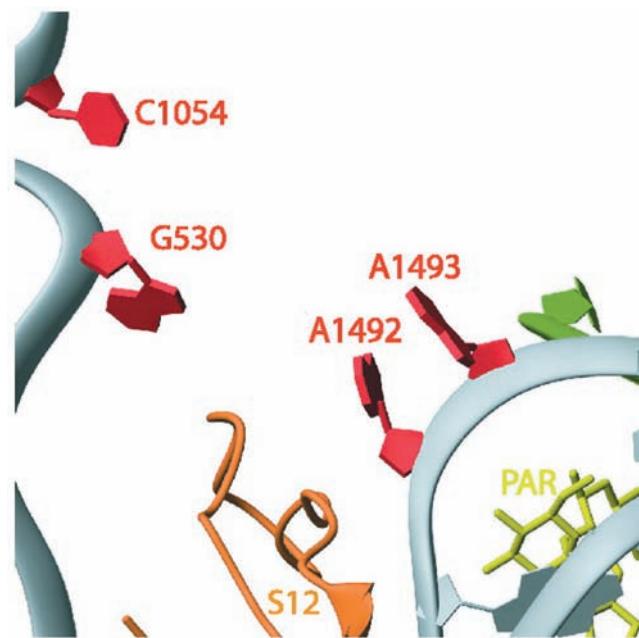


Figure 32-66 X-ray structure of the 30S ribosome in complex with the antibiotic paromomycin. The view and coloring are the same as those in Fig. 32-59 with the paromomycin (PAR) drawn in stick form in yellow-green. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBID 1IBK.]

EF-Tu–catalyzed GTP hydrolysis has occurred, resulting in the release of the tRNA. Hence, in addition to preventing protein synthesis, tetracycline binding causes the unproductive hydrolysis of GTP which, since this occurs every time a cognate aminoacyl-tRNA binds to the ribosome, poses an enormous energetic drain on the cell. The nucleotides forming the tetracycline binding site are poorly conserved in eukaryotic ribosomes, thereby accounting for tetracycline’s bacterial specificity.

Tetracycline also blocks the stringent response (Section 31-3I) by inhibiting (p)ppGpp synthesis. This indicates that deacylated tRNA must bind to the A site in order to activate stringent factor.

Tetracycline-resistant bacterial strains have become quite common, thereby precipitating a serious clinical problem. Resistance is often conferred by a decrease in bacterial cell membrane permeability to tetracycline rather than any alteration of ribosomal components.

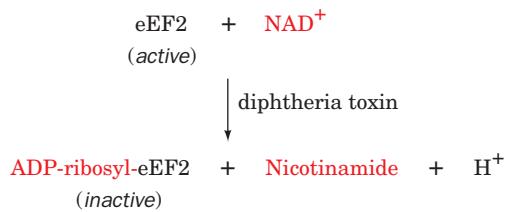
e. Diphtheria Toxin

Diphtheria is a disease resulting from bacterial infection by *Corynebacterium diphtheriae* that harbor the bacteriophage **corynephage β** . Diphtheria was a leading cause of childhood death until the late 1920s when immunization became prevalent. Although the bacterial infection is usually confined to the upper respiratory tract, the bacteria secrete a phage-encoded protein, known as **diphtheria toxin (DT)**, which is responsible for the disease’s lethal effects. *Diphtheria toxin* specifically inactivates the eukaryotic elongation factor eEF2, thereby inhibiting eukaryotic protein synthesis.

The pathogenic effects of diphtheria are prevented, as was discovered in the 1880s, by immunization with **toxoid** (formaldehyde-inactivated toxin). Individuals who have contracted diphtheria are treated with antitoxin from horse serum, which binds to and thereby inactivates DT, as well as with antibiotics to combat the bacterial infection.

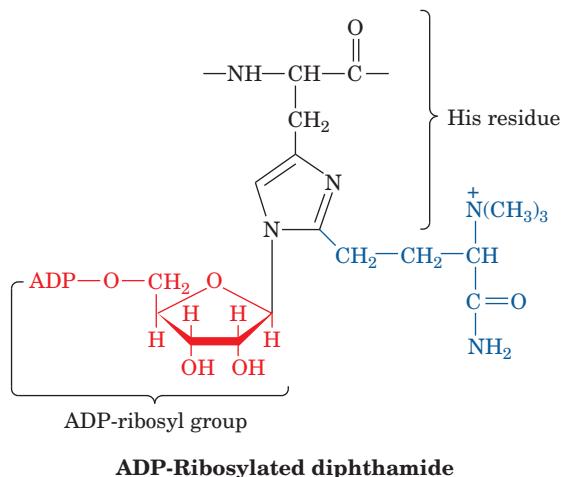
DT is a member of the family of bacterial toxins that includes cholera toxin (CT) and pertussis toxin (PT; Section 19-2Ce). It is a monomeric 535-residue protein that is readily cleaved past its Arg residues 190, 192, and 193 by trypsin and trypsinlike enzymes. This hydrolysis occurs around the time diphtheria toxin encounters its target cell, yielding two fragments, A and B, which, nevertheless, remain linked by a disulfide bond. The B fragment's C-terminal domain binds to a specific receptor on the plasma membrane of susceptible cells, thereby inducing DT's uptake into the endosome (Fig. 12-91) via receptor-mediated endocytosis (Section 12-5Bc; free fragment A is devoid of toxic activity). The endosome's low pH of 5 triggers a conformational change in the B fragment's N-terminal domain, which then inserts into the endosomal membrane so as to facilitate the entry of the A fragment into the cytoplasm. The disulfide bond linking the A and B subunits is then cleaved by the cytoplasm's reducing environment.

Within the cytosol, the A fragment catalyzes the **ADP-ribosylation** of eEF2 by NAD^+ ,



thereby inactivating this elongation factor. Since the A fragment acts catalytically, *one molecule is sufficient to ADP-ribosylate all of a cell's eEF2s, which halts protein synthesis and kills the cell*. Only a few micrograms of diphtheria toxin are therefore sufficient to kill an unimmunized individual.

Diphtheria toxin specifically ADP-ribosylates a modified His residue on eEF2 known as **diphthamide**:



Diphthamide occurs only in eEF2 (not even in its bacterial counterpart, EF-G), which accounts for the specificity of diphtheria toxin in exclusively modifying eEF2 (recall that CT ADP-ribosylates a specific Arg residue on G_{sa} and PT ADP-ribosylates a specific Cys residue on G_{ia} ; Section 19-2C). Since diphthamide occurs in all eukaryotic eEF2s, it probably is essential to eEF2 activity. Yet, certain mutant cultured animal cells, which have unimpaired capacity to synthesize proteins, lack the enzymes that post-translationally modify His to diphthamide (although mutating the diphthamide His to Asp, Lys, or Arg inactivates translation). Perhaps the diphthamide residue has a control function.

4 CONTROL OF EUKARYOTIC TRANSLATION

The rates of ribosomal initiation on prokaryotic mRNAs differ by factors of up to 100, a variation that is largely a consequence of their different Shine-Dalgarno sequences. Moreover, the genes forming an operon are often expressed in decreasing molar amounts from the operon's 5' end to its 3' end. For example, the proteins specified by the *E. coli lac* operon (Section 31-1Ab), β -galactosidase, galactose permease, and thiogalactoside transacetylase, are produced in molar ratios of 10:5:2. Such **polarity** may arise when the initiation codon of a gene that lacks a Shine-Dalgarno sequence is very near the Stop codon of its upstream gene, a situation that occurs most often when the Stop codon overlaps the initiation codon as in the sequence AUGA. The translation of the upstream gene will then be required for the translation of the downstream gene, a phenomenon termed **translational coupling**. The polarity arises because a ribosome often dissociates from the mRNA on encountering the upstream gene's Stop codon. Alternatively, an mRNA may fold in a way that masks an internal Shine-Dalgarno sequence, for example, by the base pairing of a segment adjacent to the Shine-Dalgarno sequence to a downstream element of the preceding gene. Such Shine-Dalgarno sequences only become available when a ribosome that is translating the preceding gene disrupts the folded structure.

Genetic expression in prokaryotes is largely transcriptionally controlled (Section 31-3). This is apparently because prokaryotic mRNAs have lifetimes of only a few minutes and, hence, it is a more efficient use of resources to control their transcription. Nevertheless, the expression of certain prokaryotic genes is translationally controlled, most notably those encoding the ribosomal proteins (which comprise 10% of cellular proteins), which must be produced in equimolar amounts. The production of ribosomal proteins is controlled, in part, through a process in which a ribosomal protein binds to the mRNA of the operon encoding it in the vicinity of a translational start site located near the mRNA's 5' end so as to inhibit its translational initiation. However, each such protein binds

more tightly to an rRNA in forming the ribosome. Consequently, only when there is an excess of that protein will it inhibit its own translation as well as those of other proteins encoded by its operon.

Eukaryotic cells, whose mRNAs have lifetimes of hours or days, respond to many of their needs through translational control. In this section, we examine how eukaryotic translation is regulated via the phosphorylation/dephosphorylation of eIF2 and eIF4E. We then consider translational control by mRNA masking and cytoplasmic polyadenylation and end by discussing the uses of antisense oligonucleotides.

A. Regulation of eIF2

Four important pathways for the regulation of translation in eukaryotes involve the phosphorylation of the conserved Ser 51 on the α subunit of eIF2 (**eIF2 α** ; recall that eIF2 is an $\alpha\beta\gamma$ trimer that conducts Met-tRNA_i^{Met} to the 40S ribosomal subunit, and the resulting complex scans the bound mRNA for the initiating AUG codon to form the 48S preinitiation complex; Section 32-3Cd). The so-called **eIF2 α kinases** that do so share a conserved kinase domain but have unique regulatory domains.

a. Heme Availability Controls Globin Translation

Reticulocytes synthesize protein, almost exclusively hemoglobin, at an exceedingly high rate and are therefore a favorite subject for the study of eukaryotic translation. Hemoglobin synthesis in fresh reticulocyte lysates proceeds normally for several minutes but then abruptly stops because of the inhibition of translational initiation and the consequent polysome disaggregation. This process is prevented by the addition of heme [a mitochondrial product (Section 26-4A) that this *in vitro* system cannot synthesize], thereby indicating that *globin synthesis is regulated by heme availability*. The inhibition of globin translational initiation is also reversed by the addition of the eukaryotic initiation factor eIF2 and by high levels of GTP.

In the absence of heme, reticulocyte lysates accumulate an eIF2 α kinase named **heme-regulated inhibitor [HRI]**; also called heme-controlled repressor (**HCR**). HRI is a homodimer whose 629-residue subunits each contain two heme-binding sites. When heme is plentiful, both of these sites are occupied and the protein, which is autophosphorylated at several Ser and Thr residues, is inactive. However, when heme is scarce, one of these sites loses its bound heme, thereby activating HRI to autophosphorylate itself at several additional sites and to phosphorylate Ser 51 of eIF2 α .

Phosphorylated eIF2 can participate in the ribosomal initiation process in much the same way as unphosphorylated eIF2. This puzzling observation was clarified by the discovery that GDP does not dissociate from phosphorylated eIF2 at the completion of the initiation process as it normally does through a process facilitated by eIF2B acting as a GEF (Fig. 32-44). This is because phosphorylated eIF2 forms a much tighter complex with eIF2B than does unphosphorylated eIF2. This sequesters eIF2B (Fig. 32-

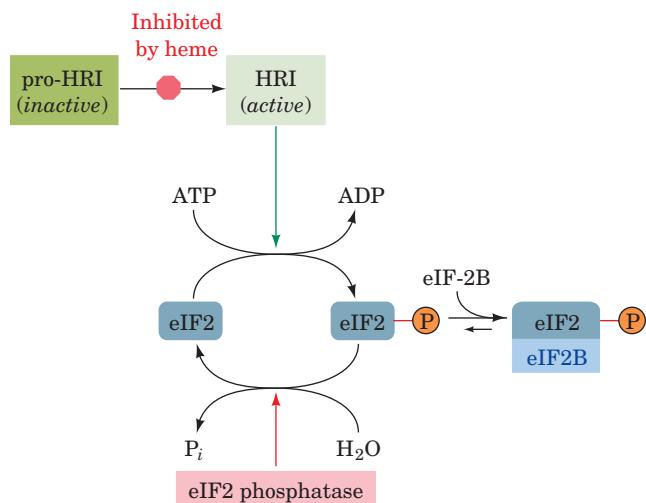


Figure 32-67 Model for heme-controlled protein synthesis in reticulocytes.

67), which is present in lesser amounts than eIF2, thereby preventing the regeneration of the eIF2 · GTP required for translational initiation. The presence of heme reverses this process by inhibiting HRI, whereon the phosphorylated eIF2 molecules are reactivated through the action of **eIF2 phosphatase**, which is unaffected by heme. The reticulocyte thereby coordinates its synthesis of globin and heme.

b. Interferons Protect against Viral Infection

Interferons are cytokines that are secreted by virus-infected vertebrate cells. On binding to surface receptors of other cells, interferons convert them to an antiviral state, which inhibits the replication of a wide variety of RNA and DNA viruses. Indeed, the discovery of interferons in the 1950s arose from the observation that virus-infected individuals are resistant to infection by a second type of virus.

There are three families of interferons: **type α** or **leukocyte interferon** (165 residues; leukocytes are white blood cells), the related **type β** or **fibroblast interferon** (166 residues; fibroblasts are connective tissue cells), and **type γ** or **lymphocyte interferon** (146 residues; lymphocytes are immune system cells). *Interferon synthesis is induced by the double-stranded RNA (dsRNA) that is generated during infection by both DNA and RNA viruses, as well as by the synthetic dsRNA poly(I) · poly(C).* Interferons are effective antiviral agents in concentrations as low as $3 \times 10^{-14} M$, which makes them among the most potent biological substances known. Moreover, they have far wider specificities than antibodies raised against a particular virus. They have therefore elicited great medical interest, particularly since some cancers are virally induced (Section 19-3B). Indeed, they are in clinical use against certain tumors and viral infections. These treatments are made possible by the production of large quantities of these otherwise quite scarce proteins through recombinant DNA techniques (Section 5-5G).

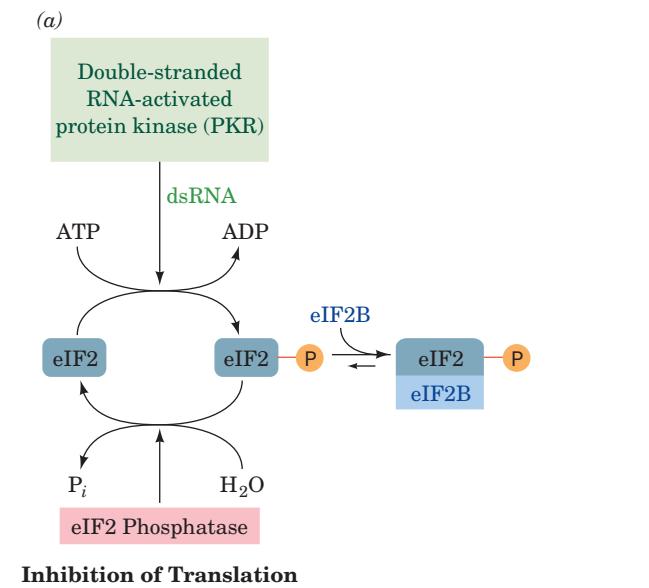
Interferons prevent viral proliferation largely by inhibiting protein synthesis in infected cells (lymphocyte interferon

also modulates the immune response). They do so in two independent ways (Fig. 32-68):

1. Interferons induce the production of an eIF2 α kinase, **double-stranded RNA-activated protein kinase [PKR; also known as double-stranded RNA-activated inhibitor (DAI); 551 residues]**, which on binding dsRNA, dimerizes and autophosphorylates itself. This activates PKR to phosphorylate eIF2 α at its Ser 51, thereby inhibiting ribosomal initiation and hence the proliferation of viruses in virus-infected cells. The importance of PKR to cellular antiviral defense is indicated by the observation that many viruses express inhibitors of PKR.

2. Interferons also induce the synthesis of **(2',5')-oligoadenylate synthetase (2,5A synthetase)**. In the presence of dsRNA, this enzyme catalyzes the synthesis from ATP of the unusual oligonucleotide **pppA(2'p5'A)_n** where $n = 1$ to 10. This compound, **2,5-A**, activates a preexisting endonuclease, **RNase L**, to degrade mRNA, thereby inhibiting protein synthesis. 2,5-A is itself rapidly degraded by an enzyme named **(2',5')-phosphodiesterase** so that it must be continually synthesized to maintain its effect.

The independence of the 2,5-A and PKR systems is demonstrated by the observation that the effect of 2,5-A on protein synthesis is reversed by added mRNA but not by added eIF2. [Recall that RNA interference (RNAi; Section 31-4At) constitutes an alternative dsRNA-based antiviral defense.]



c. PERK Prevents the Buildup of Unfolded Proteins in the ER

PERK-like endoplasmic reticulum kinase (PERK), a 1087-residue transmembrane protein, resides in the endoplasmic reticulum (ER) membrane of all multicellular eukaryotes. It is repressed by its binding to the ER-resident chaperone BiP (Section 12-4Bf). When the ER contains an excessive amount of unfolded proteins (caused by various forms of stress such as high temperatures), BiP dissociates from PERK, thereby activating PERK to phosphorylate eIF2 α at its Ser 51 and hence inhibit translation. Thus PERK functions to protect the cell from the irreversible damage caused by the accumulation of unfolded proteins in the ER.

Wolcott–Rallison syndrome is a genetic disease characterized mainly by insulin-dependent (type I) diabetes that develops in early infancy (type I diabetes usually first appears in childhood; Section 27-4B). It is caused by mutations in the catalytic domain of PERK. This results in the death of pancreatic β cells, in which PERK is particularly abundant. Multiple systemic disorders subsequently occur including **osteoporosis** (reduction in the quantity of bone) and growth retardation.

d. GCN2 Regulates Amino Acid Biosynthesis

GCN2 (1590 residues), the sole eIF2 α kinase in yeast, is a transcriptional activator of the gene encoding **GCN4**, a transcriptional activator of numerous yeast genes, many of which encode enzymes that participate in amino acid biosynthetic pathways. The C-terminal domain of GCN2, which resembles histidyl-tRNA synthetase (HisRS), preferentially binds uncharged tRNAs (whose presence is indicative of an insufficient supply of amino acids). The binding of an uncharged tRNA to this HisRS-like domain activates the adjacent eIF2 α kinase domain and thereby inhibits translational initiation, although at only a modest level.

Despite this inhibition of yeast protein synthesis, activated GCN2 induces the expression of GCN4. This seemingly paradoxical property of GCN2, as Alan Hinnebusch explained, arises from the fact that GCN4 mRNA contains four short so-called **upstream open reading frames**

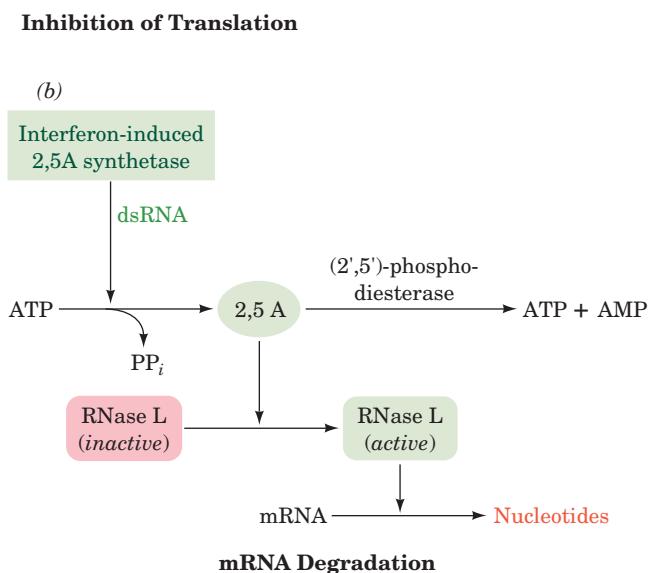


Figure 32-68 The action of interferon. In interferon-treated cells, the presence of dsRNA, which normally results from a viral infection, causes (a) the inhibition of translational initiation and (b) the degradation of mRNA, thereby blocking translation and preventing virus replication.

(uORFs), uORF1 to uORF4, in its 5' leader that precedes the sequence encoding GCN4. Under the normal nutrient conditions in which GCN2 is inactive, the ribosome binds to the mRNA near its 5' cap and scans for the nearest AUG initiation codon (which is in uORF1), where it forms the 48S preinitiation complex (Fig. 32-44) and commences the translation of uORF1 (Section 32-3Cd). On terminating translation at uORF1's Stop codon, the presence of the surrounding A + U-rich sequences causes the ribosome to resume scanning for the next AUG codon, where it initiates the translation of uORF2. This process repeats until the ribosome terminates at the end of uORF4, where its Stop codon's surrounding G + C-rich sequences induce the ribosome to disengage from the mRNA. Hence GCN4 is only expressed at a low basal level. However, under the low nutrient conditions in which GCN2 phosphorylates eIF2 α at its Ser 51, the resulting reduced level of the eIF2 · Met-tRNA $^{\text{Met}}$, GTP ternary complex causes the 40S subunit to scan longer distances before it can form the 48S preinitiation complex. Consequently, ~50% of the ribosomes scan past uORF2, uORF3, and uORF4 and only initiate translation at the *GCN4* AUG codon, which is therefore translated at a high level (uORF2 and uORF3 can be mutationally eliminated without significantly affecting translational control).

Mammalian homologs of GCN2 are activated under conditions of amino acid starvation. This suggests that the foregoing process has been conserved throughout eukaryotic evolution.

B. Regulation of eIF4E

eIF4E (cap-binding protein) binds to the m⁷G cap of eukaryotic mRNAs and thereby participates in translational initiation by helping to identify the initiating AUG codon (Section 32-3Cd). When mammalian cells are treated with hormones, cytokines, **mitogens** (substances that induce mitosis), and/or growth factors, Ser 209 of human eIF4E is phosphorylated via a Ras-activated MAP kinase cascade (Sections 19-3C and 19-3D), thereby increasing eIF4E's affinity for capped mRNA and hence stimulating translational initiation. Ser 209 occupies a surface position on eIF4E adjacent to the binding site for the β phosphate group of the m⁷GDP and flanking the putative binding cleft for mRNA (Fig. 32-45b). The structure of eIF4E suggests that the phosphoryl group of phosphorylated Ser 209 forms a salt bridge with Lys 159, which occupies the other side of the putative mRNA-binding cleft, so as to form a clamp that would help stabilize the bound mRNA. The importance of regulating eIF4E activity is indicated by the observations that the overexpression of eIF4E causes the malignant transformation of rodent cell lines and that eIF4E expression is elevated in several human cancers.

The homologous ~120-residue proteins known as **4E-BP1**, **4E-BP2**, and **4E-BP3** (BP for *binding protein*; the first two are also known as **PHAS-I** and **PHAS-II**) inhibit cap-dependent translation. They do so by binding on the

opposite side of eIF4E from its mRNA-binding site, presumably to a patch of seven highly conserved surface residues, and hence do not prevent eIF4E from binding the m⁷G cap. Rather, they block eIF4E from binding to eIF4G and thereby interfere with the formation of the eIF4F complex that positions the 40S ribosomal subunit-bound Met-tRNA $^{\text{Met}}$ on the mRNA's initiating AUG codon (Section 32-3Cd). In fact, the 4E-BPs and eIF4G all possess the sequence motif YXXXXL ϕ (where ϕ is an aliphatic residue, most often L but also M or F) through which they bind to eIF4E.

The treatment of responsive cells with insulin or any of several protein growth factors causes the 4E-BPs to dissociate from eIF4E. This is because the presence of these hormones induces the phosphorylation of the 4E-BPs at six Ser/Thr residues via the signal transduction pathway involving PI3K, PKB, and mTOR (Fig. 19-67). Evidently, the phosphorylation of eIF4E and the 4E-BPs have similar if not synergistic effects in the hormonal regulation of translation in eukaryotes.

C. mRNA Masking and Cytoplasmic Polyadenylation

It has been known since the nineteenth century that early embryonic development in animals such as sea urchins, insects, and frogs is governed almost entirely by information present in the oocyte (egg) before fertilization. Indeed, sea urchin embryos exposed to sufficient actinomycin D (Section 31-2Cc) to inhibit RNA synthesis without blocking DNA replication develop normally through their early stages without a change in their protein synthesis program. This is in part because an unfertilized egg contains large quantities of mRNA that is "masked" by associated proteins to form ribonucleoprotein particles, thereby preventing the mRNAs' association with the ribosomes that are also present. On fertilization, this mRNA is "unmasked" in a controlled fashion, quite possibly by the dephosphorylation of the associated proteins, and commences directing protein synthesis. Development of the embryo can therefore start immediately on fertilization rather than wait for the synthesis of paternally specified mRNAs. Thus, gene expression in the early stages of development is entirely translationally controlled; transcriptional control only becomes important when transcription is initiated.

a. Cytoplasmic Polyadenylation

Another mechanism of translational control in oocytes and early embryos involves the polyadenylation of mRNAs in the cytoplasm (polyadenylation usually occurs in the nucleus, following which the mRNA is exported to the cytoplasm; Section 31-4Ab). A substantial number of maternally supplied mRNAs in oocytes have relatively short poly(A) tails (20–40 nt versus a usual length of ~250 nt). The 3' untranslated region of these mRNAs contains both the AAUAAA polyadenylation signal (which is required for polyadenylation in the nucleus; Section 31-4Ab)

together with a so-called **cytoplasmic polyadenylation element (CPE)**, which has the consensus sequence UUUUUUAU. The CPE is recognized by **CPE-binding protein (CPEB)**, which contains two RNA recognition motifs (RRMs) as well as a **zinc finger** motif (Section 34-3B) that contribute to its binding to the mRNA. Joel Richter discovered that CPEB recruits a 931-residue protein named **maskin** which, in turn, binds the eIF4E (cap-binding protein) that is bound to the mRNA's 5' cap (Fig. 32-69a). Maskin contains the same YXXXXL ϕ motif through which the 4E-BPs and eIF4G bind to eIF4E (Section 32-4B), thereby blocking the binding of eIF4G to eIF4E and hence preventing the formation of the 48S preinitiation complex (Fig. 32-44).

In the maturation of *Xenopus laevis* oocytes, a process that precedes fertilization and is stimulated by the steroid hormone progesterone (Section 19-1Gb), a variety of mRNAs, including those encoding several cyclins (which participate in cell cycle control; Section 34-4Da) are translationally activated. Soon after exposure to progesterone, a protein kinase named **aurora** phosphorylates the mRNA-bound CPEB at its Ser 174. This increases CPEB's affinity for cleavage and polyadenylation specificity factor (CPSF; Section 31-4Ab), which then binds to the mRNA's AAUAAA sequence, where it recruits poly(A) polymerase (PAP) to lengthen the mRNA's poly(A) tail (Fig. 32-69b).

Translational initiation and cytoplasmic polyadenylation occur simultaneously, which suggests that these processes are linked. Indeed, Richter has shown that this occurs through the binding to poly(A) of poly(A)-binding protein (PABP; Section 31-4Ab), which as we saw (Section 32-3Cd), also binds to eIF4G to circularize the mRNA. The eIF4G in this complex displaces maskin from eIF4E, thereby permitting the formation of the 48S preinitiation complex and hence the mRNA's translation (Fig. 32-69b).

Mammalian cells also exhibit cell cycle-dependent cytoplasmic polyadenylation of mRNAs. This suggests that translational control by polyadenylation is a general feature in animal cells.

D. Antisense Oligonucleotides

Since ribosomes cannot translate double-stranded RNA or DNA–RNA hybrid helices, the translation of a given mRNA can be inhibited by a segment of its complementary RNA or DNA, that is, an **antisense RNA** or an **antisense oligodeoxynucleotide**, which are collectively known as **antisense oligonucleotides**. Moreover, endogenous RNase H's (enzymes that cleave the RNA strand of an RNA–DNA duplex; Section 31-4C) cleave an mRNA–oligodeoxynucleotide duplex on its mRNA strand, leaving the antisense oligodeoxynucleotide intact for binding to another mRNA.

Since the human genome consists of ~3 billion bp, an ~15-nt oligonucleotide (which is easily synthesized; Section 7-6Aa) should ideally be able to target any segment of the human genome. This exquisite specificity provides the delivery of an antisense oligonucleotide to, or its expression in, a selected tissue or organism with enormous biomedical and biotechnological potential. However, care must be taken that an antisense oligonucleotide does not also eliminate nontarget mRNAs.

Methods for the delivery of a therapeutically useful antisense oligonucleotide to a target tissue are as yet in their infancy. This is in large part because oligonucleotides are readily degraded by the many nucleases present in an organism and because they do not readily pass through cell membranes. Moreover, a target mRNA is likely to be associated with cellular proteins and hence not available for binding to other molecules. The nuclease resistance of oligonucleotides can be increased by derivatizing them, for example, by replacing a nonbridging oxygen at each phosphate group with a methyl group or an S atom so as to yield **methylphosphonate** or **phosphorothioate oligonucleotides**, although this reduces their antisense activity. The expression of antisense oligonucleotides in the specified tissues would, of course, circumvent the delivery problem but has all the difficulties associated with gene therapy (Section 5-5H).

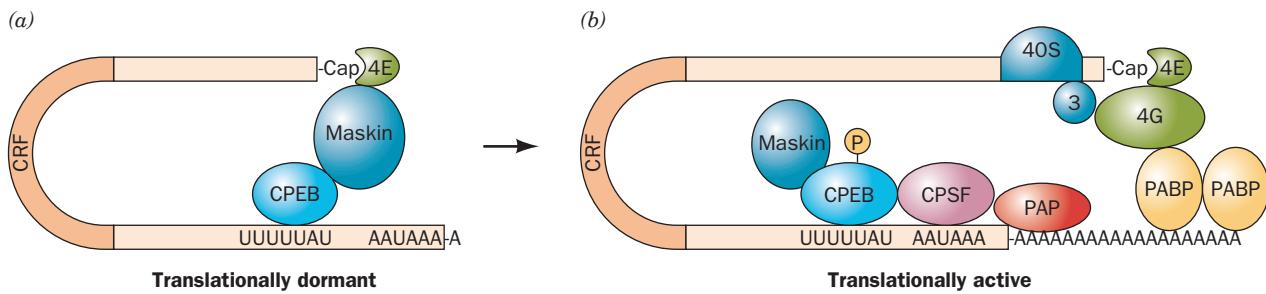


Figure 32-69 CPEB-mediated translational control. (a) In immature *Xenopus laevis* oocytes, an mRNA containing the CPE (UUUUUAU) is bound by CPEB, which binds maskin, which then binds eIF4E so as to prevent it from binding eIF4G, thereby maintaining the mRNA in a translationally dormant (masked) state. (b) In the maturation process, CPEB is phosphorylated by an aurora protein kinase. The phosphorylated CPEB binds CPSF, which recruits PAP to extend the mRNA's heretofore short poly(A) tail. PABP binds to the newly lengthened poly(A) tail and simultaneously binds to eIF4G so as to displace maskin. This permits the 48S preinitiation complex to assemble and hence the translation of the mRNA to proceed. [Based on a drawing by Mendez, R. and Richer, J.D., *Nature Rev. Mol. Cell Biol.* 2, 521 (2001).]

which recruits PAP to extend the mRNA's heretofore short poly(A) tail. PABP binds to the newly lengthened poly(A) tail and simultaneously binds to eIF4G so as to displace maskin. This permits the 48S preinitiation complex to assemble and hence the translation of the mRNA to proceed. [Based on a drawing by Mendez, R. and Richer, J.D., *Nature Rev. Mol. Cell Biol.* 2, 521 (2001).]

Despite the foregoing, antisense technology is beginning to show success. **Fomivirsen** (trade name **Vitravene**), a 21-nt phosphorothioate oligonucleotide that is complementary to an mRNA expressed by **cytomegalovirus (CMV)**, is effective in the treatment of retinitis (inflammation of the retina) caused by CMV infection in individuals with AIDS (CMV is an opportunistic pathogen that rarely infects individuals with normally functioning immune systems). It was approved for human use in 1998 by the FDA, the first antisense drug so approved. A number of antisense oligonucleotides that are mainly targeted against genes that are overexpressed in specific cancers and autoimmune diseases as well as other viral infections are in clinical trials (Section 15-4Bb), although additional antisense drugs have not yet been approved for human use.

Antisense technology has also had some success in the arena of biotechnology. For example, in tomatoes and other fruits, the enzyme **polygalacturonase (PG)**, which is expressed during ripening, depolymerizes the pectin (mainly polygalacturonic acid) in the cell wall. This results in a softening of tomatoes to the point that vine-ripened (and hence better tasting) tomatoes are unable to withstand the rigors of shipping and hence must be picked before they are ripe. The introduction into a tomato, via genetic engineering techniques, of a gene expressing antisense PG RNA yielded the so-called Flavr Savr tomato that had substantially reduced PG expression and hence remained firm after vine ripening.

5 POST-TRANSLATIONAL MODIFICATION

To become mature proteins, polypeptides must fold to their native conformations, their disulfide bonds, if any, must form, and, in the case of multisubunit proteins, the subunits must properly combine. Moreover, as we have seen throughout this text, many proteins are modified in enzymatic reactions that proteolytically cleave certain peptide bonds and/or derivatize specific residues. In this section we shall review some of these **post-translational modifications**.

A. Proteolytic Cleavage

Proteolytic cleavage is the most common type of post-translational modification. Probably all mature proteins have been so modified, if by nothing else than the proteolytic removal of their leading Met (or fMet) residue shortly after it emerges from the ribosome. Many proteins, which are involved in a wide variety of biological processes, are synthesized as inactive precursors that are activated under proper conditions by limited proteolysis. Some examples of this phenomenon that we have encountered are the conversion of trypsinogen and chymotrypsinogen to their active forms by trypic cleavages

of specific peptide bonds (Section 15-3E), and the formation of active insulin from the 84-residue proinsulin by the excision of its internal 33-residue C chain (Section 9-1Aa). Inactive proteins that are activated by removal of polypeptides are called **proproteins**, whereas the excised polypeptides are termed **propeptides**.

a. Propeptides Direct Collagen Assembly

Collagen biosynthesis is illustrative of many facets of post-translational modification. Recall that collagen, a major extracellular component of connective tissue, is a fibrous triple-helical protein whose polypeptides each contain the amino acid sequence $(\text{Gly-X-Y})_n$ where X is often Pro, Y is often 4-hydroxyproline (Hyp), and $n \approx 340$ (Section 8-2B). The polypeptides of **procollagen** (Fig. 32-70) differ from those of the mature protein by the presence of both N-terminal and C-terminal propeptides of ~ 100 residues whose sequences, for the most part, are unlike those of mature collagen. The procollagen polypeptides rapidly assemble, *in vitro* as well as *in vivo*, to form a collagen triple helix. In contrast, polypeptides extracted from mature collagen will reassemble only over a period of days, if at all. *The collagen propeptides are apparently necessary for proper procollagen folding.*

The N- and C-terminal propeptides of procollagen are respectively removed by **amino-** and **carboxyprocollagen**

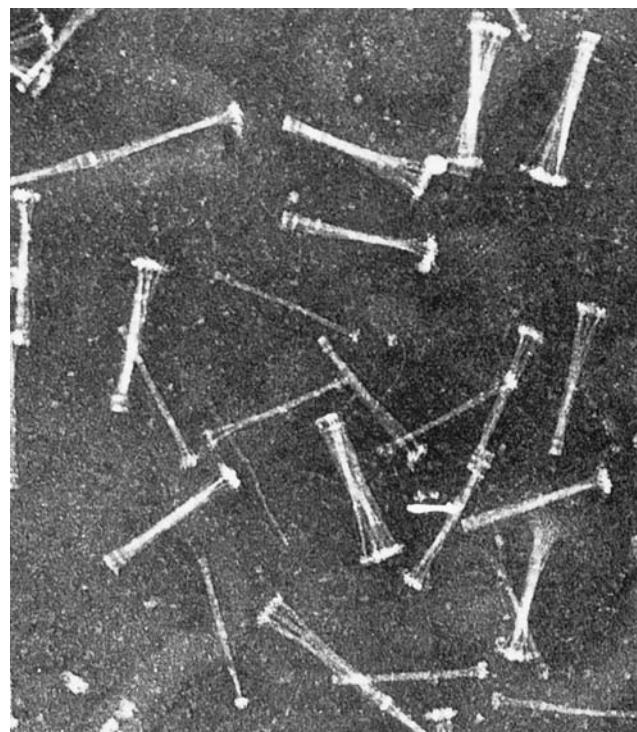


Figure 32-70 Electron micrograph of procollagen aggregates that have been secreted by fibroblasts into the extracellular medium. [Courtesy of Jerome Gross, Massachusetts General Hospital, Harvard Medical School.]

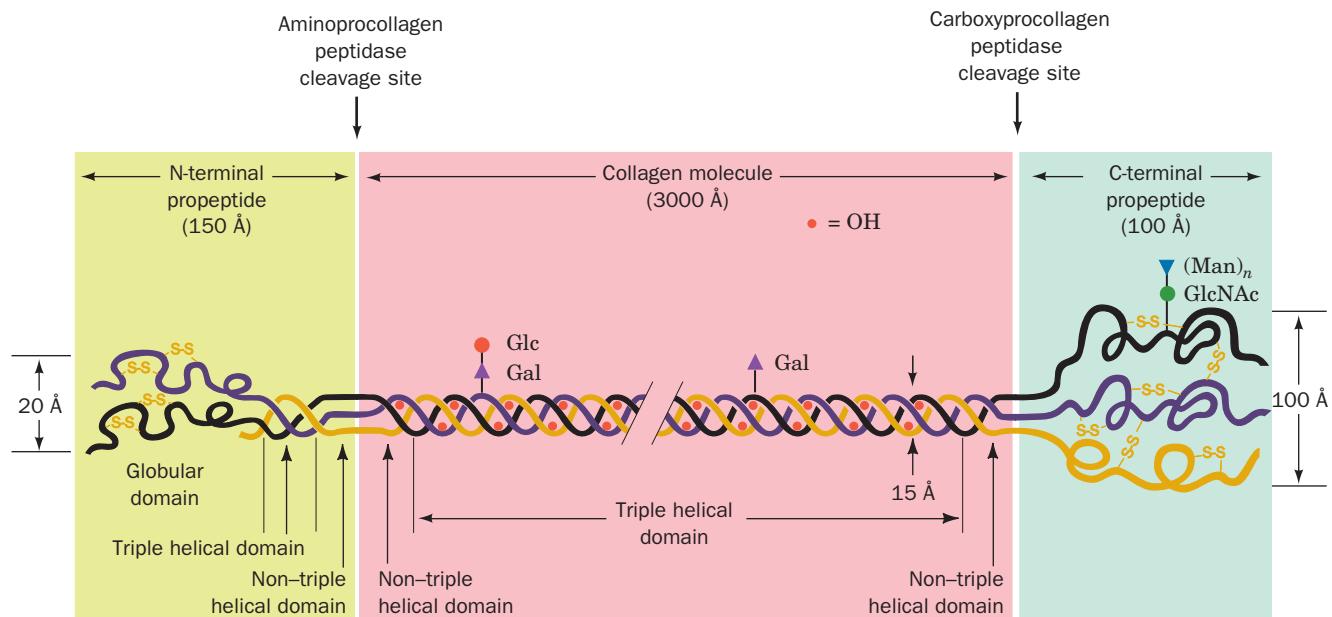


Figure 32-71 Schematic representation of the procollagen molecule. Gal, Glc, GlcNAc, and Man, respectively, denote galactose, glucose, *N*-acetylglucosamine, and mannose residues. Note that the N-terminal propeptide has intrachain disulfide

bonds while the C-terminal propeptide has both intrachain and interchain disulfide bonds. [After Prockop, D.J., Kivirikko, K.I., Tuderman, L., and Guzman, N.A., *New Engl. J. Med.* **301**, 16 (1979).]

peptidases (Fig. 32-71), which may also be specific for the different collagen types. An inherited defect of aminoprocollagen peptidase in cattle and sheep results in a bizarre condition, **dermatosparaxis**, that is characterized by extremely fragile skin. An analogous disease in humans, **Ehlers–Danlos syndrome VII**, is caused by a mutation in one of the procollagen polypeptides that inhibits the enzymatic removal of its aminopropeptide. Collagen molecules normally spontaneously aggregate to form collagen fibrils (Figs. 8-31 and 8-32). However, electron micrographs of dermatosparaxic skin show sparse and disorganized collagen fibrils. *The retention of collagen's aminopropeptides apparently interferes with proper fibril formation.* (The dermatosparaxis gene was bred into some cattle herds because heterozygotes produce tender meat.)

b. Signal Peptides Are Removed from Nascent Proteins by a Signal Peptidase

Many transmembrane proteins or proteins that are destined to be secreted are synthesized with an N-terminal **signal peptide** of 13 to 36 predominantly hydrophobic residues. As we saw in our discussion of the **secretory pathway** (Section 12-4B), a signal peptide is recognized by a **signal recognition particle (SRP)**. The SRP binds a ribosome synthesizing a signal peptide to a protein pore known as the **translocon** that is embedded in the membrane [the rough endoplasmic reticulum (RER) in eukaryotes and the plasma membrane in bacteria] and conducts the signal peptide and its following nascent polypeptide through the translocon.

Proteins bearing a signal peptide are known as **preproteins** or, if they also contain propeptides, as **prepropeptides**.

Once the signal peptide has passed through the membrane, it is specifically cleaved from the nascent polypeptide by a membrane-bound **signal peptidase**. Both insulin and collagen are secreted proteins and are therefore synthesized with leading signal peptides in the form of **preproinsulin** and **preprocollagen**. These and many other proteins are therefore subject to three sets of sequential proteolytic cleavages: (1) the deletion of their initiating Met residue, (2) the removal of their signal peptides, and (3) the excision of their propeptides.

c. Polyproteins

Some proteins are synthesized as segments of **polyproteins**, polypeptides that contain the sequences of two or more proteins. Examples include many polypeptide hormones (Section 34-3C); the proteins synthesized by many viruses, including those causing polio (Section 33-2C) and AIDS (Section 15-4Cb); and **ubiquitin**, a highly conserved eukaryotic protein involved in protein degradation (Section 32-6Bb). Specific proteases post-translationally cleave polyproteins to their component proteins, presumably through the recognition of the cleavage site sequences. Some of these proteases are conserved over remarkable evolutionary distances. For instance, ubiquitin is synthesized as several tandem repeats (**polyubiquitin**) that *E. coli* properly cleave even though prokaryotes lack ubiquitin. Other proteases have more idiosyncratic cleavage sequences. This has allowed medicinal chemists to design inhibitors of **HIV protease** (which catalyzes an essential step in the viral life cycle) that have been highly effective in attenuating if not preventing the progression of AIDS (Section 15-4Cd).

B. Covalent Modification

Proteins are subject to specific chemical derivatizations, both at the functional groups of their side chains and at their terminal amino and carboxyl groups. Over 150 different types of side chain modifications, involving all side chains but those of Ala, Gly, Ile, Leu, Met, and Val, are known (Section 4-3A). These include acetylations, glycosylations, hydroxylations, methylations, nucleotidylations, phosphorylations, and ADP-ribosylations as well as numerous “miscellaneous” modifications.

Some protein modifications, such as the phosphorylation of glycogen phosphorylase (Section 18-1A) and the ADP-ribosylation of eEF2 (Section 32-3Ge), modulate protein activity. Several side chain modifications covalently bond cofactors to enzymes, presumably to increase their catalytic efficiency. Examples of linked cofactors that we have encountered are N^e -lipoyllysine in dihydrolipoyl transacetylase (Section 21-2Ac) and 8 α -histidylflavin in succinate dehydrogenase (Section 21-3F). The attachment of complex carbohydrates, which occur in almost infinite variety, alter the structural properties of proteins and form recognition markers in various types of targeting and cell-cell interactions (Sections 11-3C, 12-3E, and 23-3B). Modifications that cross-link proteins, such as occur in collagen (Section 8-2Bc), stabilize supramolecular aggregates. The functions of most side chain modifications, however, remain enigmatic.

a. Collagen Assembly Requires Chemical Modification

Collagen biosynthesis (Fig. 32-72) is illustrative of protein maturation through chemical modification. As the nascent procollagen polypeptides pass into the RER of the fibroblasts that synthesized them, the Pro and Lys residues are hydroxylated to Hyp, 3-hydroxy-Pro, and 5-hydroxy-Lys

(Hyl). The enzymes that do so are sequence specific: **Prolyl 4-hydroxylase** and **lysyl hydroxylase** act only on the Y residues of the Gly-X-Y sequences, whereas **prolyl 3-hydroxylase** acts on the X residues but only if Y is Hyp. Glycosylation, which also occurs in the RER, subsequently attaches sugar residues to Hyl residues (Section 8-2Bb). The folding of three polypeptides into the collagen triple helix must follow hydroxylation and glycosylation because the hydroxylases and glycosyl transferases do not act on helical substrates. Moreover, the collagen triple helix denatures below physiological temperatures unless stabilized by hydrogen bonding interactions involving Hyp residues (Section 8-2B). Folding is also preceded by the formation of specific interchain disulfide bonds between the carboxylpropeptides. This observation bolsters the previously discussed conclusion that collagen propeptides help select and align the three collagen polypeptides for proper folding.

The procollagen molecules pass into the Golgi apparatus where they are packaged into **secretory vesicles** (Sections 12-4C and 12-4D) and secreted into the extracellular spaces of connective tissue. The aminopropeptides are excised just after procollagen leaves the cell and the carboxypropeptides are removed sometime later. The collagen molecules then spontaneously assemble into fibrils, which suggests that an important propeptide function is to prevent intracellular fibril formation. Finally, after the action of the extracellular enzyme **lysyl oxidase**, the collagen molecules in the fibrils spontaneously cross-link (Fig. 8-33).

C. Protein Splicing: Inteins and Exteins

Protein splicing is a post-translational modification process in which an *internal protein* segment (an **intein**) excises itself from a surrounding *external protein*, which it ligates to form the mature **extein**. The portions of the unspliced

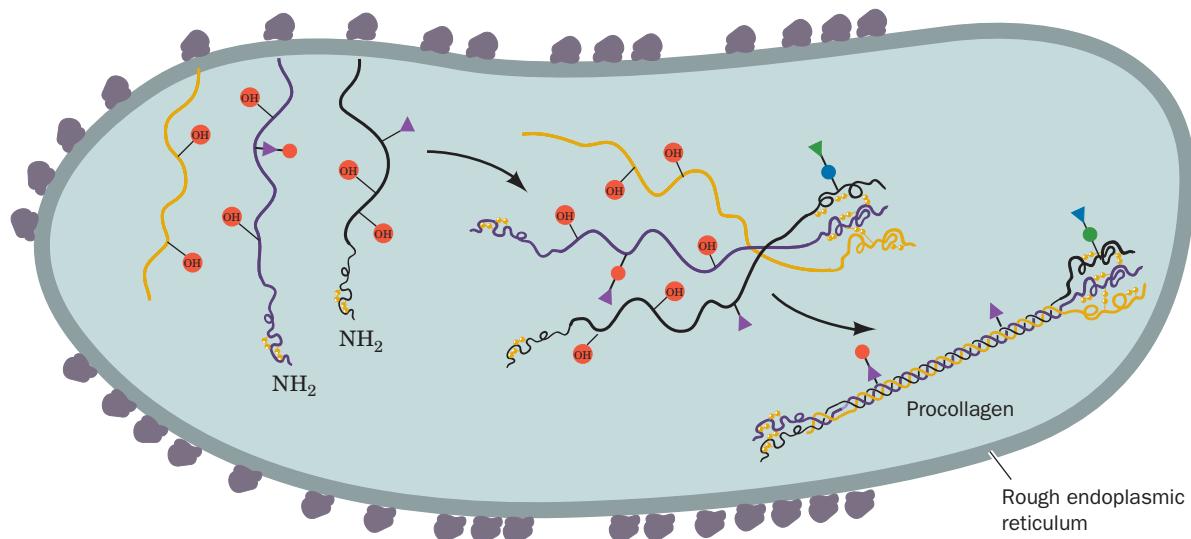


Figure 32-72 Schematic representation of procollagen biosynthesis. Saccharides are represented as in Fig. 32-71. The diagram does not indicate the removal of signal peptides. [After

Prockop, D.J., Kivirikko, K.I., Tuderman, L., and Guzman, N.A., *New Engl. J. Med.* **301**, 18 (1979).]

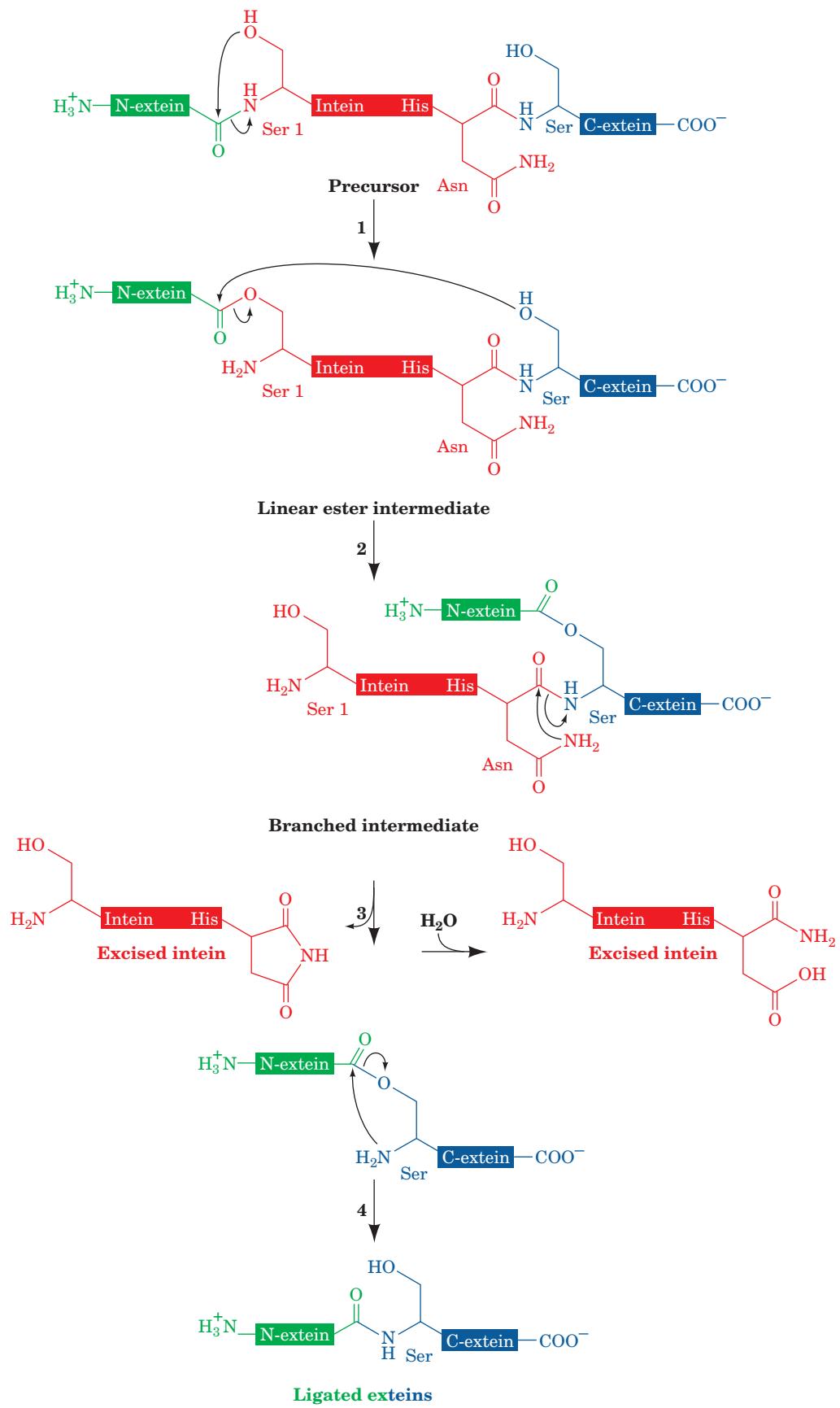


Figure 32-73 Series of reactions catalyzed by inteins to splice themselves out of a polypeptide chain.
See the text for details.

extein on the N- and C-terminal sides of the intein are called the **N-extein** and the **C-extein**. Over 500 putative inteins, ranging in length from 100 to 1650 residues, have so far been identified in archaeabacteria, eubacteria, single-celled eukaryotes, and viruses (and are registered in the Intein Database at <http://www.neb.com/neb/inteins.html>). The various exteins in which these inteins are embedded exhibit no significant sequence similarity and, in fact, can be replaced by other polypeptides, thereby indicating that exteins do not contain the catalytic elements that mediate protein splicing. In contrast, the ~130-residue splicing elements of inteins exhibit significant sequence similarity. All of them have four conserved splice-junction residues: (1) a Ser/Thr/Cys at the intein's N-terminus; and (2 and 3) a His–Asn/Gln dipeptide at the intein's C-terminus; which is immediately followed by (4) a Ser/Thr/Cys at the N-terminus of the C-extein.

Protein splicing occurs via a reaction sequence that involves four successive nucleophilic displacements, the first three of which are mediated by the intein (Fig. 32-73):

1. Attack by the N-terminal intein residue (Ser, Thr, or Cys; shown in Fig. 32-73 as Ser) on its preceding carbonyl group, yielding a linear (thio)ester intermediate.

2. A transesterification reaction in which the —OH or —SH group on the C-extein's N-terminal residue (shown in Fig. 32-73 as Ser) attacks the above (thio)ester linkage, thereby yielding a branched intermediate in which the N-extein has been transferred to the C-extein.

3. Cleavage of the amide linkage connecting the intein to the C-extein by cyclization of the intein's C-terminal Asn or Gln (shown in Fig. 32-73 as Asn). The succinimide ring of the excised intein then spontaneously hydrolyzes to regenerate Asn (or iso-Asn).

4. Spontaneous rearrangement of the (thio)ester linkage between the ligated exteins to yield the more stable peptide bond.

The X-ray structure of the 198-residue **GyrA intein** from *Mycobacterium xenopi*, determined by James Sacchetti, indicates how this intein catalyzes the foregoing splicing reactions. This intein's N-terminal residue, Cys 1, was replaced by an Ala–Ser dipeptide with the expectation that the mutant protein would resemble the intein's pre-splicing state (the new N-terminal residue, Ala 0, presumably represents the C-terminal residue of the N-extein). The X-ray structure reveals that this monomeric protein consists primarily of β strands, two of which curve about the periphery of the entire protein to give it the shape of a flattened horseshoe (Fig. 32-74). The intein's catalytic site is located at the bottom of a broad and shallow cleft near the center of this so-called **β -horseshoe**, where the intein's N-terminal and C-terminal residues are in close proximity. The Ala 0–Ser 1 peptide bond, the bond cleaved in Reaction 1 of the protein splicing process (Fig. 32-73) assumes the cis conformation (Fig. 8-2), a rare high-energy conformation (except when the peptide bond is followed by Pro) that destabilizes this bond. Its amide nitrogen atom is hydrogen bonded to the side chain of the highly conserved

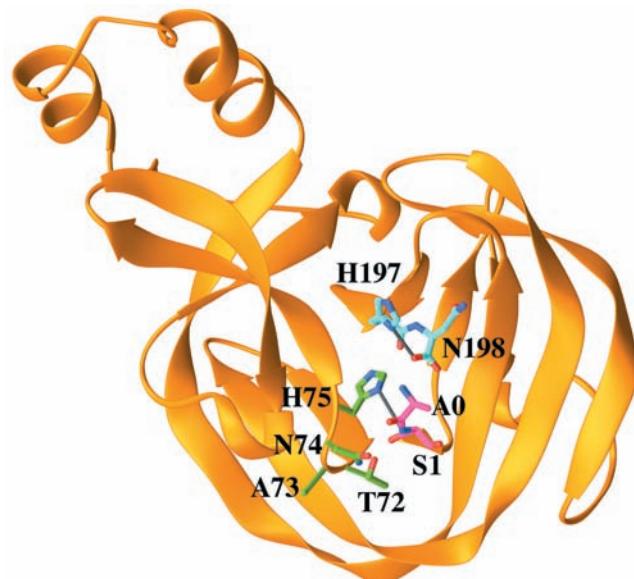


Figure 32-74 X-ray structure of the *M. xenopi* Gyr A intein in which Cys 1 was replaced by an Ala 0–Ser 1 dipeptide. The protein is drawn in ribbon form with its N-terminal Ala 0–Ser 1 dipeptide and its C-terminal His 197–Asn 198 dipeptide as well as the side chains of residues 72 through 75 drawn in stick form colored according to atom type (C of residues 0–1 magenta, C of residues 72–75 green, C of residues 197–198 cyan, N blue, and O red). Hydrogen bonds are represented by thin gray bonds. [Based on an X-ray structure by James Sacchetti, Texas A&M University. PDBid 1AM2.]

His 75. Hence His 75 is well positioned to donate a proton that would promote the breakdown of the tetrahedral intermediate in Reaction 1. The side chains of Thr 72 and Asn 74 appear well positioned to stabilize this tetrahedral intermediate in a manner resembling that of the oxyanion hole in serine proteases (Section 15-3Db). The position of Ser 1 and a modeled Thr at the intein's C-terminus is consistent with Reaction 2 of the splicing process. The side chain of the invariant His 197 is hydrogen bonded to the carboxylate of the C-terminal Asn 198 and hence is positioned to protonate the peptide bond cleaved in Reaction 3.

a. Most Inteins Encode a Homing Endonuclease

What is the biological function of inteins? Nearly all inteins contain polypeptide inserts forming so-called **homing endonucleases**. These are site-specific endonucleases that make a double-strand break in genes that are homologous to their corresponding extein but which lack inteins. The break initiates the double-strand break repair of the DNA via recombination (Section 30-6Ag). Since the intein-containing gene is likely to be the only other gene in the cell containing extein-like sequences, the intein gene is copied into the break. Thus, most inteins mediate a highly specific transposition or “homing” of the genes that insert them in similar sites. The intein's protease and endonuclease activities appear to have a symbiotic relationship: The protease activity excises the intein from the host protein, thereby preventing deleterious effects on the host, whereas the endonuclease

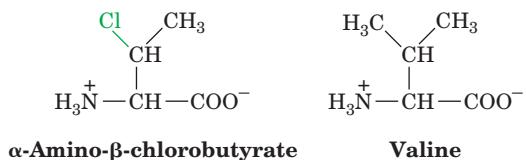
activity assures the mobility of the intein gene. Thus intein genes appear to be molecular parasites (junk DNA) that only function to propagate themselves. Indeed, homing endonucleases are also encoded by certain types of introns.

6 PROTEIN DEGRADATION

The pioneering work of Henry Borsook and Rudolf Schoenheimer around 1940 demonstrated that the components of living cells are constantly turning over. For example, adult humans normally turn over $\sim 2\%$ of their proteins per day. Proteins have lifetimes that range from as short as a few minutes to weeks or more. In any case, *cells continuously synthesize proteins from and degrade them to their component amino acids*. The function of this seemingly wasteful process is 2-fold: (1) to eliminate abnormal proteins whose accumulation would be harmful to the cell, and (2) to permit the regulation of cellular metabolism by eliminating superfluous enzymes and regulatory proteins. Indeed, since the level of an enzyme depends on its rate of degradation as well as its rate of synthesis, *controlling a protein's rate of degradation is as important to the cellular economy as is controlling its rate of synthesis*. In this section we consider the processes of intracellular protein degradation and their consequences.

A. Degradation Specificity

Cells selectively degrade abnormal proteins. For example, hemoglobin that has been synthesized with the valine analog α -amino- β -chlorobutyrate



has a half-life in reticulocytes of ~ 10 min, whereas normal hemoglobin lasts the 120-day lifetime of the red cell (which makes it perhaps the longest lived cytoplasmic protein). Likewise, unstable mutant hemoglobins are degraded soon after their synthesis, which, for reasons explained in Section 10-3A, results in the hemolytic anemia characteristic of these molecular disease agents. Bacteria also selectively degrade abnormal proteins. For instance, *amber* and *ochre* mutants of β -galactosidase have half-lives in *E. coli* of only a few minutes, whereas the wild-type enzyme is almost indefinitely stable. Most abnormal proteins, however, probably arise from the chemical modification and/or spontaneous denaturation of these fragile molecules in the cell's reactive environment rather than by mutations or the rare errors in transcription or translation. *The ability to eliminate damaged proteins selectively is therefore an essential recycling mechanism that prevents the buildup of substances that would otherwise interfere with cellular processes.*

Normal intracellular proteins are eliminated at rates that depend on their identities. A given protein is elimi-

Table 32-11 Half-Lives of Some Rat Liver Enzymes

Enzyme	Half-Life (h)
Short-Lived Enzymes	
Ornithine decarboxylase	0.2
RNA polymerase I	1.3
Tyrosine aminotransferase	2.0
Serine dehydratase	4.0
PEP carboxylase	5.0
Long-Lived Enzymes	
Aldolase	118
GAPDH	130
Cytochrome <i>b</i>	130
LDH	130
Cytochrome <i>c</i>	150

Source: Dice, J.F. and Goldberg, A.L., *Arch. Biochem. Biophys.* **170**, 214 (1975).

nated with first-order kinetics, indicating that the molecules being degraded are chosen at random rather than according to their age. The half-lives of different enzymes in a given tissue vary substantially as is indicated for rat liver in Table 32-11. Remarkably, *the most rapidly degraded enzymes all occupy important metabolic control points, whereas the relatively stable enzymes have nearly constant catalytic activities under all physiological conditions. The susceptibilities of enzymes to degradation have evidently evolved together with their catalytic and allosteric properties so that cells can efficiently respond to environmental changes and metabolic requirements*. The criteria through which native proteins are selected for degradation are considered in Section 32-6B.

The rate of protein degradation in a cell also varies with its nutritional and hormonal state. Under conditions of nutritional deprivation, cells increase their rate of protein degradation so as to provide the necessary nutrients for indispensable metabolic processes. The mechanism that increases degradative rates in *E. coli* is the stringent response (Section 31-3I). A similar mechanism may be operative in eukaryotes since, as happens in *E. coli*, increased rates of degradation are prevented by antibiotics that block protein synthesis.

B. Degradation Mechanisms

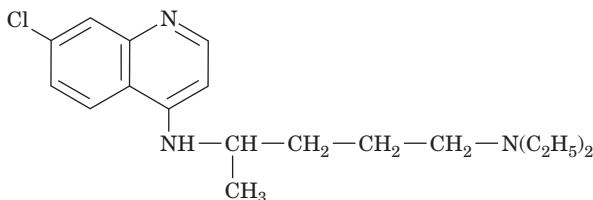
Eukaryotic cells have dual systems for protein degradation: lysosomal mechanisms and ATP-dependent cytosolically based mechanisms. We consider both mechanisms below.

a. Lysosomes Mostly Degrade Proteins Nonselectively

Lysosomes are membrane-encapsulated organelles (Section 1-2Ad) that contain ~ 50 hydrolytic enzymes, including a variety of proteases known as **cathepsins**. The lysosome maintains an internal pH of ~ 5 and its enzymes have acidic pH optima. This situation presumably protects

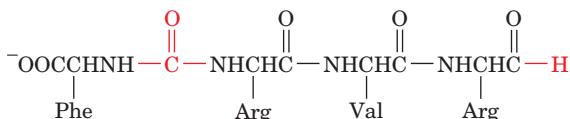
the cell against accidental lysosomal leakage since lysosomal enzymes are largely inactive at cytosolic pH's.

Lysosomes recycle intracellular constituents by fusing with membrane-enclosed bits of cytoplasm known as **autophagic vacuoles** and subsequently breaking down their contents. They similarly degrade extracellular substances that the cell takes up via endocytosis (Section 12-5Bc). The existence of these processes has been demonstrated through the use of lysosomal inhibitors. For example, the antimalarial drug **chloroquine**



Chloroquine

is a weak base that, in uncharged form, freely penetrates the lysosome where it accumulates in charged form, thereby increasing the intralysosomal pH and inhibiting lysosomal function. The treatment of cells with chloroquine reduces their rate of protein degradation. Similar effects arise from treatment of cells with cathepsin inhibitors such as the polypeptide antibiotic **antipain**.



Antipain

Lysosomal protein degradation in well-nourished cells appears to be nonselective. Lysosomal inhibitors do not affect the rapid degradation of abnormal proteins or short-lived enzymes. Rather, they prevent the acceleration of nonselective protein breakdown on starvation. However, the continued nonselective degradation of proteins in starving cells would rapidly lead to an intolerable depletion of essential enzymes and regulatory proteins. Lysosomes therefore also have a selective pathway, which is activated only after a prolonged fast, that takes up and degrades proteins containing the pentapeptide Lys-Phe-Glu-Arg-Gln (KFERQ) or a closely related sequence. Such KFERQ proteins are selectively lost in fasting animals from tissues that atrophy in response to fasting (e.g., liver and kidney) but not from tissues that do not do so (e.g., brain and testes). KFERQ proteins are specifically bound in the cytosol and delivered to the lysosome by a 73-kD **peptide recognition protein (prp73)**, a member of the 70-kD heat shock protein (Hsp70) family (Section 9-2C).

Both normal and pathological processes are associated with increased lysosomal activity. **Diabetes mellitus** (Section 27-4B) stimulates the lysosomal breakdown of proteins. Similarly, muscle wastage caused by disuse, denervation, or traumatic injury arises from increased lysosomal activity. The regression of the uterus after childbirth, in which this muscular organ reduces its mass from 2 kg to

50 g in 9 days, is a striking example of this process. Many chronic inflammatory diseases, such as **rheumatoid arthritis**, involve the extracellular release of lysosomal enzymes that break down the surrounding tissues.

b. Ubiquitin Marks Proteins Selected for Degradation

It was initially assumed that protein degradation in eukaryotic cells is primarily a lysosomal process. Yet, reticulocytes, which lack lysosomes, selectively degrade abnormal proteins. The observation that protein breakdown is inhibited under anaerobic conditions led to the discovery of a cytosolically based ATP-dependent proteolytic system that is independent of the lysosomal system. This phenomenon was thermodynamically unexpected since peptide hydrolysis is an exergonic process.

The analysis of a cell-free rabbit reticulocyte system demonstrated that **ubiquitin** (Fig. 32-75) is required for ATP-dependent protein degradation. This 76-residue monomeric protein, so named because it is ubiquitous as well as abundant in eukaryotes, is the most highly conserved protein known: It is identical in such diverse organisms as humans, toad, trout, and *Drosophila* and differs in only three residues between humans and yeast. Evidently, ubiquitin is all but uniquely suited to an essential cellular function.

Proteins that are selected for degradation are so marked by covalently linking them to ubiquitin. This process, which is reminiscent of amino acid activation (Section 32-2C),

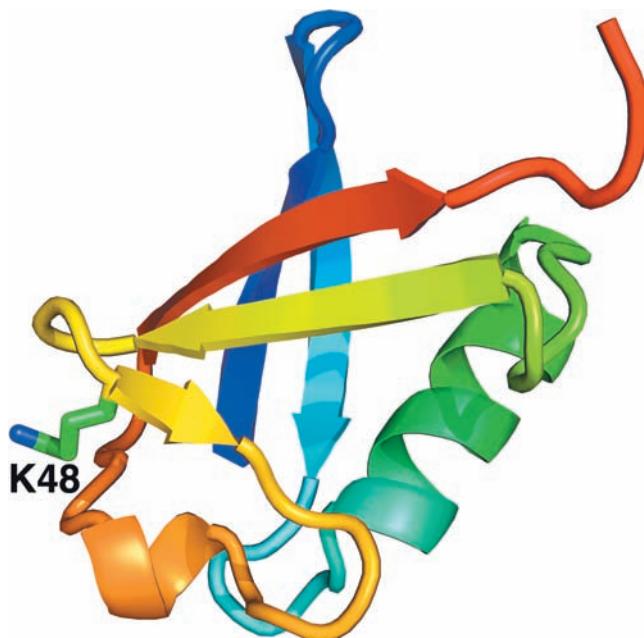


Figure 32-75 X-ray structure of human ubiquitin. The polypeptide is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red) with the side chain of Lys 48 shown in stick form with C green and N blue. This $\alpha + \beta$ architecture is known as a **β -Grasp** fold because its β sheets appear to grasp its α helix. [Based on an X-ray structure by Charles Bugg, University of Alabama at Birmingham. PDBId 1UBO.]  See Interactive Exercise 46.

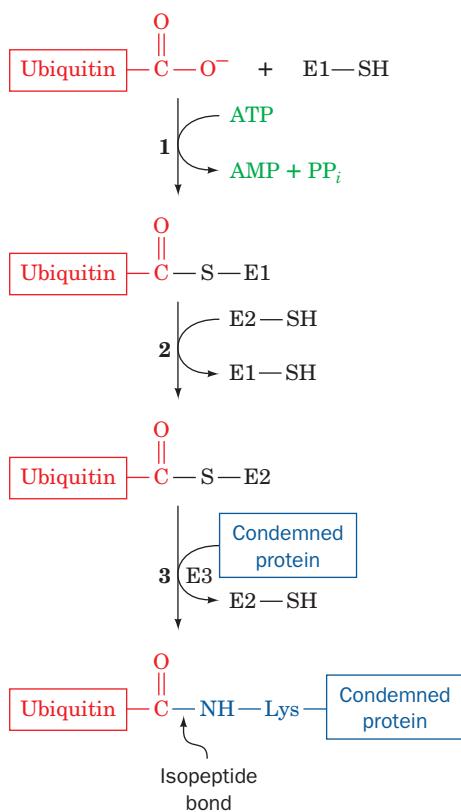


Figure 32-76 Reactions involved in the attachment of ubiquitin to a protein. In the first part of the process, ubiquitin's terminal carboxyl group is joined, via a thioester linkage, to E1 in a reaction driven by ATP hydrolysis. The activated ubiquitin is subsequently transferred to a sulfhydryl group of an E2 and then, in a reaction catalyzed by an E3, to a Lys ϵ -amino group on a condemned protein, thereby flagging the protein for proteolytic degradation by the 26S proteasome.

occurs in a three-step pathway that was elucidated notably by Avram Hershko, Aaron Ciechanover, and Irwin Rose (Fig. 32-76):

1. In an ATP-requiring reaction, ubiquitin's terminal carboxyl group is conjugated, via a thioester bond, to **ubiquitin-activating enzyme (E1)**, a homodimer of \sim 1050-residue subunits. In this process, the substrate protein's terminal carboxyl group is initially adenylated and then transferred to the E1 Cys—SH group with the elimination of AMP. Most organisms, including yeast and humans, have only one type of E1.

2. The ubiquitin is then transferred to a specific Cys sulfhydryl group on one of numerous proteins named **ubiquitin-conjugating enzymes (E2s)** (11 in yeast and over 20 in mammals). The various E2's are characterized by \sim 150-residue catalytic cores containing the active site Cys that exhibit at least 25% sequence identities and which mainly vary by the presence or absence of N- and/or C-terminal extensions that exhibit little sequence identity to each other. The X-ray and NMR structures of several species of E2 reveal that their catalytic cores all assume closely similar α/β structures (e.g., Fig. 32-77) in which

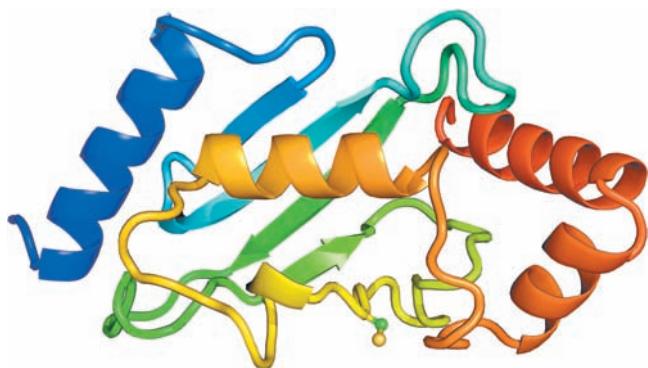


Figure 32-77 X-ray structure of an E2 protein from *Arabidopsis thaliana*. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chain of Cys 88, to which ubiquitin is covalently linked, is shown in ball-and-stick form with C green and S yellow. [Based on an X-ray structure by William Cook, University of Alabama at Birmingham. PDBid 2AAK.]

most of the identical residues are clustered on one surface near the ubiquitin-accepting Cys residue, where they presumably interact with ubiquitin and E1.

3. **Ubiquitin-protein ligase (E3)** transfers the activated ubiquitin from E2 to a Lys ϵ -amino group of its target protein, thereby forming an **isopeptide bond**. Each of the many E3s present in eukaryotic cells mediates the ubiquitination (alternatively, ubiquitylation) of a specific set of proteins and thereby marks them for degradation. Each E3 is served by one or a few specific E2s. Most E3s are members of two unrelated families, those containing a **HECT** domain (HECT for homologous to *E6AP* C-terminus) and those containing a so-called **RING** domain (also called a **RING-finger** domain; RING for really interesting new gene), although some E2s react well with members of both families. The human genome contains 28 HECT genes and 616 RING genes, more than its number of protein kinase genes (518), which is indicative of the specialized and varied functions of E3s (although not all RING domain-containing proteins are E3s). HECT domain E3s are modularly constructed with a unique N-terminal domain that interacts with its target proteins via their so-called **ubiquitination signals** (usually short polypeptide segments; see below) and an \sim 350-residue HECT domain that mediates E2 binding and catalyzes the ubiquitination reaction. RING domains, which are implicated in recognizing a substrate protein's ubiquitination signal, are 40- to 60-residue motifs that each bind two structurally but not catalytically implicated Zn²⁺ ions via a total of 8 Cys and His residues in a characteristic consensus sequence (much like the zinc finger motifs in certain DNA-binding proteins, Section 34-3B). RING domain-containing E3s may consist of a single subunit or may be multisubunit proteins in which the RING domain is contained in one subunit. HECT E3-mediated ubiquitination occurs via the transfer of ubiquitin from E2 to a conserved Cys residue on the HECT domain followed by its transfer to the substrate protein Lys side chain. In contrast, RING domain E3s act

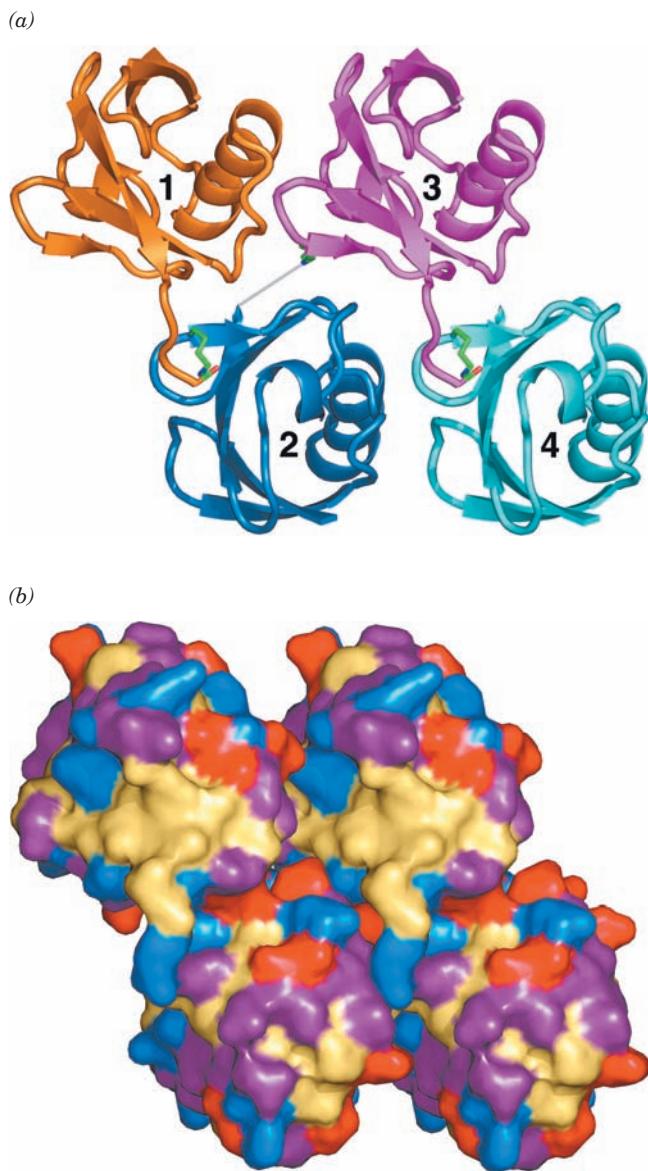


Figure 32-78 X-ray structure of human tetraubiquitin. (a) A ribbon drawing in which the isopeptide bonds connecting successive ubiquitin molecules, together with the Lys side chains making them, are drawn in stick form with C green, N blue, and O red. However, since the C-terminal three residues of ubiquitin 2 are disordered, the isopeptide bond connecting it to ubiquitin 3 is represented by a thin gray bond (this isopeptide bond nevertheless exists, as was demonstrated by SDS-PAGE of dissolved crystals). The monomer units in a multiubiquitin chain of any length are likely to be arranged with the repeating symmetry of the tetraubiquitin structure, although the weak interactions between adjacent ubiquitin units suggests that this chain is conformationally flexible. (b) A surface diagram, viewed as in Part a, in which basic residues (Arg, Lys, His) are blue, acidic residues (Asp, Glu) are red, uncharged polar residues (Gly, Ser, Thr, Asn, Gln) are purple, and hydrophobic residues (Ile, Leu, Val, Ala, Met, Phe, Tyr, Pro) are tan (ubiquitin lacks Cys and Trp residues). Note the unusually large solvent-exposed surface occupied by the hydrophobic residues. [Based on an X-ray structure by William Cook, University of Alabama at Birmingham, and Cecile Pickart, Johns Hopkins University. PDBid 1TBE.]

as adapters that position the reactive E2-ubiquitin thioester bond for the direct transfer of the ubiquitin to the substrate protein Lys side chain.

In order for a target protein to be efficiently degraded, it must be linked to a chain of at least four tandemly linked ubiquitin molecules in which Lys 48 of each ubiquitin forms an isopeptide bond with the C-terminal carboxyl group of the succeeding ubiquitin (Fig. 32-78). These **polyubiquitin (polyUb)** chains, which can reach lengths of 50 or more ubiquitin molecules, are generated by the E3s, although how they switch from transferring a ubiquitin to the target protein to processively synthesizing a polyubiquitin chain is unknown.

c. Ubiquitinated Proteins Are Hydrolyzed in the Proteasome

A ubiquitinated protein is proteolytically degraded to short peptides in an ATP-dependent process mediated by a large (2000 kD, 26S) multisubunit protein complex named the **26S proteasome** (sometimes spelled “proteosome”) that electron micrographic studies reveal has the shape of a bi-capped hollow barrel (Fig. 32-79). Proteolysis occurs inside the barrel, which permits this process to be extensive and processive, while preventing nonspecific proteolytic damage to other cellular components. PolyUb chains are the signals that target a protein to the proteasome; the identity of the target protein has little effect on the efficiency with which it is degraded by the proteasome. Nevertheless, the proteasome does not degrade ubiquitin molecules; they are returned to the cell. The size and functional complexity of this entire proteolytic system, which occurs in the nucleus as well as the cytosol, rivals that of the ribosome (Section 32-3) and the spliceosome (Section 31-4A) and hence is

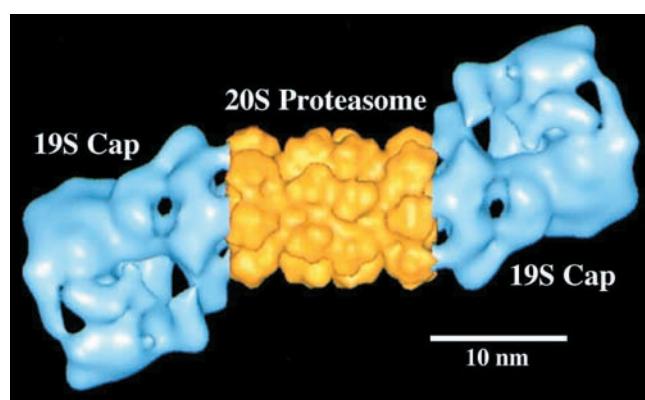


Figure 32-79 Electron microscopy-based image of the *Drosophila melanogaster* 26S proteasome. The complex is around $450 \times 190 \text{ \AA}$. The central portion of this 2-fold symmetric multiprotein complex (yellow), the **20S proteasome**, consists of four stacked 7-membered rings of subunits that form a hollow barrel in which the proteolysis of ubiquitin-linked proteins occurs. The **19S caps** (cyan), which may attach to one or both ends of the 20S proteasome, control the access of condemned proteins to the 20S proteasome (see below). [Courtesy of Wolfgang Baumeister, Max Planck Institute of Biochemistry, Martinsreid, Germany.]

indicative of the importance of properly managing protein degradation. Indeed, ~5% of the proteins expressed by yeast participate in protein degradation. We discuss the structure and function of the 26S proteasome below.

d. Many E3s Have Elaborate Modular Structures

The proto-oncogene product **c-Cbl** (906 residues) is a single-subunit, RING domain-containing E3 that functions to ubiquitinate certain activated receptor tyrosine kinases (RTKs; Section 19-3A), thereby terminating their signaling. Nikola Pavletich determined the X-ray structure of the N-terminal half of c-Cbl (residues 47–447) in its ternary complex with the E2 protein **UbcH7** (which consists of little more than the ~150-residue E2 catalytic core) and a 9-residue peptide containing the ubiquitination signal from a nonreceptor tyrosine kinase (NRTK) named **ZAP-70** (Fig. 19-44). The structure (Fig. 32-80) reveals that UbcH7 and c-Cbl's RING domain and SH2-containing tyrosine kinase-binding (TKB) domains interact with one another across multiple interfaces to form a compact and apparently rigid structure. The RING domain consists of a 3-stranded β sheet, an α helix, and two large loops that are held together by two tetrahedrally coordinated Zn^{2+} ions. UbcH7 adopts the characteristic α/β fold of other E2s of known structure (e.g., Fig. 32-77). The ZAP-70 peptide is bound on the opposite side of the TKB domain from the UbcH7 active site Cys residue (Cys 86) and is ~60 Å distant from it.

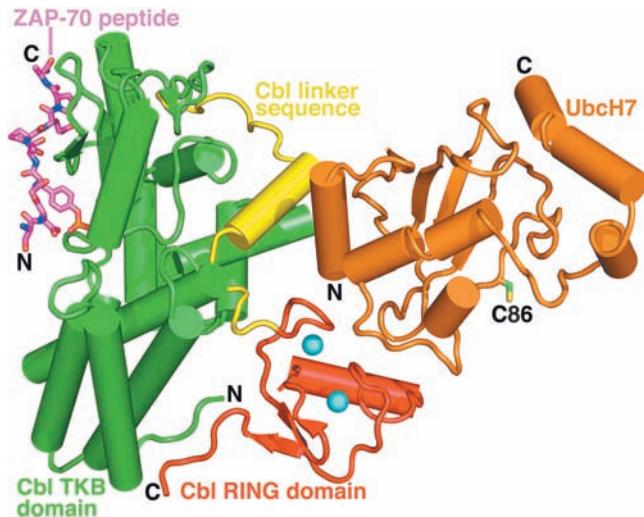


Figure 32-80 X-ray structure of the human c-Cbl–UbcH7–ZAP-70 peptide ternary complex drawn in tube-and-arrow form.

UbcH7, an E2 that consists almost entirely of the E2 catalytic core, is colored orange with the side chain of its active site Cys 86 shown in stick form with C green and S yellow. c-Cbl (residues 47–447 of the 903-residue protein), a monomeric RING E3, is colored according to domain with its TKB domain green, its RING domain red, and the linker joining them yellow. The RING domain's two bound Zn^{2+} ions are represented by cyan spheres. The 9-residue ubiquitination site of the RTK ZAP-70, whose fourth residue is phospho-Tyr, is drawn in stick form with C magenta, N blue, and O red, and P orange. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1FBV.]

SCF complexes are multisubunit RING E3s that consist of **Cul1** (a member of the **cullin** family; 776 residues), **Rbx1** (which contains the complex's RING domain; 108 residues), **Skp1** (163 residues), and a member of the **F-box protein** family (~430 to >1000 residues; SCF for Skp1–cullin–F-box protein). Rbx1 and Cul1 form the complex's catalytic core that binds E2; F-box proteins consist of an ~40-residue **F-box** that binds Skp1 followed by protein–protein interaction modules such as **leucine-rich repeats (LRRs)** or WD40 repeats (Section 19-2C) that bind substrate protein; and Skp1 functions as an adapter that links the F-box to Cul1. Cells contain numerous different F-box proteins (at least 38 in humans) that presumably permit the specific ubiquitination of a diverse variety of protein substrates (see below).

Pavletich has also determined the X-ray structures of two segments of the **SCF^{Skp2}** complex (where the superscript identifies the complex's F-box protein, here **Skp2**, 436 residues). The structure of the Skp1–Skp2 complex (Fig. 32-81) reveals that it has the shape of a sickle with the Skp1 and the 3-helix F-box of Skp2 forming the handle and its 10 LRRs (~26 residues each) forming the curved blade. The structure of the Cul1–Rbx1–Skp1–F-box^{Skp2} quaternary complex (Fig. 32-82) shows that Cul1 is an elongated protein that consists of a long stalk formed by three repeats of a novel five-helix motif known as a cullin repeat followed by a globular domain that binds Rbx1. Apparently Cul1 acts like a rigid scaffold that organizes the Skp1–F-box^{Skp2} complex and Rbx1 so as to hold them over 100 Å

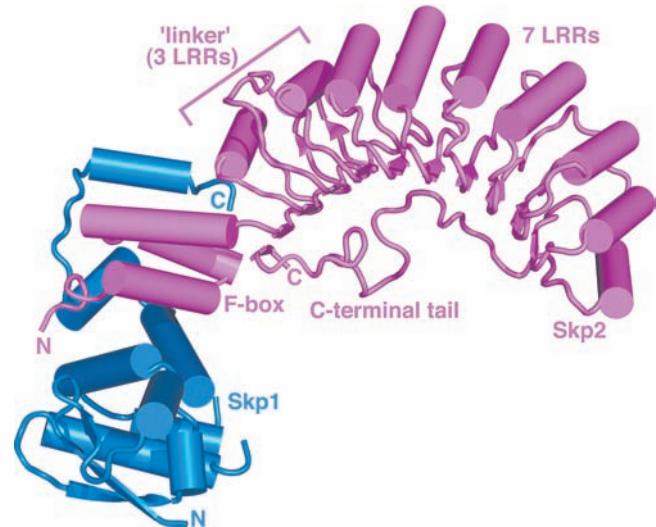


Figure 32-81 X-ray structure of the human Skp1–Skp2 complex. Skp1 and Skp2 are drawn in tube-and-arrow form in blue and magenta. Skp2 consists of an N-terminal F-box that forms three helices, followed by 3 noncanonical so-called linker leucine-rich repeats (LRRs) that are contiguous with 7 LRRs that were predicted from their amino acid sequences for a total of 10 LRRs. After the tenth LRR, Skp2's ~30-residue C-terminal tail extends back past the first LRR by packing under the concave surface of the LRR domain. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1FQV.]

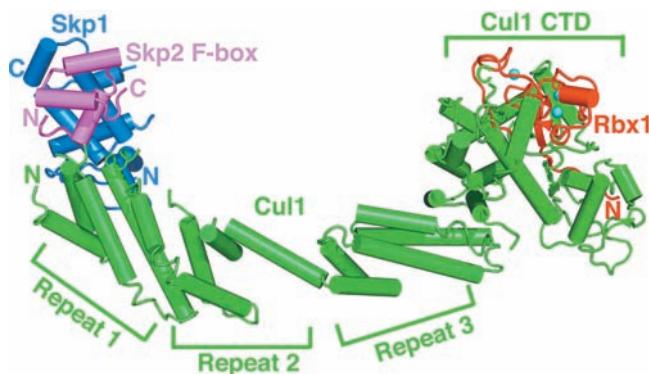


Figure 32-82 X-ray structure of the human Cul1-Rbx1-Skp1-F-box^{Skp2} quaternary complex.

Cul1-Rbx1-Skp1-F-box^{Skp2} quaternary complex. Cul1, Rbx1, Skp1, and the Skp2 F-box are drawn in tube-and-arrow form and respectively colored green, red, blue, and magenta. The three cullin repeats of Cul1 are indicated. The three Zn²⁺ ions bound to Rbx1 are represented by cyan spheres. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1LDK.]

apart. The Rbx1 RING domain contains a 20-residue insert that forms the binding site for a third tetrahedrally liganded Zn²⁺ ion.

The apparent rigidity of the foregoing three structures has enabled Pavletich to construct a model of the intact SCF^{Skp2}-E2 complex by superimposing Skp1-Skp2 on Cul1-Rbx1-Skp1-F-box^{Skp2} and docking the E2 UbcH7 onto the Rbx1 RING domain based on the c-Cbl-UbcH7 structure (Fig. 32-83). The model indicates that E2 and the LRR-containing domain of Skp2 are on the same side of the SCF complex but separated by a distance of ~50 Å.

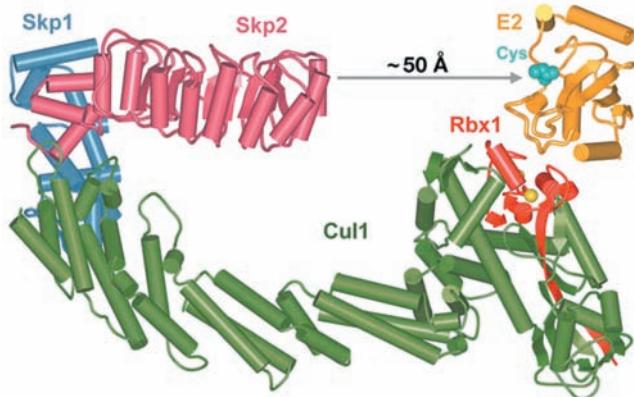


Figure 32-83 Model of the SCF^{Skp2}-E2 complex. This model, which is based on the X-ray structures in Figs. 32-80, 32-81, and 32-82, is colored and viewed as in Fig. 32-82. E2 is yellow with its active site Cys residue, to which ubiquitin would be covalently linked, drawn in space-filling form in cyan. The Zn²⁺ ions associated with the Rbx1 RING domain are represented by yellow spheres. The gray arrow indicates the 50-Å gap between the tip of the Skp2 LRR domain and the E2 active site. [Courtesy of Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York.]

This suggests that Cul1's long stalk functions to separate the complex's substrate-binding and catalytic sites so that substrates with different sizes and various distances between their ubiquitinated Lys residues and their ubiquitination signals can be accommodated.

e. The Ubiquitin System Has Both Housekeeping and Regulatory Functions

Until the mid-1990s, it appeared that the ubiquitin system functioned mainly in a “housekeeping” capacity to maintain the proper balance among metabolic proteins and to eliminate damaged proteins. Indeed, as Alexander Varshavsky discovered, *the half-lives of many cytoplasmic proteins vary with the identities of their N-terminal residues* (Table 32-12). Thus, in a selection of 208 cytoplasmic proteins known to be long lived, all have a “stabilizing” residue, Met, Ser, Ala, Thr, Val, or Gly, at their N-termini. This so-called **N-end rule** is true for both eukaryotes and prokaryotes, which suggests the system that selects proteins for degradation is conserved in eukaryotes and prokaryotes, even though prokaryotes lack ubiquitin. The N-end rule results from the actions of the single-subunit, RING E3 named **E3α** (~1950 residues; also known as **Ubr1**) whose ubiquitination signals are the destabilizing N-terminal residues in Table 32-12.

Similarly, it has long been known that proteins with segments rich in Pro (P), Glu (E), Ser (S), and Thr (T), the so-called **PEST proteins**, are rapidly degraded. This is because these PEST elements often contain phosphorylation sites that target their proteins for ubiquitination.

Table 32-12 Half-Lives of Cytoplasmic Enzymes as a Function of Their N-Terminal Residues

N-Terminal Residue	Half-Life
<i>Stabilizing</i>	
Met	>20 h
Ser	
Ala	
Thr	
Val	
Gly	
<i>Destabilizing</i>	
Ile	~30 min
Glu	
Tyr	~10 min
Gln	
<i>Highly Destabilizing</i>	
Phe	~3 min
Leu	
Asp	
Lys	
Arg	~2 min

Source: Bachmair, A., Finley, D., and Varshavsky, A., *Science* **234**, 180 (1986).

It is now clear, however, that the ubiquitin system is far more sophisticated than a simple garbage disposal system. Thus, the known E3s each respond to certain ubiquitination signals that often occur on a quite limited range of target proteins, many of which have regulatory functions. For example, *the ubiquitination system has an essential function in cell cycle progression*. The cell cycle, as we have seen in Section 30-4Aa and will further discuss in Section 34-4D, is regulated by a series of proteins known as cyclins. A given cyclin, which is expressed immediately preceding and/or during a specific phase of the cell cycle, binds to a corresponding **cyclin-dependent protein kinase (Cdk)**, which then phosphorylates its target proteins so as to activate them to carry out the processes of that phase of the cell cycle. Moreover, many cyclins also inhibit the transition to the subsequent phase of the cell cycle (e.g., DNA replication or mitosis). Consequently, for a cell to progress from one phase of the cell cycle to the next, the cyclin(s) governing that phase must be eliminated. This occurs via the specific ubiquitination of the cyclin, thereby condemning it to be destroyed by the proteasome. The E3s responsible for this process are the SCF complexes containing F-box proteins targeted to a corresponding cyclin and a multisubunit complex known as the **anaphase-promoting complex (APC)**; alternatively the **cyclosome**; Section 34-4Da). APC, an ~1500-kD RING domain-containing particle that in yeast consists of 11 subunits, specifically ubiquitinates proteins that contain the 9-residue consensus sequence RTALGDIGN, the so-called **destruction box**, near their N-termini.

The transcription factor **NF- κ B**, which plays a central role in immune and inflammatory responses (Section 34-3Bs), is maintained in an inactive state in the cytosol through its binding to the inhibitor **I κ B α** in a way that occludes the short internal basic sequence that directs NF- κ B's import into the nucleus (its **nuclear localization signal; NLS**). However, the stimulation of cell-surface receptors by proinflammatory cytokines such as **tumor necrosis factor- α (TNF α** ; Section 19-3Db) and **interleukin-1 (IL-1**; Section 19-3Eb) initiate a signal transduction pathway (Section 19-3D) that phosphorylates I κ B α bound to NF- κ B at both Ser residues in the sequence DSGLDS. This phosphorylated sequence is the ubiquitination signal for the SCF complex containing the F-box protein **β -TrCP** (605 residues), which mediates the ubiquitination of the phosphorylated I κ B α . The consequent destruction of I κ B α exposes the NLS of NF- κ B, which is then translocated to the nucleus where it activates the transcription of its target genes (Section 34-3Bs).

Some viruses usurp the ubiquitin system. Oncogenic forms of **human papillomavirus (HPV)**, the cause of nearly all cervical cancers (a leading cause of death of women in developing countries), encode the ~150-residue **E6 protein**, which combines with the 875-residue cellular protein named **E6-associated protein (E6AP)**; the first E3 known to contain a HECT domain) to ubiquitinate **p53**, thereby marking it for destruction. This latter protein is a transcription factor that monitors genome integrity and

hence is important in preventing malignant transformation and the proliferation of cancer cells (Section 34-4Ca), that is, it is a **tumor suppressor** (a protein whose loss of function is a cause of cancer). Consequently, HPV provokes the uncontrolled growth of the cells it infects and hence its own proliferation. E6AP normally functions to ubiquitinate certain members of the Src family of protein tyrosine kinases (Section 19-3Ba), including Src itself. The deletion of the segment of chromosome 15 that contains the E6AP gene causes Angelman syndrome, which as we have seen (Section 30-7d) is characterized by severe mental retardation and is exclusively maternally inherited due to genomic imprinting.

The foregoing are only a few examples of the numerous cellular processes that are regulated by the ubiquitin-mediated proteolysis system. Not surprisingly, therefore, many pathological conditions in humans, including inflammatory, neurodegenerative, and muscle-wasting diseases, are attributable to malfunctioning ubiquitination systems.

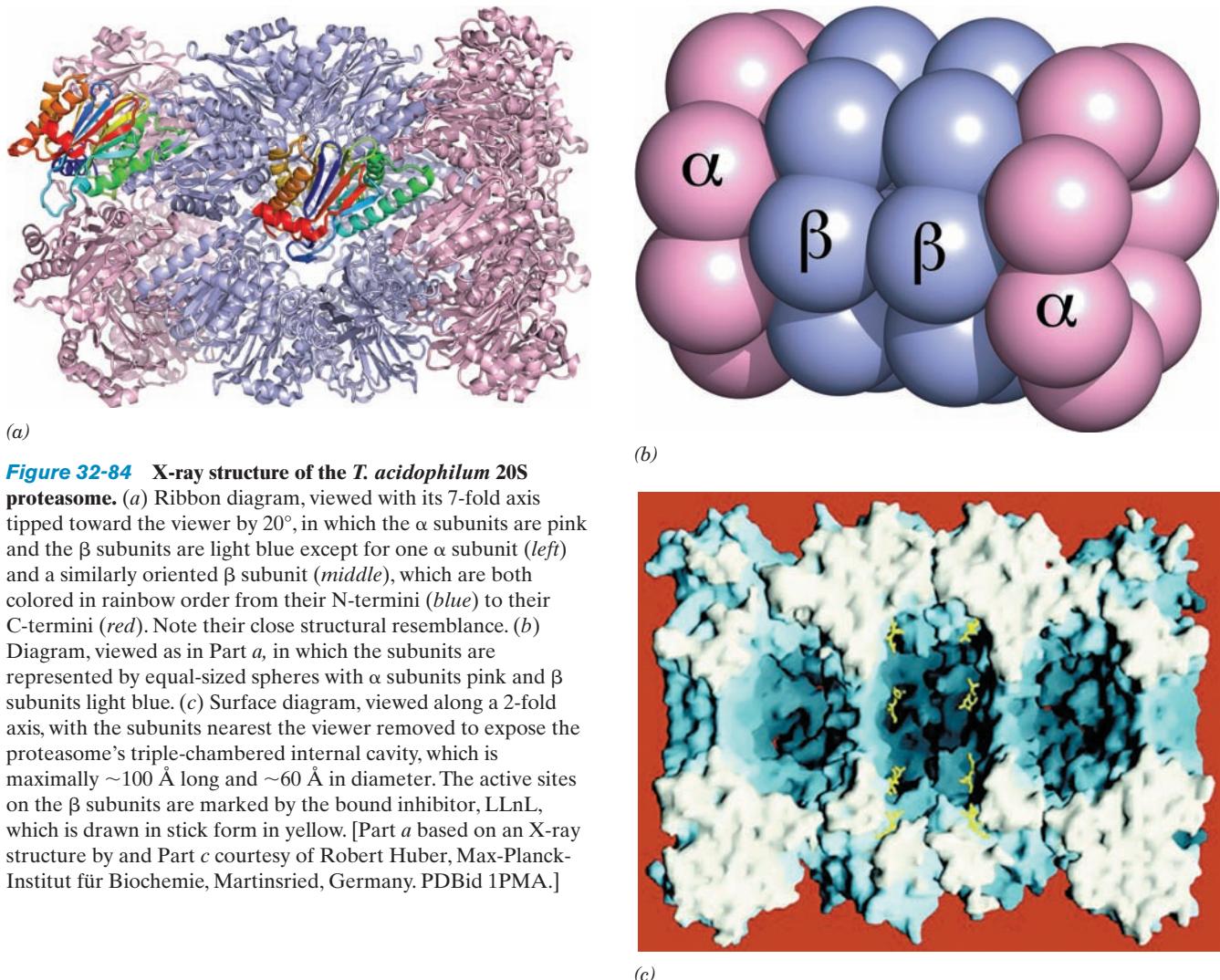
f. The 20S Proteasome Catalyzes Proteolysis Inside a Hollow Barrel

The 26S proteasome (Fig. 32-79) is an ~2100-kD multisubunit protein that catalyzes the ATP-dependent hydrolysis of ubiquitin-linked proteins. This yields oligopeptides with lengths of 4 to 25 residues and averaging 7 to 9 residues that are subsequently degraded to their component amino acids by cytosolic exopeptidases. The 26S proteasome consists of a **20S proteasome** (~670 kD), the barrel-shaped catalytic core of the 26S proteasome, and its **19S caps** (~700 kD; also known as **PA700** and the **19S regulator**), which associate with the ends of the 20S proteasome and stimulate its activity (PA for *proteasome activator*). The 20S proteasome only hydrolyzes unfolded proteins in an ATP-independent manner; the 19S caps function to identify and unfold the ubiquitinated protein substrates.

The 20S proteasome occurs in the nuclei and cytosol of all eukaryotic cells and in all archaeabacteria yet examined. However, the only eubacteria in which it occurs are those of the class Actinobacteria, which suggests that they obtained it via horizontal gene transfer from some other organism.

The 20S proteasome of *Thermoplasma acidophilum* (an archaeabacterium) consists of 14 copies each of α and β subunits (233 and 203 residues) that electron microscopy studies revealed form a 150- \AA long and 110- \AA -diameter barrel in which the subunits are arranged in four stacked rings (as is evident in the central portion of the 26S proteasome seen in Fig. 32-79). The α and β subunits are 26% identical in sequence except for an ~35-residue N-terminal tail of the α subunit, which the β subunit lacks. Eukaryotic 20S proteasomes are more complex in that they consist of 7 different α -like and 7 different β -like subunits versus only one of each type for the *T. acidophilum* 20S proteasome.

The X-ray structure of the *T. acidophilum* 20S proteasome, determined by Baumeister and Robert Huber, reveals that its two inner rings each consist of 7 β subunits and its two outer rings each consist of 7 α subunits, all arranged with D_7 symmetry (Fig. 32-84). Thus the overall



(a)

(b)

(c)

Figure 32-84 X-ray structure of the *T. acidophilum* 20S proteasome. (a) Ribbon diagram, viewed with its 7-fold axis tipped toward the viewer by 20°, in which the α subunits are pink and the β subunits are light blue except for one α subunit (left) and a similarly oriented β subunit (middle), which are both colored in rainbow order from their N-termini (blue) to their C-termini (red). Note their close structural resemblance. (b) Diagram, viewed as in Part a, in which the subunits are represented by equal-sized spheres with α subunits pink and β subunits light blue. (c) Surface diagram, viewed along a 2-fold axis, with the subunits nearest the viewer removed to expose the proteasome's triple-chambered internal cavity, which is maximally ~ 100 Å long and ~ 60 Å in diameter. The active sites on the β subunits are marked by the bound inhibitor, LLNL, which is drawn in stick form in yellow. [Part a based on an X-ray structure by and Part c courtesy of Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBid 1PMA.]

structure of the 20S proteasome superficially resembles that of the unrelated molecular chaperone GroEL (Section 9-2Ca). The structures of the α and β subunits are remarkably similar (Fig. 32-84a) except, of course, for the α subunit's N-terminal tail (blue in Fig. 32-84a), which extends radially inward to contact the N-terminal tail of a neighboring α subunit. This accounts for the observation that α subunits alone spontaneously assemble into 7-membered rings (a capacity that is abolished by the deletion of their N-terminal 35 residues), whereas β subunits alone remain monomeric.

The central cavity of the *T. acidophilum* 20S proteasome consists of three large chambers (Fig. 32-84c): Two are located at the interfaces between adjoining rings of α and β subunits, with the third, larger chamber centrally located between the two rings of β subunits. Unfolded polypeptide substrates enter the central chamber of the barrel (where the proteasome's active sites are located; see below) through ~ 13 -Å-diameter axially located apertures in the α rings that are lined with hydrophobic residues. This allows only unfolded

proteins to enter the central chamber, thereby protecting properly folded proteins from indiscriminate degradation by this omnivorous protein-dismantling machine.

The X-ray structure of the yeast 20S proteasome, determined by Huber, demonstrates that its outer and inner rings respectively consist of seven different α -type subunits and seven different β -type subunits, all of which are uniquely arranged (Fig. 32-85). The α -like subunits have folds that are similar to one another as well as to that of the *T. acidophilum* 20S proteasome and likewise for the β -like subunits. Consequently, this 28-subunit, 6182-residue protein complex has exact 2-fold rotational symmetry relating its two pairs of rings but only pseudo-7-fold rotational symmetry relating the subunits within each ring. The narrow axial apertures in the α rings through which unfolded polypeptides enter the hydrolytic chamber (Fig. 32-85d) are occluded in the closed state (Fig. 32-85c) by a plug formed by the interdigititation of its α subunits' N-terminal tails. This indicates that the 19S caps of the 26S proteasome, which have been shown to activate the 20S proteasome,

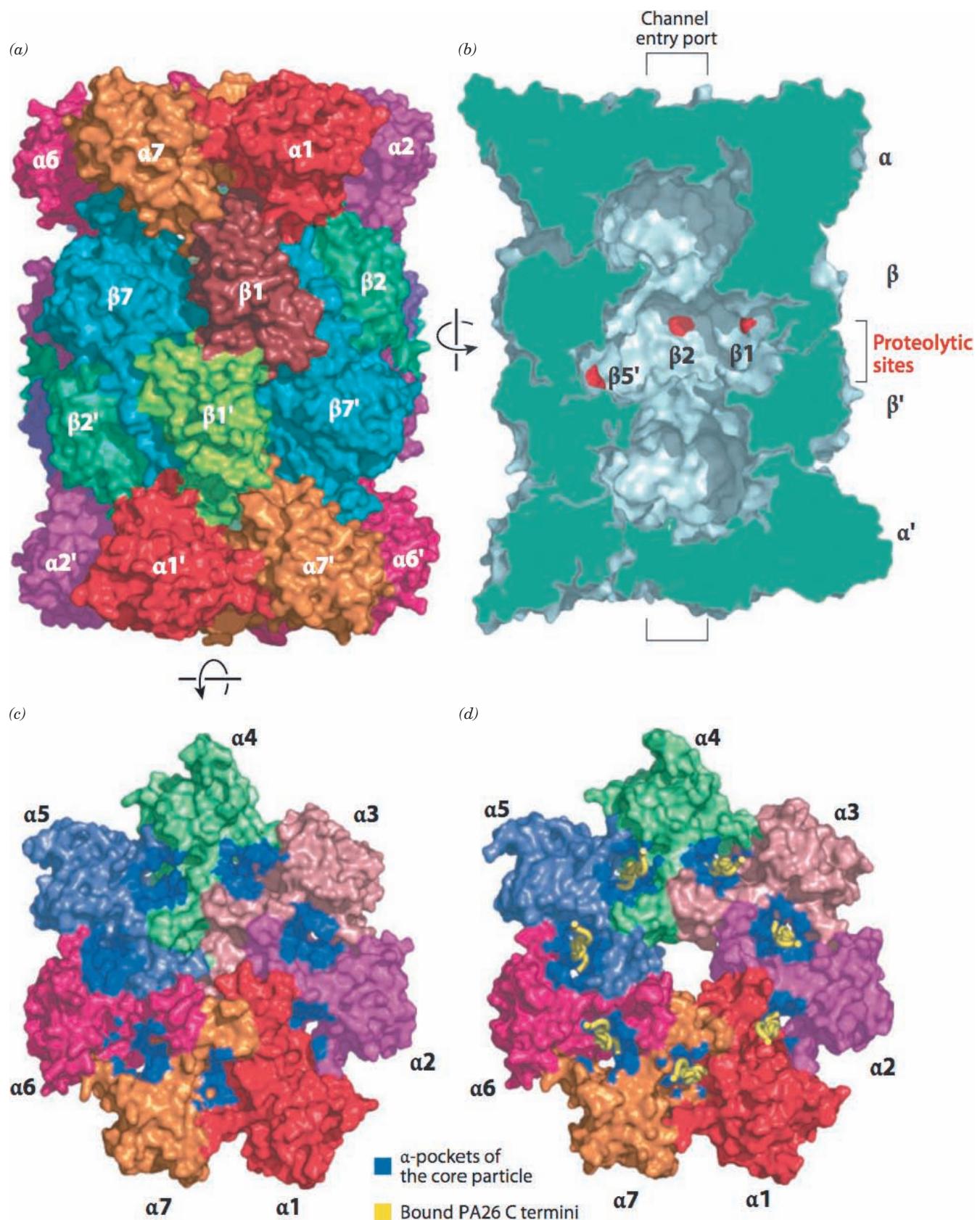
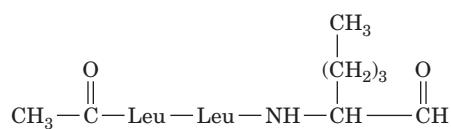


Figure 32-85 (Opposite) X-ray structure of the yeast 20S proteasome. (a) Surface diagram viewed along the 28-mer's 2-fold axis. Each pair of identical subunits has the same color except for β 1 and the symmetry related β 1', which are colored differently. (b) Cutaway surface diagram sliced along the complex's pseudo-7-fold axis. The slice surface is green and the active sites of the β 1, β 2, and β 5' subunits are marked in red. The brackets indicate the approximate positions of the channel entry ports as seen in the open state. (c) End view of the 20S proteasome showing its α -ring in its closed state represented as in Part a. The pockets where PA26 binds (see below) are highlighted in blue. (d) The open state represented as in Part c and with the C-termini of its bound PA26 subunits shown in worm form in yellow. [Courtesy of Daniel Finley, Harvard Medical School. Based on X-ray structures by Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany (closed state), and Christopher Hill, University of Utah (open state). PDBIDs 1RYP and 1FNT.]

control the access to it by inducing conformational changes in its α rings (see below). The X-ray structure of the bovine 20S proteasome, determined by Tomonari Tsukihara, reveals that its arrangement of seven α -type and seven β -type subunits is similar to that in yeast.

g. The Proteasome Catalyzes Peptide Hydrolysis via a Novel Mechanism

The X-ray structure of the *T. acidophilum* 20S proteasome in complex with the aldehyde inhibitor **acetyl-Leu-Leu-norleucinal (LLnL)**



reveals that its active sites are on the inner surfaces of its rings of β subunits, with the aldehyde function of the LLnL close to the side chain of the highly conserved Thr 1 β . Deletion of this Thr or its mutation to Ala yields properly assembled 20S proteasomes that are completely inactive. Evidently, 20S proteasomes catalyze peptide hydrolysis by a novel mechanism in which the hydroxyl group of its Thr 1 β is the attacking nucleophile. This, as yet, poorly understood mechanism, in which the amino group at the N-terminus and possibly a bound water molecule act to nucleophilically activate the hydroxyl side chain, is now known to be employed by other hydrolases (e.g., glutamate synthase; Section 26-5Aa), which are collectively known as the **N-terminal nucleophile (Ntn)** family of hydrolases. The *T. acidophilum* β subunits preferably cleave polypeptides after hydrophobic residues. However, in the yeast and the bovine 20S proteasomes, only subunits β 1, β 2, and β 5 are catalytically active. Their respective preferences for cleavage after acidic (caspase-like), basic (trypsin-like), and hydrophobic (chymotrypsin-like) residues are explained by

the respective basic, acidic, and nonpolar characters of their pockets that bind the side chain of the residue preceding the scissile peptide bond, although this specificity is relatively low. The functions of the four different catalytically inactive β subunits are unknown, although mutagenically modifying an inactive β subunit can abolish the catalytic activity of an active β subunit.

h. The 19S Caps Control the Access of Ubiquitinated Proteins to the 20S Proteasome

The 20S proteasome probably does not exist alone *in vivo*; it is most often in complex with two 19S caps that function to recognize ubiquitinated proteins, unfold them, and feed them into the 20S proteasome in an ATP-dependent manner (it may also associate with other regulatory complexes; see below). The 19S cap, which consists of ~18 different subunits, is poorly characterized due in large part to its low intrinsic stability. Its so-called base complex consists of 9 different subunits, 6 of which are ATPases that form a ring that abuts the α ring of the 20S proteasome (Fig. 32-79). Each of these ATPases contains an ~230-residue ATPase module that is a member of the AAA+ family (Section 30-2Ca). Cecile Pickart demonstrated via cross-linking experiments that one of these ATPases, named **S6'** (alternatively **Rpt5**), contacts the polyUb signal that targets a condemned protein to the 26S proteasome. This suggests that the recognition of the polyUb chain as well as substrate protein unfolding are ATP-driven processes. Moreover, the ring of ATPases must function to open (gate) the otherwise closed axial aperture of the 20S proteasome so as to permit the entry of the unfolded substrate protein.

Eight additional subunits form the so-called lid complex, the portion of the 19S cap that is more distal to (distant from) the 20S proteasome. The functions of the lid subunits are largely unknown, although a truncated 26S proteasome that lacks the lid subunits is unable to degrade polyubiquitinated substrates. Several other subunits may be transiently associated with the 19S cap and/or with the 20S proteasome.

i. Deubiquitinating Enzymes Have Several Functions

The enzymes that hydrolytically cleave the isopeptide bonds linking successive ubiquitin units in polyUb are known as **deubiquitinating enzymes (DUBs)**. Cells contain a surprisingly large number of DUBs (at least 17 in yeast and ~100 in humans). Nearly all known DUBs are **cysteine proteases**, enzymes whose catalytic mechanism resembles that of serine proteases (Section 15-3C) but whose attacking nucleophile is Cys—S[—] rather than Ser—OH.

DUBs may release entire polyUb chains from a condemned protein or sequentially release ubiquitin units from the chain terminus. It has been proposed that this latter process functions as a clock to time the protein degradation process. If a polyUb chain is trimmed to less than four ubiquitin units before degradation begins, then its attached protein is likely to escape destruction. This would spare proteins that had been inappropriately tagged with only short polyUb chains.

The mammalian 19S lid subunit known as **POH1** (**Rpn11** for the 65% identical yeast subunit) appears to be responsible for the deubiquitination of target proteins prior to their degradation; its inactivation prevents target protein degradation. Curiously, this DUB is a Zn^{2+} -dependent protease (as is carboxypeptidase A; Fig. 15-42) rather than a cysteine protease.

Certain DUBs function to dismember polyUb chains that have been released from substrate proteins by sequentially removing ubiquitin units from the end of the chain that is nearest to the substrate protein (that with a free C-terminus). Consequently, these DUBs cannot remove ubiquitin units from polyUb chains that are still attached to substrate proteins, thereby preventing their premature removal.

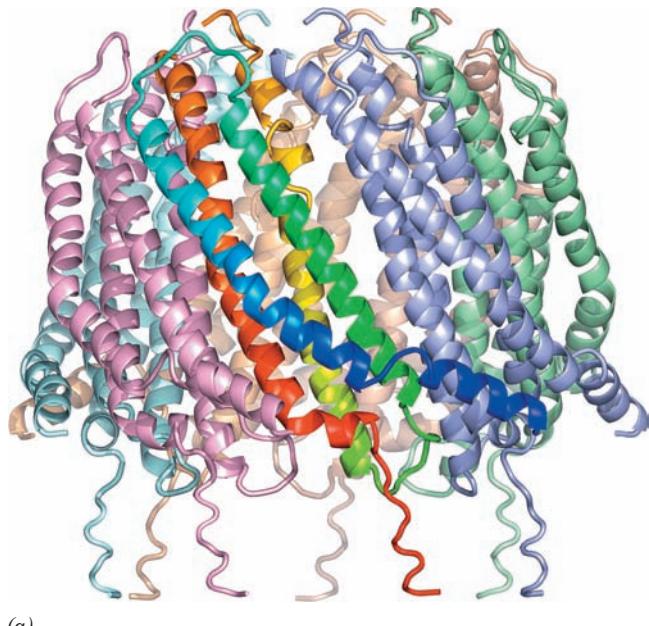
Cells express ubiquitin as polyproteins containing several ubiquitin units (Section 32-5Ac) or with ubiquitin fused to certain ribosomal subunits (there is no gene that encodes a single ubiquitin unit). These polyproteins are rapidly processed by certain DUBs to yield free ubiquitin.

into the 20S proteasome in an ATP-independent manner so as to permit the entrance of polypeptides (but not folded proteins). The mammalian 11S activator, which functions in the generation of peptides for presentation to the immune system (Section 35-2E), is named **REG** (alternatively **PA28**). It is a heteroheptameric complex of two ~245-residue subunits, **REG α** and **REG β** , that exhibit ~50% sequence identity except for a highly variable internal 18-residue segment that is thought to confer subunit-specific properties. Indeed **REG α** alone forms a heptamer whose biochemical properties are similar to that of **REG** (although both subunits must be present *in vivo*).

The trypanosome *Trypanosoma brucei*, which lacks 19S caps, expresses a homoheptameric 11S activator named **PA26** that is only 14% identical to human **REG α** . Nevertheless, the various 11S activators activate 20S proteasomes from widely divergent species. Thus, rat 20S proteasome is activated by **PA26** and the yeast 20S proteasome is activated by human **REG α** despite the fact that yeast lacks 11S activators.

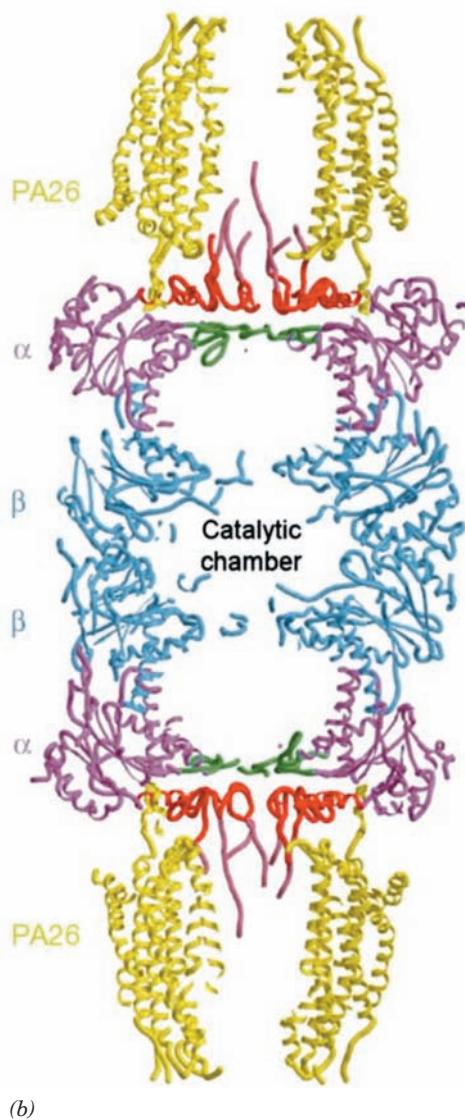
j. The 11S Activator Forms a Heptameric Barrel That Opens the 20S Proteasome

Higher eukaryotes contain an **11S activator** (alternatively, **11S regulator**) that functions to open the channel



(a)

Figure 32-86 X-ray structure of *T. brucei* PA26 in complex with the yeast 20S proteasome. (a) The PA26 heptamer in ribbon form viewed with its 7-fold axis vertical. Each of its subunits are differently colored with that closest to the viewer colored in rainbow order from its N-terminus (blue) to its C-terminus (red). (b) Cutaway diagram of the entire complex drawn in worm form and viewed with its 7-fold axis vertical. The PA26 is yellow, the α and β subunits of the 20S proteasome are magenta and blue, its α -annulus is green, and its N-terminal segments that are ordered and partially disordered are red and pink. [Part a based on an X-ray structure by and Part b courtesy of Christopher Hill, University of Utah. PDBid 1FNT.]



(b)

The X-ray structure of PA26 in complex with the yeast 20S proteasome, determined by Christopher Hill, reveals that each PA26 monomer consists of an up-down-up-down 4-helix bundle. These monomers form a 7-fold symmetric heptameric barrel that is 90 Å in diameter, 70 Å long, and has a 33-Å-diameter central pore (Fig. 32-86a) and which closely resembles the previously determined X-ray structure of human REGα. Two PA26 barrels associate coaxially with the 20S proteasome, one at each end (Fig. 32-86b). The conformation of the 20S proteasome in this complex, for the most part, is closely similar to that of the 20S proteasome alone (Fig. 32-85). However, the C-terminal tails of the PA26 subunits insert into pockets on the 20S proteasome's α subunits in a way that induces conformational changes in its N-terminal tails that clear the 20S proteasome's otherwise blocked central aperture (Fig. 32-85c,d), thus permitting unfolded polypeptides to enter the proteasome's central chamber.

k. Bacteria Contain a Variety of Self-Compartmentalized Proteases

Nearly all eubacteria lack 20S proteasomes. Nevertheless they have ATP-dependent proteolytic assemblies that share the same barrel-shaped architecture and carry out similar functions. For example, in *E. coli*, two proteins known as **Lon** and **Clp** mediate up to 80% of the bacterium's protein degradation, with additional contributions from at least three other proteins including **heat shock locus UV (HslUV)**. Thus, *all cells appear to contain proteases*

whose active sites are only available from the inner cavity of a hollow particle to which access is controlled. These so-called **self-compartmentalized proteases** appear to have arisen early in the history of cellular life, before the advent of eukaryotic membrane-bound organelles such as the lysosome, which similarly carry out degradative processes in a way that protects the cell contents from indiscriminate destruction.

Clp protease consists of two components, the proteolytically active **ClpP** and one of several ATPases, which in *E. coli* are **ClpA** and **ClpX**. The X-ray structure of ClpP, determined by John Flanagan, reveals that it oligomerizes to form an ~90-Å-long and -wide hollow barrel that consists of two back-to-back 7-fold symmetric rings of 193-residue subunits (Fig. 32-87) and thereby has the same D_7 symmetry as does the 20S proteasome. Nevertheless, the ClpP subunit has a novel fold that is different from that of the 20S proteasome's homologous α and β subunits. The ClpP active site, which is only exposed on the inside of the barrel, contains a catalytic triad composed of Ser 97, His 122, and Asp 171, and hence is a serine protease (Section 15-3Ab).

HslUV protease appears to be a hybrid of Clp and the 26S proteasome. Its **HslV** subunits in *Haemophilus influenzae* (174 residues) are 18% identical to the β subunits of the *T. acidophilum* 20S proteasome, whereas its regulatory **HslU** caps (444 residues) have ATPase activity and are homologous to *E. coli* ClpX. The X-ray structure of *H. influenzae* HslUV, determined by David McKay, indicates that HslV forms a dimer of hexameric rather than

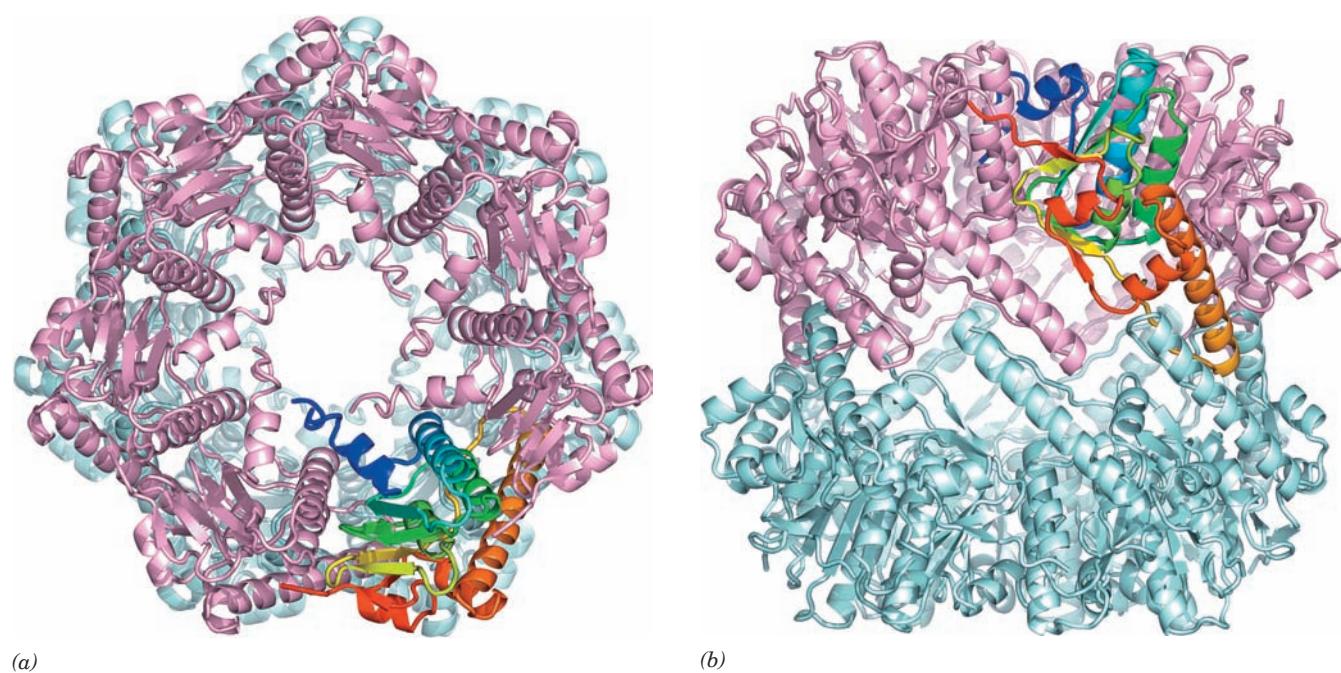


Figure 32-87 X-ray structure of *E. coli* ClpP. (a) View of the heptameric complex along its 7-fold axis, drawn in ribbon form in which the lower ring is pale cyan and the upper ring is pink with one subunit colored in rainbow order from its N-terminus (blue)

to its C-terminus (red). (b) View along the complex's 2-fold axis (rotated 90° about a horizontal axis with respect to Part a). [Based on an X-ray structure by John Flanagan, Brookhaven National Laboratory, Upton, New York. PDBid 1TYF.]

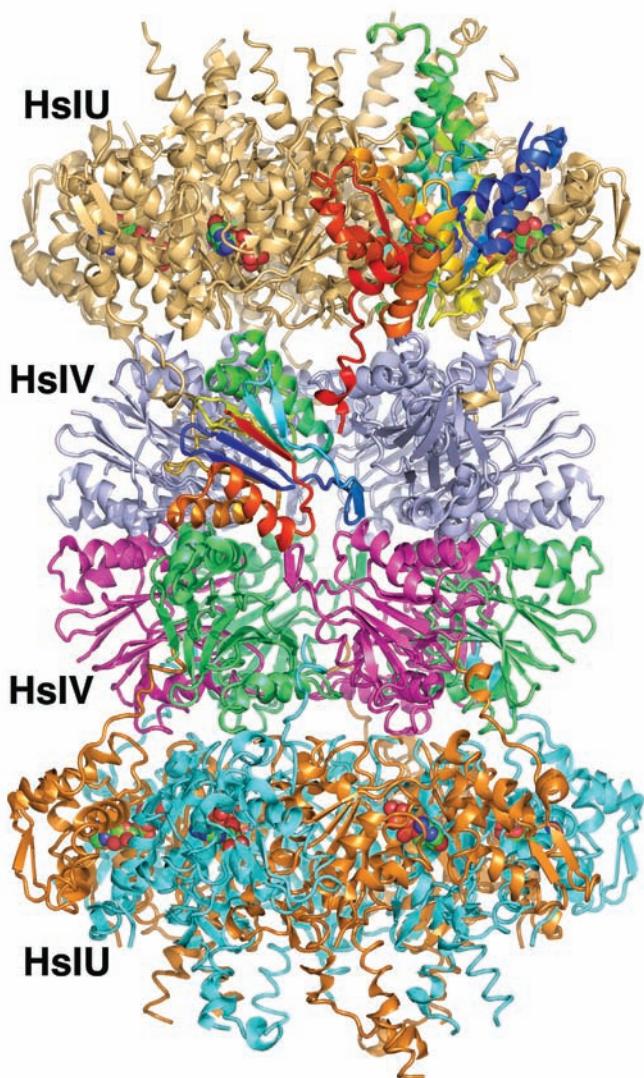


Figure 32-88 X-ray structure of *H. influenzae* HsIVU in complex with ATP. The 821-kD complex is drawn in ribbon form viewed along a 2-fold axis with its 6-fold axis vertical. The D_6 symmetric dodecamer of HsIV subunits is coaxially bound at both ends by C_6 symmetric HsIU hexamers to yield a complex with overall D_6 symmetry. The subunits of the lower HsIV hexamer are alternately orange and cyan, whereas those of the upper HsIV hexamer are pale orange except for one subunit, which is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The subunits of the lower HsIV hexamer are alternately green and magenta, whereas those of the upper hexamer are pale blue except for one subunit, which is colored in rainbow order. The ATPs, which are bound at the interfaces between neighboring HsIU subunits, are drawn in space-filling form with C green, N blue, O red, and P orange. [Based on an X-ray structure by David McKay, Stanford University School of Medicine. PDB 1G3I.]

heptameric rings (Fig. 32-88). A hexameric ring of HsIU subunits binds to both ends of the HsIV dodecamer to form a 24-subunit assembly with D_6 symmetry, rather than the D_7 symmetry of the 26S proteasome. Nevertheless, both the fold and the intersubunit contacts of the HsIV subunits are

closely similar to those of the 20S proteasome β subunits. In addition, both have N-terminal Thr residues. Thus, HsIV can be regarded as the eubacterial homolog of archaeabacterial and eukaryotic 20S proteasomes.

Thermoplasma acidophilum contains another large proteolytic complex, which is unrelated to the proteasome. The X-ray structure of this protease (Fig. 32-89), determined by Huber, indicates that it forms a 730-kD toroidal hexameric ring with D_3 symmetry that has a peculiar triangular shape reminiscent of a tricorn (a hat whose brim is turned up on three sides) and hence was named **tricorn protease**. Cryo-EM studies indicate that 20 of these tricorn hexamers associate to form a 14,600-kD hollow icosahedron (Fig. 32-89c; an icosahedron is shown in Fig. 8-65c), making it by far the largest homooligomeric enzyme complex known (it is even larger than some virus particles, many of which also have icosahedral symmetry; Section 33-2Aa).

I. Ubiquitination Has Multiple Proteasome-Independent Functions

Proteins may be monoubiquitinated or polyubiquitinated or even monoubiquitinated on more than one Lys residue. Moreover, ubiquitin has seven Lys residues so that seven types of polyubiquitin chains are possible.

Many types of ubiquitination mediate processes other than directing their associated proteins to the proteasome. This occurs through mechanisms reminiscent of protein phosphorylation, but instead of being recognized by specialized phosphoprotein-binding motifs such as SH2 (Section 19-3Cb), ubiquitinated proteins are recognized by conserved **ubiquitin-binding domains (UBDs)**. Moreover, whereas protein phosphorylation is reversible through the action of protein phosphatases (Section 19-3F), ubiquitination is reversible through the agency of DUBs.

Ubiquitination participates in regulating such diverse cellular processes as endocytosis, protein trafficking, DNA repair, intracellular signaling, and transcription. For example, during S phase of the cell cycle, the monoubiquitination of PCNA (the sliding clamp associated with the eukaryotic DNA replication fork; Section 30-4Ba) at its Lys 164 recruits an error-prone translesion DNA polymerase to the replication fork at a DNA damage site, whereas the polyubiquitination of the same site with a Lys 63-linked chain recruits DNA polymerases that mediate error-free lesion repair. In another example, in the signal transduction pathway that activates NF- κ B (Section 32-6Be), I κ B α is phosphorylated by **I κ B kinase (IKK)**, which has a regulatory subunit named **NEMO** (for *NF- κ B* essential modulator). The binding of cytokines such as TNF α and IL-1 to their transmembrane receptors in this pathway activates the receptor to Lys 63-linked polyubiquitin NEMO, which in turn activates IKK to phosphorylate I κ B α . This, as we saw, induces the formation of the Lys 48-linked polyubiquitination signal that leads to the destruction of I κ B α and hence the translocation of NF- κ B to the nucleus. *Yersinia pestis*, the bacterium that causes bubonic plague (Section 19-3Fc), produces a virulence factor named **YopJ**, which functions as a DUB that prevents the activation of NF- κ B (which is an

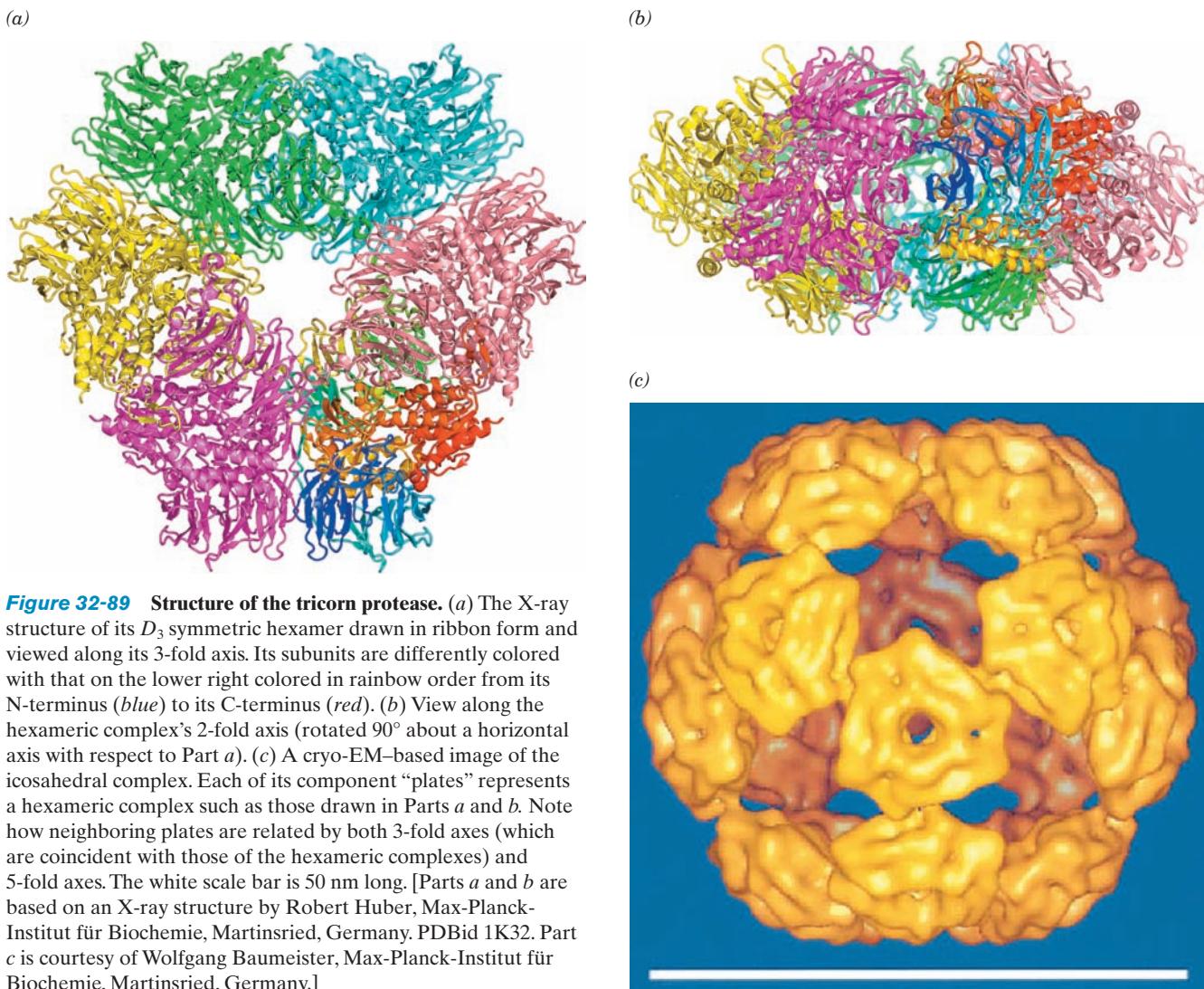


Figure 32-89 Structure of the tricorn protease. (a) The X-ray structure of its D_3 symmetric hexamer drawn in ribbon form and viewed along its 3-fold axis. Its subunits are differently colored with that on the lower right colored in rainbow order from its N-terminus (blue) to its C-terminus (red). (b) View along the hexameric complex's 2-fold axis (rotated 90° about a horizontal axis with respect to Part a). (c) A cryo-EM-based image of the icosahedral complex. Each of its component “plates” represents a hexameric complex such as those drawn in Parts a and b. Note how neighboring plates are related by both 3-fold axes (which are coincident with those of the hexameric complexes) and 5-fold axes. The white scale bar is 50 nm long. [Parts a and b are based on an X-ray structure by Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBid 1K32. Part c is courtesy of Wolfgang Baumeister, Max-Planck-Institut für Biochemie, Martinsried, Germany.]

important immune system activator). We shall see in Section 34-3Bff that the monoubiquitination of certain histones functions to regulate transcription.

m. Ubiquitinlike Modifiers Participate in a Variety of Regulatory Processes

Eukaryotic cells express several proteins that have the same β -Grasp fold as ubiquitin and are similarly conjugated to other proteins, although their surface residues and charge distributions differ significantly. These **ubiquitinlike modifiers (UbIs)**, which participate in a variety of fundamental cellular processes, each have a corresponding activating enzyme (E1), at least one conjugating enzyme (E2), and one or more ligases (E3s), that function to link the Ubl to its target protein(s) in a manner closely resembling that of ubiquitin.

Two of the most extensively studied UbIs are **SUMO** (small ubiquitin-related modifier; 18% identical to ubiquitin) and **RUB1** (related-to-ubiquitin 1; called **NEDD8** in

vertebrates; 50% identical to ubiquitin), proteins that are highly conserved from yeast to humans. The sumoylation of PCNA at the same Lys residue at which it is ubiquitinated promotes normal DNA replication during S phase. The sumoylation of $\text{I}\kappa\text{B}\alpha$ at the same residue (Lys 21) at which it is ubiquitinated blocks its ubiquitination and subsequent degradation and thereby prevents the translocation of NF- κ B to the nucleus. Evidently, there is a complex regulatory interplay between the ubiquitination and sumoylation of both PCNA and $\text{I}\kappa\text{B}\alpha$. SUMO also modifies two mammalian glucose transporters, GLUT1 and GLUT4 (Section 20-2E), and in doing so, increases the availability of GLUT4 but decreases that of GLUT1.

All known RUB1 targets are cullins, all of which are subunits of SCF complexes, the multisubunit RING E3s (Section 32-6Bd). In fact, β -TrCP, the E3 that directs the ubiquitination of $\text{I}\kappa\text{B}\alpha$, must be conjugated to RUB1 before it can do so, thereby adding further complexity to the control of NF- κ B.

CHAPTER SUMMARY

1 The Genetic Code Point mutations are caused by either base analogs that mispair during DNA replication or by substances that react with bases to form products that mispair. Insertion/deletion (frameshift) mutations arise from the association of DNA with intercalating agents that distort the DNA structure. The analysis of a series of frameshift mutations that suppress one another established that the genetic code is an unpunctuated triplet code. In a cell-free protein synthesizing system, poly(U) directs the synthesis of poly(Phe), thereby demonstrating that UUU is the codon specifying Phe. The genetic code was elucidated through the use of polynucleotides of known composition but random sequence, by the ability of defined triplets to promote the ribosomal binding of tRNAs bearing specific amino acids, and through the use of synthetic mRNAs of known alternating sequences. The latter investigations have also demonstrated that the 5' end of mRNA corresponds to the N-terminus of the polypeptide it specifies and have established the sequences of the Stop codons. Degenerate codons differ mostly in the identities of their third base. Small single-stranded DNA phages such as ϕ X174 contain overlapping genes in different reading frames. The genetic code used by mitochondria differs in several codons from the “standard” genetic code.

2 Transfer RNA and Its Aminoacylation Transfer RNAs consist of 54 to 100 nucleotides that can be arranged in the cloverleaf secondary structure. As many as 10% of a tRNA's bases may be modified. Yeast tRNA^{Phe} forms a narrow, L-shaped, three-dimensional structure that resembles that of other tRNAs. Most of the bases are involved in stacking and base pairing associations including nine tertiary interactions that appear to be essential for maintaining the molecule's native conformation. Amino acids are appended to their cognate tRNAs in a two-stage reaction catalyzed by the corresponding aminoacyl-tRNA synthetase (aaRS). There are two classes of aaRSs, each containing 10 members. Class I aaRSs have two conserved sequence motifs that occur in the Rossmann fold common to the catalytic domain of these enzymes. Class II aaRSs have three conserved sequence motifs that occur in the 7-stranded antiparallel β sheet-containing fold that forms the core of their catalytic domains. In binding only their cognate tRNAs, aaRSs recognize only an idiosyncratic but limited number of bases (identity elements) that are, most often, located at the anticodon and in the acceptor stem. The great accuracy of tRNA charging arises from the proofreading of the bound amino acid by certain aminoacyl-tRNA synthetases via a double-sieve mechanism and at the expense of ATP hydrolysis.

Many organisms and organelles lack a GlnRS and instead synthesize Gln-tRNA^{Gln} by the GluRS-catalyzed charging of tRNA^{Gln} with glutamate followed by its transamidation using glutamine as the amido group source in a reaction mediated by Glu-tRNA^{Gln} amidotransferase (Glu-Adt). Ribosomes select tRNAs solely on the basis of their anticodons. Sets of degenerate codons are read by a single tRNA through wobble pairing. The UGA codon, which is normally the *opal* Stop codon may, depending on its context in mRNA, specify a selenoCys (Sec) residue, which is carried by a specific tRNA (tRNA^{Sec}), thereby forming a selenoprotein. Nonsense mutations may be suppressed by tRNAs whose anticodons have mutated to recognize a Stop codon.

3 Ribosomes and Polypeptide Synthesis The ribosome consists of a small and a large subunit whose complex shapes have been revealed by cryoelectron microscopy and X-ray crystallography. The three RNAs and 52 proteins comprising the *E. coli* ribosome self-assemble under proper conditions. Both ribosomal subunits consist of an RNA core in which the proteins are embedded, mainly as globular domains on the back and sides of the particle, with long basic polypeptide segments that infiltrate between the RNA helices so as to neutralize their anionic charges. Eukaryotic ribosomes are larger and more complex than those of prokaryotes.

Ribosomal polypeptide synthesis proceeds by the addition of amino acid residues to the C-terminal end of the nascent polypeptide. The mRNAs are read in the 5' \rightarrow 3' direction. mRNAs are usually simultaneously translated by several ribosomes in the form of polysomes. The ribosome has three tRNA-binding sites: the A site, which binds the incoming aminoacyl-tRNA; the P site, which binds the peptidyl-tRNA; and the E site, which transiently binds the outgoing deacylated tRNA. During polypeptide synthesis, the nascent polypeptide is transferred to the aminoacyl-tRNA, thereby lengthening the nascent polypeptide by one residue. The newly deacylated tRNA is translocated to the E site and the new peptidyl-tRNA, with its associated codon, is translocated to the P site. In prokaryotes, the initiation sites on mRNA are recognized through their Shine-Dalgarno sequences and by their initiating codon. Prokaryotic initiating codons specify fMet-tRNA_f^{Met}. Initiation involves the participation of three initiation factors that induce the assembly of the ribosomal subunits with fMet-tRNA_f^{Met} in the P site and mRNA. Eukaryotic initiation is a far more complicated process that requires the participation of at least 11 initiation factors. The system binds the mRNA's 5' cap and scans along the mRNA until it finds its AUG initiation codon, usually the mRNA's first AUG, through codon-anticodon interactions with the initiating tRNA, Met-tRNA_i^{Met}.

Polypeptides are elongated in a three-part cycle, consisting of aminoacyl-tRNA decoding, transpeptidation, and translocation, that requires the participation of elongation factors and is vectorially driven by GTP hydrolysis. EF-Tu, which functions to escort aminoacyl-tRNAs into the ribosomal A site, undergoes a major conformational change on hydrolyzing its bound GTP. The X-ray structure of the 50S subunit clearly shows that the ribosomal peptidyl transferase center is distant from any protein and hence that the ribosome is a ribozyme. Peptide bond formation is catalyzed by a substrate-assisted mechanism in which the ribosome functions as an entropy trap. Translocation is motivated through the EF-G-catalyzed hydrolysis of GTP. EF-G \cdot GDP, which binds to the same ribosomal site as aminoacyl-tRNA \cdot EF-Tu \cdot GTP, is a macromolecular mimic of this complex. Translocation occurs via intermediate states, the A/P and P/E states, in which the newly formed peptidyl-tRNA and the newly deacylated tRNA are respectively bound to the A and P subsites of the 30S subunit and to the P and E subsites of the 50S subunit, following which EF-G hydrolyzes its bound GTP and shifts these tRNAs to the P/P and E/E states. The ribosome initially selects an aminoacyl-tRNA whose anticodon is cognate to its A-site-bound codon through interactions involving three universally con-

served 30S subunit bases while the tRNA is in the A/T binding state. The codon–anticodon interaction is then proofread in an independent process that follows the hydrolysis of the EF-Tu-bound GTP and which occurs when the tRNA has shifted to the A/A binding state, a process called accommodation.

Termination codons bind release factors that induce the hydrolysis of the peptidyl-tRNA bond. Eukaryotic elongation and termination resemble those of prokaryotes. Ribosomal inhibitors, many of which are antibiotics, are medically important and biochemically useful in elucidating ribosomal function. Streptomycin causes mRNA misreading and inhibits prokaryotic chain initiation, chloramphenicol inhibits prokaryotic peptidyl transferase, paromomycin causes codon misreading, tetracycline inhibits aminoacyl-tRNA binding to the prokaryotic 30S subunit, and diphtheria toxin ADP-ribosylates eEF2.

4 Control of Eukaryotic Translation Several mechanisms of translational control have been elucidated in eukaryotes. eIF2 α kinases catalyze the phosphorylation of eIF2 α , which then tightly binds eIF2B so as to prevent it from exchanging eIF2-bound GDP for GTP and hence inhibits translational initiation. These eIF2 α kinases include heme-regulated inhibitor (HRI), which functions to coordinate globin synthesis with heme availability; double-stranded RNA-activated protein kinase (PKR), an interferon-induced protein that functions to inhibit viral proliferation; and PKR-like endoplasmic reticulum kinase (PERK), which functions to protect the cell from the irreversible damage caused by the accumulation of unfolded proteins in the ER. GCN2, in contrast, is an eIF2 α kinase that, when amino acids are scarce, stimulates the translation of the transcriptional activator GCN4 by causing the 40S ribosomal subunit to scan across four upstream open reading frames (uORFs) in the *GCN4* mRNA, thereby permitting the ribosome to initiate translation at the GCN4 coding sequence. The phosphorylation of eIF4E (cap-binding protein) by a MAP kinase cascade increases eIF4E's affinity for capped mRNA and thereby stimulates translational initiation. The binding of 4E-BPs to eIF4E blocks its binding of eIF4G and hence prevents initiation. However, the insulin-induced phosphorylation of the 4E-BPs causes them to dissociate from eIF4E. The mRNAs in certain animal oocytes are masked by the binding of proteins, which prevents their translation.

Many oocyte mRNAs have short poly(A) tails that are preceded by a cytoplasmic polyadenylation element (CPE) that is bound by CPE-binding protein (CPEB). CPEB binds maskin which binds eIF4E, thereby inhibiting translational initiation. However, when CPEB is phosphorylated, it recruits poly(A) polymerase (PAP), which extends the mRNA's poly(A) tail such that it is bound by poly(A)-binding protein (PABP). PABP then binds to eIF4G, which in turn displaces maskin from eIF4E, thereby permitting the translation of the mRNA. Antisense oligonucleotides can be used to inhibit the translation of their complementary mRNAs. Although the delivery of antisense oligonucleotides to their sites of action has proved to be a difficult problem, their use is starting to show some medical and biotechnological successes.

5 Post-Translational Modification Proteins may be post-translationally modified in a variety of ways. Proteolytic cleavages, usually by specific peptidases, activate proproteins. The signal peptides of preproteins are removed by signal peptidases. Covalent modifications alter many types of side chains in a variety of ways that modulate the catalytic activities of enzymes, provide recognition markers, and stabilize protein structures.

Protein splicing occurs via the intein-catalyzed self-excision between an N-extein and a C-extein accompanied by the ligation of the N- and C-exteins via a peptide bond. Most inteins contain a homing endonuclease that makes a double-strand nick in a gene similar to that encoding the corresponding extein, thereby triggering a recombinational double-strand DNA repair process that copies the gene encoding the intein into the break. Inteins therefore appear to be molecular parasites.

6 Protein Degradation Proteins in living cells are continually turning over. This controls the level of regulatory enzymes and disposes of abnormal proteins that would otherwise interfere with cellular processes. Proteins are degraded by lysosomes via a nonspecific process as well via a process specific for KFERQ proteins that is stimulated during starvation. A cytosolically based ATP-dependent system degrades normal as well as abnormal proteins in a process that flags these proteins by the covalent attachment of Lys 48-linked polyubiquitin chains to their Lys residues. This process is mediated by three consecutively acting enzymes: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin-protein ligase (E3). Most cells have one species of E1, several species of E2, and numerous species of E3, each of which is served by one or a few E2s. The polyubiquitinated protein is proteolytically degraded in the 26S proteasome.

E3s can have complicated modular structures, each having different specificities for target proteins. SCF complexes, one of whose several subunits contains a RING domain, are particularly elaborate. The RING E3 known as E3 α functions to ubiquitinate proteins that satisfy the N-end rule. The transcription factor NF- κ B is activated through the ubiquitination and subsequent destruction of its otherwise bound inhibitor I κ B α by the SCF β -TrCP, which is activated through phosphorylation via a signal transduction cascade. Cyclins, which mediate the cell cycle, are destroyed in a programmed manner through ubiquitination by their cognate E3s, one of which is anaphase-promoting complex (APC).

The 26S proteasome consists of a hollow protein barrel formed by two rings of seven α subunits flanking two rings of seven β subunits known as the 20S proteasome, which is bound at each end by 19S caps that each consist of ~18 subunits. The active sites of the β subunits, which are members of the N-terminal nucleophile (Ntn) family of hydrolases, are inside the barrel. Ubiquitinated proteins are selected by the 19S caps, which unfold them in an ATP-dependent manner and then feed them into the 20S proteasome via its axial channel.

The polyubiquitin (polyUb) chains are excised from the condemned protein by proteasome-associated deubiquitinating enzymes (DUBs), while other DUBs dismember the polyUb chains to their component ubiquitin units, thereby recycling them. The 11S activator is a heptameric complex that, on binding to one end of a 20S proteasome, opens its axial channel in an ATP-independent manner, thereby permitting polypeptides, but not folded proteins, to enter the 20S proteasome. Eubacteria, nearly all of which lack proteasomes, nevertheless express a variety of self-compartmentalized proteases, including ClpP, heat shock locus UV (HsUV), and tricorn protease, that function to proteolytically dispose of their cellular proteins. Monoubiquitination and polyubiquitination with chains linked through ubiquitin residues other than Lys 48 regulate a variety of cellular processes. Ubiquitinlike modifiers (Ubls), such as SUMO and RUB1, also participate in numerous regulatory processes.

REFERENCES

General

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R., *Molecular Biology of the Gene* (6th ed.), Chap. 14 and 15, Cold Spring Harbor Laboratory Press (2008).

The Genetic Code

Attardi, G., Animal mitochondrial DNA: an extreme example of genetic economy, *Int. Rev. Cytol.* **93**, 93–145 (1985).

Benzer, S., The fine structure of the gene, *Sci. Am.* **206**(1), 70–84 (1962).

Crick, F.H.C., The genetic code, *Sci. Am.* **207**(4), 66–74 (1962) [The structure of the code as determined by phage genetics]; and The genetic code: III, *Sci. Am.* **215**(4), 55–62 (1966). [A description of the nature of the code after its elucidation was almost complete.]

Crick, F.H.C., Burnett, L., Brenner, S., and Watts-Tobin, R.J., General nature of the genetic code for proteins, *Nature* **192**, 1227–1232 (1961).

Fox, T.D., Natural variation in the genetic code, *Annu. Rev. Genet.* **21**, 67–91 (1987).

Judson, J.F., *The Eighth Day of Creation*, Expanded Edition, Part II, Cold Spring Harbor Laboratory Press (1996). [A fascinating historical narrative on the elucidation of the genetic code.]

Khorana, H.G., Nucleic acid synthesis in the study of the genetic code, *Nobel Lectures in Molecular Biology, 1933–1975*, pp. 303–331, Elsevier (1977).

Knight, R.D., Freeland, S.J., and Landweber, L.F., Selection, history and chemistry: the three faces of the genetic code, *Trends Biochem. Sci.* **24**, 241–247 (1999).

Nirenberg, M., The genetic code, *Nobel Lectures in Molecular Biology, 1933–1975*, pp. 335–360, Elsevier (1977).

Nirenberg, M., Historical review: deciphering the genetic code—a personal account, *Trends Biochem. Sci.* **29**, 46–54 (2004).

Nirenberg, M. and Leder, P., RNA code words and protein synthesis, *Science* **145**, 1399–1407 (1964). [The determination of the genetic code by the ribosomal binding of tRNAs using specific trinucleotides.]

Nirenberg, M.W. and Matthaei, J.H., The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides, *Proc. Natl. Acad. Sci.* **47**, 1588–1602 (1961). [The landmark paper reporting the finding that poly(U) stimulates the synthesis of poly(Phe).]

Singer, B. and Kusmirek, J.T., Chemical mutagenesis, *Annu. Rev. Biochem.* **51**, 655–693 (1982).

The Genetic Code, Cold Spring Harbor Symp. Quant. Biol. **31** (1966). [A collection of papers describing the establishment of the genetic code. See especially the articles by Crick, Nirenberg, and Khorana.]

Yanofsky, C., Establishing the triplet nature of the genetic code, *Cell* **128**, 815–818 (2007).

Yarus, M., Caporaso, J.G., and Knight, R., Origins of the genetic code: the escaped triplet theory, *Annu. Rev. Biochem.* **74**, 179–198 (2005).

Transfer RNA and Its Aminoacylation

Alexander, R.W. and Schimmel, P., Domain–domain communication in aminoacyl-tRNA synthetases, *Prog. Nucleic Acid Res. Mol. Biol.* **69**, 317–349 (2001).

Ambrogelly, A., Paliora, S., and Söll, D., Natural expansion of the genetic code, *Nature Chem. Biol.* **3**, 29–35 (2007).

Björk, G.R., Ericson, J.U., Gustafsson, C.E.D., Hagervall, T.G.,

Jösson, Y.H., and Wikström, P.M., Transfer RNA modification, *Annu. Rev. Biochem.* **56**, 263–287 (1987).

Böck, A., Forschhammer, K., Heider, J., and Baron, C., Selenoprotein synthesis: an expansion of the genetic code, *Trends Biochem. Sci.* **16**, 463–467 (1991).

Crick, F.H.C., Codon–anticodon pairing: the wobble hypothesis, *J. Mol. Biol.* **19**, 548–555 (1966).

Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D.G., and Yokoyama, S., Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA^{Val} and valyl-tRNA synthetase, *Cell* **103**, 793–803 (2000).

Ibba, M. and Söll, D., Aminoacyl-tRNA synthesis, *Annu. Rev. Biochem.* **69**, 617–650 (2000).

Ibba, M.A., Francklyn, C., and Cusack, S. (Eds.), *The Aminoacyl-tRNA Synthetases*, Landes Bioscience/Eurekah.com (2005). [Contains reviews on each of the twenty aminoacyl-tRNA synthetases.]

Jacquin-Becker, C., Ahel, I., Ambrogelly, A., Ruan, B., Söll, D., and Stathopoulos, C., Cysteinyl-tRNA formation and prolyl-tRNA synthetase, *FEBS Lett.* **514**, 34–36 (2002).

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.M.J., Seeman, N.C., and Rich, A., Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science* **185**, 435–440 (1974); and Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C., and Klug, A., Structure of yeast phenylalanine tRNA at 3 Å resolution, *Nature* **250**, 546–551 (1974). [The landmark papers describing the high-resolution structure of a tRNA.]

Krzycki, J.A., The direct genetic encoding of pyrrolysine, *Curr. Opin. Microbiol.* **8**, 706–712 (2005); and Nozawa, K., O'Donoghue, P., Gundlapalli, S., Araiso, Y., Ishitani, R., Umebara, T., Söll, D., and Nureki, O., Pyrolysyl-tRNA synthetase-tRNA^{PyL} structure reveals the molecular basis of orthogonality, *Nature* **457**, 1163–1167 (2009).

Ling, J., Reynolds, N., and Ibba, M., Aminoacyl-tRNA synthesis and translational control, *Annu. Rev. Microbiol.* **63**, 61–78 (2009).

Nureki, O., Vassylyev, D.G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P., and Yokoyama, S., Enzyme structure with two catalytic sites for double-sieve selection of substrate, *Science* **280**, 578–582 (1998). [The X-ray structures of IleRS in complexes with isoleucine and valine.]

Rould, M.A., Perona, J.J., and Steitz, T.A., Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase, *Nature* **352**, 213–218 (1991).

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D., Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA^{Asp}, *Science* **252**, 1682–1689 (1991).

Schimmel, P. and Beebe, K., Aminoacyl tRNA synthetases: from the RNA to the theater of proteins. In Gesteland, R.F., Cech, T.R., and Atkins, J.F. (Eds.), *The RNA World* (3rd ed.), pp. 227–255, Cold Spring Harbor Laboratory Press (2006).

Silvian, L.F., Wang, J., and Steitz, T.A., Insights into editing from an Ile-tRNA synthetase structure with tRNA^{Ile} and mupirocin, *Science* **285**, 1074–1077 (1999).

Söll, D. and RajBhandary, U.L. (Eds.), *tRNA: Structure, Biosynthesis, and Function*, ASM Press (1995).

Stadtman, T.C., Selenocysteine, *Annu. Rev. Biochem.* **65**, 83–100 (1996).

Ribosomes and Polypeptide Synthesis

- Aitken, C.E., Petrov, A., and Puglisi, J.D., Single ribosome dynamics and the mechanism of translation, *Annu. Rev. Biophys.* **39**, 491–513 (2010).
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T., The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution, *Science* **289**, 905–920 (2000).
- Bashan, A. and Yonath, A., Correlating ribosome function with high-resolution structures, *Trends Microbiol.* **16**, 326–335 (2008).
- Bell, C.E. and Eisenberg, D.E., Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide, *Biochemistry* **35**, 1137–1149 (1996).
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W., The native 3D organization of bacterial polysomes, *Cell* **136**, 261–271 (2009).
- Brodersen, D.E., Clemons, W.M., Jr., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V., The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit, *Cell* **103**, 1143–1154 (2000).
- Czworkowski, J., Wang, J., Steitz, J.A., and Moore, P.B., The crystal structures of elongation factor G complexed with GDP, at 2.7 Å resolution; and Ævarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Yu., Al-Karadaghi, S., Svensson, L.A., and Liljas, A., Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus*, *EMBO J.* **13**, 3661–3668 and 3669–3677 (1994).
- Dintzis, H.M., Assembly of the peptide chains of hemoglobin, *Proc. Natl. Acad. Sci.* **47**, 247–261 (1961); and The wandering pathway to determining N to C synthesis of proteins: Some recollections concerning protein structure and biosynthesis, *Biochem. Mol. Biol. Educ.* **34**, 241–246 (2006). [The determination of the direction of polypeptide biosynthesis.]
- Dunkle, J.A. and Cate, J.H.D., Ribosome structure and dynamics during translocation and termination, *Annu. Rev. Biophys.* **39**, 227–244 (2010).
- Frank, J., Single-particle imaging of macromolecules by cryo-electron microscopy, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 303–319 (2002).
- Frank, J. and Agrawal, R.K., A ratchet-like inter-subunit reorganization of the ribosome during translocation, *Nature* **406**, 318–322 (2000).
- Frank, J. and Gonzalez, R.L., Jr., Structure and dynamics of a progressive Brownian motor: the translating ribosome, *Annu. Rev. Biochem.* **79**, 381–412 (2010).
- Gao, H., et al., RF3 induces ribosomal conformational changes responsible for dissociation of class I release factors, *Cell* **129**, 929–941 (2007).
- Gingras, A.-C., Raught, B., and Sonnberg, N., eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation, *Annu. Rev. Biochem.* **68**, 913–963 (1999).
- Held, W.A., Ballou, B., Mizushima, S., and Nomura, M., Assembly mapping of 30S ribosomal proteins from *Escherichia coli*, *J. Biol. Chem.* **249**, 3103–3111 (1974).
- Holbrook, S.R., Structural principles from large RNAs, *Annu. Rev. Biophys.* **37**, 445–464 (2008).
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V., The mechanism of eukaryotic translation and principles of its regulation, *Nature Rev. Mol. Cell Biol.* **10**, 113–127 (2010).
- Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S., and Leberman, R., The structure of the *Escherichia coli* EF-Tu · EF-Ts complex at 2.5 Å resolution, *Nature* **379**, 511–518 (1996).
- Kiel, M.C., Kaji, H., and Kaji, A., Ribosome recycling, *Biochem. Mol. Biol. Educ.* **35**, 40–44 (2007); and Hirokawa, G., Demeshkin, N., Iwakura, N., Kaji, H., and Kaji, A., The ribosome-recycling step: consensus or controversy? *Trends Biochem. Sci.* **31**, 143–149 (2006).
- Kjeldgaard, M. and Nyborg, J., Refined structure of elongation factor EF-Tu from *Escherichia coli*, *J. Mol. Biol.* **223**, 721–742 (1992); and Kjeldgaard, M., Nissen, P., Thirup, S., and Nyborg, J., The crystal structure of elongation factor EF-Tu from *Thermus aquaticus* in the GTP conformation, *Structure* **1**, 35–50 (1993).
- Korostelev, A. and Noller, H.F., The ribosome in focus: new structures bring new insights, *Trends Biochem. Sci.* **32**, 434–441 (2007); and Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F., Crystal structure of the 70S ribosome-tRNA complex reveals functional interactions and rearrangements, *Cell* **126**, 1065–1077 (2006).
- Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., and Noller, H.F., Structural basis for translation termination on the 70S ribosome, *Nature* **454**, 852–857 (2008); Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W.G., and Noller, H.F., Crystal structure of a translation termination complex formed with release factor RF2, *Proc. Natl. Acad. Sci.* **105**, 19684–19689 (2008); and Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R.M., Petry, S., Kelley, A.C., and Ramakrishnan, V., Insights into translational termination from the structure of RF2 bound to the ribosome, *Science* **322**, 953–956 (2008).
- Marcotrigiano, J., Gingras, A.-C., Sonenberg, N., and Burley, S.K., Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP, *Cell* **89**, 951–961 (1997).
- Moazed, D. and Noller, H.F., Intermediate states in the movement of transfer RNA in the ribosome, *Nature* **342**, 142–148 (1989).
- Moore, P.B. and Steitz, T.A., The involvement of RNA in ribosome function, *Nature* **418**, 229–235 (2002); and The structural basis of large ribosomal subunit function, *Annu. Rev. Biochem.* **72**, 813–850 (2003).
- Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M.T., and Blanchard, S.C., Navigating the ribosome's metastable energy landscape, *Trends Biochem. Sci.* **31**, 390–399 (2009).
- Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B.F.C., and Nyborg, J., Crystal structure of the ternary complex of Phe-tRNA^{Phen}, EF-Tu, and a GTP analog, *Science* **270**, 1464–1472 (1995).
- Noller, H.F., Hoffarth, V., and Zimniak, L., Unusual resistance of peptidyl transferase to protein extraction procedures, *Science* **256**, 1416–1419 (1992); and Noller, H.F., Peptidyl transferase: protein, ribonucleoprotein, or RNA? *J. Bacteriol.* **175**, 5297–5300 (1993).
- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V., Recognition of cognate transfer RNA by the 30S ribosomal subunit, *Science* **292**, 897–902 (2001); and Ogle, J.M., Carter, A.P., and Ramakrishnan, V., Insights into the decoding mechanism from recent ribosome structures, *Trends Biochem. Sci.* **28**, 259–266 (2003).
- Pioletti, M., et al., Crystal structure of complexes of the small ribosomal subunit with tetracycline, edeine and IF3, *EMBO J.* **20**, 1829–1839 (2001).
- Rodnina, M.V., Beringer, M., and Wintermeyer, W., How the ribosome makes peptide bonds, *Trends Biochem. Sci.* **32**, 20–26 (2007).
- Schlunzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A., Structure of functionally activated small ribosomal subunit at 3.3 Å resolution, *Cell* **102**, 615–623 (2000).
- Schmeing, T.M., and Ramakrishnan, V., What recent ribosome structures have revealed about the mechanism of translation,

- Nature* **461**, 1234–1242 (2009); Simonovic, M. and Steitz, T.A., A structural view of the mechanism of the ribosome-catalyzed peptide bond mechanism, *Biochim. Biophys. Acta* **1789**, 612–623 (2009); and Zimmerman, E. and Yonath, A., Biological implications of the ribosome's stunning stereochemistry, *ChemBioChem* **10**, 63–72 (2009). [Authoritative reviews.]
- Schmeing, T.M., Voorhees, R.M., Kelley, A.C., Gao, Y.-G., Murphy, F.V., IV, Weir, J.R., and Ramakrishnan, V., The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA; and Gao, Y.-G., Selmer, M., Dunham, M., Weixlbaumer, A., Kelley, A.C., and Ramakrishnan, V., The structure of the ribosome with elongation factor G trapped in the post-translocational state, *Science* **326**, 688–693 and 694–699 (2009).
- Selmer, M., Dunham, C.M., Murphy, F.V., IV, Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V., Structure of the 70S ribosome complexed with mRNA and tRNA, *Science* **313**, 1935–1942 (2006); and Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C., and Ramakrishnan, V., Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome, *Nature Struct. Mol. Biol.* **16**, 528–533 (2009).
- Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R., The ribosome as an entropy trap. *Proc. Natl. Acad. Sci.* **101**, 7897–7901 (2004).
- Sonenberg, N., eIF4E, the mRNA cap-binding protein: from basic discovery to translational research, *Biochem. Cell Biol.* **86**, 178–183 (2008).
- Sonenberg, N. and Dever, T.E., Eukaryotic translation initiation factors and regulators, *Curr. Opin. Struct. Biol.* **13**, 56–63 (2003).
- Spanh, C.M.T., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., and Frank, J., Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit–subunit interactions, *Cell* **107**, 373–386 (2001).
- Steitz, J.A. and Jakes, K., How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S RNA and the mRNA during initiation of protein synthesis in *Escherichia coli*, *Proc. Natl. Acad. Sci.* **72**, 4734–4738 (1975).
- Sykes, M.T. and Williamson, J.R., A complex assembly landscape for the 30S ribosomal subunit, *Annu. Rev. Biophys.* **38**, 197–215 (2009); and Woodson, S.A., RNA folding and ribosome assembly, *Curr. Opin. Struct. Biol.* **12**, 667–673 (2008).
- Vestergaard, B., Van, L.B., Andersen, G.R., Nyborg, J., Buckingham, R.H., and Kjeldgaard, M., Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1, *Mol. Cell* **8**, 1375–1382 (2001).
- Weixlbaumer, A., Petry, S., Dunham, C.M., Selmer, M., Kelley, A.C., and Ramakrishnan, V., Crystal structure of the ribosome recycling factor bound to the ribosome, *Nature Struct. Mol. Biol.* **14**, 733–737 (2007).
- Wimberly, B.T., Broderson, D.E., Clemons, W.M., Jr., Morgan-Warren, R., von Rhein, C., Hartsch, T., and Ramakrishnan, V., Structure of the 30S ribosomal subunit, *Nature* **407**, 327–339 (2000); and Broderson, D.E., Clemons, W.M., Jr., Carter, A.P., Wimberly, B.T., and Ramakrishnan, V., Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16S RNA, *J. Mol. Biol.* **316**, 725–768 (2002).
- Yonath, A., The search and its outcome: High resolution structures of ribosomal particles from mesophilic, thermophilic, and halophilic bacteria at various functional states, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 257–273 (2002).
- Yonath, A., Antibiotics targeting ribosomes: resistance, selectivity, synergism, and cellular regulation, *Annu. Rev. Biochem.* **74**, 649–679 (2005).
- Youngman, E.M., McDonald, M.E., and Green, R., Peptide release on the ribosome: mechanism and implications for translational control, *Annu. Rev. Microbiol.* **62**, 33–373 (2008).
- Yusupova, G.Z., Yusupov, M.M., Cate, J.D.H., and Noller, H.F., The path of messenger RNA through the ribosome, *Cell* **106**, 233–241 (2001).

Control of Eukaryotic Translation

- Branch, A.D., A good antisense molecule is hard to find, *Trends Biochem. Sci.* **23**, 45–50 (1998).
- Calkhoven, C.F., Müller, C., and Leutz, A., Translational control of gene expression and disease, *Trends Mol. Med.* **8**, 577–583 (2002).
- Chen, J.-J. and London, I.M., Regulation of protein synthesis by heme-regulated eIF-2 α kinase, *Trends Biochem. Sci.* **20**, 105–108 (1995).
- Clemens, M.J., PKR—A protein kinase regulated by double-stranded RNA, *Int. J. Biochem. Cell Biol.* **29**, 945–949 (1997).
- Dever, T.E., Gene-specific regulation by general translation factors, *Cell* **108**, 545–556 (2002).
- Gray, N.K. and Wickens, M., Control of translation initiation in animals, *Annu. Rev. Cell Dev. Biol.* **14**, 399–458 (1998).
- Lawrence, J.C., Jr. and Abraham, R.T., PHAS/4E-BPs as regulators of mRNA translation and cell proliferation, *Trends Biochem. Sci.* **22**, 345–349 (1997).
- Lebedeva, I. and Stein, C.A., Antisense oligonucleotides: promise and reality, *Annu. Rev. Pharmacol. Toxicol.* **41**, 403–419 (2001).
- Matthews, M.B., Sonnenberg, N., and Hershey, J.W.B. (Eds.), *Translational Control in Biology and Medicine*, Cold Spring Harbor Laboratory Press (2007).
- Mendez, R. and Richter, J.D., Translational control by CPEB: a means to the end, *Nature Rev. Mol. Cell Biol.* **2**, 521–529 (2001).
- Phillips, M.I. (Ed.), *Antisense Technology: Part A. General Methods, Methods of Delivery, and RNA Studies; and Part B. Applications, Methods Enzymol.* **313** and **314** (2000).
- Sen, G.C. and Lengyel, P., The interferon system, *J. Biol. Chem.* **267**, 5017–5020 (1992).
- Sheehy, R.E., Kramer, M., and Hiatt, W.R., Reduction of polygalacturonase activity in tomato fruit by antisense RNA, *Proc. Natl. Acad. Sci.* **85**, 8805–8809 (1988).
- Sonenberg, N. and Hinnebusch, A.G., Regulation of translation initiation in eukaryotes: mechanisms and biological targets, *Cell* **136**, 731–745 (2009).
- Tafuri, S.R. and Wolffe, A.P., Dual roles for transcription and translation factors in the RNA storage particles of *Xenopus* oocytes, *Trends Cell. Biol.* **3**, 94–98 (1993).
- Tamm, I., Dörken, B., and Hartmann, G., Antisense therapy in oncology: new hope for an old idea? *Lancet* **358**, 489–497 (2001).

Post-Translational Modification

- Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E., Inteins: structure, function, and evolution, *Annu. Rev. Microbiol.* **56**, 263–287 (2002).
- Harding, J.J., and Crabbe, M.J.C. (Eds.), *Post-Translational Modifications of Proteins*, CRC Press (1992).
- Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R., Jr., and Sacchettini, J.C., Crystal structure of Gyr A protein from *Mycobacterium xenopi* reveals structural basis of splicing, *Nature Struct. Biol.* **5**, 31–36 (1998).
- Liu, X.-Q., Protein-splicing intein: genetic mobility, origin, and evolution, *Annu. Rev. Genet.* **34**, 61–76 (2000).
- Saleh, L. and Perler, F.B., Protein splicing in *cis* and in *trans*, *Chem. Rec.* **6**, 183–193 (2006).
- Wold, F., *In vivo* chemical modification of proteins, *Annu. Rev. Biochem.* **50**, 783–814 (1981).

Wold, F. and Moldave, K. (Eds.), *Posttranslational Modifications, Parts A and B, Methods Enzymol.* **106** and **107** (1984). [Contains extensive descriptions of the amino acid "zoo."]

Protein Degradation

- Bochtler, M., Ditzel, L., Groll, M., Hartmann, C., and Huber, R., The proteasome, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 295–317 (1999).
- Brandstetter, H., Kim, J.-S., Groll, M., and Huber, R., Crystal structure of the tricorn protease reveals a protein disassembly line, *Nature* **414**, 466–470 (2001); Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W., and Koster, A.J., Tricorn protease exists as an icosahedral supermolecule *in vivo*, *Mol. Cell* **1**, 59–65 (1997); and Walz, J., Koster, A.J., Tamura, T., and Baumeister, W., Capsids of tricorn protease studied by cryo-microscopy, *J. Struct. Biol.* **128**, 65–68 (1999).
- Cook, W.J., Jeffrey, L.C., Kasperek, E., and Pickart, C.M., Structure of tetraubiquitin shows how multiubiquitin chains can be formed, *J. Mol. Biol.* **236**, 601–609 (1994).
- Cook, W.J., Jeffrey, L.C., Sullivan, M.L., and Vierstra, R.D., Three-dimensional structure of a ubiquitin-conjugating enzyme (E2), *J. Biol. Chem.* **267**, 15116–15121 (1992).
- DeMartino, G.N. and Gillette, T.G., Proteasomes: machines for all reasons, *Cell* **129**, 659–662 (2007).
- Deshaines, R.J. and Joazeiro, C.A.P., RING domain E3 ubiquitin ligases, *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Dye, B.T. and Schulman, B.A., Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 131–150 (2007).
- Finley, D., Recognition and processing of ubiquitin–protein conjugates by the proteasome, *Annu. Rev. Biochem.* **78**, 477–513 (2009).
- Glickman, M.H. and Ciechanover, A., The ubiquitin–proteasome proteolytic pathway: destruction for the sake of construction, *Physiol. Rev.* **82**, 373–428 (2002).
- Hershko, A. and Ciechanover, A., The ubiquitin system, *Annu. Rev. Biochem.* **67**, 425–479 (1998).
- Jentsch, S. and Pyrowalakis, G., Ubiquitin and its kin: how close are the family ties, *Trends Cell Biol.* **10**, 335–342 (2003). [Discusses Ubls.]
- Liu, F. and Walters, K.J., Multitasking with ubiquitin through multivalent interactions, *Trends Biochem. Sci.* **35**, 352–360 (2010).
- Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R., Crystal structure of the 20S proteasome from the archeon *T. acidophilum* at 3.4 Å resolution, *Science* **268**, 533–539 (1995); and Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R., Structure of 20S proteasome from yeast at 2.4 Å resolution, *Nature* **386**, 463–471 (1997).
- Manchado, E., Eguren, M., and Malumbres, M., The anaphase-promoting complex/cyclosome (APC/C): cell-cycle-dependent and -independent functions, *Biochem. Soc. Trans.* **38**, 65–71 (2010).
- Mukhopadhyay, D. and Riezman, H., Proteasome-independent functions of ubiquitin in endocytosis and signaling, *Science* **315**, 201–205 (2007).
- Navon, A. and Ciechanover, A., The 26S proteasome: from basic mechanism to drug targeting, *J. Biol. Chem.* **284**, 33713–33718 (2009).
- Page, A.M. and Hieter, P., The anaphase-promoting complex: new subunits and regulators, *Annu. Rev. Biochem.* **68**, 583–609 (1999).
- Pickart, C.M., Back to the future with ubiquitin, *Cell* **116**, 181–190 (2004); and Mechanisms underlying ubiquitination, *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Schwartz, A.L. and Ciechanover, A., Targeting proteins for destruction by the ubiquitin system: implication for human pathobiology, *Annu. Rev. Pharmacol. Toxicol.* **49**, 73–96 (2009).
- Senahdi, V.-J., Bugg, C.E., Wilkinson, K.D., and Cook, W.J., Three-dimensional structure of ubiquitin at 2.8 Å resolution, *Proc. Natl. Acad. Sci.* **82**, 3582–3585 (1985).
- Skaug, B., Jiang, X., and Chen, Z.J., The role of ubiquitin in NF-κB regulatory pathways, *Annu. Rev. Biochem.* **78**, 769–796 (2009).
- Sousa, M.C., Trame, C.B., Tsuruta, H., Wilbanks, S.M., Reddy, V.J., and McKay, D.B., Crystal and solution structures of an Hs1UV protease–chaperone complex, *Cell* **103**, 633–643 (2000).
- Unno, M., Mizushima, T., Morimoto, Y., Tomisugi, Y., Tanaka, K., Yasuoka, N., and Tsukihara, T., The structure of the mammalian proteasome at 2.75 Å resolution, *Structure* **10**, 609–618 (2002).
- VanDemark, A.P. and Hill, C.P., Structural basis of ubiquitylation, *Curr. Opin. Struct. Biol.* **12**, 822–830 (2002).
- Varshavsky, A., Regulated protein degradation, *Trends Biochem. Sci.* **30**, 283–286 (2005).
- Varshavsky, A., Turner, G., Du, F., and Xie, Y., The ubiquitin system and the N-end rule, *Biol. Chem.* **381**, 779–789 (2000).
- Voges, D., Zwickl, P., and Baumeister, W., The 26S proteasome: a molecular machine designed for controlled proteolysis, *Annu. Rev. Biochem.* **68**, 1015–1068 (1999).
- Wang, J., Hartling, J.A., and Flanagan, J.M., The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis, *Cell* **91**, 447–456 (1997).
- Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., and Hill, C.P., Structural basis for the activation of 20S proteasomes by 11S regulators, *Nature* **408**, 115–120 (2000); and Förster, A., Whitby, F.G., and Hill, C.P., The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation, *EMBO J.* **22**, 4356–4354 (2003).
- Zheng, N., et al., Structure of the Cul1–Rbx1–Skp1–F-box^{Skp2} SCF ubiquitin ligase complex, *Nature* **41**, 703–709 (2002); Schulman, B.A., et al., Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex, *Nature* **408**, 381–386 (2000); and Zheng, N., Wang, P., Jeffrey, P.D., and Pavletich, N.P., Structure of a c-Cbl–UbcH7 complex: RING domain function in ubiquitin–protein ligases, *Cell* **102**, 533–539 (2000).
- Zwickl, P., Seemüller, E., Kapelari, B., and Baumeister, W., The proteasome: A supramolecular assembly designed for controlled proteolysis, *Adv. Prot. Chem.* **59**, 187–222 (2002).

PROBLEMS

1. What is the product of the reaction of guanine with nitrous acid? Is the reaction mutagenic? Explain.

2. What is the polypeptide specified by the following DNA antisense strand? Assume translation starts at the first initiation codon.

5'-TCTGACTATTGAGCTCTGGCACATAGCA-3'

***3.** The fingerprint of a protein from a phenotypically revertant mutant of bacteriophage T4 indicates the presence of an altered tryptic peptide with respect to the wild type. The wild-type and mutant peptides have the following sequences:

Wild type	Cys-Glu-Asp-His- Val-Pro-Gln-Tyr-Arg
Mutant	Cys-Glu-Thr- Met-Ser- His-Ser- Tyr-Arg

Indicate how the mutant could have arisen and give the base sequences, as far as possible, of the mRNAs specifying the two peptides. Comment on the function of the peptide in the protein.

4. Explain why the various classes of mutations can reverse a mutation of the same class but not a different class.

5. Which amino acids are specified by codons that can be changed to an *amber* codon by a single point mutation?

6. The mRNA specifying the α chain of human hemoglobin contains the base sequence

...UCCAAUACGUUAAGCUGGA...

The C-terminal tetrapeptide of the normal α chain, which is specified by part of this sequence, is

-Ser-Lys-Tyr-Arg

In hemoglobin Constant Spring, the corresponding region of the α chain has the sequence

-Ser-Lys-Tyr-Arg-Gln-Ala-Gly...

Specify the mutation that causes hemoglobin Constant Spring.

7. Explain why a minimum of 32 tRNAs are required to translate the "standard" genetic code.

8. Draw the wobble pairings not in Fig. 32-25a.

9. A colleague of yours claims that by exposing *E. coli* to HNO_2 she has mutated a tRNA^{Gly} to an *amber* suppressor. Do you believe this claim? Explain.

***10.** Deduce the anticodon sequences of all suppressors listed in Table 32-6 except UGA-1 and indicate the mutations that caused them.

11. How many different types of macromolecules must be minimally contained in a cell-free protein synthesizing system from *E. coli*? Count each type of ribosomal component as a different macromolecule.

12. Why do oligonucleotides containing Shine-Dalgarno sequences inhibit translation in prokaryotes? Why don't they do so in eukaryotes?

13. Why does m⁷GTP inhibit translation in eukaryotes? Why doesn't it do so in prokaryotes?

14. What would be the distribution of radioactivity in the completed hemoglobin chains on exposing reticulocytes to ³H-labeled leucine for a short time followed by a chase with unlabeled leucine?

15. Design an mRNA with the necessary prokaryotic control sites that codes for the octapeptide Lys-Pro-Ala-Gly-Thr-Glu-Asn-Ser.

***16.** Indicate the translational control sites in and the amino acid sequence specified by the following prokaryotic mRNA.

5'-CUGAUAGGAUUAAGGUCAUCACGAGCUAAUCACGUU CGAC-3'

17. What is the energetic cost, in ATP equivalents, for the *E. coli* synthesis of a polypeptide chain of 100 residues starting from amino acids and mRNA? Assume that no losses are incurred as a result of proofreading.

18. It has been suggested that Gly-tRNA synthetase does not require an editing mechanism. Why?

19. Explain why prokaryotic ribosomes can translate a circular mRNA molecule, whereas eukaryotic ribosomes normally cannot, even in the presence of the required cofactors.

20. EF-Tu binds all aminoacyl-tRNAs with approximately equal affinity so that it can deliver them to the ribosome with the same efficiency. Based on the experimentally determined binding constants for EF-Tu and correctly charged and mischarged aminoacyl-tRNAs (see table), explain how the tRNA-EF-Tu recognition system could prevent the incorporation of the wrong amino acid during translation.

Aminoacyl-tRNA	Dissociation Constant (nM)
Ala-tRNA ^{Ala}	6.2
Gln-tRNA ^{Ala}	0.05
Gln-tRNA ^{Gln}	4.4
Ala-tRNA ^{Gln}	260

Source: LaRiviere, F.J., Wolfson, A.D., and Uhlenbeck, O.C., *Science* **294**, 167 (2001).

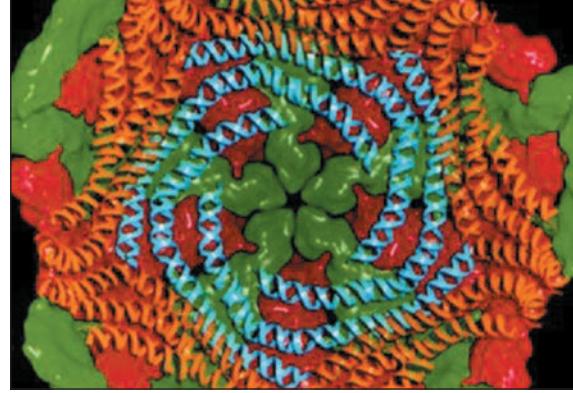
21. All cells contain an enzyme called **peptidyl-tRNA hydrolase**, and cells that are deficient in this enzyme grow very slowly. What is the probable function of this enzyme and why is it necessary?

22. An antibiotic named fixmycin, which you have isolated from a fungus growing on ripe passion fruit, is effective in curing several types of sexually transmitted diseases. In characterizing fixmycin's mode of action, you have found that it is a bacterial translational inhibitor that binds exclusively to the large subunit of *E. coli* ribosomes. The initiation of protein synthesis in the presence of fixmycin results in the generation of dipeptides that remain associated with the ribosome. Suggest a mechanism of fixmycin action.

23. Heme inhibits protein degradation in reticulocytes by allosterically regulating ubiquitin-activating enzyme (E1). What physiological function might this serve?

24. Genbux Inc., a biotechnology firm, has cloned the gene encoding an industrially valuable enzyme into *E. coli* such that the enzyme is produced in large quantities. However, since the firm wishes to produce the enzyme in ton quantities, the expense of isolating it would be greatly reduced if the bacterium could be made to secrete it. As a high-priced consultant, what general advice would you offer to solve this problem?

Viruses: Paradigms for Cellular Functions



CHAPTER 33

1 Tobacco Mosaic Virus

- A. Structure
- B. Assembly

2 Icosahedral Viruses

- A. Virus Architecture
- B. Tomato Bushy Stunt Virus
- C. Picornaviruses
- D. Simian Virus 40 (SV40)
- E. Bacteriophage MS2
- F. Bluetongue Virus
- G. Paramecium bursaria Chlorella Virus

3 Bacteriophage λ

- A. The Lytic Pathway
- B. Virus Assembly
- C. The Lysogenic Mode
- D. Mechanism of the λ Switch

4 Influenza Virus

- A. Virus Structure and Life Cycle
- B. Mechanism of Antigenic Variation
- C. Mechanism of Membrane Fusion

tigations have revealed that viruses are highly abundant in oceans, where they have an average concentration of $\sim 3 \times 10^9$ particles/L. Thus, viruses comprise the second largest component of the oceanic biomass after prokaryotes (they are estimated to collectively contain $\sim 10^{12}$ kg of carbon) and, since they probably infect all oceanic organisms, have a major influence on the biosphere.

An intact virus particle, which is referred to as a **virion**, consists of a nucleic acid molecule encased by a protein **capsid**. In some of the more complex virions, the capsid is surrounded by a lipid bilayer and glycoprotein-containing **envelope**, which is derived from a host cell membrane. Since the small size of a viral nucleic acid severely limits the number of proteins that can be encoded by its genome, its capsid, as Francis Crick and James Watson pointed out in 1957, must be built up of one or a few kinds of protein subunits that are arranged in a symmetrical or nearly symmetrical fashion. There are two ways that this can occur:

1. In the **helical viruses** (Section 33-1), the coat protein subunits associate to form helical tubes.

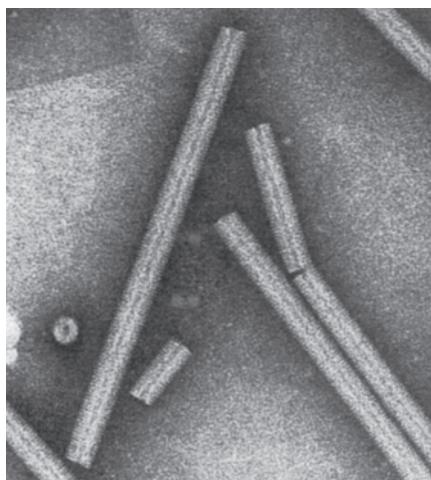
2. In the **icosahedral viruses** (also known as **spherical viruses**; Section 33-2), coat proteins aggregate as closed polyhedral shells.

In both cases, the viral nucleic acid occupies the capsid's central region. In many viruses, the coat protein subunits may be "decorated" by other proteins so that the capsid exhibits spikes and, in certain larger bacteriophages, a complex tail. These assemblies are involved in recognizing the host cell and delivering the viral nucleic acid into its interior. Figure 33-1 is a "rogues' gallery" of viruses of varying sizes and morphologies.

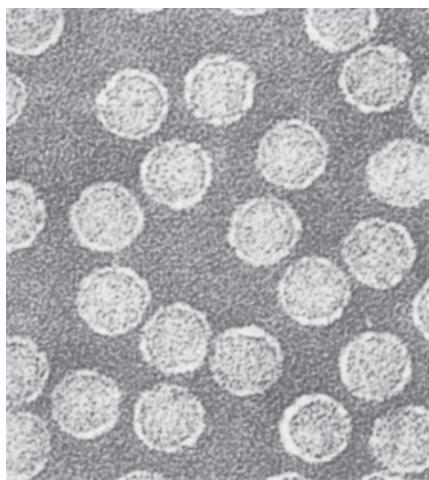
The great simplicity of viruses in comparison to cells made them invaluable tools in the elucidation of gene structure and function, as well as some of our best characterized models for the assembly of biological structures. Although all viruses use ribosomes and other host factors for the RNA-instructed synthesis of proteins, their modes of genome replication are far more varied than that of cellular life. In contrast to cells, in which the hereditary molecules are invariably double-stranded DNA, viruses contain either single- or double-stranded DNA or RNA. In RNA viruses, the viral RNA may be directly replicated or act as a template in the synthesis of DNA. The RNA of single-stranded RNA viruses may be the positive strand (the

Viruses (*Latin: poison*) are parasitic entities, consisting of nucleic acid molecules with protective coats, that are replicated by the enzymatic machinery of suitable host cells. Since they lack metabolic apparatus, viruses are not considered to be alive (although this is a semantic rather than a scientific distinction). They range in complexity from **satellite tobacco mosaic virus (STMV**; Section 33-2Bc), whose 1059-nt genome encodes only one 159-residue protein, to **mimivirus**, whose 1181-kb genome encodes ~ 911 proteins [to put this latter number into perspective, the smallest known genome of a cellular organism, that of the obligate insect endosymbiont *Carsonella ruddii* (Table 7-3), consists of 160 kb and encodes 182 proteins].

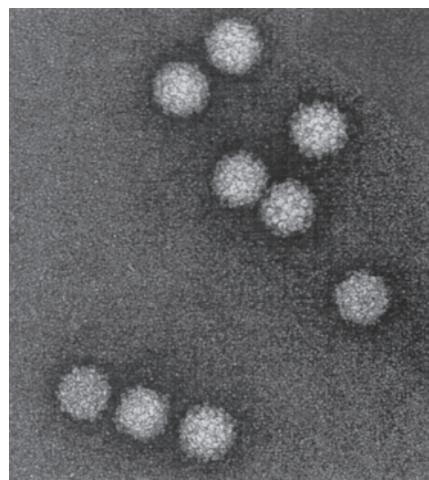
Viruses were originally characterized at the end of the nineteenth century as infectious agents that could pass through filters that held back bacteria. Yet viral diseases, varying in severity from smallpox and rabies to the common cold, have no doubt plagued mankind since before the dawn of history. It is now known that viruses can infect plants and bacteria as well as animals. Each viral species has a very limited **host range**; that is, it can reproduce in only a small group of closely related species. Recent investi-



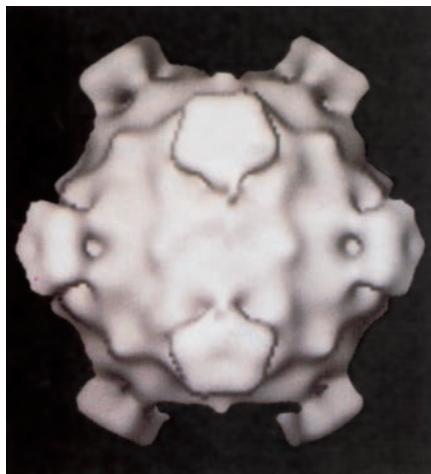
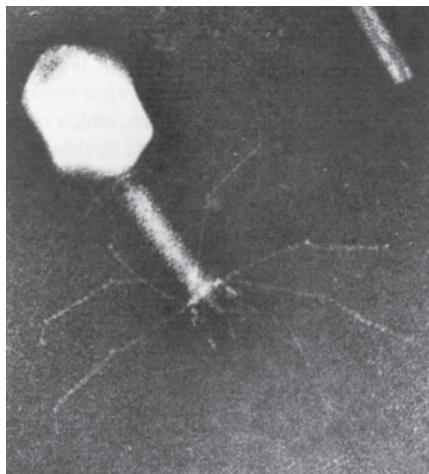
(a) Tobacco mosaic virus (TMV)



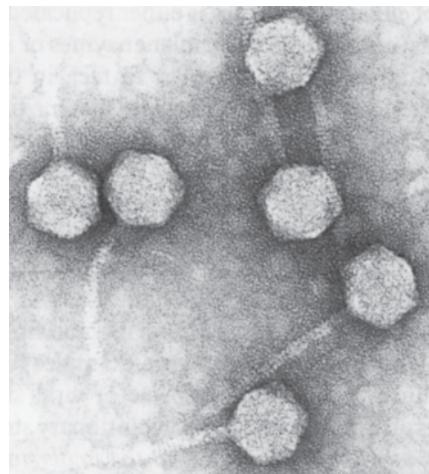
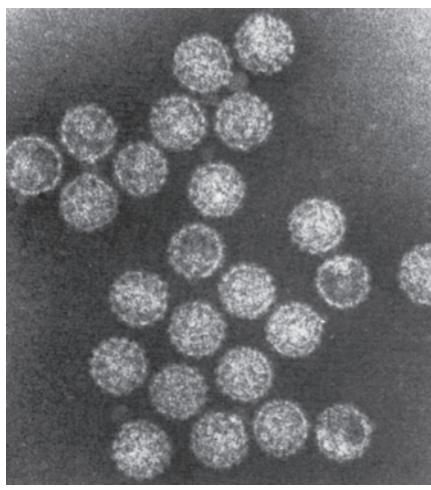
(b) Bacteriophage MS2



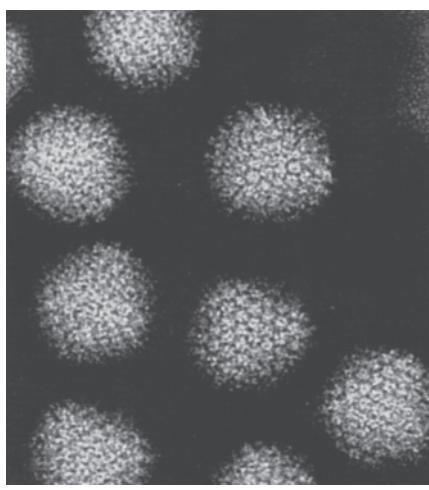
(c) Tobacco bushy stunt virus (TBSV)

(d) Bacteriophage ϕ X174

(e) Bacteriophage T4

(f) Bacteriophage λ 

(g) Simian virus 40 (SV40)



(h) Adenovirus



(i) Influenza virus

Figure 33-1 Electron micrographs of a selection of viruses. TMV, MS2, TBSV, and influenza virus are single-stranded RNA viruses; ϕ X174 is a single-stranded DNA virus; and λ , T4, SV40, and adenovirus are double-stranded DNA viruses. Bacteriophage M13, a filamentous, single-stranded DNA coliphage, is shown in

Fig. 5-43. [Parts a–c and f–i courtesy of Robley Williams, University of California at Berkeley, and Harold Fisher, University of Rhode Island; Part d courtesy of Michael Rossmann, Purdue University; and Part e courtesy of John Finch, Cambridge University, U.K.]

mRNA) or the negative strand (complementary to the mRNA). Viral DNA may replicate autonomously or be inserted in the host chromosome for replication with the host DNA. The DNA of eukaryotic viruses is either replicated and transcribed in the cell nucleus by cellular enzymes or in the cytoplasm by virally specified enzymes. In fact, in the case of negative strand RNA viruses, enzymes that mediate viral RNA transcription must be carried by the virion because most cells lack the ability to transcribe RNA.

This chapter is a discussion of the structures and biology of a variety of viruses. In it, we examine mainly **tobacco mosaic virus (TMV)**, a helical RNA virus; several **icosahedral viruses**; **bacteriophage λ** , a tailed DNA bacteriophage; and **influenza virus**, an enveloped RNA virus. These examples have been chosen to illustrate important aspects of viral structure, assembly, molecular genetics, and evolutionary strategy. *Much of this information is relevant to the understanding of the corresponding cellular phenomena.*

1 TOBACCO MOSAIC VIRUS

Tobacco mosaic virus causes leaf mottling and discoloration in tobacco and many other plants. It was the first virus to be discovered (by Dmitri Iwanowsky in 1892), the first virus to be isolated (by Wendell Stanley in 1935), and even now is among the most extensively investigated and well-understood viruses from the standpoint of structure and assembly. In this section, we discuss these aspects of TMV.

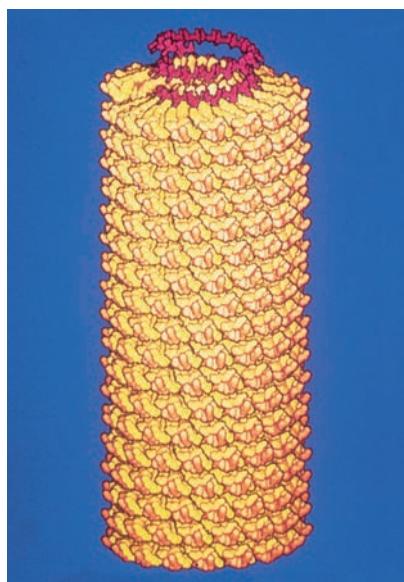


Figure 33-2 Model of tobacco mosaic virus (TMV) illustrating the helical arrangement of its coat protein subunits and RNA molecule. The RNA is represented by the red chain exposed at the top of the viral helix. Only 18 turns (415 Å) of the TMV helix are shown, which represent ~14% of the TMV rod. [Courtesy of Gerald Stubbs and Keiichi Namba, Vanderbilt University; and Donald Caspar, Brandeis University.]

A. Structure

TMV is a rod-shaped particle (Fig. 33-1a) that is ~3000 Å long, 180 Å in diameter, and has a particle mass of 40 million D. Its ~2130 identical copies of coat protein subunits (158 amino acid residues; 17.5 kD) are arranged in a hollow right-handed helix that has 16 1/3 subunits/turn, a pitch (rise per turn) of 22.92 Å, and a 40-Å-diameter central cavity (Fig. 33-2). TMV's single RNA strand (~6400 nt; ~2 million D) is coaxially wound within the turns of the coat protein helix such that 3 nt are bound to each protein subunit (Fig. 33-2).

a. TMV Coat Protein Aggregates to Form Viruslike Helical Rods

The aggregation state of TMV coat protein is both pH and ionic strength dependent (Fig. 33-3). At slightly alkaline pH's and low ionic strengths, the coat protein forms complexes of only a few subunits. At higher ionic strengths, however, the subunits associate to form a double-layered disk of 17 subunits/layer, a number that is nearly equal to the number of subunits per turn in the intact virion. At neutral pH and low ionic strengths, the subunits form short helices of slightly more than two turns (39 ± 2 subunits) termed “protohelices” (also known as “lockwashers”). If

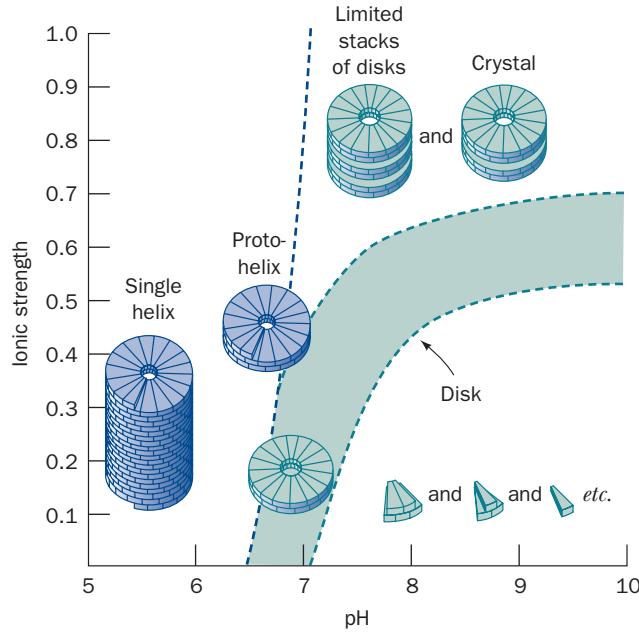


Figure 33-3 Aggregation state of TMV coat protein as a function of pH and ionic strength. Under basic conditions, the subunits aggregate into small clusters. Around neutrality and at high ionic strengths, the protein forms a 34-subunit double-layered disk. Under acidic conditions and at low ionic strengths, the subunits form protohelices that stack to form long helices. At neutral pH and low ionic strength, which resembles physiological conditions, the protein forms helices only in the presence of TMV RNA. [After Durham, A.C.H., Finch, J.T., and Klug, A., *Nature New Biol.* **229**, 38 (1971).]

the pH of these protohelices is shifted to ~ 5 , they stack in imperfect register and eventually anneal to form indefinitely long helical rods that, although they lack RNA, resemble intact virions (Fig. 33-4). These observations, as we shall see below, lead to the explanation of how TMV assembles.

b. TMV Coat Protein Interacts Flexibly with Viral RNA

X-ray studies of TMV have been pursued on two fronts. The virus itself does not crystallize but forms a highly oriented gel of parallel viral rods. The X-ray analysis of this gel by Kenneth Holmes and Gerald Stubbs yielded a

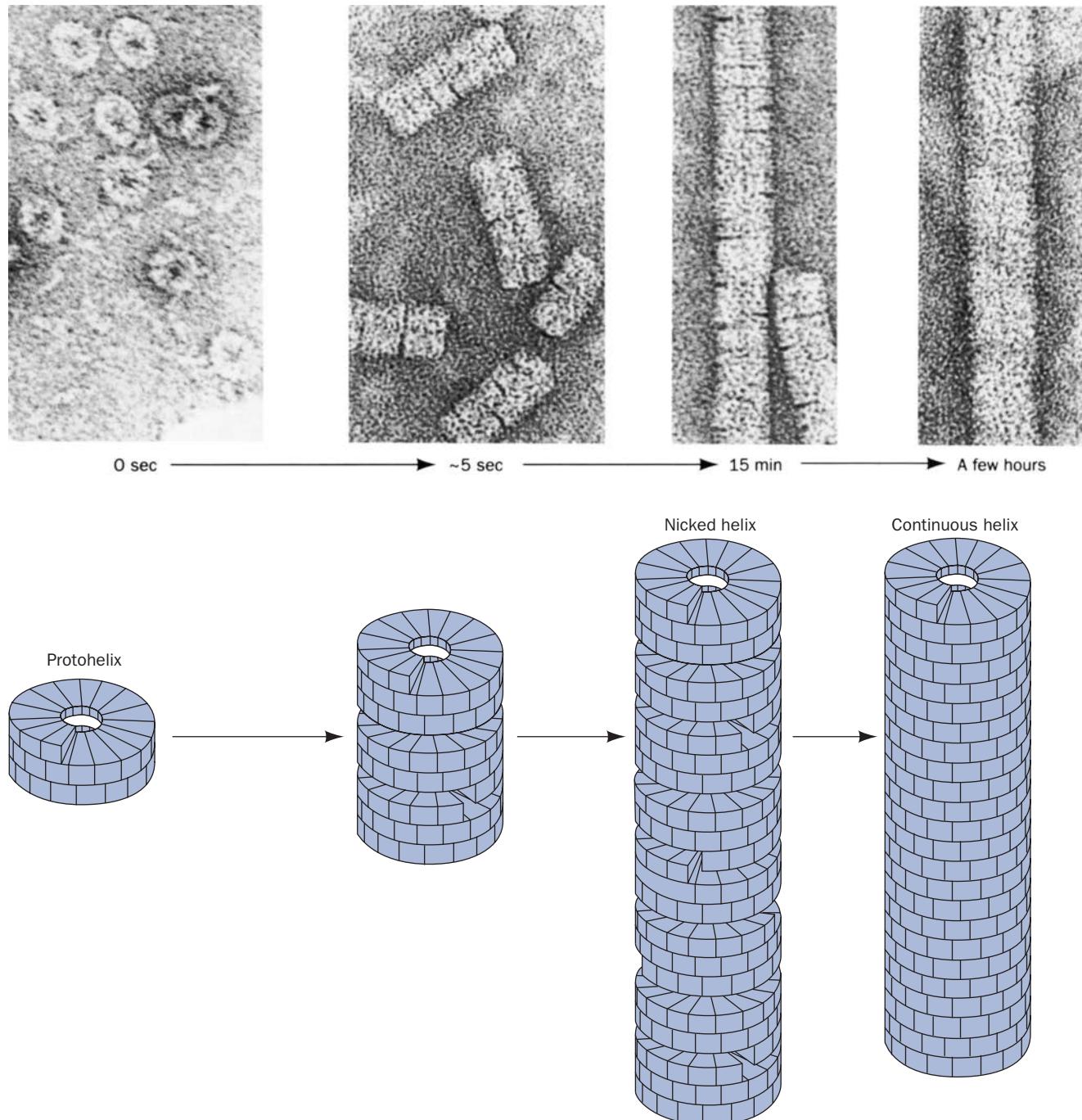


Figure 33-4 Growth of TMV coat protein rods. Electron micrographs (above) and their interpretive diagrams (below) show TMV coat protein aggregates following a rapid change in pH from 7 to 5 at low ionic strength. This pH shift causes the

protohelices to form “nicked” (imperfectly stacked) helices that, within a few hours, anneal to yield continuous helical protein rods. [Courtesy of Aaron Klug, MRC Laboratory of Molecular Biology, Cambridge, U.K.]

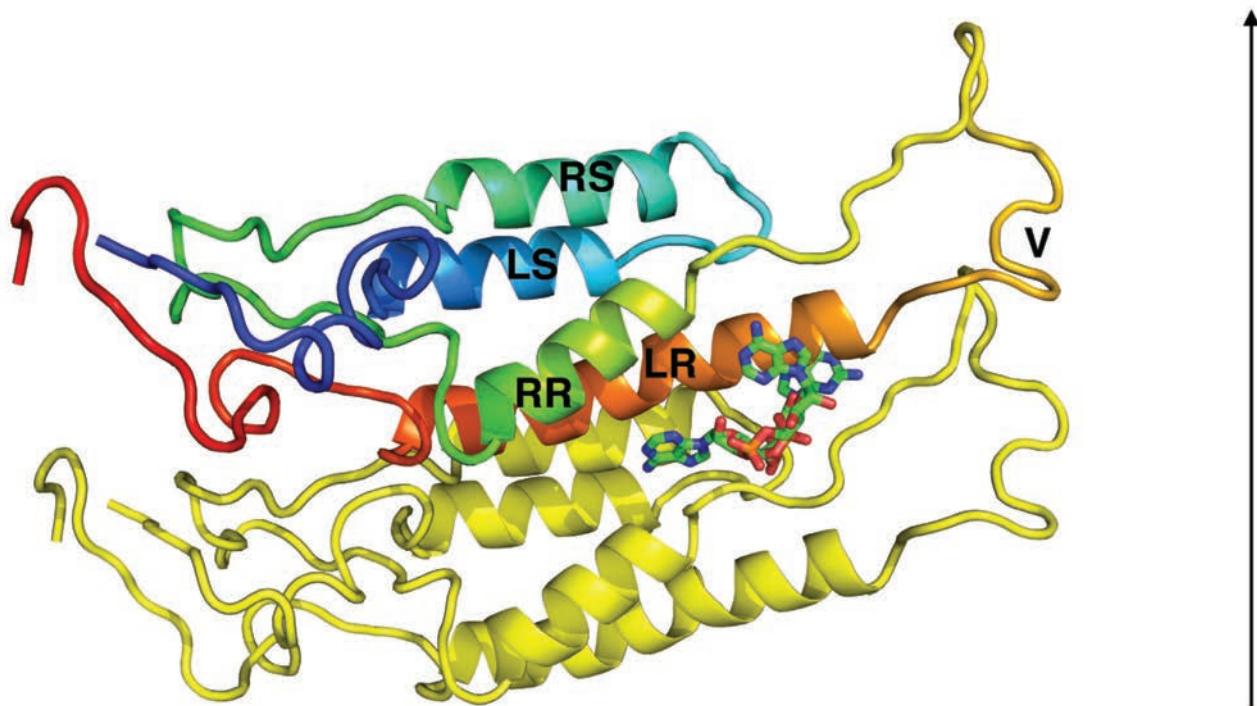


Figure 33-5 X-ray structure of two vertically stacked TMV subunits. The two subunits are shown in ribbon form with the upper subunit colored in rainbow order from its N-terminus (blue) to its C-terminus (red), and with the lower subunit yellow. The structure is viewed perpendicular to the virus helix axis (vertical arrow on the right). Each subunit has four approximately radially extending helices (LR, RR, LS, and RS), as well as a short vertical segment (V), which comprises part of the flexible

loop in the disk structure (dashed lines in Fig. 33-7). A trinucleotide segment of the RNA, here represented by GAA and drawn in stick form with C green, N blue, O red, and P orange, is bound between the two subunits. Its three bases lie flat against the LR helix of the upper subunit so as to grasp it in a clawlike manner. [Based on an X-ray structure by Gerald Stubbs and Keiichi Namba, Vanderbilt University. PDBid 2TMV.]

structure of sufficient resolution (2.9 Å) to reveal the folding of the protein and the RNA (Figs. 33-5 and 33-6). This study is complemented by Aaron Klug's X-ray crystal structure determination, at 2.8-Å resolution, of the 34-subunit coat protein disk (Fig. 33-7).

A major portion of each subunit consists of a bundle of four alternately parallel and antiparallel α helices that project more or less radially from the virus axis (Figs. 33-5 to 33-7). In the disk, one of the inner connections

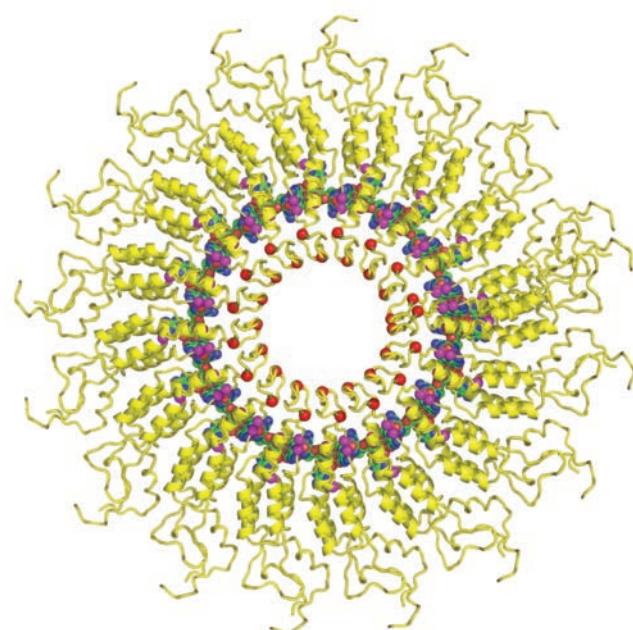


Figure 33-6 Top view of 17 TMV coat protein subunits comprising slightly more than one helical turn in complex with a 51-nucleotide RNA segment of sequence (GAA)₁₇. The subunits are shown in ribbon form in yellow and the RNA is drawn in space-filling form with C green, N blue, O red, and P orange. The C_α atoms of Glu 95 and 106 are represented by red spheres and those of Arg 41, 90, and 92 are represented by magenta spheres. Note that the acidic residues form an ~25-Å-radius helix that lines the virion's inner cavity and the basic residues form an ~40-Å-radius helix that interacts with the RNA's anionic sugar-phosphate chain. [Based on an X-ray structure by Gerald Stubbs and Keiichi Namba, Vanderbilt University. PDBid 2TMV.]

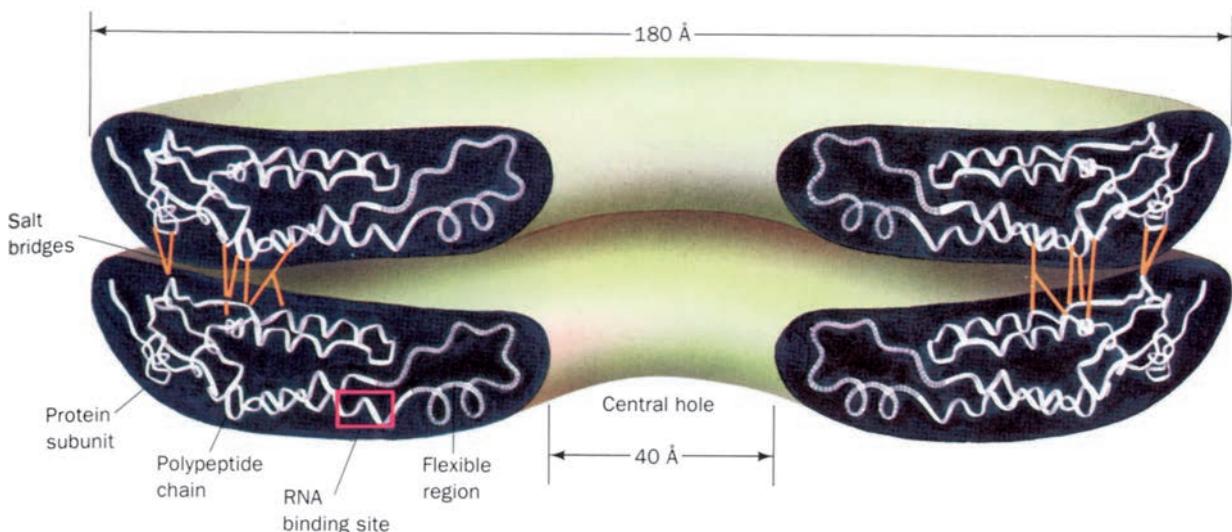


Figure 33-7 Structure of the TMV protein disk in cross section showing its polypeptide chains as ribbon diagrams. The dashed lines represent disordered loops of polypeptide chain that are therefore not visible in the disk X-ray structure. The stacked

protein rings interact along their outer rims through a system of salt bridges (red lines). [After Butler, P.J.G. and Klug, A., *Sci. Am.* **239**(5): 67 (1978). Copyright © 1978 by Scientific American, Inc.]

between these α helices, a 24-residue loop (residues 90–113; dashed line in Fig. 33-7), is not visible, apparently because it is highly mobile. This disordered loop is also present in the protohelix as shown by NMR studies. In the virus, however, the loop adopts a definite conformation containing a series of reverse turns arranged such that the overall direction of this polypeptide segment is approximately parallel to the virus axis (V in Fig. 33-5). This conformational change, as we shall see, is an important aspect of virus assembly.

In the virus, the RNA is helically wrapped between the coat protein subunits at a radius of ~ 40 Å. The triplet of bases binding to each subunit forms a clawlike structure around one of the radial helices (LR; Fig. 33-5) with each base occupying a hydrophobic pocket in which it lies flat against LR. Arg residues 90 and 92, which are invariant in the several known TMV strains and which are part of the disk and protohelix's disordered loop, as well as Arg 41, form salt bridges with the RNA phosphate groups (Fig. 33-6).

B. Assembly

How is the TMV virion assembled from its component RNA and coat protein subunits? *The assembly of any large molecular aggregate, such as a crystal or a virus, generally occurs in two stages: (1) nucleation, the largely random aggregation of subunits to form a quasi-stable nucleation complex, which is almost always the rate-determining step of the assembly process; followed by (2) growth, the cooperative addition of subunits to the nucleation complex in an orderly arrangement that usually proceeds relatively rapidly.* For TMV, it might reasonably be expected that the nucleation complex minimally consists of the viral

RNA in association with the 17 or 18 subunits necessary to form a stable helical turn, which could then grow by the accumulation of subunits at one or both ends of the helix. The low probability for the formation of such a complicated nucleation complex from disaggregated subunits accounts for the observed 6-h time necessary to complete this *in vitro* assembly process. Yet, the *in vivo* assembly of TMV probably occurs much faster. A clue as to the nature of this *in vivo* process was provided by the observation that if protohelices rather than disaggregated subunits are mixed with TMV RNA, complete virus particles are formed in 10 min. Other RNAs do not have this effect. Evidently, *the in vivo nucleation complex in TMV assembly is the association of a protohelix with a specific segment of TMV RNA.* (Although it was originally assumed that the double-layered disk rather than the protohelix formed the nucleating complex, experimental evidence indicates that the disk does not form under physiological conditions and that its rate of conversion to the protohelix under these conditions is too slow to account for the rate of TMV assembly. Other experiments, however, suggest that it is the disk that predominates at pH 7.0, the pH at which TMV most rapidly assembles from its component protein and RNA. Thus, keep in mind that the question as to whether TMV assembles from protohelices, as we state here, or from double-layered disks has not been fully resolved.)

a. TMV Assembly Proceeds by the Sequential Addition of Protohelices

The specific region of the TMV RNA responsible for initiating the virus particle's growth was isolated using the now classic nuclease protection technique. The RNA is

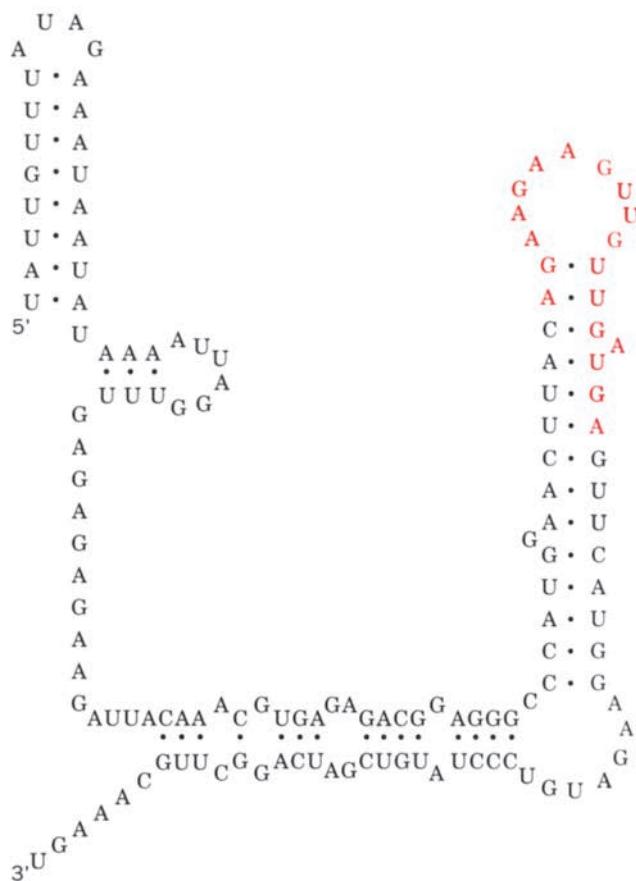


Figure 33-8 Initiation segment of TMV RNA. It probably forms a weakly base paired hairpin, as drawn, that is thought to begin TMV assembly by specifically binding to a coat protein protohelix. Note that this RNA's loop region has an 18-nt segment (red) with a G every third residue (each coat protein subunit binds three nucleotides) but no C's.

mixed with a small amount of coat protein so as to form a nucleation complex that cannot grow because of the lack of coat protein. The RNA that is not protected by coat protein is then digested away by RNase, leaving intact only the initiation sequence. This RNA fragment forms a hairpin loop whose 18-nucleotide apical sequence, AGAAGAAGUUGUUGAUGA, has a G at every third residue (recall that each coat protein subunit binds three nucleotides) but no C's (Fig. 33-8). Site-directed mutagenesis studies have confirmed that this initiation sequence is sufficient to direct TMV assembly and that the regularly spaced G's and lack of C's are important for its function. TMV's high binding affinity for this initiation sequence is explained, in part, by the observations that coat protein subunits bind every third nucleotide in the unusual syn conformation and that G assumes this conformation more easily than any other nucleotide (Section 29-2Aa). The lack of C's perhaps prevents the involvement of these G's in base pairing associations.

The above initiation complex is located some 1000 nucleotides from the 3' end of the TMV RNA. Hence, the simple model of viral assembly in which the RNA is sequentially coated by protein from one end to the other cannot be correct. Rather, the RNA initiation hairpin must insert itself between the protohelix's protein layers from its central cavity (Fig. 33-9a). The RNA binding, for reasons explained below, induces the ordering of the disordered loop, thereby trapping the RNA (Fig. 33-9b). Growth then proceeds by a repetition of this process at the "top" of the complex, thereby incrementally pulling the RNA's 5' end up through the central cavity of the growing viral helix (Fig. 33-9c).

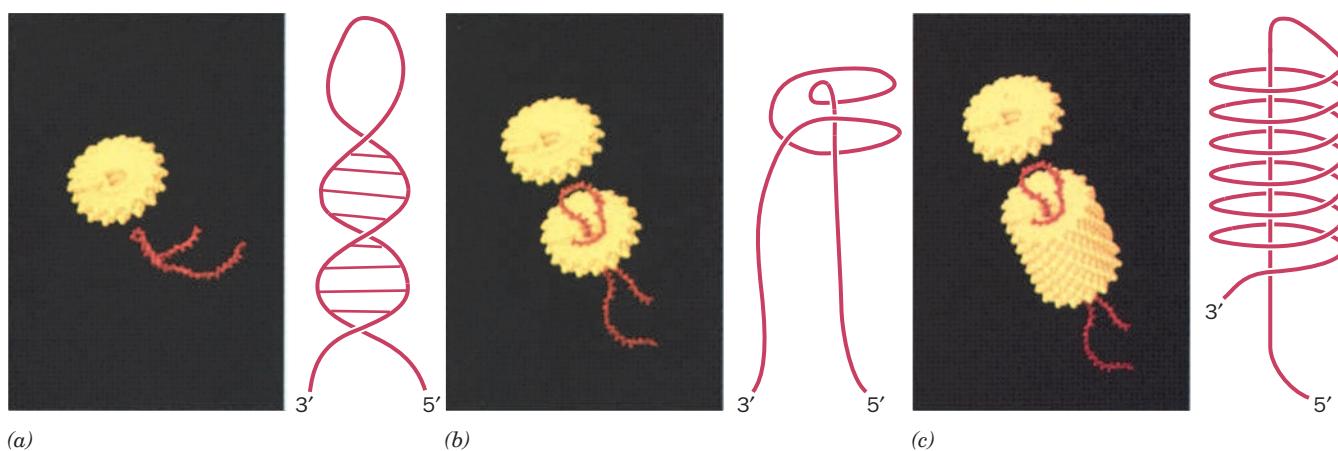


Figure 33-9 Assembly of TMV. (a) The process begins by the insertion of the hairpin loop formed by the initiation sequence of the viral RNA into the protohelix's central cavity. (b) The RNA then intercalates between the layers of the protohelix, thereby ordering the disordered loop and trapping the RNA. (c) Elongation proceeds by the stepwise addition of protohelices

to the “top” of the viral rod. The consequent binding of the RNA to each protohelix, which converts it to the helical form, pulls the RNA’s 5’ end up through the virus’ 40-Å-diameter central cavity to form a traveling loop at the viral rod’s growing end. [Viral images courtesy of Hong Wang and Gerald Stubbs, Vanderbilt University.]

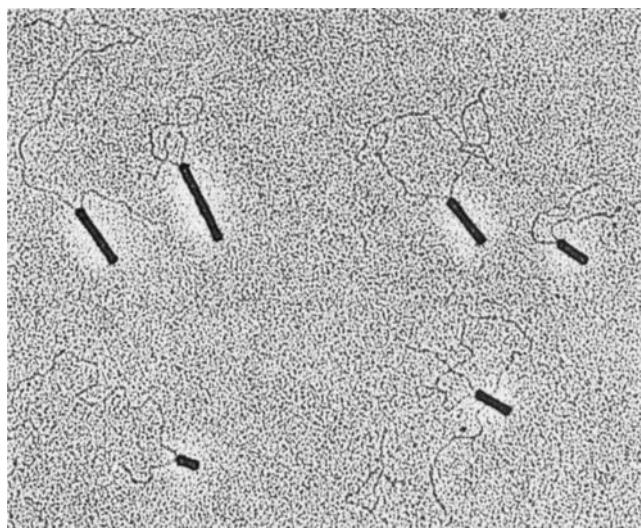


Figure 33-10 Electron micrograph of partially reconstituted TMV particles showing that their two RNA tails emerge from the same end of the growing viral rod. An analysis of these particles indicates that the length of one of the tails, probably the 3' end, is constant (720 ± 80 nucleotides), whereas that of the other tail is inversely proportional to the length of its incomplete rod. [Courtesy of K.E. Richards, CNRS, France.]

The above assembly model has been corroborated by several experimental observations:

1. Electron micrographs reveal that partially assembled rods (Fig. 33-10) have two RNA “tails” projecting from one end.
2. The length of the longer tail, presumably the 5' end, decreases linearly with the length of the rod, whereas the shorter tail maintains a more or less constant length.
3. Nuclease digestion experiments on partially assembled rods indicate that the RNA is protected in increments of ~ 100 nucleotides, as is expected for elongation steps consisting of the addition of a protohelix to the growing rod.

The coating of the 3' end of the RNA is a much slower process than the coating of its 5' end and hence probably occurs by the successive addition of single subunits. The RNA, which acts as the viral mRNA, carries the gene specifying the coat protein near its 3' end. Perhaps this assembly mechanism allows coat protein synthesis during all but the final stages of assembly, thereby permitting the completion of this process.

b. Electrostatic Repulsions and Steric Interactions

Prevent Helix Formation in the Absence of RNA

What is the mechanism that prevents the formation of TMV coat protein helices in the absence of viral RNA but triggers virus assembly in its presence (and, conversely, how does intact TMV disassemble to initiate an infection)? Structural considerations suggest that the coat protein subunit's disordered loop sterically prevents the protohelix from growing longer. Moreover, as we have seen (Fig. 33-3), the state of coat protein aggregation varies with pH.

Titration studies show that each subunit has two ionizations with pK 's near 7, which must each be attributed to anomalously basic carboxyl groups because coat protein has no His residues. The most plausible candidates for these anomalously basic carboxyls are two intersubunit pairs of carboxyl groups: Glu 95–Glu 106, disordered loop members which interact across a side-to-side subunit interface (Fig. 33-6); and Glu 50–Asp 77, which interact across a top-to-bottom subunit interface. Moreover, Asp 116 is close to an RNA phosphate group. The electrostatic repulsions between these closely spaced negative charges promotes the formation of the disordered loop and therefore favors the protohelix conformation. The binding of the RNA initiation sequence to the protein apparently provides sufficient free energy to overcome these repulsions, thereby triggering helix formation (a process that partially protonates the anomalously basic carboxyl groups; recall the similar conformationally induced pK changes in the Bohr effect of hemoglobin; Section 10-2E). Indeed, site-directed mutagenesis of Glu 50 \rightarrow Gln or Asp 77 \rightarrow Asn both increases virion stability and decreases its infectivity (presumably by inhibiting viral disassembly). Further growth of the viral rod can occur on RNA segments that lack this sequence as a consequence of the additional binding interactions between adjacent protohelices. *The carboxyl groups evidently act as a negative switch to prevent the formation of a protein helix in the absence of RNA under physiological conditions.*

2 ICOSAHEDRAL VIRUSES

The simpler **icosahedral viruses**, being uniform molecular assemblies, crystallize in much the same way as proteins. The techniques of X-ray crystallography can therefore be brought to bear on determining virus structures. In this section we consider the results of such studies.

A. Virus Architecture

The very limited genomic resources of the simpler viruses in many cases limit them to having but one type of protein in their capsid. Since these coat protein subunits are chemically identical, they must all assume the same or nearly the same conformations and have similar interactions with their neighbors. What geometrical constraints does this limitation impose on viral architecture?

We have already seen that TMV solves this problem by assuming a helical geometry (Fig. 33-2). The coat protein subunits in such a long but finite helix, although geometrically distinguishable, have, with the exception of the subunits at the helix ends, virtually identical environments. Such subunits are said to be **quasi-equivalent** to indicate that they are not completely indistinguishable as they would be in an object whose elements are all related by exact symmetry.

a. Icosahedral Viruses Have Icosahedral Capsids

A second arrangement of equivalent subunits that can encapsulate a nucleic acid is that of a polyhedral shell. There are only three polyhedral symmetries in which all the

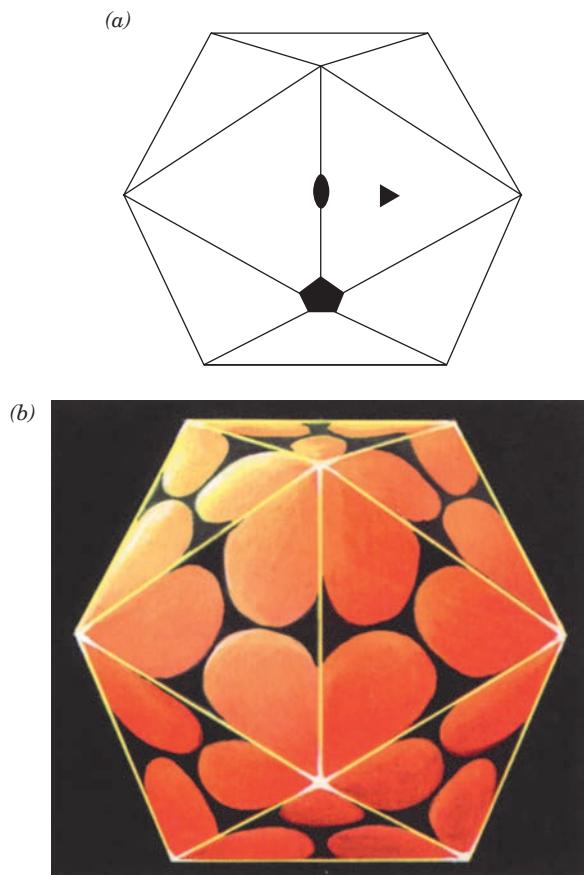


Figure 33-11 Icosahedron. (a) This regular polyhedron has 12 vertices, 20 equilateral triangular faces of identical size, and 30 edges. It has a 5-fold axis of symmetry through each vertex, a 3-fold axis through the center of each face, and a 2-fold axis through the center of each edge (also see Fig. 8-65c). (b) A drawing of 60 identical subunits (*lobes*) arranged with icosahedral symmetry. [Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]

elements are indistinguishable: those of a tetrahedron, a cube, and an icosahedron (Fig. 8-65c). Capsids with these symmetries would have 12, 24, or 60 subunits identically arranged on the surface of a sphere. For example, an icosahedron (Fig. 33-11a) has 20 triangular faces, each with 3-fold symmetry, for a total of $20 \times 3 = 60$ equivalent positions (each represented by a lobe in Fig. 33-11b). Of these polyhedra, the icosahedron encloses the greatest volume per subunit. Indeed, electron microscopy of the so-called icosahedral viruses (such as Fig. 33-1b–h) first demonstrated that they have icosahedral symmetry.

b. Viral Capsids Resemble Geodesic Domes

A viral nucleic acid, if it is to be protected effectively against a hostile environment, must be completely covered by coat protein. Yet, many viral nucleic acids occupy so large a volume that their coat protein subunits would have to be prohibitively large if their capsids were limited to the 60 subunits required by exact icosahedral symmetry. In fact, nearly all viral capsids have considerably more than 60 chemically identical subunits. How is this possible?

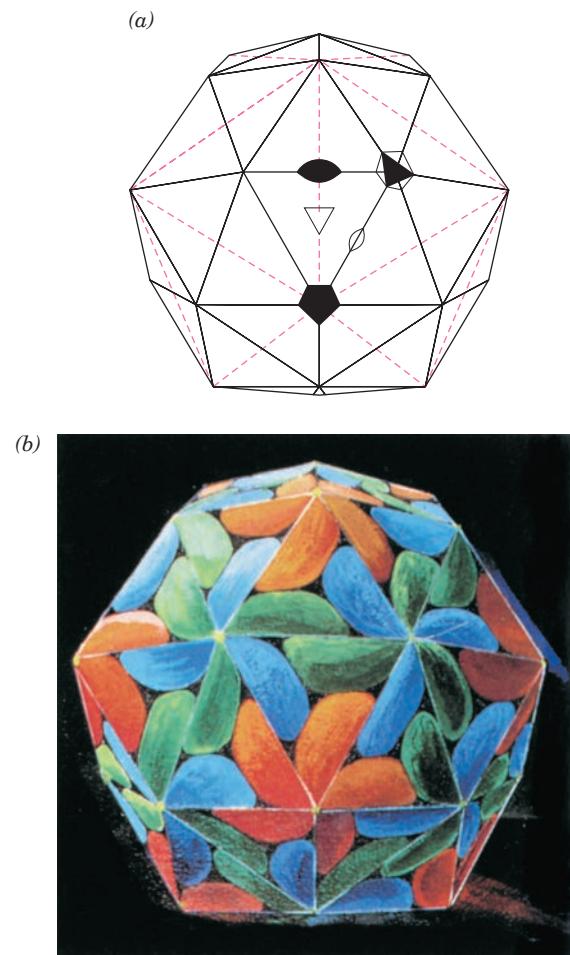


Figure 33-12 $T = 3$ icosadeltahedron. (a) This polyhedron has the exact rotational symmetry of an icosahedron (*solid symbols*) together with local 6-fold, 3-fold, and 2-fold rotational axes (*hollow symbols*). Note that the edges of the underlying icosahedron (*dashed red lines*), are not edges of this polyhedron and that its local 6-fold axes are coincident with its exact 3-fold axes. (b) A drawing of a $T = 3$ icosadeltahedron showing its arrangement of three quasi-equivalent sets of 60 icosahedrally related subunits (*lobes*). The A lobes (*orange*) pack about the icosadeltahedron's exact 5-fold axes, whereas the B and C lobes (*blue and green*) alternate about its local 6-fold axes. TBSV's chemically identical coat protein subunits are arranged in this manner. [Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]

Donald Caspar and Klug pointed out the solution to this dilemma. *The triangular faces of an icosahedron can be subdivided into integral numbers of equal sized equilateral triangles* (e.g., Fig. 33-12a). The resulting polyhedron, an **icosadeltahedron**, has “local” symmetry elements relating its subunits (lobes in Fig. 33-12b) in addition to its exact icosahedral symmetry. By local symmetry, we mean that the symmetry is only approximate so that, in contrast to the case for exact symmetry, it breaks down over larger distances. For instance, the subunits (lobes) in Fig. 33-12b that are distributed about each exact triangular vertex form clusters whose members are related by a local 6-fold axis of symmetry. *Adjacent subunits in these clusters are not exactly*



Figure 33-13 Geodesic dome built on the plan of a $T = 36$ icosadeltahedron. One of its pentagonal vertices is clearly visible in the upper center of this photograph. [Steve Vidler/SuperStock]

equivalent; they are quasi-equivalent. In contrast, the subunits clustered about the twelve 5-fold axes of icosahedral symmetry are exactly equivalent. The interactions between the subunits clustered about the local 6-fold axes are therefore essentially distorted versions of those about the exact 5-fold axes. Consequently, the coat protein subunits of any viral capsid with icosadeltahedral symmetry must make alternative sets of intersubunit associations and/or have sufficient conformational flexibility to accommodate these distortions.

Icosadeltahedra are familiar figures. The faceted surface of a soccer ball is an icosadeltahedron. Likewise, **geodesic domes** (Fig. 33-13), which were originally designed by Buckminster Fuller, are portions of icosadeltahedra. It was, in fact, Fuller's designs that inspired Caspar and Klug. *Geodesic domes are inherently rigid shell-like structures that are constructed from a few standard parts, make particularly efficient use of structural materials, and can be rapidly and easily assembled. Presumably the evolution of icosahedral virus capsids was guided by these very principles.*

The number of subunits in an icosadeltahedron is $60T$, where T is called the **triangulation number**. The permissible values of T are given by $T = h^2 + hk + k^2$, where h and k are positive integers. An icosahedron, the simplest icosadeltahedron, has $T = 1$ ($h = 1, k = 0$) and therefore 60 subunits. The icosadeltahedron with the next level of complexity has a triangulation number of $T = 3$ ($h = 1, k = 1$) and hence 180 subunits (Fig. 33-12). A capsid with this geometry has three different sets of icosahedrally related subunits that are quasi-equivalent to each other (lobes A, B, and C in Fig. 33-12b). The X-ray structures of viruses with capsids consisting of $T = 1, 3, 4, 7$, and 13 icosadeltahedra have been determined. Some of the larger icosahedral viruses form icosadeltahedra with even greater triangulation numbers (see below). However, some of them are

based on somewhat different assembly principles (Section 33-2D). The T value for any particular capsid, presumably, depends on its subunit's innate curvature.

B. Tomato Bushy Stunt Virus

Tomato bushy stunt virus (TBSV; Fig. 33-1c) is a $T = 3$ icosahedral virus with a radius of 177 Å. It consists of 180 identical coat protein subunits, each of 386 residues (43 kD), encapsulating a single-stranded RNA molecule of ~4800 nt (1500 kD; the positive or message strand) and a single copy of an ~85-kD protein. The X-ray crystal structure of TBSV, the first of a virus to be determined at atomic resolution, was reported in 1978 by Stephen Harrison. TBSV's coat protein subunits have three domains (Fig. 33-14): P, the C-terminal domain, which projects outward from the virus; S, which forms the viral shell; and R, the protein's inwardly extending N-terminal domain, which is attached to the S domain via a connecting arm. The S domain is almost entirely composed of an 8-stranded antiparallel β

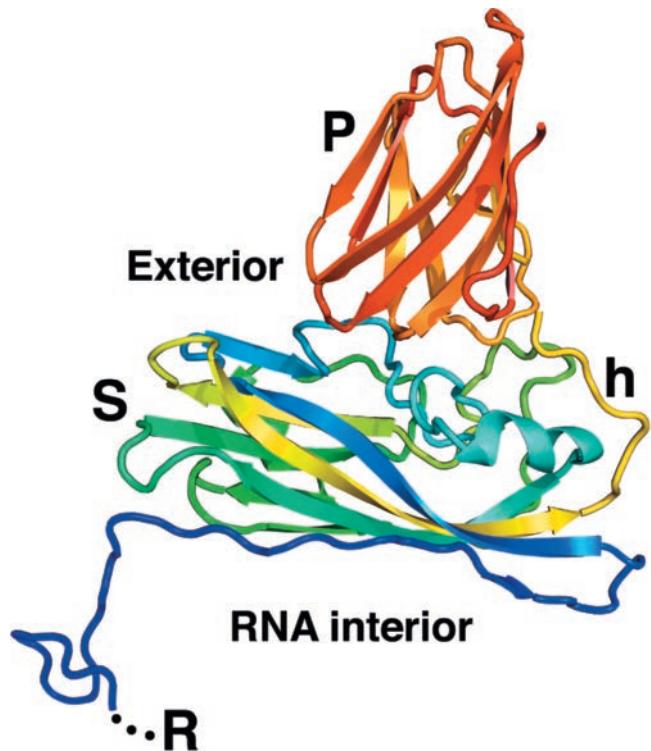


Figure 33-14 X-ray structure of the tomato bushy stunt virus (TBSV) coat protein subunit. It is drawn in ribbon form colored in rainbow order from N-terminus (blue) to C-terminus (red) and viewed with the RNA interior below. Each subunit consists of three domains: P, which projects from the virion's surface (yellow through red); S, which forms the capsid (blue through yellow); and R, which extends below the capsid surface where it participates in binding the viral RNA. The S domain is largely comprised of an 8-stranded antiparallel β barrel that has the jelly roll or Swiss roll topology (Section 8-3B). The P domain is also composed largely of an antiparallel β sheet, whereas the R domain is not visible in the viral X-ray structure so its tertiary structure is unknown. The flexible hinge region linking the P and S domains is denoted by h. [Based on an X-ray structure by Stephen Harrison, Harvard University. PDBid 2TBV.]

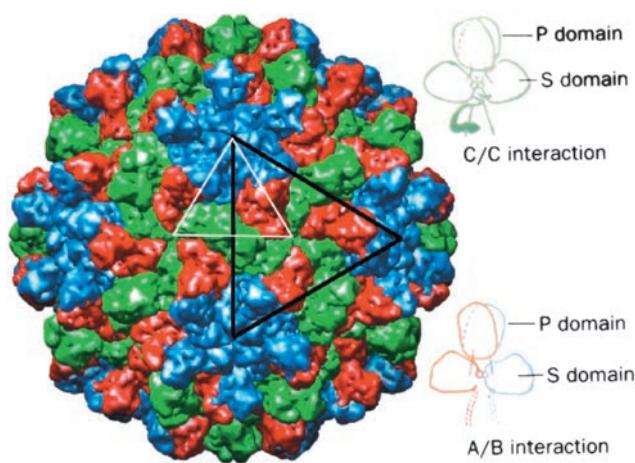


Figure 33-15 X-ray structure of TBSV showing the $T = 3$ icosadeltahedral arrangement of its coat protein subunits. The subunits occur in three quasi-equivalent packing environments, A, B, and C. The A subunits (blue) pack around exact 5-fold axes, whereas the B subunits (red) alternate with the C subunits (green) about exact 3-fold axes (local 6-fold axes). The C subunits are also disposed about the strict 2-fold axes, whereas the A and B subunits are related by local 2-fold axes. A triangular face of the $T = 3$ icosadeltahedron is outlined in white and that of the underlying icosahedron is outlined in black. The subunits respond to the different conformational requirements of their three quasi-equivalent positions through flexion at the hinge region between their S and P domains (right). Compare this drawing to Fig. 33-12. [After Harrison, S.C., *Trends Biochem. Sci.* **9**, 348, 349 (1984).]

barrel, which we shall see occurs in the coat proteins of the majority of icosahedral viruses with known structures.

a. TBSV's Identical Subunits Associate through Nonidentical Contacts

The chemically identical TBSV coat protein subunits occupy three symmetrically distinct environments denoted A, B, and C (Fig. 33-15). How does the protein accommodate the different contacts required by its several sets of analogous but nonidentical associations? TBSV's structure reveals that *analogous intersubunit contacts vary both through alternative sets of interactions and by conformational distortions of the same interactions*. Perhaps the most remarkable alternative interaction is the interdigitation of the arms connecting the R and S domains of the C subunits. These arms extend toward each icosahedral 3-fold axis (quasi-6-fold axis) in the clefts between the adjacent C and B subunits and then spiral downward about this 3-fold axis to form a β -sheetlike arrangement that resembles the overlapping flaps of a cardboard carton: chain 1 over chain 2 over chain 3 over chain 1 (Fig. 33-16a). This interaction, together with a strong association between neighboring C subunits across the icosahedral 2-fold axis (Fig. 33-15), organizes the 60 C subunits into a coherent network (Fig. 33-16b) that determines the triangulation number of the TBSV capsid: *The C subunits can be thought of as forming a $T = 1$ icosahedral shell whose gaps are filled in by the A and B subunits*. In response, the three sets of quasi-equivalent subunits assume somewhat different conformations: The three- or four-residue “hinge” connecting the S and P domains (h in Fig. 33-14) has an $\sim 30^\circ$ greater dihedral angle in the A and B subunits than in the C subunits (Fig. 33-15, right). This, in turn, permits the interactions between P domains to be identical in the AB and CC dimers (projecting dimeric knobs in Fig. 33-15). Evidently, interdomain associations between subunits are stronger in TBSV than those within subunits.

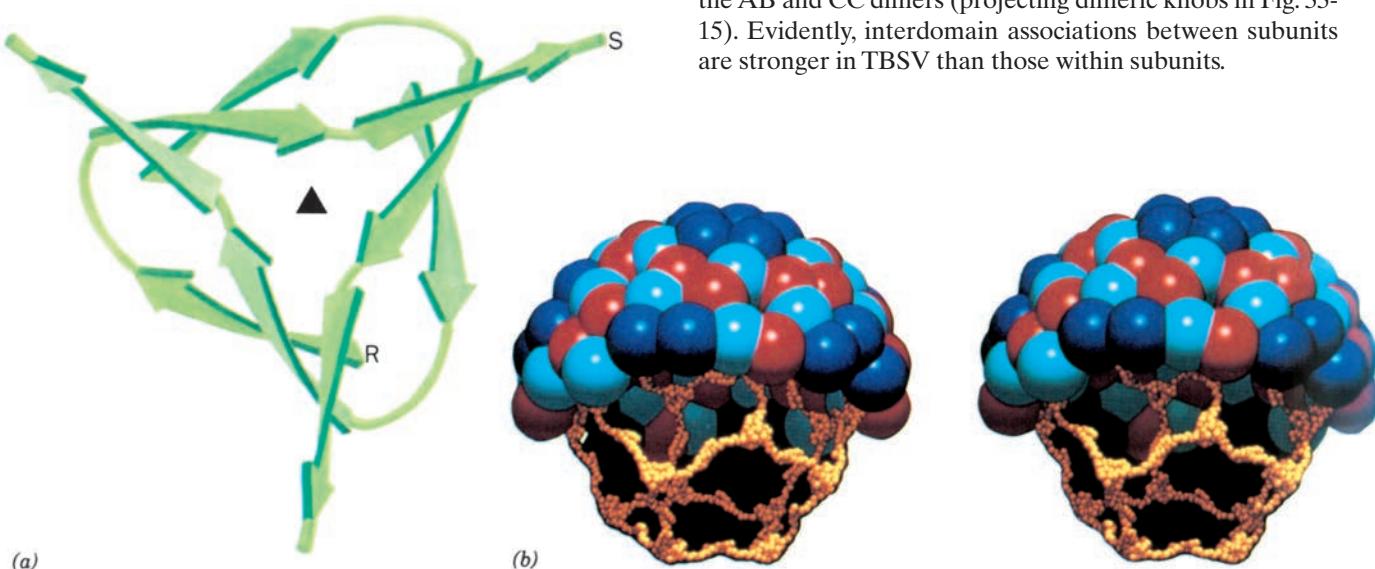


Figure 33-16 Architecture of the TBSV capsid. (a) The C subunit arms of TBSV protein pack about the capsid's exact 3-fold axes (triangle) and associate as β sheets. The view is from outside the capsid. (b) A stereo cutaway drawing showing the capsid's internal scaffolding of C subunit arms. The chemically identical A (blue), B (cyan), and C subunits (red) are represented by large spheres, whereas the residues comprising the C subunit arms are represented by the smaller yellow spheres. The C

subunit arms associate to form an icosahedral ($T = 1$) framework that apparently plays a major role in holding together the viral capsid. Directions for viewing stereo drawings are given in the Appendix to Chapter 8. [Part a after a drawing by Jane Richardson, Duke University, and Part b courtesy of Arthur Olson, The Scripps Research Institute, La Jolla, California. Based on an X-ray structure by Stephen Harrison, Harvard University. PDBid 2TBV.]

b. TBSV's RNA-Containing Core Is Disordered

The entire connecting arm between the R and S domains in the A and B subunits, as well as their first few residues in the C subunits, are not visible in TBSV's X-ray structure, thereby indicating that these polypeptide segments have no fixed conformations. The R domains are therefore flexibly tethered to the S domains so that they are also absent from the X-ray structure, even though these domains probably have a fixed conformation. Neutron scattering studies, nevertheless, suggest that protein, constituting perhaps half of the R domains, forms a 50- to 80-Å-radius inner shell. The remaining R domains are thought to project into the space between the inner and outer shells.

The viral RNA is absent from the X-ray structure, which indicates that it too is disordered. The above neutron scattering studies reveal that this RNA is sandwiched between the virus' inner and outer protein shells (Fig. 33-17). The volume constraints imposed by this arrangement require that the RNA be tightly packed. This packing is made possible because most of the negative charges of the RNA phosphate groups are neutralized by the numerous positively charged Arg and Lys residues of the R domains, the inner faces of the S domains, and their connecting arms.

c. Many Other RNA Viruses Are Remarkably Similar to TBSV

The structures of numerous other RNA plant viruses have been elucidated, including those of **southern bean mosaic virus (SBMV)** by Michael Rossmann and **satellite tobacco mosaic virus (STMV)** by Alexander McPherson.

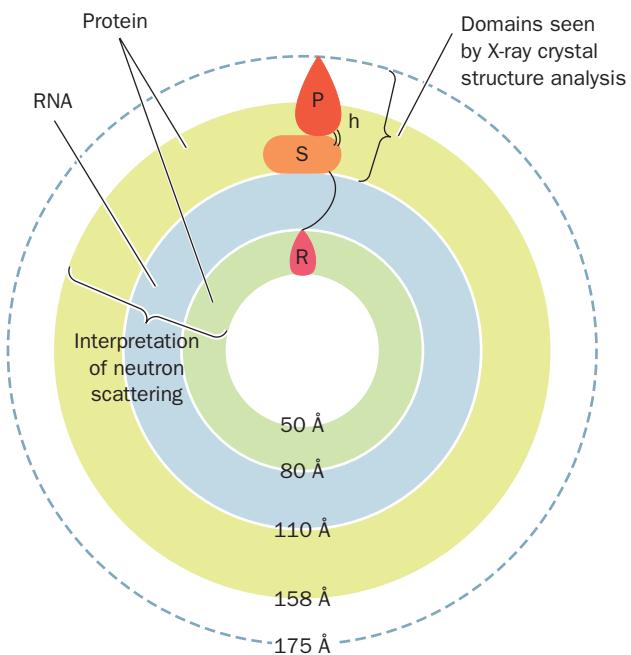


Figure 33-17 Radial organization of TBSV indicating the distribution of its protein and RNA components. The R domain positions are inferred from their known chain length. Only about half of the R domains are contained in the inner protein shell. [After Harrison, S.C., *Biophys. J.* **32**, 140 (1980).]

SBMV is a $T = 3$ virus that closely resembles TBSV in its quaternary structure. Moreover, SBMV's 260-residue coat protein subunit, although it entirely lacks a P domain, has an S domain whose polypeptide backbone is nearly superimposable on that of TBSV (Fig. 33-18a). The RNA in SBMV, as is that in TBSV, is disordered.

STMV's quaternary structure differs from those of TBSV or SBMV: It is a $T = 1$ RNA virus. This 172-Å-diameter particle, which is the smallest known virion, encloses a 1059-nt RNA that encodes only one protein, its 159-residue viral coat protein (STMV can only multiply in cells that are coinfect with the more complex TMV, one of only four known examples of a parasitic relationship between viruses). Nevertheless, STMV's coat protein, which also lacks a P domain, has an S domain that structurally resembles those of SBMV and TBSV. Evidently, these biochemically dissimilar viruses arose from a common ancestor.

d. Most of STMV's RNA Is Visible

The most striking aspect of the STMV structure is that nearly 60% of its RNA is visible (Fig. 33-19). The RNA largely takes the form of 30 double helical 10-bp segments that lie on the icosahedron's 2-fold axes and which are linked by mainly disordered single-stranded regions. Computerized searches indicate up to 68% of STMV RNA could simultaneously form base pairs. However, it seems unlikely that a unique structure with extensive and non-repetitive pairing between bases that are distant in sequence could be made to fold in a manner consistent with STMV's icosahedral symmetry. Rather, it appears that the double helical segments seen in Fig. 33-19 represent a series of somewhat different local stem-loop structures. Thus, inside the viral capsid, STMV RNA assumes a structure that is probably not its lowest free energy state. More likely, it assumes one of numerous relatively low free energy states that transiently form during viral assembly and become trapped through interactions with the viral protein coat. Indeed, the observation that STMV coat protein does not form capsids in the absence of RNA suggests that the RNA, which lacks icosahedral symmetry, nevertheless directs the formation of the icosahedral viral particle.

C. Picornaviruses

The X-ray structures of **poliovirus**, the cause of **poliomyelitis**, and **rhinovirus**, the cause of **infectious rhinitis** (the common cold), were respectively determined by James Hogle and Rossmann. Both of these human pathogens are **picornaviruses**, a large family of animal viruses that also includes the agents causing human **hepatitis A** and **foot-and-mouth disease**. Picornaviruses (*pico*, small + *rna*) are among the smallest RNA-containing animal viruses: They have a particle mass of $\sim 8.5 \times 10^6$ D of which $\sim 30\%$ is a single-stranded RNA of ~ 7500 nucleotides. Their icosahedral protein shell, which is ~ 300 Å in diameter, contains 60 protomers, each consisting of four structural proteins, **VP1**, **VP2**, **VP3**, and **VP4**. These four proteins are synthesized by an infected cell as a single polyprotein, which is cleaved to the individual subunits during virion

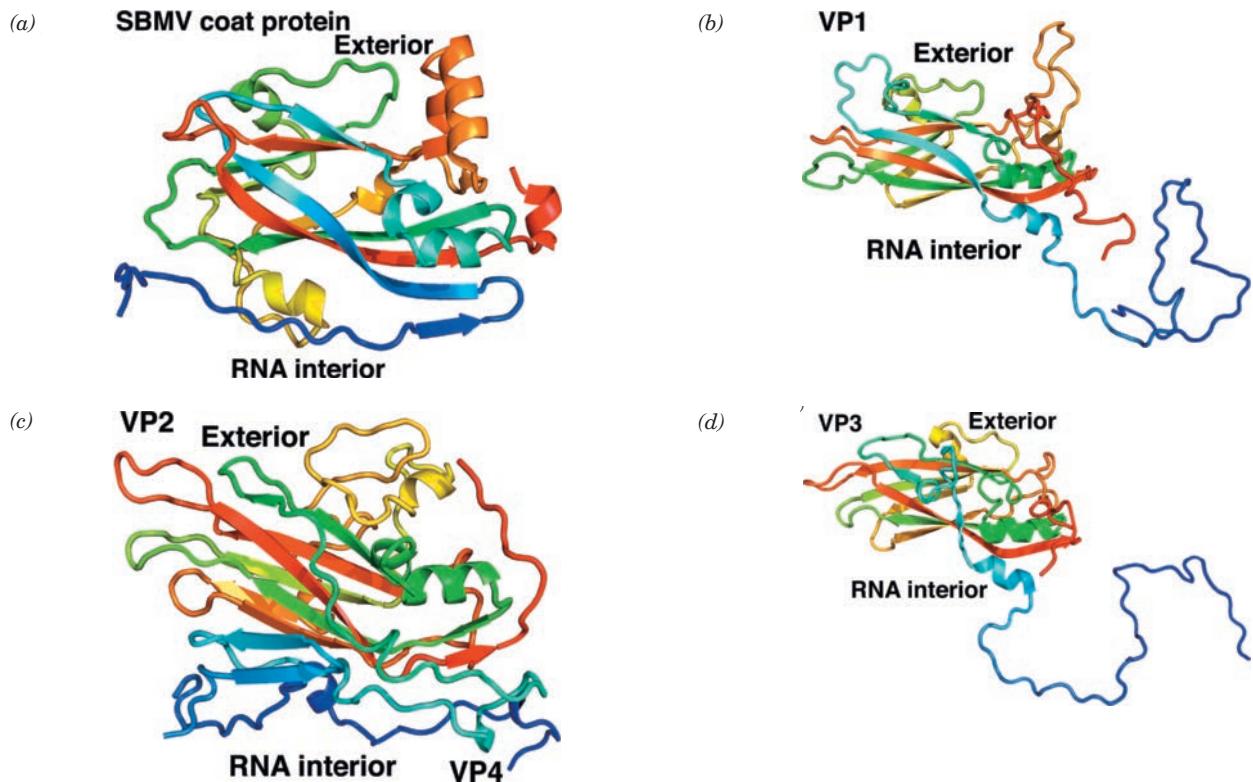


Figure 33-18 Comparison of the X-ray structures of southern bean mosaic virus (SBMV) and human rhinovirus coat proteins. (a) SBMV coat protein, and (b) VP1, (c) VP2 (together with VP4), and (d) VP3 proteins of human rhinovirus. Each polypeptide is shown in ribbon form colored in rainbow order from N-terminus (blue) to C-terminus (red), with VP2 + VP4 treated as a single

chain. Note the close structural similarities of their 8-stranded β -barrel cores and that of TBSV's S domain (Fig. 33-14). The VP1, VP2, and VP3 proteins of poliovirus also have this fold. [Based on X-ray structures by Michael Rossmann, Purdue University. PDBids 4BSV and 4RHV.]

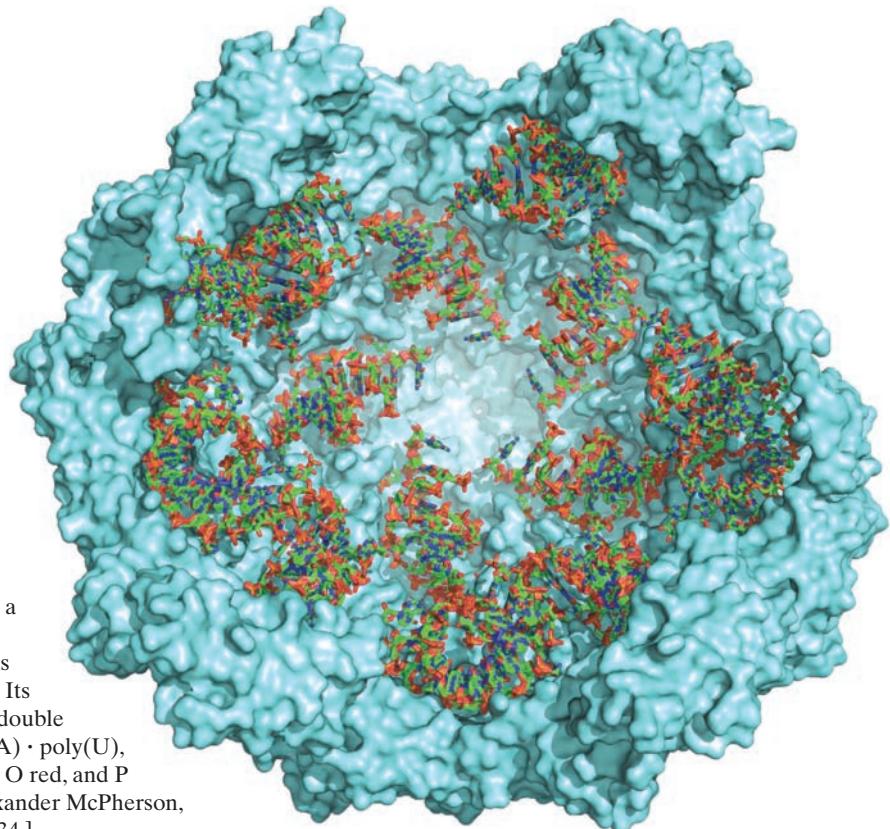


Figure 33-19 X-ray structure of satellite tobacco mosaic virus (STMV). Half of the icosahedral ($T = 1$) viral capsid, which forms a shell between the radii of 57 and 86 \AA , is represented by its molecular surface (cyan) as viewed toward its interior along a 5-fold axis. Its visible RNA, which consists mainly of 10-bp double helical segments that were modeled as poly(A) \cdot poly(U), are drawn in stick form with C green, N blue, O red, and P orange. [Based on an X-ray structure by Alexander McPherson, University of California at Irvine. PDBid 1A34.]

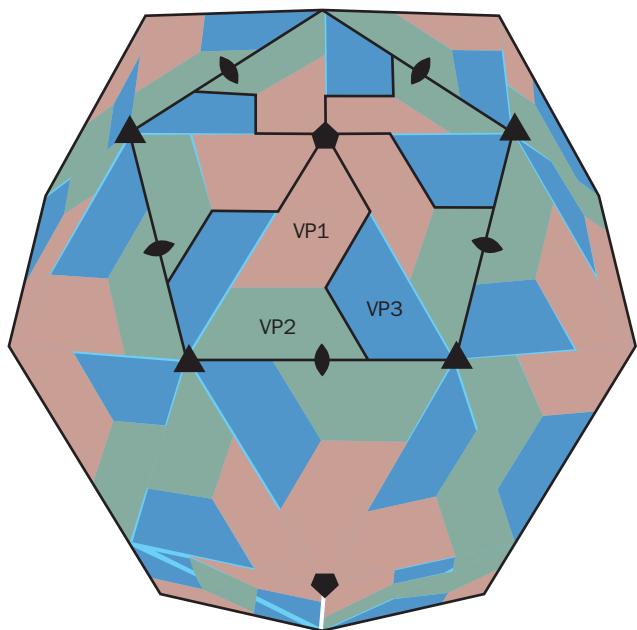


Figure 33-20 Arrangement of the 60 trimers (triangles) of pseudo-equivalent VP1, VP2, and VP3 subunits in the icosahedral capsid of human rhinovirus. This arrangement resembles that of TBSV in which 180 chemically identical subunits are quasi-symmetrically related to form a $T = 3$ icosadeltahedron (Figs. 33-12 and 33-15). The positions of the icosahedron's exact 5-, 3-, and 2-fold axes are marked. [After Rossmann, M.G., et al., *Nature* **317**, 147 (1985).]

assembly. Picornaviruses can be highly specific as to the cells they infect; for example, poliovirus binds to receptors that occur only on certain types of primate cells.

The structures of poliovirus, rhinovirus, and **foot-and-mouth disease virus (FMDV)** (determined by David Stuart) are remarkably alike, both to each other and to TBSV and SBMV. Although VP1, VP2, and VP3 of picornaviruses have no apparent sequence similarities with each other or with the coat proteins of TBSV and SBMV, these proteins all exhibit striking structural similarities (Figs. 33-14 and 33-18; VP4, which is much smaller than the other subunits, forms, in effect, an N-terminal extension of VP2). Indeed, the picornaviruses' chemically distinct VP1, VP2, and VP3 subunits are pseudosymmetrically related by pseudo-3-fold axes passing through the center of each triangular face of the icosahedron ($T = 1$) virion, which therefore has pseudo- $T = 3$ symmetry (Fig. 33-20). The chemically identical but conformationally distinct A, B, and C subunits of the $T = 3$ plant viruses are similarly quasi-symmetrically related by analogously located local 3-fold axes (Fig. 33-15). These structural similarities strongly suggest that the picornaviruses and the icosahedral plant viruses all diverged from a common ancestor.

The protein capsids of poliovirus, rhinovirus, and FMDV form a hollow shell enclosing a disordered core composed of the viral RNA and some protein, much as in the icosahedral plant viruses. This arrangement is vividly illustrated in Fig. 33-21, which shows both the inner and outer views of the poliovirus capsid. Note that VP4 largely lies inside of the capsid. Also note the rugged topography of the capsid's outer surface. Some of its crevices form the receptor-binding site through which the virus is targeted to specific cells.

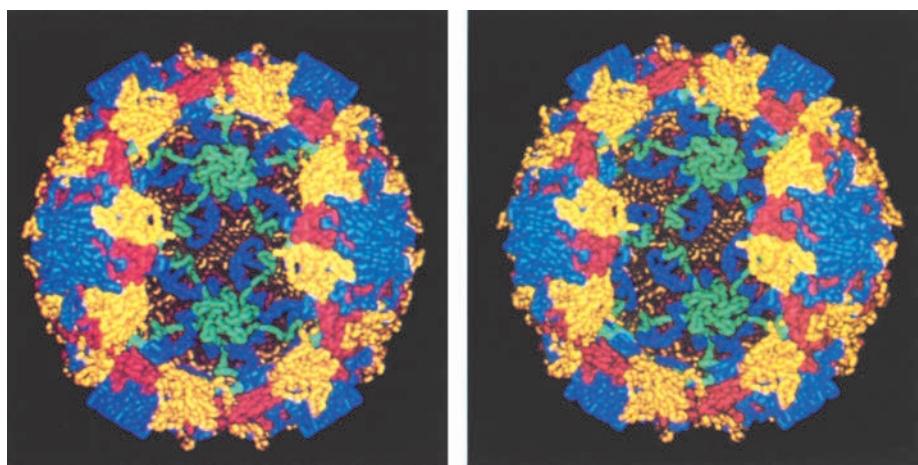


Figure 33-21 Stereo diagram of the poliovirus capsid in which its inner surface is revealed by the removal of two pentagonal faces. Here, the polypeptide chain is represented by a folded tube that approximates the volume of the protein and in which VP1 is blue, VP2 is yellow, VP3 is red, and VP4 is green. The VP4 subunits, which line the capsid's inner surface, associate about its

5-fold axes of symmetry to form a framework similar to although geometrically distinct from that formed by the C subunit arms of TBSV (Fig. 33-16). [Courtesy of Arthur Olson, The Scripps Research Institute, La Jolla, California. Based on an X-ray structure by James Hogle, Harvard Medical School. PDBid 2PLV.]

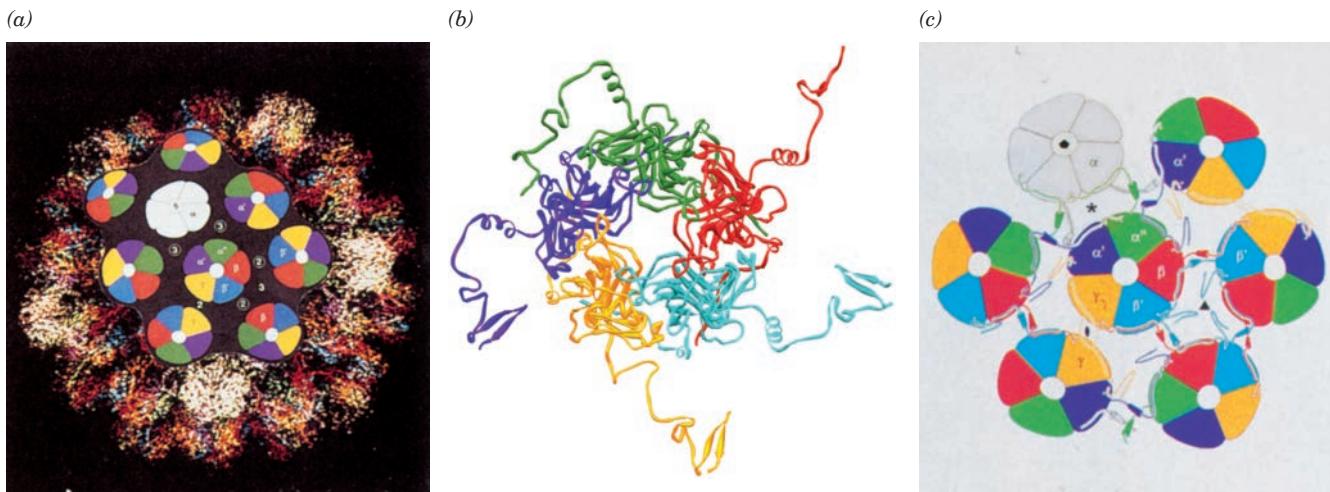


Figure 33-22 X-ray structure of simian virus 40 (SV40). (a) The SV40 virion consists of 360 copies of VP1 that are organized into 72 pentamers of which 12 (white) are 5-coordinated and 60 (colored) are 6-coordinated. Three types of interpentamer clustering are indicated on the schematic part of the drawing: The white (α), purple (α'), and green (α'') subunits form a 3-fold interaction ③; the red (β) and cyan (β') subunits form one type of 2-fold interaction ②; and the yellow subunits (γ) form a second type of 2-fold interaction (2). The icosahedral axes of symmetry are indicated by the numerals 5, 3, and 2. (b) A 6-coordinated pentamer as viewed from outside the virion. The

VP1 subunits are colored as in Part a. Note the C-terminal arms extending out from each subunit. (c) Schematic diagram showing how the C-terminal arms tie the pentamers together. The C-terminal arms are represented by lines and small cylinders (helices). The icosahedral particle's exact 5-, 3-, and 2-fold axes are represented by the conventional symbols, whereas the asterisk indicates a local 2-fold axis relating 5- and 6-coordinated pentamers. [Parts a and c courtesy of and Part b based on an X-ray structure by Stephen Harrison, Harvard University. PDBid 1SVA.]

D. Simian Virus 40 (SV40)

Simian virus 40 (SV40) is a **polyomavirus**, the simplest class of viruses containing double-stranded DNA. This \sim 500-Å external diameter icosahedral virus (Fig. 33-1g) functions to transfer a 5243-bp circular “minichromosome” (DNA in complex with histone-containing particles known as **nucleosomes**; Section 34-1B) from the nucleus of one cell to that of another. The viral capsid consists of 360 copies of a 361-residue protein, **VP1**, that are arranged with icosahedral symmetry. However, this number of particles cannot be arranged with the icosadeltahedral symmetry characteristic of TBSV, for instance, because $T = 360/60 = 6$ is a forbidden value for icosadeltahedra (for which $T = h^2 + hk + k^2$). Rather, as Caspar demonstrated through low-resolution X-ray studies of polyomaviruses, VP1 exclusively forms pentamers that take up two nonequivalent positions (Fig. 33-22a). In fact, the SV40 capsid consists of 72 VP1 pentamers that are centered on the vertices of a $T = 7$ icosadeltahedron. Twelve of these pentamers lie on the icosahedron's twelve 5-fold rotation axes, each surrounded by 5 pentamers of a different geometric class. This latter class of 60 pentamers (which in a true $T = 7$ icosadeltahedron would have to be hexamers) are each surrounded by 6 pentamers, 5 of its own class and 1 of the former class. As a consequence, each capsid contains six symmetry-inequivalent classes of the chemically identical VP1 subunits. What

conformational adjustments must the subunits make to form such a structure and, in particular, how does a pentameric structure coordinate with six other such pentamers?

The X-ray structure of SV40, determined by Harrison, indicates that VP1 consists of three modules: (1) an N-terminal arm that extends across the inside of the pentamer beneath the clockwise neighboring subunit (looking from the outside in) and whose first 15 residues are not visible in the structure (they probably extend inward to interact with the minichromosome which is likewise not visible); (2) an antiparallel β barrel with the same topology as that in RNA plant viruses and picornaviruses (Figs. 33-14 and 33-18), although oriented more or less radially with respect to the capsid rather than tangentially; and (3) a 45- to 50-residue C-terminal arm, the site of the only major conformational variation among the six symmetry-inequivalent sets of VP1 subunits. The C-terminal arms form the principal interpentamer contacts by extending from their pentamer of origin so as to invade a neighboring pentagon (Fig. 33-22b,c). Each pentamer thereby receives five invading arms from adjacent pentamers as well as donating five such arms. *It is the differing patterns of C-terminal arm exchange among the various pentamers that determines how they associate in forming the capsid.* Since these C-terminal arms are probably flexible and unstructured on a free

pentamer, the capsid's pentameric building blocks probably behave, so to speak, as if they are tied together with ropes rather than being cemented together across extended complementary surfaces. Indeed, deletion of the C-terminal arms from recombinant VP1 subunits does not prevent their associating into pentamers but precludes these pentamers from assembling into the viruslike shells that they would otherwise form.

There are many other viruses of known structure whose capsid proteins contain the foregoing 8-stranded antiparallel β barrel. These include the $T = 1$ bacteriophage ϕ X174 (a single-stranded DNA virus; Section 30-3B; Fig. 33-1d), the $T = 3$ **Norwalk virus** (which is responsible for >96% of nonbacterial gastroenteritis in the United States) and **black beetle virus** (both single-stranded RNA viruses), the $T = 4$ **Nudaurelia Capensis ω virus** (a single-stranded RNA insect virus), and the $T = 13$ **bluetongue virus** (Section 33-2F). Information concerning all icosahedral viruses of known structure is available from the *Virus Particle ExploreR* database (VIPERdb) at <http://viperdb.scripps.edu/>.

E. Bacteriophage MS2

The RNA bacteriophage MS2 infects only F^+ (male) *E. coli* (Section 31-1Ac) because infection is initiated by viral attachment to bacterial F pili. The 275- \AA -diameter MS2 virion (Fig. 33-1b) consists of 180 identical 129-residue coat protein subunits arranged with $T = 3$ icosadeltahedral symmetry encapsidating a 3569-nt single-stranded RNA molecule. The virion also contains a single copy of the 44-kD A-protein, which is thought to be responsible for viral attachment to the F pili and must therefore be exposed on the phage surface.

The X-ray structure of MS2, determined by Karin Valegård and Lars Liljas, reveals that it has a $T = 3$ protein shell formed by 60 icosahedrally related triangular protomers, each of which consists of three chemically identical subunits with slightly different conformations, much as in TBSV (Fig. 33-15). However, the MS2 coat protein does not contain the 8-stranded antiparallel β barrel present in all the previously discussed icosahedral viruses. Rather, each subunit consists of a 5-stranded antiparallel β sheet facing the interior of the particle overlaid with a short β hairpin and two α helices facing the viral exterior (Fig. 33-23). This protein fold resembles those of several other bacteriophage coat proteins.

F. Bluetongue Virus

Bluetongue virus is a member of the **orbivirus** genus within the **Reoviridae** family, one of the largest families of viruses. Members of the *Reoviridae* are responsible for significant levels of child mortality in developing countries (by the **rotaviruses**, which cause diarrhea) as well as a variety of economically important diseases of both plants and animals. The orbiviruses are icosahedral viruses whose capsids are made of two shells: an outer shell consisting of the viral proteins **VP2** and **VP5**, which is lost on cell entry; and a

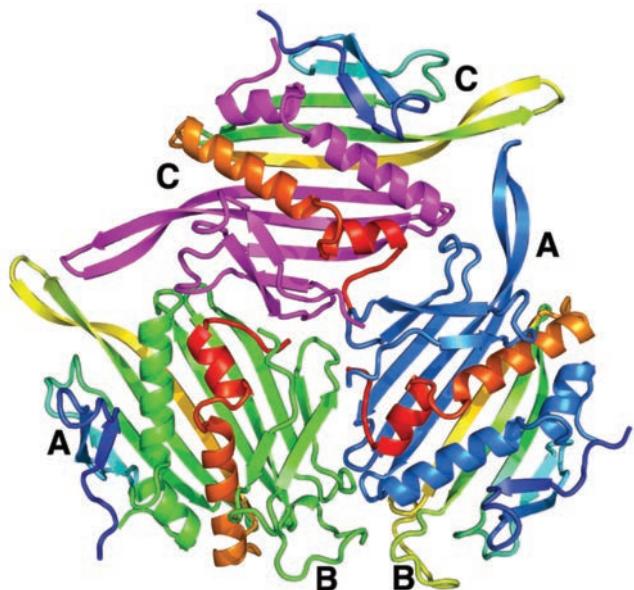
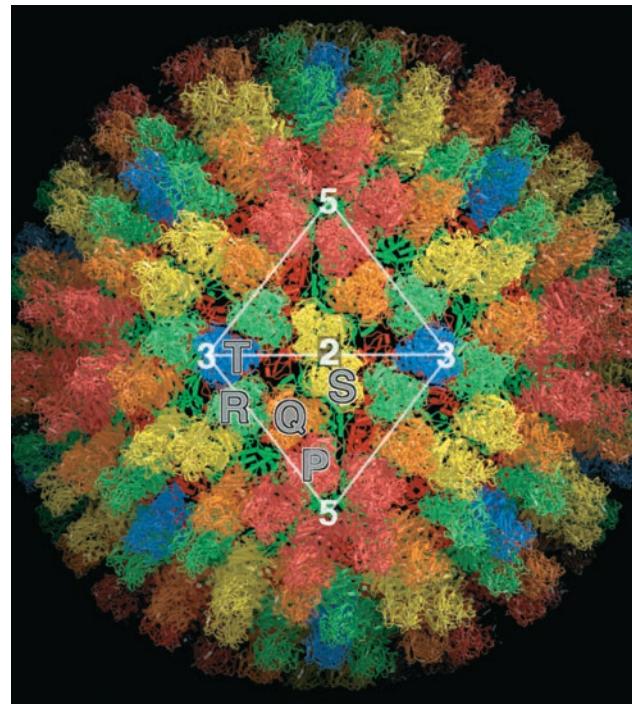


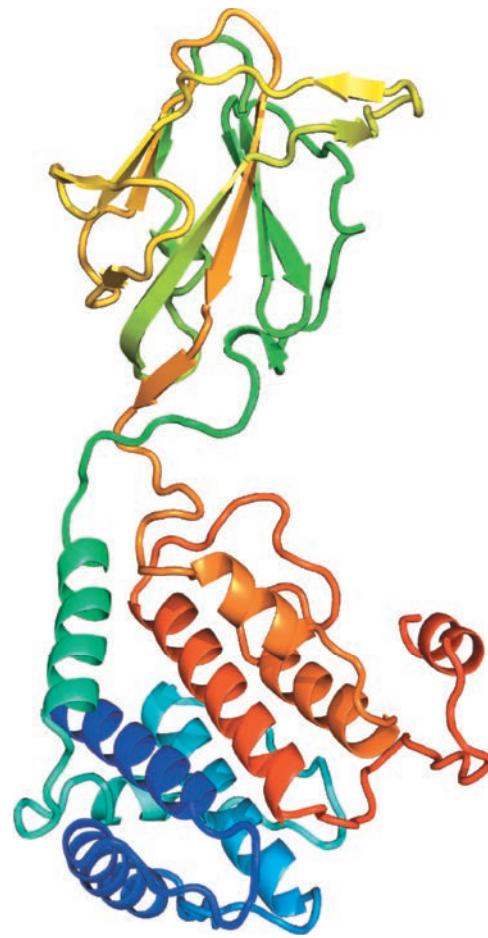
Figure 33-23 X-ray structure of bacteriophage MS2 showing three dimers related by a quasi-3-fold axis of the $T = 3$ icosadeltahedral particle. The view is along the quasi-3-fold axis from outside the capsid. The A, B, and C subunits are as defined in Fig. 33-12b with the inner subunits in this assembly colored blue, green, and magenta, respectively (those in Fig. 33-12b are differently colored) and the outer subunits colored in rainbow order from their N-termini (blue) to their C-termini (red). The two C monomers shown are related by the particle's exact 2-fold axis, whereas closely associated A and B monomers are related by quasi-2-fold axes. In all cases, each monomer's 5-stranded antiparallel β sheet is extended across the 2-fold axis and its helices interlock with those of its dimeric mate. Note the lack of structural resemblance between the MS2 subunits and the eight-stranded antiparallel β barrels that form the coat proteins of most other icosahedral viruses of known structure (e.g., Figs. 33-14 and 33-18). [Based on an X-ray structure by Karin Valegård and Lars Liljas, Uppsala University, Sweden. PDBid 2MS2.]

transcriptionally active core that is released into the cytoplasm. The core consists of two layers, an inner $T = 2$ shell constructed from 120 copies of the 100-kD protein **VP3(T2)**, and an outer $T = 13$ shell consisting of 780 copies of the 38-kD protein **VP7(T13)**. The capsid contains an aggregate of ~20 kb of double-stranded RNA (dsRNA) in usually 10 different segments, nearly all of which encode only one protein each. During infection, the dsRNA is maintained within the core because its release into the cell would trigger the interferon-mediated shutdown of translation (Section 32-4Ab), which would prevent viral proliferation. Consequently, the core also contains multiple copies of virally encoded **dsRNA-dependent RNA polymerase [VP1(Pol)]**, helicase **[VP6(Hel)]**, and capping enzyme **[VP4(Cap)]**, which are associated with each dsRNA segment so as to form active transcription complexes. The resulting mRNAs are extruded into the host cell's



(a)

Figure 33-24 X-ray structure of bluetongue virus core. (a) Its $T = 13$ outer shell. The triangular icosahedral asymmetric unit, whose edges (white lines) link the icosahedron's symmetry axes, contains 13 copies of VP7 arranged as five trimers, P, Q, R, S, and T, which are colored red, orange, green, yellow, and blue, respectively. Trimer T sits on an icosahedral 3-fold axis and hence contributes a monomer to the asymmetric unit. (b) The structure



(b)

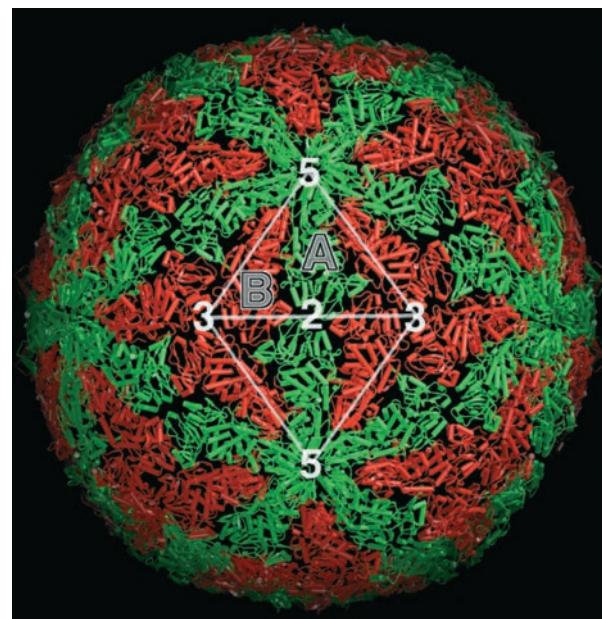
of VP7, which is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The 8-stranded antiparallel β barrel domain (above) forms the viral core's outer projections and the helical domain (below) forms its outer shell. [Part a courtesy of and Part b based on an X-ray structure by David Stuart, Oxford University, U.K. PDBid 2BTV.]

cytoplasm, where they direct the ribosomal synthesis of viral proteins. In addition, the mRNAs are encapsidated within growing cores, where they act as templates for the synthesis of negative strand RNA segments, thus forming progeny dsRNAs. Nevertheless, how each core is packaged with precisely one copy of each dsRNA segment is unknown.

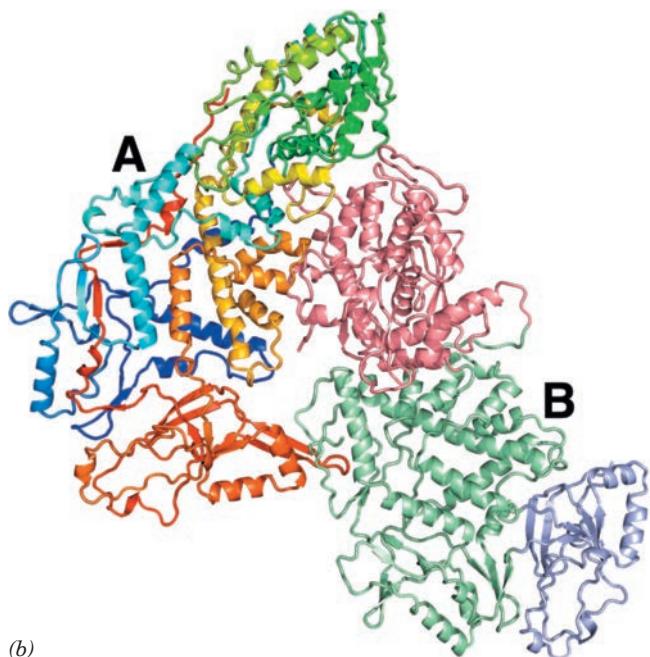
Bluetongue virus (**BTV**) infects ungulates (hoofed mammals; e.g., sheep) and is transmitted by certain blood-feeding insects. It is named for the cyanotic tongues (due to swelling) that many BTV-infected animals have. The X-ray structure of the BTV core, determined by Stuart, reveals both layers of the 700- \AA -diameter, $\sim 55,000$ -kD particle, and hence BTV is the largest particle of known X-ray structure (although the ribosome constitutes the largest asymmetric assembly of known X-ray structure; Section 32-3A). The asymmetric portion of the outer shell consists

of 13 independent copies of VP7(T13) arranged as five geometrically distinct trimers, P, Q, R, S, and T (Fig. 33-24a). The 349-residue VP7(T13) consists of two domains (Fig. 33-24b): an 8-stranded antiparallel β barrel common to many icosahedral viruses that forms the core's outer projections (and is responsible for its hedgehoglike appearance) and a helical domain that forms the core's outer shell. Note that in most icosahedral viruses that we have encountered, the β barrel domain forms the viral shell (Section 33-2B). The various geometrically distinct copies of VP7(T13) have nearly identical conformations, with maximum deviations between pairs of equivalent C_α positions of only 0.3 \AA , even though there are significant differences in the way neighboring subunits contact one another.

BTV's inner shell consists of 60 asymmetric homodimers of VP3(T2) subunits, A and B, arranged with



(a)



(b)

Figure 33-25 X-ray structure of bluetongue virus core. (a) Its $T = 2$ inner shell. It is constructed from homodimers of VP3 subunits arranged with icosahedral ($T = 1$) symmetry so that the two subunits forming the homodimer, A (green) and B (red), are symmetrically inequivalent. The triangular icosahedral asymmetric unit is indicated (white lines). Compare this structure with that of the $T = 13$ outer shell (Fig. 33-24a). (b) The structure of the VP3 asymmetric dimer. Its A subunit is colored in rainbow

order from its N-terminus (blue) to its C-terminus (red) and its B subunit is colored according to domain with its apical domain pink, its carapace domain pale green, and its dimerization (across the 2-fold axis) domain light purple. Note the somewhat different conformations and the very different structural environments of these two chemically identical subunits. [Part a courtesy of and Part b based on an X-ray structure by David Stuart, Oxford University, U.K. PDBid 2BTV.]

icosahedral ($T = 1$) symmetry (Fig. 33-25a). The 901-residue VP3(T2) consists of three domains (Fig. 33-25b): an apical domain, which contains 11 helices and 10 β strands in A but 10 helices and 11 β strands in B; a carapace domain, which contains 20 helices in A and 21 in B; and a dimerization domain, which contains 5 helices and 13 β strands in A and 4 helices and 14 β strands in B. The inner shell is relatively smooth and has few charged residues. It has 9-Å-wide pores at its icosahedral 5-fold vertices. These pores, which are lined with conserved Arg residues, are too narrow to permit the exit of mRNA, but in the presence of Mg^{2+} they open sufficiently to do so (see below).

The BTV core's outer and inner shells interact through relatively flat, predominantly hydrophobic surfaces. The symmetry mismatch between these two shells requires that they have 13 different sets of contacts, which makes the closely similar conformations of the 13 geometrically distinct VP7(T13) subunits all the more remarkable. VP3(T2) self-assembles to form a subcore but VP7(T13), although it forms trimers in solution, does not self-assemble to form an icosahedral shell. It is therefore likely that the inner shell forms a permanent scaffolding on which 260 VP7(T13) trimers crystallize in two dimensions to form the outer shell.

The X-ray structure of BTV also reveals the paths of ~80% of its 19,219-bp dsRNA (Fig. 33-26a). The dsRNA appears to be partially ordered about the icosahedral 5-fold axes through its interactions with the inside of the VP3(T2) shell (although note that the dsRNA, whose 10 segments

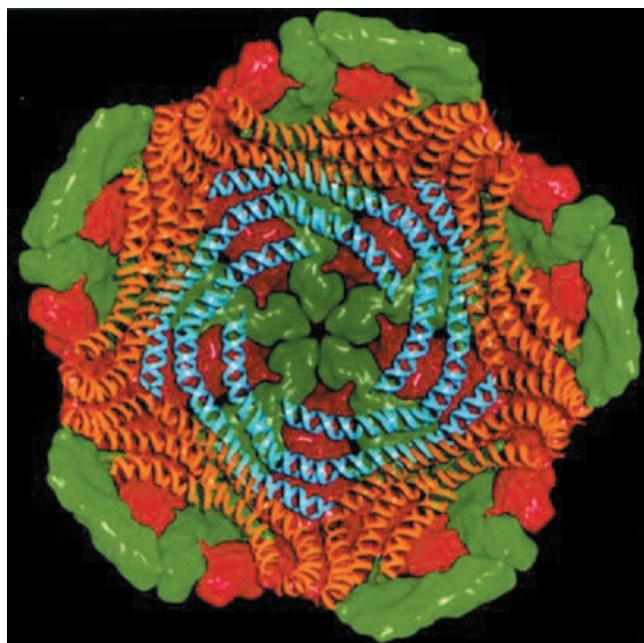
range in size from 822 to 3954 bp, cannot be arranged with true icosahedral symmetry). This suggests that each transcription complex is organized about a 5-fold axis near the inner surface of the core with its associated dsRNA segment spiraling toward the center of the core. Indeed, X-ray structures of BTV crystals that had been soaked in solutions containing 20-nt oligonucleotides reveal electron density emanating from the viral core along its 5-fold axes that presumably mimics newly synthesized mRNAs (Fig. 33-26b).

G. *Paramecium bursaria Chlorella Virus*

Viruses of the **chlorovirus** genus are among the largest and most complex known icosahedral viruses. These viruses have a layered structure comprising dsDNA surrounded by a protein core, a lipid membrane, and finally an icosahedral protein shell. **Paramecium bursaria Chlorella virus type 1 (PBCV-1)** infects certain *Chlorella*-like algae. It attaches to its host cell and, through the mediation of viral enzymes to digest the host cell wall around the point of attachment, injects its dsDNA into the cell, leaving the empty capsid on the cell surface. PBCV-1, which has a molecular mass of $\sim 10^9$ D, has a 331-kb genome that encodes 377 proteins and 10 tRNAs. Its major capsid protein, **Vp54** (a 437-residue glycoprotein), accounts for ~40% of the virion's protein mass.

The cryo-EM-based image of PBCV-1 (Fig. 33-27a), determined by Timothy Baker and Rossmann, reveals that its outer shell forms a 1900-Å-diameter, $T = 169$

(a)



(b)

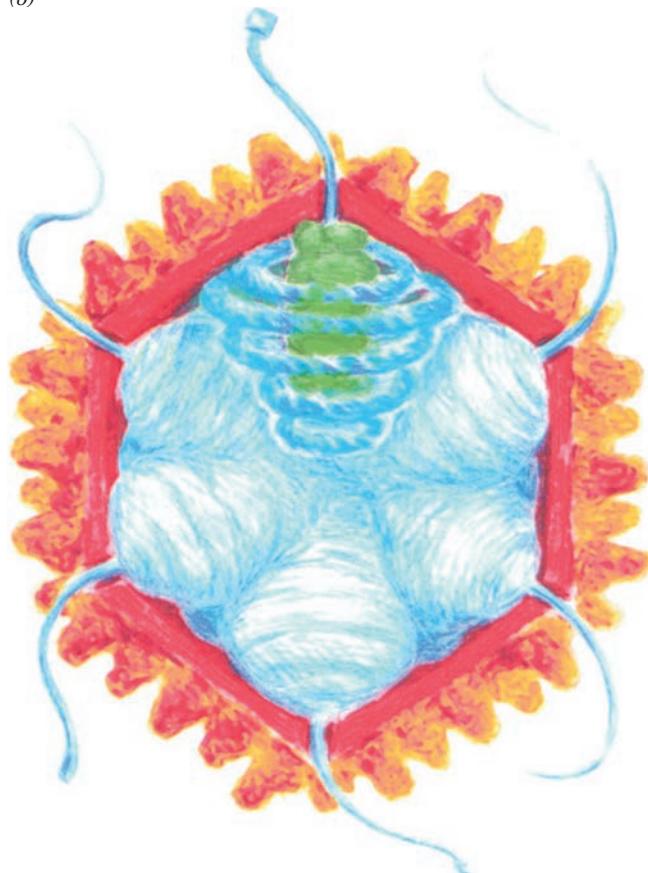
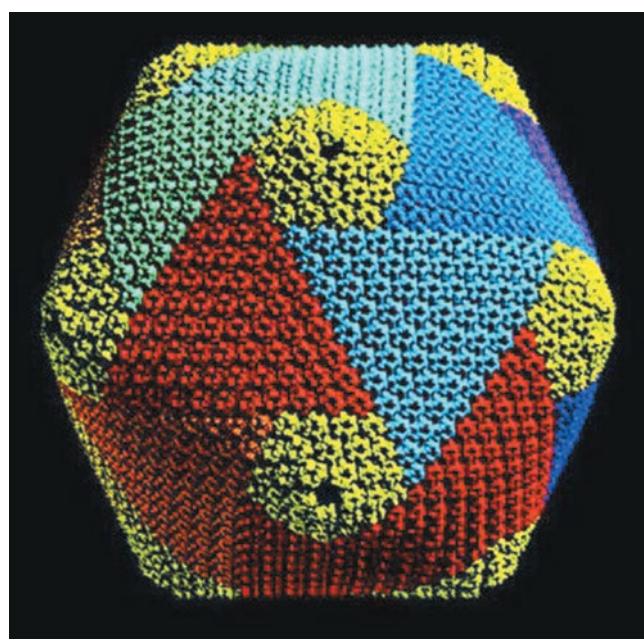


Figure 33-26 Arrangement of RNA in BTV. (a) The X-ray structure-based packing of dsRNA in the BTV core's inner shell as viewed along an icosahedral 5-fold axis and showing only the core's lower hemisphere. The relatively poorly resolved electron density has been modeled as A-form RNA with the RNA that is packed about the centrally located 5-fold axis blue and that packed about other 5-fold axes orange. The A and B subunits of the VP3(T2) forming the core's inner shell are green and red. (b) A cartoon model for the arrangement of RNA (blue) in the BTV core. The dsRNA is drawn as a coil that is wrapped about its associated transcription complex (green). Newly synthesized mRNA tails are

shown exiting the pores at the core's 5-fold vertices. Note that each transcription complex is associated with one 5-fold vertex.
[Courtesy of David Stuart, Oxford University, U.K.]

(a)



(b)

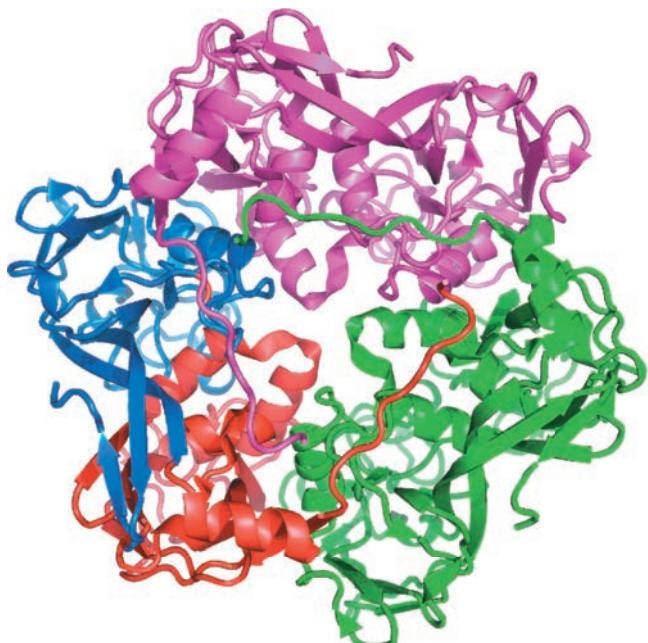


Figure 33-27 Structure of the PBCV-1 capsid. (a) A quasi-atomic model based on fitting the X-ray structure of Vp54 to the cryo-EM-based image of the capsid. The pentasymmetrons are yellow and the trisymmetrons are variously colored. (b) The X-ray structure of the Vp54 homotrimer as viewed along its

3-fold axis. The N-terminal β barrel domain of the leftmost subunit is red, its C-terminal β barrel is blue, and the remaining monomers are green and magenta. [Part a courtesy of and Part b based on an X-ray structure by Michael Rossmann, Purdue University. PDBid 1M4X.]

icosadeltahedron ($h = 7$ and $k = 8$). This enormous capsid is constructed from 20 triangular units named **trisymmetrons** and 12 pentagonal caps named **pentasymmetrons**, which respectively consist of pseudohexagonal arrays of 66 and 30 trimers of Vp54 for a total of 1680 trimers and hence 5040 monomers of Vp54 in the capsid (see below). The trisymmetrons do not correspond to the capsid's icosahedral faces. Instead, they bend around the edges of the icosahedron, leaving openings at its 5-fold vertices that are filled by the pentasymmetrons. Each pentasymmetron also contains a pentamer of a different protein at its 5-fold vertex.

The X-ray structure of Vp54 (Fig. 33-27b) reveals that it forms cyclic trimers in which each monomer consists of two consecutive antiparallel β barrels similar to those in other icosahedral viruses. The two β barrels in the Vp54 monomer are related by a 53° rotation about the trimer's 3-fold axis and hence the trimer has pseudohexagonal symmetry. These trimers have been fitted to the cryo-EM-based image of the PBCV-1 capsid, thereby yielding its quasi-atomic model (Fig. 33-27a).

3 BACTERIOPHAGE λ

Bacteriophage λ (Figs. 33-1f and 33-28), a midsize (58 million D) **coliphage** (bacteriophage that infects *E. coli*), has a 55-nm-diameter, $T = 7$, icosadeltahedral head and a flexible 15- to 135-nm-long tail that bears a single thin fiber at its end. The virion contains a 48,502-bp linear double-stranded B-DNA molecule of known sequence. Phage λ is among the most extensively characterized complex viruses with respect to its molecular biology. Indeed, as we shall see in this section, *its genetic regulatory mechanisms form one of our best paradigms for the control of development in higher organisms and its assembly is among our most well-characterized examples of the morphogenesis of biological structures*.

Bacteriophage λ adsorbs to *E. coli* through a specific interaction between the viral tail fiber and **maltoporin** (Section 20-2D; the product of the *E. coli lamB* gene), which is a component of the bacterium's outer membrane. This interaction initiates a complex and poorly understood process in which the phage DNA is injected through the viral tail into the host cell. Soon after entering the host, the λ DNA, which has complementary single-stranded ends of 12 nucleotides (cohesive ends), circularizes and is covalently closed and supertwisted by the host DNA ligase and DNA gyrase (Fig. 33-29, Stages 1 and 2).

At this stage the virus has a “choice” of two alternative life styles (Fig. 33-29):

1. It can follow the familiar **lytic** mode in which the phage is replicated by the host such that, after 45 min at 37°C , the host lyses to release ~ 100 progeny phages.

2. *The phage may take up the so-called lysogenic life cycle, in which its DNA is inserted at a specific site in the host chromosome such that the phage DNA passively replicates*

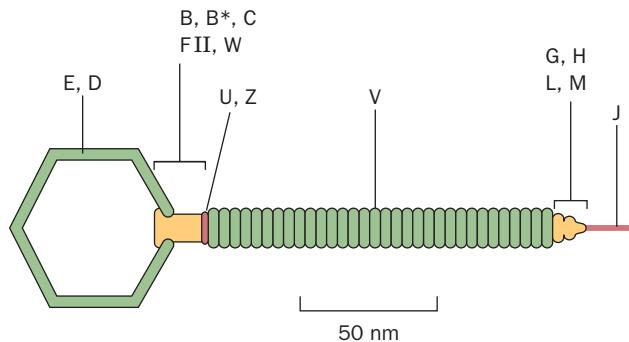


Figure 33-28 Sketch of bacteriophage λ indicating the locations of its protein components. The letters refer to specific proteins (gene products; see text). The bar represents 50 nm. [After Eiserling, F.A., in Fraenkel-Conrat, H. and Wagner, R.R. (Eds.), *Comparative Virology*, Vol. 13, p. 550, Plenum (1979).]

with the host DNA. Nevertheless, even after many bacterial generations, if conditions warrant, the phage DNA will be excised from the host DNA to initiate a lytic cycle in a process known as **induction**.

How the phage chooses between the lytic and lysogenic modes is the subject of Section 33-3D.

Phage DNA that is following a lysogenic life cycle is described as a **prophage**, whereas its host is called a **lysogen**. An intriguing property of lysogens is that they cannot be reinfected by phages of the type with which they are lysogenized: *They are immune to superinfection*. A bacteriophage that can follow either a lytic or a lysogenic life style is known as a **temperate phage**, whereas those that have only a lytic mode are said to be **virulent**. Bacteriophages that are reproducing lytically are said to be engaged in **vegetative growth**.

Over 90% of the thousands of known types of phages are temperate and, conversely, most bacteria in nature are lysogens. Yet, the presence of prophages has frequently gone unnoticed because they have little apparent affect on their hosts. For example, the K12 strain of *E. coli* had been the subject of intensive investigations for >20 years before 1951 when Esther Lederberg found it to be lysogenic for bacteriophage λ (which marks the discovery of this phage as well as the phenomenon of lysogeny).

The advantage of lysogeny is clear. A parasite that can form a stable association with its host has a better chance of long-term survival than one that invariably destroys its host. A virulent phage, on encountering a colony of its host bacteria, will multiply prodigiously. After the colony has been wiped out, however, it may be some time, if at all, before any of the progeny encounter another suitable host in a generally hostile world. In contrast, a prophage will multiply with its host indefinitely so long as the host remains viable.

But what if the host is fatally injured? Does the parasite die with the host? In the case of bacteriophage λ , it is

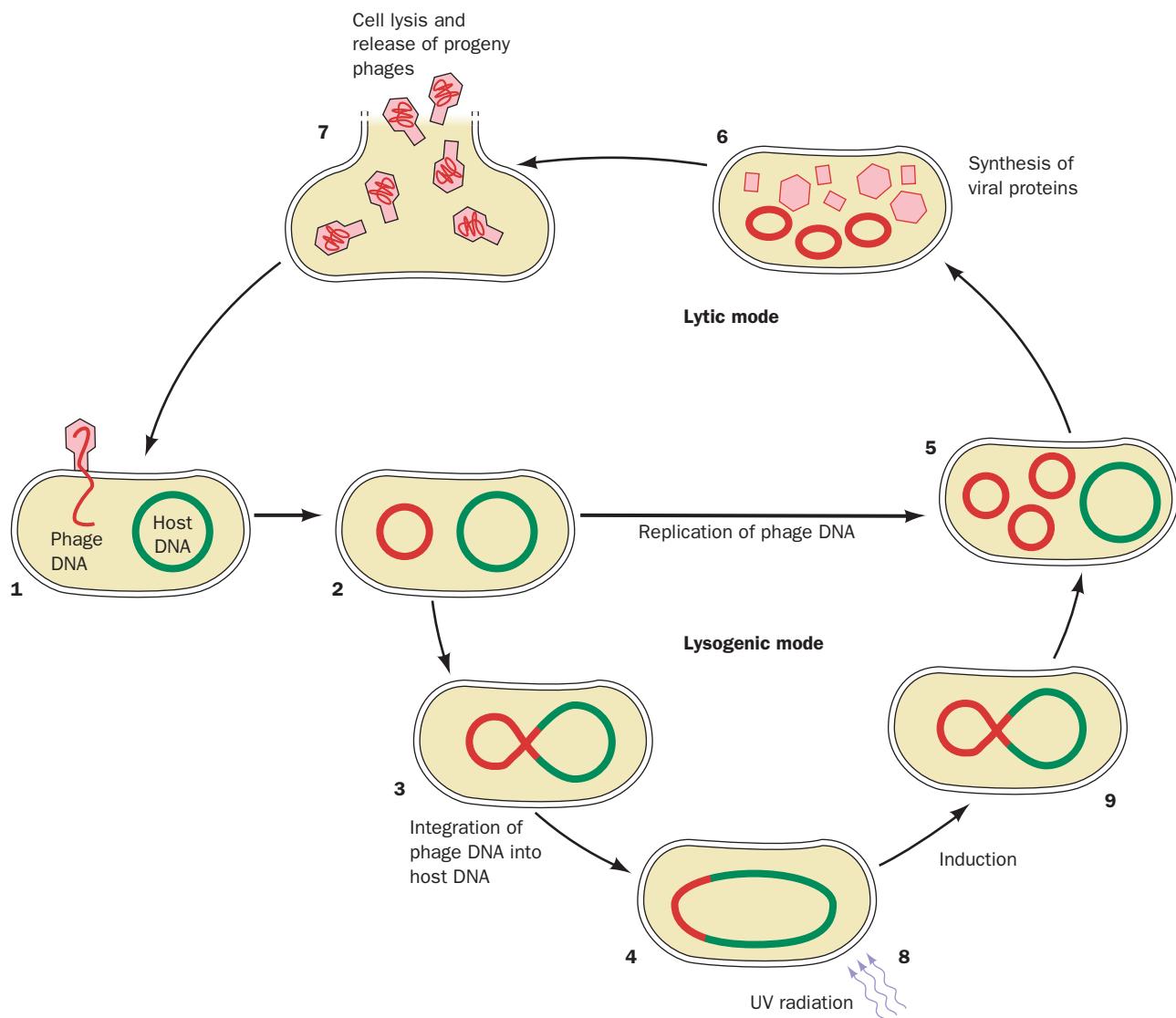


Figure 33-29 The λ phage life cycle. The infection of the bacterial host *E. coli* begins when the virus specifically adsorbs to the cell and injects its DNA (1). The linear DNA then circularizes (2) and commences directing the infection process. In the lysogenic mode, the phage DNA is stably integrated at a specific site in the host chromosome (3 and 4) so that it is passively replicated with the bacterial cell. Alternatively, the

phage may take up the lytic mode in which the DNA directs its own replication (5), as well as the synthesis of viral proteins (6), so as to result in the lysis of the host cell with the release of ~ 100 progeny phages (7). DNA damage, as is caused, for example, by UV radiation (8), induces the excision of the prophage DNA from the lysogenic bacterial chromosome (9) and causes the phage to take up the lytic mode.

precisely such traumatic conditions, exposure to agents that damage the host DNA or disrupt its replication, that induce the lytic phase. This has been described as the “lifeboat” response: The prophage escapes a doomed host through the formation of infectious viral particles that have at least some chance of further replication. Conversely, lysogeny is triggered by poor nutritional conditions for the host (phages can lytically replicate only in an actively growing host) or a large number of phages infecting

each host cell (which signals that the phages are on the verge of eliminating the host).

This section describes the genetic system that controls the orderly formation of phage particles in the lytic mode, the mechanism through which these phage particles are assembled, and the regulatory mechanism through which bacteriophage λ selects and maintains its life cycle. *Analogous systems are believed to underlie many cellular processes.*

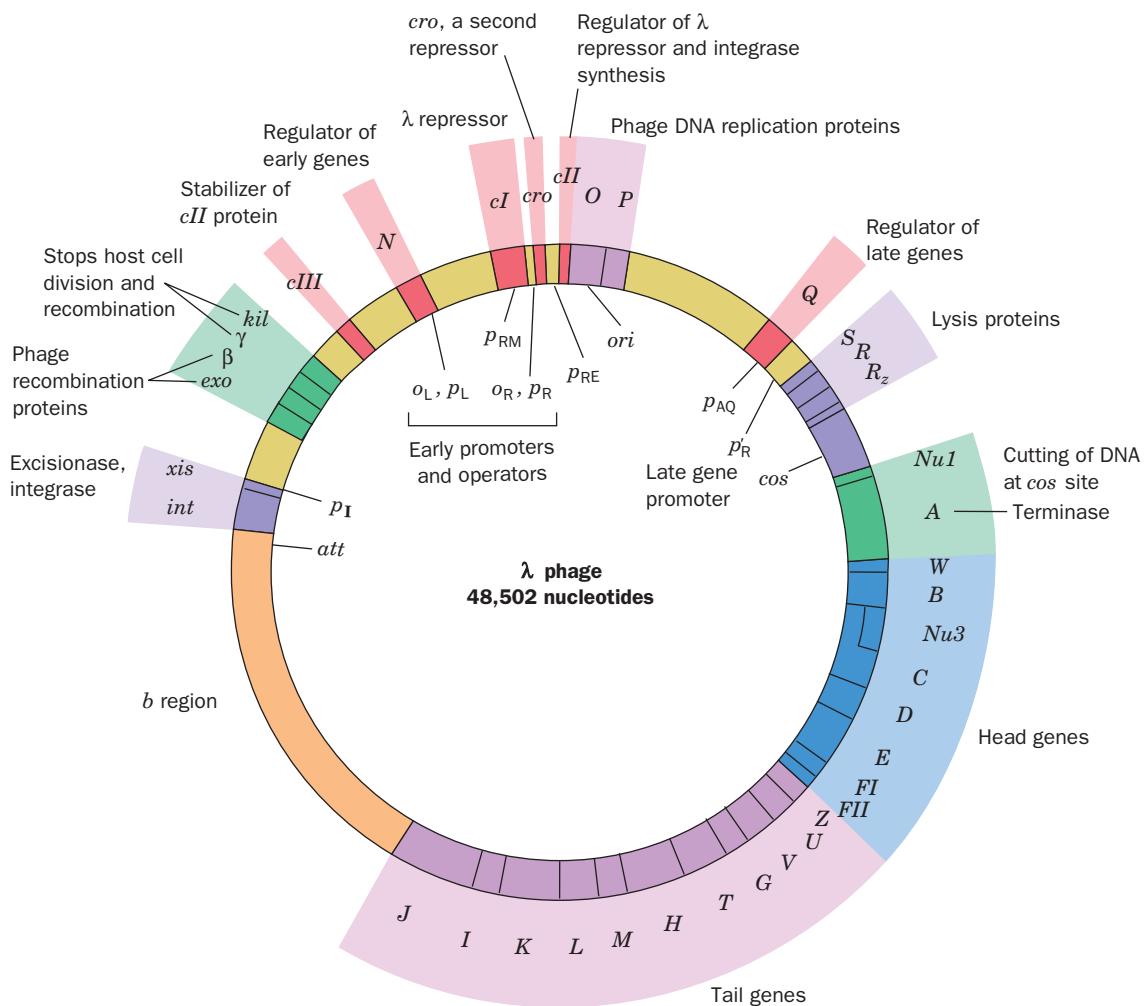


Figure 33-30 A genetic map of bacteriophage λ . Most of its structural genes (indicated outside the circle) and control sites (indicated inside the circle) are shown, with the genes encoding

regulatory proteins shaded in red. On packaging into the virion, the circular chromosome is cut at the *cos* site yielding a linear DNA.

A. The Lytic Pathway

The bacteriophage λ genome, as its genetic map indicates (Fig. 33-30), encodes ~ 50 gene products and contains numerous control sites. Note the λ chromosome's organization. Its genes are clustered according to function. For example, the genes concerned with the synthesis of phage tail proteins are tandemly arranged on the bottom of Fig. 33-30. This organization, as we shall see, enables these genes to be transcribed together, that is, as an operon. The functions of many of the λ genes and control sites, together with those of the host that are important in phage function, are tabulated in Table 33-1.

In the lytic replication of phage λ , as in love and war, proper timing is essential. This is because the DNA must be replicated in sufficient quantity before it is made unavailable by packaging into phage particles and because packaging must be completed before the host cell is enzymati-

cally lysed. The transcription of the λ genome, which is carried out by host RNA polymerase, is controlled in both the lytic and the lysogenic programs by the regulatory genes that are shaded in red in Fig. 33-30.

a. The Lytic Mode Has Early, Delayed Early, and Late Phases

The lytic transcriptional program has three phases (Fig. 33-31):

1. Early transcription Soon after phage infection or induction, *E. coli* RNA polymerase commences "leftward" transcription of the phage DNA starting at the promoter p_L and "rightward" transcription (and thus from the opposite DNA strand) from the promoters p_R and p'_R (Fig. 33-31a):

(i) The "leftward" transcript, L1, which terminates at termination site t_{L1} , encodes the *N* gene.

Table 33-1 Important Genes and Genetic Sites for Bacteriophage λ

Gene or Site	Function
Phage genes	
<i>cI</i>	λ repressor; establishment and maintenance of lysogeny
<i>cII, cIII</i>	Establishment of lysogeny
<i>cro</i>	Repressor of <i>cI</i> and early genes
<i>N, Q</i>	Antiterminators for early and delayed early genes
<i>O, P</i>	Origin recognition in DNA replication
γ	Inhibits host RecBCD
<i>int</i>	Prophage integration and excision
<i>xis</i>	Prophage excision
<i>B, C, D, E, W, Nu3, FI, FII</i>	Head assembly
<i>G, H, I, J, K, L, M, U, V, Z</i>	Tail assembly
<i>A, Nu1</i>	DNA packaging
<i>R, R_z, S</i>	Host lysis
<i>b</i>	Accessory gene region
Phage sites	
<i>attP</i>	Attachment site for prophage integration
<i>attL, attR</i>	Prophage excision sites
<i>cos</i>	Cohesive end sites in linear duplex DNA
<i>o_L, o_R</i>	Operators
<i>p_I, p_L, p_R, p_{RM}, p_{RE}, p'_R</i>	Promoters
<i>t_{L1}, t_{R1}, t_{R2}, t_{R3}, t'_R</i>	Transcriptional termination sites
<i>nutL, nutR</i>	<i>N</i> utilization sites
<i>qut</i>	<i>Q</i> utilization site
<i>ori</i>	DNA replication origin
Host genes^a	
<i>lamB</i>	Host recognition protein
<i>dnaA, dnaB</i>	DNA replication initiation
<i>lig</i>	DNA ligase
<i>gyrA, gyrB</i>	DNA gyrase
<i>rpoA, rpoB, rpoC</i>	RNA polymerase core enzyme
<i>rho</i>	Transcription termination factor
<i>nusA, nusB, nusE</i>	Necessary for <i>gpN</i> function
<i>groEL, groES</i>	Head assembly
<i>himA, himD</i>	Integration host factor
<i>hflA, hflB</i>	Degrades <i>gpcII</i>
<i>cap, cya</i>	Catabolite repressor system
<i>attB</i>	Prophage integration site
<i>recA</i>	Induction of lytic growth

^aThe genes encoding DNA polymerase I and the subunits of DNA polymerase III (Table 30-2) and the primosome (Table 30-4) are also required.

(ii) “Rightward” transcription from *p_R* terminates with ~50% efficiency at *t_{R1}*, to yield transcript R1, and otherwise at *t_{R2}* to yield transcript R2. R1 contains only the *cro* gene transcript, whereas R2 also contains the *cII, O*, and *P* gene transcripts.

(iii) “Rightward” transcription from *p'_R* terminating at *t'_R* yields a short transcript, R4, that specifies no protein.

L1, R1, and R2 are translated by host ribosomes to yield proteins whose functions are described below.

2. Delayed early transcription The second transcriptional phase commences as soon as a significant quantity of the protein *gpN* (gp for gene product) accumulates. *This protein, through a mechanism considered below, acts as a transcriptional antiterminator at termination sites *t_{L1}, t_{R1}, and t_{R2}* (Fig. 33-31b):*

(i) Leftward transcript L1 is extended to form L2, which additionally contains the transcripts of the *cIII, xis*, and *int* genes (which encode proteins involved in switching between the lytic and lysogenic modes; Sections 33-3C and 33-3D) together with the *b region* gene transcripts (which specify the so-called **accessory proteins** that, although not essential for lytic growth, increase its efficiency).

(ii) Transcript R2 is extended to form transcript R3, which additionally encodes a second antiterminator, *gpQ*, whose function is discussed below. The continuing translation of R2 and later R3 to yield *gpO* and *gpP*, proteins that are both required for λ DNA replication, stimulates viral DNA production. Similarly, the translation of R1 and later R3 yields **Cro protein (gpcro)**, a repressor of both the “rightward” and “leftward” genes (see below; *cro* stands for control of repressor and other things).

*At this stage, ~15 min postinfection, Cro protein has accumulated in sufficient quantity to bind to operators *o_L* and *o_R*, thereby shutting off transcription from *p_L* and *p_R*. This is more than just efficient use of resources; the overexpression of the early genes, as occurs in λ *cro*⁻ phage, poisons the lytic cycle’s late phase.*

3. Late transcription In the final transcriptional phase (Fig. 33-31c), the antiterminator *gpQ* acts to extend transcript R4 through *t'_R* to form transcript R5. The “gene dosage” effect of the ~30 copies of phage DNA that have accumulated by the beginning of this stage results in the rapid synthesis of the capsid-forming proteins (which are all encoded by late genes; their assembly to form mature phage particles is described in Section 33-3B), as well as *gpR*, *gpR_z*, and *gpS*, which catalyze host cell lysis [*gpR* is a transglycosidase that cleaves the bond between NAG and NAM in the host cell wall peptidoglycan (Section 11-3B); *gpR_z* is an endopeptidase that hydrolyzes a peptidoglycan peptide bond; and *gpS* forms pores in the cell membrane, thereby providing *gpR* and *gpR_z* with access to their peptidoglycan

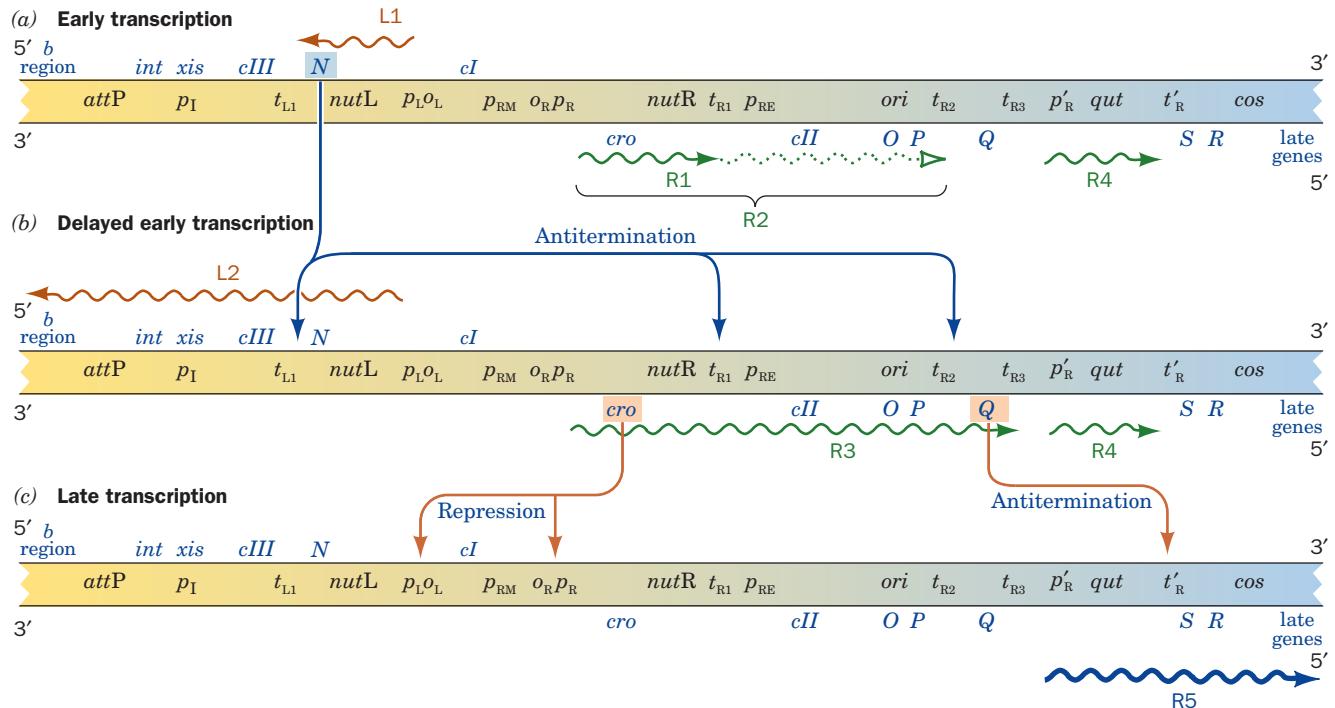


Figure 33-31 Gene expression in the lytic pathway of phage λ .

Genes specifying proteins that are transcribed to the “left” and “right” are shown above and below the phage chromosome. Control sites are indicated between the DNA strands. The genetic map is not drawn to scale and not all of the genes or control sites are indicated. Transcripts are represented by wiggly arrows pointing in the direction of mRNA elongation; the actions of regulatory proteins are denoted by arrows pointing

from each regulatory protein to the site(s) it controls. The lytic pathway has three transcriptional phases: (a) early transcription, (b) delayed early transcription, and (c) late transcription. Gene expression in each of the latter two phases is regulated by proteins synthesized in the preceding phase as is explained in the text. [After Arber, W., in Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, p. 389, Cold Spring Harbor Laboratory (1983).]

substrate]. The first phage particle is completed ~ 22 min postinfection.

b. Antitermination Requires the Action of Several Proteins

Transcriptional control in the λ lytic phase is exerted by gpN- and gpQ-mediated antitermination rather than by repressor binding at an operator site through which, for example, lac operon expression (Section 31-1Ad) is regulated. gpN (107 residues) acts at both rho-dependent and rho-independent termination sites [t_{L1} and t_{R1} are rho dependent (and are, in fact, the terminators with which rho was originally identified), whereas t_{R2} is rho independent; transcriptional termination is discussed in Section 31-2D]. Yet, gpN does not act at just any transcriptional termination site. Rather, genetic analysis of mutant phage defective for antitermination has established the existence of two so-called *nut* (for *N* utilization) sites that are required for antitermination: ***nutL***, which is located between p_L and *N*, and ***nutR***, which occurs between *cro* and t_{R1} (Fig. 33-31). These sites have closely similar sequences consisting of two elements, ***boxB***, whose transcripts can form hydrogen-bonded hairpin loops, and ***boxA*** (Fig. 33-32a).

What is the mechanism of gpN-mediated antitermination? The observation that some *E. coli* defective in antitermination have mutations that map in the *rpoB* gene (which

encodes the RNA polymerase β subunit) suggests that gpN acts at *nut* sites to render core RNA polymerase (lacking a σ subunit) resistant to termination. Indeed, gpN-modulated RNA polymerase will pass over many different

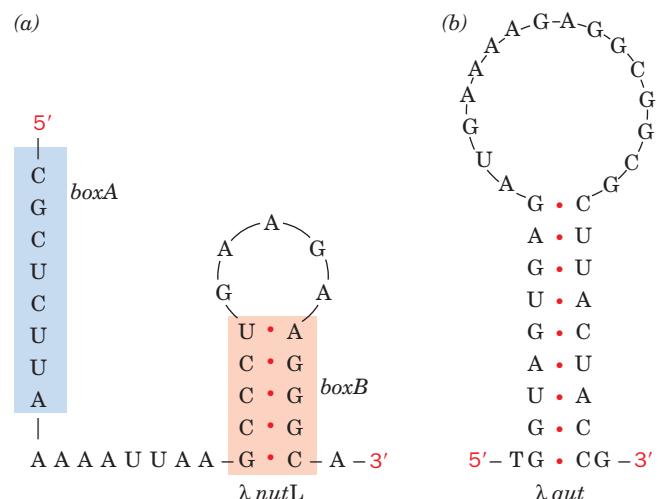


Figure 33-32 The RNA sequences of the phage λ control sites: (a) *nutL*, which closely resembles *nutR*, and (b) *qut*. Each of these control sites is thought to form a base-paired hairpin.

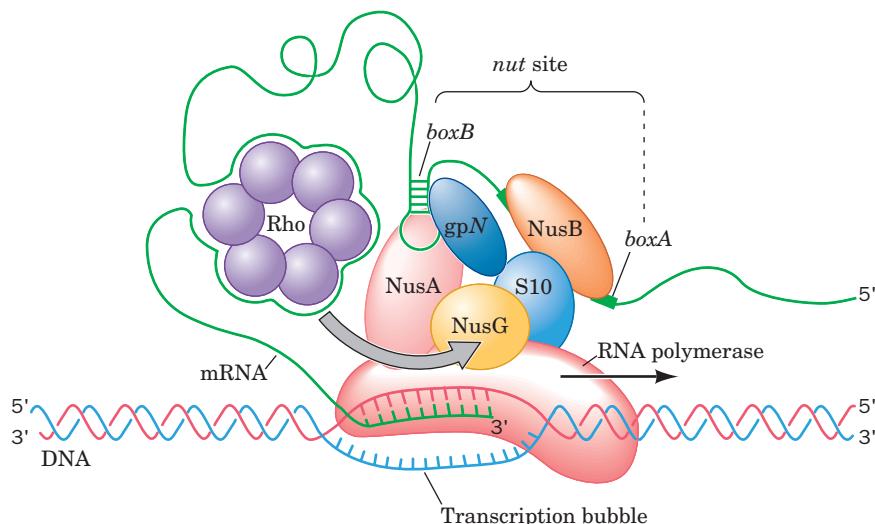


Figure 33-33 Schematic model of the antitermination complex between transcribing RNA polymerase, gpN, and Nus proteins. gpN and the Nus proteins form a complex on a *nut* site of the nascent RNA that binds to the transcribing RNA polymerase further along the looped-out RNA. This complex inhibits RNA polymerase from pausing at a transcriptional termination site,

terminators that it encounters either naturally or by experimental design. A variety of evidence, including the observation that covering *nut* RNA with ribosomes prevents antitermination, indicates that gpN recognizes this site on RNA, not DNA.

Genetic analyses have revealed that antitermination requires several other host factors termed **Nus** (for *N* utilization substance) **proteins** (Fig. 33-33): **NusA**, which specifically binds to both gpN and RNA polymerase; **NusE** (which, interestingly, is ribosomal protein S10) and **NusG**, which both bind to RNA polymerase; and **NusB**, which binds to S10. On encountering a *nut* site, gpN forms a complex with the Nus proteins and RNA polymerase that travels with this enzyme during elongation and inhibits it from pausing at termination sites. At rho-independent terminators, this deters the release of the transcript at the terminator's weakly bound poly(U) segment, whereas at rho-dependent terminators, it may prevent rho factor from overtaking RNA polymerase, thereby stopping it from unwinding and thus releasing the transcript at the transcription bubble. Alternatively, since it has been shown that NusG binds directly to rho, this interaction, as modulated by gpN, may inhibit rho from releasing the nascent transcript.

The *boxB* RNA is recognized by gpN via the latter's ~18-residue, Arg-rich, N-terminal segment. The NMR structure of the 15-nt *boxB* hairpin from the lambdoid (λ -like) **bacteriophage P22** (which grows on *Salmonella typhimurium*), in complex with the 20-residue N-terminal segment of its gpN, was determined by Dinshaw Patel. It reveals that the peptide forms a helix that binds against the major groove face of the *boxB* hairpin via electrostatic and hydrophobic interactions (Fig. 33-34). This presumably orients the opposite face of the RNA for interactions with host factors.

which may prevent rho factor from overtaking the RNA polymerase so as to release the transcript. Another possibility is that transcript release may be inhibited by a gpN-modulated direct interaction between NusG and rho factor (curved gray arrow). [After Greenblatt, J., Nodwell, J.R., and Mason, S.W., *Nature* **364**, 402 (1993).]

Transcriptional antitermination is not limited only to certain bacteriophage. Indeed, the seven ribosomal RNA (*rrn*) operons of λ 's host organism, *E. coli* (which encode

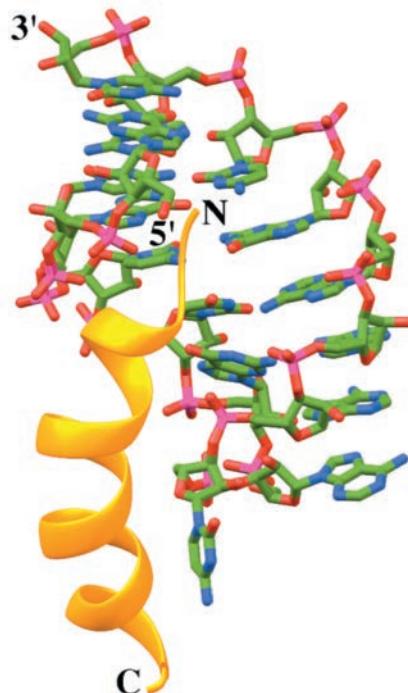


Figure 33-34 NMR structure of the bacteriophage P22 *boxB* RNA in complex with the 20-residue, Arg-rich, N-terminal segment of its gpN. The peptide is represented by a gold ribbon and the RNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P magenta). [Based on an NMR structure by Dinshaw Patel, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1A4T.]

its 5S, 16S, and 23S RNAs; Section 31-4B), each contain a *boxA*-like element, which together with the Nus proteins, mediates antitermination at *rrn* (which probably explains the function of S10 as a Nus protein). This suggests that λ *boxA* is a defective form of *rrn boxA* that requires the presence of gpN bound to *boxB* in addition to the Nus proteins to inhibit termination.

gpQ (207 residues), which overrides t'_R to permit late transcription, acts at a *qut* site (analogous to the *nut* sites) that is located some 20bp downstream from p'_R and that can form an RNA hairpin similar to those of the *nut* sites (Fig. 33-32b). Curiously, however, gpQ-mediated antitermination occurs via a mechanism that is quite different from that mediated by gpN. In fact, gpQ binds specifically to *qut* DNA, not to RNA, where together with NusA it binds to RNA polymerase holoenzyme that is paused at p'_R during the initiation phase, thereby accelerating it out of this promoter site and somehow inducing it not to terminate transcription at t'_R .

c. gpO and gpP Participate in λ DNA Replication

The course of DNA replication in phage λ is diagrammed in Fig. 33-35. Electron microscopy indicates that in the early stages of lytic infection, λ DNA replication occurs via the bidirectional θ mode (Section 30-1Aa) from a single replication origin (**ori**). However, by the late stage of the lytic program, when ~ 50 λ DNA circles have been synthesized, θ mode DNA replication ceases, probably due to exhaustion of one or more of the required host proteins. At this point, around 3 of the ~ 50 DNA circles commence replication via the rolling circle (σ) mode (Section 30-3Bb), with the accompanying synthesis of the complementary strand, although the mechanism of the switchover between the two modes of DNA replication is unclear. The host RecBCD protein (Section 30-6Ad), a nuclease that would rapidly fragment the resulting concatemeric (consisting of tandemly linked identical units) linear duplex DNA, is inactivated by the phage γ protein.

In the process of phage assembly (Section 33-3B), the concatemeric DNA is specifically cleaved in its *cos* (for cohesive-end site) site to yield the linear duplex DNA with complementary 12-nt single-stranded ends that are contained in mature phage particles. The staggered double-stranded scission is made by the so-called **terminase**, which is a complex of the phage proteins gpA (641 residues) and gpNu1 (181 residues).

Phage λ DNA is replicated by the host DNA replication machinery (Sections 30-1 to 30-3) with the participation of only two phage proteins, gpO (333 residues) and gpP (233 residues). gpO, presumably as dimers, specifically binds to four repeated 18-bp palindromic segments within the phage DNA *ori* region, whereas gpP interacts with both gpO and the DnaB protein of the host primosome. gpO and gpP, it is thought, act analogously to host DnaA and DnaC proteins, which as we saw, are required for the initiation of replication of *E. coli* DNA (Section 30-3Ca). Nevertheless, DnaA is required for λ DNA replication. The gpO and gpP proteins apparently function to recognize the λ *ori* site, which, curiously, lies within the *O* gene.

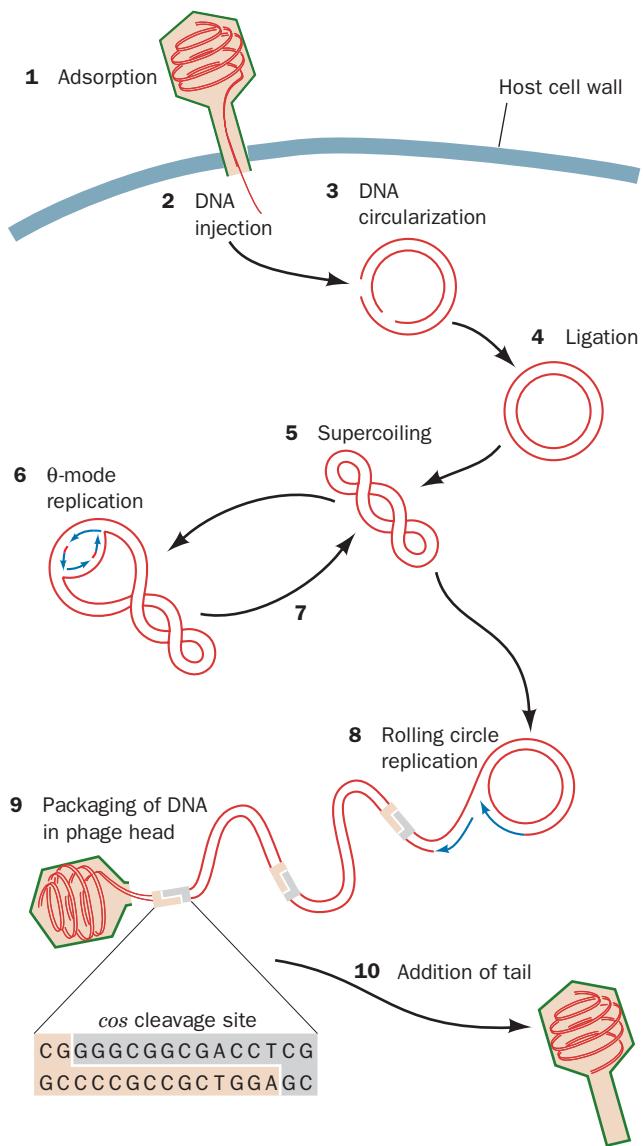


Figure 33-35 DNA replication in the lytic mode of bacteriophage λ . The phage particle adsorbs to the host cell (1) and injects its linear duplex DNA chromosome (2). The DNA circularizes by base pairing at its complementary single-stranded ends (3), and the resulting nicked circle is covalently closed (4) and supercoiled (5) by the sequential actions of host DNA ligase and host DNA gyrase. DNA replication commences according to both the bidirectional θ mode (6 and 7) and the rolling circle mode (8) but in the later stages of infection occurs exclusively by the rolling circle mode. Here curved blue arrows indicate the most recently synthesized DNA at the replication forks and the arrowheads represent the 3' ends of the growing DNA chains. The concatemeric DNA produced by the rolling circle mode is specifically cleaved at its *cos* sites (shaded boxes) and is packaged into phage heads (9). The addition of tails (10) completes the assembly of the mature phage particles, which are each capable of initiating a new round of infection. [After Furth, M.E. and Wickner, S.H., in Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, p. 146, Cold Spring Harbor Laboratory (1983).]

B. Virus Assembly

The mature λ phage head contains two major proteins: **gpE** (341 residues), which forms its polyhedral shell, and **gpD** (110 residues), which “decorates” its surface. Electron microscopy indicates that these proteins, which are present in equal numbers, are arranged on the surface of a $T = 7$ icosadeltahedron. However, the λ head also contains four major proteins, **gpB**, **gpC**, **gpFII**, and **gpW**, which form a cylindrical structure that attaches the tail to the head. This **head-tail connector** occurs at one of the head’s 5-fold vertices and thereby breaks its icosahedral symmetry. Hence, gpE and gpD are present in somewhat fewer than the 420 copies/phage in a perfect $T = 7$ icosadeltahedron.

The tail is a tubular entity that consists of 32 stacked hexagonal rings of **gpV** (246 residues) for a total of 192 subunits. The tail begins with a complex adsorption organelle composed of five different proteins, **gpG**, **gpH**, **gpL**, **gpM**, and **gpJ**, and ends with an assembly of **gpU** and **gpZ** (Fig. 33-28).

The study of complex virus assembly has been motivated by the conviction that it will provide a foundation for understanding the assembly of cellular organelles. Phage assembly is studied through a procedure developed by Robert Edgar and William Wood that combines genetics, biochemistry, and electron microscopy. Conditionally lethal mutations (either temperature-sensitive mutants, which appear normal at low temperatures but exhibit a mutant phenotype at higher temperatures; or suppressor-sensitive *amber* mutants, Section 32-2E) are generated that, under nonpermissive conditions, block phage assembly at various stages. This process results in the accumulation of intermediate assemblies or side products that can be isolated and structurally characterized through electron microscopy. The mutant protein can be identified, through a process known as **in vitro complementation** (in analogy with *in vivo* genetic complementation; Section 1-4C), by mixing cell-free extracts containing these structural intermediates with the corresponding normal protein to yield infectious phage particles.

The assembly of bacteriophage λ occurs through a branched pathway in which the phage heads and tails are formed separately and then joined to yield mature virions.

a. Phage Head Assembly

λ phage head assembly occurs in five stages (Fig. 33-36, right):

1. Two phage proteins, **gpB** (533 residues) and **gpNu3**, together with two host-supplied chaperonin proteins, GroEL and GroES, interact to form an “initiator” that consists of 12 copies of gpB arranged in a ring with a central

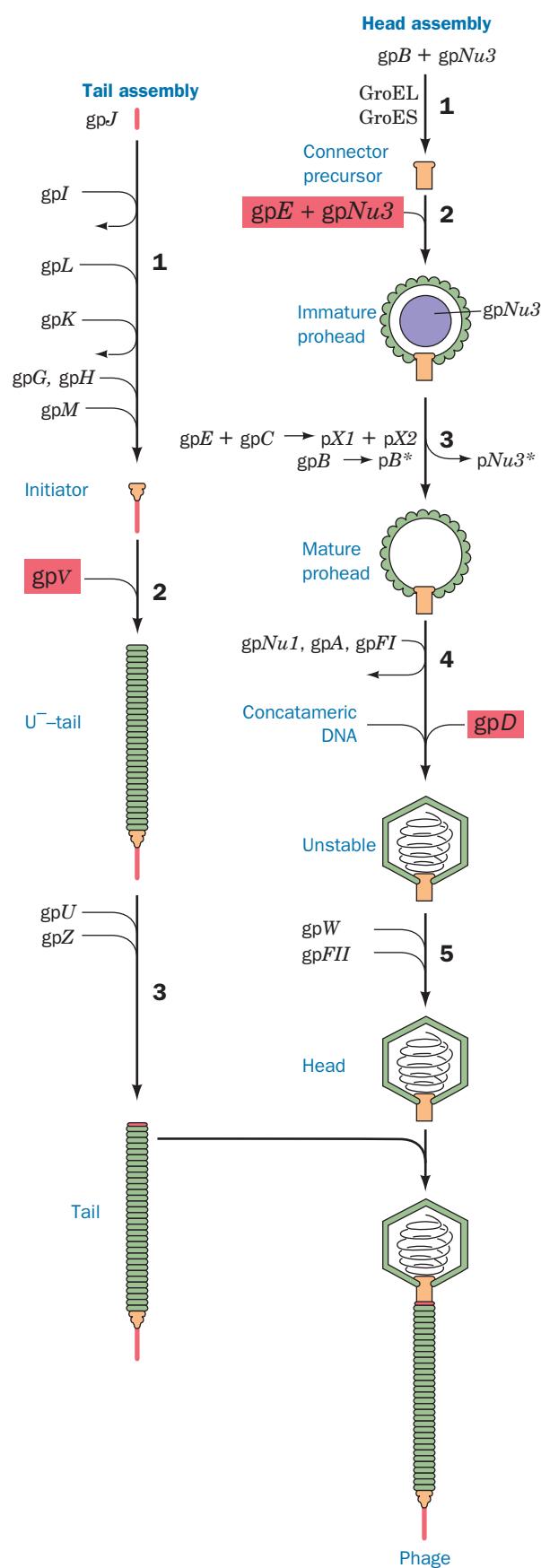


Figure 33-36 The assembly of bacteriophage λ . The heads and tails are assembled in separate pathways before joining to form the mature phage particle. Within each pathway the order of the various reactions is obligatory for proper assembly to occur. gpE, gpNu3, gpD, and gpV are highlighted in red boxes to indicate that relatively large numbers of these proteins are required for phage assembly. The numbered steps are described in the text.

orifice. This precursor of the mature phage head-tail connector (Fig. 33-28) apparently organizes the phage head's subsequent formation. GroEL and GroES, it will be recalled, provide a protected environment that facilitates the proper folding and assembly of proteins and protein complexes such as the connector precursor (Section 9-2C). In fact, these chaperonins were discovered through their role in λ assembly. gpNu3, as we shall see, also functions as a molecular chaperone in that it has but a transient role in phage head assembly.

2. gpE and additional gpNu3 associate to form a structure called an immature **prohead**. If gpB, GroEL, or GroES is defective or absent, some gpE assembles into spiral or tubular structures, which indicates that the missing proteins guide the formation of a proper shell. The absence of gpNu3 results in the formation of but a few shells that contain only gpE. gpNu3 evidently facilitates proper shell construction and promotes the association of gpE with gpB.

3. In the formation of the mature prohead, ~75% of the gpB has its N-terminal 22-residue segment excised to form **gpB***; the gpNu3 is degraded and lost from the structure; and 10 copies of gpC (439 residues) participate in a fusion-cleavage reaction with 10 additional copies of gpE to yield the hybrid proteins **pX1** and **pX2** (p for protein), which form the collar that apparently holds the connector in place. This maturation process, which involves only phage gene products that are part of the immature prohead, requires that all of the prohead components be present and functional, that is, that the immature prohead be correctly assembled to start with. The enzyme(s) that catalyzes this process has, nevertheless, not been identified.

4. The concatemeric viral DNA is packaged in the phage head and cleaved by mechanisms discussed below. During this process, the capsid proteins undergo a conformational change that results in an expansion of the phage head to twice its original volume (a process that occurs in 4M urea in the absence of DNA). gpD then attaches to newly exposed binding sites on gpE, thereby partially stabilizing the capsid's expanded structure.

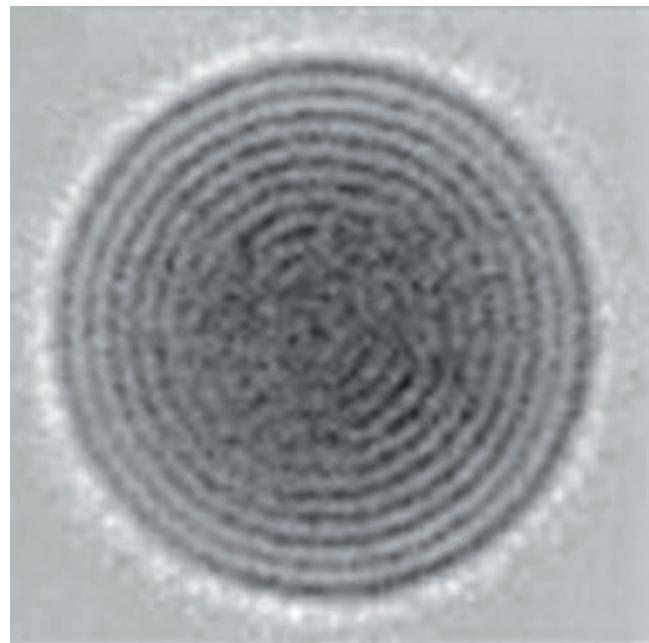
5. In the final stage of phage head assembly, gpW (68 residues) and gpFII (117 residues) add in that order to stabilize the head and form the tail-binding site.

These stages of phage head assembly, as well as some of their component reactions, must proceed in an obligatory order for proper assembly to occur. Of particular interest is that *the components of the mature phage head are not entirely self-assembling as are, for example, TMV (Section 33-1B) and ribosomes (Section 32-3Ac)*. Rather, the *E. coli* proteins GroEL and GroES, as we saw, facilitate head-tail connector assembly. Moreover, gpNu3, which occurs in ~200 copies inside the immature prohead but is absent from the mature prohead, evidently acts as a "scaffolding" protein that organizes gpE to form a properly assembled phage head. Finally, *since phage assembly involves several proteolytic reactions, it must also be considered to occur via enzyme-directed processes*.

b. DNA Is Tightly Packed in the Phage Head

An intriguing question of tailed phage assembly is, how does a relatively small phage head (55 nm in diameter in λ) package a far longer (16,500 nm in λ), stiff dsDNA molecule? Cryo-EM studies of DNA-filled phage heads from **bacteriophage T7** (a tailed coliphage) appear to answer this question. The images of these phage heads seen in axial view (along the line through the capsid vertex containing the head-tail connector to the center of the particle) reveal a striking pattern of at least 10 concentric rings, with only the outer ring, which is slightly thicker than the others, representing the protein shell (Fig. 33-37a). In contrast, the

(a)



(b)



Figure 33-37 The packing of dsDNA inside a T7 phage head. (a) A cryo-EM-based image of a dsDNA-filled T7 phage head as viewed along a line from the head-tail connector to the center of the particle. The outer, somewhat thicker ring represents the phage's protein capsid. The nine inner rings, whose spacing is 2.5 nm, represent coiled dsDNA. [Courtesy of Alasdair Steven, NIH, Bethesda, Maryland.] (b) A drawing of the concentric shell model in which the DNA is wound inward like a spool of twine about the phage's long axis. [After Harrison, S.C., *J. Mol. Biol.* **171**, 579 (1983).]

side views of these phage heads show only punctate (marked with dots or spots) patterns that in some places form linear features. Computer modeling indicates that these patterns can be accounted for by the spooling of the DNA in concentric shells around the axis through the connector (Fig. 33-37b). Six such shells are required to accommodate the entire 40-kb T7 DNA into its 55-nm-diameter T7 phage head. This does not contradict the observation of at least nine concentric rings of DNA because, in a given shell, the DNA is coiled more tightly toward the poles of the phage head and therefore appears in projection as multiple rings of lower radii. Since the DNA linearly enters the phage prohead through the head-tail connector (see below), it has been proposed that its stiffness causes it to first coil against the inner wall of the rigid protein shell and then to wind concentrically inward, much like a spool of twine. Nevertheless, the DNA's detailed winding path varies randomly from particle to particle as is indicated by the observation that packaged DNA can be cross-linked to the capsid along its entire length.

c. DNA Is “Pumped” into the Phage Head by an ATP-Driven Process

The packaging of λ DNA begins when terminase (gpA + gpNu1) binds to its recognition sequences on a randomly selected \sim 200-bp *cos* site. The resulting complex then binds to the prohead so as to introduce the DNA into it through the orifice in its head-tail connector. The “left” end of the DNA chromosome enters the prohead first as is indicated by the observation that only this end of the chromosome is packaged by an *in vitro* system when λ DNA restriction fragments are used. Whether the cutting of the initial *cos* site precedes or follows the initiation of packaging is unknown. However, at least *in vitro*, this process requires the binding to *cos* of the *E. coli* histonelike protein known as **integration host factor (IHF)**. IHF binds specific sequences of duplex DNA, which it wraps around its surface, thereby inducing a U-turn in the DNA (see below).

The packing of double-stranded DNA (dsDNA) inside a phage head is not only entropically unfavorable but is also enthalpically opposed due to dsDNA's stiffness and its large intramolecular charge repulsions in the coiled state. In fact, the pressure inside a dsDNA virus is as high as 50 atm, over 10 times that in any other steady-state biological system (and much higher than a lipid bilayer can withstand). Consequently, the dsDNA must be actively “pumped” into the phage head, which occurs via an ATP-driven process. Force measurements indicate that bacteriophage dsDNA packaging motors are among the most powerful known biological motors; the force they generate is over 20-fold greater than that by a myosin molecule in muscle (Section 35-3Bd). Indeed, since the bacteriophage λ packages dsDNA at an average rate of \sim 600 bp/s (and hence packages its entire 48.5-kb chromosome in \sim 80 s), it generates around twice the power per unit weight as a typical automobile engine. The injection of λ DNA into a host bacterium by a mature phage is a spontaneous process that, once it has been triggered, is driven by the free energy stored in the compacted DNA.

The structure of the head-tail connector of bacteriophage λ is unknown. However, Baker and Rossmann have determined the X-ray structure of the head-tail connector of **bacteriophage ϕ 29** (a tailed dsDNA-containing phage that infects *Bacillus subtilis*), which as in bacteriophage λ , is the portal through which dsDNA enters and exits the phage head. The structure reveals that the ϕ 29 head-tail connector consists of a funnel-shaped cyclic dodecamer of identical 309-residue subunits that is 75 Å high, has a maximum width of 69 Å, and encloses a central channel whose diameter is 36 Å at its narrow end and 60 Å at its wide end (Fig. 33-38).

Cryo-EM studies of ϕ 29 by Rossmann indicate that the connector is mounted at a pentagonal vertex of the phage prohead with its narrow end protruding into the exterior

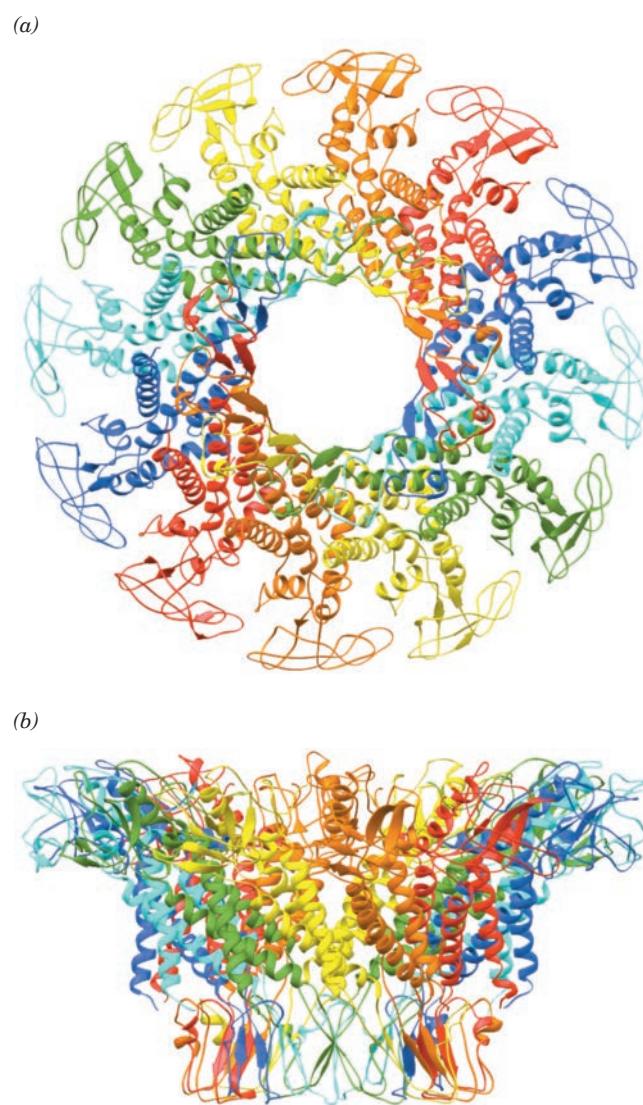


Figure 33-38 X-ray structure of the bacteriophage ϕ 29 head-tail connector. The cyclic dodecameric protein is shown in ribbon form with its identical subunits in different colors. (a) View along the 12-fold axis. (b) View perpendicular to Part a. [Based on an X-ray structure by Timothy Baker and Michael Rossmann, Purdue University. PDBid 1FOU.]

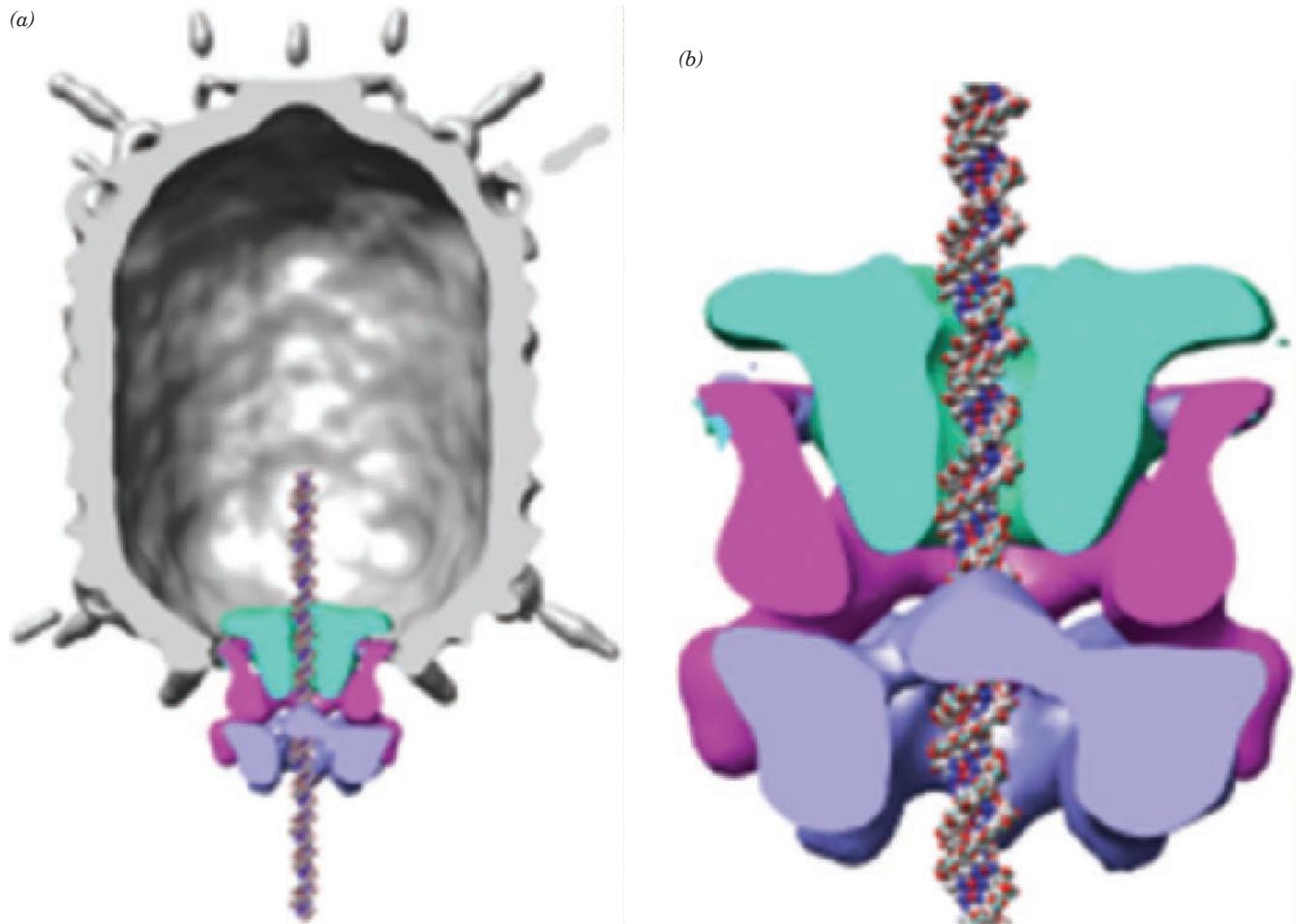


Figure 33-39 Cryo-EM of the bacteriophage φ29 prohead. (a) Cutaway diagram of the empty prohead (gray) showing the positions of the connector (green), the pRNA (magenta), and ATPase (lavender). Canonical dsDNA, drawn in space-filling

form with C white, N blue, O red, and P green, has been modeled into the central channel of the packaging motor. (b) Close-up of the connector-pRNA-ATPase-dsDNA assembly. [Courtesy of Michael Rossmann, Purdue University.]

(Fig. 33-39a). The prohead, in addition, contains a virally encoded 174-nt RNA known as **pRNA** (p for prohead), whose presence is essential for DNA packaging but which is absent in the mature virion. Cryo-EM studies reveal that the pRNA forms a cyclic homopentamer that surrounds the narrow end of the dodecameric head-tail connector and is sandwiched between the connector and a cyclic homopentamer of ATPases that drive the DNA packaging (Fig. 33-39b). This has led to a model in which the five ATPases fire successively around the cyclic pentamer, thereby generating a traveling wave of conformational changes that, in essence, causes the ATPase to walk along the dsDNA's helical groove(s) so as to pump it through the connector into the prohead (a mechanism that resembles that postulated for hexagonal helicases such as E1 protein; Section 30-2Ca). This would translocate the DNA by one-fifth of its helical pitch for every ATP hydrolyzed, which for canonical B-DNA (which has 10 bp/turn) is the height of two base pairs, a quantity that

is consistent with the observed ATP consumption of φ29 during packaging. The DNA packaging engine of bacteriophage λ most likely has a similar mechanism, although φ29 and its close relatives are the only phages known to have a pRNA.

The final step in the bacteriophage λ DNA packaging process is the recognition and cleavage of the next *cos* site (Fig. 33-35) on the concatemeric DNA by terminase, possibly with the participation of **gpFI**. Phage λ therefore contains a unique segment of DNA (in contrast to some phages in which the amount of DNA packaged is limited by a “headful” mechanism that results in their containing somewhat more DNA than an entire chromosome). In fact, the λ packaging system will efficiently package a DNA that is 75 to 105% the length of the wild-type λ DNA as long as it is flanked by *cos* sites (the central third of the phage DNA, which encodes the dispensable accessory genes, can be replaced by other sequences, thereby making phage λ a useful cloning vector; Section 5-5Bb).

d. Tail Assembly

Tail assembly, which occurs independently of head assembly, proceeds, as a comparison of Figs. 33-28 and 33-36 indicates, from the 200-Å-long tail fiber toward the head-binding end. This strictly ordered series of reactions can be considered to have three stages (Fig. 33-36, *left*):

1. The formation of the “initiator,” which ultimately becomes the adsorption organelle, requires the sequential actions on gpJ (the tail fiber protein) of the products of phage genes **I**, **L**, **K**, **G**, **H**, and **M**. Of these, only **gpI** and **gpK** are not components of the mature tail.

2. The initiator forms the nucleus for the polymerization of gpV, the major tail protein, to form a stack of 32 hexameric rings. The length of this stack is thought to be regulated by gpH (853 residues), which the available evidence suggests becomes extended along the length of the growing tail and somehow limits its growth. λ tail length is apparently specified in much the same way that the helical length of TMV is governed (Section 33-1Ba), although in TMV the regulating template is an RNA molecule rather than a protein.

3. In the termination and maturation stage of tail assembly, **gpU** attaches to the growing tail, thereby preventing its further elongation. The resultant immature tail has the same shape as the mature tail and can attach to the head. In order to form an infectious phage particle, however, the immature tail must be activated by the action of **gpZ** before joining the head.

The completed tail then spontaneously attaches to a mature phage head to form an infectious λ phage particle (Fig. 33-36, *bottom*).

e. The Assembly of Other Double-Stranded DNA Phages Resembles That of λ

The assembly of several other double-stranded DNA bacteriophages has been studied in detail, notably that of coliphages **T4** (Fig. 33-1e) and **T7** and phage **P22**. All of them are formed in assembly processes that closely resemble that of phage λ . For example, their head assembly processes proceed in obligatory reaction sequences through an initiation stage; the scaffolded assembly of a prohead; an ATP-driven DNA packaging process, in which the DNA assumes a tightly packed conformation and the prohead undergoes an expansion; and a final stabilization. The mature phages then form by the attachment of separately assembled tails to the completed and DNA-filled heads.

C. The Lysogenic Mode

Lysogeny is established by the integration of viral DNA into the host chromosome accompanied by the shutdown of all lytic gene expression. With phage λ , integration takes place through a site-specific recombination process that differs from homologous recombination (Section 30-6A) in that it occurs only between the chromosomal sites designated *attP*

on the phage and *attB* on the bacterial host (Fig. 33-40). These two attachment sites have a 15-bp identity (Fig. 33-41), so they can be represented as having the sequences *POP'* for *attP* and *BOB'* for *attB*, where *O* denotes their common sequence. *attB* comprises only the 23 bp drawn in Fig. 33-41, whereas the entire *attP* sequence spans >250 bp. Phage integration occurs through a process that yields the inserted phage chromosome flanked by the sequence *BOP'* on the “left” (the *attL* site) and *POB'* on the “right” (the *attR* site; Fig. 33-40). The nature of the crossover site was determined through the use of ^{32}P -labeled bacterial DNA and unlabeled phage DNA. The crossover site occurs at a unique position on each strand that is displaced with respect to its complementary strand so as to form a staggered recombination joint (Fig. 33-41).

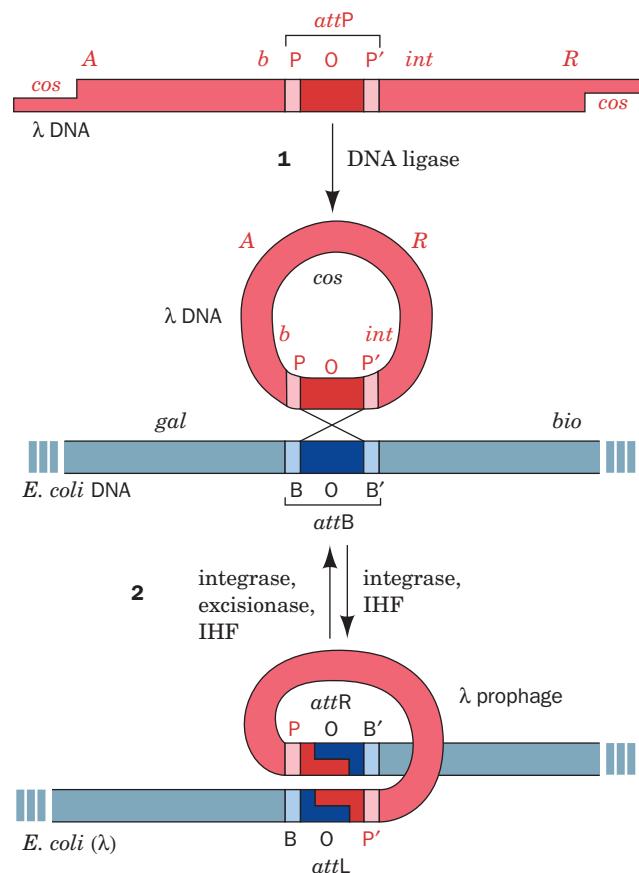


Figure 33-40 Site-specific recombination in bacteriophage λ . This schematic diagram shows (1) the circularization of the linear phage λ DNA through base pairing between its complementary ends to form the *cos* site; and (2) the integration/excision of this DNA into/from the *E. coli* chromosome through site-specific recombination between the phage *attP* and host *attB* sites. The darker colored regions in the *att* sites represent the identical 15-bp crossover sequences (*O*), whereas the lighter colored regions symbolize the unique sequences of bacterial (*B* and *B'*) and phage (*P* and *P'*) origin. [After Landy, A. and Weisberg, R.A., in Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, p. 212, Cold Spring Harbor Laboratory (1983).]

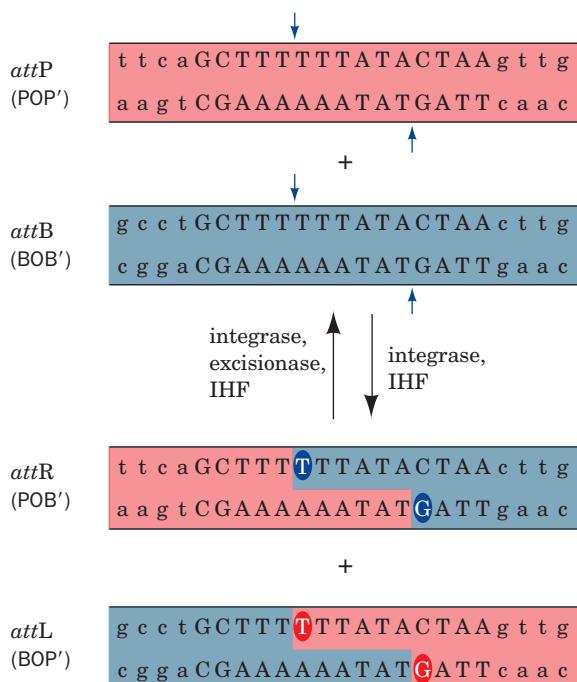


Figure 33-41 The site-specific recombination process that inserts/excises phage λ DNA into/from the chromosome of its *E. coli* host. Exchange occurs between the phage *attP* site (red; only its 23-bp core is shown) and the bacterial *attB* site (blue), and the prophage *attL* and *attR* sites. The strand breaks occur at the approximate positions indicated by the short blue arrows. The sources of the more darkly shaded bases in *attR* and *attL* are uncertain. The uppercase letters represent bases in the O region common to the phage and bacterial DNAs, whereas lowercase letters symbolize bases in the flanking B, B', P, and P' sites.

a. λ Integrase and IHF Mediate λ DNA Integration, whereas Excisionase Is Additionally Required for λ DNA Excision

Phage integration is catalyzed by λ integrase (356-residue subunits; the λ *int* gene product), a homotetrameric enzyme that is homologous to and mediates a site-specific recombination reaction similar to that mediated by the Cre recombinase of bacteriophage P1 (whose structure and mechanism, which involves a Holliday junction intermediate, are discussed in Section 30-6Bg). Integration also requires the participation of IHF, a heterodimer of 30% identical 99- and 94-residue subunits that binds to specific subsites on *attP*. The X-ray structure of IHF in complex with a 35-bp target DNA, determined by Phoebe Rice and Howard Nash, reveals that this pseudo-2-fold symmetric protein wraps the DNA around it in a $>160^\circ$ bend (Fig. 33-42). Most of this bend arises from two large kinks, separated by 9 bp, that are each formed by the intercalation of a highly conserved Pro side chain between two consecutive base pairs.

Since the integration reaction is energy-neutral, how is it made irreversible? λ integrase differs from Cre recombi-

nase in that, in addition to the sequences forming the Holliday junction intermediate, it binds to specific sequences on the outer arms of the *attP* DNA. The simultaneous binding of all these sites to integrase is made possible by U-turns in the *attP* DNA imposed by the binding of IHF. The resulting asymmetry of the integrase tetramer allosterically induces it to resolve the Holliday junction intermediate as the lysogenic integration product rather than reversing the reaction that formed the Holliday junction (recall that Holliday junctions can be resolved in two ways; Section 30-6Aa).

Despite the foregoing, λ integrase also mediates prophage excision. This is possible because excision also requires the participation of **excisionase** (72 residues; the λ *xis* gene product, which is only expressed during the delayed early stage of lytic growth; Section 33-3De) together with IHF and **Fis** (a DNA-binding host protein that also stimulates Hin-mediated gene inversion; Section 30-6Bf). Excisionase (which, by itself, has no catalytic activity and hence is misnamed) binds to a specific site on *POB'*, where it induces a sharp bend in this DNA. The resulting arrangement of the DNA on the integrase during excision, which differs from that during integration, biases the resolution of the Holliday junction intermediate in favor of excision.

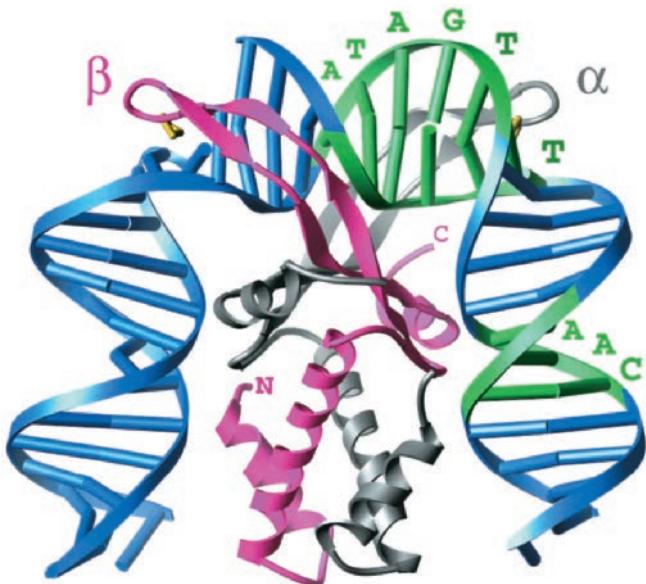


Figure 33-42 X-ray structure of integration host factor (IHF) in complex with a 35-bp target DNA. The structure is viewed with its pseudo-2-fold axis vertical. The α subunit of IHF is gray and its β subunit is magenta. The “top” strand of the 35-bp *att* site DNA, which is shown in ladder form, was synthesized in two segments. The consensus sequence to which IHF binds is highlighted in green and interacts mainly with the β ribbon arm of the α subunit and the body of the β subunit. The Pro side chains near the tip of each β ribbon arm that intercalate between base pairs are drawn in yellow. [Courtesy of Phoebe Rice, University of Chicago. PDBid 1IHF.]

b. The Relative Levels of Cro Protein and *cI* Repressor Determine the λ Phage Life Cycle

The establishment of lysogeny in phage λ is triggered by high concentrations of *gpcII* (see below). This early gene product stimulates “leftward” transcription from two promoters, p_I (I for integrase) and p_{RE} (RE for repressor establishment; Fig. 33-43a):

1. Transcription initiated from p_I , which is located within the *xis* gene, results in the production of integrase but not excisionase. λ DNA is consequently integrated into the host chromosome to form the prophage.

2. The transcript initiated from p_{RE} encodes the *cI* gene whose product is called the λ or *cI* **repressor**. The λ repressor, as does Cro protein (Section 33-3Aa), binds to the o_L and o_R operators, thereby blocking transcription from p_L and p_R , respectively (Fig. 33-43b; note that these operators are upstream from their corresponding promoters rather than downstream as in the *lac* operon; Fig. 31-2). *Both repressors therefore act to shut down the synthesis of early gene products, including Cro protein and gpcII.*

gpcII is metabolically unstable with a half-life of ~ 1 min (see below) so that *cI* transcription from p_{RE} soon ceases. λ repressor bound at o_R , but not Cro protein, however, stimulates “leftward” transcription of *cI* from p_{RM} (RM for repressor maintenance; Fig. 33-43b). In other words, *Cro protein represses all mRNA synthesis, whereas λ repressor stimulates transcription of its own gene while repressing all other mRNA synthesis. This conceptually simple difference between the actions of λ repressor and Cro protein forms the basis of a genetic switch that stably maintains phage λ in either the lytic or the lysogenic state.* The molecular mechanism of this switch is described in Section 33-3D. In the following paragraphs we discuss how this switch is “thrown” from one state to another. You should recognize, however, that, *once the switch is thrown in favor of the lytic cycle, that is, when Cro protein occupies o_L and*

o_R , *the phage is irrevocably committed to at least one generation of lytic growth.*

c. *gpcII* Is Activated when Phage Multiplicity Is High or Nutritional Conditions Are Poor

The reason why a high *gpcII* concentration is required to establish lysogeny is that this early gene product can stimulate transcription from p_I and p_{RE} only when it is in oligomeric form. This phenomenon accounts for the observation that lysogeny is induced when the **multiplicity of infection** (ratio of infecting phages to bacteria) is large (≥ 10) since this gene dosage effect results in *gpcII* being synthesized at a high rate.

gpcII is metabolically unstable because it is preferentially proteolyzed by host proteins, notably ***gphfIA*** and ***gphfIB***. However, *gpcIII* somehow protects *gpcII* from the action of *gphfIA*, which is why its presence enhances lysogenization (Fig. 33-43a). The activity of *gphfIA* is dependent on the host cAMP-activated catabolite repression system (Section 31-3C) as is indicated by the observation that *E. coli* mutants defective in this system lysogenize with less than normal frequency. Yet, if these mutant strains are also *hflA*⁻, they lysogenize with greater than normal frequency. Apparently the *E. coli* catabolite repression system, which is known to regulate the transcription of many bacterial genes, controls *hflA* activity, perhaps by directly repressing this protein’s synthesis at high cAMP concentrations. *This explains why poor host nutrition, which results in elevated cAMP concentrations, stimulates lysogenization.*

Once a prophage has been integrated in the host chromosome, lysogeny is stably maintained from generation to generation by λ repressor. This is because λ repressor stimulates its own synthesis at a rate sufficient to maintain lysogeny in the bacterial progeny while repressing the transcription of all other phage genes. In fact, λ repressor is synthesized in sufficient excess to also repress transcription

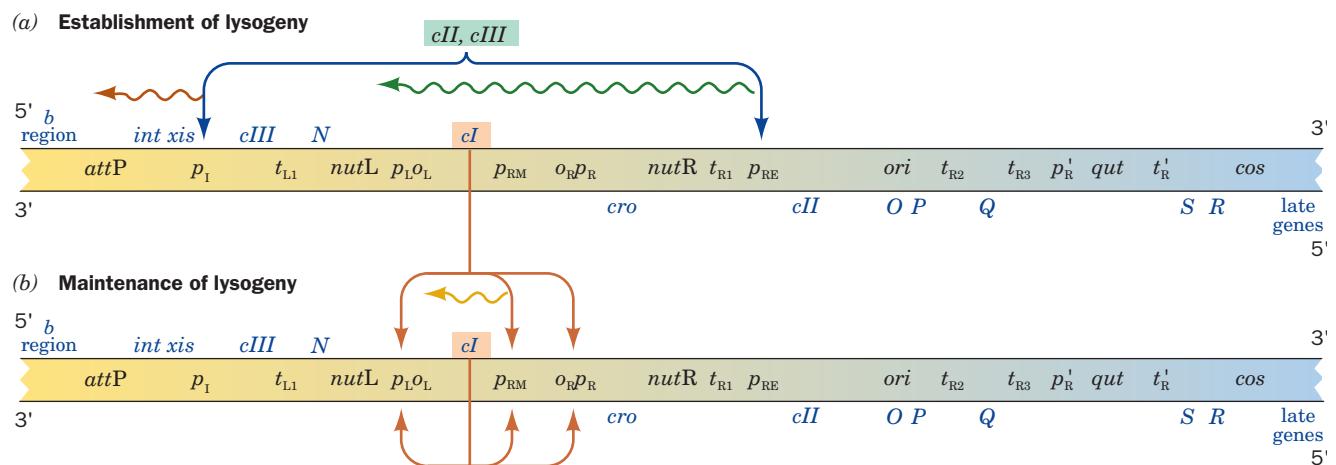


Figure 33-43 Control of gene expression in bacteriophage λ . (a) The establishment of lysogeny. (b) The maintenance of lysogeny. The symbols used are described in the legend of

Fig. 33-31. [After Arber, W., in Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, p. 389, Cold Spring Harbor Laboratory (1983).]

from superinfecting λ phage, thereby accounting for the phenomenon of immunity. We shall see below how induction occurs.

D. Mechanism of the λ Switch

The lysogenic cycle is a highly stable mode of phage λ replication; under normal conditions lysogens spontaneously induce only about once per 10^5 cell divisions. Yet, transient exposure to inducing conditions triggers lytic growth in almost every cell of a lysogenic bacterial culture. In this section, we consider how this genetic switch, whose mechanism was largely elucidated by Mark Ptashne, can so tightly repress lytic growth and yet remain poised to turn it on efficiently.

a. o_R Consists of Three Homologous Palindromic Subsites

Both of the operators to which λ repressor and Cro protein bind, o_L and o_R , consist of three subsites (Fig. 33-44). These are designated o_{L1} , o_{L2} , and o_{L3} for o_L , and o_{R1} , o_{R2} , and o_{R3} for o_R . Each of these subsites consists of a similar 17-bp segment that has approximate palindromic symmetry. However, as we shall see, o_L plays only a minor role in the λ switch relative to that of o_R .

b. λ Repressor and Cro Protein Structurally Resemble Other Repressors

λ repressor binds to DNA as a dimer so that its 2-fold symmetry matches those of the operator subsites to which it binds. The monomer's 236-residue polypeptide chain is folded into two roughly equal sized domains connected by an \sim 30-residue segment that is readily cleaved by proteolytic enzymes. The isolated N-terminal domains retain their ability to bind specifically to operators (although with only half of the binding energy of the intact repressor) but do not dimerize in solution. The C-terminal domains can still dimerize but lack the capacity to bind DNA. Evidently, repressor's N-terminal domain binds operator, whereas its

C-terminal domain provides the contacts for dimer formation.

Although the λ repressor has not been crystallized, its N-terminal domain comprising residues 1 to 92, as excised by treatment with the papaya protease **papain**, does crystallize. The X-ray structure of this protein, both alone and in complex with a 20-bp DNA containing the o_{L1} sequence, was determined by Carl Pabo. The N-terminal domain crystallizes as a symmetric dimer with each subunit containing an N-terminal arm and five α helices (Fig. 33-45a). Two of these helices, $\alpha 2$ and $\alpha 3$, form a helix-turn-helix (HTH) motif, much like those in many other prokaryotic repressors of known structure (Section 31-3Da). The $\alpha 3$ helix, the recognition helix, protrudes from the protein surface such that the two $\alpha 3$ helices of the dimeric protein fit into successive major grooves of the operator DNA. Similar associations are observed in the X-ray structures of the closely related **bacteriophage 434 repressor** N-terminal fragment in complex with a 20-bp DNA containing its operator sequence (Fig. 31-32). The X-ray structure of the C-terminal domain (residues 132–236) of the λ repressor, determined by Mitchell Lewis, reveals how this domain dimerizes (Fig. 33-45b). The intact λ repressor presumably dimerizes on its target DNA as is drawn in Fig. 33-45c.

Cro protein also forms dimers. In contrast to λ or 434 repressor, however, this 66-residue polypeptide forms but one domain that contains both its operator recognition site and its dimerization contacts. The X-ray structure of Cro in complex with a 17-bp tight-binding operator DNA, determined by Brian Matthews, reveals that this dimer likewise contains a pair of HTH units (Fig. 33-46), but which bind to the DNA such that they induce it to bend about the protein by 40° . The sequence-specific binding predicted by this structure is supported by Robert Sauer's genetic studies indicating that mutant varieties of Cro, in which the proposed DNA-contacting residues have been changed, are defective in operator binding. Moreover, this structure closely resembles that of the related **phage 434 Cro protein** in complex with a 20-bp DNA containing its operator sequence (Fig. 31-33).



Figure 33-44 Base sequences of the operator regions of the phage λ chromosome. (a) o_L and (b) o_R . Each of these operators consists of three homologous 17-bp subsites separated by short AT-rich spacers. Each subsite has approximate palindromic (2-fold

symmetry as is demonstrated by the comparison of the two sets of red letters in each subsite. The wiggly arrows mark the transcriptional start sites and directions at the indicated promoters.

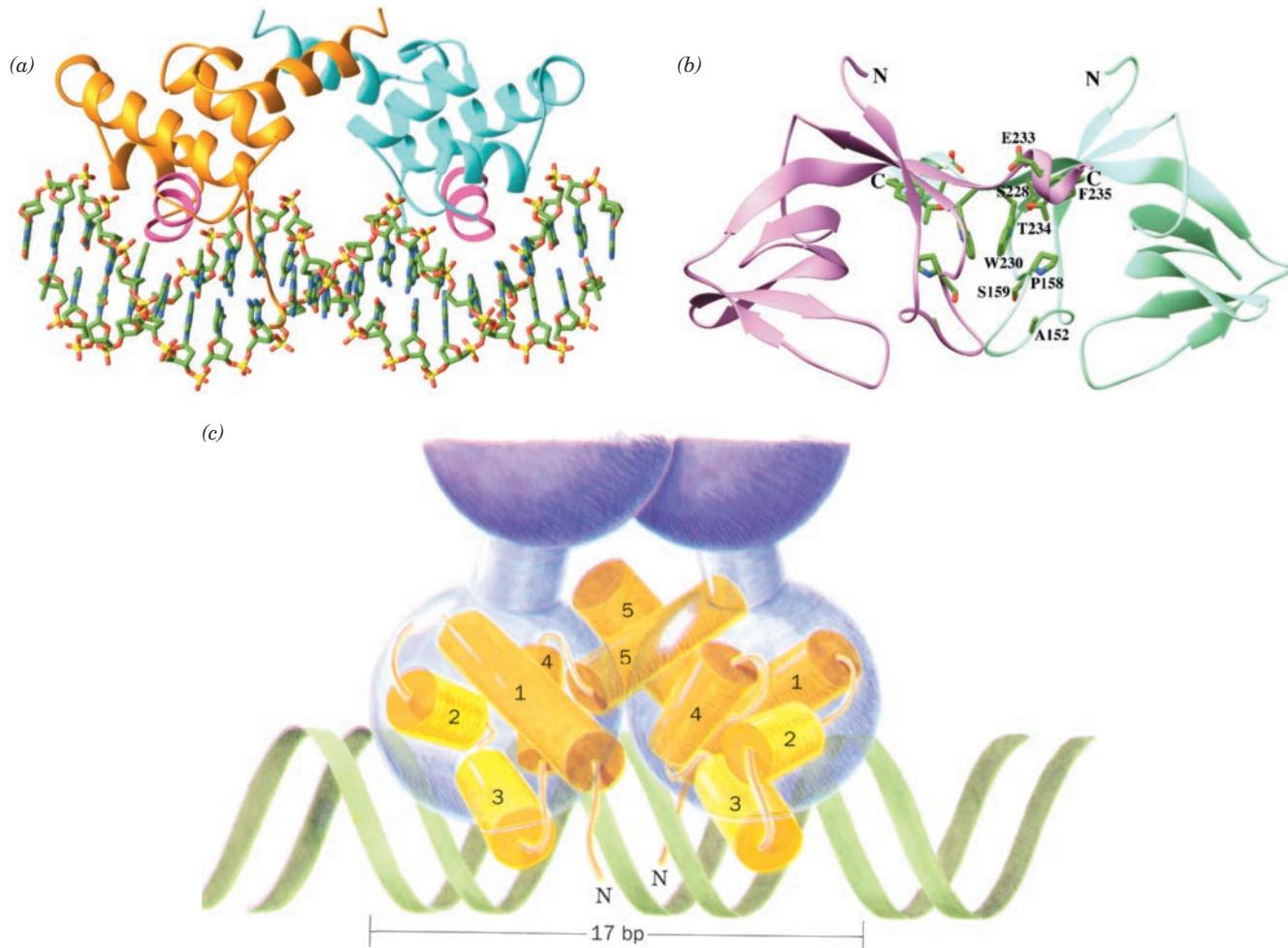
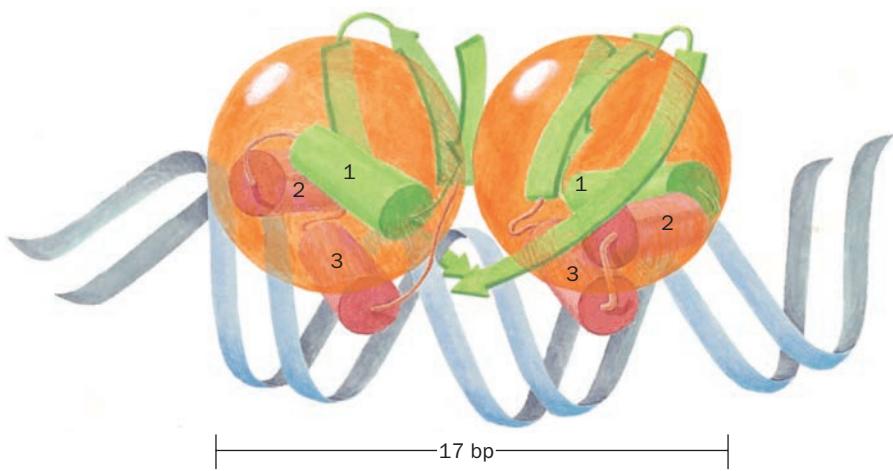


Figure 33-45 Structure of the λ repressor. (a) The X-ray structure of the N-terminal domain homodimer in complex with B-DNA. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P yellow). The two protein subunits are drawn as gold and cyan ribbons with their recognition helices magenta. Note that the protein's N-terminal arms wrap around the DNA. This accounts for the observation that the G residues in the major groove on the repressor-operator complex's "back side" are protected from methylation only when these N-terminal arms are intact. [Based on an X-ray structure by Carl Pabo, The Johns Hopkins University. PDBid 1LMB.] (b) The X-ray structure of the C-terminal domain dimer. Mutations of the residues that are drawn in ball-and-stick form (and which

are labeled for the green subunit) interfere with dimerization. [Based on an X-ray structure by Mitchell Lewis, University of Pennsylvania. PDBid 1F39.] (c) An interpretive drawing indicating how contacts between the repressor's C-terminal domains (*upper lobes*) maintain the intact protein's dimeric character. The λ repressor binds to the 17-bp operator subsites of o_L and o_R as symmetric dimers with the N-terminal domain of each subunit specifically binding to a half-subsite. Note how the α_3 recognition helices of the symmetry related α_2 - α_3 HTH units (*light yellow*) fit into successive turns of the DNA's major groove. [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 36, Cold Spring Harbor Laboratory Press (2004).]  See Interactive Exercise 47

Figure 33-46 X-ray structure of the Cro protein dimer in its complex with B-DNA. Note that the λ repressor (Fig. 33-45), although otherwise dissimilar, contains HTH units that also bind in successive turns of the DNA's major groove. [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 38, Cold Spring Harbor Laboratory Press (2004).]  See Interactive Exercise 48



c. Repressor Stimulates Its Own Synthesis while Repressing All Other λ Genes

Chemical and nuclease protection experiments have indicated that λ repressor has the following order of intrinsic affinities for the subsites of o_R (Fig. 33-47):

$$o_{R1} > o_{R2} > o_{R3}$$

Despite this order, o_{R1} and o_{R2} are filled nearly together. This is because λ repressor bound at o_{R1} cooperatively binds

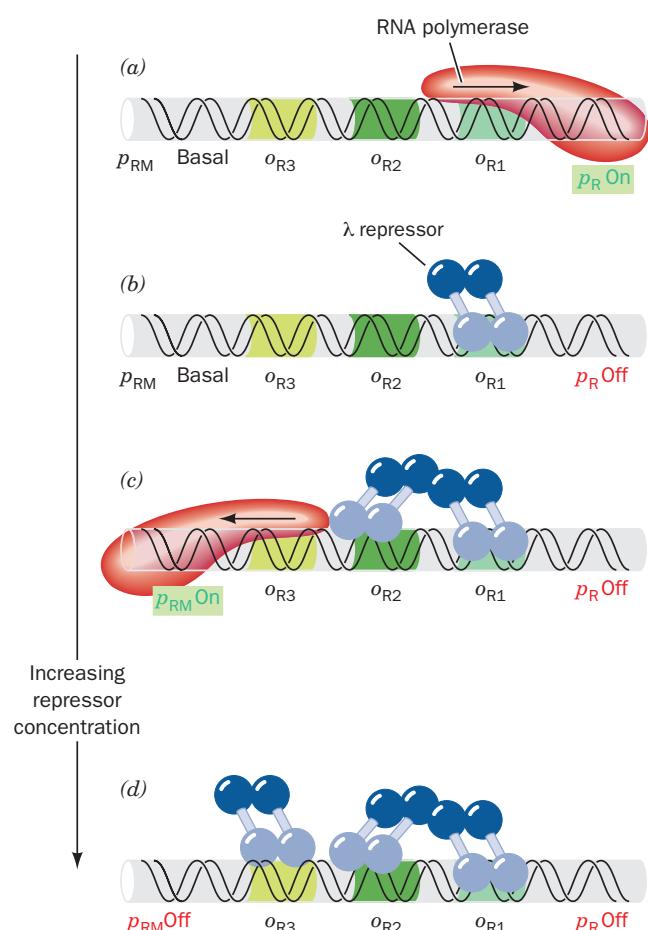


Figure 33-47 The binding of λ repressor to the three subsites of o_R . (a) In the absence of repressor, RNA polymerase initiates transcription at a high level from p_R (right) and at a basal level from p_{RM} . (b) Repressor has ~ 10 times higher affinity for o_{R1} than it does for o_{R2} or o_{R3} . Repressor dimer therefore first binds to o_{R1} so as to block transcription from p_R . (c) A second repressor dimer binds to o_{R2} at only slightly higher repressor concentrations due to specific binding between the C-terminal domains of neighboring repressors. In doing so, it stimulates RNA polymerase to initiate transcription from p_{RM} at a high level (left). (d) At high repressor concentrations, repressor binds to o_{R3} so as to block transcription from p_{RM} . Note that although Parts c and d are drawn with interdimer contacts between the C-terminal domains of only two repressor monomers, this interaction may involve contacts between the C-terminal domains of all four repressor monomers. [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 21, Cold Spring Harbor Laboratory Press (2004).]

repressor at o_{R2} through associations between their C-terminal domains (Fig. 33-47c). o_{R1} and o_{R2} are therefore both occupied at low λ repressor concentrations, whereas o_{R3} becomes occupied only at higher repressor concentrations.

The binding of λ repressor to o_R , as we previously mentioned, abolishes transcription from p_R and stimulates it from p_{RM} (Fig. 33-47c). At high concentrations of λ repressor, however, transcription from p_{RM} is also repressed (Fig. 33-47d). These phenomena have been clearly demonstrated through the construction of a series of hybrid operons that permit the effect of λ repressor on a promoter to be studied in a controlled manner. The system has two elements (Fig. 33-48):

1. A plasmid bearing the *lacI* gene (which encodes *lac* repressor; Section 31-1Ad) and the *lac* operator–promoter sequence fused to the *cl* gene. This construct permits the amount of λ repressor produced to be directly controlled by varying the concentration of the *lac* inducer IPTG (Section 31-1Aa).

2. A prophage containing o_R and either p_{RM} , as Fig. 33-48 indicates, or p_R fused to the *lacZ* gene. The amount of the *lacZ* gene product, β -galactosidase, produced, which can be readily assayed, reflects the activity of p_{RM} (or p_R).

The manipulation of these systems has demonstrated that at intermediate λ repressor concentrations (when o_{R1} and

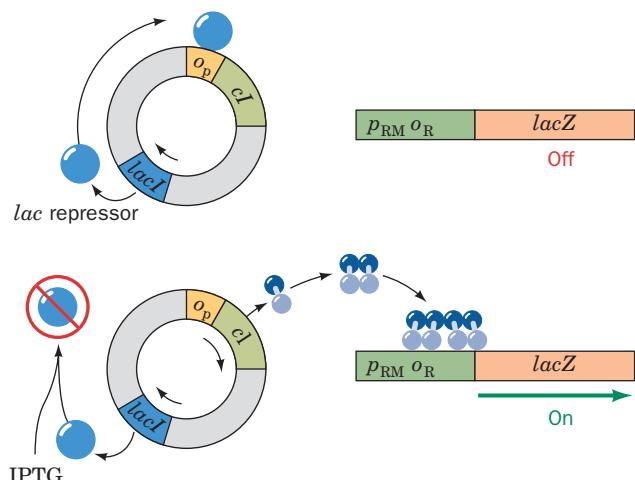


Figure 33-48 The genetic system used to study the effect of λ repressor on p_{RM} . The bacterium contains two hybrid operons. The first (left) is a plasmid bearing the *lac* operator–promoter (O_p) fused to the λ *cl* gene so as to provide a source of repressor. The *lacI* gene, which encodes *lac* repressor, is also incorporated in the plasmid so that the level of λ repressor in the bacterium may be controlled by the concentration of the *lac* inducer IPTG. The second operon (right) is carried on a prophage that contains the promoter p_{RM} fused to the *lacZ* gene. The level of β -galactosidase (β -galactosidase) in these cells therefore reflects the activity of p_{RM} . In similar experiments, the *cro* gene was substituted for λ *cl* and/or p_{RM} was replaced by p_R . [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 88, Cold Spring Harbor Laboratory Press (2004).]

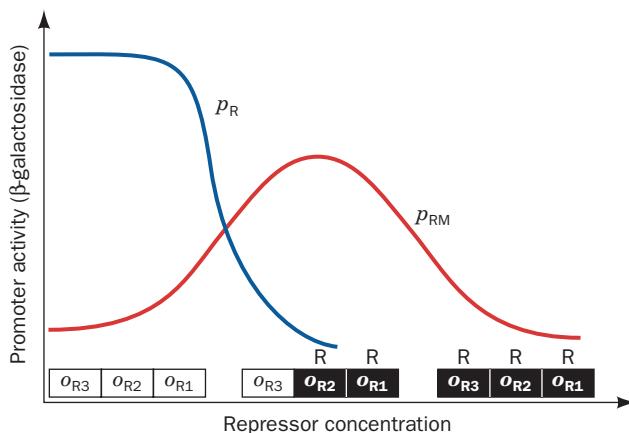


Figure 33-49 The response of p_{RM} and p_R to the λ repressor level. The p_{RM} curve was derived using the system diagrammed in Fig. 33-48, whereas the p_R curve was obtained using a similar system but with p_R rather than p_{RM} fused to *lacZ*. The amount of λ repressor that maximally stimulates p_{RM} is approximately that which occurs in a λ lysogen. At least 5-fold more repressor is required to half-maximally repress p_{RM} . The boxes indicate the states of each o_R subsite at the various repressor concentrations; black represents repressor occupancy. [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 89, Cold Spring Harbor Laboratory Press (2004).]

o_{R2} are occupied), transcription from p_R is indeed repressed, whereas that from p_{RM} is stimulated (Fig. 33-49). Transcription from p_{RM} only becomes repressed at high levels of λ repressor (when o_{R3} is also occupied). The stimulation of transcription from p_{RM} is abolished by mutations in o_{R2} that prevent repressor binding, whereas its repression at high repressor concentrations is relieved by mutations in o_{R3} . Thus, *occupancy of o_{R2} by λ repressor stimulates transcription from p_{RM} , whereas occupancy of o_{R3} prevents it by excluding RNA polymerase from p_{RM}* (Fig. 33-47c,d). By the same token, occupancy of o_{R1} and/or o_{R2} prevents transcription from p_R . In this way, λ repressor prevents the synthesis of all phage gene products but itself. Yet, at high repressor concentrations, its synthesis is also repressed, thereby maintaining the repressor concentration within reasonable limits.

What is the basis of λ repressor's remarkable property of inhibiting transcription from one promoter while stimulating it from another? Knowledge of the sizes and shapes of repressor and RNA polymerase, as well as their positions on the DNA as demonstrated by chemical protection experiments, indicate that repressor at o_{R2} and RNA polymerase at p_{RM} are in contact (Fig. 33-50). Evidently, *repressor stimulates RNA polymerase activity through their cooperative binding to DNA*. This model was corroborated by the analysis of repressor mutants that bind normally (or nearly so) to operators but fail to stimulate the binding of RNA polymerase: All of the mutated residues occur either in helix $\alpha 2$ or in the link connecting it to helix $\alpha 3$ and lie on the surface of the protein that is thought to face the RNA polymerase-binding site (Fig. 33-50).

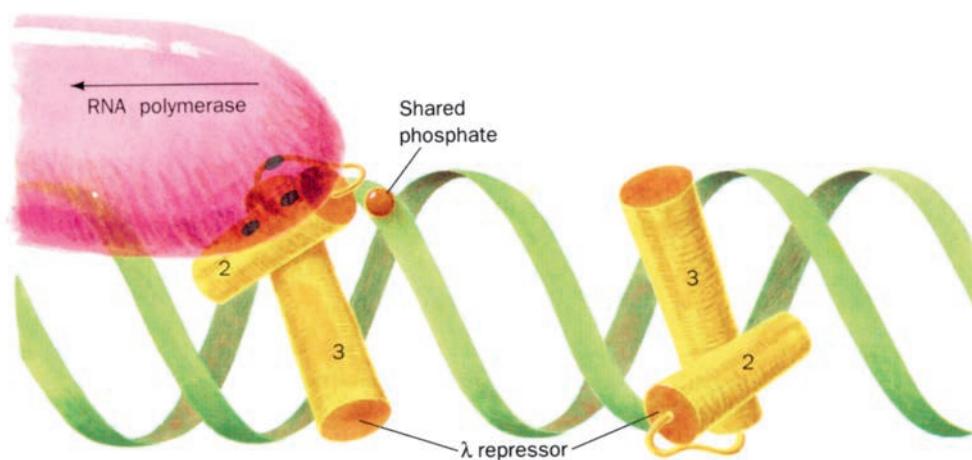


Figure 33-50 Interactions between the λ repressor and RNA polymerase when bound to o_{R2} . Repressor bound at o_{R2} is proposed to stimulate transcription at p_{RM} through a specific association with RNA polymerase that helps the polymerase bind to the promoter. This model is supported by the locations of the altered residues (blue dots) in three mutant repressors that

bind normally to o_{R2} but fail to stimulate transcription at p_{RM} . The relative positions of repressor and RNA polymerase are established by the location of a phosphate group (orange sphere) whose ethylation interferes with the binding of both proteins to the DNA. For the sake of clarity, only the α_2 - α_3 helix-turn-helix units of the repressor dimer are shown.

d. Cro Protein Binding to o_R Represses All λ Genes

Cro protein binds to the subsites of o_R in an order opposite to that of λ repressor (Fig. 33-51):

$$o_{R3} > o_{R2} \approx o_{R1}$$

This binding is noncooperative. Through experiments similar to that diagrammed in Fig. 33-48, but with *cro* in place of *cl*, the binding of Cro protein to o_{R3} was shown to abolish transcription from p_{RM} . Additional Cro binding to o_{R2} and/or o_{R1} turns off transcription from p_R .

e. The SOS Response Induces the RecA-Mediated Cleavage of λ Repressor

A final piece of information allows us to understand the workings of the λ switch. *The lytic phase is induced by agents that damage host DNA or inhibit its replication.* These are just the conditions that induce *E. coli*'s SOS response: The resulting fragments of single-stranded DNA activate RecA protein to stimulate the self-cleavage of LexA protein, the SOS gene repressor, at an Ala—Gly bond (Section 30-5Da). *Activated RecA protein likewise stimulates the autocatalytic cleavage of λ repressor monomer's Ala 111—Gly 112 bond, which occurs in the polypeptide segment linking the λ repressor's two domains.* The ability of λ repressor to cooperatively bind to o_{R2} is thereby abolished (Fig. 33-52a,b); the C-terminal domains can still dimerize but they no longer link the DNA-binding N-terminal domains. The consequent reduction in concentration of intact free monomers shifts the monomer–dimer equilibrium such that the operator-bound dimers dissociate to form monomers, which are then cleaved through the influence of activated RecA before they can rebind to their target DNA.

In the absence of repressor at o_R , the λ early genes, including *cro*, are transcribed (Fig. 33-52c). As Cro accumulates, it first binds to o_{R3} so as to block even basal levels of

λ repressor synthesis (Fig. 33-52d). Thus, *there being no mechanism for selectively inactivating Cro, the phage irreversibly enters the lytic mode:* The λ switch, once thrown, cannot be reset. The prophage is subsequently excised from the host chromosome by the integrase and excisionase that are produced in the delayed early phase (Fig. 33-31).

f. The λ Switch's Responsiveness to Conditions Arises from Cooperative Interactions among Its Components

The complexity of the above switch mechanism endows it with a sensitivity that is not possible in simpler systems. The degree of repression at p_R is a steep function of repressor concentration (Fig. 33-53, right): The repression of p_R in a lysogen is normally 99.7% complete but drops to half this level on inactivation of 90% of the repressor. This steep sigmoid binding curve arises from the much greater operator affinity of repressor dimers compared to monomers. This situation, in turn, results from the cooperative linking of the monomer–dimer equilibrium, the binding of dimer to operator, and the association of dimers bound at o_{R1} and o_{R2} to form a tetramer. In fact, this cooperative effect is further enhanced by the similar binding of repressor tetramer to o_{L1} and o_{L2} , which through DNA looping, forms an octamer with the repressor tetramer bound at o_{R1} and o_{R2} [a phenomenon that likewise increases the repression of p_L ; DNA looping, as we have seen, is also an important element in the control of expression of the *araBAD* operon (Section 31-3E) and the *lac* operon (Section 31-3Fa)]. In contrast, a 99.7% repressed promoter controlled by a stably oligomeric repressor binding to a single operator site requires 99% repressor inactivation for 50% expression (Fig. 33-53, left). *The cooperativity of λ repressor oligomerization and multiple operator site binding are therefore responsible for the remarkable responsiveness of the λ switch to the health of its host.*

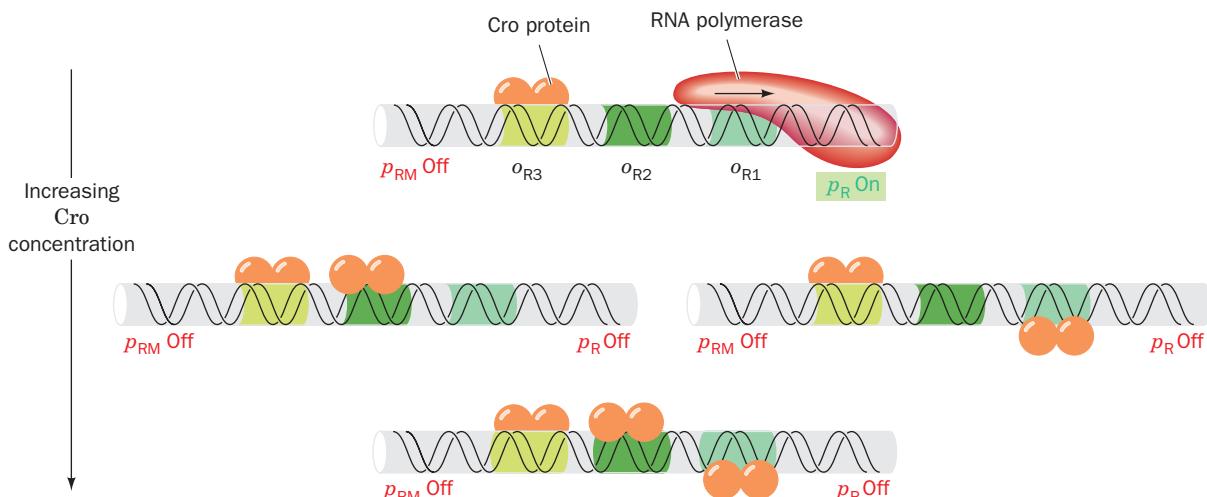
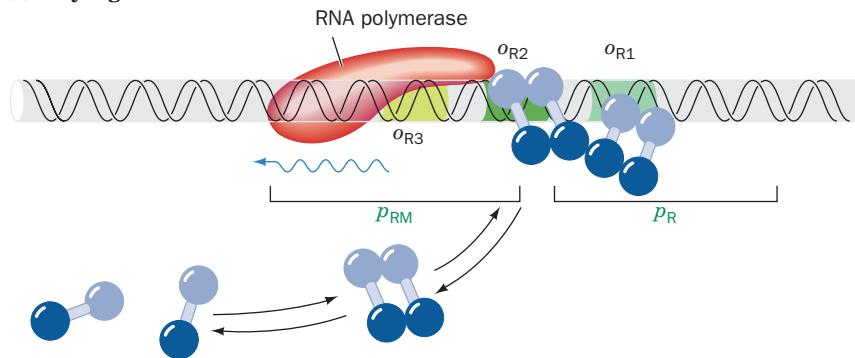


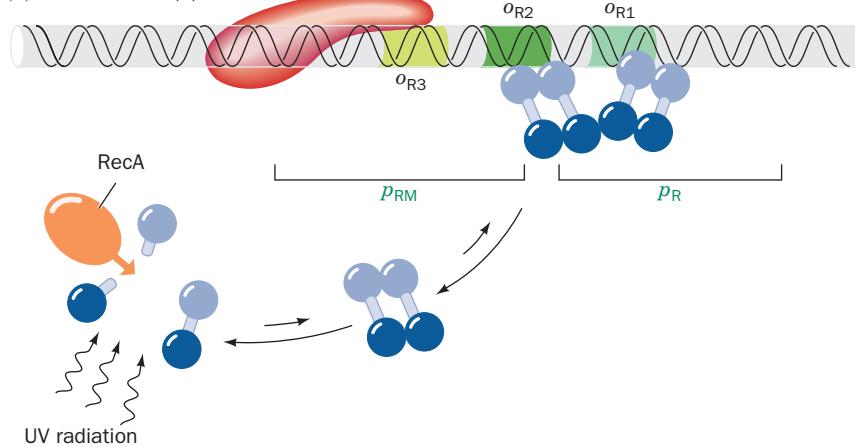
Figure 33-51 The binding of Cro protein to the three o_R subsites. o_{R3} binds Cro \sim 10 times more tightly than does o_{R1} or o_{R2} . Cro dimer therefore first binds to o_{R3} . A second dimer then binds to either o_{R1} or o_{R2} and in each case blocks transcription

from p_R . At high Cro concentrations, all three operator subsites are occupied. Compare this binding sequence with that of λ repressor (Fig. 33-47). [After Ptashne, M., *A Genetic Switch* (3rd ed.), p. 25, Cold Spring Harbor Laboratory Press (2004).]

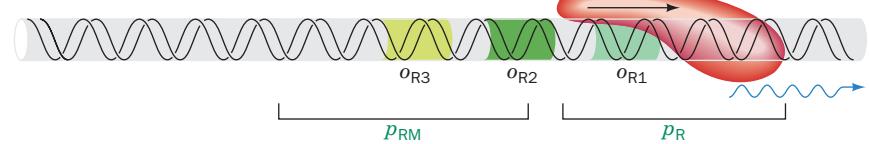
(a) Lysogenic mode



(b) Induction (1)



(c) Induction (2)



(d) Early lytic growth

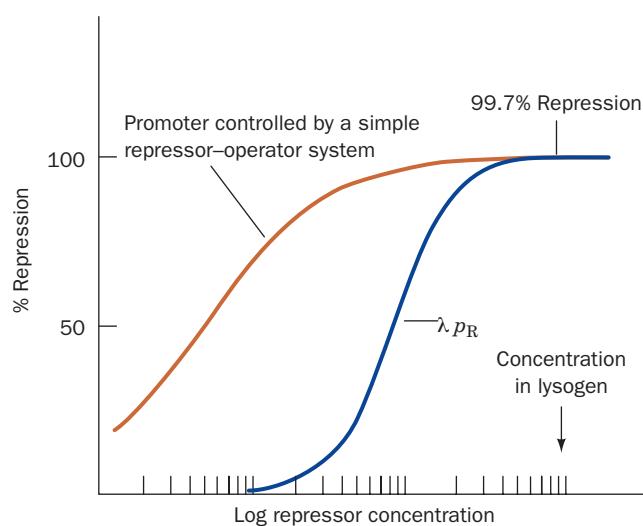
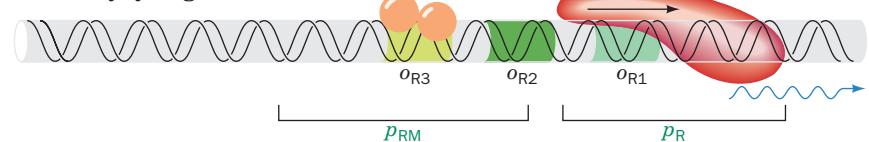


Figure 33-53 Theoretical repression curves for λ p_R (right) and a simple repressor-operator system such as that of the *lac* operon (left). Note the greater sensitivity of the λ system to a decrease in repressor concentration. [After Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K., and Ptashne, M., *Nature* **294**, 221 (1981).]

Figure 33-52 The λ switch. (a) In the lysogenic mode, two dimeric molecules of λ repressor cooperatively bind, through associations between their C-terminal domains, to o_{R1} and o_{R2} . This blocks host RNA polymerase from gaining access to p_R . However, the repressor bound to o_{R2} associates, through its N-terminal domain, with RNA polymerase at p_{RM} , thereby inducing the transcription of *cI*, the λ repressor gene, from this promoter (wiggly arrow). (b) Damage to the host DNA, as caused, for example, by UV radiation, activates host RecA protein to stimulate the self-cleavage of λ repressor monomers at a specific Ala-Gly bond between their two domains. (c) The consequent degradation of repressor monomers shifts the monomer-dimer equilibrium so as to free the o_{R1} and o_{R2} subsites for binding by RNA polymerase. This results in the transcription of the early genes (Fig. 33-31), including *cro*, from p_R (wiggly arrow). (d) The Cro protein thus synthesized preferentially binds at o_{R3} so as to block further transcription of *cI* from p_{RM} . Lytic growth is thereby irreversibly induced.

4 INFLUENZA VIRUS

Influenza is one of the few common infectious diseases that is poorly controlled by modern medicine. Its annual epidemics, one of which was recorded by Hippocrates in 412 B.C., are occasionally punctuated by devastating pandemics that infect 20 to 40% of the world's population. For example, the influenza pandemic of 1918, the so-called Spanish flu, which killed 40 to 50 million people worldwide (~3% of the world's population at the time) was the most lethal plague ever recorded (since it often killed previously healthy young adults, it lowered the average life expectancy in the United States by over 10 years). Since that time there have been four other pandemics of lesser severity, the so-called Asian flu of 1957, the Hong Kong flu of 1968, the Russian flu of 1977, and the swine flu of 2009 (the historical record suggests that there have been 13 pandemics in the past 400 years). All of these pandemics were characterized by the appearance of a new strain of influenza virus to which the human population had little resistance and against which previously existing influenza virus vaccines were ineffective. Moreover, between pandemics, influenza virus undergoes a gradual antigenic variation that degrades the level of immunological resistance against renewed infection. Even in nonpandemic years, influenza is responsible for the deaths of 250,000 to 500,000 mainly elderly people worldwide. In the United States, an average of 5% to 20% of the population contracts influenza each year, of which an average of ~40,000 people die. What characteristics of the influenza virus permit it to evade human immunological defenses? In this section we shall discuss this question and, in doing so, examine the structure and life cycle of the influenza virus.

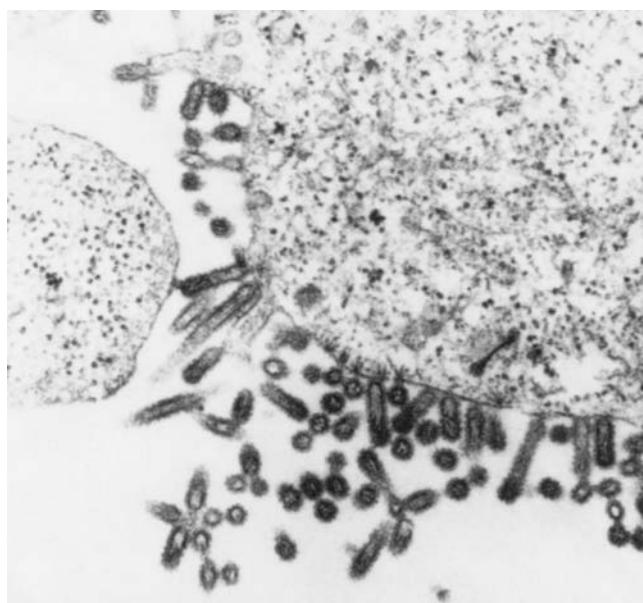


Figure 33-54 Electron micrograph of influenza viruses budding from infected chick embryo cells. [From Sanders, F.K., *The Growth of Viruses*, p. 15, Oxford University Press (1975).]

A. Virus Structure and Life Cycle

Electron micrographs of influenza virus (Fig. 33-54) reveal a collection of nonuniform spheroidal particles that range from 80 to 120 nm in diameter and whose surfaces are densely studded with radially projecting "spikes." The influenza virion, which grows by budding from the plasma membrane of an infected cell (Fig. 33-54), is an example of an **enveloped virus**. Its outer envelope consists of a lipid bilayer of cellular origin that is pierced by virally specified integral membrane glycoproteins, the "spikes." There are two types of these surface spikes (Fig. 33-55):

1. A rod-shaped spike composed of **hemagglutinin (HA)**, so named because it causes erythrocytes to agglutinate (clump together). HA mediates influenza target cell recognition by specifically binding to cell-surface receptors (glycophorin A molecules in erythrocytes; Section 12-3Aa) bearing terminal *N*-acetylneurameric acid (sialic acid; Fig. 11-11) residues. Each virion contains ~500 copies of HA.
2. A mushroom-shaped spike known as **neuraminidase (NA)**, which catalyzes the hydrolysis of the linkage joining a terminal sialic acid residue to a D-galactose or a D-galactosamine residue. NA probably facilitates the transport of the virus to and from the infection site by permitting its passage through mucin (mucus) and preventing viral self-aggregation. Each virion incorporates ~100 copies of NA.

In addition, the membranous outer envelope contains small amounts of **matrix protein 2 (M2)**, a proton channel that plays an important role in viral uncoating, assembly, and budding (see below).

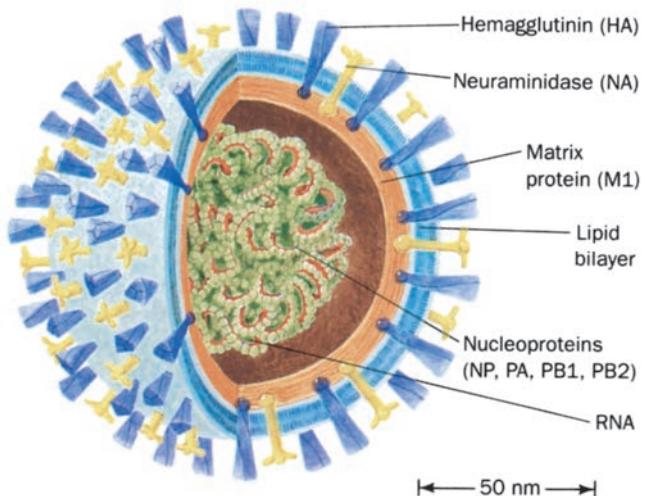


Figure 33-55 Cutaway diagram of the influenza virion. The HA and NA spikes are embedded in a lipid bilayer that forms the virion's outer envelope. M1 coats the underside of this membrane. The virion core contains the eight single-stranded RNA segments that comprise its genome in complex with the proteins NP, PA, PB1, and PB2 to form helical structures named nucleocapsids. [After Kaplan, M.M. and Webster, R.G., *Sci. Am.* 237(6): 91 (1977). Copyright © 1977 by Scientific American, Inc.]

Just beneath the viral membrane is a 6-nm-thick protein shell composed of ~3000 copies of **matrix protein 1 (M1)**, the virion's most abundant protein. M1 interacts with **nuclear export protein [NEP]**; formerly known as **nonstructural protein 2 (NS2)**.

The influenza virus genome is unusual in that it consists of eight different sized segments of single-stranded RNA. These RNA molecules are negative strands; that is, they are complementary to the viral mRNAs. In the viral core, these RNAs occur in complex with four different proteins: **nucleocapsid protein (NP)**, which occurs in ~1000 copies, and **polymerase acidic protein (PA)**, **polymerase basic protein 1 (PB1)**, and **polymerase basic protein 2 (PB2)**, which are present in 30 to 60 copies each. The resulting **nucleocapsids** have the appearance of twisted rods (10–15 nm in diameter and 30–130 nm long) that are folded back and coiled on themselves.

The eight viral RNAs, which vary in length from 890 to 2341 nucleotides, have all been sequenced. They encode the virus' nine structural proteins (HA, NA, M1, M2, NEP, NP, PA, PB1, and PB2) and **nonstructural protein 1 (NS1)**, which occurs only in infected cells. The sizes of the RNAs and the proteins they encode are listed in Table 33-2. About 10% of the viral mRNAs encoding M1 and NS1 are processed by the host cell splicing machinery to yield smaller mRNAs that respectively encode M2 and NEP but in mainly different reading frames from M1 and NS1.

a. Virus Life Cycle

The influenza infection of a susceptible cell begins with the HA-mediated adsorption of the virus to specific cell-surface receptors. The virus is then taken into the cell via endocytosis (Section 12-5Bc), whereupon the endocytic vesicle fuses with the endosome (Fig. 12-91). In the acidic (pH ~5) medium of the endosome, the viral M2 protein, a proton channel, admits protons into the virion, which induces the separation of the nucleocapsids from M1. The

Table 33-2 The Influenza Virus Genome

RNA Segment	Length (nt)	Polypeptide(s) Encoded
1	2341	PB2
2	2341	PB1
3	2233	PA
4	1778	HA
5	1565	NP
6	1413	NA
7	1027	M1, M2
8	890	NS1, NEP

Source: Lamb, R.A. and Choppin, P.W., *Annu. Rev. Biochem.* **52**, 473 (1983).

viral and endosome membranes then fuse through a mechanism discussed in Section 33-4C, thereby introducing the nucleocapsids into the cytosol. By ~20 min postinfection, in a process mediated by NP, the still intact nucleocapsids have been transported to the cell nucleus, where they commence transcription of the viral RNAs (**vRNAs**). Cellular enzyme systems are incapable of mediating such RNA-directed RNA synthesis. Rather, it is carried out by a viral RNA transcriptase system that consists of the nucleocapsid proteins.

The transcription of the influenza virus genome is terminated if infected cells are treated with inhibitors of RNA polymerase II (which synthesizes cellular mRNA precursors; Section 31-2E) such as actinomycin D or α -amanitin. Yet, none of these agents affects the viral transcriptase's *in vitro* activity. The resolution of this seeming paradox is that *in vivo* viral mRNA synthesis is primed by newly synthesized cellular mRNA fragments consisting of a 7-methyl-G cap (Section 31-4Aa) followed by a 9- to 17-nt chain ending in A or G (Fig. 33-56, top). Viral mRNAs, as do most mature cellular mRNAs, have poly(A) tails appended to

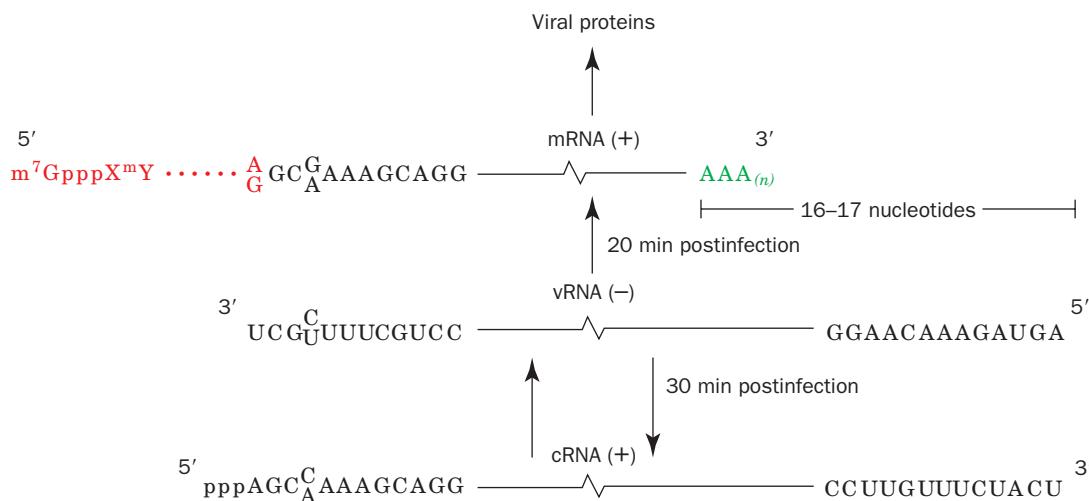


Figure 33-56 The biosynthesis of influenza vRNA, mRNA, and cRNA. The conserved nucleotides at the ends of the RNA segments are indicated. The viral mRNA's host-derived capped 5'

head and 3' poly(A) tail are shown in color. [After Lamb, R.A. and Choppin, P.W., *Annu. Rev. Biochem.* **52**, 490 (1983).]

their 3' ends by the cellular polyadenylation machinery (Section 31-4Ab).

The synthesis of the viral mRNAs is terminated 16 to 17 nt from the 5' ends of their vRNA templates by the presence of a sequence of 5 to 7 U's on the vRNAs. Consequently, the viral mRNAs cannot act as templates in vRNA replication. Rather, in an alternative transcription process that begins some 30 min postinfection, complete vRNA complements are synthesized. These so-called **cRNAs**, whose synthesis does not require a primer, begin with pppA at their 5' ends and lack poly(A) tails (Fig. 33-56, *bottom*). Hence cRNAs, unlike viral mRNAs, do not associate with ribosomes in infected cells. The synthesis of cRNAs, in contrast to that of viral mRNAs, requires the presence of NS1, which inhibits the processing of cellular pre-mRNAs and interferes with the synthesis of poly(A) tails. The cRNAs are the templates for vRNA synthesis. The resulting dsRNA would normally induce an interferon-mediated antiviral state in the infected cell (Section 32-4Ab). However, NS1 also functions as an interferon antagonist, thereby permitting viral proliferation.

The influenza virus transcription complex is a trimer consisting of PB1, PB2, and PA, which, together with NP, bind the RNA template. PB1 is the polymerase that catalyzes both the initiation and the elongation of the RNA transcript. PB2 binds to the 5' caps of cellular pre-mRNAs, although the endonuclease function that cleaves the capped primers from them appears to reside on PB1. Mutational experiments indicate that PA is required for vRNA but not mRNA synthesis, although its role in vRNA synthesis is poorly understood. The abundance of NP suggests that it has a structural role in the nucleocapsid, although it has also been implicated in the antitermination required to synthesize cRNAs rather than vRNAs.

The mechanism of influenza virus assembly is not well characterized. The viral spike glycoproteins, HA and NA, are ribosomally synthesized on the rough endoplasmic reticulum, further processed in the Golgi apparatus (Section 12-4C), and then transported, presumably in clathrin-coated vesicles, to areas of the plasma membrane containing lipid rafts (Section 12-3Cb). There, they aggregate in sufficient numbers to exclude host proteins (Fig. 33-57a,b). In the nucleus, the vRNAs combine with PA, PB1, and PB2 to form the nucleocapsids, which then interact with M1 protein. Nuclear export protein (NEP), as its name implies, mediates the export of nucleocapsids from the nucleus, which it does in partnership with both M1 and a cellular export factor. M1 then forms a nucleocapsid-enclosing shell that binds to HA and NA on the inside of the plasma membrane (Fig. 33-57b). This binding process causes the entire assembly to bud from the cell surface, thereby forming the mature virion (Fig. 33-57c). The complete infection cycle occupies ~8 to 12 h.

One of the mysteries of influenza virus assembly is how each virion acquires a complete set of the eight vRNAs. In mixed infections with various influenza strains, the reassortment of their genomic segments occurs with high frequency. This suggests that the nucleocapsids are randomly selected but that each virion contains sufficient numbers of

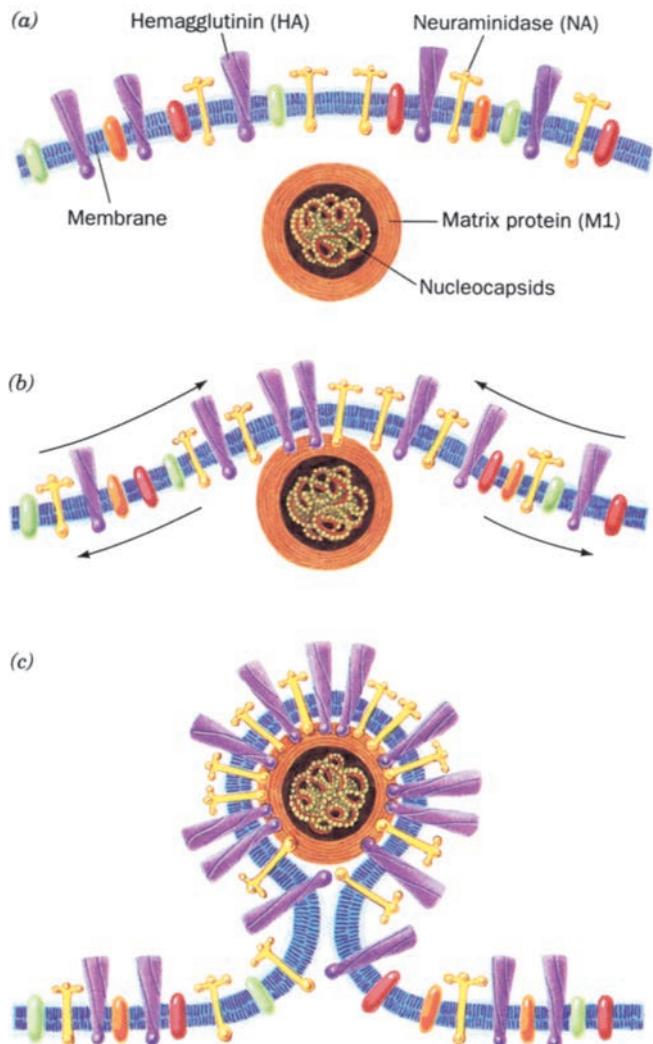


Figure 33-57 The budding of influenza virus from the host cell membrane. (a) The viral glycoproteins, HA and NA, are inserted into the plasma membrane of the host cell and the matrix protein, M1, forms the nucleocapsid-containing shell. (b) The binding of the matrix protein to the cytoplasmic domains of HA and NA results in the aggregation of these glycoproteins so as to exclude host cell membrane proteins (*arrows*). (c) This binding process induces the membrane to envelop the matrix protein shell such that the mature virion buds from the host cell surface. [After Wiley, D.C., Wilson, I.A., and Skehel, J.J., in Jurnak, F.A. and McPherson, A. (Eds.), *Biological Macromolecules and Assemblies*, Vol. 1: *Virus Structures*, Wiley (1984).]

vRNAs to ensure a reasonable probability that a given particle be infectious. This hypothesis is in agreement with the observation that aggregates of influenza virus have enhanced infectivity, a process that presumably occurs through the complementation of their vRNAs. However, electron microscopic studies by Yoshihiro Kawaoka reveal that the nucleocapsids in budding influenza A viruses have a distinct organization with seven parallel rods of different

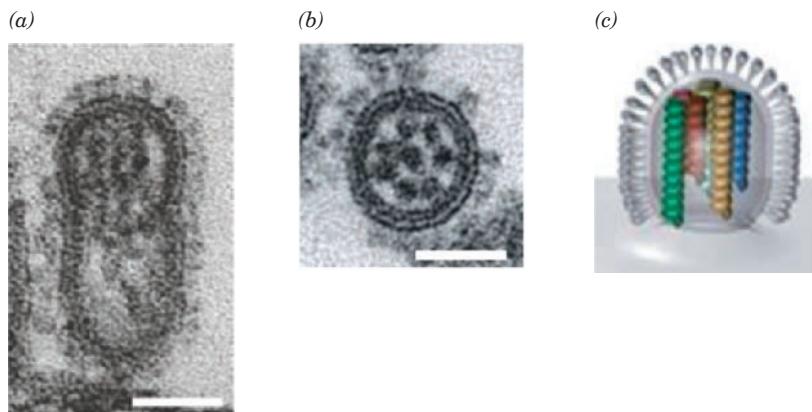


Figure 33-58 Nucleocapsids in budding influenza A virions.

(a) Electron micrograph showing the rodlike nucleocapsids of different lengths associated with the viral envelope at the distal end of the budding virion. (b) Electron micrograph of a transverse section through the nucleocapsid rods. Note that the central rod is surrounded by seven other rods. (c) Model of the

budding virion indicating its nucleocapsids, which are drawn in different colors. The cell surface from which the virion is budding is represented by the gray band. Scale bars 50 nm. [Parts a and b courtesy of Takeshi Noda and Part c courtesy of Yuko Kawaoka, University of Tokyo, Japan.]

lengths (maximally 130 nm) surrounding a single rod (Fig. 33-58) and with this assembly associated with and oriented perpendicularly to the budding viral tip. This supports a model in which the eight vRNAs are selected by an orderly process in which the nucleocapsids contain specific signals that cause them to be recruited as a complete set.

B. Mechanism of Antigenic Variation

Influenza viruses are classified into three immunological types, A, B, and C, depending on the antigenic properties of their differing nucleoproteins and matrix proteins. The A virus has caused all of the major pandemics in humans and has therefore been more extensively investigated than the B and C viruses. The B and C viruses infect mainly humans. However, the A virus infects a wide variety of mammalian and avian species in addition to humans (e.g., avian influenza is nearly 100% lethal to domestically raised chickens and turkeys). Indeed, it is thought that migratory birds (and, more recently, jet planes) are the major vectors that transport influenza A viruses around the world. The species specificity of a particular viral strain presumably arises from the binding specificity of its HA for cell-surface glycolipids.

a. HA Residue Changes Are Responsible for Most of the Antigenic Variation in Influenza Viruses

HA, being the influenza virus' major surface protein, is largely responsible for stimulating the production of the antibodies that neutralize the virus. Consequently, the different influenza virus subtypes arise mainly through the variation of HA. Antigenic variation in NA, the virus' other major surface protein, also occurs but this has lesser immunological consequences.

Two distinct mechanisms of antigenic variation have been observed in influenza A viruses:

1. Antigenic shift, in which the gene encoding one HA species is replaced by an entirely new one. This change may or may not be accompanied by a replacement of NA. It is thought that these new viral strains arise from the reassortment of genes among animal and human flu viruses. *Antigenic shift is responsible for influenza pandemics because the human population's immunity against previously existing viral strains is ineffective against the newly generated strain.* Evidently, these viruses had retained the (largely unknown) genetic traits responsible for their virulence in humans.

2. Antigenic drift, which occurs through a succession of point mutations in the HA gene, resulting in an accumulation of amino acid residue changes that attenuate the host's immunity. This process occurs in response to the selective pressure brought about by the buildup in the human population of immunity to the extant viral strains. HA varies in this manner by an average of 3.5 accepted amino acid changes per year.

Influenza A viruses are classified into subtypes according to the similarities of their HA and NA. There are 16 known subtypes of HA (H1 through H16) and 9 of NA (N1 through N9) that occur in mammals and birds. Avian virus subtypes occur in nearly all combinations, whereas only a few combinations have been found in humans. For example, human influenza A viruses circulating before 1957 were designated H1N1, those of the 1957 pandemic were H2N2, those of the 1968 pandemic were H3N2, and those of the 1977 pandemic were again H1N1 (and hence affected mainly young people who had not been exposed to

pre-1918 viruses) as were those of the 2009 pandemic. Since 1977, H3N2 and H1N1 viruses have been cocirculating.

Humans are rarely infected by avian flu viruses and such viruses do not appear to be readily transmitted between humans. Thus, the so-called bird flu, an H5N1 strain that has circulated since around 2000, infected mainly people who were in close contact with poultry, and while its infection was ~60% fatal, it was rarely transmitted between humans. However, phylogenetic studies indicate that pigs (swine) and birds can exchange influenza viruses as can pigs and humans. This suggests that pigs serve as “mixing vessels” for the creation of new pandemic flu viruses, which in turn, explains why Southeast Asia, where humans, pigs, and birds (ducks and chickens) often live in close proximity, is where most flu pandemics appear to have originated. Indeed, the 2009 H1N1 swine flu contained genes from human, pig, and bird strains of flu viruses (the first known triple hybrid virus), although epidemiological evidence indicates that it arose in Mexico.

b. HA Is an Elongated Trimeric Transmembrane Glycoprotein

HA plays a central role in both the viral infection process and in the immunological measures and countermeasures taken in the continuing biological contest between host and parasite. This has motivated considerable efforts to elucidate the structural basis of its properties. HA is a homotrimer of 550-residue subunits that is 19% carbohydrate by weight. The protein has three domains (Fig. 33-59):

1. A large hydrophilic, carbohydrate-bearing N-terminal domain that occupies the viral membrane’s external surface and that contains its sialic acid–binding site.
2. A hydrophobic 24- to 28-residue membrane-spanning domain that is located near the polypeptide’s C-terminus.
3. A hydrophilic C-terminal domain that occurs on the membrane’s inner side and that consists of the protein’s 10 C-terminal residues.

HA, which is synthesized as a single polypeptide designated HA0, is post-translationally cleaved by host-secreted proteases by the excision of Arg 329, thereby yielding two chains, HA1 and HA2, that are linked by a disulfide bond. This cleavage, which does not affect HA’s receptor-binding affinity, is required for the fusion of the virus with the host cell and therefore activates viral infectivity (see below). Indeed, the cleavability of HA is one of the major factors determining the virulence of influenza viruses.

HA can be removed from the virion by treatment with detergent but the resulting solubilized protein has not been made to crystallize. However, treatment of HA from a Hong Kong-type (H3) virus with the pineapple protease **bromelain**, which cleaves the polypeptide 9 residues before the membrane-spanning segment, yields a water-soluble protein named **BHA** that has been crystallized. X-ray analysis of these crystals by John Skehel and Don Wiley revealed an unusual structure (Fig. 33-60). The monomer consists of a long fibrous stalk extending from the membrane surface on which is perched a globular region. The fibrous stalk consists of segments from HA1 and HA2 and includes a remarkable 76-Å-long (53 residues in 14 turns) α helix. The globular region, which is comprised of only HA1 residues, contains an 8-stranded antiparallel β -sheet structure (a distorted jelly roll barrel; Section 8-3Bg) that forms the sialic acid–binding pocket.

The dominant interaction stabilizing BHA’s trimeric structure is a triple-stranded coiled coil consisting of the 76-Å α helices from each of its subunits (Fig. 33-60c). The BHA trimer is therefore an elongated molecule, some 135 Å in length, with a triangular cross section that varies in radius from 15 to 40 Å. The carbohydrate chains, which are attached to the protein via *N*-glycosidic linkages at each of its subunit’s seven Asn-X-Thr/Ser sequences (Section 11-3Ca), are located almost entirely along the trimer’s lateral surfaces. The role of the carbohydrates is unclear despite the fact that they cover some 20% of the protein’s surface. However, the observation that the mutational

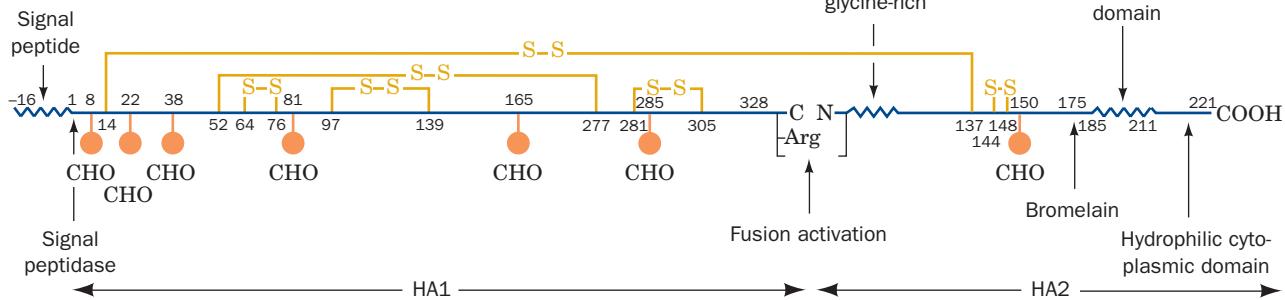


Figure 33-59 The primary structure of the 1968 Hong Kong influenza virus hemagglutinin. Its external domain (all of HA1 and HA2 through 185), its membrane anchoring domain (185–211 of HA2), and its cytoplasmic domain (212–221 of HA2) are indicated as are the positions of the signal peptide directing

the protein’s insertion into the membrane, the S–S bridges, the carbohydrate (CHO) attachment sites, the fusion activation site, and the bromelain cleavage site. [After Wilson, I.A., Skehel, J.J., and Wiley, D.C., *Nature* **289**, 367 (1981).]

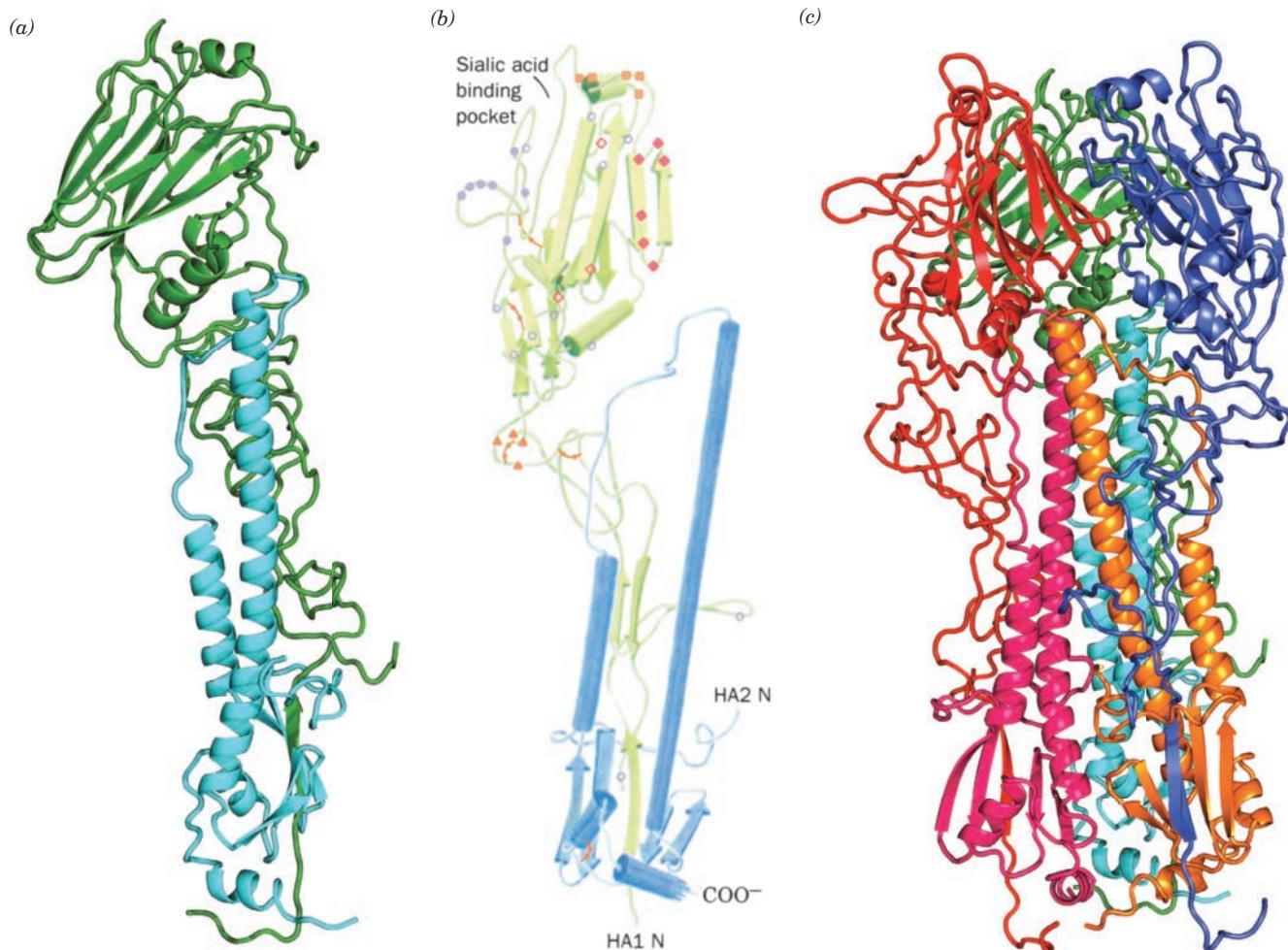


Figure 33-60 X-ray structure of influenza hemagglutinin.

(a) The polypeptide backbone of the monomer drawn in ribbon form. HA1 is green and HA2 is cyan. (b) A cartoon diagram of the monomer from a somewhat different point of view than Part *a* but similarly colored. The pairs of linked, small, filled circles represent disulfide groups. The positions of the mutant residues at the four antigenic sites are indicated by filled circles, squares, triangles, and diamonds. Open symbols represent antigenically

neutral residues. Note the position of the sialic acid-binding pocket. (c) A ribbon diagram of the HA trimer. Each HA1 and HA2 chain is drawn in a different color. The orientation of the green HA1 and cyan HA2 are the same as in Part *a*. [Based on an X-ray structure by John Skehel, National Institute for Medical Research, London, U.K., and Don Wiley, Harvard University. Part *b* after a drawing by Hidde Ploegh, in Wilson, I.A., Skehel, J.J., and Wiley, D.C., *Nature* **289**, 366 (1981). PDBid 4HMG.]

generation of a new oligosaccharide attachment site blocks antibody binding to HA suggests that carbohydrates modulate HA's antigenicity.

c. Antigenic Variation Results from Surface Residue Changes

HA's antigenic sites have been identified by mapping HA sequence changes on the protein's three-dimensional structure. The HA residues that mutated in an antigenically significant manner in Hong Kong-type viruses during the period 1968 to 1977 are indicated in Fig. 33-60b. These residues all occur on the protein's surface, often in polypeptide loops, where their mutational

variation affects the protein's surface character but apparently not its overall structure or stability. The variable residues are clustered in four sites surrounding HA's receptor-binding pocket, which is formed from amino acid residues that are largely conserved in numerous influenza virus strains. The strains responsible for the major flu epidemics between 1968 and 1975 had at least one mutation in each of these four antigenic sites. This degree of antigenic variation appears necessary to reinfect individuals previously infected with the same viral type. Evidently, antibodies directed against even conserved residues in HA's receptor-binding pocket, which would otherwise prevent HA from binding to its receptor, are dislodged by the

antigenic variation that so readily occurs about the rim of this binding pocket (we study antibody–antigen interactions in Section 35-2Bf).

d. NA Is a Tetrameric Transmembrane Glycoprotein

Influenza virus neuraminidase (NA) is a homotetrameric glycoprotein of 469-residue subunits. It has a

box-shaped globular head attached to a slender stalk that is anchored in the viral membrane. Digestion by a mixture of relatively nonspecific proteases from *Streptomyces griseus* known as **pronase** cleaves NA before residues 74 or 77, after the membrane attachment site, to yield an enzymatically active and crystallizable protein. The X-ray structure of this protein (Fig. 33-61), determined by Peter

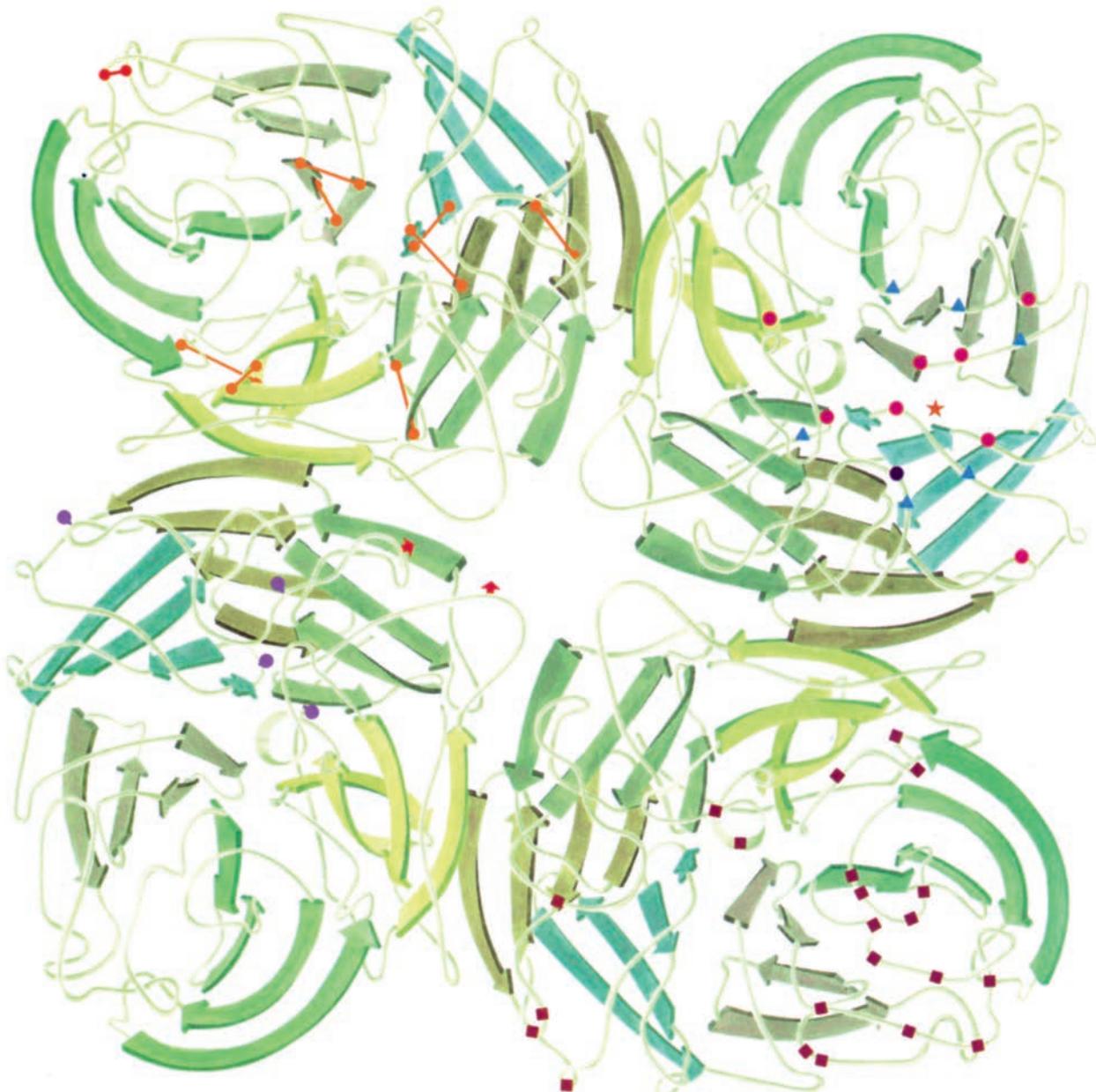


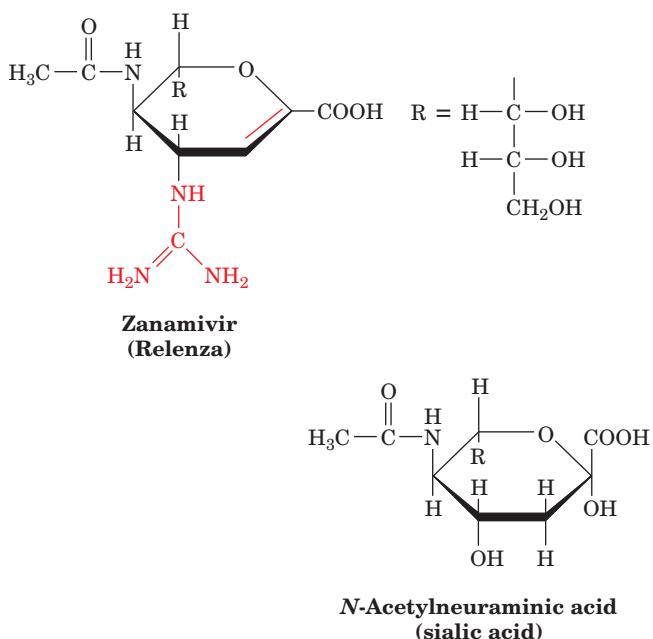
Figure 33-61 X-ray structure of the influenza neuraminidase tetramer. The view is along the 4-fold axis of the homotetrameric protein, looking toward the viral membrane. In each monomer unit, each of the six topologically equivalent 4-stranded antiparallel β sheets is differently shaded. The positions of the disulfide bonds are indicated in the upper left subunit. In the lower left subunit, the four carbohydrate attachment sites are indicated by filled purple circles and the Asp residues that ligand

Ca^{2+} ions are represented by red arrows. In the upper right monomer, the filled red circles and blue triangles, respectively, represent the conserved acidic and basic residues surrounding the enzyme's sialic acid-binding site, which is represented by a red star. In the lower right monomer, the positions of the mutated residues in NA's antigenic variants are flagged by filled brown squares. [After Varghese, J.N., Laver, W.G., and Colman, P.M., *Nature* **303**, 35 (1983). PDBid 1NN2.]

Colman and Graeme Laver, shows it to have 4-fold symmetry. Each subunit is composed of six topologically identical 4-stranded antiparallel β sheets arranged like the blades of a propeller. This so-called β propeller structurally resembles the 7-bladed β propellers of the clathrin heavy chain (Fig. 12-63b) and the heterotrimeric G protein G β subunit (Fig. 19-19b). Sugar residues are linked to NA at four of its five potential Asn-X-Ser/Thr N-glycosylation sites.

NA's sialic acid-binding site is located in a large pocket on the top of each monomer (star on the upper right subunit in Fig. 33-61). Its bound sialic acid residue interacts, through an extensive hydrogen bonding network, with 16 polar residues that are conserved in all known NA sequences (HA, in contrast, has only 2 polar residues in its sialic acid-binding site). Sequence changes in antigenic variants of NA occur in 7 chain segments that form a nearly continuous surface that encircles the catalytic site (squares on the lower right subunit in Fig. 33-61) in a manner similar to that of HA's receptor-binding site. Between 1968 and 1975, NA exhibited the same number of residue changes in its putative antigenic determinants as did HA. Antibodies against NA, nevertheless, do not neutralize infectivity. Rather, they restrict multiple cycles of viral replication and thus probably attenuate illness.

The realization that the structure of the NA catalytic site is strain-invariant together with the X-ray structure of NA in complex with sialic acid has led to the design of clinically effective inhibitors of NA (drug design is discussed in Section 15-4). For example, **zanamivir** (trade name **Relenza**),



a sialic acid mimic and a potent inhibitor of NA ($K_I = 0.1$ nM), is an effective antiviral agent, both in tissue culture and when delivered as an orally inhaled powder. Although the use of zanamivir does not prevent influenza infection, if

it is administered within 2 days of the onset of flulike symptoms, it significantly diminishes the length and severity of these symptoms and reduces the incidence of secondary bacterial infections.

C. Mechanism of Membrane Fusion

An influenza virus infection begins with the binding of HA to its cell-surface receptor. Then, as we discussed in Section 12-5Bc, the bound virus is taken into the cell via receptor-mediated endocytosis, a process that is accompanied by the fusion of the virus' enveloping membrane with that of the cell, thereby injecting the viral nucleocapsids into the cytosol. What is the mechanism of this membrane fusion?

Membrane fusion is mediated by HA but only after it has been exposed to the ~ 5.0 pH of the endosomal vesicle (Fig. 12-91). A variety of studies have implicated an ~ 25 -residue conserved hydrophobic segment at the N-terminus of HA2, the so-called fusion peptide, with mediating membrane fusion by inserting into the cellular membrane. Yet, BHA's X-ray structure indicates that the fusion peptide is buried in the protein's hydrophobic interior, ~ 100 Å from the receptor binding site at the "top" of protein, the region in closest proximity to the cellular membrane. Thus, HA must undergo an extensive conformational change before it can initiate membrane fusion.

At pH 5.0, BHA indeed undergoes a conformational change but one that causes it to aggregate in a manner unsuitable for crystallographic studies. However, the successive proteolytic digestion of BHA at pH 5.0 by trypsin and thermolysin (Table 7-2) yields a crystallizable protein fragment named **TBHA₂**, which consists of residues 1 to 27 of HA1 and residues 38 to 175 of HA2 that are linked by a disulfide bond.

The X-ray structure of TBHA₂, determined by Skehel and Wiley, reveals that this protein has dramatically refolded relative to BHA in a way that involves extensive changes in both its secondary and tertiary structural elements (Fig. 33-62). Thus, segments A and B at the N-terminus of TBHA₂ (Fig. 33-62c; the red and orange segments in Fig. 33-62) undergo a jackknife-like movement of ~ 100 Å in a way that extends the top of the long helix by ~ 10 helical turns toward the cell membrane [although the long helix is shortened from the bottom by a similar but not so extensive shift of segment D (green in Fig. 33-62) so as to partially replace the flipped out helix A]. This conformational change is irreversible (HA does not revert to its original form when the pH is raised) and hence has been described as occurring via a "spring-loaded" mechanism. The rearrangement of segments A and B had been predicted by Peter Kim who noted that segments A, B, and C have the heptad repeat characteristic of coiled coils (Section 8-2A). Such conformational shifts in intact HA translocate the fusion peptide (which would extend beyond the N-terminus of TBHA₂ at the top of segment A) by at least 100 Å from its position in BHA. There it could bridge the viral and cellular membranes so as to facilitate their fusion in a manner similar to that postulated for

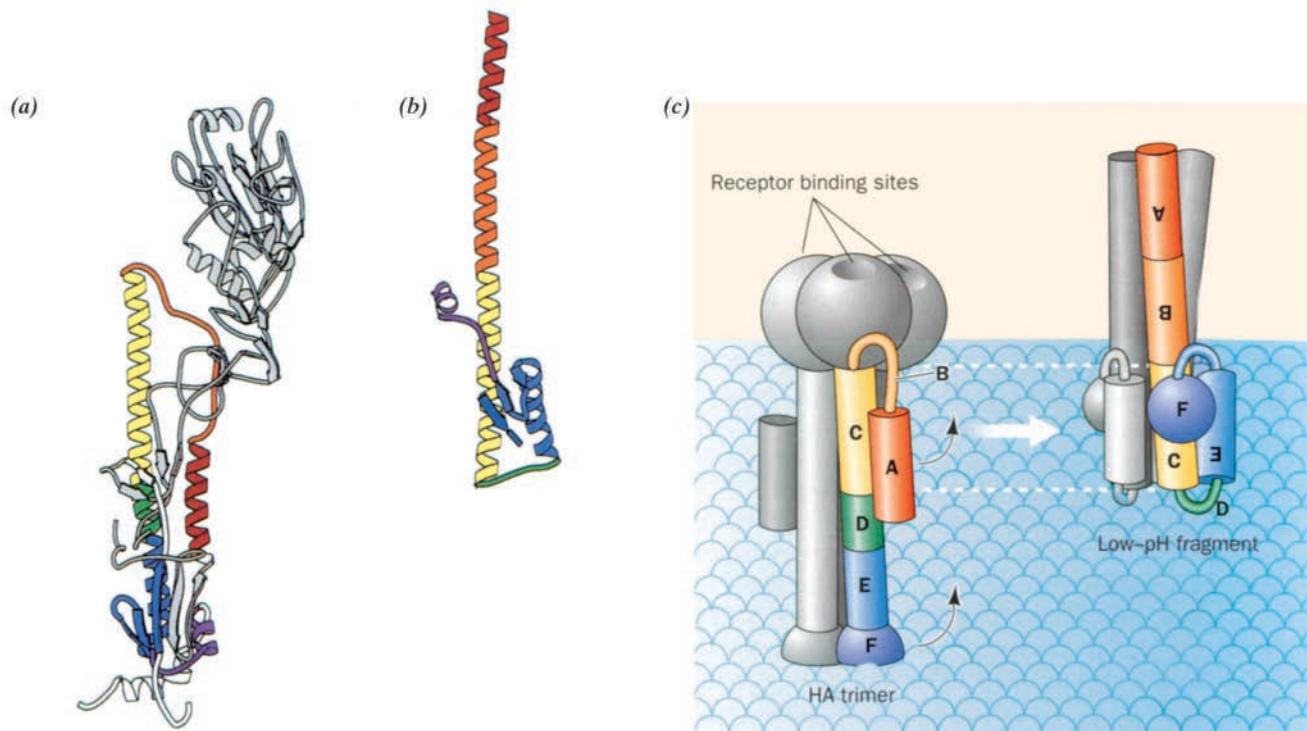


Figure 33-62 Comparison of the X-ray structures of BHA and TBHA₂. (a) Ribbon diagram of BHA in which the structural elements of the HA2 chain in TBHA₂ are colored in rainbow order (red to violet) from N- to C-terminus and the HA1 segment of TBHA₂, which is disulfide-linked to HA2, is blue. Regions of BHA that are proteolytically excised to form TBHA₂ are gray and those that are apparently disordered in TBHA₂ are white. (b) Ribbon diagram of TBHA₂ colored as in Part a. The heights of the various structural elements relative to the yellow helix segment, which is common to both BHA and TBHA₂, are

indicated. (c) Schematic diagram showing the positions and heights above the viral membrane surface of TBHA₂'s various structural elements in the HA trimer and in the low-pH fragment. The structural elements are colored as in Parts a and b. In the low-pH fragment, the fusion peptide (*not shown*) would protrude well above the receptor-binding heads where it would presumably insert itself into the cellular membrane. [Parts a and b courtesy of Don Wiley, Harvard University. PDBids 4HMG and 1HTM.]

SNARE complexes (Section 12-4De). The conformational change is probably triggered, at least in part, by the protonation at pH 5 of the six Asp side chains in the 19-residue

segment B, thereby reducing the charge–charge repulsions that apparently prevent this segment from forming a coiled coil at higher pH's.

CHAPTER SUMMARY

1 Tobacco Mosaic Virus Viruses are complex molecular aggregates that exhibit many attributes of living systems. Their structural and genetic properties have therefore served as valuable paradigms for the analogous cellular functions. The tobacco mosaic virus (TMV) virion consists of a helix of identical and therefore largely quasi-equivalent coat protein subunits enclosing a coaxially wound single strand of RNA. X-ray studies of TMV gels reveal that this RNA is bound, with three nucleotides per subunit, between the subunits of the protein helix. In the absence of TMV RNA, the subunits aggregate at high ionic strengths to form double-layered disks, and at low ionic strengths to form protohelices, which stack to form helical rods under acidic conditions. The virus' innermost polypeptide loop is disordered in both the disk and

the protohelix. Virus assembly is initiated when a protohelix (or possibly a double-layered disk) binds to the initiation sequence of TMV RNA, which is located ~1000 nucleotides from the RNA's 3' end. Interactions between the RNA and protohelix trigger the ordering of the disordered loop, thereby converting the protohelix to the helical form. Elongation of the virus particle then proceeds by the sequential addition of protohelices (disks) to the “top” of the assembly so as to pull the 5' end of the RNA up through the center of the growing viral helix.

2 Icosahedral Viruses Viral capsids are formed from one or a few types of coat protein subunits. These must be either helically arranged, as in TMV, or quasi-equivalently arranged in a polyhedral shell so as to enclose the viral nucleic acid. The

coat proteins of many icosahedral viruses are arranged in icosadeltahedra consisting of $60T$ subunits, where T is the triangulation number. The coat protein of tomato bushy stunt virus (TBSV) is arranged in a $T = 3$ icosadeltahedron so that TBSV subunits occupy three symmetrically distinct positions. The subunits must therefore associate through several sets of nonidentical intersubunit contacts. Some of the R domains form a structurally disordered inner protein shell. The viral RNA together with the remaining R domains is tightly packed in the space between the inner and outer protein shells. Other spherical plant viruses, southern bean mosaic virus (SBMV) and satellite tobacco mosaic virus (STMV), have tertiary and quaternary structures that are clearly related to those of TBSV. STMV's single-stranded RNA appears to form a series of stem and loop structures with the viral capsid. The structurally similar VP1, VP2, and VP3 coat proteins of poliovirus, rhinovirus, and foot-and-mouth disease virus (FMDV) are likewise icosahedrally arranged. However, the simian virus 40 (SV40) capsid consists of 72 pentagons of identical subunits in two different environments that are linked together by differing arrangements of their C-terminal arms in a non-icosadeltahedral arrangement. Although the coat proteins of most spherical viruses consist mainly of structurally similar 8-stranded antiparallel β barrels, that of bacteriophage MS2, a $T = 3$ virion, has an unrelated fold. Among the large icosahedral viruses of known structure are the bluetongue virus (BTV) core, which has a $T = 13$ outer shell and a $T = 2$ inner shell that envelops its largely visible dsRNA genome; and *Paramaecium bursaria Chlorella virus type 1* (PBCV-1), which has a $T = 169$ outer shell consisting of 1680 trimers of Vp54.

3 Bacteriophage λ Lytic growth of bacteriophage λ in *E. coli* is controlled by the sequential syntheses of antiterminators, which inhibit both rho-independent and rho-dependent transcriptional terminators. Thus gpN, which is synthesized in the early stage of growth, permits the synthesis of gpQ in the delayed early stage which, in turn, permits the synthesis of the capsid proteins in the late stage. Early gene transcription is repressed in the delayed early stage by Cro protein. DNA replication, which commences in the early stage, is mediated by the host DNA replication machinery with the aid of the phage proteins gpO and gpP. DNA synthesis initially occurs by both the θ and rolling circle (σ) modes but eventually switches entirely to the rolling circle mode.

The λ virion heads and tails are separately assembled. Head assembly is a complex process involving the participation of many phage gene products, not all of which are part of the mature virion. Phage heads are not self-assembling in that their formation is guided by host chaperonins and a viral scaffolding protein and requires several enzymatically catalyzed protein modification reactions. The mature phage head is a $T = 7$ icosadeltahedron of gpE, which is decorated by an equal number of gpD subunits. Just before the final stage of its assembly, the phage head is filled with a linear double strand of DNA in a process that is driven by ATP hydrolysis. The packaged DNA is wound in a spool that winds from outside to inside. Tail assembly occurs in a stepwise process from the tail fiber to the head-binding end. The body of the tail consists of a stack of hexameric rings of gpV. The completed heads and tails spontaneously join to form the mature virion.

Lysogeny is established by site-specific recombination between the phage *attP* site and the bacterial *attB* sites in a

process mediated by phage integrase (*gpint*) and integration host factor (IHF). Induction, in which this process is reversed, requires the additional action of phage excisionase (*gpxis*). Lysogeny is induced by a high level of *gpcII*, which stimulates the transcription of *int* and the λ repressor gene, *cI*. Repressor, as does Cro, binds to the o_L and o_R operators to shut down early gene transcription, including that of *cro* and *cII*. Each of these dimeric proteins, like other repressors of known structure, contains two symmetrically related helix-turn-helix (HTH) units that bind in successive turns of B-DNA's major groove. However, repressor, but not Cro, induces its own synthesis from the promoter *p_{RM}* by binding to o_{R2} so as to interact with RNA polymerase. The induction of repressor synthesis therefore throws the genetic switch that stably maintains the phage in the lysogenic state from generation to generation. Damage to host DNA, nevertheless, stimulates host RecA protein to mediate λ repressor cleavage so as to release repressor from o_L and o_R . This initiates the synthesis of early gene products, including *gpint* and *gpxis*, from *p_L* and *p_R* and thus triggers induction. If sufficient Cro protein is then synthesized to repress the synthesis of repressor, the phage becomes irrevocably committed to at least one generation of lytic growth. The tripartite character of o_R , the site of the λ switch, together with the cooperative nature of repressor binding to o_R and o_L , confers the λ switch with a remarkable sensitivity to the health of its host.

4 Influenza Virus The influenza virion's enveloping membrane is studded with protein spikes consisting of hemagglutinin (HA), which mediates host recognition, and neuraminidase (NA), which facilitates the passage of the virus to and from the infection site. Inside the membrane is a shell of matrix protein that contains the virus' genome of eight single-stranded RNAs, each in a separate protein complex known as a nucleocapsid. These vRNAs are templates for the transcription of mRNAs as catalyzed by the nucleocapsid proteins. This process is primed by host-derived 7-methyl-G-capped mRNA fragments. The viral mRNAs, which have poly(A) tails, lack the sequences complementary to the vRNA's 5' ends. The vRNAs, however, also act as templates for the transcription of the corresponding cRNAs which, in turn, are the templates for vRNA synthesis. The virus is assembled in and near the plasma membrane and forms by budding from the cell surface.

Influenza viruses infect a variety of mammals besides humans as well as many birds. Variation in the antigenic character of HA has been mainly responsible for the different influenza subtypes. Antigenic variation in HA occurs by either antigenic shift, in which the HA gene from an animal virus replaces that from a human virus, or antigenic drift, which occurs by a succession of point mutations in the HA gene. NA may vary in a similar fashion. HA is an elongated trimeric glycoprotein. Its surface has four antigenic sites that surround its sialic acid-binding pocket and that, in the viruses which caused the major epidemics between 1968 and 1975, all exhibit at least one mutational change. NA is a mushroom-shaped tetrameric glycoprotein. Its antigenic variations occur on a surface that also encircles its active site. HA mediates the fusion of the viral and host endosome membranes through a dramatic conformational change that translocates its fusion peptides to the vicinity of the endosome membrane into which they then insert.

REFERENCES

General

- Cann, A.J., *Principles of Modern Virology* (4th ed.), Elsevier Academic Press (2005).
- Carter, J.B. and Saunders, V.A., *Virology. Principles and Application*, Wiley (2007).
- Chiu, W., Burnett, R.M., and Garcea, R.L. (Eds.), *Structural Biology of Viruses*, Oxford University Press (1997).
- Dimmock, N.J., Easton, A.J., and Leppard, K.N., *Introduction to Modern Virology* (6th ed.), Blackwell Publishing (2007).
- Flint, S.J., Enquist, L.W., Racaniello, V.R., and Skalka, A.M., *Principles of Virology* (3rd ed.), Vol. 1. Molecular Biology, ASM Press (2009).
- Mahy, B.W.J. and van Regenmortel, H.M.V., *Desk Encyclopedia of General Virology*, Elsevier Academic Press (2010).
- Suttle, C.A., Viruses in the sea, *Nature* **437**, 356–361 (2005).

Tobacco Mosaic Virus

- Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., and Klug, A., Protein disk of tobacco mosaic virus at 2.8 Å showing the interactions within and between subunits, *Nature* **276**, 362–368 (1978).
- Butler, P.J.G., Self-assembly of tobacco mosaic virus: The role of an intermediate aggregate in generating both specificity and speed, *Philos. Trans. R. Soc. London* **B354**, 537–550 (1999).
- Butler, P.J.G., Bloomer, A.C., and Finch, J.T., Direct visualization of the structure of the “20 S” aggregate of coat protein of tobacco mosaic virus, *J. Mol. Biol.* **224**, 381–394 (1992). [Evidence indicating that the TMV coat protein double-layered disk predominates over the protohelix at pH 7.0.]
- Kendall, A., McDonald, M., and Stubbs, G., Precise determination of the helical repeat of tobacco mosaic virus, *Virology* **369**, 226–227 (2007).
- Klug, A., The tobacco mosaic virus particle: Structure and assembly, *Philos. Trans. R. Soc. London* **B354**, 531–535 (1999).
- Namba, K., Pattanayek, R., and Stubbs, G., Visualization of protein–nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å by X-ray fiber diffraction, *J. Mol. Biol.* **208**, 307–325 (1989).
- Raghavendra, K., Kelly, J.A., Khairallah, L., and Schuster, T.M., Structure and function of disk aggregates of the coat protein of tobacco mosaic virus, *Biochemistry* **27**, 7583–7588 (1988). [Evidence indicating that the TMV coat protein disks do not convert to the protohelices that nucleate TMV assembly.]
- Stubbs, G., Molecular structures of viruses from the tobacco mosaic group, *Semin. Virol.* **1**, 405–412 (1990).
- Stubbs, G., Tobacco mosaic virus particle structure and the initiation of disassembly, *Philos. Trans. R. Soc. London* **B354**, 551–557 (1999).

Icosahedral Viruses

- Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., and Tsukihara, T., Structure of southern bean mosaic virus at 2.8 Å resolution, *Nature* **286**, 33–39 (1980).
- Acharya, R., Fry, E., Stuart, D., Fox, G., and Brown, F., The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution, *Nature* **337**, 709–716 (1989).
- Arnold, E. and Rossmann, M.G., Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å, *J. Mol. Biol.* **211**, 763–801 (1990); and Rossmann, M.G., et al., Structure of a human common cold virus

and relationship to other picornaviruses, *Nature* **317**, 145–153 (1985).

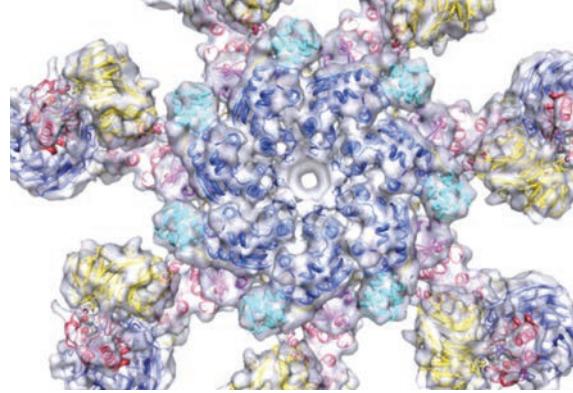
- Caspar, D.L.D. and Klug, A., Physical principles in the construction of regular viruses, *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24 (1962). [The classic paper formulating the geometric principles governing the construction of icosahedral viruses.]
- Dokland, T., Freedom and restraint: Themes in virus capsid assembly, *Structure* **8**, R157–R162 (2000).
- Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Ziéntara, S., Mertens, P.P.C., and Stuart, D.I., The atomic structure of the bluetongue virus core, *Nature* **395**, 470–478 (1998); Gouet, P., Diprose, J.M., Grimes, J.M., Malby, R., Burroughs, J.N., Ziéntara, S., Stuart, D.I., and Mertens, P.P.C., The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography, *Cell* **97**, 481–490 (1999); and Diprose, J.M., et al., Translocation portals for the substrates and products of a viral transcription complex: the bluetongue virus core, *EMBO J.* **20**, 7229–7239 (2001).
- Harrison, S.C., Common features in the structures of some icosahedral viruses: a partly historical view, *Semin. Virol.* **1**, 387–403 (1990).
- Harrison, S.C., The familiar and unexpected in structures of icosahedral viruses, *Curr. Opin. Struct. Biol.* **11**, 195–199 (2001).
- Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., and Bricogne, G., Tomato bushy stunt virus at 2.9 Å resolution, *Nature* **276**, 368–373 (1978). [The first report of a high-resolution virus structure.]
- Hogle, J.M., Chow, M., and Filman, D.J., The structure of poliovirus, *Sci. Am.* **256**(3), 42–49 (1987); and Three-dimensional structure of poliovirus at 2.9 Å resolution, *Science* **229**, 1358–1365 (1985).
- Hurst, C.J., Benton, W.H., and Enneking, J.M., Three dimensional model of human rhinovirus type 14, *Trends Biochem. Sci.* **12**, 460 (1987). [A “paper doll”-type cutout with accompanying assembly directions for constructing an icosahedral model of human rhinovirus. This useful learning device may also be taken as a $T = 3$ icosahedron.]
- Larson, S.B. and McPherson, A., Satellite tobacco mosaic virus RNA: structure and implications for assembly, *Curr. Opin. Struct. Biol.* **11**, 59–65 (2001); and Larson, S.B., Day, J., Greenwood, A., and McPherson, A., Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution, *J. Mol. Biol.* **277**, 37–59 (1998).
- Morgan, G.J., Historical review: Viruses, crystals, and geodesic domes, *Trends Biochem. Sci.* **28**, 86–90 (2003).
- Munshi, S., Liljas, L., Cavarelli, J., Bomu, W., McKinney, B., Reddy, V., and Johnson, J.E., The 2.8 Å structure of a $T = 4$ animal virus and its implications for membrane translocation of RNA, *J. Mol. Biol.* **261**, 1–10 (1996). [The X-ray structure of Nudaurelia Capensis ω virus.]
- Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., and Rossmann, M.G., The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus, *Proc. Natl. Acad. Sci.* **99**, 14758–14763 (2002); and Yan, X., Olson, N.H., Van Etten, J.L., Bergoin, M., Rossmann, M.G., and Baker, T.S., Structure and assembly of large lipid-containing dsDNA viruses, *Nature Struct. Biol.* **7**, 101–103 (2000). [The structure of PBCV-1.]
- Rossmann, M.G. and Johnson, J.E., Icosahedral RNA virus structure, *Annu. Rev. Biochem.* **58**, 533–573 (1989).
- Stehle, T., Gamblin, S.J., Yan, Y., and Harrison, S.C., The structure of simian virus 40 refined at 3.1 Å, *Structure* **4**, 165–182 (1996);

- and Liddington, R.C., Yan, Y., Mouli, J., Sahli, R., Benjamin, T.L., and Harrison, S.C., Structure of simian virus 40 at 3.8-Å resolution, *Nature* **354**, 278–284 (1991).
- Valegård, K., Liljas, L., Fridborg, K., and Unge, T., The three-dimensional structure of the bacterial virus MS2, *Nature* **345**, 36–41 (1990); and Golmohammadi, R., Valegård, K., Fridborg, K., and Liljas, L., The refined structure of bacteriophage MS2 at 2.8 Å resolution, *J. Mol. Biol.* **234**, 620–639 (1993).
- Bacteriophage λ**
- Albright, R.A. and Matthews, B.W., Crystal structure of λ -Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol.* **280**, 137–151 (1998); and Brennan, R.G., Roderick, S.L., Takeda, Y., and Matthews, B.W., Protein–DNA conformational changes in the crystal structure of λ Cro–operator complex, *Proc. Natl. Acad. Sci.* **87**, 8165–8169 (1990).
- Azaro, M.A. and Landy, A., λ Integrase and λ Int family, in Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (Eds.), *Mobile DNA II*, pp. 118–148, ASM Press (2002).
- Beamer, L.J. and Pabo, C.O., Refined 1.8 Å crystal structure of the λ repressor–operator complex, *J. Mol. Biol.* **227**, 177–196 (1992); and Jordan, S.R. and Pabo, C.O., Structure of the lambda complex at 2.5 Å resolution: details of the repressor–operator interactions, *Science* **242**, 893–899 (1988).
- Bell, C.E., Frescura, P., Hochschild, A., and Lewis, M., Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding, *Cell* **101**, 801–811 (2000).
- Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T., A structural basis for allosteric control of DNA recombination by λ integrase, *Nature* **435**, 1059–1066 (2009).
- Brüssow, H. and Hendrix, R.W., Phage genomics: Small is beautiful, *Cell* **108**, 13–16 (2002).
- Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A., and Patel, D.J., Solution structure of P22 transcriptional antitermination N peptide–box B RNA complex, *Nature Struct. Biol.* **5**, 203–212 (1998); and Legault, P., Li, J., Mogridge, J., Kay, L.E., and Greenblatt, J., NMR structure of the bacteriophage λ N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif, *Cell* **93**, 289–299 (1998).
- Cerritelli, M.E., Cheng, N., Rosenberg, A.H., McPherson, C.E., Booy, F.P., and Steven, A.C., Encapsidated conformation of bacteriophage T7 DNA, *Cell* **91**, 271–280 (1997).
- Dodd, J.B., Shearwin, K.E., and Egan, J.B., Revisited gene regulation in bacteriophage λ , *Curr. Opin. Genet. Devel.* **15**, 145–152 (2005).
- Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J.A., Duda, R.L., Hendrix, R.W., Komives, E.A., and Johnson, J.E., An unexpected twist in viral capsid maturation, *Nature* **458**, 646–650 (2009).
- Greenblatt, J., Nodwell, J.R., and Mason, S.W., Transcriptional antitermination, *Nature* **364**, 401–406 (1993).
- Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, Cold Spring Harbor Laboratory (1983). [A compendium of review articles on many aspects of bacteriophage λ .]
- Johnson, J.E., Virus particle maturation: insights into elegantly programmed nanomachines, *Curr. Opin. Struct. Biol.* **20**, 210–216 (2010).
- Morais, M.C., Koti, J.S., Bowman, V.D., Reyes-Aldrete, E., Anderson, D.L., and Rossmann, M.G., Defining molecular and domain boundaries in the bacteriophage ϕ 29 DNA packaging motor, *Structure* **16**, 1267–1274 (2008).
- Muriel, H., Bacteriophage lambda DNA maturation and packaging, *Annu. Rev. Biochem.* **60**, 125–153 (1991).
- Murray, N.E. and Gann, A., What has phage lambda ever done for us? *Curr. Biol.* **17**, R305–R312 (2007).
- Pabo, C.O. and Lewis, M., The operator-binding domain of λ repressor: structure and DNA recognition, *Nature* **298**, 443–447 (1982).
- Ptashne, M., *A Genetic Switch* (3rd ed.), Phage λ Revisited, Cold Spring Harbor Laboratory Press (2004). [An authoritative review of the λ switch.]
- Rao, V.B., and Feiss, M., The bacteriophage DNA packaging motor, *Annu. Rev. Genet.* **42**, 647–681 (2008).
- Roberts, J.W., RNA and protein elements of *E. coli* and λ transcription antitermination complexes, *Cell* **72**, 653–655 (1993).
- Simpson, A.A., et al., Structure of the bacteriophage ϕ 29 DNA packaging motor, *Nature* **409**, 745–750 (2000).
- Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M., Crystal structure of the λ repressor and a model for pairwise cooperative operator binding, *Nature* **452**, 1022–1025 (2008); and Hochschild, A. and Lewis, M., The bacteriophage λ CI finds an asymmetric solution, *Curr. Opin. Struct. Biol.* **19**, 79–86 (2009).
- Swinger, K.K. and Rice, P.A., IHF and HU: flexible architects of bent DNA, *Curr. Opin. Struct. Biol.* **14**, 28–35 (2004); and Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H., Crystal structure of an IHF–DNA complex: A protein-induced DNA turn, *Cell* **87**, 1295–1306 (1996).
- Taylor, K. and Wegrzyn, G., Replication of coliphage lambda DNA, *FEMS Microbiol. Rev.* **17**, 109–119 (1995).
- Influenza Virus**
- Air, G.M. and Laver, W.G., The molecular basis of antigenic variation in influenza virus, *Adv. Virus Res.* **31**, 53–102 (1986).
- Barry, J.M., *The Great Influenza*, Viking (2004). [A detailed historical account of the 1918 influenza pandemic and the hunt for its cause.]
- Bullough, P.A., Hughson, F.M., Skehel, J.J., and Wiley, D.C., Structure of influenza haemagglutinin at the pH of membrane fusion, *Nature* **371**, 37–43 (1994).
- Carr, C.M. and Kim, P.S., A spring-loaded mechanism for the conformational change of influenza hemagglutinin, *Cell* **73**, 823–832 (1994).
- Colman, P., Influenza virus neuraminidase: Structure, antibodies, and inhibitors, *Protein Sci.* **3**, 1687–1696 (1994).
- Colman, P.M., Neuraminidase inhibitors as antivirals, *Vaccine* **20**, S55–S58 (2002).
- Eckert, D.M. and Kim, P.S., Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem.* **70**, 777–810 (2001).
- Harris, A., Cardone, G., Winkler, D.C., Heymann, J.B., Brecher, M., White, J.M., and Steven, A.C., Influenza virus pleiomorphy characterized by cryoelectron tomography, *Proc. Natl. Acad. Sci.* **103**, 19123–19127 (2006).
- Kolata, G.B., *Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus that Caused It*, Farrar, Straus and Giroux (1999).
- Neumann, G., Noda, T., and Kawaoka, Y., Emergence and pandemic potential of swine-origin H1N1 influenza virus, *Nature* **459**, 931–939 (2009).
- Nicholson, K.G., Webster, R.G., and Hay, A.J., *Textbook of Influenza*, Blackwell Science (1998).
- Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H., and Kawaoka, Y., Architecture of ribonucleoprotein complexes in influenza A virus particles, *Nature* **439**, 490–492 (2006).
- Potter, C.W. (Ed.), *Influenza*, Elsevier (2002).

- Skehel, J.J. and Wiley, D.C., Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin, *Annu. Rev. Biochem.* **69**, 531–569 (2000).
- Skehel, J.J., Stevens, D.J., Daniels, R.S., Douglas, A.R., Knossow, M., Wilson, I.A., and Wiley, D.C., A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody, *Proc. Natl. Acad. Sci.* **81**, 1779–1783 (1984).
- Tumpey, T.M. and Belser, J.A., Resurrected pandemic influenza viruses, *Annu. Rev. Microbiol.* **63**, 79–98 (2009).
- Varghese, J.N., Laver, W.G., and Colman, P.M., Structure of the influenza virus glycoprotein antigen neuramimidase at 2.9 Å resolution, *Nature* **303**, 35–40 (1983); and Colman, P.M., Varghese, J.N., and Laver, W.G., Structure of the catalytic and antigenic sites in influenza virus neuramidase, *Nature* **303**, 41–44 (1983).
- Varghese, J.N., McKimm-Breschkin, J.L., Caldwell, J.B., Kortt, A.A., and Colman, P.M., The structure of the complex between influenza virus neuramidase and sialic acid, the viral receptor, *Proteins* **14**, 327–332 (1992).
- Webster, R.G., Laver, W.G., Air, G.M., and Schild, G.C., Molecular mechanisms of variation in influenza viruses, *Nature* **296**, 115–121 (1982).
- Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., and Wiley, D.C., Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid, *Nature* **333**, 426–431 (1988).
- Wilson, I.A., Skehel, J.J., and Wiley, D.C., Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, *Nature* **289**, 366–373 (1981); and Wiley, D.C., Wilson, I.A., and Skehel, J.J., Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and involvement in antigenic variation, *Nature* **289**, 373–378 (1981).

PROBLEMS

1. Why does a pH shift from 7 to 5 at low ionic strengths cause TMV double-layered disks to aggregate as helical rods?
2. Can a nucleic acid encode a monomeric protein large enough to enclose it? Explain.
3. Explain why the number of vertices in an icosadeltahedron always ends in the numeral “2” (e.g., 12 for $T = 1$).
4. Sketch a $T = 9$ icosadeltahedron.
5. The coat protein pentagons of SV40 are arranged at the vertices of a $T = 7$ icosadeltahedron. Yet the SV40 virion cannot have icosadeltahedral symmetry. Explain.
6. Why is it necessary to use conditionally lethal mutations in studying phage assembly rather than just lethal mutations?
7. Compare the volume contained by a λ phage head to that of λ DNA.
8. Virulent phages form clear plaques on a bacterial lawn, whereas bacteriophage λ forms turbid (cloudy) plaques. Explain.
9. What is the mutual consensus sequence of the o_L and o_R half-subsites?
10. λ repressor binds cooperatively to o_{R1} and o_{R2} but independently to o_{R3} . However, if o_{R1} is mutationally altered so that it does not bind repressor, then repressor binds cooperatively to o_{R2} and o_{R3} . Explain.
11. Bacteriophage 434 is a lambdoid phage that has both a repressor and a Cro protein. You have constructed a hybrid repressor that consists of the 434 repressor with its α 3 helix replaced by that from 434 Cro protein. Compare the pattern of contacts this hybrid protein makes with its operator, as indicated by chemical protection experiments, with those of the native 434 repressor and Cro proteins.
12. What is the probability that an influenza virion will have its proper complement of eight different RNAs if it has room for only eight nucleocapsids and binds them at random?



Eukaryotic Gene Expression

CHAPTER **34**

1 Chromosome Structure

- A. Histones
- B. Nucleosomes: The First Level of Chromatin Organization
- C. 30-nm Fibers: The Second Level of Chromatin Organization
- D. Radial Loops: The Third Level of Chromatin Organization
- E. Interphase Chromosomes

2 Genomic Organization

- A. The C-Value Paradox
- B. Repetitive Sequences
- C. Distribution of Genes
- D. Tandem Gene Clusters
- E. Gene Amplification
- F. Clustered Gene Families: Hemoglobin Gene Organization
- G. The Thalassemias: Genetic Disorders of Hemoglobin Synthesis

3 Control of Expression

- A. Chromosomal Activation and Deactivation
- B. Regulation of Transcriptional Initiation
- C. Other Expressional Control Mechanisms

4 Cell Differentiation and Growth

- A. Embryological Development
- B. The Molecular Basis of Development
- C. The Molecular Basis of Cancer
- D. The Regulation of the Cell Cycle
- E. Apoptosis: Programmed Cell Death

How does a fertilized ovum give rise to a highly differentiated multicellular organism? This question, of course, is just a sophisticated version of one that every child has asked: Where did I come from? Biologists began rational attempts to answer this question in the late nineteenth century and since that time have assembled an impressive body of knowledge concerning the general patterns of cellular differentiation and organismal development. Yet we have had the technical ability to study embryogenesis on the molecular level only in the last 40 years or so.

In order to understand cellular differentiation we must first understand the workings of the eukaryotic cell. Eukaryotic cells are, for the most part, much larger and far more complex than prokaryotic cells (Section 1-2). However, *the basic difference between these two types of cells is that eukaryotes have a nuclear membrane that separates their chromosomes from their cytoplasm, thereby physically divorcing the eukaryotic transcriptional process from that of*

translation.

In contrast, the prokaryotic chromosome is embedded in the cytosol so that the initiation of protein synthesis often occurs on mRNAs that are still being transcribed. The transcriptional and translational control processes in eukaryotes are consequently fundamentally different from those of prokaryotes. This situation is reflected in both the packaging and the genetic organization of eukaryotic DNA in comparison with that of prokaryotes. We therefore begin this chapter with a physical description of the eukaryotic chromosome. We then consider how the eukaryotic genome is organized and how it is expressed. Finally, we discuss cell differentiation, its aberration, cancer, how the cell cycle is controlled, and programmed cell death. In all these subjects, as we shall see, our knowledge is quite fragmentary. Eukaryotic molecular biology is under such intense scrutiny, however, that significant advances in its understanding are made almost daily. Thus, perhaps more so than for other subject matter considered in this textbook, it is important that the reader supplement the material in this chapter with that in the recent biochemical literature.

1 CHROMOSOME STRUCTURE

Eukaryotic chromosomes, which consist of a complex of DNA, RNA, and protein called **chromatin**, are dynamic entities whose appearance varies dramatically with the stage of the cell cycle. The individual chromosomes assume their familiar condensed forms (Figs. 1-18 and 34-1) only during cell division (M phase of the cell cycle; Section 30-4A). During interphase, the remainder of the cell cycle, when the chromosomal DNA is transcribed and replicated, the chromosomes of most cells become so highly dispersed that they cannot be individually distinguished (Fig. 34-2). Cytologists have long recognized that there are two types of this dispersed chromatin: a less densely packed variety named **euchromatin** and a more densely packed variety termed **heterochromatin** (Fig. 34-2). These two types of chromatin differ, as we shall see, in that euchromatin is genetically expressed, whereas heterochromatin is not expressed.

The 46 chromosomes in a human cell each contain a single DNA molecule of between 44 and 246 million bp (Section 5-3D), so that at 3.4 Å/bp, they have contour lengths



Figure 34-1 Scanning electron micrograph of a metaphase chromosome. It consists of two sister (identical) chromatids joined at their centromere (the constricted region near the center of a chromosome through which it attaches to the mitotic spindle; Section 1-4Aa). Chromosomes constitute a cell's largest molecular entities. [Andrew Syred/Photo Researchers, Inc.]

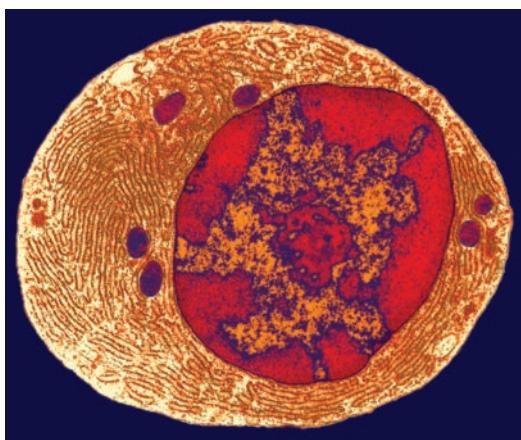


Figure 34-2 Electron micrograph of a B lymphocyte. The large body in the center of the cell is the nucleus in which heterochromatin (red) adheres to the inner nuclear membrane and the transcriptionally active euchromatin (orange) is more centrally located. The small red body at the center of the nucleus is the nucleolus. Also note the extensive rough endoplasmic reticulum filling the cytoplasm of this antibody-producing cell (Section 35-2A). Compare this figure with Fig. 1-5. [Don Fawcett/Visuals Unlimited.]

between 1.5 and 8.4 cm. Yet in metaphase, their most condensed state (Fig. 34-1), these chromosomes range in length from 1.3 to 10 μm . *The DNA in metaphase chromosomes therefore has a packing ratio (ratio of its contour length to the length of its container) of >8000.* How does the DNA attain such a high degree of condensation? Structural studies have revealed that this results from several levels of folding. We discuss these levels below, starting with the lowest level. We begin, however, by studying the proteins responsible for much of this folding.

A. Histones

The protein components of chromatin, which comprise somewhat more than half its mass, consist mostly of **histones**, which were discovered in 1884 by Albrecht Kossel and for many years were believed to be the genetic material itself. There are five major classes of these proteins, **histones H1, H2A, H2B, H3, and H4**, all of which have a large proportion of positively charged residues (Arg and Lys; Table 34-1). These proteins therefore ionically bind DNA's negatively charged phosphate groups. Indeed, histones may be extracted from chromatin by 0.5M NaCl, a salt solution of sufficient concentration to interfere with these electrostatic interactions. Histones, in contrast to most other cellular proteins, are almost entirely synthesized during the relatively short S phase of the cell cycle (when DNA is replicated), when they are needed in massive amounts for chromatin replication.

a. Histones Are Evolutionarily Conserved

The amino acid sequences of histones H2A, H2B, H3, and H4 (Table 34-1) have remarkably high evolutionary stability. For example, histones H4 from cows and peas, species that diverged 1.2 billion years ago, differ by only two conservative residue changes (Fig. 34-3), which makes histone H4, the most invariant histone, among the most evolutionarily conserved proteins known (Section 7-3Bc). *Such rigid evolutionary stability implies that the above four histones have critical functions to which their structures are so well tuned that they are all but intolerant to change.* The fifth histone, histone H1, is more variable than the other histones; we shall see below that its role differs from that of the other histones.

b. Histones May Be Modified and Have Variant Forms

Histones are subject to post-translational modifications that include acetylations, methylations, phosphorylations, ubiquitination (Section 32-6B1), sumoylation (Section 32-6Bm), and ADP-ribosylation (Section 32-3Ge) of specific Arg, Glu, His, Lys, Ser, Thr, and Tyr side chains. Most of these modifications, many of which are reversible, decrease the histones' positive charges, thereby significantly altering histone–DNA interactions. Yet, despite the histones' great evolutionary stability, their degree of modification varies enormously with the species, tissue, and the stage of the cell cycle.

Table 34-1 Calf Thymus Histones

Histone	Number of Residues	Mass (kD)	% Arg	% Lys	UEP ^a (×10 ⁻⁶ year)
H1	215	23.0	1	29	8
H2A	129	14.0	9	11	60
H2B	125	13.8	6	16	60
H3	135	15.3	13	10	330
H4	102	11.3	14	11	600

^aUnit evolutionary period: The time for a protein's amino acid sequence to change by 1% after two species have diverged (Section 7-3Bc).

Ac—Ser—Gly—Arg—Gly—Lys—Gly—Gly—Lys—Gly—Leu—10
 Gly—Lys—Gly—Gly—Ala—Lys—Arg—His—Arg—Lys—20
 Val—Leu—Arg—Asp—Asn—Ile—Gln—Gly—Ile—Thr—30
Lys—Pro—Ala—Ile—Arg—Arg—Leu—Ala—Arg—Arg—40
 Gly—Gly—Val—Lys—Arg—Ile—Ser—Gly—Leu—Ile—50
 Tyr—Glu—Glu—Thr—Arg—Gly—Val—Leu—Lys—Val—60
 Phe—Leu—Glu—Asn—Val—Ile—Arg—Asp—Ala—Val—70
 Thr—Tyr—Thr—Glu—His—Ala—Lys—Arg—Lys—Thr—80
 Val—Thr—Ala—Met—Asp—Val—Val—Tyr—Ala—Leu—90
Lys—Arg—Gln—Gly—Arg—Thr—Leu—Tyr—Gly—Phe—100
 Gly—Gly 102

Figure 34-3 The amino acid sequence of calf thymus

histone H4. This 102-residue protein's 25 Arg and Lys residues are indicated in red. Pea seedling H4 differs from that of calf thymus by conservative changes at the two shaded residues: Val 60 → Ile and Lys 77 → Arg. The underlined residues are subject to post-translational modification: Ser 1 is invariably *N*-acetylated and may also be *O*-phosphorylated; Lys residues 5, 8, 12, and 16 may be *N*-acetylated; and Lys 20 may be mono- or di-*N*-methylated. [After DeLange, R.J., Fambrough, D.M., Smith, E.L., and Bonner, J., *J. Biol. Chem.* **244**, 5678 (1969).]

Most, if not all, eukaryotes have numerous genetically distinct subtypes of the canonical histones. These variant histones have important roles in such essential cellular processes as transcriptional initiation and termination, DNA repair, homologous recombination, and telomere and centromere function. For example, **H2A.Z** (64% identical to H2A) is associated with the promoters of actively transcribed genes, **H2A.X** (95% identical to H2A) binds to DNA at double-strand breaks, thereby marking them for repair, and the H3-like **CenH3** (also called **CENP-A**; 50% identical to H3) functions in organizing chromatin structure at the centromere. In fact, the erythroid cells of chick embryos contain a histone H1 variant that differs so greatly from H1 that it is named **histone H5** (38% identical to chicken H1; avian erythrocytes, unlike those of mammals, have nuclei). Variant histones are synthesized independent of DNA replication and many are only synthesized during specific stages of embryonic development and in the differentiation of certain cell types. How they carry out their various specialized functions is, for the most part, unknown.

B. Nucleosomes: The First Level of Chromatin Organization

The first level of chromatin organization was pointed out by Roger Kornberg in 1974 through the synthesis of several lines of evidence:

1. Chromatin contains roughly equal numbers of molecules of histones H2A, H2B, H3, and H4, and no more than half that number of histone H1 molecules.
2. X-ray diffraction studies indicate that chromatin fibers have a regular structure that repeats about every

10 nm along the fiber direction. This same X-ray pattern is observed when purified DNA is mixed with equimolar amounts of all the histones except histone H1.

3. Electron micrographs of chromatin (Fig. 34-4) reveal that it consists of ~10-nm-diameter particles connected by thin strands of apparently naked DNA, rather like beads on a string. These particles are presumably responsible for the foregoing X-ray pattern.

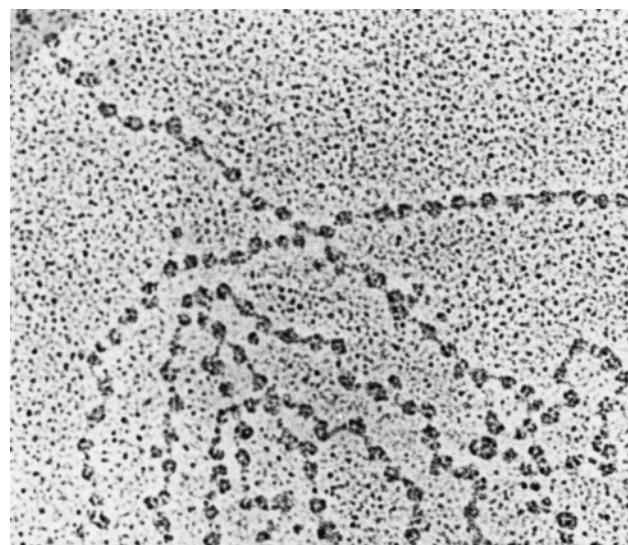


Figure 34-4 Electron micrograph of *D. melanogaster* chromatin showing that its 10-nm fibers are strings of closely spaced nucleosomes. [Courtesy of Oscar L. Miller, Jr., University of Virginia.]

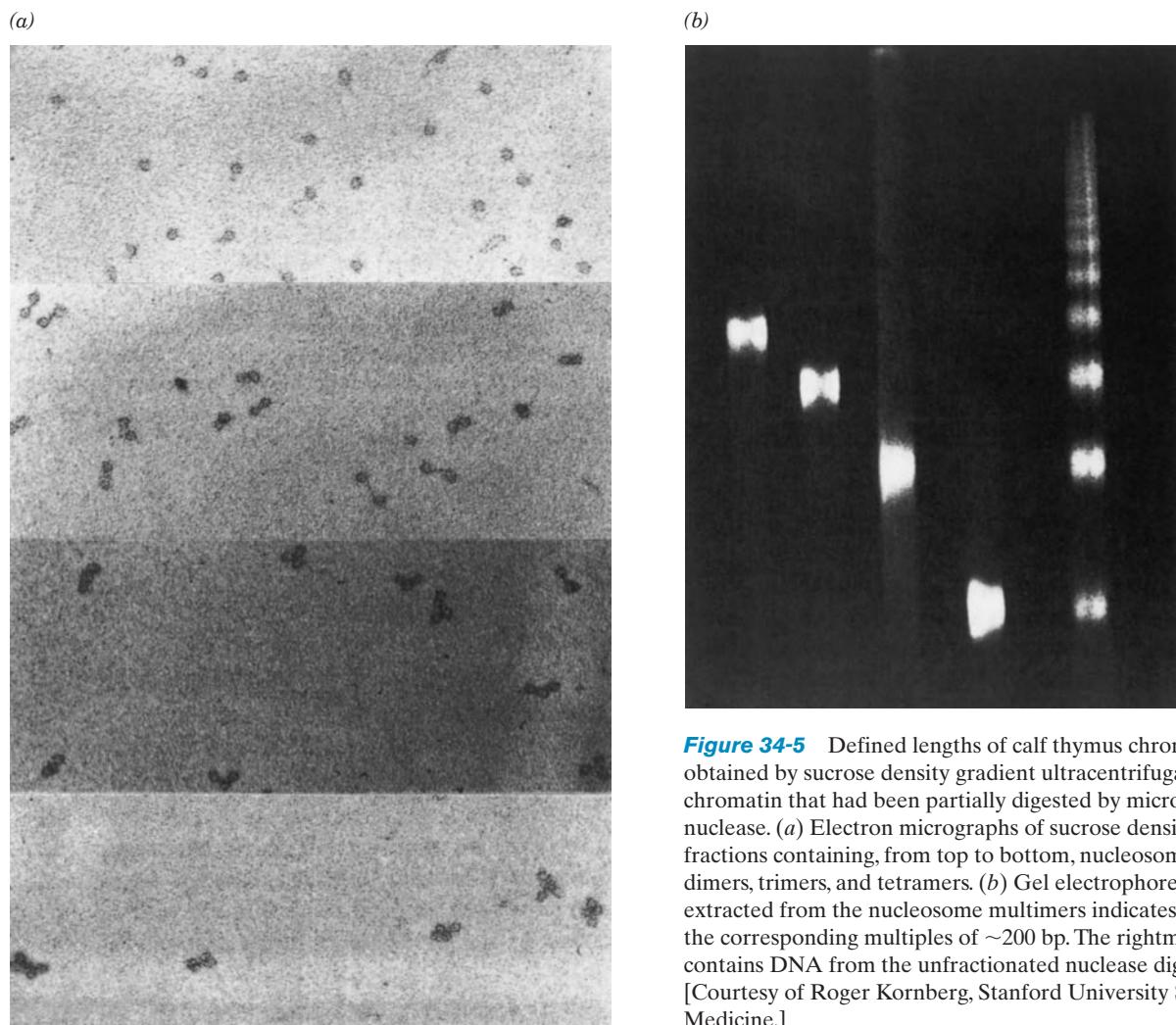


Figure 34-5 Defined lengths of calf thymus chromatin obtained by sucrose density gradient ultracentrifugation of chromatin that had been partially digested by micrococcal nuclease. (a) Electron micrographs of sucrose density gradient fractions containing, from top to bottom, nucleosome monomers, dimers, trimers, and tetramers. (b) Gel electrophoresis of DNA extracted from the nucleosome multimers indicates that they are the corresponding multiples of ~ 200 bp. The rightmost lane contains DNA from the unfractionated nuclease digest. [Courtesy of Roger Kornberg, Stanford University School of Medicine.]

4. Brief digestion of chromatin by **micrococcal nuclease** (which cleaves double-stranded DNA) cleaves the DNA between some of the above particles (Fig. 34-5a); apparently the particles protect the DNA closely associated with them from nuclease digestion. Gel electrophoresis indicates that each particle n -mer contains $\sim 200n$ bp of DNA (Fig. 34-5b).

5. Chemical cross-linking experiments, such as are described in Section 8-5Ca, indicate that histones H3 and H4 associate to form the tetramer $(H3)_2(H4)_2$ (Fig. 34-6).

These observations led Kornberg to propose that *the chromatin particles, which are called nucleosomes, consist of the octamer $(H2A)_2(H2B)_2(H3)_2(H4)_2$ in association with ~ 200 bp of DNA*. The fifth histone, H1, was postulated to be associated in some manner with the outside of the nucleosome (see below).

a. DNA Coils around a Histone Octamer to Form the Nucleosome Core Particle

Micrococcal nuclease, as described above, initially degrades chromatin to particles known as **chromatosomes** that each consist of 167 bp of DNA in complex with a histone octamer and one molecule of histone H1. On further

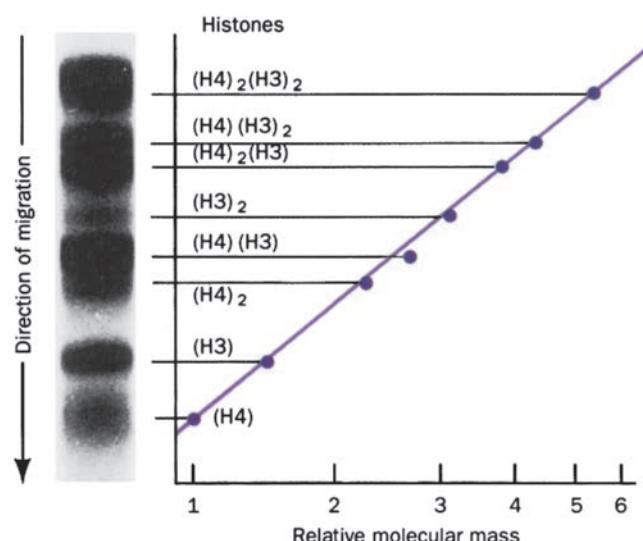
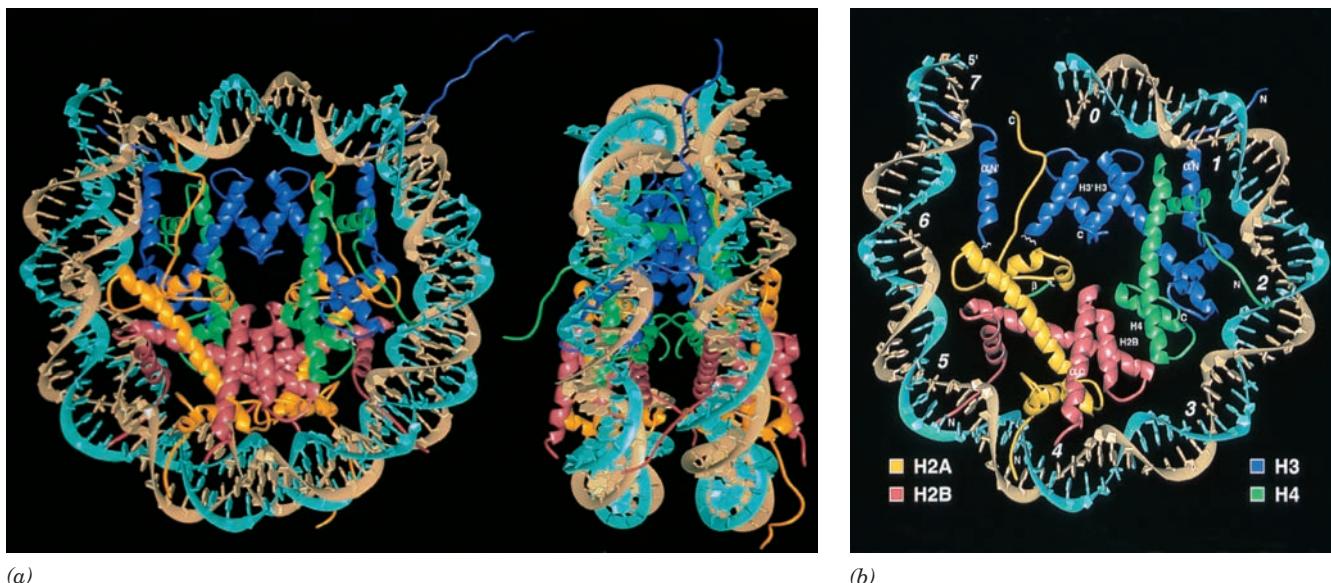


Figure 34-6 SDS-gel electrophoresis of a mixture of calf thymus histones H3 and H4 that had been cross-linked by dimethylsuberimidate. The electrophoretogram contains all the bands expected from an $(H3)_2(H4)_2$ tetramer. [Courtesy of Roger Kornberg, Stanford University School of Medicine.]



(a)

(b)

Figure 34-7 X-ray structure of the nucleosome core particle. (a) The entire core particle as viewed (left) along its superhelical axis and (right) rotated 90° about the vertical axis. The proteins of the histone octamer are drawn in ribbon form with H2A yellow, H2B red, H3 blue, and H4 green. The sugar-phosphate backbones of the 146-bp DNA are drawn as tan and cyan ribbons whose attached bases are represented by polygons of the same color. In both views, the pseudo-2-fold axis is vertical and passes through the DNA center at the top. (b) The top half of the

nucleosome core particle as viewed in Part a, left, and identically colored. The numbers 0 through 7 arranged about the inside of the 73-bp DNA superhelix mark the positions of sequential double helical turns. Those histones that are drawn in their entirety are primarily associated with this DNA segment, whereas only fragments of H3 and H2B from the other half of the particle are shown. The two four-helix bundles shown are labeled H3' H3 and H2B H4. [Courtesy of Timothy Richmond, ETH, Zürich, Switzerland. PDBid 1AOI.]

digestion, some of the chromatosome's DNA is trimmed away in a process that releases histone H1. This yields the 205-kD **nucleosome core particle**, which consists of a 147-bp strand of dsDNA in association with the above histone octamer. The DNA cumulatively removed by this digestion, which had previously joined neighboring nucleosome core particles, is known as **linker DNA**. Its average length, which varies from 10 to 50 bp from organism to organism and tissue to tissue, is ~18 bp in yeast, ~28 bp in *Drosophila*, and ~38 bp in humans.

The X-ray structures of nucleosome core particles containing 146- or 147-bp palindromic DNAs of defined sequence and histones from *X. laevis*, chicken, and *S. cerevisiae* were respectively determined by Timothy Richmond, Gerard Bunick, and Karolin Luger. These structures reveal the nucleosome core particle to be a nearly 2-fold symmetric wedge-shaped disk that has a diameter of ~110 Å and a maximum thickness of ~60 Å. The DNA, which assumes the B form, is wrapped around the outside of the histone octamer in 1.65 turns of a left-handed superhelix (Fig. 34-7). This is the origin of the negative supercoiling in eukaryotic DNA (topoisomerase II relaxes the compensatory positive supercoils generated when DNA wraps around nucleosomes).

Despite having only weak sequence similarity, all four types of histones in the histone octamer share a similar ~70-residue **histone fold** near their C-termini in which a long central helix is flanked on each side by a loop and a shorter helix (Fig. 34-8). Pairs of histone folds interdigitate in head-to-tail arrangements to form the crescent-shaped

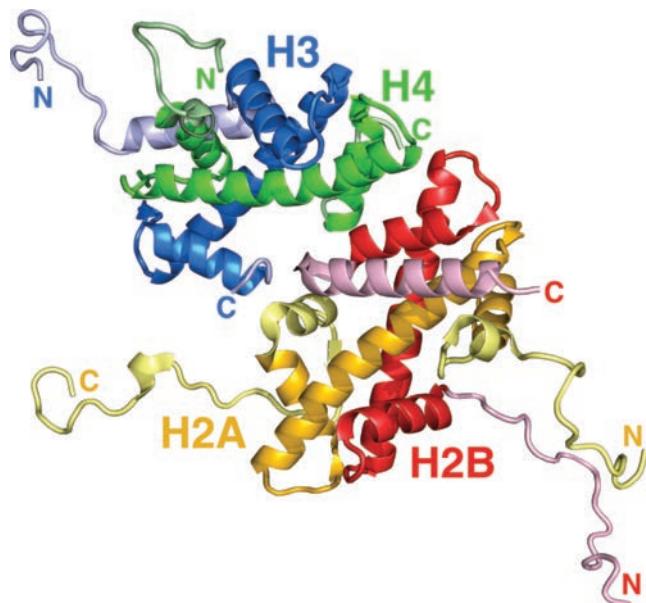


Figure 34-8 X-ray structure of a histone octamer within the nucleosome core particle. Those portions of H2A, H2B, H3, and H4 that form the histone folds are yellow, red, blue, and green, respectively, with their N- and C-terminal tails colored in lighter shades. [Based on an X-ray structure by Gerard Bunick, University of Tennessee and Oak Ridge National Laboratory, Oak Ridge, Tennessee. PDBid 1EQZ.]

heterodimers H2A–H2B and H3–H4, each of which binds 2.5 turns (27–28 bp) of dsDNA that curves around it in a 140° arc. Successive arcs are joined by 3- or 4-bp segments. The H3–H4 pairs interact, via a bundle of four helices from the two H3 histones, to form an (H3–H4)₂ tetramer with which each H2A–H2B pair interacts, via a similar four-helix bundle between H2B and H4, to form the histone octamer (Fig. 34-7b).

The histones bind exclusively to the inner face of the DNA, primarily via its sugar–phosphate backbones, through hydrogen bonds, salt bridges, and helix dipoles (their positive N-terminal ends), all interacting with phosphate oxygens, as well as through hydrophobic interactions with the deoxyribose rings. These interactions together with the nucleosome's net positive charge facilitate the bending of the negatively charged dsDNA around the nucleosome. There are few contacts between the histones and the bases, in accord with the nucleosome's lack of sequence specificity (but see Section 34-1Bf). However, an Arg side chain is inserted into the DNA's minor groove at each of the 14 positions it faces the histone octamer. The DNA is bent by an average of $(1.65 \text{ turns/nucleosome} \times 360^\circ/\text{turn})/(147 \text{ bp/nucleosome}) = 4^\circ/\text{bp}$ and its superhelix has an average radius of 42 Å and a pitch (rise per turn) of 26 Å. However, the DNA does not follow a uniform superhelical path but, rather, is bent fairly sharply at several locations due to outward bulges of the histone core. Moreover, the DNA double helix exhibits considerable conformational variation along its length such that its twist, for example, varies from 7.5 to 15.2 bp/turn with an average value of 10.2 bp/turn (vs 10.4 bp/turn for B-DNA in solution). Approximately 75% of the DNA surface is accessible to solvent and hence appears to be available for interactions with DNA-binding proteins.

The histones of the nucleosome core contain N-terminal tails that emanate from their central histone folds (Fig. 34-8) and vary in length from 23 to 43 residues (they comprise ~25% of the mass of these histones). These highly positively charged polypeptide segments, in agreement with previous biochemical studies, extend beyond the DNA. Those of H2B and H3 emerge from the nucleosome between the gyres (helical turns) of the DNA superhelix in channels formed by two vertically aligned minor grooves, whereas those of H2A and H4 do so above and below the DNA helices. The segments of the N-terminal tails that extend past the DNA are largely unstructured, that is, they are devoid of secondary structure, and substantial portions of their N-terminal segments are disordered. Nevertheless, one of the H4 N-terminal tails makes multiple hydrogen bonds and salt bridges with a highly negatively charged region on an H2A–H2B dimer of an adjacent nucleosome in the crystal structure. In addition, an H2A N-terminal tail of one nucleosome interacts with both the DNA and the H2A N-terminal tail of a neighboring nucleosome. Solution studies indicate, moreover, that the N-terminal tails interact with linker DNA. We shall see in Section 34-3Baa that the modulation of these interactions by the extensive and varied post-translational modifications of the N-terminal tails listed above are implicated in facilitating chromatin

unfolding to make its component DNA available to participate in such essential processes as transcription, DNA replication, recombination, and DNA repair. Among the foregoing histones, only H2A has an extensive C-terminal tail (39 residues), although it is entirely contained within the body of the nucleosome core and hence is unlikely to participate in internucleosomal interactions.

The archaeon *Methanothermus fervidus* expresses two closely related proteins that form a spheroidal complex with DNA, which presumably functions to prevent the thermal denaturation of this hyperthermophile's DNA. These proteins are ~30% identical in sequence to the histone fold domains of the histone octamer but lack their N- and C-terminal tails. Evidently, these tails have been added to the histone fold during the course of evolution.

b. Histone H1 “Seals Off” the Nucleosome

In the micrococcal nuclease digestion of nucleosomes, the ~200-bp DNA is first degraded to 167 bp. Then there is a pause before histone H1 is released from the chromatosomes and the DNA is further shortened to 147 bp. The 2-fold symmetry of the core particle suggests that the reduction in length of the 167-bp DNA comes about by the removal of 10 bp from each of its two ends. Since the 147-bp DNA of the core particle makes ~1.7 superhelical turns, the 167-bp intermediate should be able to make nearly two full superhelical turns, which would bring its two ends as close together as possible. Aaron Klug therefore proposed that histone H1 binds to nucleosomal DNA in a cavity formed by the central segment of its DNA and the segments that enter and leave the core particle (Fig. 34-9). This model is supported by the observation that in chromatin filaments containing H1, the DNA enters and leaves the nucleosome on the same side (Fig. 34-10a), whereas in H1-depleted chromatin, the entry and exit points are more randomly distributed and tend to occur on opposite sides

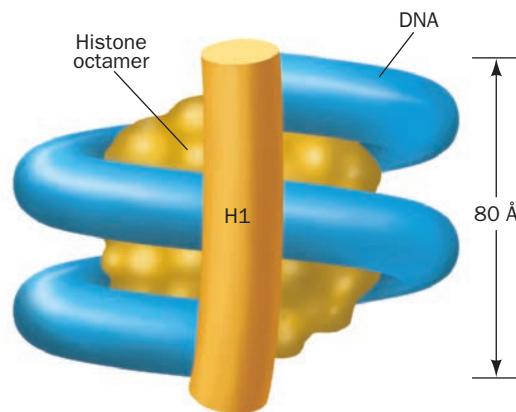


Figure 34-9 Model of the interaction of histone H1 with the DNA of the 166-bp chromatosome. The DNA's two complete superhelical turns (blue) enable H1 to bind to the DNA's two ends and its middle. Here the histone octamer is represented by the central spheroid (yellow) and the H1 molecule is represented by the cylinder (orange).

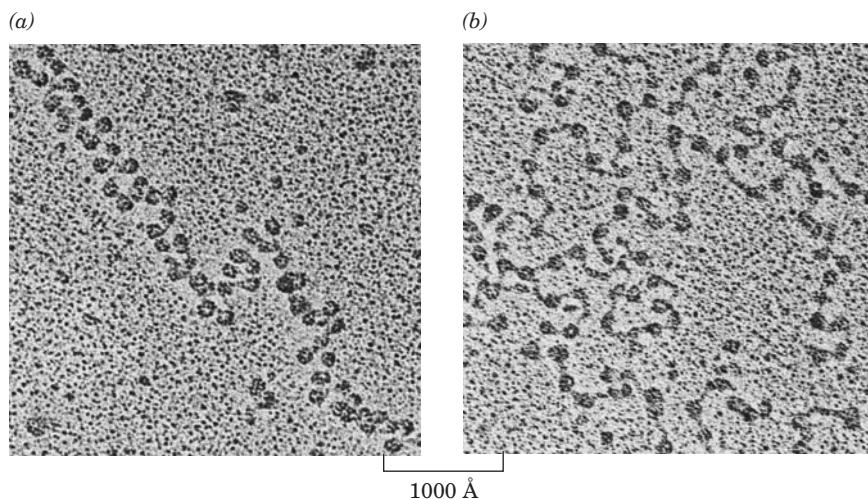


Figure 34-10 Electron micrographs of chromatin. (a) H1-containing chromatin and (b) H1-depleted chromatin, both in 5 to 15 mM salt. [Courtesy of Fritz Thoma, ETH, Zürich, Switzerland.]

of the nucleosome (Fig. 34-10b). The model also suggests that the length of the linker DNA is controlled by the sub-species of histone H1, which are collectively known as **linker histones**, bound to it.

Histone H5 is a variant of histone H1 that has several Lys → Arg substitutions and binds chromatin more tightly. The observations that the expression of histone H5 in rat sarcoma cells inhibits DNA replication, thereby arresting cells in the G₁ phase of the cell cycle, and that histone H5 more closely resembles **histone H1°** (a histone H1 variant that occurs in terminally differentiated cells) than does histone H1 itself, suggest that histone H5 is associated with replicationally and transcriptionally inactive chromatin.

Linker histones consist of a highly conserved globular, trypsin-resistant domain that is flanked by extended N- and C-terminal arms that are rich in basic residues. These basic arms, which comprise more than half of the intact protein, are therefore thought to interact with the linker DNA connecting adjacent nucleosomes even though it is the globular domain that is required for the binding of histone H1 to the nucleosome.

c. The Globular Domain of Histone H5 Structurally Resembles CAP Protein

Venki Ramakrishnan has determined the X-ray structure of chicken **GH5**, an 89-residue polypeptide that contains the 81-residue globular domain of the 190-residue histone H5 (although GH5's five N-terminal and eleven C-terminal residues are disordered). The polypeptide chain folds into a 3-helix bundle with a 2-stranded β sheet at its C-terminus (Fig. 34-11). This structure, and in particular its 3-helix bundle, is strikingly similar in conformation to that of the helix-turn-helix (HTH) motif-containing DNA-binding domain of *E. coli* catabolite activator protein (CAP; Fig. 31-31). Thus, even though there is little sequence identity between GH5 and CAP, their similar structures

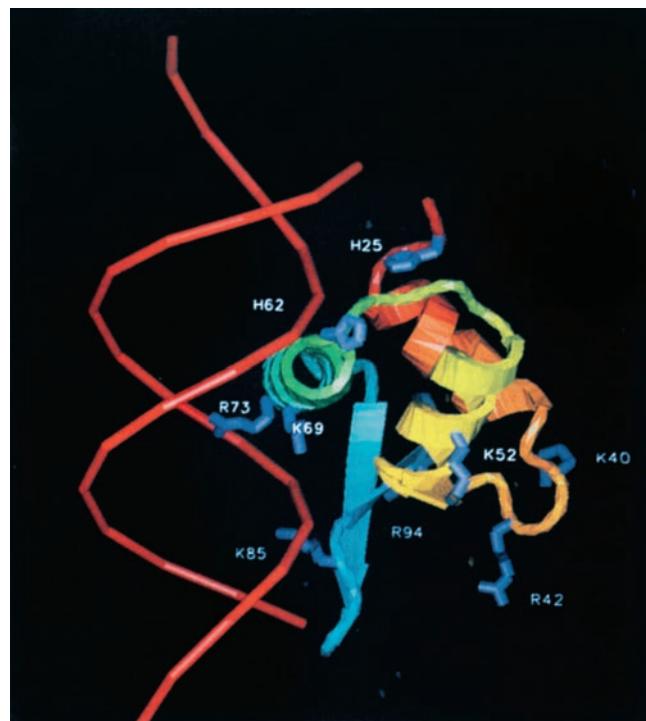


Figure 34-11 X-ray structure of GH5 drawn in hypothetical complex with DNA. This model was constructed by superimposing the structure of GH5 on that of CAP in the CAP–DNA structure (Fig. 31-31a). However, to avoid any presumptions about the nature of the DNA, that in the CAP structure, which is bent, was replaced by canonical B-DNA, which is represented here by its phosphate backbone (red). GH5 is shown in ribbon form colored in rainbow order from its N-terminus (red) to its C-terminus (blue). Conserved basic residues, as well as two His residues that have been cross-linked to DNA, are shown in stick form (dark blue). [Courtesy Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1HST.]

suggest that GH5 binds DNA in a manner analogous to CAP. Indeed, a model of the GH5–DNA complex based on the known X-ray structure of the CAP–DNA complex (Section 31-3Cb) positions GH5’s highly conserved Lys 69, Arg 73, and Lys 85 side chains to interact with the DNA (Fig. 34-11). These residues, which all have counterparts in CAP, are protected against chemical modification in chromatin. Moreover, GH5 contains a cluster of four conserved basic residues on the opposite face of the protein from its “recognition helix,” which could interact with a second segment of duplex DNA in agreement with the experimental evidence that GH5 simultaneously binds two DNA duplexes.

d. Parental Nucleosomes Are Transferred to Daughter Duplexes on DNA Replication

The *in vivo* replication of eukaryotic DNA is accompanied by its packaging into chromatin; that is, it is the chromatin that actually is replicated. What, then, is the fate of the histone octamers originally associated with the parental DNA? There are several possibilities: The “parental” octamers may remain associated with either the leading strand or the lagging strand, or they may be partitioned between the two daughter DNA duplexes, either at random or in some systematic way. Although attempts to resolve this issue at first yielded contradictory results, the weight of the evidence now indicates that parental octamers are distributed at random between the daughter duplexes. Moreover, the parental octamers remain associated with DNA during the replication process instead of dissociating from the parental DNA and later rebinding the daughter duplexes. Thus, nucleosomes either open up to permit the passage of a replication fork or parental histone octamers immediately in front of an advancing replication fork are somehow transferred to the daughter duplexes immediately behind the replication fork.

e. Nucleosome Assembly Is Facilitated by Molecular Chaperones

How are nucleosomes formed *in vivo*? *In vitro*, at high salt concentrations, nucleosomes self-assemble from the proper mixture of DNA and histones. In fact, when only H3, H4, and DNA are present, the mixture forms nucleosome-like particles by the deposition of (H3)₂(H4)₂ tetramers onto the DNA. Nucleosomes are then formed by the recruitment of H2A–H2B dimers to these particles followed by H1. A variety of evidence indicates that this is also the case *in vivo*.

At physiological salt concentrations, *in vitro* nucleosome assembly occurs much more slowly than at high salt concentrations and, unless the histone concentrations are carefully controlled, is accompanied by considerable histone precipitation. However, in the presence of **nucleoplasmin**, a 199-residue acidic protein that has been isolated from *X. laevis* oocyte nuclei, and DNA topoisomerase I (Section 29-3Cb), nucleosome assembly proceeds rapidly without histone precipitation. Nucleoplasmin binds to histones but neither to DNA nor to nucleosomes. Evidently,

nucleoplasmin functions as a molecular chaperone (Section 9-2C) to bring histones and DNA together in a controlled fashion, thereby preventing their nonspecific aggregation through their otherwise strong electrostatic interactions. In fact, nucleoplasmin, which Ron Laskey discovered in 1978, was the first molecular chaperone to be described. The topoisomerase I, no doubt, acts to provide the nucleosome with its preferred level of supercoiling.

A variety of other molecular chaperones have since been shown to facilitate the formation and disassembly of nucleosomes during the replication, repair, and transcription of eukaryotic DNA, the remodeling of nucleosomes (Section 34-3B), and the transport of histones. Among them are **Asf1** (for *anti-silencing function 1*), **CAF-1** (for *chromatin-assembly factor-1*), and **HIRA** (for *histone regulator A*), all of which interact with histone H3–H4, and **NAP1** (for *nucleosome assembly protein 1*), which interacts with histone H2A–H2B, as does nucleoplasmin.

The X-ray structure of Asf1 in complex with histones H3 and H4, independently determined by Toshiya Senda and by Mair Churchill and Jessica Tyler, reveals that Asf1 forms a sandwich of 5- and 3-stranded antiparallel β sheets (Fig. 34-12). Asf1 interacts with H3 in a way that prevents

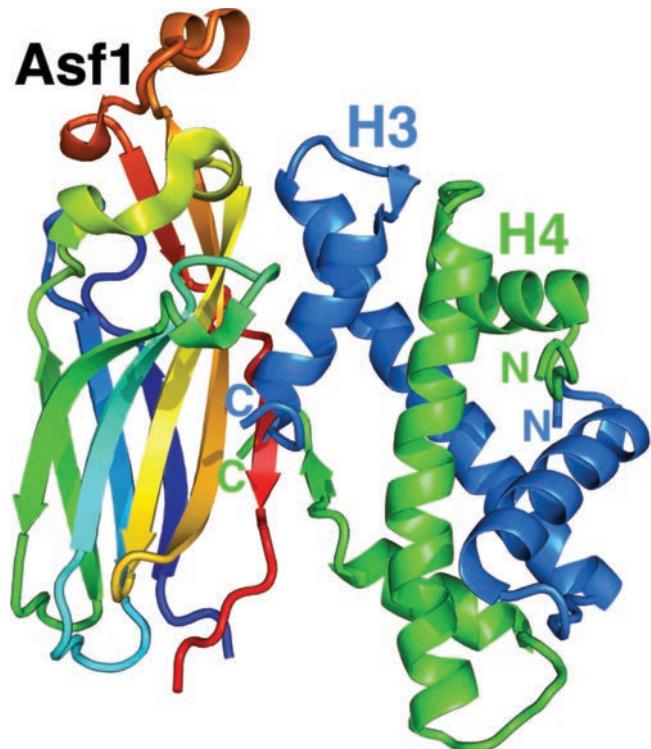


Figure 34-12 X-ray structure of the Asf1–H3–H4 heterotrimer. Human Asf1 (residues 1–159 of the 206-residue protein) is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), H3 (residues 60–135) is blue, and H4 (residues 24–100) is green. The H3–H4 dimer is viewed approximately as that in Fig. 34-7b. [Based on an X-ray structure by Toshiya Senda, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan. PDBid 2IO5.]

the formation of the 4-helix bundle between the two H3 subunits in the nucleosome and hence disrupts the (H3–H4)₂ tetramer (Fig. 34-7b). In addition, the C-terminal tail of H4 undergoes a large conformational shift from its position in the nucleosome in which its β strand forms a 2-stranded parallel β sheet with the C-terminal tail of H2A to one in which the H4 β strand extends Asf's 5-stranded β sheet (Fig. 34-12), thereby weakening the interaction of the H3–H4 dimer with the nucleosome. Thus, the assembly/disassembly of the nucleosome appears to involve the removal/insinuation of Asf1 from/into the nucleosome. Perhaps these processes are triggered by the presence/absence of DNA.

f. The Affinity of Histone Octamers for DNA Varies with the DNA's Sequence

The flexibility of dsDNA varies with its sequence and hence the affinity of a histone octamer for DNA is sequence dependent. In particular, the dinucleotides AA, TT, and TA tend to bend toward their minor grooves, whereas GC tends to bend toward its major groove. Since positions on DNA at which the minor and major grooves face inward on the nucleosome alternate by \sim 5 bp, sequences in which these dinucleotides occur every \sim 10 bp with AA/TT/AT separated by \sim 5 bp from GC facilitate the wrapping of DNA in a nucleosome.

In addition, segments of dsDNA with the sequence poly(dA) · poly(dT), which are known as **poly(dA:dT) tracts**, are unusually stiff. This is probably because their relatively narrow minor grooves are occupied by an assembly of highly ordered water molecules—a so-called **spine of hydration**—that are extensively hydrogen bonded to both the DNA and each other. Consequently, poly(dA:dT) tracts tend to resist incorporation into nucleosomes. In fact, poly(dA:dT) tracts of 10 to 20 bp are highly enriched in eukaryotic genomes but not in prokaryotic genomes. These tracts are associated with promoters and origins of DNA replication, where nucleosome depletion apparently increases the availability of these sites to the proteins that mediate these processes.

Despite the foregoing, DNA sequence does not appear to be the controlling factor in determining nucleosome organization *in vivo*. Rather, as we shall discuss in Section 34-3B, histone tail modifications, the distributions of histone subtypes, and the actions of ATP-dependent protein complexes that actively remodel chromatin have important roles in determining the arrangement of nucleosomes at the various genomic sites. Thus, as we shall see, *nucleosomes are highly dynamic entities whose positional fluctuations strongly influence gene expression*.

C. 30-nm Fibers: The Second Level of Chromatin Organization

The 167-bp nucleosomal DNA has a packing ratio of \sim 7 (its 560-Å contour length is wound into an \sim 80-Å-high supercoil). Clearly, the filament of nucleosomes, which only occurs at low ionic strengths and hence is unlikely to have

an independent existence *in vivo*, represents only the first level of chromosomal DNA compaction. Only at physiological ionic strengths does the next level of chromosomal organization become apparent.

As the salt concentration is raised, the H1-containing nucleosome filament initially folds to a zigzag conformation (Fig. 34-10a), whose appearance suggests that nucleosomes interact through contacts between their H1 molecules. Then, as the salt concentration approaches the physiological range, chromatin forms an \sim 30-nm-diameter fiber in which the nucleosomes are visible (Fig. 34-13). Two models for this **30-nm fiber** are given the greatest credence.

The first model (Fig. 34-14a), proposed by Richmond, is based on the X-ray structure that he determined of a tetranucleosome in which the DNA has a strong nucleosome positioning sequence (see below) that repeats every 167 bp. In this model, the nucleosomes follow a zigzag path in which all the odd-numbered nucleosomes and all the even-numbered nucleosomes stack next to each other. This \sim 25-nm-diameter structure, in which the two stacks wind around each other in two left-handed helices (a so-called **two-start helix**; the DNA forms a right-handed supersuperhelix), has 18.9 nucleosomes per turn, a pitch of 31.6 nm, and hence a rise of 0.6 nucleosome/nm. The short 20-bp linker DNAs cross back and forth through the center of the fiber between successive nucleosomes, which is only possible because these linker DNAs are unconstrained by linker histones. Longer linker DNAs would allow the binding of linker histones and thus increase the fiber diameter, but would require the rigid dsDNA to bend and/or disrupt the contacts between nucleosomes.

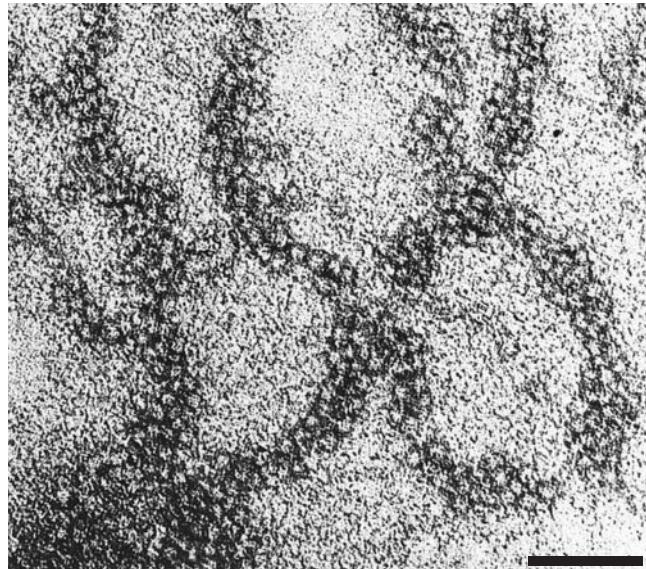
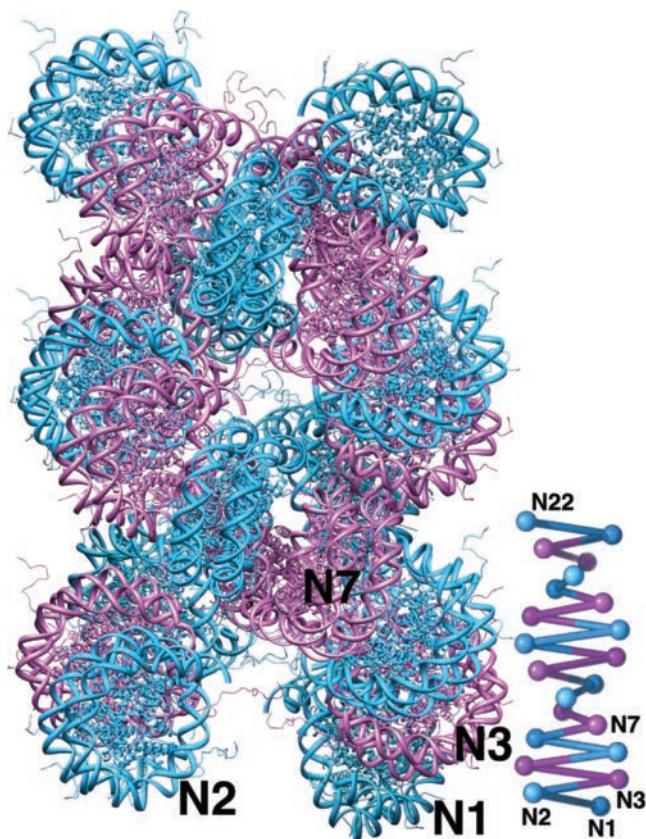


Figure 34-13 Electron micrograph of 30-nm chromatin fibers. Note that the fibers are two to three nucleosomes across. The bar represents 1000 Å. [Courtesy of Jerome B. Rattner, University of Calgary, Canada.]

(a)



(b)

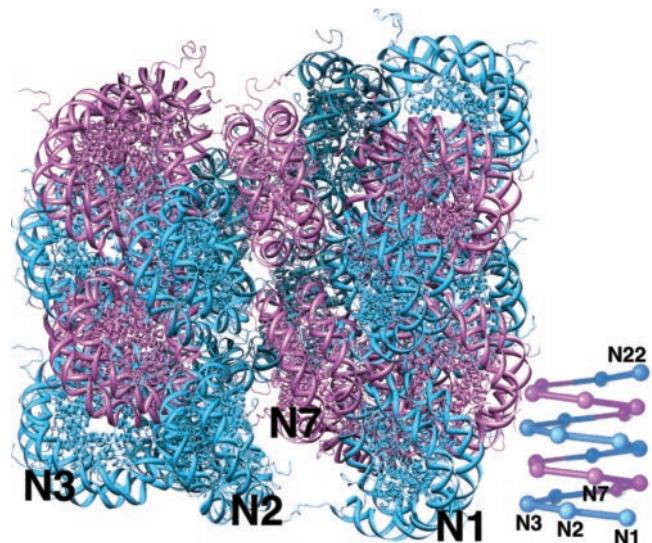


Figure 34-14 Models of the 30-nm chromatin fiber consisting of 22 nucleosomes in which the fiber axis is vertical. (a) A model based on the X-ray structure of a tetranucleosome that lacks linker histones. Alternate pairs of nucleosomes are colored cyan and pink and the DNA is represented by a wider ribbon than the polypeptide chains. Note how this model consists of a left-handed two-start helix of stacked nucleosomes. (b) A model that is consistent with the dimensions of assemblies of long strands of nucleosomes and linker histones. The nucleosomes, which are represented as in Part a, are colored cyan and pink in alternate gyres of the left-handed one-start helix. Linker histones are not

included in this model. Note how the nucleosomes are interdigitated. In both models, the first (N1), second (N2), third (N3), and seventh (N7) nucleosomes in the linear DNA sequences are indicated and the inset is a schematic diagram of the DNA's connectivity. The 2-fold axis of each nucleosome is perpendicular to and intersects the fiber axis such that its linker histone binding site occupies the interior of the fiber. [Courtesy of Daniela Rhodes, MRC Laboratory of Molecular Biology, Cambridge, U.K. The X-ray structure on which Part a is based was determined by Timothy Richmond, ETH, Zürich, Switzerland. PDBid 1ZBB.]

Electron micrographs, by Daniela Rhodes, reveal that assemblies consisting of linker histones and a long string of nucleosomes in which the DNA consists of physiologically realistic repeat lengths containing a strong nucleosome positioning sequence have dimensions that are inconsistent with the foregoing model. However, these dimensions are consistent with an alternate model of the 30-nm fiber (Fig. 34-14b), formulated by Rhodes, in which the DNA and the nucleosomes follow a single left-handed helix (a **one-start helix**; the DNA forms a left-handed super-superhelix) that has a diameter of ~34 nm, 5.4 nucleosomes per turn, a pitch of ~5.4 nm, and hence a rise of 1 nucleosome/nm. The observation that the dimensions of the nucleosome assemblies are constant over a wide range of linker DNA lengths suggests that the linker DNA, which is not resolved in the electron micrographs, is bent and occupies the interior of the fiber.

The foregoing two models suggest that there are (at least) two species of higher order chromatin structures: that represented in Fig. 34-14a, which is relatively loose, lacks linker histones, and is therefore more accessible to proteins that interact with DNA and histones; and that represented by Fig. 34-14b, which is compact, contains linker histones, and is therefore less accessible to proteins. However, Kensal van Holde has argued that the 30-nm fiber lacks a regular structure; rather, because of the varying lengths of the presumably straight and stiff linker DNAs (every base pair added to a linker DNA rotates the adjacent nucleosomes by ~36° with respect to each other), it has an irregular helixlike structure that simulations indicate forms a fiber with an average diameter of 30 nm. This would account for the difficulty in definitively determining the structure of the 30-nm fiber despite numerous attempts to do so over more than three decades.

D. Radial Loops: The Third Level of Chromatin Organization

Histone-depleted metaphase chromosomes exhibit a central fibrous protein matrix or scaffold surrounded by an extensive halo of DNA (Fig. 34-15a). The strands of DNA that can be followed are observed to form loops that enter and exit the scaffold at nearly the same point (Fig. 34-15b). Most of these loops have lengths in the range 15 to 30 μm (which corresponds to 45–90 kb), so that when condensed as 30-nm fibers they would be $\sim 0.6 \mu\text{m}$ long. Electron micrographs of chromosomes in cross section, such as Fig. 34-16a, strongly suggest that the chromatin fibers of metaphase chromosomes are radially arranged. If the observed loops correspond to these radial fibers, they would each contribute $\sim 0.3 \mu\text{m}$ to the diameter of the chromosome (a fiber must double back on itself to form a loop). Taking into account the 0.4- μm width of the scaffold, this model predicts the diameter of the metaphase chromosome to be 1.0 μm , in agreement with observation (Fig. 34-16b). A typical human chromosome, which contains ~ 140 million bp, would therefore have ~ 2000 of these ~ 70 -kb radial loops (Fig. 34-16c). The 0.4- μm -diameter scaffold of

such a chromosome has sufficient surface area along its 6- μm length to bind this number of radial loops. The radial loop model therefore accounts for DNA's observed packing ratio in metaphase chromosomes. However, keep in mind that the higher order structures of chromosomes probably vary from region to region within a given chromosome and certainly change with the stage in the cell cycle. Nevertheless, it is clear that a metaphase chromosome's single DNA molecule is roughly collinear with the chromosome's longitudinal axis.

a. Metaphase Chromosomes Are Organized by Cohesin and Condensin

The radial DNA loops are attached to the matrix via AT-rich **matrix-associated regions (MARs)**; alternatively, **scaffold attachment regions (SARs)**. They are organized by **nonhistone proteins**, whose thousands of varieties constitute $\sim 10\%$ of chromosomal proteins. Among the most prominent nonhistone proteins are the **structural maintenance of chromosomes (SMC) proteins**, which are members of a protein family that is expressed by prokaryotes as well as eukaryotes. SMC proteins are large polypeptides

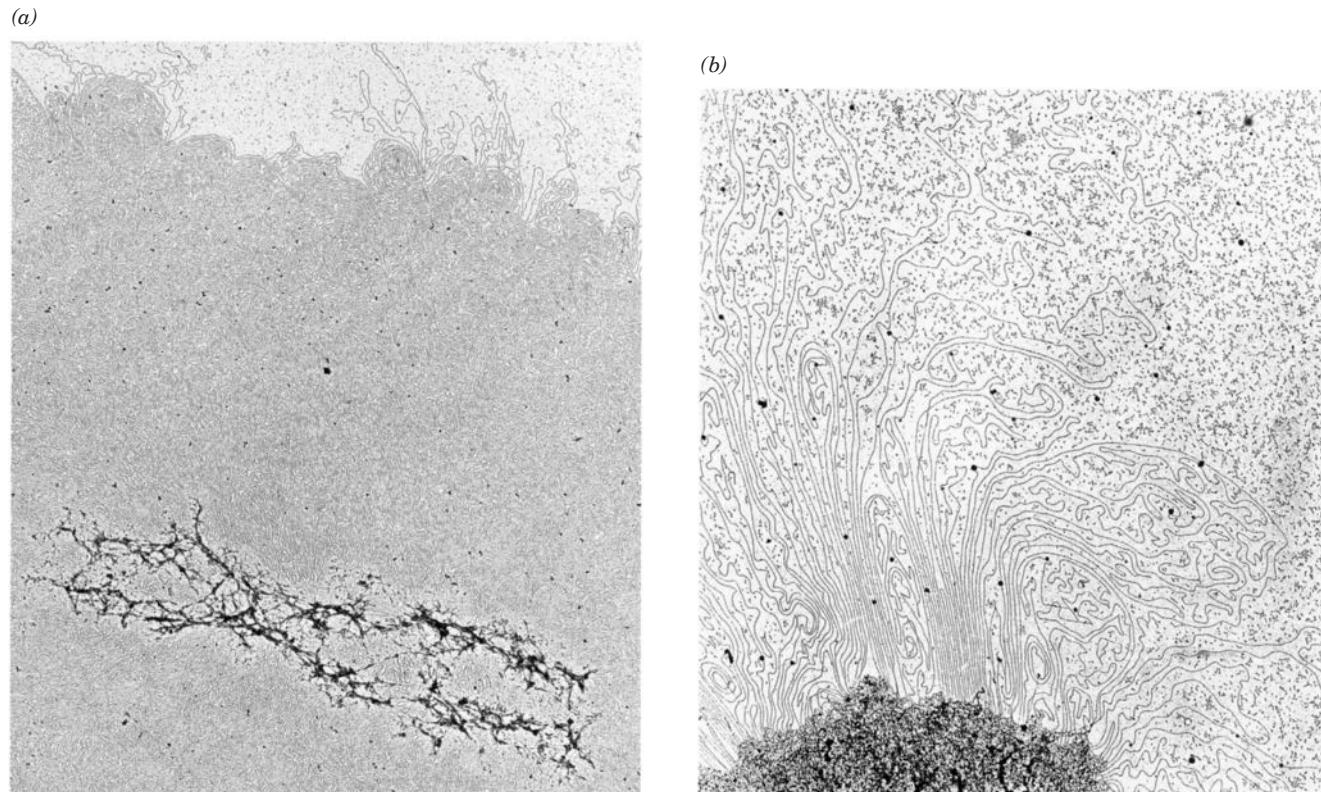


Figure 34-15 Electron micrograph of a histone-depleted metaphase human chromosome. (a) The central protein matrix (scaffold) serves to anchor the surrounding DNA. (b) At higher

magnification it can be seen that the DNA is attached to the scaffold in loops. [Courtesy of Ulrich Laemmli, University of Geneva, Switzerland.]

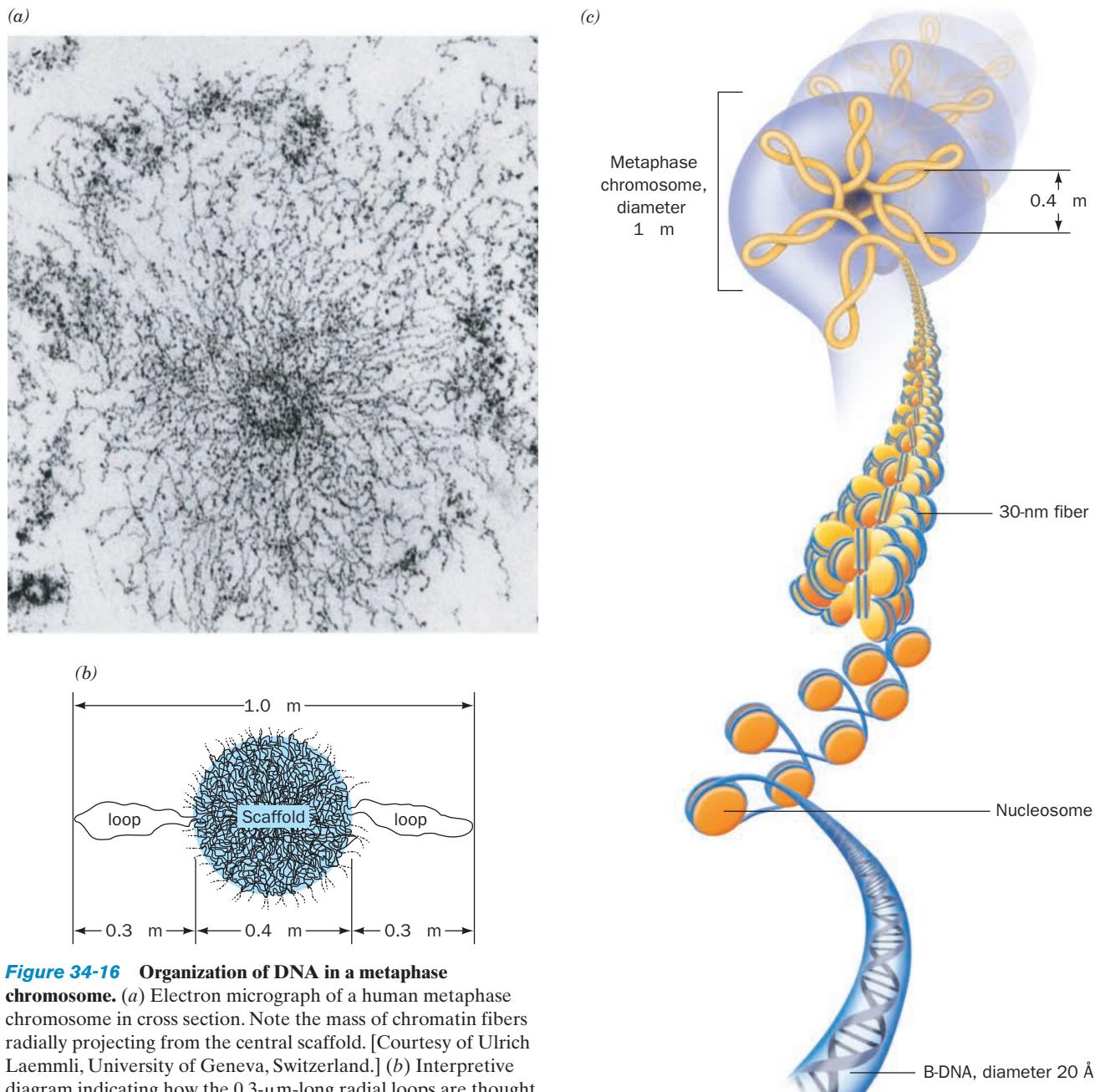
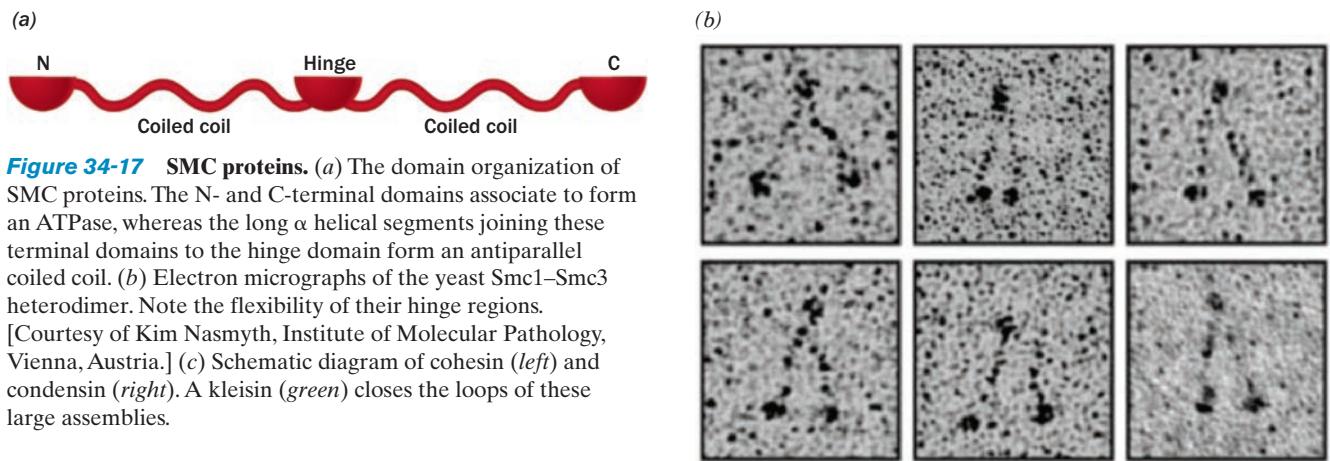


Figure 34-16 Organization of DNA in a metaphase chromosome. (a) Electron micrograph of a human metaphase chromosome in cross section. Note the mass of chromatin fibers radially projecting from the central scaffold. [Courtesy of Ulrich Laemmli, University of Geneva, Switzerland.] (b) Interpretive diagram indicating how the 0.3- μm -long radial loops are thought to combine with the 0.4- μm -wide scaffold to form the 1.0- μm -diameter metaphase chromosome. (c) Model diagramming the various levels of metaphase chromatin organization. In this model, rosettes of twisted radial loops form additional levels of coiling and thus the DNA has five levels of coiling (DNA double helix, nucleosomes, 30 nm fibers, radial loops, and coils of rosettes).

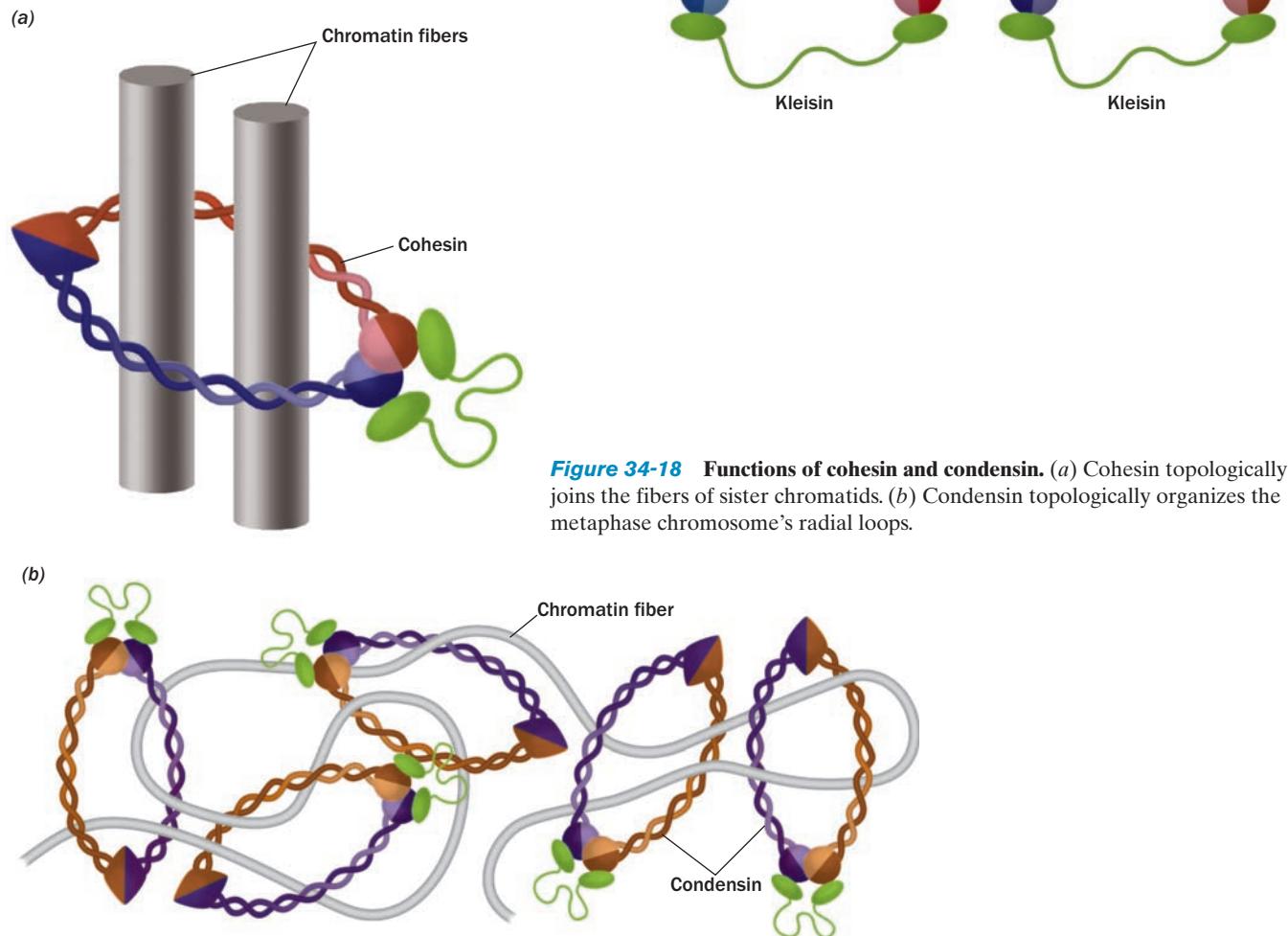
(1000–1300 residues) that have a unique domain organization (Fig. 34-17a): Their N- and C-terminal domains, which collectively form an ATPase domain resembling that of ABC transporters (Section 20-3E), are linked by two ~50-nm-long α helical segments that come together to form an antiparallel coiled coil and which are joined by a nonhelical, so-called hinge domain. Two different family members join at their hinge regions to form a V-shaped heterodimer

(Fig. 34-17b). Eukaryotes have two major species of SMC proteins, **cohesin**, whose SMC subunits are **Smc1** and **Smc3**, and **condensin**, whose SMC subunits are **Smc2** and **Smc4** (Fig. 34-17c). A third protein, a member of the **kleisin** (Greek: *kleisimo*, closure) superfamily, binds to the ATPase domains of both members of the SMC heterodimer to form a closed loop whose diameter is large enough to encompass two 10-nm fibers.

During the G₁ phase of the cell cycle (Section 30-4A) cohesin loads onto the dispersed chromatin fibers in an ATP-dependent manner. When the DNA is subsequently replicated during S phase the replisome passes through the



cohesin rings, thus keeping the resulting sister chromatids together (Fig. 34-18a). As the cell enters M phase (mitosis; Fig. 1-19), condensin organizes the radial loops of the metaphase chromosomes (Fig. 34-18b). Concurrently, most of the cohesin rings open up to allow the parallel sister chromatids to separate everywhere but at their centromeres (Fig. 34-1), a process that also requires the action of topoisomerase II to decatenate (untangle) the two dsDNA



strands. Then, as the sister chromatids complete their separation during the anaphase portion of mitosis, the remaining cohesin rings open up. Finally, as the two daughter cells enter interphase and their chromosomes disperse, the condensin rings open up to release the chromatin fibers.

The release of cohesin during the onset of M phase is induced by the phosphorylation, via a poorly understood process, of control proteins that bind to cohesin's kleisin subunit. A small population of cohesin bound near the centromere escapes this release; during anaphase, however, their kleisin subunits are cleaved by a protease named **separase**, thereby permitting the sister chromatids to fully separate.

E. Interphase Chromosomes

a. Polytene Chromosomes Contain Numerous Parallel DNA Strands

Until the 1980s, the diffuse structure of most interphase chromosomes (Fig. 34-2) made it all but impossible to characterize them at the level of individual genes. However, this difficulty was greatly ameliorated by the existence of “giant” banded chromosomes in certain terminally differenti-

ated (nondividing) secretory cells of dipteran (two-winged) flies (Fig. 34-19). These chromosomes, of which those from the salivary glands of *D. melanogaster* larvae were the most extensively studied, are produced by multiple replications of a synapsed (joined in parallel) diploid pair in which the replicas remain attached to one another and in register (see below). Each diploid pair may replicate in this manner as many as nine times so that the final **polytene** (Greek: *tainia*, band or ribbon) **chromosome** contains up to $2 \times 2^9 = 1024$ DNA strands. The function(s) of polytene chromosomes is unclear although probably they permit a greatly increased rate of transcription of certain genes.

The four giant chromosomes of *D. melanogaster* have an aggregate length of ~ 2 mm so that its haploid genome of 1.80×10^8 bp has an average packing ratio in these chromosomes of ~ 30 . About 95% of this DNA is concentrated in chromosomal bands (Fig. 34-20). These bands (more properly, **chromomeres**), as microscopically visualized through staining, form a pattern that is characteristic of each *D. melanogaster* strain. Indeed, chromosomal rearrangements such as duplications, deletions, and inversions result in a corresponding change in the banding pattern. *A polytene chromosome's banding pattern therefore forms a cytological map that parallels its genetic map.*

The characteristic banding pattern of each polytene chromosome suggests that its component DNA molecules are precisely aligned. This hypothesis was corroborated by the application of *in situ* (on site) **hybridization**. In this technique, developed by Mary Lou Pardue and Joseph Gall, an immobilized chromosome preparation is treated with NaOH to denature its DNA; it is then hybridized with a purified species of radioactively labeled mRNA (or its corresponding cDNA), and the chromosomal binding site of the radioactive probe is determined by autoradiography. A given mRNA hybridizes with one, or no more than a few, chromosomal bands (Fig. 34-21).

D. melanogaster's four polytene chromosomes exhibit an aggregate of ~ 5000 bands. It originally appeared that the number of *D. melanogaster* genes was roughly equal to

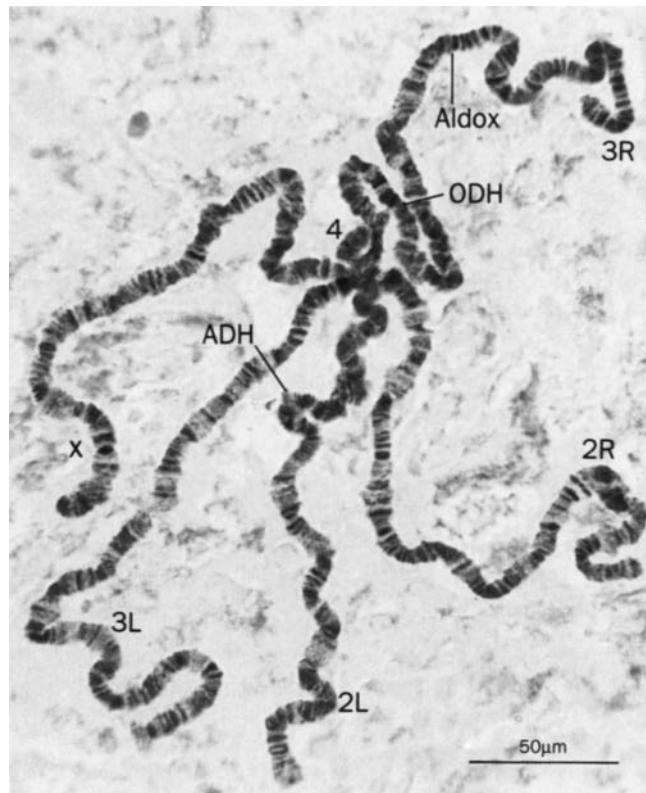


Figure 34-19 Photomicrograph of the stained polytene chromosomes from *D. melanogaster* salivary gland. Such chromosomes consist of darkly staining bands interspersed with light-staining interband regions. All four chromosomes in a single cell are held together by their centromeres. The chromosomal positions for the genes specifying alcohol dehydrogenase (ADH), aldehyde oxidase (Aldox), and octanol dehydrogenase (ODH) are indicated. [Courtesy of B.P. Kaufmann, University of Michigan.]

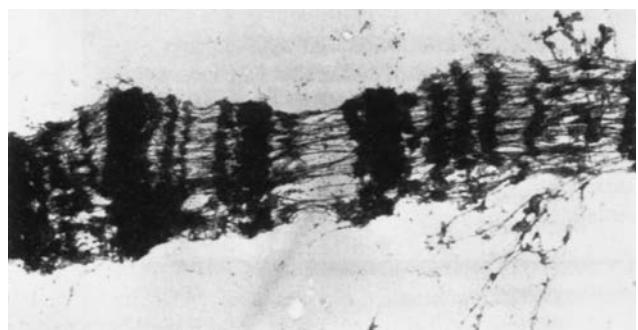


Figure 34-20 Electron micrograph of a segment of polytene chromosome from *D. melanogaster*. Note that its interband regions consist of chromatin fibers that are more or less parallel to the long axis of the chromosome, whereas its bands, which contain $\sim 95\%$ of the chromosome's DNA, are much more highly condensed. [Courtesy of Gary Burkholder, University of Saskatchewan, Canada.]

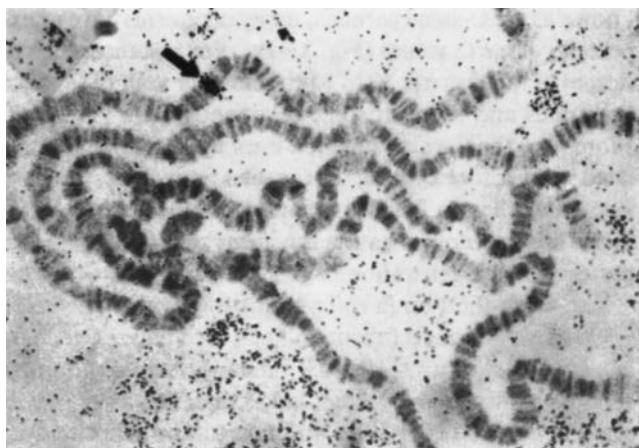


Figure 34-21 Autoradiograph of a *D. melanogaster* polytene chromosome that has been *in situ* hybridized with radioactively labeled yolk protein cDNA. The dark grains (arrow) identify the chromosomal location of the yolk protein gene. [From Barnett, T., Pachl, C., Gergen, J.P., and Wensink, P.C., *Cell* 21, 735 (1980). Copyright © 1980 by Cell Press.]

this number of bands and hence it was thought that each band corresponds to a single gene. However, the genome sequence of *D. melanogaster* contains \sim 14,000 genes, nearly three times its number of bands. In fact, genes have been shown to be located in both band and interband regions, with some bands containing several genes and others containing none. Thus, it appears that the banding pattern of polytene chromosomes is a consequence of different levels of gene expression due to variations in chromatin structure (Section 34-3B), with the genes in the relatively open interband regions presumably more highly expressed than those in the more condensed and hence less accessible bands.

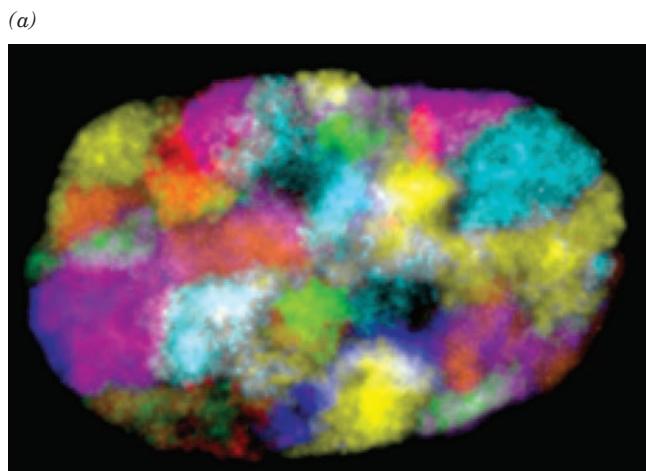


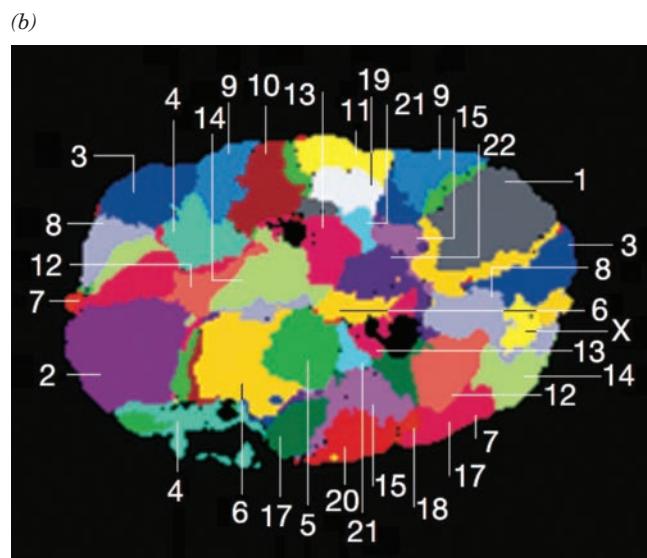
Figure 34-22 The chromosomal organization in a human interphase nucleus as determined by three-dimensional FISH. (a) A computer-generated section through the nucleus in which each species of chromosome has been assigned a different false color. The dark regions correspond to unstained nucleoli. A stack of such images locates each of the 46 chromosomes in three

b. Interphase Chromosomes Occupy Discrete Nuclear Territories

The experimental difficulties and lack of sensitivity of using radioactive probes stimulated the development of an improved method of *in situ* hybridization known as **fluorescence *in situ* hybridization (FISH)**. In this technique, the probe mRNA or DNA is fluorescently labeled and the chromosomal site to which it hybridized is identified by fluorescence microscopy.

How are the chromosomes in an interphase nucleus distributed? This question was answered using a method known as **three-dimensional FISH** in which the chromosomes in an interphase nucleus are each “painted” with a mixture of probes specific for sites that are scattered throughout the chromosome and in which the probes are labeled by a chromosome-specific combination of **fluorophores** (fluorescent groups). Then, using a sophisticated form of fluorescence microscopy that acquires a series of images at different levels in the nucleus and spectrally differentiates the various combinations of fluorophores, the three-dimensional locations of the various chromosomes in a nucleus are determined. The resulting images reveal that each chromosome occupies a specific space or **territory** in the nucleus that does not greatly overlap the territories of other chromosomes (Fig. 34-22). The territory occupied by a particular chromosome does not greatly vary during the lifetime of a cell, but it changes after cell division. However, the observation that **translocations**, in which a segment from one chromosome becomes attached to another, are more frequent between certain pairs of chromosomes suggests that their territories are often in proximity.

Images such as Fig. 34-22 might naively suggest that the nucleus is tightly packed with chromatin. However, a rough



dimensions. (b) A schematic diagram that identifies each of the chromosomes by its number. Note that homologous chromosomes (e.g., the two copies of chromosome 7) occupy different although usually similarly placed territories in the nucleus. [Courtesy of Thomas Cremer and Andreas Bolzer, University of Munich, Germany.]

calculation indicates that chromatin as 10-nm fibers occupies only ~17% of the nuclear volume. Moreover, electron micrographs of interphase nuclei reveal that their chromatin is dispersed mainly as 10- and 30-nm fibers, and FISH experiments have shown that RNA transcripts are distributed throughout chromosomal territories. Evidently, there is sufficient space within chromosomal territories to allow free access of the transcriptional machinery and presumably the replicational machinery to its target DNA.

2 GENOMIC ORGANIZATION

Higher organisms contain a great variety of cells that differ not only in their appearances (e.g., Fig. 1-10) but in the proteins they synthesize. Pancreatic acinar cells, for example, synthesize copious quantities of digestive enzymes, including trypsin and chymotrypsin, but no insulin, whereas the neighboring pancreatic β cells produce large amounts of insulin but no digestive enzymes. Clearly, each of these different types of cells expresses different genes. Yet most of a multicellular organism's somatic cells contain the same genetic information as the fertilized ovum from which they are descended (a phenomenon described as **totipotency**) as

is demonstrated, for example, by the ability to raise a mammal such as a sheep, cow, or mouse from an enucleated oocyte into which the nucleus from an adult cell had been inserted. Similarly, a single cell from a plant can give rise to the normal plant. Evidently, cells have enormous expressional flexibility. Nevertheless, only a small fraction of the DNA in higher eukaryotic genomes is expressed. What is the nature of the remaining unexpressed sequences and do they have any function? In this section we describe the genetic organization of the eukaryotic chromosome. How eukaryotic gene expression is controlled is the subject of Section 34-3.

A. The C-Value Paradox

One might reasonably expect the morphological complexity of an organism to be roughly correlated with the amount of DNA in its haploid genome, its **C value**. After all, the morphological complexity of an organism must reflect an underlying genetic complexity. Nevertheless, in what is known as the **C-value paradox**, many organisms have unexpectedly large C values (Fig. 34-23). For instance, the genomes of lungfish are 10 to 15 times larger than of those of mammals and those of some salamanders are yet

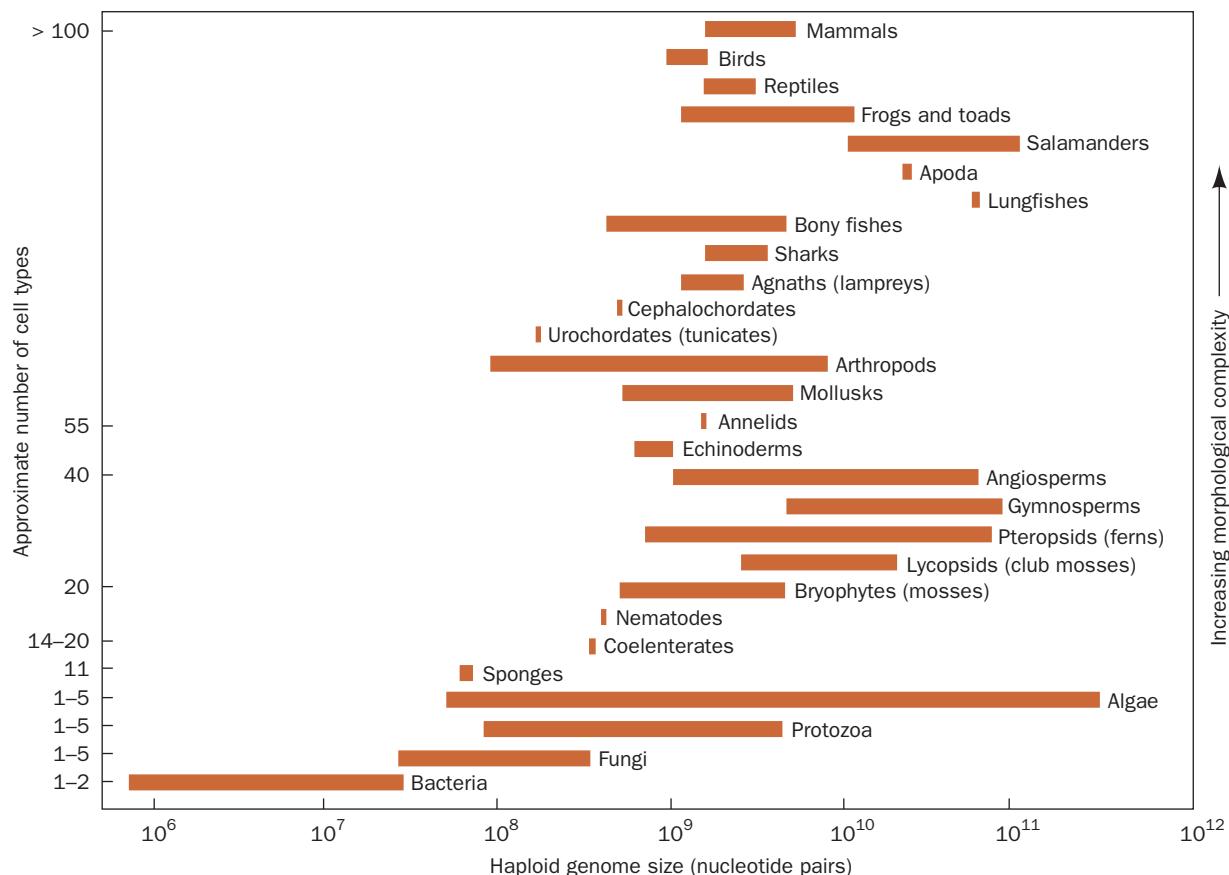


Figure 34-23 The range of haploid genome DNA contents in various categories of organisms indicating the C-value paradox. The morphological complexity of the organisms, as estimated

according to their number of cell types, increases from bottom to top. [After Raff, R.A. and Kaufman, T.C., *Embryos, Genes, and Evolution*, p. 314, Macmillan (1983).]

Table 34-2 Genome Size and Gene Number in a Variety of Organisms

Organism	Genome Size (kb)	Number of Genes
<i>Haemophilus influenzae</i> (bacterium)	1,830	1,740
<i>Escherichia coli</i> (bacterium)	4,639	4,289
<i>Saccharomyces cerevisiae</i> (yeast)	12,070	6,034
<i>Caenorhabditis elegans</i> (nematode)	97,000	19,099
<i>Oryza sativa</i> (rice)	389,000	~35,000
<i>Arabidopsis thaliana</i> (mustard weed)	119,200	~26,000
<i>Drosophila melanogaster</i> (fruit fly)	180,000	13,061
<i>Mus musculus</i> (mouse)	2,500,000	~23,000
<i>Homo sapiens</i> (human)	3,038,000	~23,000

larger. In fact, the C-value paradox even applies to closely related species; for example, the C values for several species of *Drosophila* have a 2.5-fold spread. Does the “extra” DNA in the larger genomes have a function, and if not, why is it preserved from generation to generation?

The number of genes encoded by an organism’s genome roughly increases with the organism’s complexity (Table 34-2) but does so at a much lower rate than had been previously anticipated. For example, the human genome, which is ~650 times larger than that of *E. coli*, encodes only ~5.4 times as many **structural** (protein-coding) **genes**. Certainly the control of genetic expression in eukaryotes is a far more elaborate process than it is in prokaryotes. Yet does much of the unexpressed DNA in the human genome, at least 98% of the total, function in the control of genetic expression? The recent and ongoing eukaryotic genome sequence determinations are beginning to provide answers to the foregoing questions.

a. C_0t Curve Analysis Indicates DNA Complexity

The rate at which DNA renatures is indicative of the lengths of its unique sequences. If DNA is sheared into uniform fragments of 300 to 10,000 bp (Section 5-3D), denatured, and kept at a low concentration so that the effects of mechanical entanglement are small, the rate-determining step in renaturation is the collision of complementary sequences. Once the complementary sequences have found each other through random diffusion, they rapidly zip up to form duplex molecules. The rate of renaturation of denatured DNA is therefore expressed

$$\frac{d[A]}{dt} = -k[A][B] \quad [34.1]$$

where A and B represent complementary single-stranded sequences and k is a second-order rate constant (Section 14-1Bb). Since $[A] = [B]$ for duplex DNA, Eq. [34.1] integrates to

$$\frac{1}{[A]} = \frac{1}{[A]_0} + kt \quad [34.2]$$

where $[A]_0$ is the initial concentration of A.

It is convenient to measure the fraction f of unpaired strands:

$$f = \frac{[A]}{[A]_0} \quad [34.3]$$

Combining Eqs. [34.2] and [34.3] yields

$$f = \frac{1}{1 + [A]_0 kt} \quad [34.4]$$

The concentration terms in these equations refer to unique sequences since the collision of noncomplementary sequences does not lead to renaturation. Hence, if C_0 is the initial concentration of base pairs in solution, then

$$[A]_0 = \frac{C_0}{x} \quad [34.5]$$

where x is the number of base pairs in each unique sequence and is known as the DNA’s **complexity**. For example, the repeating sequence $(AGCT)_n$ has a complexity of 4, whereas an *E. coli* chromosome, which consists of ~4.6 million bp of unrepeated sequence, has a complexity of ~4.6 million. Combining Eqs. [34.4] and [34.5] yields

$$f = \frac{1}{1 + C_0 kt/x} \quad [34.6]$$

When half of the molecules in the sample have renatured, $f = 0.5$, so that

$$C_0 t_{1/2} = \frac{x}{k} \quad [34.7]$$

where $t_{1/2}$ is the time for this to occur. The rate constant k is characteristic of the rate at which single strands collide in solution under the conditions employed, so it is independent of the complexity of the DNA and, for reasonably short DNA fragments, the length of a strand. Consequently, for a given set of conditions, the value of $C_0 t_{1/2}$ depends only on the complexity x of the DNA. This situation is indicated in

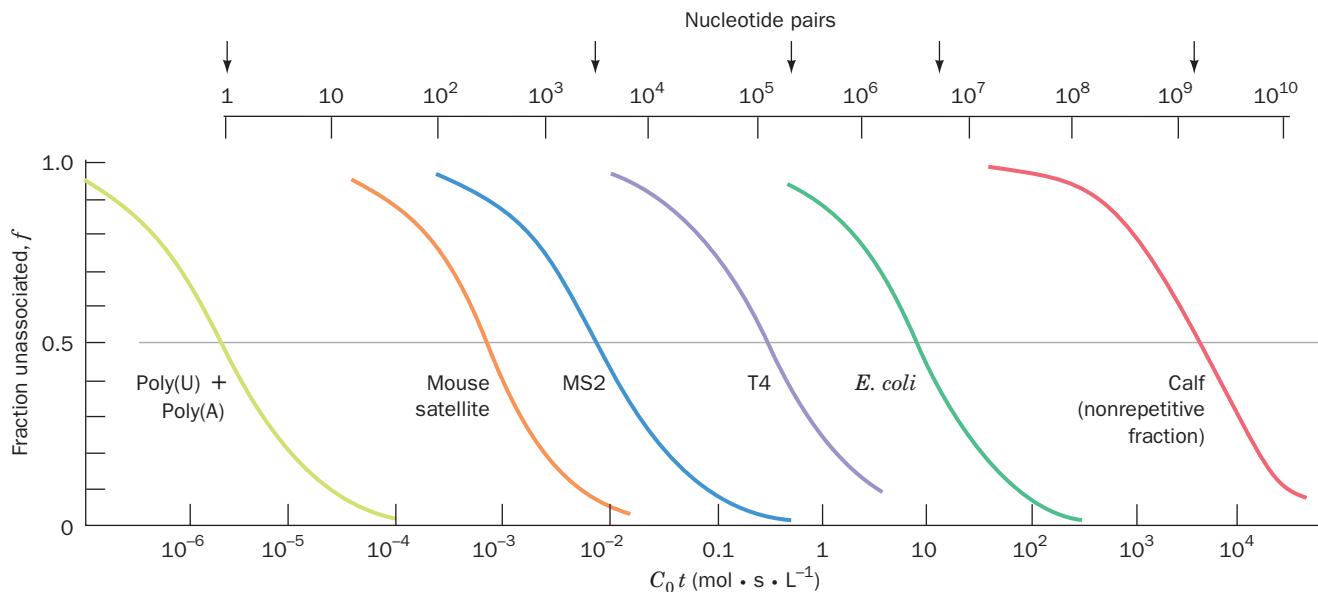


Figure 34-24 The reassociation ($C_0 t$) curves of duplex DNAs from various sources. The DNA was dissolved in a solution containing $0.18M$ Na^+ and sheared to an average length of

400 bp. The upper scale indicates the genome sizes of some of the DNAs (MS2 and T4 are bacteriophages). [After Britten, R.J. and Kohne, D.E., *Science* **161**, 530 (1968).]

Fig. 34-24, which is a series of plots of f versus $C_0 t$ for various DNAs. Such plots are referred to as $C_0 t$ (pronounced “cot”) curves. The complexities of the DNAs in Fig. 34-24 vary from 1 for the synthetic duplex poly(A) · poly(U) to $\sim 3 \times 10^9$ for some fractions of mammalian DNAs. Their corresponding values of $C_0 t_{1/2}$ vary accordingly.

The speed and sensitivity of $C_0 t$ curve analysis is greatly enhanced through the hydroxyapatite fractionation of the renaturing DNA. Hydroxyapatite, it will be recalled (Section 6-6Ba), binds double-stranded DNA at a higher phosphate concentration than it binds single-stranded DNA. The single- and double-stranded DNAs in a solution of renaturing DNA may therefore be separated by hydroxyapatite chromatography and the amounts of each measured. The single-stranded DNA can then be further renatured and the process repeated. If the renaturing DNA is radioactively labeled, much smaller quantities of it can be detected than is possible by spectroscopic means. Thus, through the hydroxyapatite chromatography of radioactively labeled DNA, the $C_0 t$ curve analysis of a DNA of such a high complexity that its $t_{1/2}$ is days or weeks can be conveniently measured in a small fraction of that time.

B. Repetitive Sequences

Consider a sample of DNA that consists of sequences with varying degrees of complexity. Its $C_0 t$ curve, Fig. 34-25 for example, is the sum of the individual $C_0 t$ curves for each complexity class of DNA. $C_0 t$ curve analysis (and, more recently, genome sequencing) has demonstrated that viral and prokaryotic DNAs have few, if any, repeated sequences (e.g., Fig. 34-24 for MS2, T4, and *E. coli*). In contrast, eukaryotic DNAs exhibit complicated $C_0 t$ curves (e.g., Fig. 34-26) that

must arise from the presence of DNA segments of several different complexities.

Eukaryotic $C_0 t$ curves may be attributed to the presence of five somewhat arbitrarily defined classes of DNAs: (1) **inverted repeats**, (2) **highly repetitive sequences** ($>10^6$ copies per haploid genome), (3) **moderately repetitive sequences** ($<10^6$ copies per haploid genome), (4) **segmental duplications** (blocks of 1–200 kb that have been copied to one or more regions of the genome that may be within the

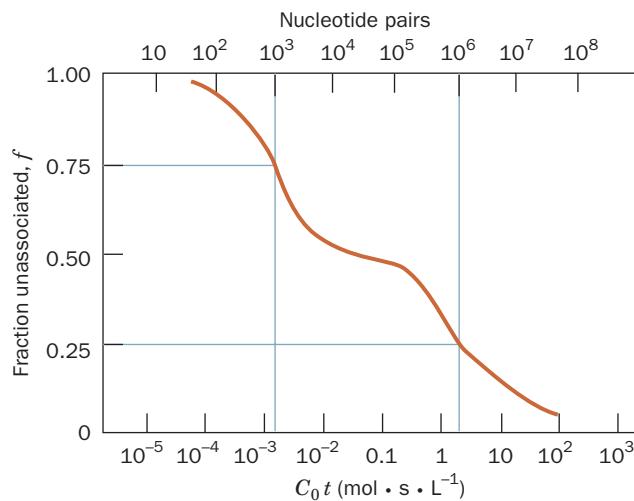


Figure 34-25 $C_0 t$ curve of a hypothetical DNA molecule. Before fragmentation, this DNA was 2 million bp in length and consisted of a unique sequence of 1 million bp and 1000 copies of a 1000-bp sequence. Note the curve’s biphasic nature.

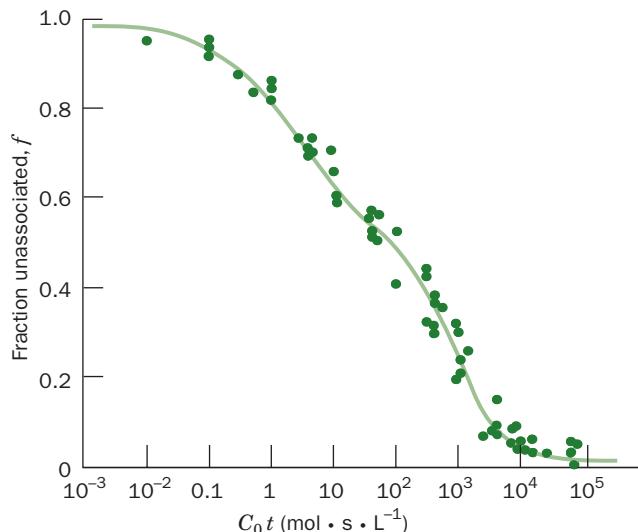


Figure 34-26 The C_0t curve of DNA from *Strongylocentrotus purpuratus* (a sea urchin). [After Galau, G.A., Britten, R.J., and Davidson, E.H., *Cell* 2, 11 (1974).]

same chromosome or on different chromosomes; they constitute $\sim 5\%$ of the human genome), and (5) **unique sequences** (~ 1 copy per haploid genome). The sequences and chromosomal distributions of these DNA segments vary with the species, so a unifying description of their arrangements cannot be made. Nevertheless, several broad generalizations are possible, as we shall see below.

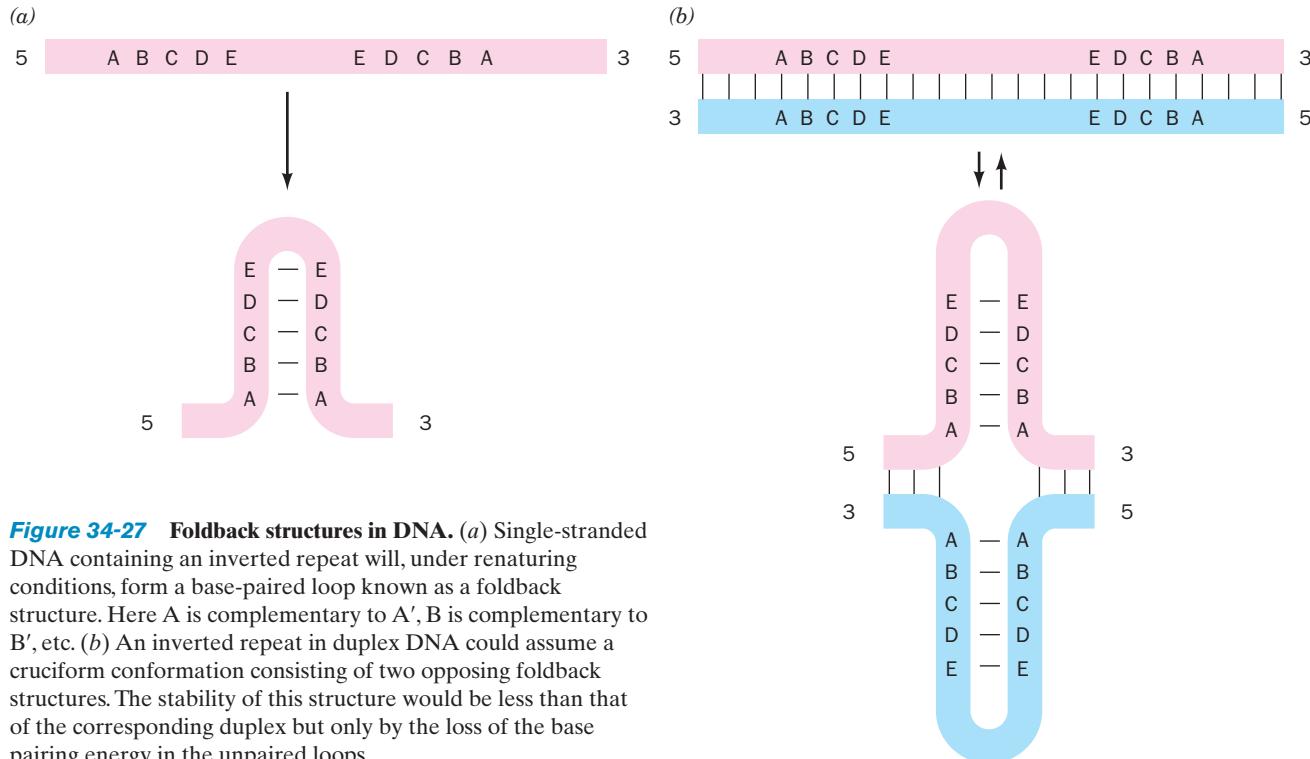


Figure 34-27 Foldback structures in DNA. (a) Single-stranded DNA containing an inverted repeat will, under renaturing conditions, form a base-paired loop known as a foldback structure. Here A is complementary to A', B is complementary to B', etc. (b) An inverted repeat in duplex DNA could assume a cruciform conformation consisting of two opposing foldback structures. The stability of this structure would be less than that of the corresponding duplex but only by the loss of the base pairing energy in the unpaired loops.

a. Inverted Repeats Form Foldback Structures

The most rapidly reassociating eukaryotic DNA, which represents as much as 10% of some genomes, renatures with first-order kinetics. This DNA contains inverted (self-complementary) sequences in close proximity, which can fold back on themselves to form hairpinlike **foldback structures** (Fig. 34-27a). Inverted sequences may be isolated by adsorbing the duplex DNA formed at very low C_0t values to hydroxyapatite and subsequently degrading its single-stranded loop and tails with **S1 nuclease** (an endonuclease from *Aspergillus oryzae* that preferentially cleaves single strands). The resulting inverted repeats range in length from 100 to 1000 bp, sizes much too large to have evolved at random. *In situ* hybridization studies on metaphase chromosomes using these inverted repeats as probes indicate that they are distributed at many chromosomal sites.

The function of inverted repeats, some 2 million of which occur in the human genome, is unknown. However, since the cruciform structures formed by paired foldback structures (Fig. 34-27b) are only slightly less stable than the corresponding normal duplex DNA, it has been suggested that the inverted repeats function in chromatin as some sort of molecular switch.

b. Highly Repetitive DNA Is Clustered at Telomeres and Centromeres

Highly repetitive DNA consists of short sequences that are tandemly repeated, either perfectly or slightly imperfectly, often thousands of times. Such **short tandem repeats** [STRs; alternatively, **simple sequence repeats** (SSRs)] can often be separated from the bulk of the chromosomal

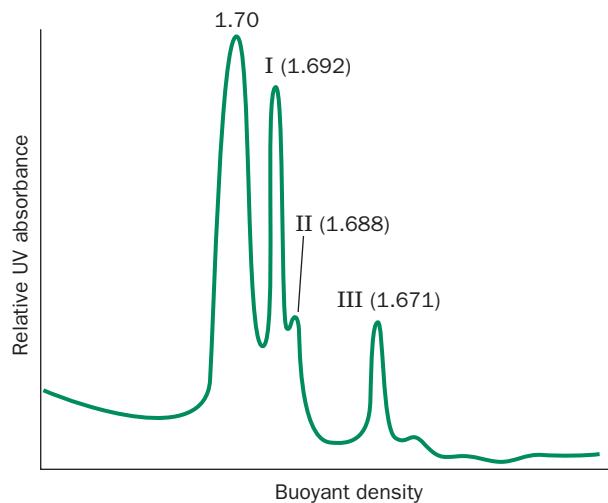
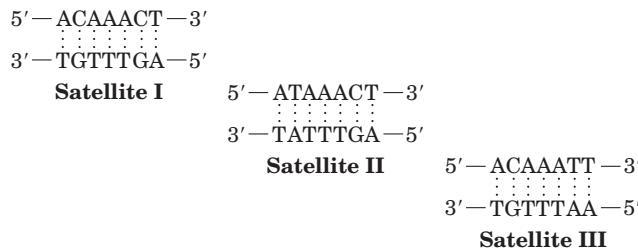


Figure 34-28 The buoyant density pattern of *Drosophila virilis* DNA centrifuged to equilibrium in neutral CsCl. Three prominent bands of satellite DNA ($\rho = 1.692, 1.688$, and 1.671 g/cm^3) are present, in addition to the main DNA band ($\rho = 1.70 \text{ g/cm}^3$). [After Gall, J.G., Cohen, E.H., and Atherton, D.D., *Cold Spring Harbor Symp. Quant. Biol.* **38**, 417 (1973).]

DNA by shear degradation followed by density gradient ultracentrifugation in CsCl since their distinctive base compositions cause them to form “satellites” to the main DNA band (Fig. 34-28; recall that the buoyant density of DNA in CsCl increases with its G + C content; Section 6-6D). The sequences of these STRs, which are also known as **satellite DNAs**, are species specific (STRs with a short repeat unit of $n = 1-13$ nt are often called **microsatellites**, whereas those with $n = 14-500$ are often called **minisatellites**). For example, the crab *Cancer borealis* has an STR comprising 30% of its genome in which the repeating unit is the dinucleotide AT. The DNA of *Drosophila virilis* exhibits three satellite bands (Fig. 34-28), which each consist of a different although closely related repeating heptanucleotide sequence:



These comprise 25, 8, and 8% of the 3.1×10^8 -bp *D. virilis* genome, so that these sequences are repeated 11, 3.6, and 3.6 million times, respectively.

Telomeres, as we have seen (Section 30-4D), consist of G-rich STRs. In addition, the *in situ* hybridization of mouse chromosomes with ^3H -labeled RNA synthesized on mouse simple sequence DNA templates established that STR DNA is concentrated in the heterochromatic region associated with the chromosomal centromere [Fig. 34-29; the

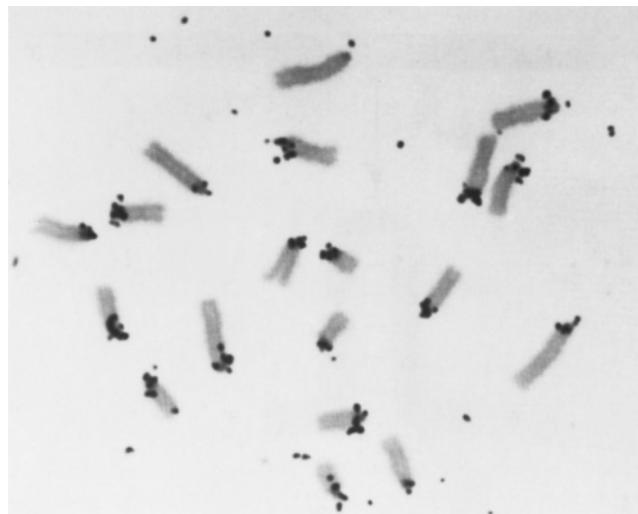


Figure 34-29 Autoradiograph of mouse chromosomes showing the centromeric location of their STR DNA through *in situ* hybridization. Note that the centromeres in mouse chromosomes are all located at one end of the chromosome (no genes lie beyond the mouse centromeres). In human and yeast chromosomes, however, centromeres occupy more internal positions (e.g., Fig. 34-1), whereas *D. melanogaster* has chromosomes of both types. [Courtesy of Joseph Gall, Carnegie Institution of Washington.]

centromere is the constricted segment of the chromosome at which sister chromatids are joined (Fig. 34-1) and at which the chromosome attaches to the mitotic spindle (Fig. 1-19)]. This observation suggests that centromeric STR DNA, which is not transcribed *in vivo*, functions to align homologous chromosomes during meiosis (Fig. 1-20) and/or to facilitate their recombination. This hypothesis is supported by the observation that satellite DNAs are largely or entirely eliminated from the somatic cells of a variety of eukaryotes (which are consequently no longer totipotent) but not from their germ cells.

STRs comprise ~3% of the human genome, with the greatest contribution (0.5%) provided by dinucleotide repeats, most frequently $(\text{CA})_n$ and $(\text{TA})_n$. STRs appear to have arisen by template slippage during DNA replication. This occurs more frequently with short repeats, which therefore have a high degree of length polymorphism in the human population. Consequently, genetic markers based on the lengths of STRs, particularly $(\text{CA})_n$ repeats, have been a mainstay of human genetic studies (Section 5-5Fa).

c. Moderately Repetitive DNAs Are Arranged in Dispersed Repeats

Moderately repetitive DNAs occur in segments of 100 to several thousand base pairs that are interspersed with larger blocks of unique DNA. Some of this repetitive DNA consists of tandemly repeated groups of genes that specify products that cells require in large quantities, such as rRNAs, tRNAs, and histones. The organization of these repeated genes is discussed in Section 34-2D.

Around 42% of the human genome consists of retrotransposons (transposable elements that propagate through the intermediate synthesis of RNA; Section 30-6Bh). Three major types of retrotransposons inhabit the human genome (Table 34-3):

1. Long interspersed nuclear elements (LINEs), which comprise 20.4% of the human genome, are 6- to 8-kb-long segments that encode the proteins that mediate their transposition (Section 30-6Bh), although the vast majority (>99%) of LINEs have accumulated mutations that render them transpositionally inactive. LINEs, which are derived from RNA polymerase II-generated transcripts, are dispersed throughout the genomes of all mammals, which suggests that the ancestral LINE became associated with the mammalian genome very early in its evolution. The most common LINE in the human genome, **LINE-1 (L1)**, consists of ~6.1 kb containing a 5' untranslated region (UTR), two open reading frames (ORFs), the second of which contains a reverse transcriptase gene, and a 3' UTR ending in a poly(A) tail. However, the average L1 has a length of ~900 bp because most L1s consist of truncated fragments of the full-length retrotransposon. Two other LINEs occur in the human genome, **L2** and **L3**, which are distantly related to L1. However, L1 is the only LINE in the human genome that is still transpositionally active.

2. Short interspersed nuclear elements (SINEs), which comprise 13.1% of the human genome, consist of 100- to 400-bp elements that are derived from RNA polymerase

III-generated transcripts. SINEs each contain an RNA polymerase III promoter but, in contrast to LINEs, do not encode proteins; they are apparently propagated by LINE-encoded enzymes. The most common SINEs in the human genome are members of the ***Alu* family**, which are so named because most of their ~300-bp segments contain a cleavage site for the restriction endonuclease *Alu*I (AGCT; Table 5-4). *Alu* elements consist of two imperfect tandem repeats that are ~90% identical in sequence to portions of the 7S RNA of the signal recognition particle (SRP; Section 12-4Ba) but both end in poly(A) segments that are not present in SRP 7S RNA. *Alu* elements occur only in primates, which indicates that they are of relatively recent origin. However, *Alu*-like elements occur in such distantly related organisms as slime molds, echinoderms, amphibians, and birds. All other types of SINEs are derived from tRNA sequences.

3. LTR retrotransposons, which contain long terminal repeats (LTRs) flanking *gag* and *pol* genes, are propagated via cytoplasmic retrovirus-like particles (Section 30-6Bh). They comprise 8.3% of the human genome. Only the vertebrate-specific **endogenous retroviruses (ERVs)** appear to have been active in the mammalian genome.

In addition, the human genome contains **DNA transposons** (Table 34-3) that resemble bacterial transposons (Section 30-6B). They comprise 2.8% of the human genome. Hence, *a total of ~45% of the human genome consists of widely dispersed and almost entirely inactive transposable elements*.

Table 34-3 Moderately Repetitive Sequences in the Human Genome^a

Type of Repeat	Number of Copies ($\times 1000$)	Total Number of Nucleotides (Mb)	Fraction of the Genome Sequence (%)
LINEs	868	559	20.4
L1	516	462	16.9
L2	315	88	3.2
L3	37	8	0.3
SINEs	1558	360	13.1
Alu	1090	290	10.6
MIR	393	60	2.2
MIR3	75	9	0.3
LTR Retrotransposons	443	227	8.3
ERV-class I	112	79	2.9
ERV(K)-class II	8	8	0.3
ERV(L)-class III	83	40	1.4
MaLR	240	100	3.6
DNA Transposons	294	78	2.8
HAT group	195	42	1.6
Tc-1 group	75	32	1.2
Unclassified	22	3.2	0.1
Total		1227	44.8

^aThese numbers are approximate and are likely to be underestimates.

Source: International Human Genome Sequencing Consortium, *Nature* **409**, 880 (2001).

d. Moderately Repetitive DNAs Are Probably Selfish DNA

It would seem likely, considering their ranges of segment lengths and copy numbers, that nonexpressed, moderately repetitive DNAs have several different functions. There is, however, little experimental evidence in support of any of the various proposals that have been put forward in this regard. The proposal that is usually given the most credence is that moderately repetitive DNAs function as control sequences that participate in coordinately activating nearby genes. Another possibility, which is based on the observation that *Alu* elements contain a segment that is homologous to the **papovavirus** replication origin, is that certain families of moderately repetitive DNAs act as DNA replication origins. A third class of proposed functions for moderately repetitive DNAs is that they increase the evolutionary versatility of eukaryotic genomes by facilitating chromosomal rearrangements and/or forming reservoirs from which new functional sequences can be recruited.

Considering both the enormous amount of repetitive DNA in most eukaryotic genomes and the dearth of confirmatory evidence for any of the above proposals, a possibility that must be seriously entertained is that most repetitive DNA serves no useful purpose for its host. Rather, it is **selfish** or **junk DNA**, molecular parasites that, over many generations, have disseminated themselves throughout the genome via various transpositional processes. The theory of natural selection indicates that the increased metabolic burden imposed by the replication of an otherwise harmless selfish DNA would eventually lead to its elimination. Yet, for slowly growing eukaryotes, the relative disadvantage of replicating, say, an additional 1000 bp of selfish DNA in an \sim 1 billion-bp genome would be so slight that its rate of elimination would be balanced by its rate of propagation. The C-value paradox may therefore simply indicate that a significant fraction, if not the great majority, of each eukaryotic genome is selfish DNA.

C. Distribution of Genes

a. A Gene May Be Difficult to Identify

The major goal of the human genome project was to provide a catalog of all human genes and their encoded proteins. Even with the finished sequence now in hand, this is by no means a simple task. A structural gene may be identified as an open reading frame (ORF), a sequence that is not interrupted by Stop codons and that exhibits the same codon-usage preferences as known structural genes in the organism. In organisms with small genomes, such as bacteria and yeast, ORFs are readily identified because these genomes contain relatively little unexpressed DNA. For instance, \sim 89% of the *E. coli* genome consists of expressed sequences, with the remainder consisting of the regulatory sequences that separate individual genes, sites that govern the origin and termination of replication, insertion sequences (Section 30-6Ba), and the inactive remnants of prophages (Section 33-3). Similarly, \sim 70% the yeast genome is expressed. However, *only 1.1 to 1.4% of*

the human genome is expressed, with \sim 24% of the genome consisting of introns and \sim 75% consisting of **intragenic sequences** (untranscribed sequences between genes). Consequently, our incomplete knowledge of the features through which cells recognize genes combined with the fact that human genes consist of relatively short exons (averaging \sim 150 nt) interspersed by much longer introns (averaging \sim 3500 nt and often much longer) greatly increases the difficulty (decreases the signal-to-noise ratio) of identifying genes. Hence, computer programs for sequence-based gene identification have had but limited success. Gene prediction algorithms therefore rely on sequence alignments with **expressed sequence tags (ESTs; cDNAs that have been reverse transcribed from mRNAs; Section 7-2Ba)** together with alignments with known genes from other organisms (which is often successful for highly conserved genes but is less so for genes that are rapidly evolving).

An important clue as to the occurrence of a gene is provided by the presence of a **CpG island**. 5-Methylcytosine (m^5C), as we have seen in Section 30-7, occurs largely in the CG dinucleotides of various eukaryotic palindromic sequences, where it is implicated in switching off gene expression. Since the spontaneous deamination of m^5C yields a normal T and thereby often results in a C · G \rightarrow T · A mutation, CG dinucleotides occur in the human genome at about one-fifth of their randomly expected frequency (which, since human DNA is 42% G + C, is $0.21 \times 0.21 \times 100 = 4\%$ of dinucleotides). Nevertheless, the human genome contains \sim 29,000 \sim 1-kb CpG islands in which unmethylated CG dinucleotides occur at close to their randomly expected frequency. About 56% of human genes are associated with CpG islands, which overlap the promoter regions of these genes and extend up to \sim 1 kb into their coding regions. Hence the presence of a CpG island in a vertebrate chromosome is strongly indicative of the presence of an associated gene.

Around 23,000 putative genes have been identified in the human genome. The discrepancy between this number and previous estimates of 50,000 to 140,000 genes is largely attributed to a much greater prevalence of alternative splicing than had previously been surmised (Section 31-4Am). The gene density along the lengths of the various chromosomes is highly variable. Thus, although the average gene frequency in the human genome is \sim 1 gene per 100 kb of DNA, this value varies from 0 to 64 genes per 100 kb (e.g., Fig. 34-30).

Many genes are transcribed to RNAs that are not translated. These so-called **noncoding RNAs (ncRNAs)** consist of tRNAs, rRNAs, small nuclear RNAs (snRNAs, which are components of spliceosomes; Section 31-4Ah), small nucleolar RNAs (snoRNAs, which participate in nucleolar RNA processing and base modification; Section 31-4Bb), microRNAs (miRNAs; Section 31-4At), as well as a variety of miscellaneous RNAs including the RNA components of the signal recognition particle (Section 12-4Ba), RNase P (Section 31-4Ca), and telomerase (Section 30-4D). The distribution of the rRNA and tRNA genes is discussed in Section 34-2D.

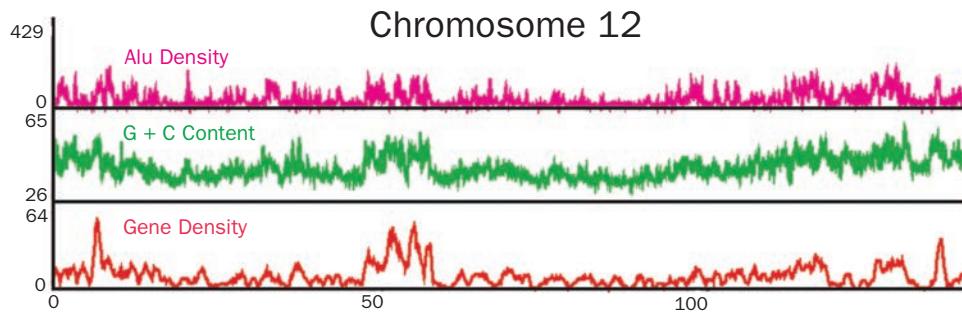


Figure 34-30 Density of structural features along the length of human chromosome 12. Gene density (orange) in this 133-megabase pair (Mb) chromosome is calculated per 1-Mb

window, the percent G + C content (green) is calculated per 100-kb window, and the density of *Alu* elements (magenta) is calculated per 100-kb window. [Courtesy of J. Craig Venter.]

b. A Surprisingly Large Fraction of Genes Have Unknown Functions

The ~23,000 predicted structural genes in the human genome have been classified according to molecular function through sequence comparisons at both the level of protein families and of domains (Fig. 34-31). Note that nearly 42% of them are classified as having unknown functions (and are therefore known as **orphan genes**), as is likewise the case with most other genomes of known sequence, including those of prokaryotes. It can be seen from Fig. 34-31 that the most common molecular functions are those of transcription factors, proteins that mediate nucleic acid metabolism (nucleic acid enzymes), and receptors. Other common functions are those of kinases, hydrolases (most of which are proteases), proto-oncogenes (Section 19-3Bb), and select regulatory proteins (proteins that partici-

pate in signal transduction). About three-quarters of all known human genes appear to have counterparts in other species. About one-quarter are present only in other vertebrates, and one-quarter are present in prokaryotes as well as eukaryotes. As expected, the human genome contains approximately the same number of “housekeeping” genes (genes required for the most fundamental cellular activities) as other eukaryotes. However, comparison of the structural genes in the human genome with those in the genomes of *D. melanogaster* and the nematode worm *Caenorhabditis elegans* reveals that the greatest expansions of gene families occurred in those encoding proteins involved in developmental regulation (Section 34-4B), neuronal structure and function, hemostasis (blood clotting and related processes; Section 35-1), the acquired immune response (Section 35-2), and cytoskeletal complexity.

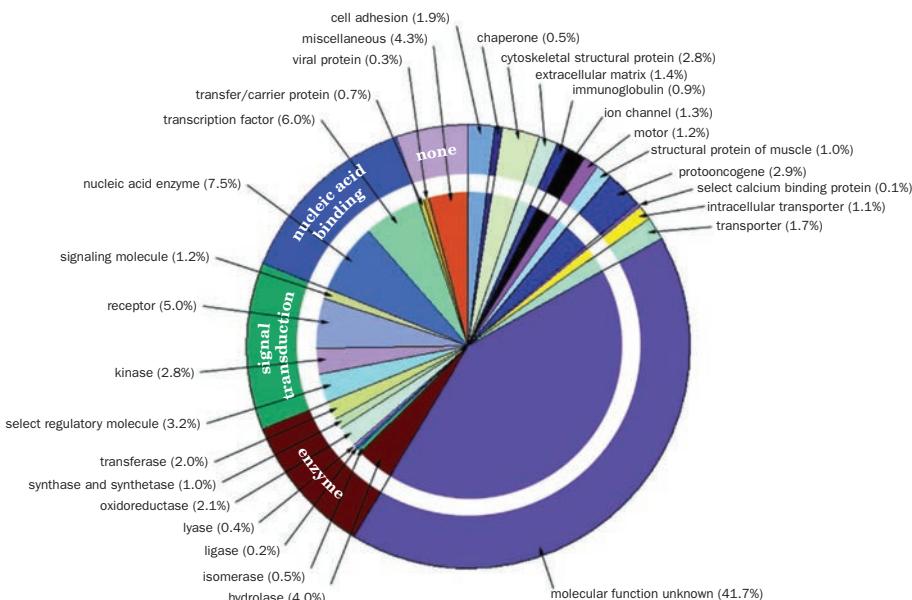


Figure 34-31 Distribution of molecular functions of the putative structural genes in the human genome. Each wedge of this pie chart lists, in parentheses, the percentage of the genes assigned to the indicated category of molecular function. The

outer circle indicates the general functional categories whereas the inner circle provides a more detailed breakdown of these categories. [Courtesy of J. Craig Venter.]

c. A Significant Portion of the Human Genome Is Transcribed to RNA

Despite the fact that the proportion of exonic DNA in the human genome is relatively small, as much as 80% of the genome is transcribed. In addition to the products of 4000 or so known genes for rRNAs, tRNAs, miRNAs, and other small RNAs, these transcripts comprise tens of thousands of other ncRNAs, most of which have no known function. Some ncRNAs are polyadenylated, as are mRNAs (Section 31-4Ab), but do not correspond to any known genes. One possibility is that this RNA represents transcriptional noise arising from promoterlike elements scattered throughout the genome. However, some ncRNAs exhibit conservation between species and tissue-specific splice variants and, moreover, the incidence of ncRNAs increases with developmental complexity. Such observations have led John Mattick to theorize that ncRNAs perform essential regulatory functions through a highly complex network of interactions with the genome that is only beginning to come to light (see, for example, Section 34-4Df). If this turns out to be the case, then there may be no such thing as junk DNA but, instead, the mammalian genome would be better viewed as consisting of islands of protein-encoding segments in a sea of regulatory sequences.

d. Many Disease Genes Have Been Identified

The availability of human genomic data has facilitated the identification of sequence variants that are associated with particular diseases. Over 12,500 such genes are described in the **Online Mendelian Inheritance in Man (OMIM)** database (<http://www.ncbi.nlm.nih.gov/omim/>). However, monogenetic diseases are relatively rare. Most diseases result from interactions among several genes and from environmental factors. One goal of genomics is to identify genetic features that can be linked to susceptibility to disease or infection. To that end, a catalog of human DNA sequence variations has been compiled as a database of **single nucleotide polymorphisms (SNPs; Section 7-2Ca; <http://www.ncbi.nlm.nih.gov/snp/>)**. Over 115 million SNPs, or single-base differences between humans, are known. A SNP between two individuals occurs every \sim 1250 bp on average. Although less than 1% of them result in protein variants, and fewer still have functional consequences, it is becoming increasingly apparent that SNPs are largely responsible for an individual's susceptibility to many diseases as well as to adverse reactions to drugs (side effects; Section 15-4B). Moreover, SNPs can serve as genetic markers for nearby disease-related genes.

D. Tandem Gene Clusters

Most genes occur but once in an organism's haploid genome. This is sufficient, even for genes specifying proteins required in large amounts, through the accumulation of their corresponding mRNAs. However, the great cellular demand for rRNAs (which comprise \sim 80% of a cell's RNA) and tRNAs, which are all ncRNAs, can only be satisfied through the expression of multiple copies of the genes specifying them. In this subsection we discuss the organization

of the genes coding for rRNAs and tRNAs. We shall also consider the organization of histone genes, the only protein-encoding genes that occur in multiple identical copies.

a. rRNA Genes Are Organized into Repeating Sets

We have seen in Sections 31-4B and 31-4C that even the *E. coli* genome, which otherwise consists of unique sequences, contains multiple copies of rRNA and tRNA genes. In eukaryotes, the genes specifying the 18S, 5.8S, and 28S rRNAs are invariably arranged in that order, reading 5' \rightarrow 3' on the RNA strand, and separated by short transcribed spacers to form a single transcription unit of \sim 7500 bp (Fig. 34-32). (Recall that the primary transcript of this gene cluster is a 45S RNA from which the mature rRNAs are derived by post-transcriptional cleavage; Section 31-4Bb.) *Indeed, this rRNA gene arrangement is universal since the 5' end of prokaryotic 23S rRNA is homologous to eukaryotic 5.8S rRNA (Section 32-3Af).*

Electron micrographs, such as Fig. 34-33, indicate that *the blocks of transcribed eukaryotic rRNA genes are arranged in tandem repeats that are separated by untranscribed spacers (Fig. 34-32)*. These tandem repeats are typically \sim 12,000 bp in length, although the untranscribed spacer varies in length between species and, to a lesser extent, from gene to gene. Quantitative measurements of the amounts of radioactively labeled rRNAs that can hybridize with the corresponding nuclear DNA (**rDNA**) and more recently genomic sequencing indicate that these rRNA genes, which may be distributed among several chromosomes, vary in haploid number from less than 50 to over 10,000, depending on the species. Humans, for example, have 150 to 200 blocks of rDNA spread over 5 chromosomes.

b. The Nucleolus Is the Site of rRNA Synthesis and Ribosome Assembly

In a typical interphase cell nucleus, the rDNA condenses to form a single nucleolus (Fig. 1-5). There, as Fig. 34-33 suggests, these genes are rapidly and continuously transcribed by RNA polymerase I (Section 31-2E). The nucleolus, as demonstrated by radioactive labeling experiments, is also the site where these rRNAs are post-transcriptionally processed and assembled with cytoplasmically synthesized ribosomal proteins into immature ribosomal subunits. Final assembly of the ribosomal subunits only occurs as they are being transferred to the cytoplasm,

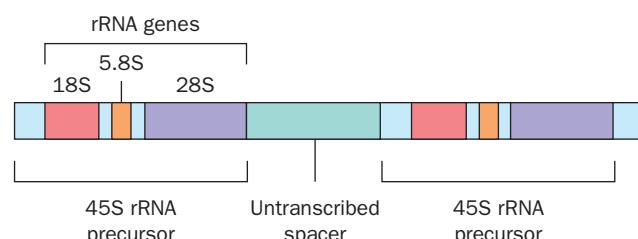


Figure 34-32 The 18S, 5.8S, and 28S rRNA genes are organized in tandem repeats in which sequences encoding the 45S rRNA precursor are interspersed by untranscribed spacers.

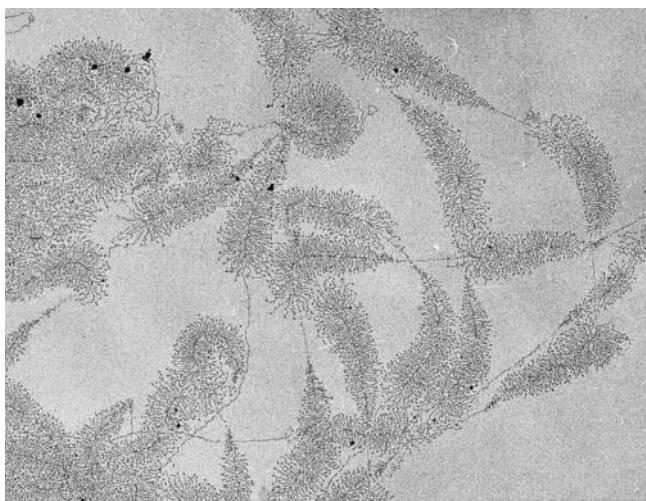


Figure 34-33 Electron micrograph of tandem arrays of actively transcribing 18S, 5.8S, and 28S rRNA genes from the nucleoli of the newt *Notophthalmus viridescens*. The axial fibers are DNA. The fibrillar “Christmas tree” matrices, which consist of newly synthesized RNA strands in complex with proteins, outline each transcriptional unit. Note that the longest ribonucleoprotein branches of each “Christmas tree” are only $\sim 10\%$ the length of their corresponding DNA stem. Apparently, the RNA strands are compacted through secondary structure interactions and/or protein associations. The matrix-free segments of DNA are the untranscribed spacers. [Courtesy of Oscar L. Miller, Jr., and Barbara R. Beatty, University of Virginia.]

which presumably prevents the premature translation of partially processed mRNAs (hnRNAs) in the nucleus.

c. 5S rRNA and tRNA Genes Occur in Multiple Clusters

The genes encoding the 120-nucleotide 5S rRNAs, much like the other rRNA genes, are arranged in clusters that contain a total of several hundred to several hundred thousand tandem repeats distributed among one or more chromosomes. In *X. laevis*, for example, the repeating unit consists of the 5S rRNA gene, a nearby **pseudogene** (a 101-bp segment of the 5S rRNA gene that, curiously, is not transcribed), and an untranscribed spacer of variable length but averaging ~ 400 bp (Fig. 34-34). The 5S rRNA genes are transcribed outside of the nucleolus by RNA polymerase III (Section 31-2E). 5S rRNA must therefore be trans-

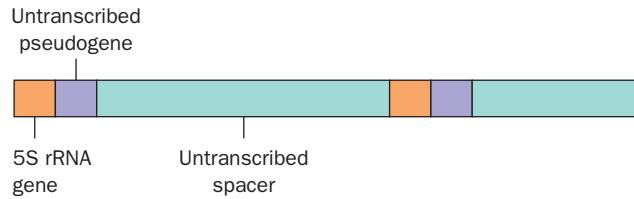


Figure 34-34 The organization of the 5S RNA genes in *Xenopus laevis*. Each of the ~ 750 -nt tandemly repeated units consists of a 5S rRNA gene trailed by an untranscribed spacer in which a pseudogene closely follows the 5S gene.

ported into the nucleolus for incorporation into the large ribosomal subunit.

The 497 tRNA genes that have been identified in the human genome are likewise transcribed by RNA polymerase III. They are also multiply reiterated and clustered, with $>25\%$ of them occurring in a 4-Mb region on chromosome 6 and most of the remainder clustered on many but not all chromosomes.

d. Histone Genes Are Reiterated

Histone mRNAs have relatively short cytoplasmic lifetimes because of their lack of the poly(A) tails that are appended to other eukaryotic mRNAs (Section 31-4Ab). Yet the canonical histones must be synthesized in large amounts during S phase of the cell cycle (when DNA is synthesized). *This process is made possible through the multiple reiteration of histone genes, which in most organisms are the only identically repeated genes that encode proteins.* This organization, it is thought, permits the sensitive control of histone synthesis through the coordinate transcription of sets of histone genes. Histone genes also differ from nearly all other eukaryotic genes in that almost all histone sequences lack introns. The significance of this observation is unknown.

There is little relationship between a genome’s size and its total number of histone genes. For example, birds and mammals have 10 to 20 copies of each of the five histone genes, *D. melanogaster* has ~ 100 , and sea urchins have several hundred. This suggests that the efficiency of histone gene expression varies with species. In many organisms the histone genes are organized into tandemly repeated quintets consisting of a gene coding for each of the five different histones interspersed by untranscribed spacers (Fig. 34-35). The gene order and the direction of transcription in these quintets are preserved over large evolutionary distances. Corresponding spacer sequences vary widely among species and, to a limited extent, among the repeating quintets within a genome. In birds and mammals, this repetitious organization has broken down; their histone genes occur in clusters but in no particular order.

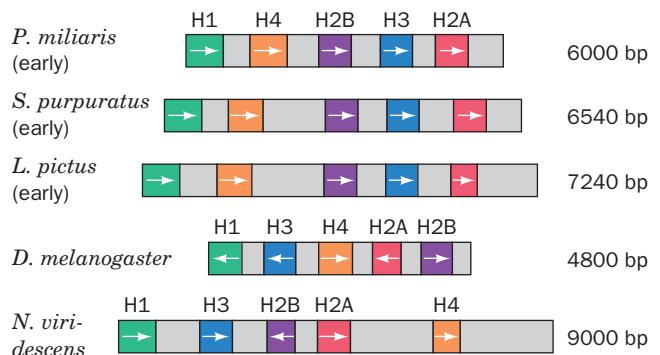


Figure 34-35 The organization and lengths of the histone gene cluster repeating units in a variety of organisms. Coding regions are indicated in color and spacers are gray. The arrows denote the directions of transcription (the top three organisms are distantly related sea urchins).

e. Reiterated Sequences May Be Generated and Maintained by Unequal Crossovers and/or Gene Conversion

How do reiterated genes maintain their identity? The usual mechanism of Darwinian selection would seem ineffective in accomplishing this since deleterious mutations in a few members of a multiply repeated set of identical genes would have little phenotypic effect. Indeed, many mutations do not affect the function of a gene product and are therefore selectively neutral. Reiterated gene sets must therefore maintain their homogeneity through some additional mechanism. Two such mechanisms seem plausible:

1. In the **unequal crossover** mechanism (Fig. 34-36a), recombination occurs between homologous segments of misaligned chromosomes, thereby excising a segment from one of the chromosomes and adding it to the other. Computer simulations indicate that such repeated expansions and contractions of a chromosome will, by random processes, generate a cluster of reiterated sequences that have been derived from a much smaller ancestral cluster. Unequal crossing-over is also thought to be the mechanism that generated segmental duplications (Section 34-2B).

2. In the **gene conversion** mechanism (Fig. 34-36b), one member of a reiterated gene set “corrects” a nearby variant through a process resembling recombination repair (Section 30-6A).

Since point mutations are rare events compared to crossovers, either mechanism would eventually result in a newly arisen variant copy of a repeated sequence either being eliminated or taking over the entire cluster. If a mutation that has been so concentrated is deleterious, it will be eliminated by Darwinian selection. In contrast, variant spacers, which are not as subject to selective pressure, would be eliminated at a slower rate. The existence of reiterated sets of identical genes separated by somewhat heterogeneous spacers may therefore be reasonably attributed to either homogenization model.

E. Gene Amplification

The selective replication of a particular set of genes, a process known as **gene amplification**, normally occurs only at specific stages of the life cycle of certain organisms. In the following subsections, we outline what is known about this phenomenon.

a. rRNA Genes Are Amplified During Oogenesis

The rate of protein synthesis during the early stages of embryonic growth is so great that in some species the normal genomic complement of rRNA genes cannot satisfy the demand for rRNA. In these species, notably certain insects, fish, and amphibians, the rDNA is differentially replicated in developing oocytes (immature egg cells). In one of the most spectacular examples of this process, the rDNA in *X. laevis* oocytes is amplified by ~1500 times its amount in somatic cells to yield some 2 million sets of rRNA genes

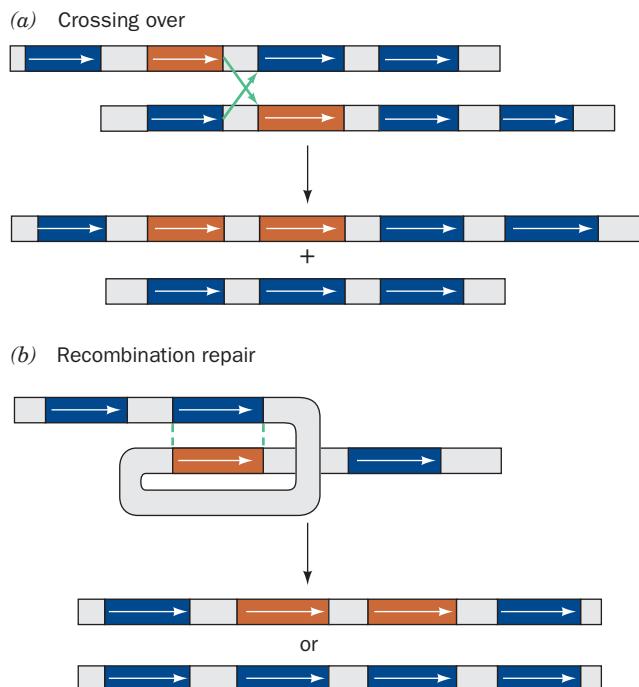


Figure 34-36 Two possible mechanisms for maintaining the homogeneity of a tandem multigene family. (a) Unequal crossing-over between mispaired but similar genes results in an unpaired DNA segment being deleted from one chromosome and added to the other. (b) Gene conversion “corrects” one member of a tandem array with respect to the other via a recombination repair mechanism. Repeated cycles of either process may either eliminate a variant gene or spread it throughout the entire tandem array.

comprising nearly 75% of the total cellular DNA. The amplified rDNA occurs as extrachromosomal circles, each containing one or two transcription units that are organized into hundreds of nucleoli (Fig. 34-37). Mature *Xenopus* oocytes therefore contain $\sim 10^{12}$ ribosomes, 200,000 times the number in most larval cells. This is so many that mutant zygotes (fertilized ova) that lack nucleoli (and thus cannot synthesize new ribosomes; the oocyte’s extra nucleoli are destroyed during its first meiotic division) survive to the swimming tadpole stage with only their maternally supplied ribosomes.

What is the mechanism of rDNA amplification? An important clue is that the untranscribed spacers from a given extrachromosomal nucleolus all have the same length, whereas we have seen that the corresponding chromosomal spacers exhibit marked length heterogeneities. This observation suggests that the rDNA circles in a single nucleolus are all descended from a single chromosomal gene. Gene amplification has been shown to occur in two stages: A low level of amplification in the first stage followed by massive amplification in the second stage. It therefore seems likely that, in the first stage, no more than a few chromosomal rRNA genes are replicated by an unknown mechanism and the daughter strands are released as extrachromosomal

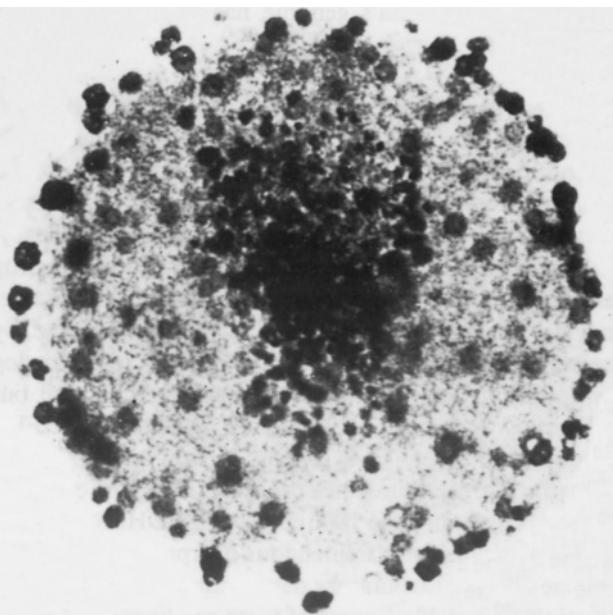


Figure 34-37 Photomicrograph of an isolated oocyte nucleus from *X. laevis*. Its several hundred nucleoli, which contain amplified rRNA genes, appear as darkly staining spots. [Courtesy of Donald Brown, Carnegie Institution of Washington.]

circles. Then, in the second stage, these circles are multiply replicated by the rolling circle mechanism (Section 30-3Bb). In support of this hypothesis are electron micrographs of amplified genes showing the “lariat” structures postulated to be rolling circle intermediates (Fig. 30-26).

b. Chorion Genes Are Amplified

The only other known example of programmed gene amplification is that of the *D. melanogaster* ovarian follicle cell genes that code for **chorion** (egg shell) **proteins** (ovarian follicle cells surround and nourish the maturing egg). Prior to chorion synthesis, the entire haploid genome of each ovarian follicle cell is replicated 16-fold. This process is followed by an ~10-fold selective replication of only the

chorion genes to form a multiply branched (partially polytene) structure in which the amplified chorion genes remain part of the chromosome (Fig. 34-38). Interestingly, chorion gene amplification does not occur in silk moth oocytes. Rather, this organism’s genome has multiple copies of chorion genes.

c. Drug Resistance Can Result from Gene Amplification

In cancer chemotherapy, a common observation is that the continued administration of a cytotoxic drug causes an initially sensitive tumor to become increasingly drug resistant to the point that the drug loses its therapeutic efficacy. One mechanism by which a cell line can acquire such drug resistance is through the overproduction of the drug’s target enzyme. Such a process can be observed, for example, by exposing cultured animal cells to the dihydrofolate analog methotrexate. This substance, it will be recalled, all but irreversibly binds to dihydrofolate reductase (DHFR), thereby inhibiting DNA synthesis (Section 28-3Be). Slowly increasing the methotrexate dose yields surviving cells that ultimately contain up to 1000 copies of the DHFR gene and are thereby capable of tremendous overproduction of this enzyme, a clear laboratory demonstration of Darwinian selection. Members of some of these cell lines contain extrachromosomal elements known as **double minute chromosomes** that each bear one or more copies of the DHFR gene, whereas in other cell lines the additional DHFR genes are chromosomally integrated. The mechanism of gene amplification in either cell type is not well understood, although it is worth noting that this phenomenon is only known to occur in cancer cells. Both types of amplified genes are genetically unstable; further cell growth in the absence of methotrexate results in the gradual loss of the extra DHFR genes.

F. Clustered Gene Families: Hemoglobin Gene Organization

Few proteins in a given organism are really unique. Rather, like the digestive enzymes trypsin, chymotrypsin, and elastase (Section 15-3), or the various collagens (Section 8-2B), they are usually members of families of structurally and functionally related proteins. In many cases, the family of genes specifying such proteins are clustered together in a single chromosomal region. In the following subsections, we consider the organization of two well-characterized clustered gene families, those encoding the two types of human hemoglobin subunits. The clustered gene families that encode immune system proteins are discussed in Section 35-2C.

a. Human Hemoglobin Genes Are Arranged in Two Developmentally Ordered Clusters

Human adult hemoglobin (HbA) consists of $\alpha_2\beta_2$ tetramers in which the α and β subunits are structurally related (Section 10-2B). The first hemoglobin made by the human embryo, however, is a $\zeta_2\epsilon_2$ tetramer (**Hb Gower 1**) in which ζ and ϵ are α - and β -like subunits, respectively

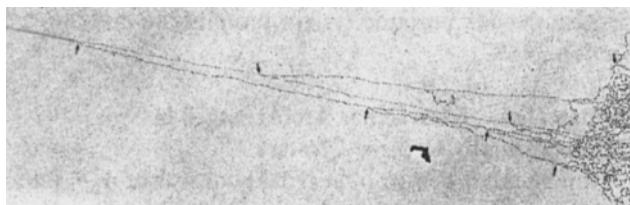


Figure 34-38 Electron micrograph of a chorion gene-containing chromatin strand from an oocyte follicle cell of *D. melanogaster*. The strand has undergone several rounds of partial replication (arrows at replication forks) to yield a multiforked structure containing several parallel copies of chorion genes. [Courtesy of Oscar L. Miller, Jr. and Yvonne Osheim, University of Virginia.]

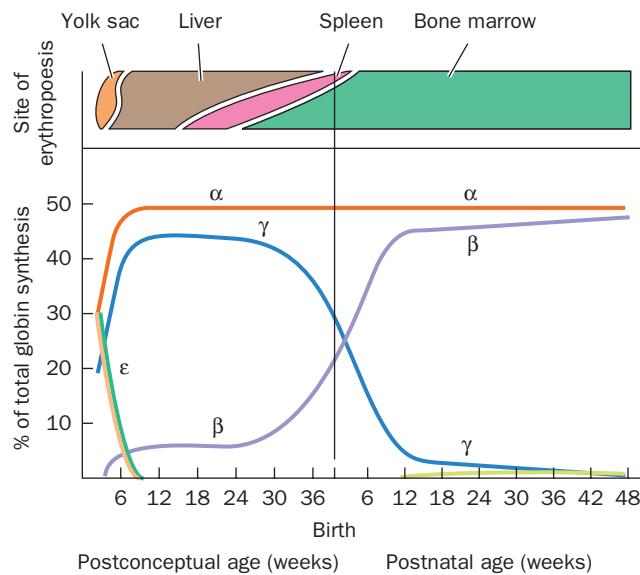


Figure 34-39 The progression of human globin chain synthesis with embryonic and fetal development. Note that any red blood cell contains only one type each of α - and β -like subunits. The progression in the sites of erythropoiesis (red cell formation), which is indicated in the upper panel, corresponds roughly to the major switches in hemoglobin types. [After Weatherall, D.J. and Clegg, J.B., *The Thalassaemia Syndromes* (3rd ed.), p. 64, Blackwell Scientific Publications (1981).]

(Fig. 34-39; their evolutionary relationships are diagrammed in Fig. 7-25). By around 8 weeks postconception, the embryonic subunits have been supplanted (in newly formed erythrocytes) by the α subunit and the β -like γ subunit to form fetal hemoglobin (HbF), $\alpha_2\gamma_2$ (the hemoglobins present during the changeover period, $\alpha_2\epsilon_2$ and $\zeta_2\gamma_2$, are named **Hb Gower 2** and **Hb Portland**, respectively). The γ subunit is gradually superseded by β starting a few weeks before birth. Adult blood normally contains $\sim 97\%$ HbA, 2% **HbA₂** ($\alpha_2\delta_2$ in which δ is a β variant), and 1% HbF.

In mammals, the genes specifying the α - and β -like hemoglobin subunits form two different gene clusters that occur on separate chromosomes. In humans and many other mammals, the genes in each globin cluster are arranged, 5' \rightarrow 3' on the coding strands, in the order of their developmental expression (Fig. 34-40). This ordering is common in

mammals but not universal; in the mouse β gene cluster, for instance, the adult genes precede the embryonic genes.

The β -globin gene cluster (Fig. 34-40), which spans >60 kb, contains five functional genes: the embryonic ϵ gene, two fetal genes, $^G\gamma$ and $^A\gamma$ (duplicated genes that encode polypeptides differing only by having either Gly or Ala at their positions 136), and the two adult genes, δ and β . The β -globin cluster also contains one **pseudogene**, $\Psi\beta$ (an untranscribed relic of an ancient gene duplication that is $\sim 75\%$ identical to the β gene; the human genome contains $\sim 20,000$ pseudogenes), eight *Alu* elements, and two *L1* elements (Section 34-2Bc).

The α -globin gene cluster (Fig. 34-40), which spans ~ 28 kb, contains three functional genes: the embryonic ζ gene and two slightly different α genes, α_1 and α_2 , which encode identical polypeptides. The α cluster also contains four pseudogenes, $\Psi\zeta$, $\Psi\alpha_2$, $\Psi\alpha_1$, and $\Psi\theta$, and three *Alu* elements.

b. Hemoglobin Genes All Have the Same Exon-Intron Structure

Protein-coding sequences represent $<5\%$ of either globin gene cluster. This situation is largely a consequence of the heterogeneous collection of untranscribed spacers separating the genes in each cluster. In addition, *all known vertebrate globin genes, including that of myoglobin and most hemoglobin pseudogenes, consist of three nearly identically placed exons separated by two somewhat variable introns* (Fig. 34-41). This gene structure apparently arose quite early in vertebrate history, well over 500 million years ago. Indeed, much of this structure even predates the divergence of plants and animals. The structure of the gene encoding leghemoglobin (a plant globin that functions in legumes to protect nitrogenase from O_2 poisoning; Section 26-6c) differs from that of vertebrates only in that the central exon of vertebrate globins is split by a third intron in the leghemoglobin gene. Quite possibly the central exon in vertebrate globins arose through the fusion of the two interior exons in a leghemoglobin-like ancestral gene.

c. DNA Polymorphisms Can Establish Genealogies

Unexpressed sequences, which are subject to little selective pressure, evolve so much faster than expressed sequences that they accumulate significant numbers of sequence **polymorphisms** (variations) within a single species.

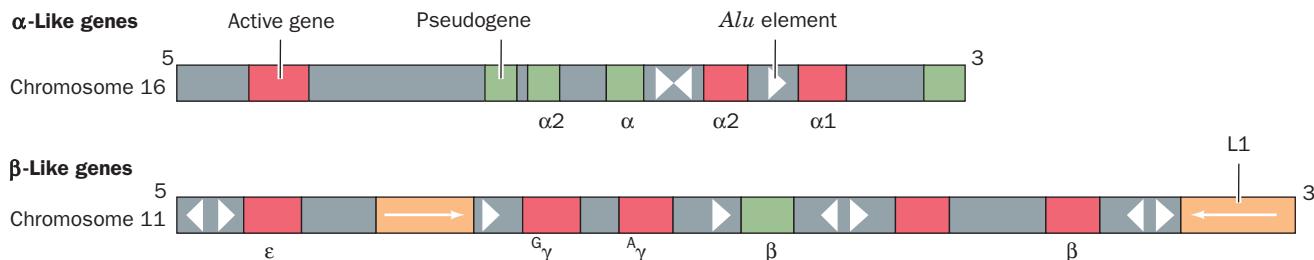


Figure 34-40 The organization of human globin genes on their respective sense strands. Red boxes represent active genes; green boxes represent pseudogenes; yellow boxes represent L1 sequences, with the arrows indicating their relative orientations;

and triangles represent *Alu* elements in their relative orientations. [After Karlsson, S. and Nienhuis, A.W., *Annu. Rev. Biochem.* **54**, 1074 (1985).]

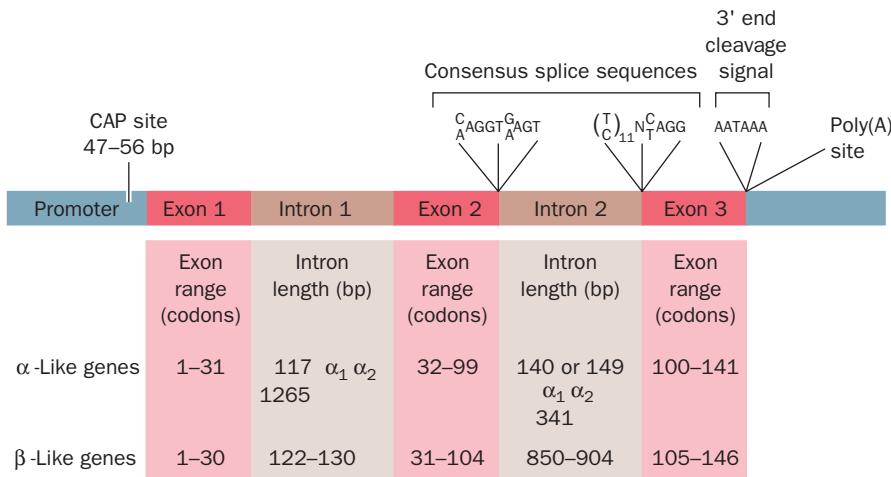


Figure 34-41 Structure of the prototypical hemoglobin gene. The conserved sequences at the exon–intron boundaries (splice sequences) and at the 3' end of the gene (polyadenylation site) are indicated. The length range of each exon (in codons) and each intron (in base pairs) is given. [Courtesy J. Craig Venter]

Consequently, the evolutionary relationships among populations within a species can be established by determining how a series of polymorphic DNA sequences are distributed among them. For example, Masatoshi Nei inferred the genealogy of 26 representative human populations from

around the world from the sequence variations in 29 polymorphic loci. The resulting phylogenetic tree (Fig. 34-42) indicates that non-African (Eurasian) populations are

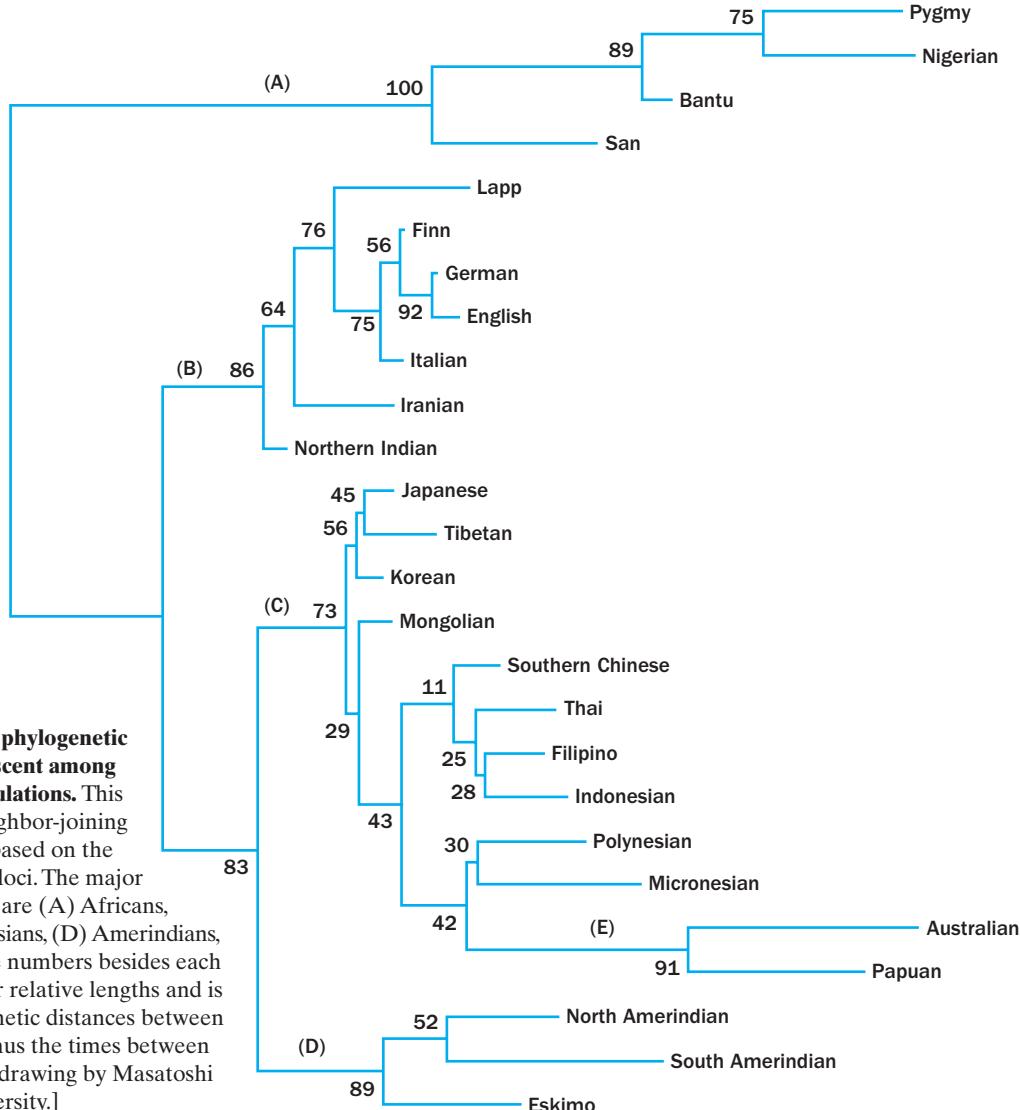


Figure 34-42 An unrooted phylogenetic tree indicating the lines of descent among 26 representative human populations. This tree was generated by the neighbor-joining (NJ) method (Section 7-4C) based on the sequences of 29 polymorphic loci. The major groups of human populations are (A) Africans, (B) Caucasians, (C) Greater Asians, (D) Amerindians, and (E) Australopapuans. The numbers besides each of the branches indicates their relative lengths and is therefore indicative of the genetic distances between the related populations and thus the times between their divergence. [Based on a drawing by Masatoshi Nei, Pennsylvania State University.]

much more closely related to each other than they are to African populations. Fossil evidence indicates that anatomically modern man arose in Africa ~200,000 years ago and spread throughout that continent. The phylogenetic tree therefore suggests that all Eurasian populations are descended from a surprisingly small “founder population” (perhaps only a few hundred individuals) that left Africa ~100,000 years ago. A similar analysis indicates that the sickle-cell variant of the β gene arose on at least three separate occasions in geographically distinct regions of Africa.

G. The Thalassemias: Genetic Disorders of Hemoglobin Synthesis

The study of mutant hemoglobins (Section 10-3) provided invaluable insights into structure–function relationships in proteins. Likewise, the study of defects in hemoglobin expression greatly facilitated our understanding of eukaryotic gene expression.

The most common class of inherited human disease results from the impaired synthesis of hemoglobin subunits. These anemias are named **thalassemias** (Greek: *thalassa*, sea) because they commonly occur in the region surrounding the Mediterranean Sea (although they are also prevalent in Central Africa, India, and the Far East). The observation that malaria is or was endemic in these same areas (Fig. 7-21) led to the realization that heterozygotes for thalassemic genes (who appear normal or are only mildly anemic; a condition known as **thalassemia minor**) are resistant to malaria. Thus, as we saw in our study of sickle-cell anemia (Section 7-3Ab), mutations that are seriously debilitating or even lethal in homozygotes (who are said to suffer from **thalassemia major**) may offer sufficient selective advantage to heterozygotes to ensure the propagation of the mutant gene.

Thalassemia can arise from many different mutations, each of which causes a disease state of characteristic severity. In α^0 - and β^0 -thalassemias, the indicated globin chain is absent, whereas in α^+ - and β^+ -thalassemias, the normal globin subunit is synthesized in reduced amounts. In what follows, we shall consider thalassemias that are illustrative of several different types of genetic lesions.

a. α -Thalassemias

Most α -thalassemias are caused by the deletion of one or both of the α -globin genes in an α gene cluster (Fig. 34-40). A variety of such mutations have been cataloged. In the absence of equivalent numbers of α chains, the fetal γ chains and the adult β chains form homotetramers: **Hb Bart's** (γ_4) and **HbH** (β_4). Neither of these tetramers exhibits any Bohr effect or cooperativity (Sections 10-1C and 10-2C), which makes their oxygen affinities so high that they cannot release oxygen under physiological conditions. Consequently, α^0 -thalassemia occurs with four degrees of severity depending on whether an individual has 1, 2, 3, or 4 missing α -globin genes:

1. Silent-carrier state: The loss of one α gene is an asymptomatic condition. The rate of expression of the re-

maining α genes largely compensates for the less than normal α gene dosage so that, at birth, the blood contains only ~1 to 2% Hb Bart's.

2. α -Thalassemia trait: With two missing α genes (either one each deleted from both α gene clusters or both deleted from one cluster), only minor anemic symptoms occur. The blood contains ~5% Hb Bart's at birth.

3. Hemoglobin H disease: Three missing α genes results in a mild to moderate anemia. Affected individuals can usually lead normal or nearly normal lives.

4. Hydrops fetalis: The lack of all four α genes is lethal. Unfortunately, the synthesis of the embryonic ζ chain continues well past the 8 weeks postconception when it normally ceases (Fig. 34-39), so the fetus usually survives until around birth.

α -Thalassemias caused by nondeletion mutations are relatively uncommon. One of the best characterized such lesions changes the UAA Stop codon of the $\alpha 2$ -globin gene to CAA (a Gln codon), so that protein synthesis continues for the 31 codons beyond this site to the next UAA. The resultant **Hb Constant Spring** is produced in only small amounts because, for unknown reasons, its mRNA is rapidly degraded in the cytosol. Another point mutation in the $\alpha 2$ gene changes Leu H8(125) α to Pro, which presumably disrupts the H helix. The consequent α^+ -thalassemia results from the rapid degradation of this abnormal **Hb Quong Sze**.

b. β -Thalassemias

Heterozygotes of β -thalassemias are usually asymptomatic. Homozygotes become so severely anemic, however, that once their HbF production has diminished, many require frequent blood transfusions to sustain life and all require them to prevent the severe skeletal deformities caused by bone marrow expansion. The anemia results not only from the lack of β chains but also from the surplus of α chains. The latter form insoluble membrane-damaging precipitates that cause premature red cell destruction (Section 10-3A). The coinheritance of α -thalassemia therefore tends to lessen the severity of β -thalassemia major.

In β -thalassemia, there may be an increased production of the δ and γ chains, so that the consequent extra HbA₂ and HbF can compensate for some of the missing HbA. In **$\delta\beta$ -thalassemia**, the neighboring δ and β genes have both been deleted, so that only increased production of the γ chain is possible. Yet many adult $\delta\beta$ -thalassemics, for reasons that are not understood, produce so much HbF that they are asymptomatic. Such individuals are said to have **hereditary persistence of fetal hemoglobin (HPFH)**. This condition is therefore of medical interest because it could also alleviate the symptoms of β -thalassemia and sickle-cell anemia.

The so-called Greek form of HPFH is associated with a G → A mutation at position -117 of the γ globin gene (its promoter region). In an effort to establish whether this mutation does, in fact, cause HPFH, the mutated γ globin gene was introduced into mice. The resulting fetal and adult

transgenic animals synthesized γ globin at a high level, with a concomitant decrease in the synthesis of the β globin gene. These changes in gene expression correlate with the loss of binding of the transcription factor **GATA-1** to the γ globin promoter, thereby suggesting that this protein is a negative regulator of the γ globin gene expression in normal human adults (transcription factors are discussed in Section 34-3B).

β^0 -Thalassemias caused by deletions are rare compared to those causing α^0 -thalassemias. This is probably because the long repeated sequences in which the α -globin genes are embedded make them more prone to unequal crossing-over than the β globin gene. Nevertheless, a β -thalassemic lesion causing the production of **Hb Lepore** is a particularly clear instance of this deletion mechanism. This lesion, the consequence of a deletion extending from within the δ gene to the corresponding position of its neighboring β gene, yields a δ/β hybrid subunit. Such deletions almost certainly arose through unequal crossovers between the β gene on one chromosome and the δ gene on another (Fig. 34-43; the two genes are 93% identical in sequence). The second product of such crossovers, a chromosome containing a β/δ hybrid flanked by normal δ and β genes (Fig. 34-43), is known as **Hb anti-Lepore**. Homozygotes for Hb Lepore have symptoms similar to those of β -thalassemia major, whereas homozygotes for Hb anti-Lepore, which have the full complement of normal globin genes, are symptom free and have only been detected through blood tests.

Most β -thalassemias are caused by a wide variety of point mutations that affect the production of β chains. These include the following:

1. Nonsense mutations that convert normal codons to the Stop codon UAG.
2. Frameshift mutations that insert/delete one or more base pairs into/from an exon.
3. Point mutations in the β gene's promoter region, either in its TATA box or in its CACCC box (Section 31-2Ee). These attenuate transcriptional initiation.

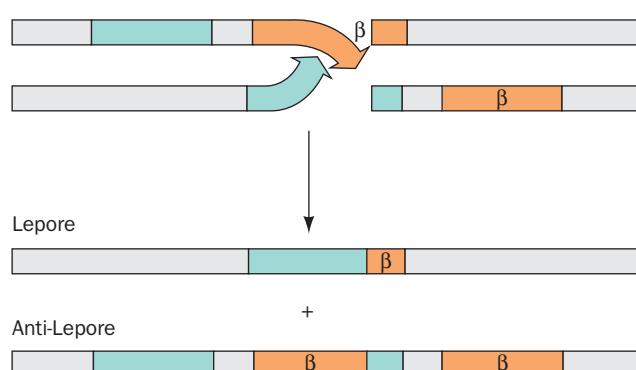


Figure 34-43 The formation of Hb Lepore and Hb anti-Lepore. This occurs by unequal crossing-over between the β -globin gene on one chromosome and the δ -globin gene on its homolog.

4. Point mutations that alter the sequence at an exon-intron junction (Section 31-4Ad). These diminish/abolish splicing and/or activate a **cryptic splice site** (an exon-intron junction-like sequence that normally is not spliced) to pair with the altered intron's unaltered end.

5. A point mutation that alters an intron's lariat-branch site (Section 31-4Ad). This activates a cryptic 3' splice site upstream of the original site, leading to the excision of a shorter than normal intron.

6. Point mutations that create new splice sites. These either compete with the neighboring normal splice site or pair with a nearby cryptic splice site.

7. A point mutation that alters the AAUAAA cleavage signal at the mRNA's 3' end (Section 31-4Ab).

Consideration of the effects of these mutations, particularly those involving gene splicing, has confirmed and extended our understanding of how eukaryotic genes are constructed and expressed.

3 CONTROL OF EXPRESSION

The elucidation of the mechanisms controlling gene expression in eukaryotes had lagged at least 20 years behind those of prokaryotes. In addition to the far greater complexity of eukaryotic systems, this is largely because the types of genetic analyses that have been so useful in characterizing prokaryotic systems (which require the detection of very rare events) are precluded in metazoa (multicellular animals) by their much slower reproductive rates. Compounding this problem are the difficulties in selecting for mutations in essential genes; the missing product of a defective enzyme in a metazoan usually cannot be replaced by simply adding that product to the diet as is often possible with, say, *E. coli*. This latter difficulty can be partially overcome by the rather laborious task of growing cells from metazoa in tissue culture. Since somatic cells do not normally undergo genetic recombination, however, genetic manipulations cannot be carried out in tissue culture the way they can in a bacterial culture.

What has made genetic manipulations of metazoa feasible is the development, in the 1970s, of molecular cloning techniques (Section 5-5). The gene encoding a particular eukaryotic protein can be identified in genomic or cDNA libraries through Southern blotting (Section 5-5D), PCR (Section 5-5F) using an oligonucleotide probe or primer whose sequence encodes a segment of the protein (a process termed reverse genetics; Section 7-2D), or on microarrays (Section 7-6B). Alternatively, if the organism's genome has been sequenced, the gene may be identified *in silico* (computationally). The gene may then be modified, for example, through site-directed mutagenesis (Section 5-5Gc), and the effects of the modification analyzed in an expression vector such as *E. coli* or yeast, or alternatively, *in vitro*.

The expression of foreign genes in metazoa (a gain of function) has been made possible through the development

of a process in which DNA is microinjected into the nucleus of a fertilized ovum (Fig. 5-58). Such DNA often integrates into the chromosome of the resulting zygote, that then undergoes normal development to form a **transgenic** individual whose cells each contain the foreign genes (in *Xenopus*, this merely involves allowing the transfected egg to hatch, whereas in mice the fertilized ovum must be implanted in the uterus of a properly prepared foster mother; see Fig. 5-5 for a striking example of a transgenic mouse). Alternatively, a normal gene may be selectively inactivated (knocked out; a loss of function) through the use of DNA specifying the defective gene, which then recombines with the normal gene. The genome of an already multicellular organism may be altered, in a technique that holds great promise for gene therapy, through the use of defective (unable to reproduce) retroviruses that contain the genes to be transferred (Section 5-5Hb). Thus, the genomes of metazoans can now be manipulated, albeit with considerable clumsiness. We are, however, becoming more adept at these procedures as we gain further understanding of how eukaryotic chromosomes are organized and expressed.

Single-celled eukaryotes, particularly yeasts, are exceptions to the foregoing discussion because they can be grown and manipulated in much the same way as bacteria. Indeed, much of our knowledge of eukaryotic molecular biology has been obtained through molecular genetic analyses of budding (baker's) yeast (*Saccharomyces cerevisiae*). In this chapter, unless otherwise indicated, the term "yeast" refers to *S. cerevisiae*.

In this section we consider the molecular basis of the enormous expressional variation that eukaryotic cells exhibit. In doing so, we shall first study the nature of transcriptionally active chromatin, then discuss how genetic expression in eukaryotes is mainly regulated through the control of transcriptional initiation, and finally consider the other means by which eukaryotes control genetic expression. Eukaryotic gene regulation, as we shall see, is an astoundingly complex process that requires the participation of well over 100 polypeptides that form assemblies with molecular masses of several million daltons. In the following section, we take up the molecular basis of normal cell differentiation, its aberration, cancer, and programmed cell death.

A. Chromosomal Activation and Deactivation

Interphase chromatin, as is mentioned in Section 34-1, may be classified in two categories: the highly condensed and transcriptionally inactive heterochromatin, and the diffuse and transcriptionally active or activatable euchromatin (Fig. 34-2). Two types of heterochromatin have been distinguished:

1. Constitutive heterochromatin, which is permanently condensed in all cells and consists mostly of the highly repetitive sequences clustered near the chromosomal centromeres and telomeres (Section 34-2Bb). Constitutive heterochromatin is transcriptionally inert.

2. Facultative heterochromatin, which varies in a tissue-specific manner. Presumably the condensation of faculta-

tive heterochromatin functions to transcriptionally inactivate large chromosomal blocks.

a. Mammalian Cells Have Only One Active X Chromosome

Female mammalian cells contain two X chromosomes, whereas male cells have one X and one Y chromosome. *Female somatic cells, however, maintain only one of their X chromosomes in a transcriptionally active state.* Consequently, males and females make approximately equal amounts of X chromosome-encoded gene products, a phenomenon known as **dosage compensation**. The inactive X chromosome is visible during interphase as a heterochromatin structure known as a **Barr body** (Fig. 34-44). In marsupials (pouched mammals), the Barr body is always the paternally inherited X chromosome, an epigenetic phenomenon (Section 30-7c). In placental mammals, however, one randomly selected X chromosome in every somatic cell is inactivated when the embryo consists of 32 to 64 cells. The progeny of each of these cells epigenetically maintain the same inactive X chromosome. *Female placental mammals are therefore mosaics composed of clonal groups of cells in which the active X chromosome is either paternally or maternally inherited.* This situation is particularly evident in human females who are heterozygotes for the X-linked congenital sweat gland deficiency **anhidrotic ectodermal dysplasia**. The skin of these women consists of patches lacking sweat glands, in which only the X chromosome containing the mutant gene is active, alternating with normal patches in which only the other X chromosome is active. Similarly, calico cats, whose coats have patches of black fur and orange fur, are almost always females whose

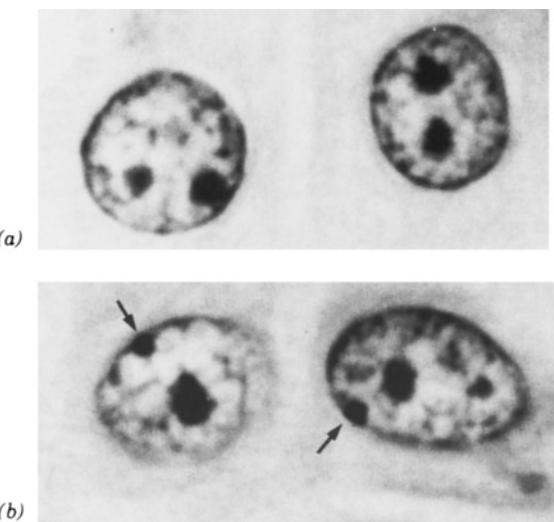


Figure 34-44 Photomicrographs of stained nuclei from human oral epithelial cells. (a) From a normal XY male showing no Barr body. (b) From a normal XX female showing a single Barr body (arrow). The presence of Barr bodies permits the rapid determination of an individual's chromosomal sex. [From Moore, K.L. and Barr, M.L., *Lancet* 2, 57 (1955). Reprinted with permission from Elsevier.]



Figure 34-45 A calico cat. The orange and black patches of fur are manifestations of the expression of different alleles in the X chromosome of each patch. The white fur is due to the effects of an autosomal gene. [© Hank Delespinasse/Age Fotostock America, Inc.]

two X chromosomes contain different alleles of a gene specifying fur color (Fig. 34-45).

The X chromosome contains the **X-inactivation center (XIC)**, which has multiple regulatory sites together with the *Xist* gene, which occurs only in placental mammals and is transcribed only from the inactive X chromosome (Xi). The 25-kb *Xist* RNA binds at or near XIC where it induces the heterochromatization of that region of Xi via a process described below. The resulting heterochromatin recruits additional *Xist* RNA to a neighboring site, which in turn, causes its heterochromatization. This process propagates in both directions from XIC until Xi is “painted” over its whole length by *Xist* RNA and the entire chromosome is heterochromatized and thereby transcriptionally inactivated at all sites but *Xist*. XIC is necessary and sufficient to initiate X inactivation as is demonstrated by the observations that the insertion of XIC into an autosome results in its inactivation and X chromosomes that lack XIC are not inactivated.

Localized *Xist* RNA does not by itself inactivate Xi. However, *Xist* RNA recruits enzyme systems that derivatize histones. In comparison to the active X chromosome (Xa), Xi has low levels of H3 Lys 4 methylation, high levels of H3 Lys 9 methylation, low levels of histone acetylation, and high levels of DNA methylation, all of which are associated with gene inactivation (histone modification and DNA methylation are discussed in Sections 34-3B and 30-7). In addition, *Xist* RNA recruits the H2A variant **macroH2A1**, which has a large C-terminal globular domain that H2A lacks. Some or all of these changes are required for X chromosome inactivation and, moreover, provide the epigenetic imprint that is responsible for maintaining Xi's heterochromatic state in subsequent cell generations. In fact, once an X chromosome has been inactivated, *Xist* RNA plays only a minor role in the maintenance of this state.

What is the mechanism through which one X chromosome is inactivated but not the other? XIC also contains a

gene named *Tsix* (*Xist* spelled backwards) that overlaps *Xist* but is transcribed from the opposite DNA strand. *Tsix* RNA therefore acts as an antisense regulator (Section 31-4As) that prevents the transcription of *Xist* and vice versa. Thus Xi expresses *Xist* but not *Tsix*, whereas Xa expresses *Tsix* but not *Xist*—at least until the heterochromatic state of Xi has been firmly established. If *Tsix* is inactivated on one X chromosome, that chromosome will be inactivated. Apparently, the balance of transcription between *Xist* and *Tsix* selects which X chromosome is inactivated (much like the balance of expression between Cro and the λ repressor stably maintains the λ phage in either the lytic or the lysogenic state; Section 33-3Cb).

b. Chromosome Puffs and Lampbrush Chromosomes Are Transcriptionally Active

The condensed state of facultative heterochromatin renders it transcriptionally inactive by making its DNA inaccessible to the proteins mediating transcription. Conversely, *transcriptionally active chromatin has a relatively open structure*. Such decondensed chromatin occurs in the **chromosome puffs** that emanate from single bands of giant polytene chromosomes (Fig. 34-46). These puffs reproducibly form and regress as part of the normal larval development program and in response to such physiological stimuli as hormones and heat. Autoradiographic studies with ^3H -labeled uridine and immunofluorescence studies using antibodies against RNA polymerase II clearly demonstrate that *puffs are the major sites of RNA synthesis in polytene chromosomes*.

In amphibian oocytes, the analogous decondensation of nonpolytene chromosomes occurs most conspicuously in

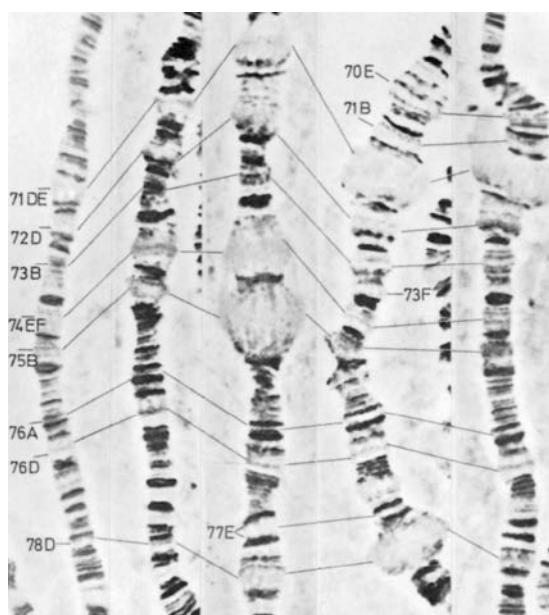


Figure 34-46 Formation and regression of chromosome puffs (lines) in a *D. melanogaster* polytene chromosome over a 22-h period of larval development. The very large puffs in this series of photomicrographs are also known as **Balbiani rings**. [Courtesy of Michael Ashburner, Cambridge University.]

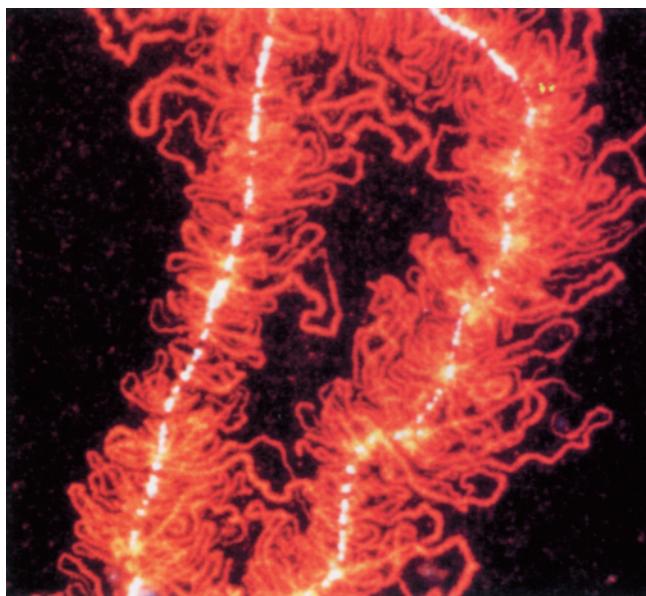


Figure 34-47 Immunofluorescence micrograph of a lampbrush chromosome from an oocyte nucleus of the newt *Notophthalmus viridescens*. The chromosome's numerous transcriptionally active loops give rise to the name "lampbrush" (an obsolete implement for cleaning kerosene lamps). [From Roth, M.B. and Gall, J.G., *J. Cell Biol.* **105**, 1049 (1987). Copyright © 1987 by Rockefeller University Press.]

the so-called **lampbrush chromosomes** (Fig. 34-47). During their prolonged meiotic prophase I (Fig. 1-20), these previously condensed chromosomes loop out segments of transcriptionally active DNA that electron micrographs such as Fig. 34-48 indicate are often single transcription units.

B. Regulation of Transcriptional Initiation

The foregoing observations suggest that selective transcription is mainly responsible for the differential protein

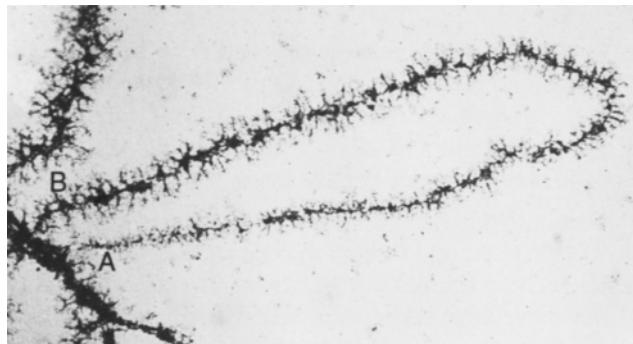


Figure 34-48 Electron micrograph of a single loop of a lampbrush chromosome. The ribonucleoprotein matrix coating the loop increases in thickness from one end of the loop (A) to the other (B), which indicates that the loop comprises a single transcriptional unit. [Courtesy of Oscar L. Miller, Jr., University of Virginia.]

synthesis among the various types of cells in the same organism. It was not until 1981, however, that James Darnell actually demonstrated this to be the case, as follows. Experimentally useful amounts of mouse liver genes were obtained by inserting the cDNAs of mouse liver mRNAs (some 95% of which are cytosolic) into plasmids and replicating them in *E. coli* (Section 5-5Ba). By hybridizing the resulting cloned cDNAs with radioactively labeled mRNAs from various mouse cell types, the *E. coli* colonies containing liver-specific genes were distinguished from colonies containing genes common to most mouse cells. In this way, 12 liver-specific cDNA clones and three common cDNA clones were obtained. The question was then asked, does a eukaryotic cell transcribe only the genes encoding the proteins it synthesizes, or does it transcribe all of its genes but only process properly the transcripts it translates? This question was answered by hybridizing the cloned mouse genes with freshly synthesized and therefore unprocessed RNAs (hnRNAs) obtained from the nuclei of mouse liver, kidney, and brain cells (Fig. 34-49). Only the RNAs extracted from liver nuclei hybridized with the 12 liver-specific cDNAs that were probed. The RNAs from all three cell types, however, hybridized with the DNA from the three clones containing the common mouse genes. Evidently, *liver-specific genes are not transcribed by brain or kidney cells. This strongly suggests that the control of genetic expression in eukaryotes is primarily exerted at the level of transcription.*

In more recent times, the use of DNA microarray technology (DNA chips; Section 7-6B) has enormously increased the number of genes whose levels of transcription may be simultaneously monitored as well as greatly reduced the effort required to do so. For example, **hepatocellular carcinoma (HCC)**, the most common liver malignancy, is among the five leading causes of cancer deaths in the world and is closely associated with chronic infections by hepatitis B or C viruses although the nature of this association is unclear. David Botstein and Patrick Brown characterized the gene expression patterns in HCC tumors and normal liver tissues by isolating their mRNAs and reverse transcribing them to cDNAs, which were then coupled to a fluorescent dye. The labeled cDNAs were then hybridized to DNA microarrays containing ~17,400 human genes. The level of transcription of each of these genes in a given tissue sample was determined from the fluorescence intensity at the corresponding position on the microarray relative to that of a reference cDNA that had a different fluorescent label. The results of this exhaustive analysis are presented in Fig. 34-50, which clearly indicates that HCC tumors have transcriptional patterns that differ from those of nontumor liver tissues. Surprisingly, however, different HCC nodules from the same patient exhibited gene expression patterns whose similarities were no greater than those of tumors from different patients. Nevertheless, certain genes are consistently expressed at high levels in HCC tumors, and hence those whose products are secreted or membrane associated may serve as serological markers for the early detection of liver cancers and/or as potential therapeutic targets.

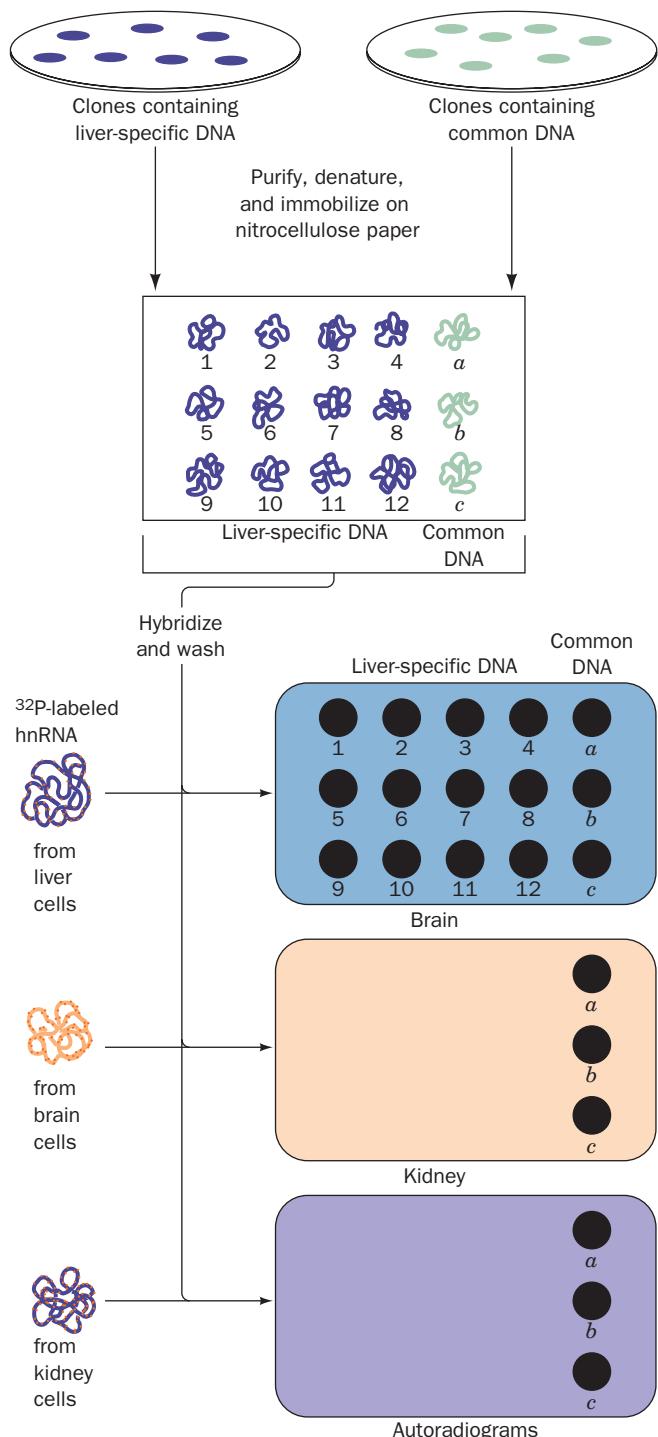


Figure 34-49 Determination of the primary role of selective transcription in the control of eukaryotic gene expression. This was established as follows. Cloned cDNAs encoding 12 different mouse liver-specific proteins (1–12) and 3 different proteins common to most mouse cells (*a*–*c*) were purified, denatured, and spotted onto filter paper (*top*). The DNAs were hybridized with newly formed and therefore unprocessed radioactively labeled RNAs produced by either mouse liver, kidney, or brain nuclei (*lower left*). Autoradiography showed that the liver RNAs hybridized with all 12 liver-specific cDNAs and all 3 common cDNAs but that the kidney and brain RNAs only hybridized with the common cDNAs (*right*).

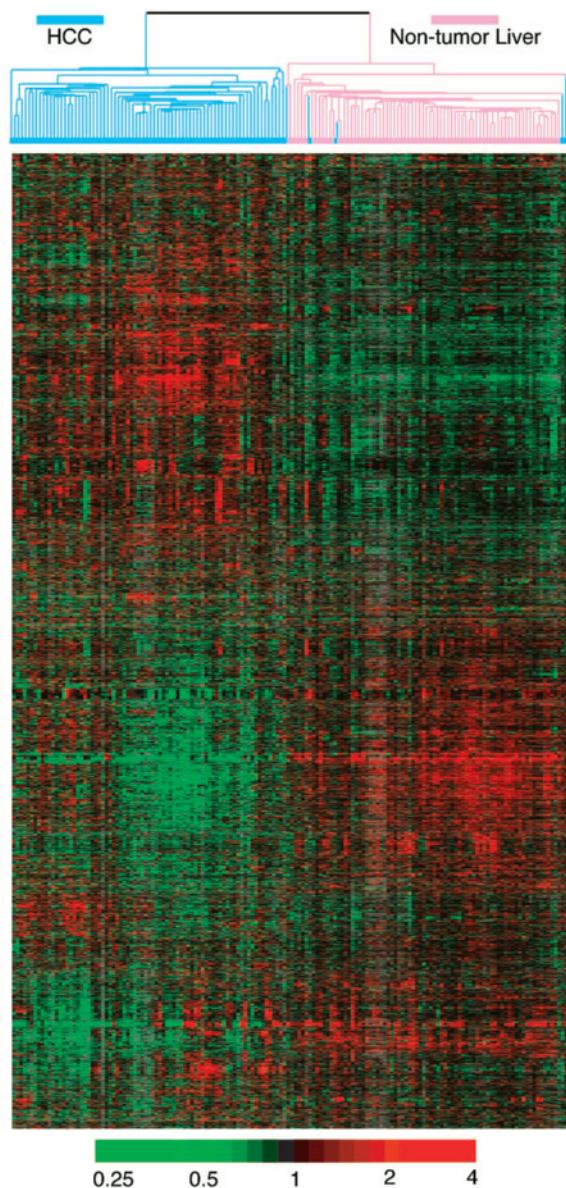


Figure 34-50 Relative transcriptional activities of the genes in hepatocellular carcinoma (HCC) tumors as determined using DNA microarrays. The data are presented in matrix form, with each column representing one of 156 tissue samples (82 HCC tumors and 74 nontumor liver tissues) and each row representing one of 3180 genes (those of the ~17,400 genes on the DNA microarray with the greatest variation in transcriptional activity among the various tissue samples). The data are arranged so as to group the genes as well as the tissue samples on the basis of similarities of their expression patterns. The color of each cell indicates the expression level of the corresponding gene in the corresponding tissue relative to its mean expression level in all the tissue samples with bright red, black, and bright green indicating expression levels of 4, 1, and 1/4 times that of the mean for that gene (as indicated on the scale below). The dendrogram at the top of the matrix indicates the similarities in expression patterns among the various tissue samples. [Courtesy of David Botstein and Patrick Brown, Stanford University School of Medicine.]

a. The Transcriptional Initiation of Structural Genes Involves Three Classes of Transcription Factors

Transcriptional initiation in eukaryotes has been most widely studied in protein-encoding genes, that is, genes that are transcribed by RNA polymerase II (**RNAP II**). In the following paragraphs, we concentrate on the major findings of these studies.

Differentiated eukaryotic cells possess a remarkable capacity for the selective expression of specific genes. The synthesis rates of a particular protein in two cells of the same organism may differ by as much as a factor of 10^9 ; that is, unexpressed eukaryotic genes are completely turned off. In contrast, simply repressible prokaryotic systems such as the *E. coli lac* operon (Section 31-3B) exhibit no more than a 1000-fold range in their transcriptional rates; they have significant basal levels of expression. Nevertheless, as we shall see below, *the basic mechanism of expressional control in eukaryotes resembles that in prokaryotes: the selective binding of proteins to specific genetic control sequences so as to modulate the rate of transcriptional initiation*.

RNAP II, unlike prokaryotic RNA polymerase holoenzyme (Section 31-2), has little if any inherent ability to bind to its promoters. Rather, three different classes of so-called **transcription factors** have been implicated in regulating transcriptional initiation by RNAP II:

1. General transcription factors (GTFs), which are required for the synthesis of all mRNAs, select the transcriptional initiation site and deliver RNAP II to it, thereby forming a complex that initiates transcription at a basal rate.

2. Upstream transcription factors are proteins that bind to specific DNA sequences upstream of the initiation site so as to stimulate or repress transcriptional initiation by GTF-complexed RNAP II. The binding of upstream factors to DNA is unregulated; that is, they bind to any available DNA containing their target sequence. Those that are

present in a cell vary with its developmental state and its needs; their synthesis is also regulated.

3. Inducible transcription factors function similarly to upstream transcription factors but must be activated (or inhibited), either by phosphorylation or by specific ligands, in order bind to their target DNA sites and influence transcriptional initiation. They are synthesized and/or activated in specific tissues at particular times and therefore mediate gene expression in a positionally and temporally specific manner.

We discuss these transcription factors below.

b. The Preinitiation Complex Is a Large and Complex Assembly

Extensive research in numerous laboratories has revealed that *the accurate transcriptional initiation of most structural genes requires the presence of six GTFs, most of which are multiprotein complexes, named TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH* (Table 34-4; TF for transcription factor and II for **class II genes**, those that are transcribed by RNAP II). These GTFs combine, in an ordered pathway, with RNAP II and promoter-containing DNA near the transcriptional start site to form a so-called **preinitiation complex (PIC)** that supports a basal level of transcription. The so-called **core promoters** of structural genes are typically 40 to 50 nt in length and are located largely upstream of the transcriptional start site. They often contain a **TATA box**, a segment of the sense strand (the DNA strand with the same sequence as its corresponding mRNA; Section 31-2A) that is centered at around position -27 and whose consensus sequence is TATA_TA_T^A (Fig. 31-26). Some sequence motifs that commonly occur in core promoters are indicated in Fig. 34-51. The remaining portions of the promoter, to which various transcription factors are idiosyncratically targeted, are known as **upstream activation sequences (UASs)**. An entire eukaryotic promoter typically extends over ~ 100 bp.

Table 34-4 Properties of the General Transcription Factors

Factor	Number of Unique Subunits in Yeast	Mass in Yeast (kD)	Number of Unique Subunits in Humans	Mass in Humans (kD)	Functions
TFIIA	2	46	3	69	Stabilizes TBP and TAF binding
TFIIB	1	38	1	35	Stabilizes TBP binding; recruits RNAP II; influences start site selection
TFIID					Recognizes TATA box; recruits
TBP	1	27	1	38	TFIIA and TFIIB; has positive
TAFs	14	824	16	1084	and negative regulatory functions
TFIIE	2	184	2	165	An $\alpha_2\beta_2$ heterotetramer; recruits TFIH and stimulates its helicase activity; enhances promoter melting
TFIIF	3	156	2	87	Facilitates promoter targeting; stimulates elongation
TFIIH	9	525	10	490	Contains an ATP-dependent helicase that functions in promoter melting and clearance

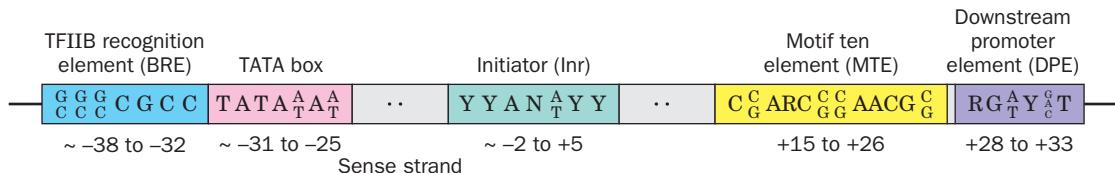


Figure 34-51 Some core promoter sequences for RNA polymerase II. The consensus sequence of each motif is given (some positions can accommodate two or more nucleotides; R, Y, and N represent purine, pyrimidine, and any nucleotide, respectively). The approximate positions of its various motifs

relative to the transcriptional initiation site (+1) are indicated below. The **BRE** is an upstream extension of the **TATA** box, whereas the **DPE** and **MTE** act cooperatively with the **Inr**. Note that a specific core promoter may contain all, some, or none of these motifs.

The assembly of the preinitiation complex, which is diagrammed in Fig. 34-52, begins with the binding of **TATA box-binding protein (TBP)** to the **TATA** box, thereby identifying the transcriptional start site (recall that eliminating the **TATA** box does not necessarily eliminate transcription but does result in heterogeneities in the transcriptional start site; Section 31-2Ee). **TBP** is then joined by a series of ~16 **TBP-associated factors (TAFs)** to form the ~1000-kD multisubunit complex **TFIID**. **TFIIA** then binds to the **TFIID**-DNA complex so as to stabilize it, followed by **TFIIB** (although recent evidence indicates that **TFIIA** is not a universal component of the **PIC** and should therefore be classified as a loosely associated **TAF**-like coactivator rather than a **GTF**). At this point, **TFIIF** recruits **RNAP II** to the promoter in a manner reminiscent of the way that σ factors interact with core RNA polymerase in bacteria (Sections 31-2Aa and 31-2B). Indeed, the smaller of human **TFIIF**'s two subunits exhibits substantial sequence homology with σ^{70} (the predominant bacterial σ factor) and, moreover, can specifically interact with bacterial RNAPs (although it does not participate in promoter recognition). Finally, **TFIIE** and **TFIIF** join, in that order, to form the **PIC**. Once this complex has been assembled, an ATP-dependent activation step, probably mediated by **TFIIF**'s helicase function to melt the promoter, is required to initiate transcription at a basal rate. You should note that the human **PIC**, exclusive of the ~12-subunit, ~600-kD **RNAP II**, contains 32 subunits with an aggregate mass of ~1900 kD. Indeed, many of the proteins in the **PIC** are the targets of transcriptional regulators.

c. TBP Greatly Distorts Its Bound TATA Box DNA

TBP has a highly conserved (81% identical between yeast and humans) C-terminal domain of 180 residues that contains two ~40% identical direct repeats of 66 or 67 residues separated by a highly basic segment. In contrast, **TBP**'s N-terminal domain is widely divergent, both in length and sequence, and, in fact, is unnecessary for **TBP** function *in vitro*. Curiously, the human N-terminal domain contains an uninterrupted run of 38 Gln residues, whereas that of *D. melanogaster* contains two blocks of 6 and 8 Gln residues separated by 32 residues, and that of yeast entirely lacks such sequences. Perhaps the N-terminal domains of **TBPs** have evolved to satisfy species-specific functions.

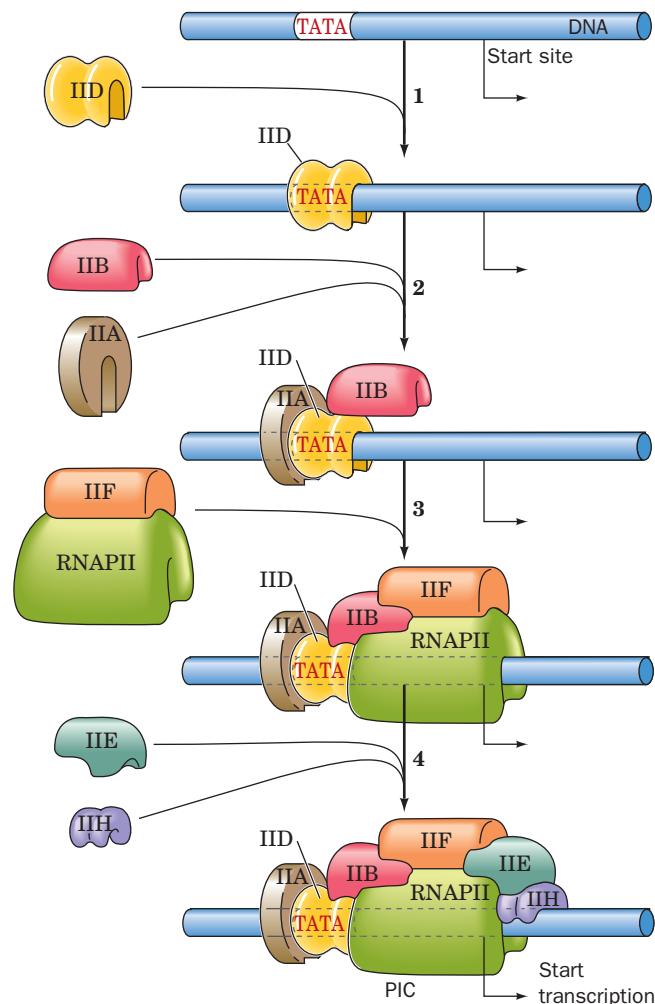


Figure 34-52 Assembly of the preinitiation complex (PIC) on a TATA box-containing promoter. (1) **TFIID** assembles on the **TATA** box beginning with the binding of **TATA box-binding protein (TBP)** to the **TATA** box. (2) **TFIIA** and **TFIIB** then bind to the growing complex. (3) **TFIIF** then binds to **RNAP II** and escorts it to the complex. (4) Finally, **TFIIE** and **TFIIF** are sequentially recruited to the complex, thereby completing the **PIC**. [After Zawel, L. and Reinberg, D., *Curr. Opin. Cell Biol.* **4**, 490 (1992).]

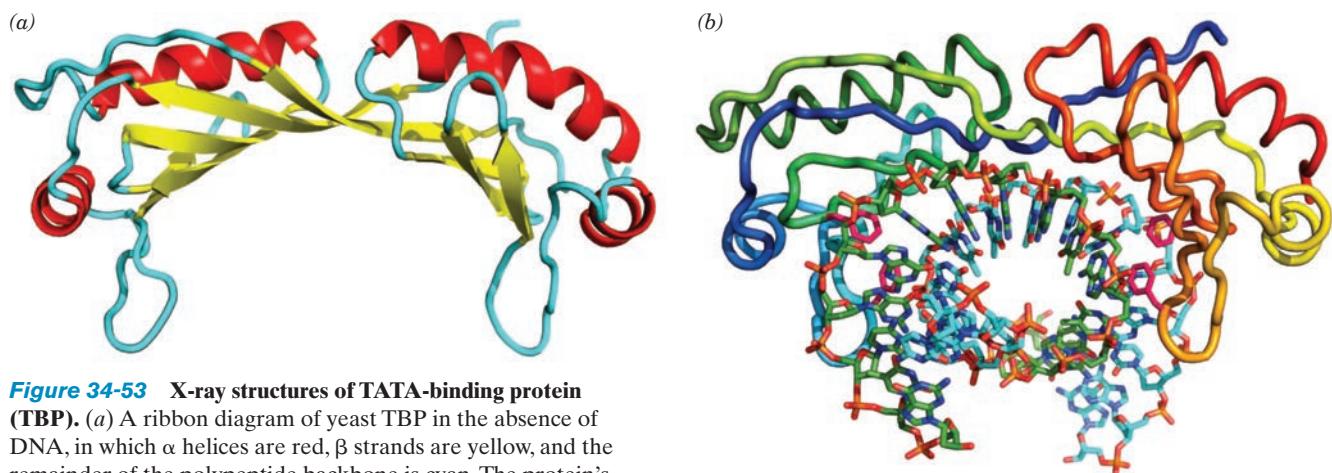


Figure 34-53 X-ray structures of TATA-binding protein (TBP). (a) A ribbon diagram of yeast TBP in the absence of DNA, in which α helices are red, β strands are yellow, and the remainder of the polypeptide backbone is cyan. The protein's pseudo-2-fold axis of symmetry is vertical. Note that the protein seems to be precisely the proper size and shape to sit astride a 20-Å-diameter cylinder of B-DNA. This, however, is not what happens. [Based on an X-ray structure by Roger Kornberg, Stanford University. PDBid 1TBP.] (b) The structure of human TBP in complex with a 16-bp TATA box-containing duplex DNA (sense strand sequence CTGCTATAAAAGGCTG, with its TATA box in bold) viewed as in Part a. The DNA, which is largely in the B form, is drawn in stick form and colored according to atom type with sense strand C green, antisense strand C cyan, N blue, O red, and P orange. It enters its binding site from behind with the 5' end of the sense strand below the saddle on the right and exits toward the front on the left with its

helix axis nearly perpendicular to the page. The protein is drawn in worm form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of Phe residues 193, 210, 284, and 301, which induce sharp kinks in the DNA, are drawn in stick form (magenta). Between the kinks, which are located at each end of the TATA box, the DNA is partially unwound with the protein's central 8 strands of its 10-stranded antiparallel β sheet inserted into and tracked along the DNA's greatly widened minor groove. The TBP does not contact the DNA's major groove. [Based on an X-ray structure by Stephen Burley, Structural GenomiX, Inc., San Diego, California. PDBid 1CDW.] See the Interactive Exercises.

The X-ray structures of TBP from yeast (only its C-terminal domain) and from the flowering plant *Arabidopsis thaliana* (whose N-terminal domain consists of only 18 residues), by Kornberg and by Stephen Burley, reveal a saddle-shaped molecule (Fig. 34-53a) that consists of two structurally similar and topologically identical domains, each composed of one of the direct repeats. These are arranged with pseudo-2-fold symmetry such that the protein consists of a 10-stranded antiparallel β sheet, 5 strands from each domain, flanked at each end by two α helices and a loop that is reminiscent of a stirrup hanging from the protein saddle. The curvature of the β pleated sheet saddle is such that it appears that TBP, in agreement with biochemical and genetic evidence, could fit snugly astride the DNA. However, the X-ray structures of the DNA complexes tell quite a different story.

Two closely similar X-ray structures of TBP–DNA complexes were determined: one by Paul Sigler of yeast TBP in complex with a 27-nt DNA that forms an 11-bp TATA box-containing stem whose ends are joined by a 5-nt loop; and one by Burley of *Arabidopsis* TBP in complex with a 16-bp TATA box-containing duplex DNA. The DNA indeed binds to the concave surface of TBP but with its duplex axis nearly perpendicular rather than parallel to the saddle's "cylindrical" axis (Fig. 34-53b). The DNA is kinked by $\sim 45^\circ$ between the first two and the last two base pairs of its 8-bp TATA element. Between these kinks, the DNA is severely, although smoothly, bent with a radius of curvature

of ~ 25 Å and unwound by $\sim 1/3$ of a turn. This permits the protein's antiparallel β sheet to bind in the DNA's greatly widened and more shallow minor groove through hydrogen bonding and van der Waals interactions (the protein does not contact the DNA's major groove). A noteworthy aspect of this remarkable structure is that each kink in the DNA is stabilized by a wedge of two Phe side chains extending from the adjacent stirrup that pries apart the base pairs flanking the kink from their minor groove side and severely buckles the interior base pair. As a result of these unprecedented distortions to the DNA (the protein undergoes only slight conformational adjustments on binding DNA), there is an $\sim 100^\circ$ angle and a lateral 18-Å displacement between the helix axes of the B-form DNA entering and leaving TBP's binding site, thereby giving the DNA a cranklike shape. The DNA, nevertheless, maintains normal Watson–Crick pairing throughout the distorted region.

d. TFIIA and TFIIB Bind to DNA, TBP, and RNAP II

TFIIB consists of an N-terminal domain (**TFIIB_N**), which interacts with RNAP II, and a C-terminal domain (**TFIIB_C**), which interacts with DNA and TBP. The X-ray structures of ternary complexes of yeast TFIIA, TBP, and a TATA box-containing promoter DNA were independently determined by Richmond and Sigler and those of ternary complexes of human TFIIB_C, human or *Arabidopsis* TBP, and a TATA box-containing promoter DNA were independently determined by Sigler and by Robert Roeder

and Burley. The TBP–DNA complexes in all the foregoing binary and ternary complexes are closely similar. Burley therefore constructed a plausible model of the TFIIA–TFIIB_C–TBP–DNA quaternary complex by superimposing the TBP–DNA complexes in the TFIIA- and TFIIB_C-containing ternary complexes (Fig. 34-54a). TFIIA, a heterodimer in yeast, consists of a 6-stranded β barrel and a 4-helix bundle that, together, have a bootlike shape. The β barrel domain of TFIIA binds to TBP’s N-terminal stirrup so as to extend TBP’s β sheet to form a continuous 16-stranded β sheet. In addition, the TFIIA β barrel binds to the DNA over its major groove through salt bridges between four of its Lys and Arg side chains and the DNA’s phosphate groups. TFIIB_C consists of two similar α helical domains that are rotated by 90° with respect to one another so as to form a cleft that clamps the TBP’s C-terminal stirrup. TFIIB_C binds to the DNA via both its domains through several salt bridges with the DNA’s phosphate groups as well as base-specific contacts in both the major and minor grooves of the **TFIIB recognition element (BRE)**; consensus sequence $\text{GGG}^{\text{CCC}}\text{GCC}$, which is located immediately upstream of the TATA box (Fig. 34-51). The formation of these interactions requires the distortions that TBP binding imposes on the DNA structure and hence TFIIB_C binding is synergistic with TBP binding. Since the pseudosymmetric TBP has been shown to bind to the TATA box in

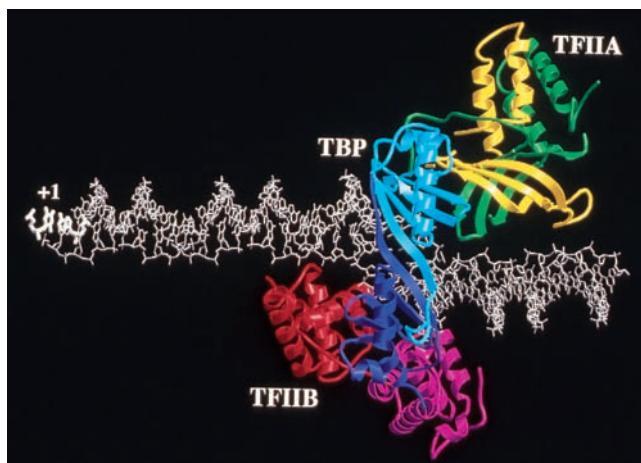
either orientation, it appears that the base-specific interactions between TFIIB_C and the promoter at the BRE function to position TFIIB to properly orient the TBP on the promoter. The model of the quaternary complex (Fig. 34-54a) indicates that its three proteins all bind to the DNA upstream of the transcriptional start site, leaving ample room for the additional protein–DNA and protein–protein interactions that regulate the frequency with which RNAP II is recruited to the promoter.

X-ray structures of the TFIIB–RNAP II complex, independently determined by Kornberg and by Patrick Cramer, reveal that TFIIB contacts RNAP II near its RNA exit channel and that TFIIB_N inserts its “B finger” loop into the active center of RNAP II where it could interact with both DNA and RNA. A model of the TFIIB–TBP–RNAP II–DNA complex based on these X-ray structures together with that of the TFIIB_C–TBP–DNA complex is shown in Fig. 34-54b.

e. TFIID Is a Horseshoe-Shaped Complex That Probably Contains a Histone-like Octamer

The electron microscopy–based structure of the human TFIID–TFIIA–TFIIB complex was determined at 35-Å resolution by Robert Tjian and Eva Nogales. TFIID is a horseshoe-shaped trilobal complex to which TFIIA and TFIIB are bound on opposite lobes that flank the central

(a)



(b)

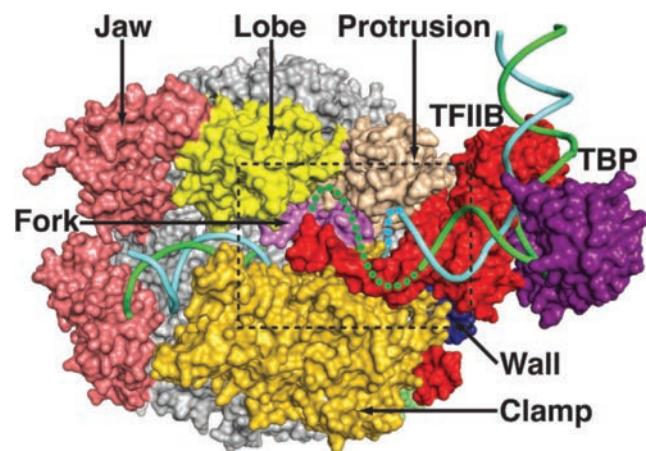


Figure 34-54 Structural models for the interactions of TFIIB. (a) Model of the TFIIA–TFIIB_C–TBP–TATA box–containing DNA quaternary complex. The arrangement of the proteins (ribbons) and DNA (white stick model) is based on the independently determined X-ray structures of the TFIIA–TBP–DNA and TFIIB–TBP–DNA ternary complexes. In the model, the DNA has been extended in both directions beyond the TATA box, with its transcription start site (+1) on the left. The TBP’s pseudosymmetrically related N- and C-terminal domains are cyan and purple, the TFIIA’s two subunits are yellow and green, and the TFIIB_C’s similar N- and C-terminal domains are red and magenta. The TBP binds to both TFIIA and TFIIB and all three proteins bind to the DNA at independent sites. [Courtesy of Stephen Burley, Structural Genomix, Inc., San Diego, California. Based on PDBids 1VOL and 1YTF.] (b) Model of the

TFIIB–TBP–RNAP II–DNA open complex based on the X-ray structures of the TFIIB–RNAP II and TFIIB_C–TBP–DNA complexes. The view of the RNAP II is approximately from the top of Fig. 31-22b followed by a ~90° rotation about the horizontal axis. The proteins are represented by their solvent-accessible surfaces with TFIIB red, TBP purple, and the various portions of RNAP II in other colors. The DNA is drawn in cartoon form with its template and nontemplate strands cyan and green and with the dotted segments representing its transcription bubble. Its upstream end is at the upper right. Note how the leftmost portion of TFIIB, its B finger, extends into the active center of RNAP II to contact the DNA. [Courtesy of Roger Kornberg, Stanford University. Based on PDBids 3K7A, 1R5U, and 1VOL.]

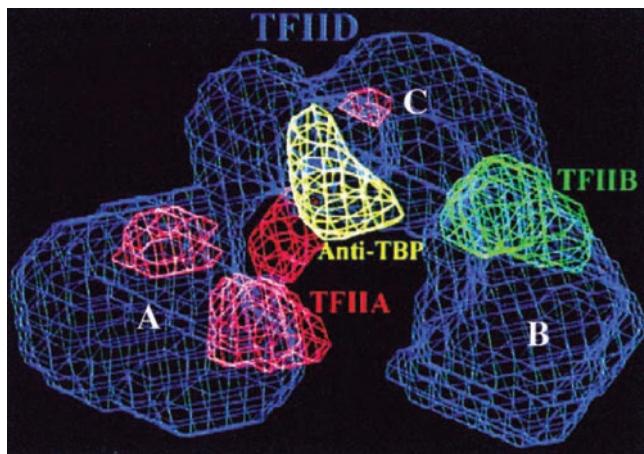


Figure 34-55 EM-based image of the human

TFIID-TFIIA-TFIIB complex at 35-Å resolution. The blue mesh outlines the entire ternary complex, which consists of three domains, A, B, and C, arranged in a horseshoe shape and roughly 200 Å wide, 135 Å high, and 110 Å thick. The red and green meshes indicate the positions of TFIIA and TFIIB as determined by comparison with the EM-based images of the TFIID-TFIIA complex and TFIID alone. The yellow mesh indicates the binding position of an anti-TBP antibody. The different shapes of TFIIA here and in Fig. 34-54a are probably due to the fact that the TFIIA in Fig. 34-54a consists only of residues 56 to 209 of the 286-residue protein. [Courtesy of Eva Nogales, University of California at Berkeley.]

cavity (Fig. 34-55). This, together with the foregoing model of the TFIIA-TFIIB-TBP-DNA quaternary complex (Fig. 34-54a), strongly suggests that TBP is located at the top of the cavity where it can contact both TFIIA and TFIIB and that the core promoter DNA passes through the cavity, where it is bound by TBP, TFIIA, and TFIIB. Indeed, the EM-based image of an anti-TBP antibody in complex with TFIID reveals that the antibody binds to TFIID in the expected position (Fig. 34-55).

The various TAFs are highly conserved from yeast to humans. Moreover, portions of 9 of the 14 known species of TAFs are homologous to nonlinker histones. For example, segments consisting of residues 17 to 86 of the 268-residue **dTAF9** (d for *Drosophila*; also called **dTAF_{II}42**, where the number indicates its nominal molecular mass in kD) and residues 1 to 70 of the 592-residue **dTAF6** (also called **dTAF_{II}60**) are, respectively, homologous to histones H3 and H4. The X-ray structure of the dTAF9(17-86)-dTAF6(1-70) complex, determined by Roeder and Burley, reveals that both of these polypeptide segments assume the histone fold (Fig. 34-56a): two short helices flanking a long central helix (Fig. 34-8). In fact, TAF9(17-86) and TAF6(1-70) associate quite similarly to H3 and H4 in the nucleosome (Fig. 34-7) to form an $\alpha_2\beta_2$ heterotetramer (Fig. 34-56a). In addition, hTAF12(57-128) (h for human; the 162-residue **hTAF12** is also called **dTAF_{II}20**), which is homologous to histone H2B, forms a complex with hTAF4(870-943) (the 1083 residue **hTAF4** is also called **dTAF_{II}135**), which is homologous to H2A. The X-ray struc-

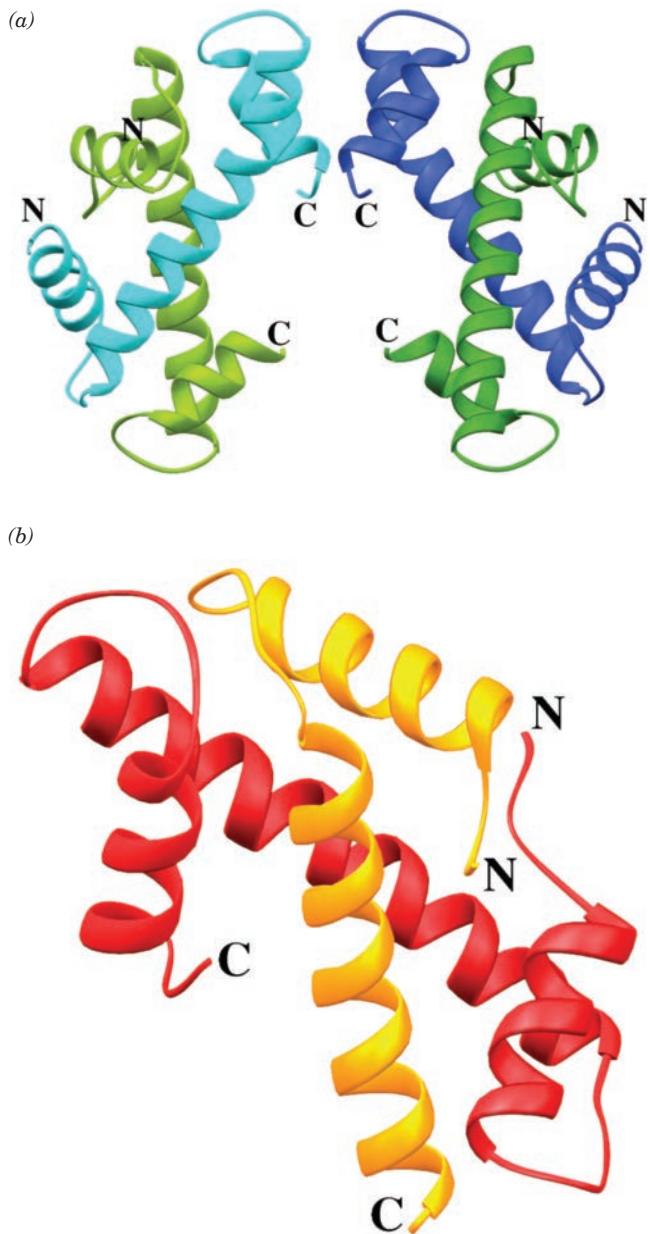


Figure 34-56 X-ray structures of TAFs that form histonelike complexes. (a) The dTAF9(17-86)-dTAF6(1-70) $\alpha_2\beta_2$ heterotetramer as viewed with its 2-fold axis vertical. Note how the H3-like TAF9 segments (blue and cyan) and the H4-like TAF6 segments (green and olive) all assume the histone fold, how TAF9-TAF6 pairs interdigitate in head-to-tail arrangements to form heterodimers, and how the two TAF9 segments interact via a four-helix bundle to form the heterotetramer, much as do histones H3 and H4 in nucleosome cores (Figs. 34-7 and 34-8). [Based on an X-ray structure by Robert Roeder and Stephen Burley, The Rockefeller University. PDBID 1TAF.] (b) The hTAF12(57-128)-hTAF4(870-943) heterodimer. Note how the H2B-like TAF12 segment (red) forms a regular histone fold but that the H2A-like TAF4 segment (gold) lacks the histone fold's C-terminal loop and helix. This is because TAF4 residues 918 to 943 are disordered. Nevertheless, the two subunits interdigitate to form a heterodimer, much as do histones H2B and H2A in nucleosome cores (Fig. 34-8). [After an X-ray structure by Dino Moras, CNRS/INSERM/ULP, Illkirch Cédex, France. PDBID 1H3O.]

ture of this complex, determined by Dino Moras, reveals that it forms a histonelike heterodimer (Fig. 34-56b) but not a histonelike tetramer. **TAF11** and **TAF13** also form a histonelike heterodimer.

Gel filtration chromatography and sedimentation measurements (Sections 6-3Ba and 6-5Aa) by Stephen Buratowski and Song Tan indicate that the heterotetramer of **yTAF6** (y for yeast) and **yTAF9** associates with two heterodimers of **yTAF12** and **yTAF4** to form a heterooctamer. The mutation to Ala or Tyr of the highly conserved Leu 464 of yTAF12 (the homolog of H2B residue Leu 77, which is located near the C-terminus of this histone's long central helix and hence occupies the hydrophobic core of the H4–H2B 4-helix bundle; Fig. 34-8) prevents the formation of this octamer. This suggests that the octamer is held together by 4-helix bundles between yTAF6 and yTAF12 similar to those between H4 and H2B in nucleosomes (Figs. 34-7b and 34-8). Indeed, a model of this interface constructed from the above two X-ray structures suggests that its putative 4-helix bundle is remarkably similar to that of the H4–H2B interface. Nevertheless, it seems unlikely that this putative TAF octamer is wrapped with DNA in the PIC as is the histone octamer in the nucleosome. This is because most of the histone residues that make critical contacts with DNA in the nucleosome have not been conserved in the foregoing TAFs and, in fact, many of them have been replaced in these TAFs by highly conserved (in the TAFs) acidic residues, which would repel the anionic DNA.

f. Many Class II Core Promoters Lack a TATA Box

The core promoters of 65% of class II genes lack TATA boxes. They are mostly “housekeeping” genes; that is, genes that are constitutively expressed in all cells at relatively low rates. How can RNAP II properly initiate transcription at these TATA-less promoters? Investigations have shown that TATA-less promoters often contain a so-called **initiator (Inr)** element that extends from positions –2 to +5 and that contains the loose consensus sequence YYAN_TYY where the A is the initiating (+1) nucleotide (Fig. 34-51). The presence of the Inr element is sufficient to direct RNAP II to the correct start site. These systems require the participation of many of the same GTFs that initiate transcription from TATA box-containing promoters. Surprisingly, they also require TBP. This suggests that with TATA-less promoters, Inr recruits TFIID such that its component TBP binds to the –30 region in a sequence-nonspecific manner. Indeed, in Inr-containing promoters that also contain a TATA box, the two elements act synergistically to promote transcriptional initiation. Nevertheless, a mutant TBP that is defective in TATA box binding will support efficient transcription from some TATA-less promoters although not from others. This suggests that the former promoters do not require a stable interaction with TBP. Some TATA-less promoters contain other sequence motifs, including the **motif ten element (MTE)** and the **downstream promoter element (DPE)** (Fig. 34-51).

The foregoing suggests that there are variants of at least some of the GTFs and TAFs. In fact, the human genome

contains multiple sequences related to TFIID and TFIID subunits as well as alternative genes for several TAFs. Some of these variant genes are only expressed in certain cell types and/or at specific developmental stages. The resulting variant transcription factors probably recognize alternative core promoter elements and/or mediate selective interactions with upstream transcription factors.

g. Class I and Class III Genes Also Require TBP for Transcriptional Initiation

RNA polymerase I (**RNAP I**, which synthesizes most rRNAs) and RNA polymerase III (**RNAP III**, which synthesizes 5S rRNA and tRNAs) require different sets of GTFs from each other and from RNAP II to initiate transcription at their respective promoters. This is not unexpected considering the very different organizations of these three classes of promoters (Sections 31-2Ed and 31-2Eg). Indeed, the promoters recognized by RNAP I (class I promoters) and nearly all those recognized by RNAP III (class III promoters) lack TATA boxes. Thus, it came as a surprise when it was demonstrated that *TBP is required for initiation by both RNAP I and RNAP III*. It participates by combining with different sets of TAFs to form the GTFs **SLI** (with class I promoters) and **TFIIB** (with class III promoters). As with certain class II TATA-less promoters, a TBP mutant that is defective for TATA box binding can still support *in vitro* transcriptional initiation by both RNAP I and RNAP III. Clearly, TBP, the only known universal transcription factor, is an unusually versatile protein.

h. Transcriptional Initiation of Class II Genes Is Mediated by Cell-Specific Upstream Transcription Factors Bound to Promoter and Enhancer Elements

The use of molecular cloning procedures permitted the demonstration that *eukaryotic promoter and enhancer elements mediate the expression of cell-specific genes* (recall that an enhancer is a gene sequence that is required for the full activity of its associated promoter but that may have a variable position and orientation with respect to that promoter; Section 31-2Ef). For example, William Rutter linked the 5'-flanking sequences of either the insulin or the chymotrypsin gene to the sequence encoding **chloramphenicol acetyltransferase (CAT)**, an easily assayed enzyme not normally present in eukaryotic cells. A plasmid containing the insulin gene recombinant elicits expression of the CAT gene only when introduced into cultured cells that normally produce insulin. Likewise, the chymotrypsin recombinant is only active in chymotrypsin-producing cells. Dissection of the insulin control sequence indicates that the segment between its positions –103 and –333 contains an enhancer: In insulin-producing cells only, it stimulates the transcription of the CAT gene with little regard to the enhancer's position and orientation relative to its promoter.

The foregoing indicates that cells contain specific transcription factors, the upstream transcription factors, that recognize the promoters and enhancers in the genes they transcribe. For instance, Tjian isolated a protein, **Sp1** (for specificity protein-1), from cultured human cells that stimulates, by factors of 10 to 50, the transcription of cellular

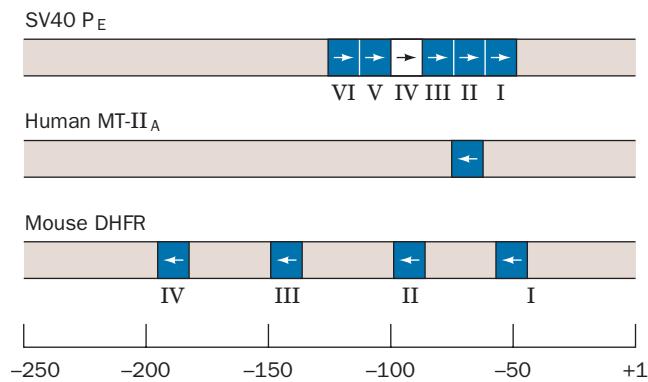


Figure 34-57 Arrangement and relative orientations of the GC boxes in the indicated promoters. Each arrow indicates the relative orientation of a GC box, which has the sequence NGGGCGGNNN. The blue boxes represent Sp1-binding sites, whereas SV40 GC box IV is shown as a white box because Sp1 bound at GC box V prevents this transcription factor from efficiently binding to GC box IV. The transcription start site is designated by +1. DHFR = dihydrofolate reductase; MT = metallothionein. [After Kadonaga, J.T., Jones, K.A., and Tjian, R., *Trends Biochem. Sci.* **11**, 21 (1986).]

and viral genes containing at least one properly positioned GC box [GGGCGG (Section 31-2Ee); Fig. 34-57]. This protein binds, for example, to the 5'-flanking region of the SV40 virus early genes so as to protect its GC boxes from DNase I digestion (Fig. 34-58a; **DNase I footprinting**) and from methylation by dimethyl sulfate (Fig. 34-58b; **DMS footprinting**). Likewise, Sp1 specifically interacts with the four GC boxes in the upstream region of the mouse dihydrofolate reductase gene and with the single GC boxes in the human **metallothionein I_A** and **II_A** promoters (metallothioneins are metal ion-binding proteins that participate in heavy metal ion detoxification processes and whose synthesis is triggered by heavy metal ions). Cryo-EM studies by Nogales and Tjian have visualized the binding sites of several activators, including Sp1, on TFIID (Fig. 34-59).

Upstream transcription factors are essential participants in controlling the differential expression of the various globin genes in the human embryo, fetus, and adult (Section 34-2Fa). A typical β -globin gene promoter, in addition to its TATA box, has two positive-acting promoter elements: a CCAAT box near the -70 to -90 region and a CACCC motif at variable sites but often near positions -95 to -120 (Section 31-2Ee). Their importance is demonstrated by the observations that individuals with point mutations in their TATA or CACCC elements have reduced β -globin levels. These promoter elements are specifically bound by upstream transcription factors. Thus, the CCAAT box is bound by the ubiquitous transcription factor **CP1** and the CACCC element is bound by Sp1, which also binds to other globin promoter sequences that resemble Sp1's consensus binding sequence. Four erythroid-specific upstream transcription factors have also been implicated in globin gene expression: **GATA-1** (so named because it

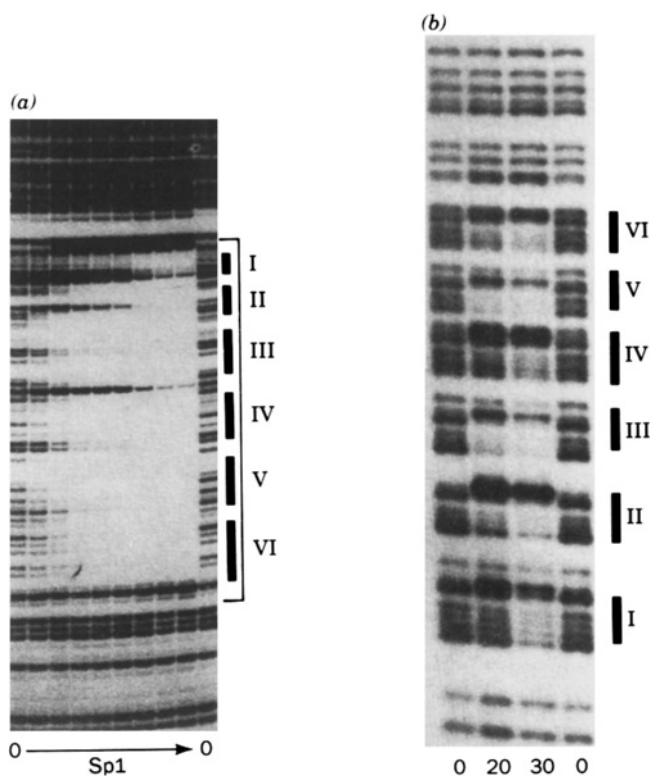


Figure 34-58 Identification of the Sp1-binding sites on the SV40 early promoter (Fig. 34-57, top). (a) Pancreatic DNase I is a relatively nonspecific endonuclease. In a DNase I footprinting assay, a DNA segment that is ³²P end-labeled on one strand is incubated with a binding protein and then lightly digested with DNase I such that, on average, each labeled DNA strand is cleaved only once. The DNA is then denatured, the resulting labeled fragments separated according to size by electrophoresis on a sequencing gel (Section 7-2Aa), and detected by autoradiography. Unprotected DNA is cleaved more or less at random and therefore appears as a "ladder" of bands, each representing an additional nucleotide (as in a sequencing ladder; Figs. 7-13 and 7-14). In contrast, the DNA sequences that the protein protects from DNase I cleavage have no corresponding bands. In the above footprint, the lanes labeled "0" are the DNase I digestion pattern in the absence of Sp1 and in the other lanes the amount of Sp1 increases from left to right. The footprint boundary is delineated by the bracket and the positions of SV40 GC boxes I to VI are indicated. [From Kadonaga, J.T., Jones, K.A., and Tjian, R., *Trends Biochem. Sci.* **11**, 21 (1986). Copyright © 1986 by Elsevier Biomedical Press.] (b) Dimethyl sulfate (DMS) methylates DNA's G residues at their N7 positions, which on treatment with weak base, excises the methylated G nucleosides from the DNA, thereby cleaving its sugar-phosphate backbone. In DMS footprinting, a protein-complexed ³²P end-labeled DNA segment is lightly treated with DMS such that each labeled DNA strand is, on average, cleaved only once. The resulting fragments are electrophoretically separated on a sequencing gel and detected by autoradiography. The DNA regions that the protein protects from methylation are not cleaved by this procedure and therefore are not represented in the resulting G residue "ladder." In the above autoradiogram, the number below each lane indicates the amount (in μ L) of an Sp1 fraction added to a fixed quantity of SV40 early promoter DNA. The positions of its GC boxes are indicated. [From Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P., and Tjian, R., *Science* **230**, 516 (1985). Copyright © 1985 by the American Society for the Advancement of Science.]

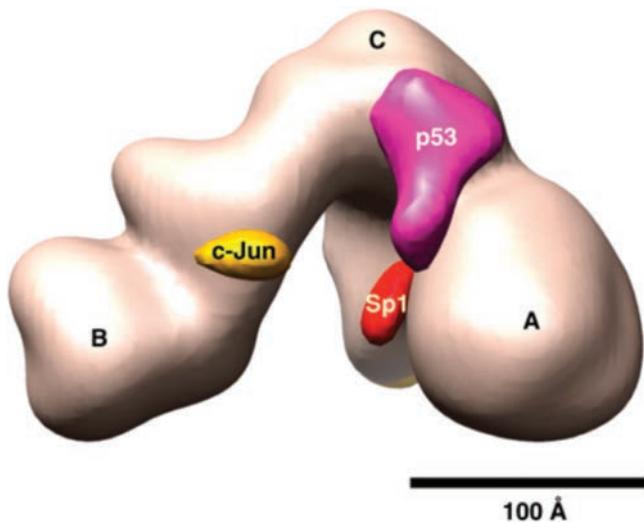


Figure 34-59 Cryo-EM-based composite model of the sites on TFIID to which certain activators bind. The TFIID (tan), whose three domains are labeled A, B, and C, is viewed from the opposite direction as in Fig. 34-55. The binding sites for the transcriptional activators Sp1 (red), c-Jun (yellow; Section 34-4Ca), and p53 (magenta; Section 34-4Df), which were individually visualized, are shown. [Courtesy of Robert Tjian, University of California at Berkeley.]

binds to sequences that contain the conserved core GATA), **NF-E2** (for *nuclear factor-erythroid 2*), **NF-E3**, and **NF-E4** (GATA-1 was previously named NF-E1).

Analysis of hereditary persistence of fetal hemoglobin (HPFH), a syndrome characterized by the inappropriate expression of γ genes in human adults (Section 34-2Gb), has provided valuable insights into the basis of stage-specific globin expression. There are several HPFH variants that differ from normal only by a point mutation in the γ gene promoter. Such mutations might result in either tighter binding of a positive transcription factor or looser binding of a negative regulator. Thus, an HPFH mutation at position -117, which is located in the more upstream of the γ gene's two CCAAT boxes, increases the resemblance of this site to CP1's consensus binding sequence and results in a 2-fold tighter binding of CP1 to the mutant site. Similarly, HPFH mutations in a GC-rich region close to position -200 result in tighter Sp1 binding.

i. Upstream Transcription Factors Interact Cooperatively with Each Other and the PIC

How do upstream transcription factors stimulate (or inhibit) transcription? *Evidently, when these proteins bind to their target DNA sites in the vicinity of a PIC (in some cases, many thousands of base pairs distant), they somehow activate (or repress) its component RNAP II to initiate transcription.* Transcription factors may bind cooperatively to each other and/or the PIC in a manner resembling the binding of two λ repressor dimers and RNA polymerase to the σ_R operator of bacteriophage λ (Section 33-3Dc),

thereby synergistically stimulating (or repressing) transcriptional initiation. Indeed, molecular cloning experiments indicate that many enhancers and **silencers** (the analogs of enhancers that function in the transcriptional repression of their associated gene) consist of segments (modules) whose individual deletion reduces but does not eliminate enhancer/silencer activity. *Such complex arrangements presumably permit transcriptional control systems to respond to a variety of stimuli in a graded manner.* In some cases, however, several transcription factors together with so-called **architectural proteins** cooperatively assemble on an ~100-bp enhancer to form a multisubunit complex, known as an **enhanceosome**, in which the absence of a single subunit all but eliminates its ability to stimulate transcriptional initiation at the associated promoter. Thus, enhanceosomes function more like on/off switches rather than providing a graded response. Architectural proteins function to bend and/or otherwise deform enhancers so as to promote the assembly of the other enhanceosome proteins. Enhanceosomes may also contain **coactivators** and/or **corepressors**, proteins that do not bind to DNA but, rather, interact with proteins that do so to activate or repress transcription.

The functional properties of many upstream transcription factors are surprisingly simple. They appear to have (at least) two domains:

1. A DNA-binding domain that binds to the protein's target DNA sequence (and whose structural properties are discussed below).
2. A domain containing the transcription factor's activation function. Sequence analysis indicates that many of these **activation domains** (also called **transactivation domains** because they act in trans with the genes they control) have conspicuously acidic surface regions whose negative charges, if mutationally increased or decreased, respectively raise or lower the transcription factor's activity. This suggests that the associations between these transcription factors and a PIC are mediated by relatively nonspecific electrostatic interactions rather than by conformationally more demanding hydrogen bonds. Other types of activation domains have also been characterized, including those with Gln-rich regions, such as Sp1, and those with Pro-rich regions.

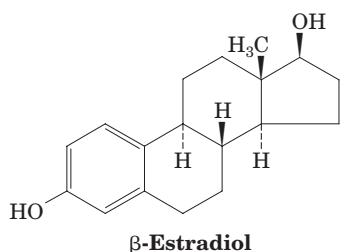
The DNA-binding and activation functions of eukaryotic transcription factors can be physically separated (which is why they are thought to occur on different domains). Thus, a genetically engineered hybrid protein, containing the DNA-binding domain of one transcription factor and the activation domain of a second, activates the same genes as the first transcription factor. Indeed, it makes little functional difference as to whether the activation domain is placed on the N-terminal side of the DNA-binding domain or on its C-terminal side. This geometric permissiveness in the binding between the activation domain and its target protein is also indicated by the observation that transcription factors are largely insensitive to the orientations and positions of their corresponding

enhancers relative to the transcriptional start site [Section 31-2Ef; it is also the basis of the two-hybrid system for identifying proteins that interact *in vivo* (Section 19-3Ca)]. Of course, the DNA between an enhancer and its distant transcriptional start site must be looped around for an enhancer-bound transcription factor to interact with the promoter-bound PIC (Section 31-2Ef).

The synergy (cooperativity) of multiple transcription factors in initiating transcription may be understood in terms of a simple recruitment model. Suppose an enhancer-bound transcription factor increases the affinity with which a PIC binds to the enhancer's associated promoter so as to increase the rate at which the PIC initiates transcription there by a factor of 10. Then, if another transcription factor binding to a different enhancer subsite likewise increases the initiation rate by a factor of 20, both transcription factors acting together will increase the initiation rate by a factor of 200. *In this way, a limited number of transcription factors can support a much larger number of transcription patterns.* Transcriptional activation, according to this model, is essentially a mass action effect: The binding of a transcription factor to an enhancer increases the transcription factor's effective concentration at the associated promoter (the DNA holds the transcription factor in the vicinity of the promoter), which consequently increases the rate at which the PIC binds to the promoter. This explains why a transcription factor that is not bound to DNA (or even lacks a DNA-binding domain) inhibits transcriptional initiation. Such unbound transcription factors compete with DNA-bound transcription factors for their target sites and thereby reduce the rate at which the PIC is recruited to the associated promoter. This phenomenon, which is known as **squelching**, is apparently why transcription factors in the nucleus are almost always bound to inhibitors unless they are actively engaged in transcriptional initiation.

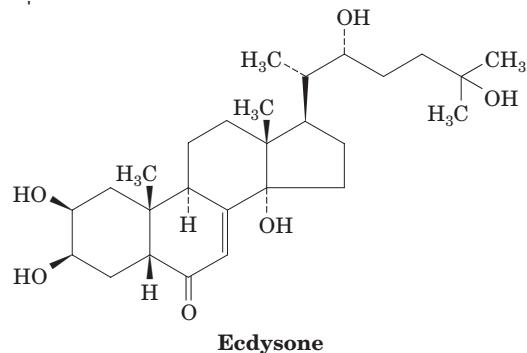
j. Steroid Receptors Are Examples of Inducible Transcription Factors

Eukaryotic cells express many cell-specific proteins in response to the presence of various hormones. Many of these hormones are **steroids** (Section 25-6C), cholesterol derivatives that mediate a wide variety of physiological and developmental responses (Section 19-1G). For example, the administration of **estrogens** (female sex hormones) such as **β -estradiol**



causes chicken oviducts to increase their ovalbumin mRNA level from \sim 10 to \sim 50,000 molecules per cell, and the amount of ovalbumin they produce rises from unde-

tectable levels to a majority of their newly synthesized protein. Similarly, the insect steroid hormone **ecdysone**



mediates several aspects of larval development (the temporal sequence of chromosome puffing shown in Fig. 34-46 can be induced by ecdysone administration).

Steroids, which are nonpolar molecules, spontaneously pass through the membranes of their target cells to bind to their corresponding steroid receptors. In the absence of their cognate steroid, these receptors are bound in large multiprotein complexes that contain chaperone proteins such as Hsp90 and Hsp70 as well as immunophilins (Sections 9-2B and 9-2C), which presumably function to maintain the receptor in its native conformation, ready to bind its cognate steroid. Depending on the identity of the receptor, these complexes mainly inhabit the nucleus or the cytosol. Steroid binding releases the receptors from these complexes, whereupon they dimerize. In the case of cytosolically located receptors, steroid binding is thought to also unmask their previously sequestered nuclear localization signals (NLS; Section 32-6Be), thereby causing the steroid–receptor complexes to be transported to the nucleus. [Most NLSs consist of a 48-residue segment of mainly basic residues or two such segments separated by an 8- to 12-residue linker that is mutation-resistant; the precise location of an NLS within a polypeptide is unimportant, unlike the case for other types of signal peptides (Section 12-4B).]

*In the nucleus, steroid–receptor complexes bind to specific segments of chromosomal enhancers known as **hormone response elements (HREs)** so as to induce, or in some cases repress, the transcription of their associated genes.* For example, receptors for **glucocorticoids** (a class of steroids that affect carbohydrate metabolism; Section 19-1Ga) bind to specific 15-bp **glucocorticoid response elements (GREs)** in the upstream regions of many genes, including those of metallothioneins. Thus, eukaryotic steroid receptors are inducible transcription factors: Their actions resemble those of prokaryotic transcriptional regulators such as the *E. coli* CAP–cAMP complex (Section 31-3Cb). However, eukaryotic systems are much more complex. For instance, different cell types may have the same receptor for a given steroid hormone and yet synthesize different proteins in response to the hormone. Apparently, only some of the genes inducible by a given steroid are made available for activation in each type of cell responsive to that steroid. Consequently, a given eukaryotic sequence-specific regulator may function as an activator or a repressor depending

on the identities of the proteins with which it is interacting. The structures of steroid receptors are discussed below.

k. Eukaryotic Transcription Factors Have a Great Variety of DNA-Binding Motifs

How do DNA-binding transcription factors recognize their target DNA sequences? In prokaryotes, as we have seen (Section 31-3Da), most repressors and activators do so via helix-turn-helix (HTH) motifs and, in a few cases, via β ribbon motifs. Eukaryotes, as we shall see, employ a far greater variety of DNA-binding motifs in their transcription factors. In the following paragraphs we discuss the structures of several of the more common of these motifs and how they bind their target DNAs.

I. Zinc Finger DNA-Binding Motifs

The first of the predominantly eukaryotic DNA-binding motifs was discovered by Aaron Klug in *Xenopus transcription factor IIIA (TFIIIA)*, a protein that binds to the internal control sequence of the 5S rRNA gene (Section 31-2Eg). This complex then sequentially binds TFIIIB (which contains TBP), **TFIIIC**, and RNA polymerase III, which, in turn, initiates transcription of the 5S rRNA gene. The 344-residue TFIIIA contains nine similar, tandemly repeated, ~30-residue modules, each of which contains two invariant Cys residues, two invariant His residues, and several conserved hydrophobic residues (Fig. 34-60). Each of these motifs binds a Zn^{2+} ion, which X-ray absorption measurements indicate is tetrahedrally liganded by the invariant Cys and His residues. Sequence analyses have since revealed that these so-called **zinc fingers** occur from 2 to over 60 times in numerous eukaryotic transcription factors, including Sp1, several *D. melanogaster* developmental regulators (Section 34-4Bb), and certain proto-oncogene proteins (proteins whose mutant forms promote cancerous growth; Section 19-3Ba), as well as the *E. coli* UvrA protein (Section 30-5Ba). In fact, it is estimated that ~1% of mammalian proteins contain zinc fingers. In some

zinc fingers, the two Zn^{2+} -liganding His residues are replaced by two additional Cys residues, whereas others have six Cys residues liganding two Zn^{2+} ions. Indeed, as we shall see, structural diversity is a hallmark of zinc finger proteins. In all cases, however, *the Zn^{2+} ions appear to knit together relatively small globular domains, thereby eliminating the need for much larger hydrophobic cores*.

m. Cys_2 - His_2 Zinc Fingers: Zif268

The first reported zinc finger-containing X-ray structure was that of an 87-residue segment of the 227-residue mouse transcription factor **Zif268**. This structure, which was determined by Carl Pabo, incorporates the protein's three zinc fingers in complex with a dsDNA containing the protein's 9-bp consensus binding sequence. The structures of the three Zif268 zinc finger motifs (Fig. 34-61a) are closely superimposable and are arranged as separate domains in a C-shaped structure that fits snugly into the DNA's major groove (Fig. 34-61b). Each zinc finger interacts in a conformationally identical manner with successive 3-bp segments of the DNA, predominantly through hydrogen bonding interactions between the zinc finger's α helix and one strand of the DNA (here, a G-rich strand). Each zinc finger makes specific hydrogen bonding contacts with two bases in the major groove. Interestingly, five of these six associations involve interactions between Arg and G residues. In addition to these sequence-specific interactions, each zinc finger hydrogen bonds with the DNA's phosphate groups via conserved Arg and His residues.

The Cys_2 - His_2 zinc finger broadly resembles the prokaryotic HTH motif as well as most other DNA-binding motifs we shall encounter (including other types of zinc finger modules) in that *all of these DNA-binding motifs provide a platform for inserting an α helix into the major groove of B-DNA*. However, Cys_2 - His_2 zinc finger proteins, unlike those containing other DNA-binding motifs, possess repeated protein modules that each contact successive DNA segments. Such a modular system can recognize extended asymmetric base sequences.

n. Cys_2 - Cys_2 Zinc Fingers: The Glucocorticoid Receptor and Estrogen Receptor DNA-Binding Domains

The **nuclear receptor superfamily**, which occurs in animals ranging from worms to humans, is composed of >150 proteins that bind a variety of hormones such as steroids (glucocorticoids, mineralocorticoids, progesterone, estrogens, and androgens; Section 19-1G), thyroid hormones (Section 19-1D), vitamin D (Section 19-1E_b), and **retinoids** (Section 34-4B₁). However, the ligands, if any, that many superfamily members bind are, as yet, unknown and hence these proteins are known as **orphan receptors**. The nuclear receptors, many of which activate distinct but overlapping sets of genes, share a conserved modular organization that includes, from N- to C-terminus, a poorly conserved trans-activation domain, a highly conserved DNA-binding domain, a connecting hinge region, and a ligand-binding domain. The DNA-binding domains contain 8 Cys residues that, in groups of four, tetrahedrally coordinate two Zn^{2+}

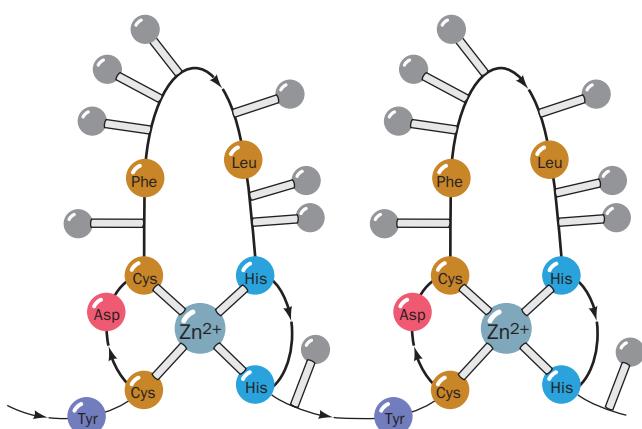


Figure 34-60 Schematic diagram of tandemly repeated Cys_2 - His_2 zinc finger motifs indicating their tetrahedrally liganded Zn^{2+} ion. Conserved amino acid residues are labeled. Gray balls represent the most probable DNA-binding side chains. [After Klug, A. and Rhodes, D., *Trends Biochem. Sci.* **12**, 465 (1988).]

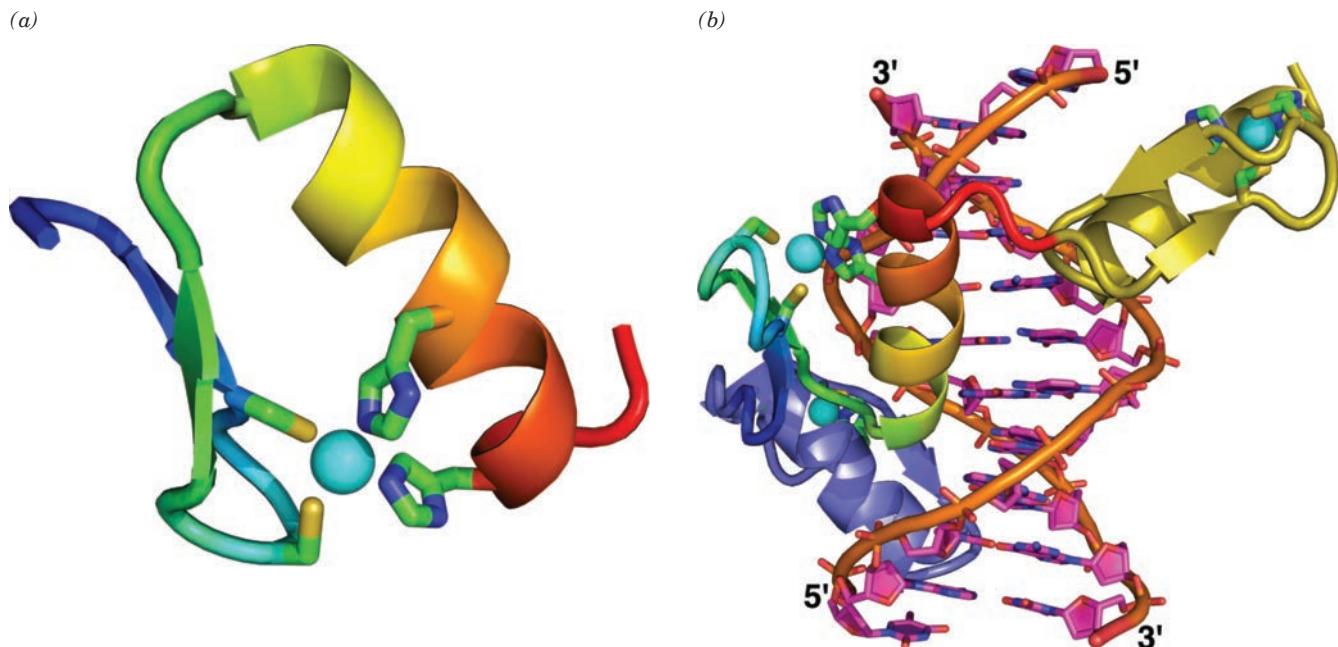


Figure 34-61 X-ray structure of the three-zinc finger segment of Zif268 in complex with a 10-bp DNA. (a) A ribbon diagram of the central zinc finger motif (finger 2). The protein segment is drawn in ribbon form and colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the Cys and His residues that tetrahedrally ligand a Zn^{2+} ion (cyan sphere) are drawn in stick form with C green, N blue, and S yellow. (b) The complex of the entire protein segment with a 10-bp DNA with a single nucleotide overhang at each end (the strand whose 5' end is on the upper right is 5'-AGCGTGGGCGT-3').

The DNA is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods. The protein is drawn in cartoon form with finger 1 lavender, finger 2 colored as in Part a, and finger 3 dark yellow. The Zn^{2+} ions are represented by cyan spheres and their liganding His and Cys side chains are drawn in stick form with C green, N blue, and S yellow. Note how the N-terminal end of each zinc finger's helix extends into the DNA's major groove to contact three base pairs. [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1ZAA.] See the Interactive Exercises.

ions. Many members of the nuclear receptor superfamily recognize hormone response elements that have the half-site consensus sequences 5'-AGAACCA-3' for **steroid receptors** and 5'-AGGTCA-3' for other nuclear receptors. These sequences are arranged in direct repeats ($\rightarrow n \rightarrow$), inverted repeats ($\rightarrow n \leftarrow$), and everted repeats ($\leftarrow n \rightarrow$), where n represents a 0- to 8-bp spacer (usually 1–5 bp) to whose length a specific receptor is targeted. Steroid receptors bind to their hormone response elements as homodimers, whereas other nuclear receptors do so as homodimers, heterodimers, and in a few cases, as monomers.

The X-ray structures of two related DNA segments complexed with the 86-residue DNA-binding domain of rat **glucocorticoid receptor (GR)** were determined by Sigler and Keith Yamamoto. One segment, designated GRE_{4S}, contains two ideal 6-bp **glucocorticoid response element (GRE)** half-sites arranged in inverted repeats about a 4-bp (nonnative) spacer ($n = 4$), whereas the other DNA, GRE_{3S}, differs from GRE_{4S} in that its spacer has the naturally occurring length of $n = 3$ bp. In both complexes, the protein forms a symmetric dimer involving protein–protein contacts even though it exhibits no tendency to dimerize in the absence of DNA (NMR measurements indicate that the dimer contact region is flexible in solution).

The X-ray structure of the DNA-binding domain of the GR subunit complexed to DNA resembles that of its NMR

structure in the absence of DNA: It consists of two structurally distinct modules, each nucleated by a Zn^{2+} coordination center, that closely associate to form a compact globular fold (Fig. 34-62a). The C-terminal module provides the entire dimerization interface as well as making several contacts with the phosphate groups of the DNA backbone. The N-terminal module, which is also anchored to the phosphate backbone, makes all of the GR's sequence-specific interactions with the GRE via three side chains extending from the N-terminal α helix, its recognition helix, which is inserted into the GRE's major groove.

In the GRE_{3S} complex, a subunit of the GR DNA-binding domain binds to each GRE half-site in a structurally identical manner, making sequence-specific contacts even though the odd number of base pairs in its spacer, which the protein does not contact, renders the DNA sequence nonpalindromic. However, in the GRE_{4S} complex (Fig. 34-62b), the protein dimer maintains a structure that is essentially identical to that in the GRE_{3S} complex so that only one of its subunits can bind to a GRE half-site in a manner resembling that in the GRE_{3S} complex. The other subunit is shifted out of register with the GRE sequence by 1 bp and hence only makes nonspecific contacts with the DNA. The dimer interactions are apparently stronger than the protein–DNA interactions, a surprising finding in light of the protein's failure to dimerize in the absence of DNA.

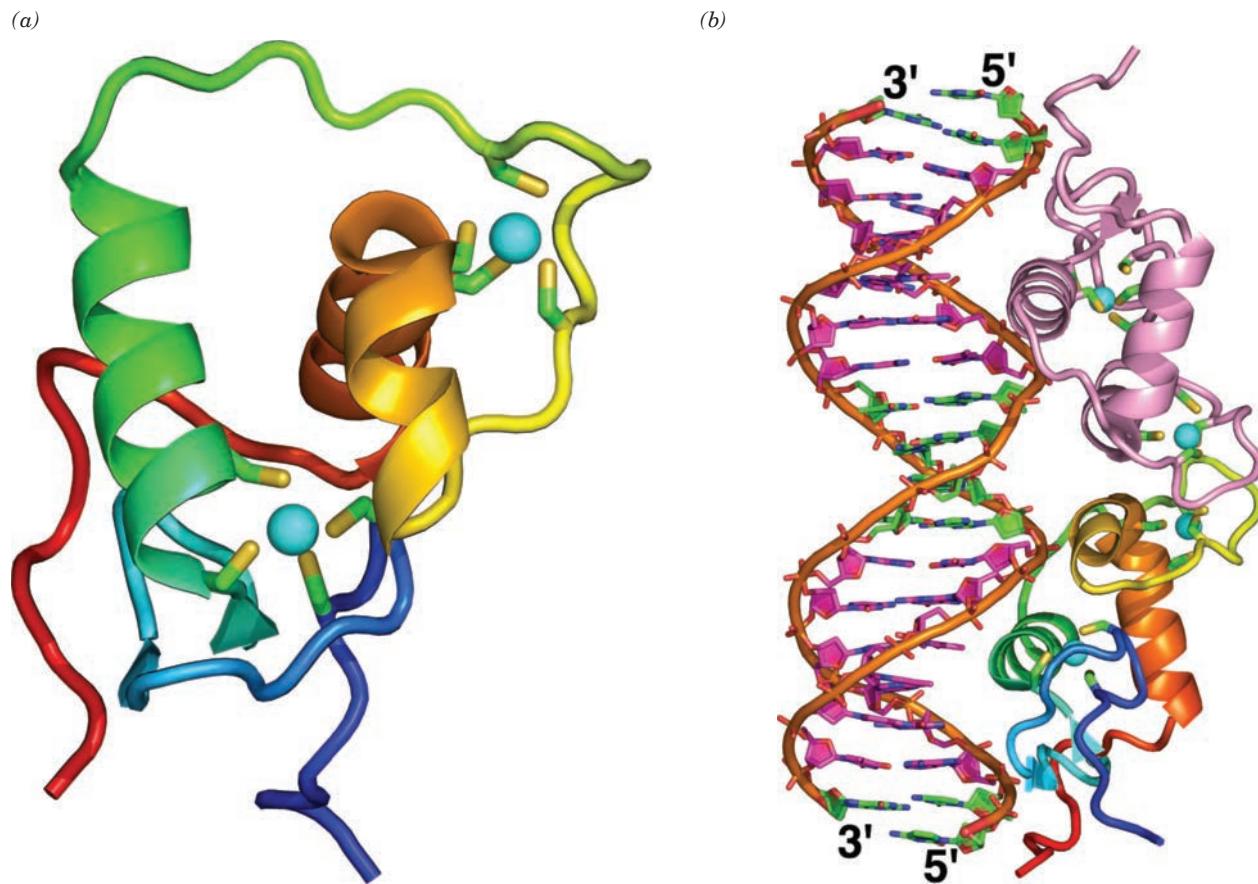


Figure 34-62 X-ray structure of the dimeric glucocorticoid receptor (GR) DNA-binding domain in complex with an 18-bp dsDNA. The DNA contains two inverted repeats of the 6-bp glucocorticoid response element (GRE) half-sites (5'-AGAACCA-3') separated by a 4-bp spacer (GRE_{4S}). (a) A ribbon diagram of a single subunit of the GR colored in rainbow order from its N-terminus (blue) to its C-terminus (red). Its two Zn²⁺ ions are represented by cyan spheres and their tetrahedrally liganding Cys side chains are shown in stick form with C green and S yellow. Compare this structure with Fig. 34-61a. (b) The complex of the homodimeric protein with GRE_{4S} DNA as viewed with its approximate 2-fold molecular axis horizontal. The protein is shown in ribbon form with its lower subunit colored as in Part a, its upper subunit pink, and its bound

Zn²⁺ ions represented by cyan spheres. The DNA is drawn in paddle form with the C atoms of its two 6-bp GRE half-sites magenta, the C atoms of the remaining nucleotides green, N blue, O red, and P orange and with successive P atoms connected by orange rods. Note how the GR's two N-terminal helices are inserted into adjacent major grooves of the DNA. However, only the lower subunit binds to the DNA in a sequence-specific manner; the upper subunit binds to the palindromic DNA one base pair closer to the center of the DNA molecule than does the upper subunit and hence does not make sequence-specific contacts with the DNA. [Based on an X-ray structure by Paul Sigler, Yale University. PDBid 1GLU.]

 See the Interactive Exercises.

Thus, the two subunits and the DNA associate in a cooperative fashion that favors the binding of the glucocorticoid receptor to targets with properly spaced half-sites.

The **estrogen response element (ERE)**, the DNA segment to which the **estrogen receptor (ER)** specifically binds, differs from the GRE only by changes in the central two base pairs in their otherwise identical 6-bp half-sites. The X-ray structure of the ER DNA-binding domain in complex with an ERE-containing DNA segment, determined by Rhodes, closely resembles that of the GR-GRE complex. However, the side chains that make base-specific contacts with each ERE half-site are quite differently arranged from those contacting the GRE_{4S} half-sites. Evidently, the discrimination of a half-site sequence is not simply a matter of substituting one or more different amino

acid residues into a common framework but, rather, involves considerable side chain rearrangement.

Members of the nuclear receptor superfamily often recognize hormone response elements with similar or even identical half-site sequences as well as different spacings. The foregoing observations provide a structural basis for the graded affinities of these receptors toward their various target genes.

o. Binuclear Cys₆ Zinc Fingers: The GAL4 DNA-Binding Domain

The yeast protein **GAL4** is a transcriptional activator of several genes that encode galactose-metabolizing proteins. This 881-residue protein binds to a 17-bp DNA segment as a homodimer. Residues 1 to 65, which contain six Cys

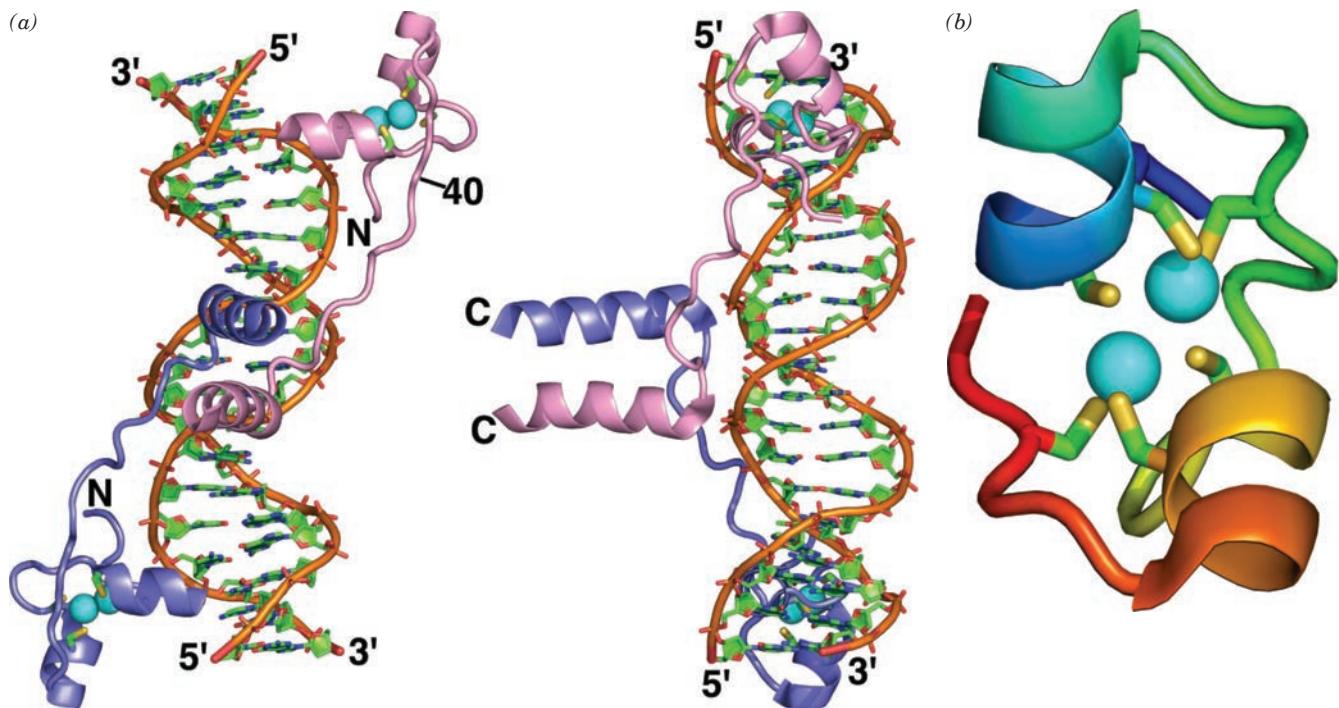


Figure 34-63 X-ray structure of the yeast GAL4 DNA-binding domain in complex with a palindromic 19-bp DNA (except for the central base pair) containing the protein's consensus binding sequence. (a) The complex of the dimeric protein with the DNA. The DNA is drawn in paddle form with C green, N blue, O red, and P orange and with successive P atoms connected by orange rods. The protein is shown in ribbon form with one subunit pink and the other lavender. The two Zn^{2+} ions bound to each subunit are represented by cyan spheres and their six liganding Cys side chains are drawn in stick form with C green and S yellow. The views are along the complex's 2-fold axis (left)

and turned 90° about the DNA helix axis so that the 2-fold axis is horizontal (right). Note how the C-terminal end of each subunit's N-terminal helix extends into the DNA's major groove. (b) A ribbon diagram of the protein's zinc finger domain (residues 8–40) colored in rainbow order from its N-terminus (residue 8; blue) to its C-terminus (red). The Cys side chains of its Zn_2Cys_6 complex is shown in stick form with C green and S yellow, and its Zn^{2+} ions are represented by cyan spheres. Compare this structure with Figs. 34-61a and 34-62a. [Based on an X-ray structure by Stephen Harrison and Ronen Marmorstein, Harvard University. PDBid 1D66.]  See the Interactive Exercises.

residues that collectively ligand two Zn^{2+} ions (Fig. 34-63), have been implicated in DNA binding; residues 65 to 94 participate in dimerization (although, as we shall see, residues 50–64 also have a weak dimerization function); and residues 94 to 106, 148 to 196, and 768 to 881 function as acidic transcriptional activating regions. The X-ray crystal structure of the 65-residue N-terminal fragment of GAL4 in complex with a symmetrical 19-bp DNA containing GAL4's palindromic 17-bp consensus sequence was determined by Mark Ptashne, Ronen Marmorstein, and Stephen Harrison.

The protein binds to the DNA as a symmetric dimer (Fig. 34-63a), although in the absence of DNA it is only monomeric. Each subunit folds into three distinct modules: a compact Zn^{2+} -ligand domain that binds specific sequences of DNA (residues 8–40), an extended linker (residues 41–49), and a short α helical dimerization element (residues 50–64). In the Zn^{2+} -ligand module (Fig. 34-63b and top and bottom of Fig. 34-63a), the two Zn^{2+} ions are each tetrahedrally coordinated by four of the six Cys residues, with two of these residues ligating both metal ions so as to form a binuclear cluster. This module's polypeptide chain forms two short α helices connected by a

loop such that the module, together with its bound Zn^{2+} ions, has pseudo-2-fold symmetry. The N-terminal helix is inserted into the DNA's major groove, thereby making sequence-specific contacts with a highly conserved CCG sequence at each end of the consensus sequence. The DNA's conformation deviates little from that of ideal B-DNA.

The dimerization helices (center of Fig. 34-63a) associate to form a short segment of parallel coiled coil in which the contact region between the coiled coil's component helices is hydrophobically stabilized by three pairs of Leu residues and a pair of Val residues (an arrangement similar to that in the so-called **leucine zipper** described below). The coiled coil is positioned over the minor groove of the DNA such that its superhelix axis coincides with the DNA's 2-fold axis. The linkers connecting the coiled coil to the DNA-binding modules wrap around the DNA, largely following its minor groove while making several nonspecific contacts with DNA phosphate groups until, on reaching the DNA-binding module, they shift over into the DNA's major groove. The two symmetrically related DNA-binding modules thereby approach the major groove from opposite sides of the DNA, ~1.5 helical turns apart, rather than from the same side of the DNA, ~1 helical turn apart,

as do, for example, HTH motifs and the glucocorticoid receptor. The resulting relatively open structure could permit other proteins to bind simultaneously to the DNA.

p. Leucine Zippers Mediate Transcription

Factor Dimerization

Transcriptional activation requires, as we have seen, the cooperative association of several proteins that bind to specific sequences on DNA. Steven McKnight discovered one way in which such associations occur. We have seen (Section 8-2A) that α helices with the 7-residue pseudorepeating sequence $(a-b-c-d-e-f-g)_n$, in which the a and d residues are hydrophobic, have a hydrophobic strip along one side, which induces them to dimerize so as to form a coiled coil. McKnight noticed that the rat liver transcription factor named **C/EBP** (for *CCAAT/enhancer binding protein*), which specifically binds to the CCAAT box (Section 31-2Ee), has a Leu at every seventh position of a 28-residue segment in its DNA-binding domain. Similar heptad repeats occur in a number of known dimeric DNA-binding proteins, including the yeast transcriptional activator **GCN4** and several DNA-binding proteins encoded by proto-oncogenes (Section 34-4Ca). McKnight suggested that these proteins form coiled coils in which the Leu side chains are interdigitated, much like the teeth of a zipper. He therefore named this motif the **leucine zipper**. The leucine zipper, as we shall see, mediates both the homodimerization and the heterodimerization of DNA-binding proteins (but note that it is not, in itself, a DNA-binding motif).

The X-ray structure of the 33-residue polypeptide corresponding to the leucine zipper of the 281-residue GCN4 was determined by Peter Kim and Thomas Alber. Its first 30 residues, which contain \sim 3.6 heptad repeats (Fig. 34-64a),

coil into an \sim 8-turn α helix that dimerizes as McKnight predicted to form \sim 1/4 turn of a parallel left-handed coiled coil (Fig. 34-64b). The dimer can be envisioned as a twisted ladder whose sides consist of the helix backbones and whose rungs are formed by the interacting hydrophobic side chains. The conserved Leu residues at heptad position d , which comprise every second rung, are not interdigitated as McKnight originally suggested but, instead, make side-to-side contacts. The alternate rungs are likewise formed by the a residues of the heptad repeat (which are mostly Val) in side-to-side contact. Each Leu side chain at position d , in addition to packing against the symmetry-related Leu side chain, d' , from the other polypeptide, packs against the side chain of the succeeding residue, e' . Similarly, each side chain at position a packs between its symmetry mate, a' , and the preceding residue, g' . These two sets of alternating layers thereby form an extensive hydrophobic interface between the coiled coil's component helices.

q. bZIP Motifs: The GCN4 DNA-Binding Domain

In many but not all leucine zipper proteins, a DNA-binding region, which is rich in basic residues, is immediately N-terminal to the leucine zipper. Sequence comparisons among 11 of these so-called **basic region leucine zipper (bZIP) proteins** revealed that the 16-residue basic sequence invariably ends 7 residues before the leucine zipper's N-terminal Leu residue. Moreover, all of these basic regions, as well as the 6-residue segment linking them to the leucine zipper, are devoid of the two strongest helix destabilizing residues, Pro and Gly (Section 9-3Aa), thereby suggesting that each bZIP polypeptide is entirely α helical.

The C-terminal 56 residues of GCN4 constitute its bZIP element. Harrison and Kevin Struhl determined the X-ray

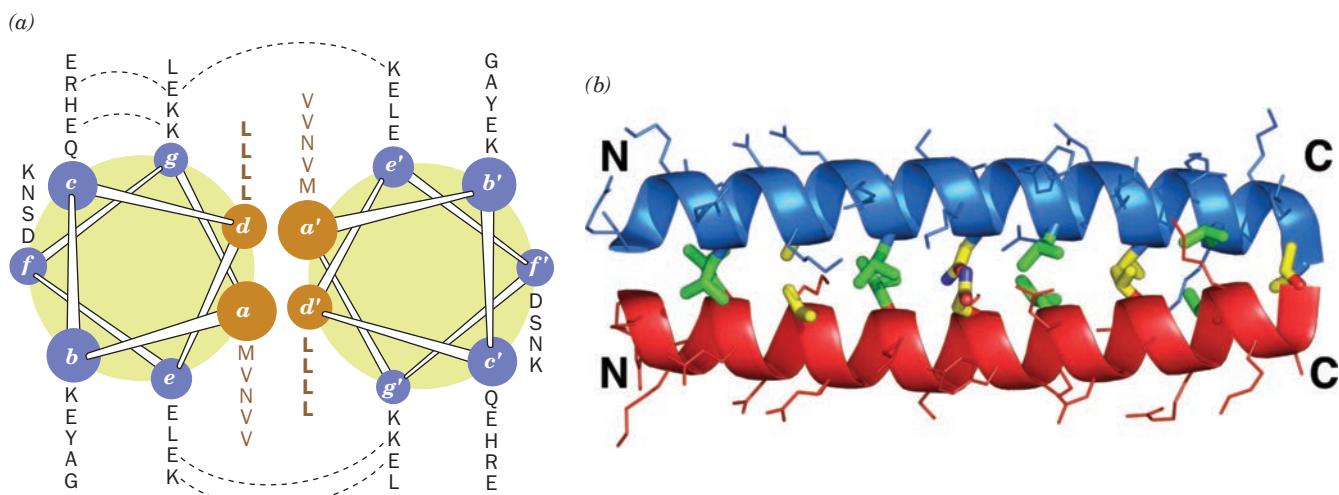


Figure 34-64 The GCN4 leucine zipper motif. (a) A helical wheel representation of the motif's two helices as viewed from their N-termini. The sequences of residues at each position are indicated by the adjacent column of one-letter codes. Residues that form ion pairs in the crystal structure are connected by dashed lines. Note that all residues at positions d and d' are Leu (L), those at positions a and a' are mostly Val (V), and those at other positions are mostly polar. [After O'Shea, E.K., Klemm,

J.D., Kim, P.S., and Alber, T., *Science* **254**, 540 (1991).] (b) The X-ray structure, in side view, in which the two identical helices are drawn in ribbon form. Side chains are shown in stick form with the contacting Leu side chains at positions d and d' yellow and side chains at positions a and a' with C green, N blue, and O red. [Based on an X-ray structure by Peter Kim, MIT, and Tom Alber, University of Utah School of Medicine. PDBID 2ZTA.]

See Kinemage Exercise 21

structure of this polypeptide segment in complex with a 19-bp-containing duplex DNA whose central 9 bp consist of GCN4's symmetrized target sequence (Fig. 34-65). The bZIP element forms a symmetric dimer in which each subunit consists, almost entirely, of a continuous α helix. The C-terminal 25 residues of two such helices associate via a leucine zipper whose geometry closely resembles that of the 33-residue GCN4 leucine zipper element alone (Fig. 34-64b). Past this point, the two α helices smoothly diverge to bind in the DNA's major groove on opposite sides of the helix, thereby clasping the DNA in a sort of scissors grip. The DNA, whose helix axis is nearly perpendicular to that of the coiled coil, maintains what is essentially a straight and undistorted B-form conformation. The basic region residues that are conserved in bZIP proteins thereby make numerous contacts with both the bases and with phosphate oxygens of the DNA target sequence.

r. bHLH Motifs: The Max DNA-Binding Domain

The **basic helix-loop-helix (bHLH) motif**, which occurs in a variety of eukaryotic transcription factors, contains a conserved DNA-binding basic region. This is immediately followed by two amphipathic helices connected by a loop that mediates the protein's dimerization. The bHLH motif in many proteins is followed by a conserved leucine zipper (Z) motif that presumably augments protein dimerization. The transcription factor **Max** is such a **bHLH/Z** protein, which, *in vivo*, forms a heterodimer with the proto-oncogene protein **Myc** and is required for both its normal and cancer-inducing activities. Max, by itself, readily homodimerizes and binds DNA with high affinity but Myc does not do so.

The X-ray structure of a truncated version of the 160-residue Max, Max(22-113), which contains the parent protein's bHLH and leucine zipper elements, was determined, by Edward Ziff and Burley, in complex with a 22-bp quasi-palindromic DNA containing Max's 6-bp central recognition element. Each subunit of this homodimeric protein consists of two long α helices connected by a loop to form a novel protein fold (Fig. 34-66). The N-terminal α helix (b/H1) contains residues from the protein's basic region (b) followed, without interruption, by those of the HLH motif's leading helix (H1). The C-terminal α helix (H2/Z), which is composed of the second HLH helix (H2) and the leucine zipper (Z), mediates the protein's homodimerization through the formation of a parallel left-handed coiled coil similar to that in GCN4 (Fig. 34-65). Each of the dimer's two b/H1 helices projects from the resulting parallel 4-helix bundle to engage the DNA in a manner reminiscent of a pair of forceps by binding in its major groove on opposite sides of the helix (much like the way GCN4 grips its target DNA, although GCN4's bZIP element consists of only two α helices rather than the four of Max). The DNA helix is essentially straight with only small deviations from the ideal B-DNA structure. Each basic region makes several sequence-specific interactions with the bases of the DNA's 6-bp recognition element as well as numerous contacts with its phosphate groups. Side chains of both the loop and the N-terminal end of the H2 helix also contact DNA phosphate groups.

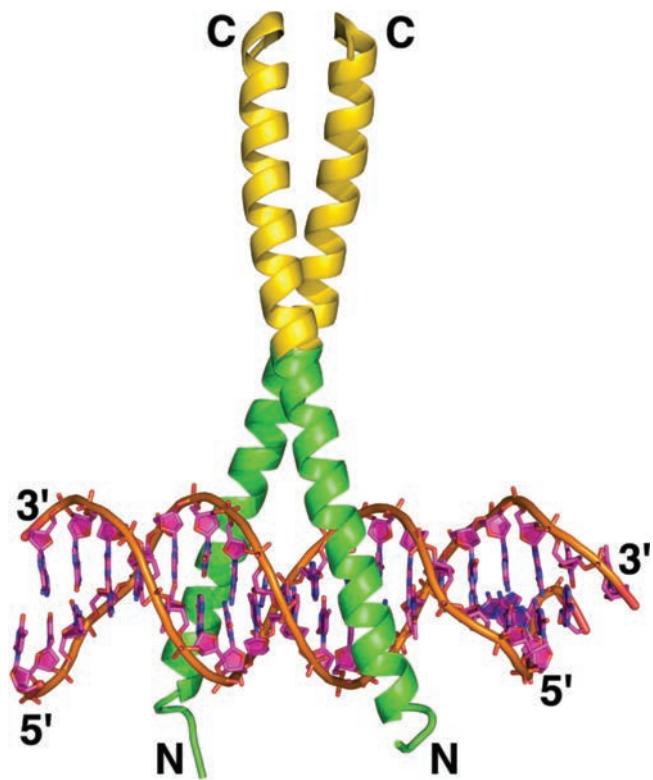


Figure 34-65 X-ray structure of the GCN4 bZIP region in complex with its target DNA. The complex is viewed with its 2-fold axis vertical. The DNA, which is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods, consists of a 19-bp segment with a single nucleotide overhang at each end and contains the protein's palindromic (except for the central base pair) 7-bp target sequence. The two identical protein subunits, shown in ribbon form, each contain a continuous 52-residue α helix. At their C-terminal ends (yellow), the two subunits associate in a parallel coiled coil (a leucine zipper), and at their basic regions (green), they smoothly diverge to each engage the DNA in its major groove at the target sequence. [Based on an X-ray structure by Stephen Harrison, Harvard University. PDBID 1YSA.] See the Interactive Exercises.

s. NF- κ B Binds DNA Differently from Other Transcription Factors

Nuclear factor κ B (NF- κ B), a transcription factor that was originally identified as an inducible nuclear activity that binds to the κ B sequence in the immunoglobulin κ light chain gene enhancer (immunoglobulin genes are discussed in Section 35-2), is present in nearly all animal cells, although its role is particularly prominent in the immune system. It is present *in vivo* mainly as a DNA-binding heterodimer of the **p50** and **p65** (alternatively **RelA**) proteins (p for protein with the number indicating its nominal molecular mass in kD), both of which contain an \sim 300-residue segment known as the **Rel homology region (RHR)** because it also occurs in the product of the **rel** oncogene. However, p50 and p65 can also form DNA-binding homodimers. RHRs, which mediate protein dimerization and DNA binding, and which contain nuclear localization

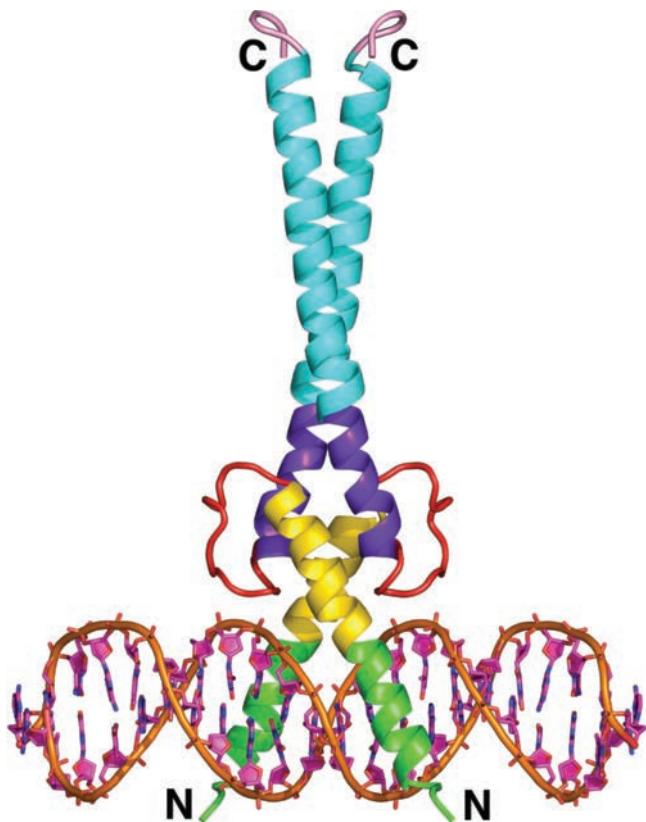


Figure 34-66 X-ray structure of the Max(22-113) dimer in complex with a 22-bp DNA containing the protein's palindromic 6-bp target sequence. The complex is viewed with its 2-fold axis vertical. The DNA is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods. The homodimeric protein is shown in ribbon form. Its N-terminal basic region (green) forms an α helix that engages its target sequence in the DNA's major groove and then merges smoothly with the H1 helix (yellow) of the helix-loop-helix (HLH) motif. Following the loop (red), the protein's two H2 helices (purple) of the HLH motif form a parallel left-handed four-helix bundle with the two H1 helices. Each H2 helix then merges smoothly with the leucine zipper (Z) motif (cyan) to form a parallel coiled coil. The protein's C-terminal ends are pink. [Based on an X-ray structure by Edward Ziff and Stephen Burley, The Rockefeller University. PDBid 1AN2.]

signals (NLSs), are present in a variety of proteins that serve as regulators of cellular defense mechanisms against stress, injury, and external pathogens, as well as of differentiation. Moreover, certain viruses, including HIV, have subverted RHRs to activate the expression of their genes. There are two classes of **Rel proteins**: those such as p65, **c-Rel**, and the *Drosophila* morphogen proteins (proteins that mediate development; Section 34-4Ba) **Dorsal** and **Dif**, whose N-terminal domains contain an RHR and whose highly variable C-terminal domains are strong transcriptional activators; and those such as p50 and the closely related **p52**, which are generated by the proteolytic processing of larger precursors and lack transactivation domains so that their homodimers function primarily as repressors.

The activity of NF- κ B is largely regulated by proteins known as **inhibitor- κ Bs (I κ Bs)**, which by binding to an NF- κ B mask its NLS so that I κ B–NF- κ B complexes reside in the cytoplasm. The I κ Bs contain multiple ankyrin repeats (Section 12-3Db) through which they bind the NF- κ Bs. The extracellular presence of a remarkable variety of external stimuli, including certain bacterial and viral products, several cytokines (Section 19-1Lb), phorbol esters (Section 19-4Ca), and oxidative and physical stress (e.g., free radicals and UV radiation), results, via signaling cascades, in I κ Bs being phosphorylated by **I κ B kinase (IKK)**. This, in turn, induces ubiquitination of the I κ Bs and their subsequent degradation by the proteasome (Section 32-6B1). The liberated NF- κ B is thereupon translocated to the nucleus, where it mediates transcriptional initiation by binding to 10-bp κ B DNA segments that have the consensus sequence GGGRNYYCC. Additional specificity may be achieved through the synergistic interaction of the NF- κ B with other DNA-bound transcription factors such as Sp1. This activation process is self-limiting: The transcription of the gene encoding the most common I κ B protein, **I κ B α** , is induced by the binding of NF- κ B to the κ B sites in this gene's promoter. The resulting newly synthesized I κ B α enters the nucleus, where it releases the NF- κ B from its complex with DNA and directs its export to the cytoplasm.

In a related mode of NF- κ B activation, p50 is synthesized as the N-terminal domain of **p105**, a protein whose C-terminal domain is an I κ B. The I κ B domain of p105 prevents both the nuclear localization and the DNA binding of p105 as well as other RHR-containing proteins. The above external stimuli also accelerate the proteolytic processing of p105 to yield a free NF- κ B and the I κ B-containing C-terminal domain of p105, which as discussed above, is phosphorylated and proteolytically degraded.

The X-ray structure of the heterodimer of mouse p50 and p65 in complex with the κ B segment of the β -interferon enhancer, determined by Gourisankar Ghosh, bears a striking resemblance to a butterfly with its homologous protein subunits forming its outspread wings and the DNA its torso (Fig. 34-67). The two protein subunits have similar structures that each consists of two domains, with the C-terminal domains forming the dimerization interface and both domains interacting with the DNA. Both their N- and C-terminal domains have immunoglobulin-like folds (a sandwich of a 3- and a 4-stranded antiparallel β sheet; Section 35-2Be) and interact with the DNA exclusively through 10 loops, 5 from each subunit, that link their β strands and fill the DNA's major groove. The p50 and p65 bind to 5-bp and 4-bp subsites at the 5' and 3' ends of the consensus sequence, respectively, with the two subsites separated by a single base pair. The protein's DNA-binding surface is much more extensive than those of other transcription factors, which accounts for the unusually high affinity of NF- κ Bs for their target sequences. This also explains the inability of deletion mutagenesis to localize NF- κ B's DNA-binding region, since changes anywhere in its structure are likely to affect the disposition of its DNA-binding loops.

Comparisons of the X-ray structures of the p50–p65 heterodimer bound to several κ B DNA segments with

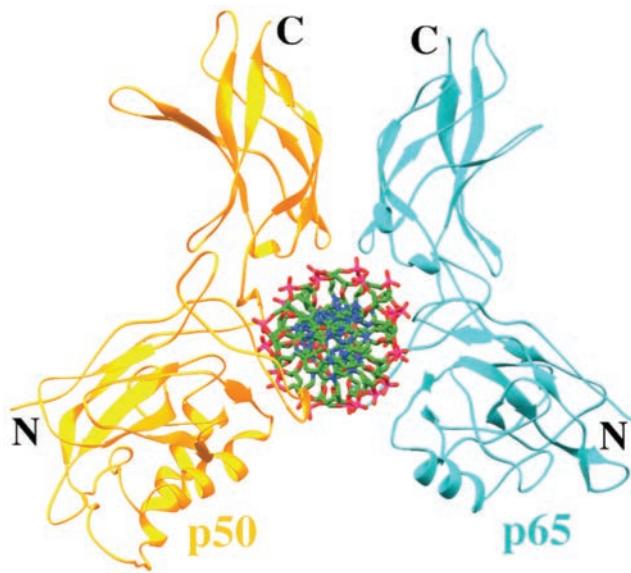


Figure 34-67 X-ray structure of the mouse NF-κB p50-p65 heterodimer bound to κB DNA from the interferon β enhancer. The structure is viewed along the helix axis of the DNA, whose two strands have the sequences 5'-TGGGAAATTCCCT-3' and 5'-AAGGAATTCCCC-3' (the duplex DNA consists of 11 bp with a 1-nt overhang at each end) and are drawn in stick form colored according to atom type (C green, N blue, O red, and P magenta). The protein is represented by ribbons with p50 (residues 39–364 of 435 residues) gold and p65 (residues 19–291 of 549 residues) cyan. [Based on an X-ray structure by Gourisankar Ghosh, University of California at San Diego. PDBid 1LE5.]

different sequences, all determined by Ghosh, reveal small but significant structural differences among the various complexes. These arise mainly from the different degrees of bending of the various κB DNAs as well as the different interactions of the proteins with the different sequences of bases, all of which result in small conformational differ-

ences among the chemically identical proteins in these complexes. This is probably why the substitution in an enhancer of one κB segment for another does not produce the same level of transcription, even though the NF-κB binds to the isolated κB segments with equal affinity. Evidently, the way in which NF-κB interacts with other proteins that are bound to the enhancer (such as the glucocorticoid receptor, which interacts with p65) affects its activational potency and hence fine-tunes the expression levels of its target genes.

NF-κB is a component of the enhanceosome (Section 34-3Bi) that controls the expression of type β interferon (IFN- β ; Section 32-4Ab). This virus-induced enhanceosome also contains the transcription factors **c-Jun** (Section 19-3Bb), **activating transcription factor 2 (ATF-2)**, which associate via a leucine zipper, and two copies each of the **interferon response factors IRF-3** and **IRF-7**. These eight polypeptides are all bound to a nearly invariant 55-bp enhancer (even more highly conserved than the gene's coding sequence) that spans the interval from -102 to -51 bp upstream of the IFN- β gene's transcription start site. Although each of these transcription factors is known to individually induce the transcription of other genes, all must be present in order to activate the IFN- β gene and hence they must act in a highly cooperative manner. The X-ray structures of three overlapping combinations of the DNA-binding domains of these transcription factors in their complexes with various portions of the IFN- β enhancer, all determined by Tom Maniatis and Harrison, permitted them to construct a composite model of the IFN- β enhanceosome (Fig. 34-68). The eight protein subunits in this \sim 160-Å-long model form a surface that recognizes a contiguous 50-bp segment of the enhancer DNA and thereby accounts for the enhancer's evolutionary invariance. Nevertheless, with the exception of the leucine zipper joining c-Jun and ATF-2 and the association between NF-κB's two subunits, the protein subunits have few direct contacts. The enhancer DNA has the B form with its helix axis following

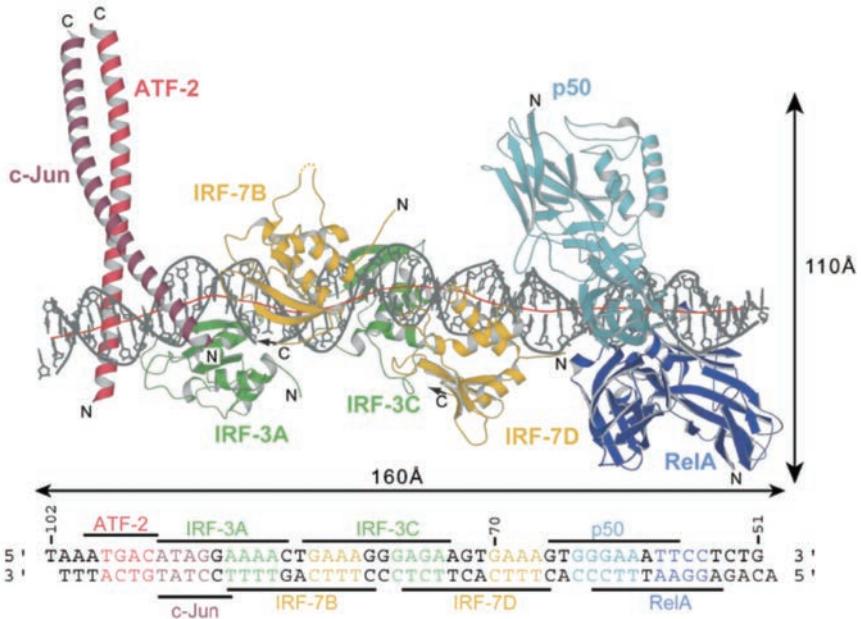


Figure 34-68 Composite model of the IFN- β enhanceosome. Its protein subunits are drawn in ribbon form with c-Jun (maroon), ATF-2 (red), both subunits of IRF-3 (green), both subunits of IRF-7 (yellow), p50 (cyan), and RelA (p65) (blue). The DNA (gray) is shown in stick form with its successive phosphate groups connected by gray rods and with the red line tracing its helix axis. The enhancer's sequence is given below with the underscores and overscores indicating its segments contacted by the various protein subunits. [Courtesy of Tom Maniatis, Harvard University; and Daniel Panne and Stephen Harrison, Harvard Medical School. PDBids 2O6G, 2O61, and 1T2K.]

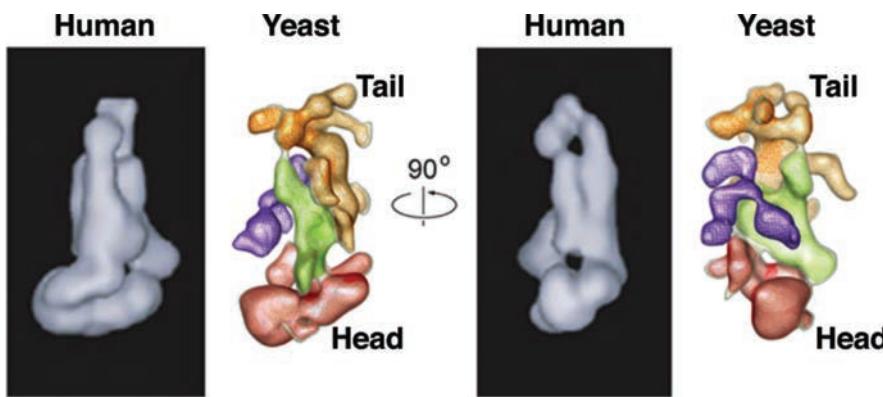


Figure 34-69 Comparison of the EM-based structures of human and yeast Mediators. In the yeast Mediator, the head (red), middle, (green), tail (orange) and arm (purple) modules are defined according to their independent mobilities. Both assemblies are $\sim 400 \text{ \AA}$ tall. [Courtesy of Francisco Asturias, The Scripps Research Institute, La Jolla, California.]

a gentle sinusoidal curve that has a net overall bend of $\sim 14^\circ$. This suggests that the cooperativity among the enhanceosome's eight protein subunits arises from their binding-induced changes in the DNA's conformation.

t. Mediator Provides the Interface between Transcriptional Activators and RNAP II

Eukaryotic genomes encode as many as several thousand transcriptional regulators for class II genes (e.g., Fig. 34-31). How does the binding of these various regulators to their cognate enhancers/silencers influence the rate at which RNAP II initiates transcription? Genetic studies have implicated TFIIB, TFIID, and TFIIH in this process *in vivo* (e.g., see Fig. 34-59). Nevertheless, activators fail to stimulate transcription by a reconstituted PIC *in vitro*. Evidently, an additional factor is required to do so. *Indeed, genetic studies in yeast by Kornberg led him to discover a 21-subunit, $\sim 1200\text{-kD}$ complex named Mediator, whose presence is required for transcription from nearly all class II gene promoters in yeast.* Mediator, which is therefore considered to be a coactivator, binds to the unphosphorylated C-terminal domain (CTD) of RNAP II's β' subunit to form the so-called **RNAP II holoenzyme** (recall that RNAP II initiates transcription only when its CTD is unphosphorylated and commences elongation when its CTD is phosphorylated; Section 31-2E). Further investigations revealed that metazoans contain similar multisubunit complexes that function similarly to yeast Mediator. These complexes were originally assigned a variety of names, including **CRSP**, **NAT**, **ARC/DRIP**, **TRAP/SMCC**, **mMED**, and **PC2**, but are now also known as Mediator. Many of their ~ 26 subunits are related, albeit distantly, to those of yeast Mediator.

A variety of evidence suggests that *Mediator acts as a central scaffold around which the PIC assembles, thus bridging the DNA-bound transcriptional regulators and RNAP II so as to influence (induce or inhibit) transcriptional initiation at the associated promoter. Mediator thereby functions to integrate the various signals implied by the binding of transcriptional regulators to their target DNAs.*

Francisco Asturias has determined the EM-based structures of yeast and human Mediators (Fig. 34-69). The two particles are similarly shaped with nearly perpendicular head and middle-tail modules. However, the EM-based structure of yeast Mediator in complex with RNAP II (Fig. 34-70), also determined by Asturias, reveal that this associ-

ation induces extensive changes in the relative positions of the various Mediator modules and, in addition, its arm module undergoes an internal rearrangement. These changes appear to maximize the contacts between Mediator and RNAP II. Mediator surrounds the entire back side of the RNAP II where TBP and TFIIB bind, consistent with the observation that Mediator binding stabilizes the PIC.

Altogether, the transcriptional machinery for class II genes in humans comprises ~ 73 subunits with an aggregate molecular mass of ~ 3.7 million D. Nevertheless, as we shall see below, this ribosome-sized assembly (the eukaryotic ribosome has a molecular mass of ~ 4.2 million; Table 32-8) requires considerable assistance from yet other large macromolecular assemblies to gain access to the DNA in chromatin.

u. Transcriptionally Active Chromatin Is Sensitive to Nuclease Digestion

Early research on the mechanism of eukaryotic transcription largely ignored the influence of chromatin. Yet, as we have seen (Section 34-3A), euchromatin but not

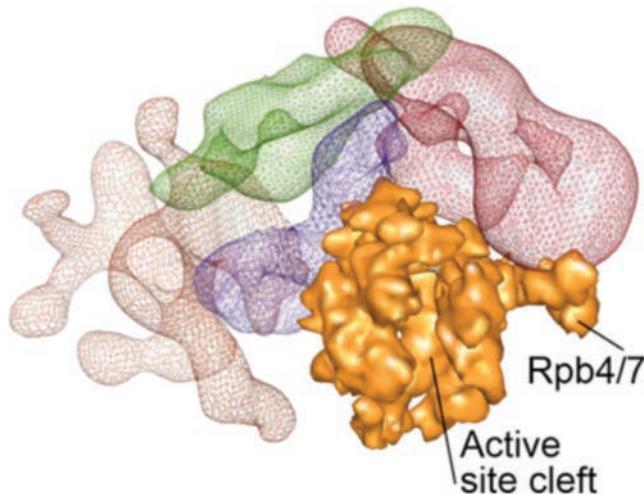


Figure 34-70 EM-based structure of yeast Mediator in complex with RNAP II. Mediator is colored as in Fig. 34-69 and the RNAP II is represented by its surface diagram (orange). Note that Mediator binding leaves RNAP II's active site cleft fully accessible. [Courtesy of Francisco Asturias, The Scripps Research Institute, La Jolla, California.]

heterochromatin is transcriptionally active. *Indeed, numerous investigations have revealed that eukaryotic cells contain elaborate systems that participate in controlling transcriptional initiation by altering chromatin structure.* In the remainder of this subsection we discuss the nature of these systems.

The open structure of transcriptionally active chromatin presumably gives the transcriptional machinery access to the active genes. This hypothesis was corroborated by Harold Weintraub's demonstration that *transcriptionally active chromatin is about an order of magnitude more susceptible to cleavage by DNase I than is transcriptionally inactive chromatin*. For example, globin genes from chicken erythrocytes (avian red cells have nuclei) are more sensitive to DNase I digestion than are those from chicken oviduct (where eggs are made), as was indicated by the loss of the abilities of these genes to hybridize with a complementary DNA probe after DNase I treatment. Conversely, the gene encoding **ovalbumin** (the major egg white protein) from oviduct is more sensitive to DNase I than is that from erythrocytes. Thus, nuclease sensitivity appears to delineate chromatin's **functional domains**, although their relationship to chromatin's structural domains (Section 34-3A) is unclear. Nevertheless, nuclease sensitivity reflects a gene's potential for transcription rather than transcription itself: The DNase I sensitivity of the oviduct ovalbumin gene is independent of whether the oviduct has been hormonally stimulated to produce ovalbumin.

The variation of a given gene's transcriptional activity with the cell in which it is located indicates that chromosomal proteins participate in the gene activation process. Yet the histones' chromosomal abundance and lack of variety make it highly unlikely that they have the specificity required for this role. Among the most common nonhistone proteins are the members of the **high mobility group (HMG)**, so named because of their high electrophoretic mobilities in polyacrylamide gels (and possibly because they were discovered by H.M. Goodwin). These highly conserved, low molecular mass (<30 kD) proteins, which have the unusual amino acid composition of ~25% basic side chains and 30% acidic side chains, are relatively abundant, with ~1 HMG molecule per 10 to 15 nucleosomes. The HMG proteins can be eluted from chick erythrocyte chromatin by 0.35M NaCl without gross structural changes to the nucleosomes. This treatment eliminates the preferential nuclease sensitivity of the erythrocyte globin genes.

v. HMG Proteins Are Architectural Proteins That Participate in Regulating Gene Expression

The HMG proteins consist of three superfamilies, **HMGB**, **HMGA**, and **HMGN**, which have the following properties:

1. The mammalian HMGB proteins, **HMGB1** and **HMGB2** (~210 residues; previously known as **HMG1** and **HMG2**), which bind DNA without regard to sequence, each consist of two tandem ~80-residue **HMG boxes**, A and B, followed by an acidic tail consisting of ~30 (HMGB1) or ~20 (HMGB2) consecutive Asp or Glu residues. However, *Drosophila* **HMG-D** and yeast **NHP6A** proteins each con-

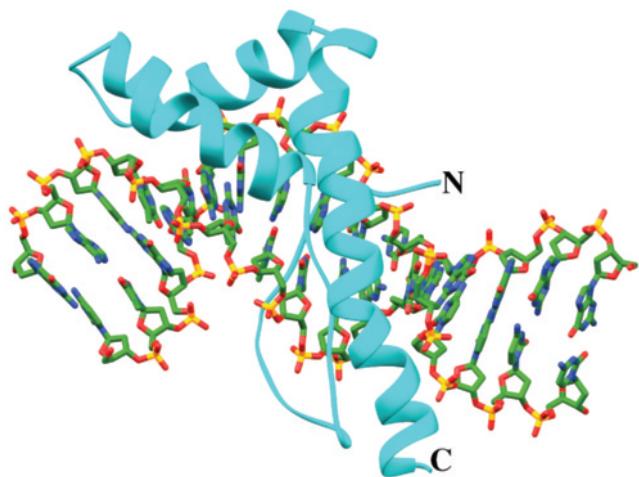


Figure 34-71 NMR structure of yeast NHP6A protein in complex with a 15-bp DNA. The protein is drawn in ribbon form (cyan) and the DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P yellow). The NHP6A's L-shaped HMG box binds in the DNA's minor groove so as to bend the DNA by ~70° toward its major groove. [Based on an NMR structure by Juli Feigon, University of California at Los Angeles. PDBid 1J5N.]

tain only one HMG box, which is closely similar to the B domain of HMGB1. The NMR structure of NHP6A in complex with a 15-bp DNA, determined by Juli Feigon, reveals, in agreement with the structures of several other HMG box-containing proteins, that the HMG box consists of three helices arranged in an L-shape with the inside of the L inserted into the minor groove of the DNA (Fig. 34-71). This induces the DNA to bend by as much as 130° toward its major groove. Apparently, nuclear HMGB proteins function as architectural proteins that induce the binding of other proteins, including various steroid receptors, to DNA and hence facilitate the assembly of nucleoprotein complexes. Indeed, NHP6A and HMGA can functionally replace the bacterial DNA-bending protein HU (Section 30-3Ca) even though HMGB and HU proteins have no structural or sequence similarity. HMG boxes also occur in several sequence-specific transcription factors, including the mammalian male sex determining factor SRY (Section 19-1Gc).

2. The HMGA superfamily consists of four proteins: the 107-, 96-, and 179-residue splice variants **HMGA1a**, **HMGA1b**, and **HMGA1c** (previously named **HMG-I**, **HMG-Y**, and **HMG-I/R**) and the homologous 109-residue **HMGA2** (previously named **HMG-C**). Each of these proteins contains three similar so-called **AT hooks** that have the invariant core sequence Arg-Gly-Arg-Pro flanked by positively charged residues and that bind to AT-rich DNA sequences. The NMR spectrum of a truncated form of HMGA1a that contains only its second and third AT hooks (residues 51–90) is indicative of a random coil. However, the NMR structure of this truncated HMGA1a in complex with a 12-bp DNA containing an AT-rich segment of the β -interferon enhancer, determined by Angela Gronenborn and Marius Clore, reveals that each of its AT hooks binds

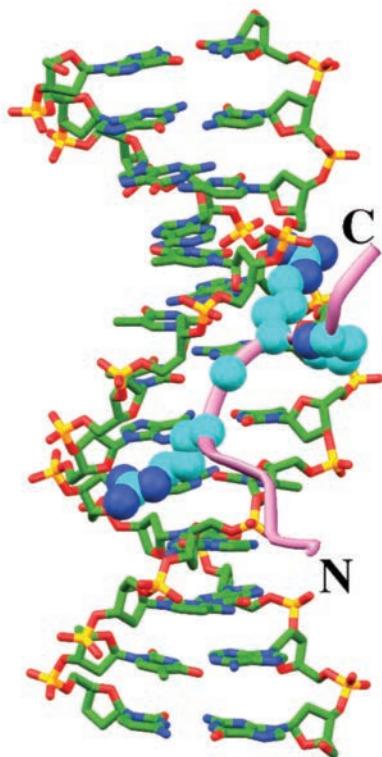


Figure 34-72 NMR structure of a truncated HMGA1a consisting of only its second and third AT hooks in complex with a 12-bp AT-rich DNA. The protein is drawn in ribbon form (*lavender*) with the side chains and C_{α} atoms of the invariant core sequence of its AT hook, Arg-Gly-Arg-Pro, drawn in space-filling form with C cyan and N blue. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P yellow). The protein's two 10-residue AT hooks bind to separate dsDNA dodecamers. Nevertheless, only one set of DNA resonances was observed, which indicates that the two AT hook–DNA structures are closely similar. The peptide segment that links the two AT hooks is not observed and hence must be highly mobile. Consequently, only the structure shown was observed. Note that the AT hook binds in the DNA's minor groove but does not cause it to bend. [Based on an NMR structure by Angela Gronenborn and Marius Clore, NIH, Bethesda, Maryland. PDBid 2EZF.]

in an extended conformation in the minor groove of a separate DNA molecule (Fig. 34-72). Despite the relatively undistorted DNA in this structure, it has been shown that full-length HMGA proteins can bend, straighten, unwind, and induce loop formation in dsDNA. HMGA proteins have been implicated in regulating the transcription of numerous genes. For example, HMGA1 proteins recruit the IFN- β enhanceosome's eight protein subunits to the IFN- β enhancer (Section 34-3Bs). However, the structure of the IFN- β enhanceosome (Fig. 34-68) indicates its various protein subunits prevent HMGA1 from binding to its four binding sites on the IFN- β enhancer. This suggests that, in assembling the IFN- β enhanceosome, HMGA1 functions as a molecular chaperone that bends the DNA in a way that facilitates subunit binding but then dissociates from the DNA as the enhanceosome forms.

3. The HMGN proteins, **HMGN1** and **HMGN2** (98 and 89 residues; previously known as **HMG14** and **HMG17**), occur in mammals but not in *Drosophila* or yeast. They are 60% identical in sequence and consist of three functional motifs: a bipartite nuclear localization signal (NLS; see above), a conserved ~30-residue, positively charged **nucleosome-binding domain (NBD)**, and a **chromatin-unfolding domain (CHUD)**. The ~30-residue, positively charged NBD, as its name implies, targets HMGN proteins to bind to nucleosome core particles as homodimers of HMGN1 or HMGN2 (but not as heterodimers) without preference for the underlying DNA sequence. This stabilizes the nucleosome core particle by bridging its two adjacent dsDNA strands. Nevertheless, HMGN proteins increase the rate of transcription and DNA replication, presumably because they loosen the structure of chromatin fibers. This apparently occurs because the CHUD domain interacts with the N-terminal tail of histone H3 (see below) and because nucleosome-bound HMGN proteins compete with histone H1 for its nucleosomal binding site. HMGN-containing nucleosomes occur as clusters averaging six adjacent nucleosomes, thereby confirming that they alter internucleosomal structure. The presumably decondensed chromatin in these clusters could provide gateways through which regulatory proteins gain access to their target DNAs.

W. RNAP III and RNAP II Use Different Mechanisms to Transcribe through Nucleosomes

Since nucleosomes bind their component DNA tightly and quite stably, how does an actively transcribing RNA polymerase, which is roughly the size of a nucleosome and must separate the strands of duplex DNA to transcribe it, get access to the DNA? Two classes of models have been proposed: The advancing RNA polymerase either (1) induces a conformational change in the nucleosome that permits its DNA to be transcribed while still associated with the nucleosome or (2) displaces the nucleosome from the DNA. These models were differentiated by Gary Felsenfeld as follows: A single nucleosome core was assembled onto a short DNA segment of defined sequence. Then, under conditions in which nucleosome cores are stable (don't decompose or move) in the absence of transcription, the resulting assembly was ligated into a plasmid between a promoter and terminators for the RNA polymerase from **bacteriophage SP6** and the DNA between these two sites was transcribed by this enzyme. This treatment caused the nucleosome to move to a different site on the same plasmid, with a small preference for the untranscribed region preceding the promoter. However, the use of a very short (227-bp) DNA template containing the SP6 promoter and a bound nucleosome revealed that nucleosome transfer occurred only to the same template molecule, 40 to 95 bp upstream of its original site, even in the presence of a large excess of competitor DNA. Evidently, the histone octamer somehow steps around a transcribing RNA polymerase so as to transfer to a nearby segment of the same DNA. Felsenfeld has proposed that this occurs via a DNA looping mechanism in which the histone octamer incrementally

spools onto its new position behind the advancing RNA polymerase as the polymerase peels the octamer away from its original position (Fig. 34-73).

How does RNA polymerase displace nucleosomes from DNA? SP6 RNA polymerase, being a phage enzyme, cannot have evolved to interact with histones but, nevertheless, appears to do so. Other prokaryotic RNA polymerases can likewise transcribe through nucleosomes. A plausible mechanism for this phenomenon is that it is promoted by the transcriptionally induced supercoiling of DNA. A moving transcription bubble, it will be recalled (Section 31-2Ca), generates positive supercoils in the DNA ahead of it and negative supercoils behind it. However, nucleosomal DNA is wound around its histone core in a left-handed toroidal coil and is therefore negatively supercoiled (Section 29-3Aa). Consequently, an advancing RNA polymerase molecule should destabilize the nucleosomes ahead of it while facilitating nucleosome assembly in its wake, precisely what is observed (although, as we shall see below, the cell employs several methods for loosening the grip of nucleosomes on DNA). Subsequent investigations revealed that yeast RNAP III, which is considerably larger than SP6 RNA polymerase, interacts with nucleosomes in a similar manner.

Despite the foregoing, transcription by RNAP II is greatly impeded by chromatin relative to that with naked DNA. This led Danny Reinberg to discover an elongation factor known as **FACT** (*facilitates chromatin transcription*), a heterodimer of two highly conserved subunits: **Spt16** (1047 residues), which binds H2A/H2B dimers, and the HMG-like protein **SSRP1** (structure-specific recognition protein 1; 709 residues), which binds H3/H4 tetramers. FACT facilitates transcription by displacing a single H2A/H2B dimer from a nucleosome during the passage of RNAP II. This apparently loosens the grip of the remaining histones on the DNA so that RNAP II can freely transcribe the resulting hexasome's bound DNA. During this process, the hexasome may change its position on the DNA. Once transcription has been completed, FACT, now functioning as a histone chaperone, facilitates the addition of an H2A/H2B dimer to the hexasome to yield an intact nucleosome. FACT is also implicated in exchanging variant histones with nucleosomes and is essential

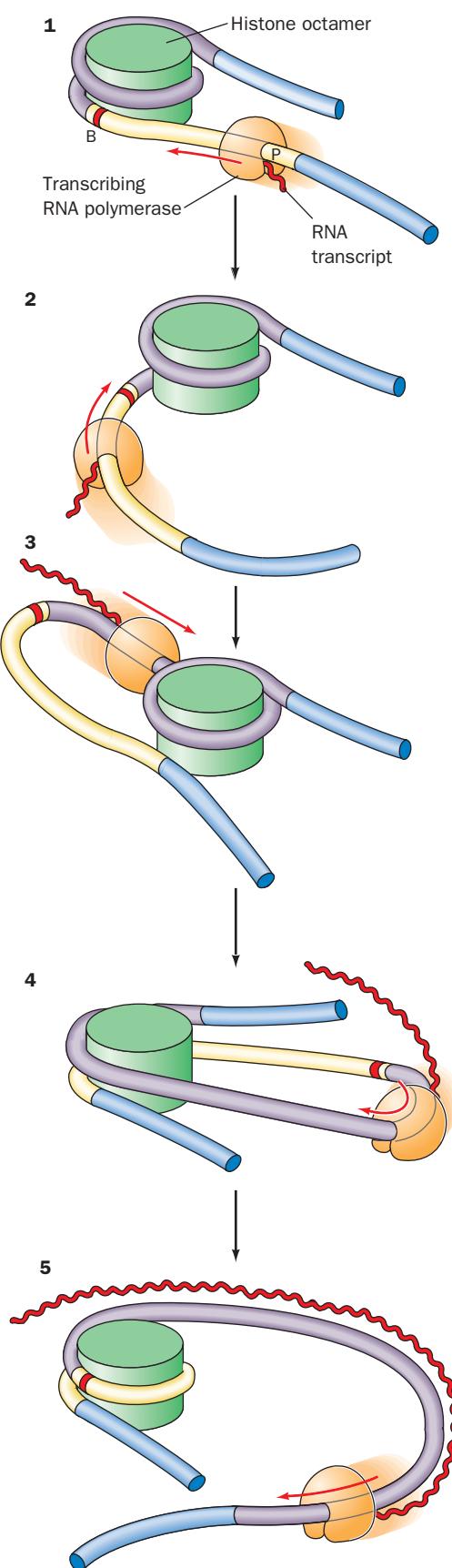


Figure 34-73 Spooling model for transcription through a nucleosome. (1) RNA polymerase commences transcription at a promoter, P; the border of the nucleosome is indicated by B. (2) As the RNA polymerase approaches the nucleosome, it induces the dissociation of the proximal (nearest) DNA, thereby exposing part of the histone octamer surface. (3) The exposed histone surface binds to the DNA behind the RNA polymerase, thus forming a loop. Note that this loop is topologically isolated from the rest of the DNA and, consequently, is subject to the superhelical stress that the advancing RNA polymerase generates (see the text). (4) As the RNA polymerase continues to advance, the DNA ahead of it peels off the histone octamer while the trailing DNA spools onto it. (5) The nucleosome is thereby re-formed behind the RNA polymerase, thus permitting the transcript to be completed. [After Studitsky, V.M., Clark, D.J., and Felsenfeld, G., *Cell* **76**, 379 (1994).]

for the replication of the DNA in chromatin, presumably by similar mechanisms.

x. Locus Control Regions Are Nuclease Hypersensitive

The very light digestion of transcriptionally active chromatin with DNase I and other nucleases has revealed the presence of **DNase I hypersensitive sites** that are about an order of magnitude more susceptible to cleavage by DNase I than are DNase I sensitive sites. These specific 100- to 200-bp DNA segments are mostly located in the 5'-flanking regions of transcriptionally active or activatable genes as well as in sequences involved in replication and recombination. Nuclease hypersensitive sites, as we shall see, are apparently the “open windows” that allow the transcriptional machinery access to DNA control sequences. This is because *DNase I hypersensitive gene segments are free of nucleosomes*. For example, in SV40-infected cells, none of the ~24 nucleosomes that are complexed to the virus’ 5.2-kb circular DNA (Fig. 34-74) incorporate the ~250-bp viral transcription initiation site, thereby rendering that site nuclease hypersensitive. However, since naked DNA is not DNase I hypersensitive, the special properties of nuclease

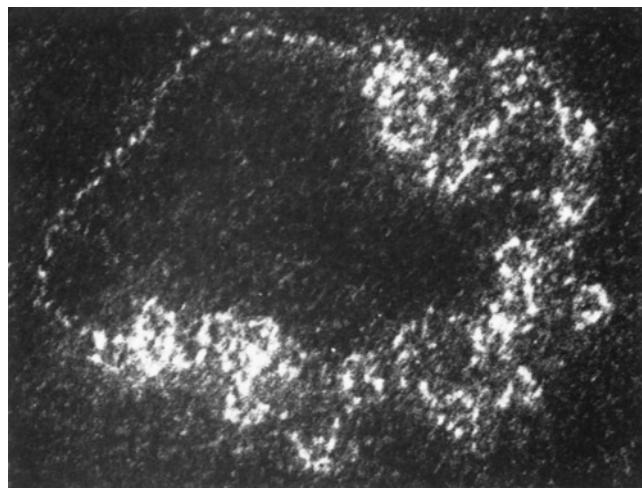


Figure 34-74 Electron micrograph of an SV40 minichromosome that has a nucleosome-free DNA segment. [Courtesy of Moshe Yaniv, Institut Pasteur, Paris, France.]

hypersensitive chromatin must arise from the sequence-specific binding of proteins so as to exclude nucleosomes.

The human β -globin cluster (Section 34-2F) has five nuclease hypersensitive sites in a region 6 to 22 kb on the 5' side of the ϵ gene as well as one hypersensitive site 20 kb on the 3' side of the β gene (Fig. 34-75). These hypersensitive sites appear to demarcate the boundaries of a large segment of transcriptionally active chromatin. Individuals with an extensive upstream deletion that eliminates the 5' hypersensitive sites, the so-called Hispanic deletion, but with normal β -like genes, have **($\gamma\delta\beta$)⁰-thalassemia** (severely reduced synthesis of γ , δ , and β globins). Similarly, mice that are transgenic for the human β -globin gene together with its local regulatory sites either fail to express or express very low levels of human β -globin. This is because a DNA segment that is randomly inserted into a genome will most often occupy a position in transcriptionally inactive heterochromatin, a phenomenon known as a **position effect**. However, mice transgenic for the entire region of the human β -globin cluster between its hypersensitive sites express high levels of human β -globin in erythroid tissues. Thus the β -globin cluster’s 5' nuclease hypersensitive sites, which are collectively known as a **locus control region (LCR)**, function to suppress position effects over large distances (e.g., the ~100 kb length of the β -globin cluster). LCRs have enhancerlike properties but, unlike enhancers, are orientation- and position-specific.

LCRs are apparently activated by proteins that are expressed only in specific cell lineages (e.g., only in erythroid cells for genes controlled by the β -globin LCR) so as to render the gene(s) under an LCR’s control susceptible to activation by transcription factors. In support of this contention, it has been shown that nonglobin genes that have been put under the control of the β -globin cluster LCR are expressed in erythroid cells but not in nonerythroid cells. LCRs occur in a growing list of mammalian genes. However, the way that they permit the expression of the genes under their control remains largely conjectural.

y. Chromatin Immunoprecipitation Reveals the DNA Binding Sites of Proteins

Throughout this section, we discuss the DNA sequences in chromatin to which specific proteins and nucleosomes bind. In many cases, these DNA sequences have been

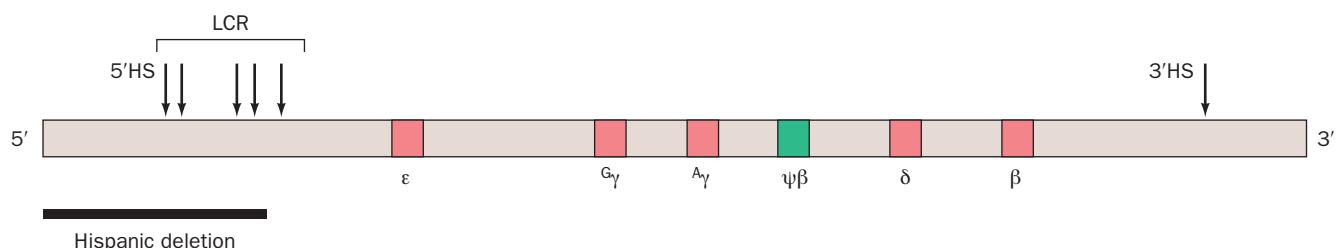
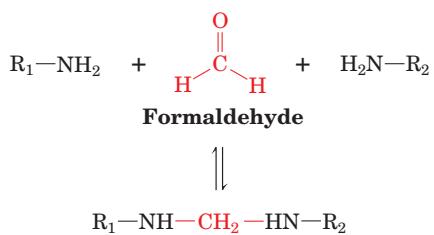


Figure 34-75 The β -globin cluster showing the positions of its genes and its DNase I hypersensitive sites (HSs; arrows). The hypersensitive sites on the 5' side of the ϵ -globin gene (5'HSs) form the locus control region (LCR), whose presence is required

for the expression of the β -like genes. The deletion of the LCR, as occurs in the Hispanic deletion, all but eliminates the expression of the β -like genes. The products of the β -globin cluster are discussed in Section 34-2Fa.

identified using a procedure known as **chromatin immunoprecipitation (ChIP)**. ChIP involves the following steps:

1. Living cells are treated with formaldehyde, which rapidly cross-links the amino and imino groups of Arg, His, and Lys residues to nearby (within ~ 2 Å) amino groups of bases, primarily those of adenine and cytosine, while preserving the chromatin structure:



2. The cells are lysed and the cross-linked chromatin is isolated and sheared into manageable (~ 500 -nt) fragments by sonication.

3. The chromatin fragments are treated with antibodies (Section 35-2B) raised against the protein of interest (which may be a histone with a specific modification such as an acetylation or a methylation at a particular site; see below). The mixture is then adsorbed to agarose gel beads to which *Staphylococcus aureus* **Protein A** has been cross-linked (Section 6-3C). Protein A binds antibodies, but only when they are bound to their target antigens, thereby permitting the isolation of only those chromatin fragments in which the DNA is cross-linked to the antibody-bound protein.

4. The DNA is released from the protein to which it is cross-linked by acidification, which reverses the formaldehyde cross-linking reaction, and the DNA is isolated.

5. The DNA can be identified by PCR (Section 5-5F) using primers corresponding to the sequence(s) of interest. More often, however, the DNA fragments are ligated to generic primers, amplified by PCR, fluorescently labeled, and then hybridized to a DNA microarray (DNA chip; Section 7-6B) that sequentially contains all the sequences in the cell's genome (a so-called **whole-genome tiling array**), thereby revealing the DNA segments to which the protein of interest binds. Such a procedure is therefore known as a **ChIP-chip** assay. Alternatively, the DNA from a genome of known sequence can be directly identified by using massively parallel DNA sequencing methods (Section 7-2C), a technique named **ChIP-seq**.

z. Insulators Isolate Genes from Distant Regulatory Elements

We have seen that enhancers function independently of their position and orientation. But then, what prevents an enhancer from affecting the transcription of all the genes in its chromosome? Conversely, the formation of heterochromatin appears to be self-nucleating. What prevents heterochromatin from spreading into neighboring segments of euchromatin so as to prevent the transcrip-

tion of their component genes? In many cases, this appears to be the job of short (<2 kb) DNA sequences known as **insulators** that thereby define the boundaries of functional domains.

Among the best characterized insulators are the *Drosophila* sequences **scs** and **scs'** (for specialized chromatin structure), which normally flank two consecutive *hsp70* heat shock genes. The transformation into *Drosophila* of the *white* gene (which confers white eye color; Section 1-4Ca) together with a minimal promoter yielded lines of flies that varied in eye color, a manifestation of the position effect. However, when the construct was flanked by **scs** and **scs'**, it yielded only flies with white eyes. Evidently, these insulators overcome the position effect. In addition, if **scs** or **scs'** is inserted between a gene and its upstream regulatory sequences, then the expression of the gene is no longer influenced by these sequences. Several other insulators have been characterized, both in *Drosophila* and in vertebrates.

The foregoing indicates that insulators resemble LCRs in that they suppress position effects. However, unlike LCRs, insulators have no enhancerlike properties; that is, they do not have positive or negative effects on the expression of the genes they control. Rather, *insulators only function to prevent regulatory elements outside the region they control from influencing the expression of the genes inside the region*. LCRs lack this property; they do not protect their associated genes from the influence of control sequences that are upstream of the LCR. In fact, the chicken β -globin cluster has an upstream insulator named **HS4** (for hypersensitive site 4) that prevents regulatory elements that are further upstream from influencing the expression of its genes.

The mechanism of insulator function is poorly understood. Presumably, it is not the insulators themselves but the proteins that bind to them that form the active insulator elements. For example, in *Drosophila*, the insertion of the transposable element *gypsy* between the promoter of the gene *yellow* (which gives flies a pale yellow body rather than the wild-type yellow-brown) and its upstream enhancers prevents these enhancers from activating *yellow* but does not affect downstream enhancers. The 12-zinc finger protein named **Su(Hw)** (for suppressor of hairy wing) specifically binds to the *gypsy* insulator and is required for its enhancer-blocking properties. **Su(Hw)** also binds to the protein **Mod(mdg4)** (for modifier of *mdg4*), and together they bind to the **nuclear matrix** (the nuclear equivalent of the cytoskeleton). In fact, immunostaining studies by Victor Corces indicate that **Su(Hw)** and **Mod(mdg4)** colocalize to several hundred sites in *Drosophila* polytene chromosomes (Fig. 34-76) and are distributed in a punctate pattern around the nuclear matrix. Similar distributions were seen with the protein **BEAF-32** (for boundary element-associated factor of 32 kD), which specifically binds to **scs'** but not to **scs**. These observations suggest that these proteins each bind to numerous insulator sites on *Drosophila* chromosomes, which, in turn, suggests that insulators function as matrix-associated

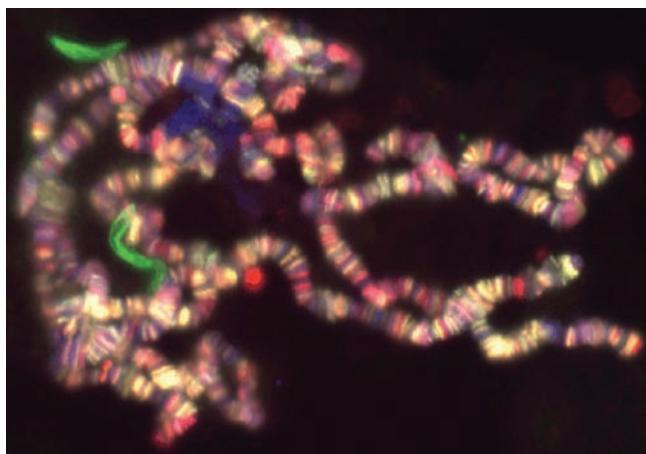


Figure 34-76 Colocalization of the Su(Hw) and Mod(mdg4) proteins on the polytene chromosomes of *Drosophila* larvae. The DNA has been stained blue, whereas the proteins have been immunostained such that sites containing Su(Hw) are green, those containing Mod(mdg4) are red, and sites where both proteins colocalize are yellow. [Courtesy of Victor Corces, The Johns Hopkins University.]

larvae. The DNA has been stained blue, whereas the proteins have been immunostained such that sites containing Su(Hw) are green, those containing Mod(mdg4) are red, and sites where both proteins colocalize are yellow. [Courtesy of Victor Corces, The Johns Hopkins University.]

and both symmetrically and asymmetrically dimethylated. All of these modifications, with the possible exception of certain methylations, are reversible.

The core histones' N-terminal tails are implicated in stabilizing the structures of both core nucleosomes and higher order chromatin (Section 34-1Bb). All of these modifications but methylations reduce (make more negative) the electronic charge of the side chains to which they are appended and hence are likely to weaken histone–DNA interactions so as to promote chromatin decondensation, although as we shall see, this is not always the case. Methyl groups, in contrast, increase the basicity and hydrophobicity of the side chains to which they are linked and hence tend to stabilize chromatin structure. Modified histone tails also interact with specific chromatin-associated nonhistone proteins in a way that changes the transcriptional accessibility of their associated genes.

Over 70 different sites for histone modification have been reported. The characterization of many of these sites led David Allis to hypothesize that *there is a “histone code” in which specific modifications evoke certain chromatin-based functions and that these marks (modifications) act sequentially or synergistically to generate unique biological*

regions (MARs; Section 34-1Da) to form structural domains. This may prevent enhancers that are outside such a domain from influencing the expression of the genes that are inside the domain and, furthermore, may inhibit heterochromatin from encroaching on and thereby transcriptionally inactivating the domain. In vertebrates, insulator function requires the binding of the highly conserved 11-zinc finger protein **CTCF** (for CCCTC-binding factor). ChIP-chip assays of the human genome have revealed that it contains nearly 14,000 CTCF-binding sites, most of which occur at sites remote from transcription start sites and which are largely invariant across different cell types. An average of 2.5 genes are bounded by pairs of CTCF binding sites.

aa. Histone Modification and Remodeling Play an Essential Role in Transcriptional Activation

The DNA packaged by chromatin must be accessible by the transcriptional machinery in order for it to be expressed. This, as we shall see, occurs via two types of processes acting in synergy: (1) the post-translational modifications of core histones, mainly their N-terminal tails, and (2) the remodeling of chromatin through the ATP-driven alteration of the position and/or properties of its nucleosomes.

The post-translational modifications to which core histones are subject include the acetylation of specific Lys side chains, the methylation of specific Lys and Arg side chains, the phosphorylation of specific Ser, Thr, and Tyr side chains, the ADP-ribosylation of specific Glu side chains, and the ubiquitination and sumoylation of specific Lys side chains (Fig. 34-77). Moreover, Lys side chains can be mono-, di-, and trimethylated and Arg side chains can be mono-

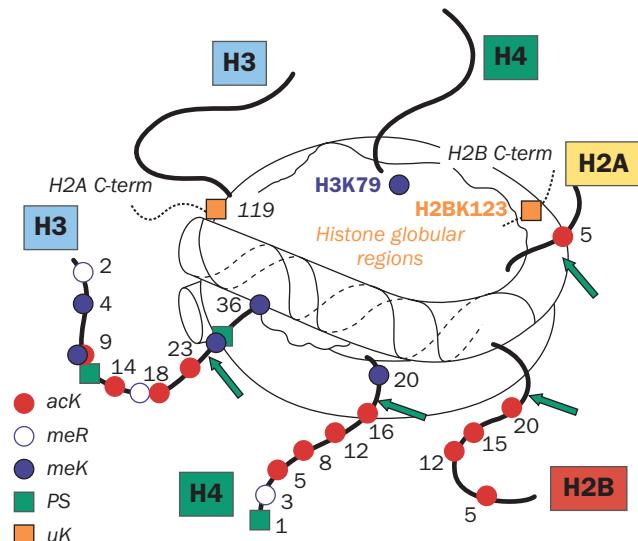


Figure 34-77 Histone modifications on the nucleosome core particle. Post-translational modification sites are indicated by the residue numbers and the colored symbols, which are defined in the key at the lower left (acK = acetyl-Lys, meR = methyl-Arg, meK = methyl-Lys, PS = phospho-Ser, and uK = ubiquitinated Lys). Note that H3 Lys 9 can be either methylated or acetylated. The N-terminal tail modifications are shown on only one of the two copies of H3 and H4 and only one molecule each of H2A and H2B are shown. The C-terminal tails of one H2A and one H2B are represented by dotted lines. The green arrows indicate the sites in intact nucleosomes that are susceptible to trypsin cleavage. This cartoon summarizes data from several organisms, some of which may lack particular modifications. [Courtesy of Bryan Turner, University of Birmingham School of Medicine, U.K.]

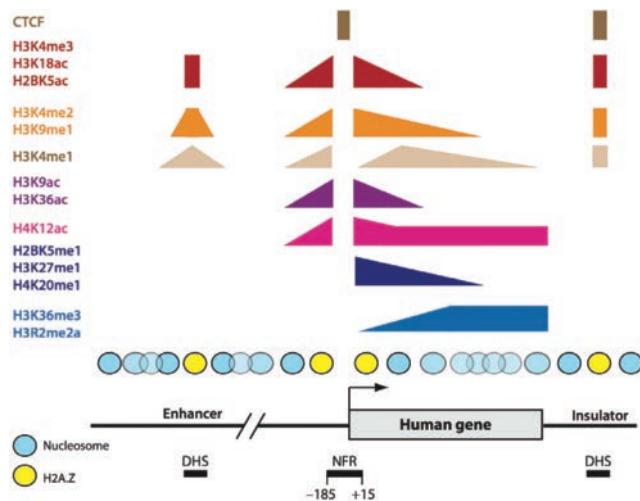


Figure 34-78 Chromatin map of a typical transcriptionally active human gene. The height of each colored strip at a given position is indicative of the prevalence of the corresponding histone mark(s) on the nucleosome at that position. For example, H3K4me3 (trimethylated H4 Lys 4) and H2BK5ac (acetylated H2B Lys 5) are associated with the nucleosomes flanking the transcriptional start site (squared off arrow) as well as portions of enhancers and insulators, whereas H3R2me2a (asymmetrically dimethylated H3 Arg 2) is associated with nucleosomes occupying the downstream region of the gene's protein encoding segment. The positions of DNase I hypersensitive sites (DHS), nucleosome free regions (NFR), CTCF binding sites (Section 34-3Bz), and nucleosomes containing the H2A variant H2A.Z are also indicated. [Courtesy of Oliver Rando, University of Massachusetts Medical School, and Howard Chang, Stanford University.]

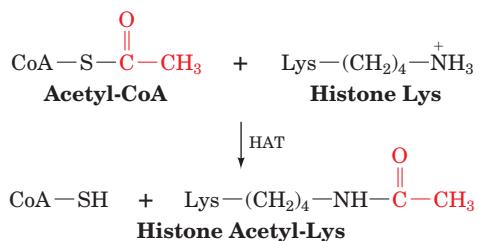
outcomes. For example, Fig. 34-78 indicates the distribution of the various histone marks on nucleosomes associated with a typical transcriptionally active human gene. It can be seen from Figs. 34-77 and 34-78 that there are a vast number of possible combinations of histone marks, although the way they interact is largely unknown.

The growth of a multicellular organism requires the proliferation of the cells in its various tissues without changing their identities (the process whereby cells progressively and irreversibly change their identities, which is known as differentiation, is discussed in Section 34-4). A particular cell type is largely defined by its characteristic pattern of gene expression. Since most cells in a multicellular organism have the same complement of DNA, how do cells maintain their identities (patterns of gene expression) from one cell generation to the next? Evidently, histone modifications are largely preserved between cell generations, that is, they are epigenetic markings in much the same way as are the methylation patterns of DNA (Section 30-7). The way in which a cell confers its histone epigenetic markings on its progeny is poorly understood, although it almost certainly involves the recruitment of histone-modifying enzymes to newly assembled nucleosomes on recently replicated DNA.

bb. Histone Acetyltransferases (HATs) Are Components of Multisubunit Transcriptional Coactivators

Canonical histones are synthesized only in S phase of the cell cycle and then rapidly acetylated on specific Lys residues. This, quite possibly, recruits chaperones (Section 34-1Be), which thereupon mediate the deposition of these histones onto DNA to form nucleosomes. The histones are then deacetylated and eventually reacetylated with new patterns that enable the nucleosomes' various functions.

Histone Lys side chains are acetylated in a sequence-specific manner by enzymes known as **histone acetyltransferases (HATs)**, all of which employ acetyl-CoA (Fig. 21-2) as their acetyl group donors:



The large number of known HATs are members of five families: (1) the **GNAT family** (for *Gcn5*-related *N-acetyltransferase*; **Gen5**, first found in yeast, is one of the best characterized HATs), whose members include *Gcn5*, its homologs **Gen5L** (for *Gcn5*-like protein) and **PCAF** [for *p300/CBP*-associated factor; **p300** and **CBP** (for *cAMP* response binding element protein) are homologous transcriptional coactivators], and **Hat1** (which acetylates histones in the cytoplasm before they are imported to the nucleus); (2) the **MYST family** (named for its founding members, **MOZ**, **Ybf2/Sas3**, **Sas2**, and **Tip60**); (3) the **p300/CBP family**; (4) the **TAF1 family** (TAF1, the largest subunit of TFIID, was formerly named **TAF_{II}250**); and (5) the **SRC family** (for steroid receptor coactivator). Each HAT acetylates one or a few specific histone Lys residues. Most HATs besides Hat1 function as transcriptional coactivators or silencers but some are implicated in regulating chromatin condensation, cell cycle progression, DNA replication, DNA repair, and transcriptional elongation.

Most if not all HATs function *in vivo* as members of often large (10–20 subunits) multisubunit complexes, many of which were initially characterized as transcriptional regulators. These include **SAGA** (for *Spt/Ada/Gcn5/acetyltransferase*), the closely similar **PCAF complex** (which contains PCAF), **STAGA** (*Spt3/TAF/Gcn5L acetyltransferase*; its HAT is *Gcn5L*), **ADA** (transcriptional *adaptor*), TFIID (which contains TAF1), **TFTC** (*TBP-free TAF-containing complex*), **SAS** (something about silencing), **NuA3**, and **NuA4** (nucleosomal acetyltransferases of H3 and H4).

Many **HAT complexes** share subunits. For example, three of ADA's four subunits, *Gcn5*, **Ada2**, and **Ada3**, are common to the 14-subunit SAGA. Likewise, SAGA and NuA4 both contain **Tra1**, a homolog of the phosphoinositide 3-kinases (Section 19-4D) that interacts with specific transcriptional activators, including **Myc** (Section 34-4Ca).

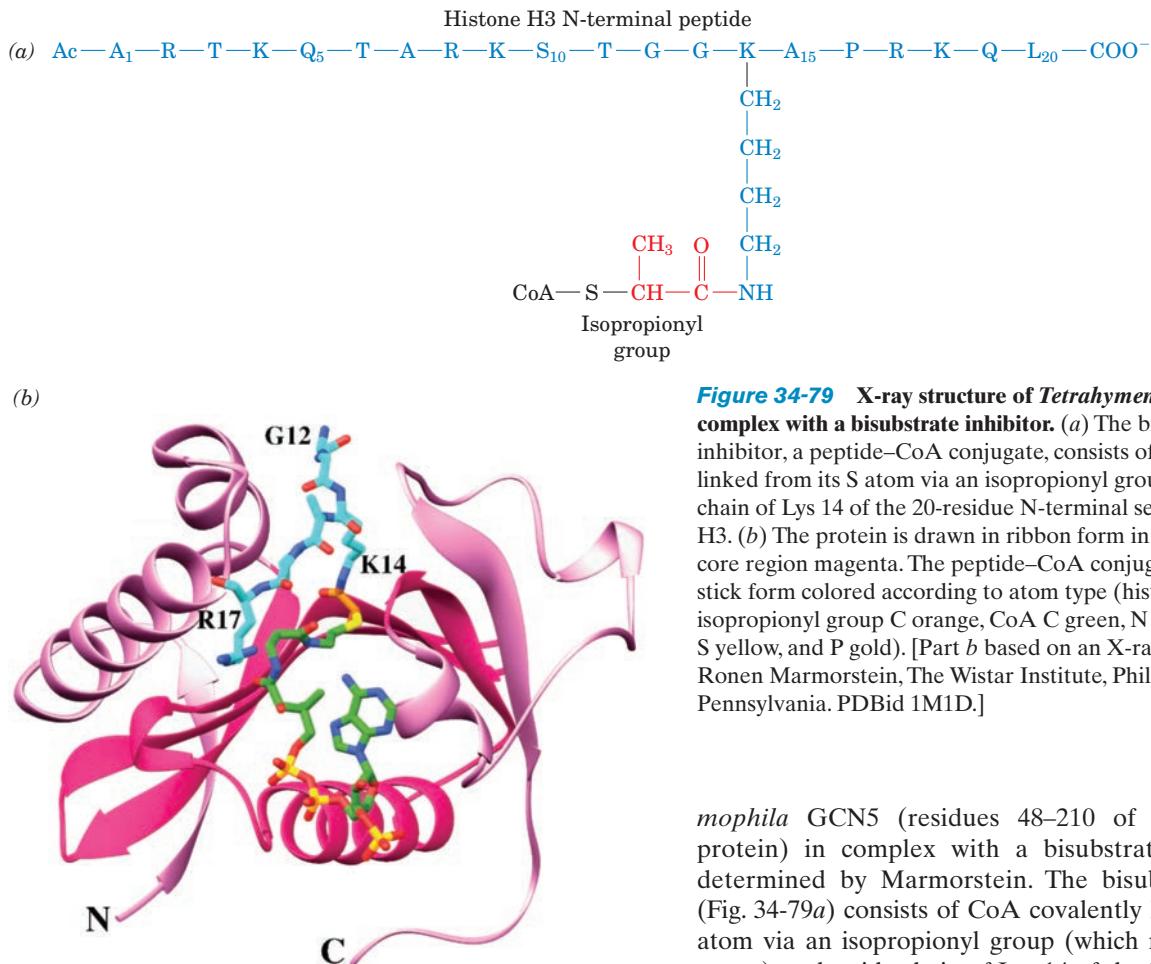


Figure 34-79 X-ray structure of *Tetrahymena* GCN5 in complex with a bisubstrate inhibitor. (a) The bisubstrate inhibitor, a peptide–CoA conjugate, consists of CoA covalently linked from its S atom via an isopropionyl group to the side chain of Lys 14 of the 20-residue N-terminal segment of histone H3. (b) The protein is drawn in ribbon form in lavender with its core region magenta. The peptide–CoA conjugate is drawn in stick form colored according to atom type (histone C blue, isopropionyl group C orange, CoA C green, N blue, O red, S yellow, and P gold). [Part b based on an X-ray structure by Ronen Marmorstein, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 1M1D.]

Intriguingly, several HAT complexes besides TFIID and TFTC contain TAFs. For example, SAGA contains **TAF5**, **TAF6**, **TAF9**, **TAF10**, and **TAF12**, as does the PCAF complex with the exception that TAF5 and TAF6 are replaced in the PCAF complex by their close homologs **PAF65 β** and **PAF65 α** (PAF for PCAF associated factor). TAF6, TAF9, and TAF12, as we discussed in Section 34-3Be, are structural homologs of histones H3, H4, and H2B, respectively. Consequently, these TAFs probably associate to form an architectural element that is common to TFIID, SAGA, and the PCAF complex and hence these complexes are likely to interact with TBP in a similar manner.

The various HAT complexes presumably target their component HATs to the promoters of active genes. Moreover, they alter the specificities of these HATs. For example, a general property of HATs is that although they can acetylate at least one type of free histone, they can only acetylate histones in nucleosomes as members of HAT complexes. Thus, free Gcn5 acetylates H3 Lys 14 and, to a lesser extent, H4 Lys 8 and 16. However, Gcn5 in SAGA expands its H3 sites to Lys 9, 14, and 18 and also acetylates H2B, whereas Gcn5 in ADA acetylates H3 at Lys 14 and 18, as well as H2B. Neither complex acetylates H4.

The X-ray structures of several HATs have been determined. That of the HAT domain of *Tetrahymena ther-*

mophila GCN5 (residues 48–210 of the 418-residue protein) in complex with a bisubstrate inhibitor was determined by Marmorstein. The bisubstrate inhibitor (Fig. 34-79a) consists of CoA covalently linked from its S atom via an isopropionyl group (which mimics an acetyl group) to the side chain of Lys 14 of the 20-residue N-terminal segment of histone H3. The structure (Fig. 34-79b) reveals the enzyme to be deeply clefted and to contain a core region common to all HATs of known structure (magenta in Fig. 34-79b) that consists of a 3-stranded antiparallel β sheet connected via an α helix to a fourth β strand that forms a parallel interaction with the β sheet. Only 6 residues of the histone tail, Gly 12 through Arg 17, are visible in the X-ray structure. The CoA moiety binds in the enzyme's cleft such that it is mainly contacted by core residues. The comparison of this structure with other Gcn5-containing structures indicates that the cleft has closed down about the CoA moiety.

cc. Bromodomains Recruit Coactivators to Acetylated Lys Residues in Histone Tails

The different patterns of histone acetylation required for different functions (the histone code) suggest that the function of histone acetylation is more complex than merely attenuating the charge–charge interactions between the cationic histone N-terminal tails and anionic DNA. In fact, there is clear evidence that specific acetylation patterns are recognized by protein modules of transcriptional coactivators in much the same way that specific phosphorylated sequences are recognized by protein modules such as the SH2 and PTB domains that mediate signal transduction via protein kinase cascades (Section 19-3). Thus, nearly all HAT-associated transcriptional coactivators

contain \sim 110-residue modules known as **bromodomains** that specifically bind acetylated Lys residues on histones. For example, Gcn5 essentially consists of its HAT domain followed by a bromodomain, whereas TAF1 consists mainly of an N-terminal kinase domain followed by a HAT domain and two tandem bromodomains.

The X-ray structure of human TAF1's double bromodomain (residues 1359–1638 of the 1872-residue protein), determined by Tjian, reveals that it consists of two nearly identical antiparallel 4-helix bundles (Fig. 34-80). A variety of evidence, including NMR structures of single bromodomains in complex with their target acetyl-Lys-containing peptides, indicates that the acetyl-Lys binding site of each bromodomain occurs in a deep hydrophobic pocket that is located at the end of its 4-helix bundle opposite its N- and C-termini. The double bromodomain's two binding pockets are separated by \sim 25 Å, which makes them ideally positioned to bind two acetyl-Lys residues that are separated by 7 or 8 residues. In fact, the N-terminal tail of histone H4 contains Lys residues at its positions 5, 8, 12, and 16 (Fig. 34-77), whose acetylation is correlated with increased transcriptional activity. Moreover, the 36-residue N-terminal peptide of histone H4, when fully acetylated, binds to the TAF1 double bromodomain in 1:1 ratio with 70-fold higher affinity than do single bromodomains but fails to bind when it is unacetylated. This is clear evidence that at least some histone marks act synergistically.

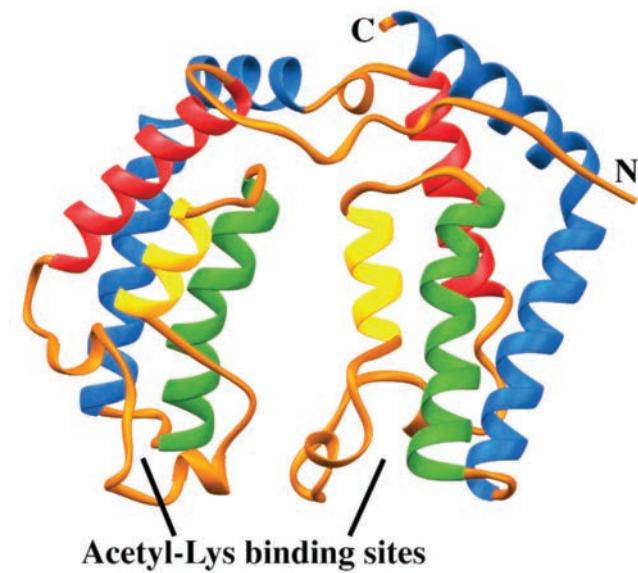


Figure 34-80 X-ray structure of the human TAF1 double bromodomain. Each bromodomain consists of an antiparallel 4-helix bundle whose helices are colored, from N- to C-termini, red, yellow, green, and blue, with the remaining portions of the protein orange. The two 4-helix bundles are related by an \sim 108° rotation about an axis that is approximately parallel to the principal axes of the 4-helix bundles (the vertical direction in this drawing). The acetyl-Lys binding sites occupy deep hydrophobic pockets at the end of each 4-helix bundle opposite its N- and C-termini. [Based on an X-ray structure by Robert Tjian, University of California at Berkeley. PDBid 1EQF.]

The foregoing structure suggests that the TAF1 bromodomains serve to target TFIID to promoters that are within or near nucleosomes (in contrast to the widely held notion that TFIID targets PICs to nucleosome-free regions). Tjian has therefore postulated that the transcriptional initiation process may begin with the recruitment of a HAT-containing coactivator complex by an upstream DNA-binding protein (Fig. 34-81). The HAT could then acetylate the N-terminal histone tails of nearby nucleosomes, which would recruit TFIID to an appropriately located promoter via the binding of its TAF1 bromodomains to the acetyl-Lys residues. Moreover, the TAF1 HAT activity could acetylate other nearby nucleosomes, thereby initiating a cascade of acetylation events that would render the DNA template competent for transcriptional initiation.

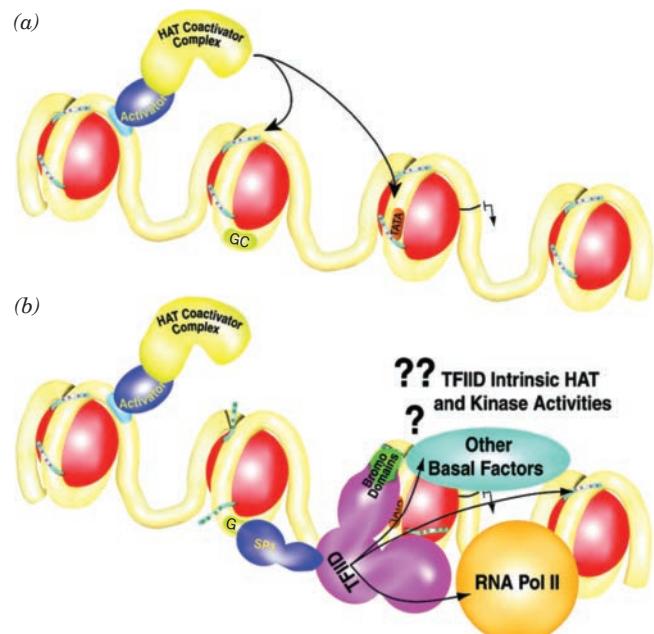
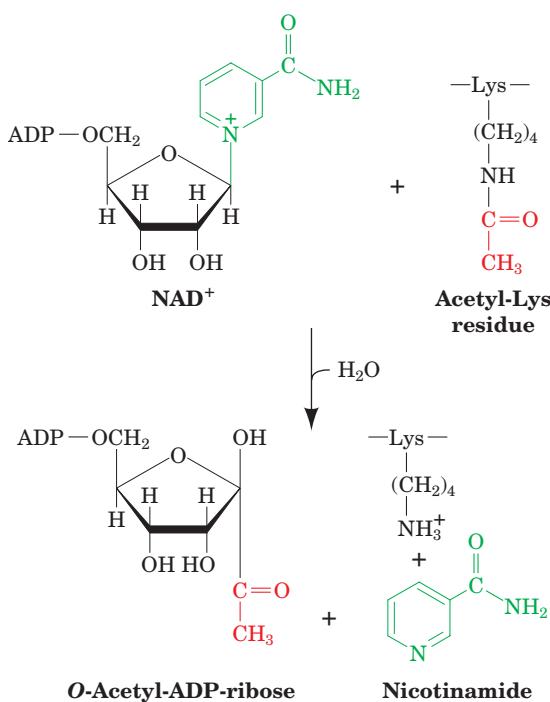


Figure 34-81 Simplified model for the assembly of a transcriptional initiation complex on chromatin-bound templates. Here the DNA is represented by a yellow worm, the histone octamers around which the DNA wraps to form nucleosomes are shown as red spheres, and their N-terminal histone tails are drawn as short cyan rods with the red and green dots representing unacetylated and acetylated Lys residues. The transcription initiation site is represented by the black ring about the DNA from which the squared-off arrow points downstream. (a) The process begins by the recruitment of a HAT-containing transcriptional coactivator complex (yellow) through its interactions with a DNA-binding activator protein (purple) that is bound to an upstream enhancer (light blue). The coactivator HAT is thereby positioned to acetylate the N-terminal tail on nearby nucleosomes (curved arrows). (b) The binding of TAF1's bromodomains to the acetylated histone tails could then help recruit TFIID (magenta) to a nearby TATA box (orange patch). Further acetylation of nearby histone tails by TAF1's HAT domain could help recruit other basal factors (cyan) and RNAP II (orange) to the promoter, thus stimulating PIC formation. Note that this model does not preclude other activation pathways such as the binding of enhancer-bound SP1 (purple) to TFIID. [Courtesy of Robert Tjian, University of California at Berkeley.]

dd. Histone Deacetylases (HDACs)

Histone acetylation is a reversible process. The enzymes that remove the acetyl groups from histones, the **histone deacetylases (HDACs)**, promote transcriptional repression and gene silencing. Eukaryotic cells from yeast to humans typically contain numerous different HDACs; 10 HDACs have been identified in yeast and 17 in humans. The HDACs consist of four protein families: Class I, which in humans contains **HDAC1, 2, 3, and 8**; Class II, which in humans contains **HDAC4-7, 9, and 10**; and Class III, which in humans contains the so-called **sirtuins, SIRT1-7** (SIRT for silent information regulator two; yeast **SIR2** was the first known sirtuin); and Class IV (**HDAC11**). Most, if not all, of the Class I HDACs are members of several multisubunit complexes. Thus, HDAC1 and HDAC2 form the catalytic cores of three complexes, **Sin3**, **NuRD** (nucleosome remodeling histone deacetylase), and **CoREST** (corepressor to *RE1* silencing transcription factor), whereas HDAC3 is the catalytic core of **N-CoR** (nuclear hormone receptor corepressor) and **SMRT** (silencing mediator of retinoid and thyroid hormone receptor). These complexes serve as **transcriptional corepressors** for numerous transcriptional repressors as well as cooperating with each other. For example, the repressor **REST** (neuron-restrictive repressor), on binding to its target DNA site, recruits CoREST and Sin3, which together repress transcription from nearby nucleosomes. Many if not all of the Class II HDACs function as transcriptional corepressors although few of them appear to be members of multisubunit complexes.

The Class III HDACs, the sirtuins, are unusual in that they contain an essential NAD^+ cofactor. Rather than simply hydrolyzing the amide bond linking the acetyl group to their target Lys side chain, they transfer it to the ADP-ribosyl group of NAD^+ , thereby yielding **O-acetyl-ADP-ribose**, nicotinamide, and the deacetylated Lys residue:



Sirtuins can also deacetylate nonhistone proteins. For ex-

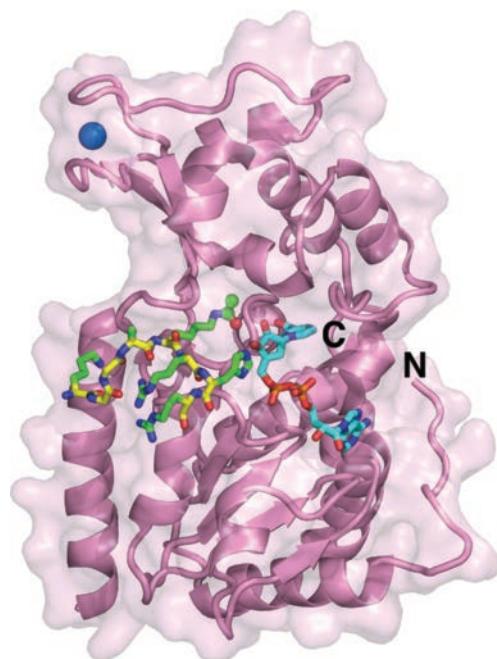
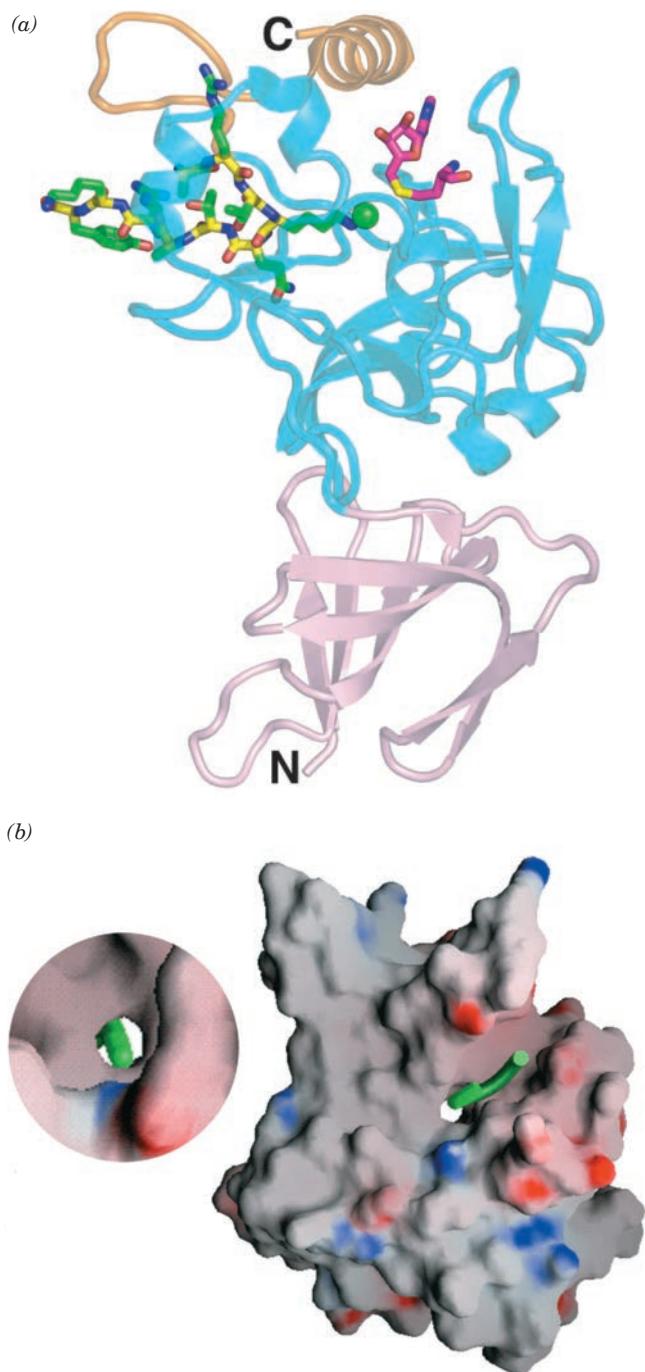


Figure 34-82 X-ray structure of the ternary complex of yeast Hst2 with carba-NAD⁺ and an acetyl-Lys-containing decapeptide from the histone H4 N-terminal tail. The Hst2 is drawn in ribbon form embedded in its semitransparent surface (pink). The H4 decapeptide (KGGA-acK-RHRKI; residues 12-21) and the carba-NAD⁺ (NAD⁺ with the ribose ring O of its NMN⁺ residue replaced by CH₂) are drawn in stick form with the H4 peptide backbone C yellow, its side chain C green, carba-NAD⁺ C cyan, N blue, O red, and P orange. The acetyl group of the acetyl-Lys residue is shown in ball-and-stick form. A bound Zn²⁺ ion is represented by a blue sphere. [Based on an X-ray structure by Ronen Marmorstein, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 1SZC.]

ample, SIRT1 targets numerous transcriptional regulators that participate in controlling metabolism, senescence, cancer, and perhaps aging. In fact, the increased life span that caloric restriction confers on many animals is associated with increased sirtuin expression.

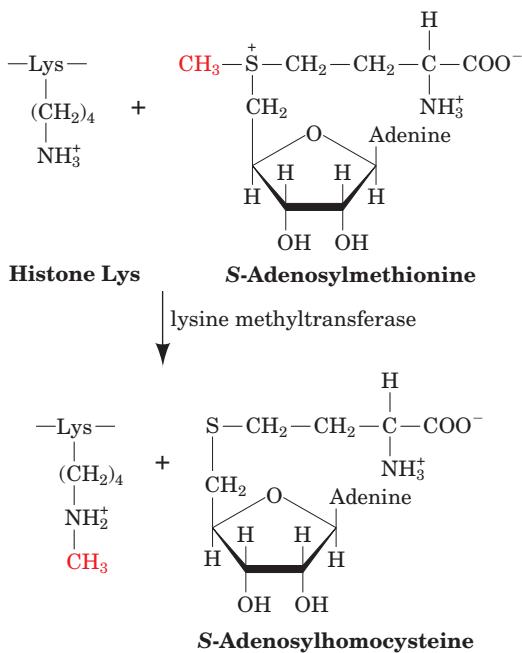
Human SIRT2, a component of transcriptionally silent chromatin, is required for gene silencing. The X-ray structure of its yeast homolog **Hst2** (homolog of SIR two 2) in complex with the nonhydrolyzable NAD⁺ analog **carba-NAD⁺** and a decapeptide segment of histone H4's N-terminal tail containing an acetyl-Lys residue (Fig. 34-82), was determined by Marmorstein. The monomeric protein consists of an elongated core that contains a conserved Rossmann fold domain and a less conserved Zn²⁺-binding domain that arises from two insertions in the Rossmann fold domain. The loops that connect these domains form a hydrophobic cleft lined with the protein's most highly conserved residues in which both the H4 peptide and the carba-NAD⁺ are bound. The Zn²⁺ ion, which is tetrahedrally liganded by four conserved Cys side chains, is too far from the enzyme's active site to have a catalytic function. Its presence is nevertheless required for catalytic activity, which suggests that it has an essential structural role.

Figure 34-83 X-ray structure of the human histone methyltransferase SET7/9 in ternary complex with its products *S*-adenosylhomocysteine and the histone H3 N-terminal decapeptide with its Lys 4 monomethylated. (a) The protein is drawn in semitransparent ribbon form in which the N-terminal domain is pink, the SET domain is cyan, and the C-terminal segment is orange. The H3 decapeptide (ART-meK-QTARKY) and *S*-adenosylhomocysteine are drawn in stick form with the H3 peptide backbone C yellow, its side chain C green, *S*-adenosylhomocysteine C magenta, N blue, O red, and S yellow. The methyl group on Lys 4 of the H3 peptide is represented by a green sphere. (b) Surface representation of the protein as seen from the same side of the protein as in Part a. The protein surface is colored according to charge (blue most positive, red most negative, and gray neutral) and the H3 decapeptide is represented by a green ribbon. Note the narrow tunnel through the protein in which the methyl-Lys side chain is inserted. The inset at the left shows a close-up of this Lys access channel containing the methyl-Lys side chain (green) as viewed from the *S*-adenosylhomocysteine binding site. [Courtesy of Steven Gamblin, National Institute for Medical Research, London, U.K. PDBid 1O9S.]



ee. Histone Methylation/Demethylation

Histone methylation at both the Lys and Arg side chains of histone H3 and H4 N-terminal tails (Fig. 34-77) tends to silence the associated genes by inducing the formation of heterochromatin. The enzymes mediating these methylations, the **histone methyltransferases (HMTs)**, all utilize *S*-adenosylmethionine (SAM; Section 26-3Ea) as their methyl donor. Thus, the lysine HMTs, the most extensively characterized HMTs, catalyze the reaction



These enzymes all have a so-called **SET domain** [Su(var)3-9, *E*(Z), Trithorax], which contains their catalytic sites.

The human lysine HMT named **SET7/9** monomethylates Lys 4 of histone H3. The X-ray structure of the SET domain of SET7/9 (residues 108–366 of the 366-residue

protein) in complex with *S*-adenosylhomocysteine (demethylated SAM; Section 26-3Ea) and the N-terminal decapeptide of histone H3 in which Lys 4 is monomethylated, was determined by Steven Gamblin. Interestingly, *S*-adenosylhomocysteine and the peptide substrate bind to opposite sides of the protein (Fig. 34-83a). However, there is a narrow tunnel through the protein into which the Lys 4 side chain is inserted such that its amine group is properly positioned for methylation by SAM (Fig. 34-83b). The arrangement of the hydrogen bonding acceptors for the Lys amine group stabilizes the methyl-Lys side chain in its observed orientation about the $\text{C}_\epsilon\text{—N}_\epsilon$ bond, thus sterically precluding the methyl-Lys group from assuming a

conformation in which it could be further methylated by SAM. Indeed, mutation of SET7/9's Tyr 245 or Tyr 305 to Phe converts it to a di- or trimethyltransferase.

Methylated histones are recognized by **chromodomains**. For example, methylated H3 Lys 9 is bound by the chromodomain-containing **heterochromatin protein 1 (HP1)**, which thereupon recruits proteins that control chromatin structure and gene expression. The NMR structure of the mouse HP1 chromodomain (residues 8–80 of the 185-residue protein) in complex with the N-terminal 18-residue tail of H3 in which Lys 4 and 9 are dimethylated (Fig. 34-84), determined by Natalia Murzina and Ernest Laue, reveals that HP1 binds the H3 tail in an extended β -strandlike conformation in a groove on its surface (Fig. 34-84). The chromodomain buries the side chain of H3 Lys 9 (but not that of H3 Lys 4) such that its two methyl groups are contained in a hydrophobic box formed by three conserved aromatic residues. In contrast, the unmethylated H3 tail does not bind to HP1.

As previously mentioned, heterochromatin has a tendency to spread, thus silencing the newly heterochromatized genes. One way in which this appears to occur is via the binding of HP1 to nucleosomes whose H3 Lys 9 residues have been methylated (which is associated with transcriptionally inactive chromatin). The bound HP1 recruits the HMT **Suv39h**, which methylates nearby nucleosomes at their H3 Lys 9 residues, which thereupon recruit additional HP1, etc. Such heterochromatin spreading, as we previously discussed, is prevented by the presence of an insulator (Section 34-3Bz). The HS4 insulator in the chicken β -globin cluster

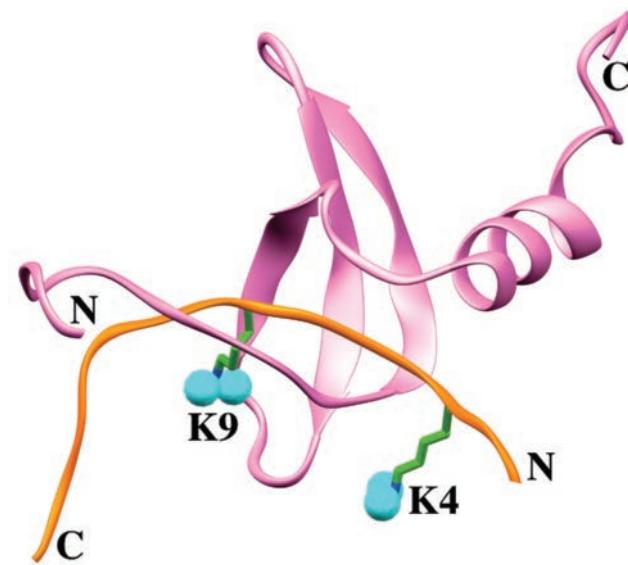


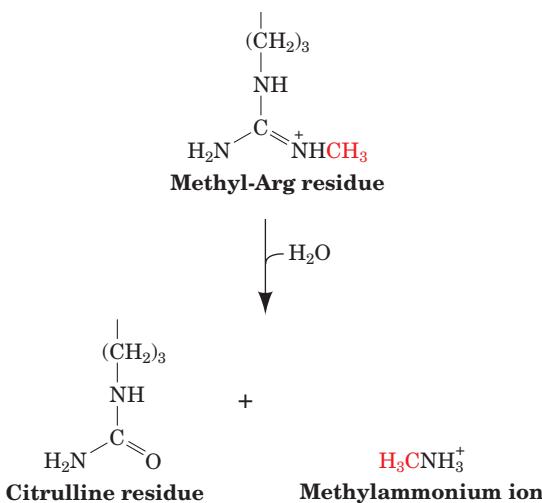
Figure 34-84 NMR structure of the mouse HP1 chromodomain in complex with the 18-residue N-terminal tail of histone H3 in which Lys 4 and Lys 9 are dimethylated. The 80-residue chromodomain is lavender, the H3 N-terminal tail is orange, and its two dimethyl-Lys side chains are drawn in stick form with C green, N blue, and with their methyl groups represented by cyan spheres. The side chain of H3 Lys 9, but not that of H3 Lys 4, is buried by the chromodomain. [Based on an NMR structure by Natalia Murzina and Ernest Laue, University of Cambridge, U.K. PDBid 1GUW.]

recruits HATs that acetylate H3 Lys 9 on nearby nucleosomes (which is associated with transcriptional activity), thereby blocking their methylation. Note that this activity is distinct from the enhancer-blocking function of HS4, a process that is mediated by the binding of CTCF to a different subsite of HS4 than that to which HATs bind.

For many years it was thought that histone methylation was irreversible. However, two types of demethylases have recently been identified:

1. Histone lysine demethylases, which catalyze oxidative reactions that remove methyl groups as formaldehyde. Two families of proteins are known to do so. Those of the **LSD1** (for *lysine-specific demethylase 1*) family, which are FAD-dependent amine oxygenases, require a protonatable methylammonium group and hence can only demethylate mono- and dimethyl-Lys. Those of the **JHDM** (for *jumonji histone demethylase*) family, which are Fe(II)-dependent dioxygenases, can also demethylate trimethyl-Lys.

2. Histone arginine demethylases, which catalyze a hydrolytic reaction that yields a citrulline residue and methylammonium ion.



Note that this reaction does not directly reverse the methylation reaction. The citrulline residue may be a histone mark itself and/or it may be reconverted to Arg by a currently unknown aminotransferase or the histone may be replaced. The way in which dimethyl-Arg residues are demethylated, if at all, is unknown.

ff. Histone Ubiquitination Functions to Regulate Transcription

Although ubiquitination mainly functions to mark proteins for destruction by the proteasome (Section 32-6Bb), it is also implicated in the control of transcription. In yeast, for example, the monoubiquitination of H2B Lys 123 (in contrast to polyubiquitination, which marks proteins for destruction), which is mediated by the ubiquitin-conjugating enzyme (E2) **Rad6** and the RING-finger-containing ubiquitin-protein ligase (E3) **Bre1**, is a prerequisite for the methylation of H3 Lys 4 and Lys 79. Together, these modifications are implicated in the silencing of genes located near telomeres. It has therefore been suggested that H2B

ubiquitination functions as a master switch that controls the site-selective histone methylation patterns responsible for telomeric gene silencing. Similarly, the TAF1 subunit of TFIID, functioning as both a ubiquitin-activating enzyme (E1) and an E2, monoubiquitinates H1, a post-translational modification that is required for the expression of genes in the correct order during *Drosophila* development. Conversely, a histone deubiquitinating enzyme (DUB) is associated with the SAGA chromatin modifying complex. Thus, although the role of histone ubiquitination is only beginning to come to light, it is clear that it is an essential transcriptional regulator.

gg. Chromatin-Remodeling Complexes

Sequence-specific DNA-binding proteins must gain access to their target DNAs before they can bind to them. Yet nearly all DNA in eukaryotes is sequestered by nucleosomes if not by higher order chromatin. How then do the proteins that bind to DNA segments gain access to their target DNAs? The answer, which has only become apparent since the mid-1990s, is that *chromatin contains ATP-driven complexes that remodel nucleosomes*, that is, they somehow disrupt the interactions between histones and DNA in nucleosomes to make the DNA more accessible. This may cause the histone octamer to slide along the DNA strand to a new location (a *cis* transfer) or relocate to a different DNA strand (a *trans* transfer). Thus, *these chromatin-remodeling complexes impose a “fluid” state on chromatin that maintains its DNA’s overall packaging but transiently exposes individual sequences to interacting factors*.

Chromatin-remodeling complexes consist of multiple subunits. The first of them to be characterized was the yeast **SWI/SNF** complex, so called because it was discovered through genetic screens as being essential for the expression of the *HO* gene, which is required for mating type switching (SWI for switching defective), and for the expression of the *SUC2* gene, which is required for growth on sucrose (SNF for sucrose nonfermenter). SWI/SNF, an 1150-kD complex of 11 different types of subunits, is only essential for the expression of ~3% of yeast genes and is not required for cell viability. However, a related complex named **RSC** (for *re*models the *struc*ture of *chrom*atin) is ~100 times more abundant in yeast and is required for cell viability. RSC

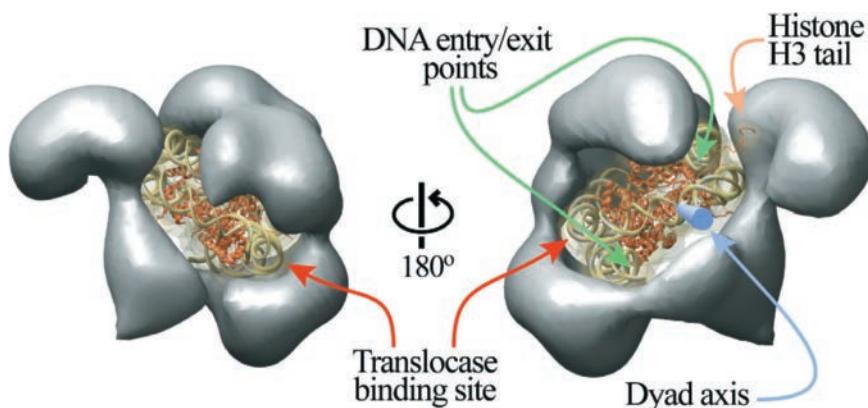
shares two subunits with SWI/SNF and many of their remaining subunits are homologs, including their ATPase subunits, which are named **Swi2/Snf2** in SWI/SNF and **Sth1** in RSC. The Swi2/Snf2 ATPase, as well as two of RSC’s subunits, contain bromodomains that are likely to facilitate the binding of their complexes to acetylated histones.

All eukaryotes contain multiple chromatin-remodeling complexes. They have been classified into four families on the basis of the similarities of their component ATPase subunits: (1) SWI/SNF, (2) **ISWI** (*imitation switch*), (3) **CHD** (*chromodomain, helicase, DNA binding*), and (4) **INO80** (*inositol-requiring protein 80*). Many of these complexes contain bromodomains, chromodomains, and/or AT hooks that presumably recruit the complexes to their target genes. Moreover, some complexes are bound by specific transcriptional activators.

Biochemical studies indicate that RSC binds tightly to nucleosomes in a 1:1 complex. Indeed, cryo-EM-based images of the yeast RSC, determined by Nogales, reveal that it has an irregular shape containing a large central cavity that is remarkably complementary size and shape to a single nucleosome core particle (Fig. 34-85).

The simultaneous release of all of the many interactions holding DNA to a histone octamer would require an enormous free energy input and hence is unlikely to occur. Then, how do chromatin-remodeling complexes function? Their various ATPase subunits share a region of homology with helicases (Section 30-2C), although they lack helicase activity. However, they are DNA translocases that “walk” up DNA strands as driven by ATP hydrolysis. In addition, they are anchored to a fixed point on the nucleosome. This puts torsional strain on the DNA in the nucleosome, thereby decreasing its local twist (DNA supercoiling is discussed in Section 29-3A). The region of decreased twist could diffuse along the DNA wrapped around the nucleosome, thereby transiently loosening the histone octamer’s grip on a segment of DNA. The torsional strain might also be partially accommodated as a writhe, which would lift a segment of DNA off the nucleosome’s surface. In either case, the resulting DNA distortion could diffuse around the surface of the nucleosome in a wave that would locally and transiently release the DNA from the histone octamer as it passed (Fig. 34-86) and hence permit the DNA to bind to its cognate

Figure 34-85 Model of the RSC with a bound nucleosome. The X-ray structure of the nucleosome (with the proteins shown as orange ribbons and the DNA backbone in dark yellow) was fitted into a cryo-EM-based image of the RSC (gray). The two views show the back and front of the complex. The DNA’s dyad (2-fold) axis is represented by a blue rod, and two possible binding sites for the ATP-hydrolyzing translocase subunit are indicated. [Courtesy of Eva Nogales and Andres Leschziner, University of California at Berkeley.]



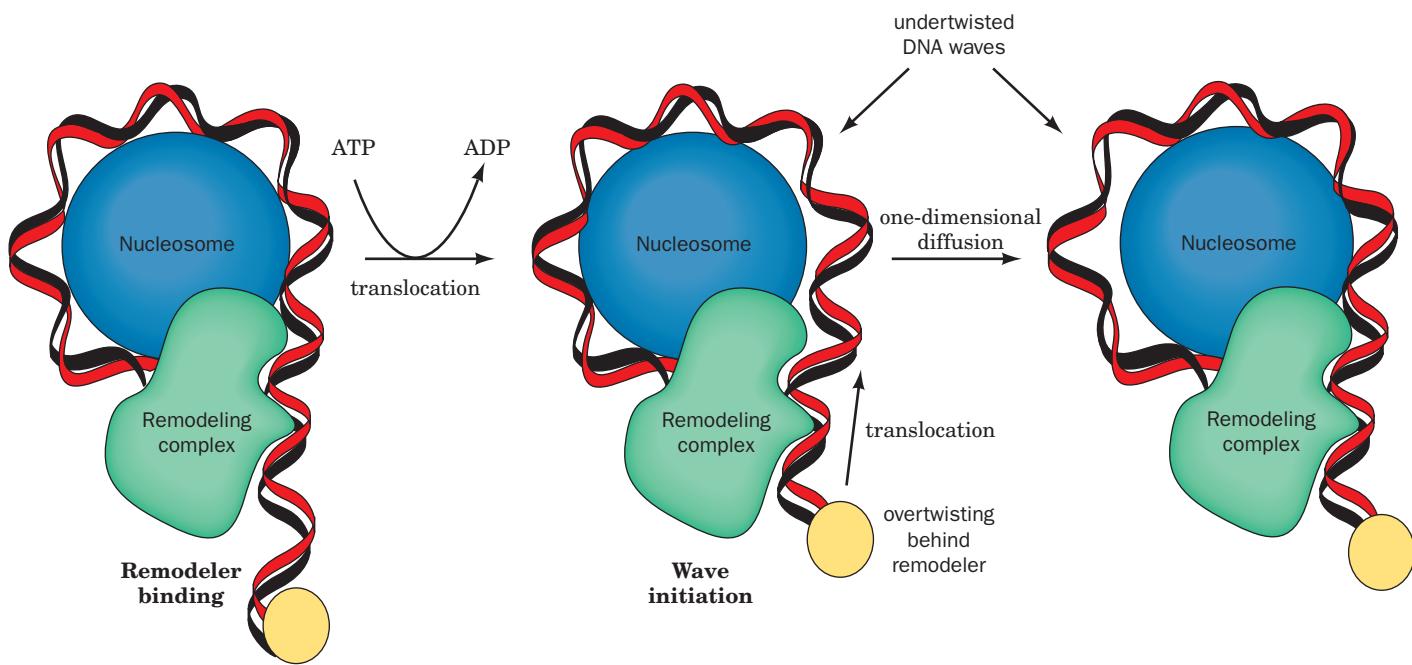


Figure 34-86 Model for nucleosome remodeling by chromatin-remodeling complexes. The chromatin-remodeling complex (green) couples the free energy of ATP hydrolysis to the translocation and concomitant twisting of the DNA in the nucleosome (blue, only half of which is shown for clarity) as depicted by the movement of a fixed point on the DNA (yellow ellipsoid). This locally breaks the contacts between the histones

DNA-binding factors. This latter mechanism resembles that proposed for the passage of RNAPs through nucleosomes (Fig. 34-73). Note that multiple cycles of ATP hydrolysis would send multiple DNA-loosening waves around the nucleosome, thereby sliding the nucleosome along the DNA.

hh. Afterword

As we have seen, eukaryotic transcriptional initiation is an astoundingly complex process that involves the synergistic participation of numerous multisubunit complexes comprising several hundred often loosely or sequentially interacting polypeptides (i.e., histones of various types and subtypes; the PIC; Mediator; a variety of transcription factors, architectural factors, coactivators, and corepressors that in some cases form enhanceosomes; several types of histone modification complexes; and chromatin-remodeling complexes), as well as large segments of DNA. Intensive investigations in many laboratories over the past three decades have, as we have discussed, identified many of these complexes, characterized their component polypeptides, and in many cases, elucidated their general functions. However, we are far from having more than a rudimentary understanding of how these various components interact *in vivo* to transcribe only those genes required by their cell under its particular circumstances in the appropriate amounts and with the proper timing. It is likely to require several additional decades of research to gain a detailed understanding of how this remarkable molecular machinery functions.

and the DNA. The position of the undertwisted and/or bulged DNA propagates around the nucleosome in a one-dimensional wave that transiently releases the DNA from the histone as it passes, thereby providing DNA-binding factors access to the DNA. [After a drawing by Saha, A., Wittmeyer, J., and Cairns, B.R., *Genes Dev.* **16**, 2120 (2002).]

C. Other Expressional Control Mechanisms

The expression of many eukaryotic genes is regulated only by the control of transcriptional initiation. However, many cellular and viral genes additionally respond to other types of control processes. The various mechanisms employed by these secondary systems are reviewed below.

1. Selection of Alternative Initiation Sites: *The expression of several eukaryotic genes is controlled, in part, through the selection of alternative transcriptional initiation sites.* For example, identical molecules of α -amylase are produced by mouse liver and salivary gland but the corresponding mRNAs synthesized by these two organs differ at their 5' ends. Comparison of the sequences of these mRNAs with that of their corresponding genomic DNA indicates that the different mRNAs arise from separate initiation sites that are ~ 2.8 kb apart (Fig. 34-87). Thus, after being spliced, the liver and salivary gland α -amylase mRNAs have different untranslated 5' leaders but the same coding sequences. The two initiation sites support different rates of initiation. This hypothesis accounts for the observation that α -amylase mRNA comprises 2% of the polyadenylated mRNA in salivary gland but only 0.02% of that in liver.

2. Selection of Alternative Splice Sites: In humans, $\sim 95\%$ of structural genes are subject to alternative splicing (Section 31-4Am). Thus, certain exons in one type of cell may be introns in another (e.g., Fig. 31-65).

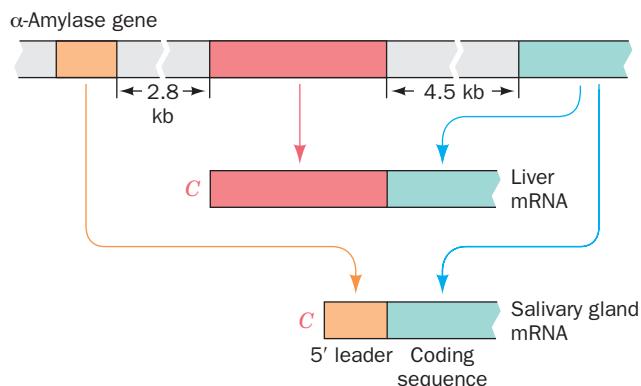


Figure 34-87 The transcription start site of the mouse *α*-amylase gene is subject to tissue-specific selection so as to yield mRNAs with different cap (C) and leader segments but the same coding sequences. [After Young, R.A., Hagenbüchle, O., and Schibler, U., *Cell* **23**, 454 (1981).]

3. Translocational Control: The observation that only ~5% of nuclear RNA ever makes its way to the cytosol suggests that differential mRNA translocation to the cytosol may be an important expressional control mechanism in eukaryotes. Evidence is accumulating that this is, in fact, the case. Cellular RNA is never “naked” but rather is always in complex with a variety of conserved proteins (Section 31-4Au). Intriguingly, nuclear and cytosolic mRNAs are associated with different sets of proteins, indicating that there is protein exchange on translocating mRNA out of the nucleus.

4. Control of mRNA Degradation: The rates at which eukaryotic mRNAs are degraded in the cytosol vary widely. Whereas most mRNAs have half-lives of hours or days, some are degraded within 30 min of entering the cytosol. A given mRNA may also be subject to differential degradation. For example, the major egg yolk protein **vitellogenin** is synthesized in chicken liver in response to estrogens (in roosters as well as in hens) and transported via the bloodstream to the oviduct. Radioactive-labeling experiments established that estrogen stimulation increases the rate of vitellogenin mRNA transcription by several hundredfold and that this mRNA has a cytosolic half-life of 480 h. When estrogen is withdrawn, the synthesis of vitellogenin mRNA returns to its basal rate and its cytosolic half-life falls to 16 h.

The poly(A) tails appended to nearly all eukaryotic mRNAs (Section 31-4Ab), and which are slowly removed in the cytosol by deadenylases (Section 31-4Av), help protect them from degradation. For example, histone mRNAs, which lack poly(A) tails, have much shorter half-lives than most other mRNAs. Histones, it will be recalled (Section 34-1A), are largely synthesized during S phase of the cell cycle (the small amounts of the canonical histones synthesized during the rest of the cell cycle are thought to be used for repair purposes). The short half-lives of histone mRNAs ensure that the rate of histone synthesis closely parallels the rate of histone gene transcription.

5. Control of Translational Initiation Rates: The rates of translational initiation of eukaryotic mRNAs, as we have

seen (Section 32-4), are responsive to the presence of certain substances, including heme (in reticulocytes) and interferon, as well as to mRNA masking (Section 32-4C).

6. Selection of Alternative Post-Translational Processing Pathways: Polypeptides synthesized in both prokaryotes and eukaryotes are subject to proteolytic cleavage and covalent modification (Section 32-5). These post-translational processing steps are important regulators of enzyme activity (e.g., Section 15-3Ea) and, in the case of glycosylations, are major determinants of a protein’s final cellular destination (Sections 12-4Cg and 23-3Bj). The selective degradation of proteins (Section 32-6B) is also a significant factor in eukaryotic gene expression.

In addition to the foregoing, most eukaryotic polypeptide hormones (whose functions are discussed in Section 19-1) are synthesized as segments of large precursor polypeptides known as **polyproteins**. These are post-translationally cleaved to yield several, not necessarily different, polypeptide hormones. *The cleavage pattern of a particular polyprotein may vary among different tissues so that the same gene product can yield different sets of polypeptide hormones.* For example, the polyprotein **pro-opiomelanocortin (POMC)**, which, in the rat, is synthesized in both the anterior and intermediate lobes of the pituitary, contains seven different polypeptide hormones (Fig. 34-88). In both of these lobes, which are functionally separate glands, post-translational processing of POMC yields an N-terminal fragment, **ACTH** and **β-LPH**. Processing in the anterior lobe ceases at this point. In the intermediate lobe, however, the N-terminal fragment is further cleaved to

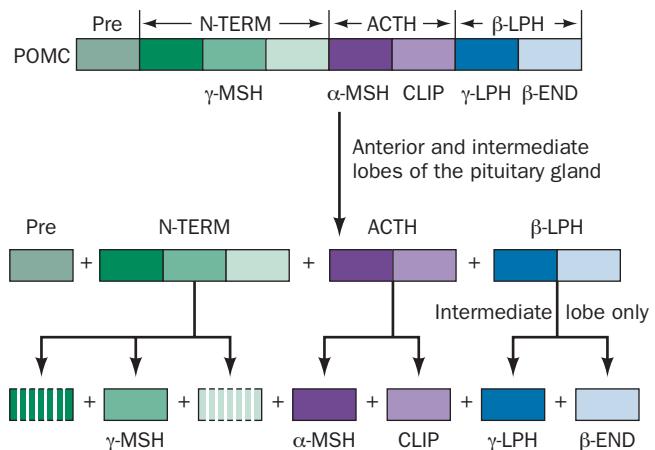


Figure 34-88 The tissue-specific post-transcriptional processing of POMC yields two different sets of polypeptide hormones. In both the anterior and intermediate lobes of the pituitary gland, POMC is proteolytically cleaved to yield its N-terminal fragment (N-TERM), **adrenocorticotrophic hormone (ACTH)**, and **β-lipotropin (β-LPH)**. In the intermediate lobe only, these polypeptide hormones are further cleaved to yield **γ-melanocyte stimulating hormone (γ-MSH)**, **α-MSH**, **corticotropin-like intermediate lobe peptide (CLIP)**, **γ-LPH**, and **β-endorphin (β-END)** (Section 19-1K). [After Douglass, J., Civelli, O., and Herbert, E., *Annu. Rev. Biochem.* **53**, 698 (1984).]

yield **γ -MSH**, ACTH is converted to **α -MSH** and **CLIP**, and β -LPH is split to **γ -LPH** and **β -END** (Fig. 34-88). These various hormones have different activities, so that the products of the anterior and intermediate lobes of the pituitary are physiologically distinct.

Most of the cleavage sites in POMC and other polyproteins consist of pairs of basic amino acid residues, Lys–Arg, for example, which suggests that cleavage is mediated by enzymes with trypsinlike activity. Indeed, the enzymes that process POMC also activate other prohormones such as proinsulin. Moreover, the observation that a yeast protease that normally functions to activate a yeast prohormone also properly processes POMC suggests that prohormone processing enzymes are evolutionarily conserved.

4 CELL DIFFERENTIATION AND GROWTH

Perhaps the most awe inspiring event in biology is the growth and development of a fertilized ovum to form an extensively differentiated multicellular organism. No outside instruction is required to do so; *fertilized ova contain all the information necessary to form complex multicellular organisms such as human beings*. Since, contrary to the beliefs of the earliest microscopists, zygotes do not contain miniature adult structures, these structures must somehow be generated through genetic specification. We begin this section by outlining how embryos develop, followed by a discussion of the best understood example of embryological development, that of the ba-

sic body plan in *Drosophila*. We then consider the genetic basis of cancer, a group of diseases caused by the proliferation of cells that have lost some of their developmental constraints. We end by discussing how the cell cycle is controlled and how unneeded or irreparably damaged cells commit suicide through programmed cell death.

A. Embryological Development

The formation of multicellular animals can be considered as occurring in four somewhat overlapping stages (Fig. 34-89):

1. Cleavage, in which the zygote undergoes a series of rapid mitotic divisions to yield many smaller cells arranged in a hollow ball known as a **blastula**.

2. Gastrulation, whereby the blastula, through a structural reorganization that includes the blastula's invagination, forms a triple-layered bilaterally symmetric structure called a **gastrula**. Cleavage and gastrulation together take from a few hours to several days depending on the organism.

3. Organogenesis, in which the body structures are formed in a process requiring various groups of proliferating cells to migrate from one part of the embryo to another in a complicated but reproducible choreography. Organogenesis occupies hours to weeks.

4. Maturation and growth, whereby the embryonic structures achieve their final sizes and functional capacities. This stage stretches into and sometimes throughout adulthood.

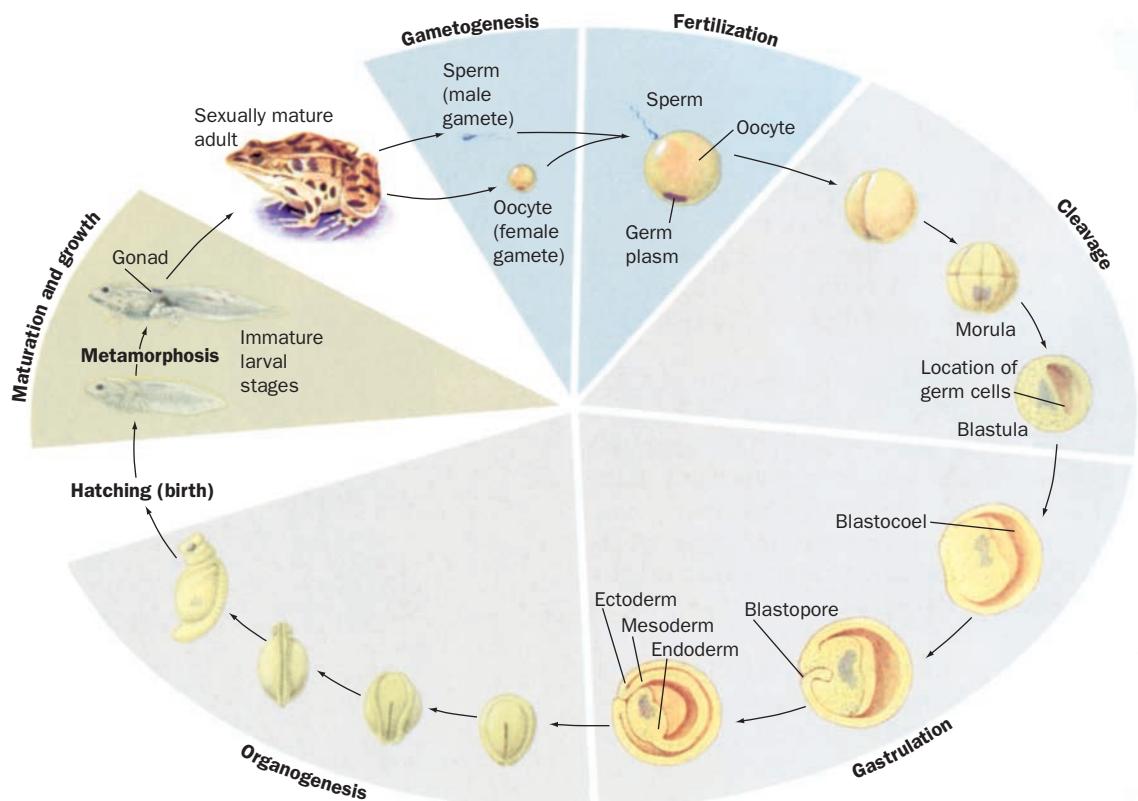


Figure 34-89 Embryogenesis in a representative animal, the frog.

a. Cell Differentiation Is Mediated by Developmental Signals

As an embryo develops, its cells become progressively and irreversibly committed to specific lines of development. What this means is that these cells undergo sequences of self-perpetuating internal changes that distinguish them and their progeny from other cells. A cell and its descendants therefore “remember” their developmental changes even when placed in a new environment. For example, the dorsal (upper) ectoderm (outer layer) of an amphibian embryo (Fig. 34-90) is normally fated to give rise to brain tissue, whereas its ventral (lower) ectoderm becomes epidermis. If a block of an early gastrula’s dorsal ectoderm is cut out and exchanged with a block of its ventral ectoderm, both blocks develop according to their new locations to yield a normal adult. If, however, this experiment is performed on the late gastrula, the transplanted tissues will differentiate as they had originally been fated, that is, as misplaced brain and epidermal tissues. Evidently, the dorsal and ventral ectoderms become committed to form brain and epidermal tissues sometime between the early and late gastrula stages.

How are developmental changes triggered; that is, what are the signals that induce two cells with identical genomes to follow different developmental pathways? To begin with, the zygote is not spherically symmetric. Rather, its yolk, as well as other substances, is concentrated toward one end. Consequently, the various cells in the early cleavage stages inherit different cytoplasmic determinants that apparently govern their further development. Even as early as an embryo’s eight-cell stage, some of its cells are demonstrably different in their developmental potential from others. However, as the above transplantation experiments indicate, cells in later stages of development also obtain developmental cues from their embryonic positions.

Cells may obtain spatial information in two ways:

1. Through direct intercellular interactions.
2. From the gradients of diffusible substances called **morphogens** released by other cells.

For most developmental programs, the interacting tissues must be in direct contact, but this is not always the case. For example, mouse ectoderm fated to become eye lens will only do so in the presence of mesenchyme (embryonic tissue that gives rise to muscle, skeleton, and connective tissue) but this process still occurs if the interacting tissues are separated by a porous filter. Lens development must therefore be mediated by diffusible substances.

Developmental signals may be recognized over great evolutionary distances. For instance, the epidermis from the back of a chick embryo, through interactions with the underlying dermis, forms feather buds that are arrayed in a characteristic hexagonal pattern. If embryonic chick epidermis is instead combined with dermis from the whiskered region of mouse embryo snout, the chick epidermis still forms feather buds but arranged in the pattern of mouse whiskers.

Even though mammals and birds diverged ~300 million years ago, mouse inducers can still activate the appropriate chicken genes, although, of course, they cannot alter the products these genes specify. In an intriguing example of

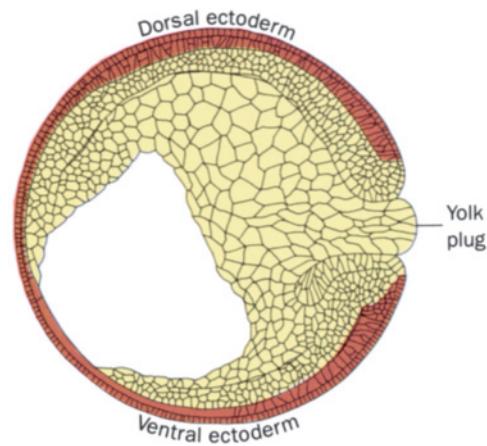


Figure 34-90 The dorsal and ventral ectoderm of an amphibian embryo.

this phenomenon, combining epithelium from the jaw-forming region of a chick embryo with molar mesenchyme from mouse embryo induces the chick tissue to grow teeth that are unlike those of mammals (Fig. 34-91). Apparently chickens, whose ancestors have been toothless for ~60 million years (a primordial bird, *Archaeopteryx*, had teeth), retain the genetic potential to grow teeth even though they lack the developmental capacity to activate these genes. This observation corroborates the hypothesis that organismal evolution proceeds largely via mutations that alter developmental programs rather than the structural genes whose expression they control (Section 7-3Bf).



Figure 34-91 The proverbial “hen’s tooth” forms in chick embryo jaw-forming epithelium under the influence of mouse embryo molar mesenchyme tissue. [Courtesy of Edward Kollar, University of Connecticut Health Center.]

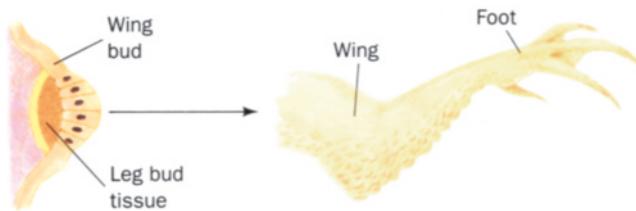


Figure 34-92 Presumptive thigh tissue from a chicken leg bud develops into a misplaced foot when implanted beneath the tip of a chicken wing bud.

b. Developmental Signals Act in Combination

An additional developmental stimulus to a previously determined cell will modulate, but not reverse, its developmental state. Consider, for example, what happens in a chicken embryo if undifferentiated tissue from the base of a leg bud, which normally gives rise to part of the thigh, is transplanted beneath the end of a wing bud, which normally develops into the handlike wing tip. The transplant does not become a wing tip or even misplaced thigh tissue; instead it forms a foot (Fig. 34-92). Apparently the same stimulus that causes the end of a wing bud to form a wing tip causes tissue that is already committed to be part of a leg to form a leg's morphological equivalent to a wing tip, a foot. Evidently, the many different tissues of a higher organism do not each form in response to a tissue-specific developmental stimulus. Rather, *a given tissue results from the effects of a particular combination of relatively nonspecific developmental stimuli*. This situation, of course, greatly reduces the number of different developmental stimuli necessary to form a complex organism and therefore simplifies the regulation of the developmental process.

B. The Molecular Basis of Development

The study of the molecular basis of cell differentiation has only become possible in recent decades with the advent of modern methods of molecular genetics. Much of what we know about this subject is based on studies of the fruit fly *D. melanogaster*. We therefore begin this section with a synopsis of embryogenesis in this genetically best characterized multicellular organism.

a. *Drosophila* Development

Almost immediately after the *Drosophila* egg (Fig. 34-93a) is laid (which, rather than the earlier fertilization, triggers development), it commences a series of synchronized nuclear divisions, one every 6 to 10 min. The DNA must therefore be replicated at a furious rate, the fastest known for eukaryotes. Most probably all of its replicons (Section 30-4Bb) are simultaneously active. The nuclear division

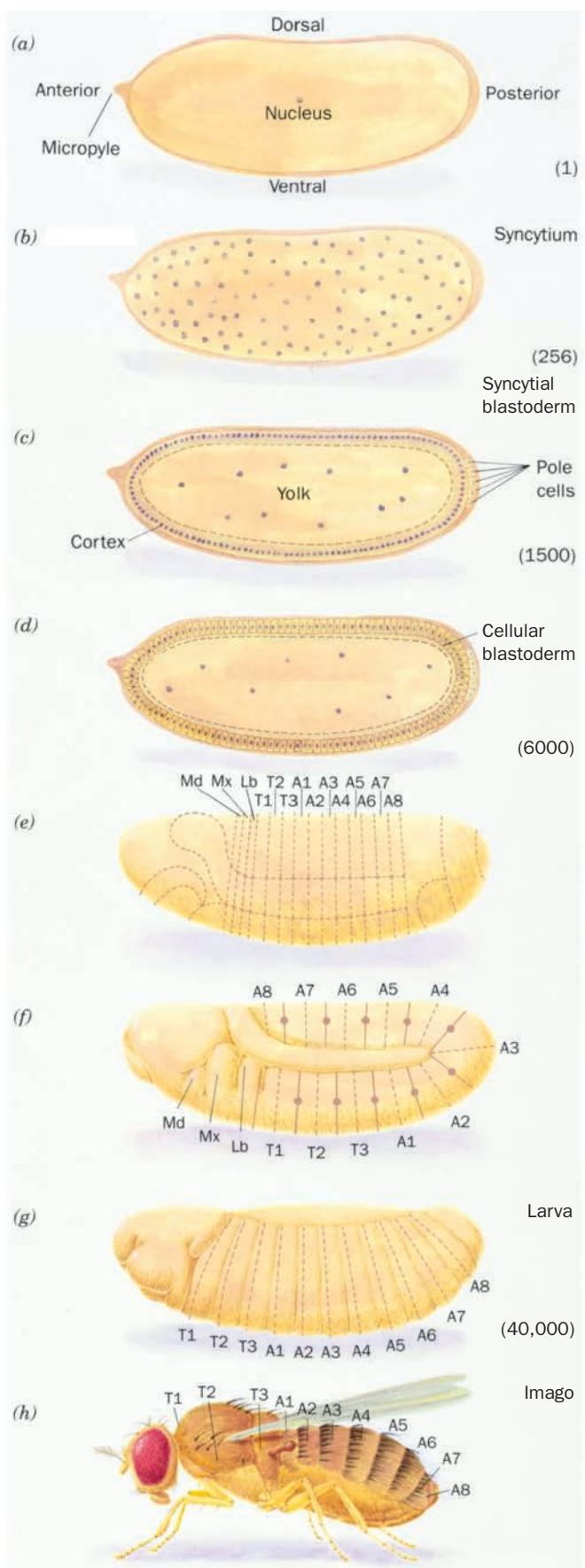


Figure 34-93 Development in *Drosophila*. The various stages are explained in the text. Note that the embryos and newly hatched larva are all the same size, ~0.5 mm long. The adult is, of course, much larger. The approximate number of cells in the early stages of development are given in parentheses.

process is unusual in that it is not accompanied by the formation of new cell membranes; the nuclei continue sharing their common cytoplasm to form a so-called **syncytium** (Fig. 34-93b), which facilitates the rapid pace of nuclear division because there is no need to increase cell mass. After the 8th round of nuclear division, the ~256 nuclei begin to migrate toward the cortex (outer layer) of the egg where, by around the 11th nuclear division, they have formed a single layer surrounding a yolk-rich core known as a **syncytial blastoderm** (Fig. 34-93c). At this stage, the mitotic cycle time begins to lengthen while the nuclear genes, which have heretofore been fully engaged in DNA replication, become transcriptionally active (a freshly laid egg contains an enormous store of mRNA that has been contributed, via cytoplasmic bridges, by the developing oocyte's 15 surrounding "nurse" cells). In the 14th nuclear division cycle, which lasts ~60 min, the egg's plasma membrane invaginates around each of the ~6000 nuclei to yield a cellular monolayer surrounding a yolk-rich core called a **cellular blastoderm** (Fig. 34-93d). At this point, after ~2.5 h of development, genomic transcriptional activity reaches its maximum in the embryo, mitotic synchrony is lost, the cells become motile, and gastrulation begins.

Until the cellular blastoderm is formed, most of the embryo's nuclei maintain the ability to colonize any portion of the cortical cytoplasm and hence to form any part of the larva or adult except its germ cells [the germ cell progenitors, the five **pole cells** (Fig. 34-93c), are set aside after the 9th nuclear division]. *It is therefore a nucleus' location within the syncytium that determines the types of cells its descendants will be become. Once the cellular blastoderm has formed, however, its cells become progressively committed to ever narrower lines of development.* This has been demonstrated, for example, by tracing the developmental fates of small clumps of cells by excising them or ablating (destroying) them with a laser microbeam and characterizing the resultant deformity.

During the embryo's next few hours, it undergoes gastrulation and organogenesis. A striking aspect of this remarkable process, in *Drosophila* as well as in higher animals, is the division of the embryo into a series of segments corresponding to the adult organism's organization (Fig. 34-93e). The *Drosophila* embryo has three segments that eventually merge to form its head (Md, Mx, and Lb for mandibular, maxillary, and labial), three thoracic segments (T1-T3), and eight abdominal segments (A1-A8). As development continues, the embryo elongates and several of its abdominal segments fold over its thoracic segments (which permits it to fit inside the eggshell; Fig. 35-93f). At this stage, the segments become subdivided into anterior (forward) and posterior (rear) compartments. The embryo then shortens and straightens to form a larva that consists of ~40,000 cells that hatches ~20 h after beginning development (Fig. 34-93g). Over the next 5 days, the larva feeds, grows, molts twice, pupates, and commences metamorphosis to form an adult (**imago**; Fig. 34-93h). In this latter process, the larval epidermis is almost entirely replaced by the outgrowth of apparently undifferentiated patches of larval epithelium known as **imaginal disks** that are commit-

ted to their developmental fates as early as the cellular blastoderm stage. These structures, which maintain the larva's segmental boundaries, form the adult's legs, wings, antennae, eyes, etc. (Fig. 34-94). About 10 days after commencing development, the adult emerges and, within a few hours, initiates a new reproductive cycle.

b. Developmental Patterns Are Genetically Mediated

What is the mechanism of embryonic pattern formation? In what follows, we discuss only the anteroposterior (head to tail) differentiation system. Keep in mind, however, that *Drosophila* also have a system that imposes dorsoventral (back to belly) differentiation.

Much of what we know about anteroposterior pattern formation stems from genetic analyses of a series of bizarre mutations in three classes of *Drosophila* genes that normally specify progressively finer regions of cellular specialization in the developing embryo:

1. Maternal-effect genes, which define the embryo's polarity, that is, its anteroposterior axis. Mutations of these genes globally alter the embryonic body pattern regardless of the paternal genotype. For instance, females homozygous for the **dicephalic** (two-headed) mutation lay eggs that develop into nonviable two-headed monsters. These are embryos with two anterior ends pointing in opposite directions and completely lacking posterior structures. Similarly, the **bicaudal** (two-tailed) and **snake mutations** give rise to mirror-symmetric embryos with two abdomens (Fig. 34-95a).

2. Segmentation genes, which specify the correct number and polarity of embryonic body segments. Investigations

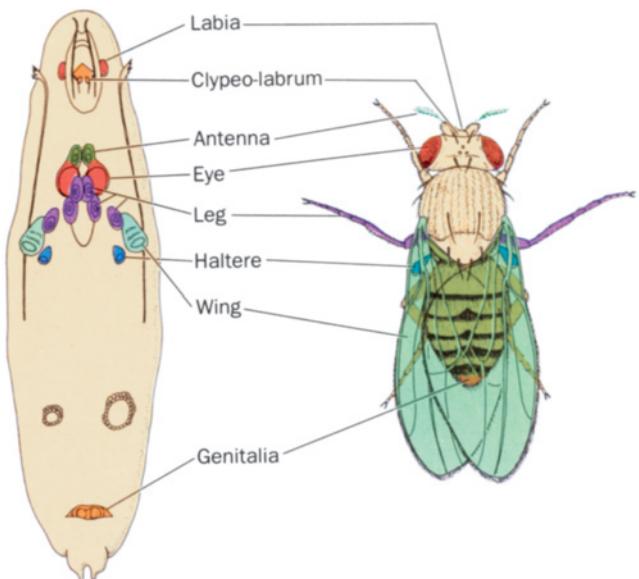


Figure 34-94 Locations and developmental fates, in *Drosophila*, of the imaginal disks (left), pouches of larval tissue that form the adult's outer structures. [After Fristrom, J.W., Raikow, R., Petri, W., and Stewart, D., in Hanly, E.W. (Ed.), *Problems in Biology: RNA in Development*, p. 382, University of Utah Press (1970).]

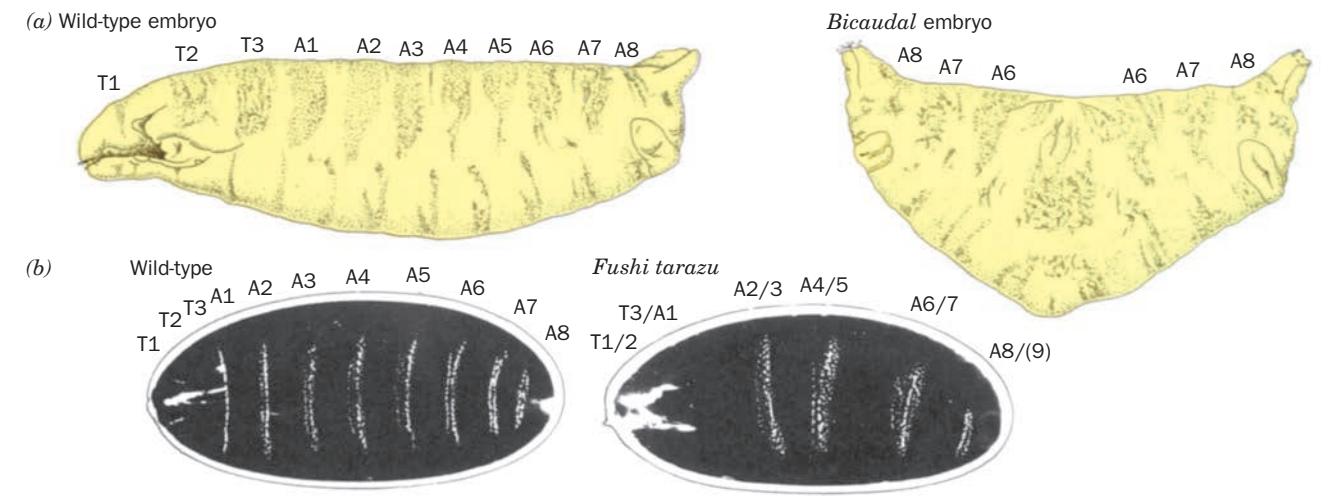


Figure 34-95 Developmental mutants of *Drosophila*. (a) The cuticle patterns of wild-type embryos (left) exhibit 11 body segments, T1 to T3 and A1 to A8 (the head segments have retracted into the body and hence are not visible here). In contrast, the nonviable “monsters” produced by homozygous *bicaudal* mutant females (right) develop only abdominal segments arranged with mirror symmetry. [After Gergen, P.J., Coulter, D., and Weischaus, E., in Gall, J.G., *Gametogenesis and the Early Embryo*, p. 200, Liss (1986).] (b) In the wild-type embryo (left), the anterior edge of each of the 11 abdominal and thoracic segments has a belt of tiny projections known as denticles (which help larvae crawl) that appear in these photomicrographs as white stripes. *Fushi tarazu* mutants (right) lack portions of alternate segments and the remaining segments are fused together (e.g., A2/3), yielding a nonviable embryo with only half of the normal number of denticle belts. [Courtesy of Walter Gehring, University of Basel, Switzerland.] (c) Head and thorax of a wild-type adult fly (left) and one that is homozygous for a mutant form of the homeotic *Antennapedia* (*Antp*) gene (right). The mutant gene is inappropriately expressed in the imaginal disks that normally form antennae (where the wild-type *Antp* gene is not expressed) so that they develop as the legs that normally occur only on segment T2. [Courtesy of Ginés Morata, Universidad Autónoma de Madrid, Spain.] (d) The correspondence (arrows) between antennae and the legs to which the *Antp* mutation transforms them. [After Postlethwait, J.H. and Schneiderman, H.A., *Devel. Biol.* **25**, 622 (1971).] (e) A four-winged *Drosophila* (it normally has two wings; Fig. 34-94) that results from the presence of three mutations in the bithorax complex, *abx*, *bx*, and *pbx*. These mutations cause the normally haltere-bearing segment T3 to develop as if it were the wing-bearing segment T2. This striking architectural change may

reflect evolutionary history: *Drosophila* evolved from more primitive insects that had four wings. [Courtesy of Edward B. Lewis, Caltech.]

by Christiane Nüsslein-Volhard and Eric Wieschaus led to their subclassification as follows:

a. **Gap genes**, the first of a developing embryo's to be transcribed, divide the embryo into several broad regions. Gap genes are so named because their mutations result in gaps in the embryo's segmentation pattern. Embryos with defective *hunchback* (*hb*) genes, for example, lack mouthparts and thorax structures.

b. **Pair-rule genes** specify the division of the embryo's broad gap regions into segments. These genes are so named because their mutations usually delete portions of every second segment. This occurs, for example, in embryos that are homozygous for mutations in the *fushi tarazu* (*ftz*; Japanese for segment deficient) gene (Fig. 34-95b).

c. **Segment polarity genes** specify the polarities of the developing segments. Thus, homozygous *engrailed* (*en*; indented with curved notches) mutants lack the posterior compartment of each segment.

3. **Homeotic selector genes**, which specify segmental identity. Homeotic mutations transform one body part into another. For instance, *Antennapedia* (*antp*, antenna-foot) mutants have legs in place of antennae (Fig. 34-95c,d), whereas the mutations *bithorax* (*bx*), *anteriorbithorax* (*abx*), and *postbithorax* (*pbx*) each transform sections of halteres (vestigial wings that function as balancers; Fig. 34-94), which normally occur only on segment T3, to the corresponding sections of wings, which normally occur only on segment T2 (Fig. 34-95e).

These genes, as we shall see, each encode transcription factors that control the expression of the succeeding genes in this list.

c. Maternal-Effect Gene Products Specify the Egg's Directionality through Gradient Formation

The properties of maternal-effect gene mutants suggest that maternal-effect genes specify morphogens whose distributions in the egg cytoplasm define the future embryo's spatial coordinate system. Indeed, immunofluorescence studies by Nüsslein-Volhard demonstrated that the product of the *bicoid* (*bcd*) gene is distributed in a gradient that decreases toward the posterior end of the normal embryo (Figs. 34-96 and 34-97a), whereas embryos with *bcd*-deficient mothers lack this gradient. The gradient, which is facilitated by the syncytium's lack of cellular boundaries, arises through the secretion, by the ovarian nurse cells, of *bcd* mRNA into the anterior end of the oocyte during oogenesis, where it attaches to the cytoskeleton via its 3' untranslated region, followed by its translation in the early embryo. The *nanos* gene mRNA is similarly deposited in the egg but it is localized near the egg's posterior pole (Fig. 34-97a). The *bcd* and *nanos* gene products, as we shall see, are transcription factors that regulate the expression of specific gap genes. Other maternal-effect genes that participate in anteroposterior axis formation specify proteins that function to trap the local-

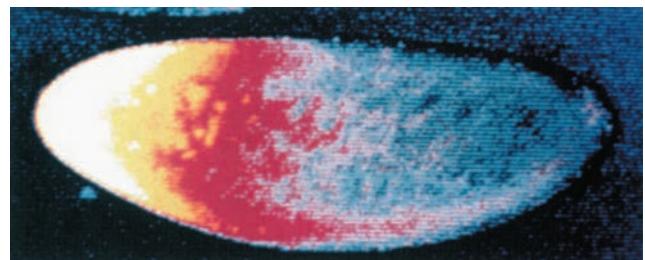


Figure 34-96 The distribution of Bicoid protein in a *Drosophila* syncytial blastoderm as revealed by immunofluorescence. High concentrations of the protein are yellow, lower concentrations are red, and its absence is black. [Courtesy of Christiane Nüsslein-Volhard, Max-Planck-Institut für Entwicklungsbiologie, Germany.]

ized mRNAs in their area of deposition. This explains why early embryos produced by females homozygous for maternal-effect mutations can often be “rescued” by the injection of cytoplasm, or sometimes just the mRNA, from early wild-type embryos. With some of these mutations, the polarity of the rescued embryo is determined by the site of the injection.

d. Gap Genes Are Expressed in Specific Regions

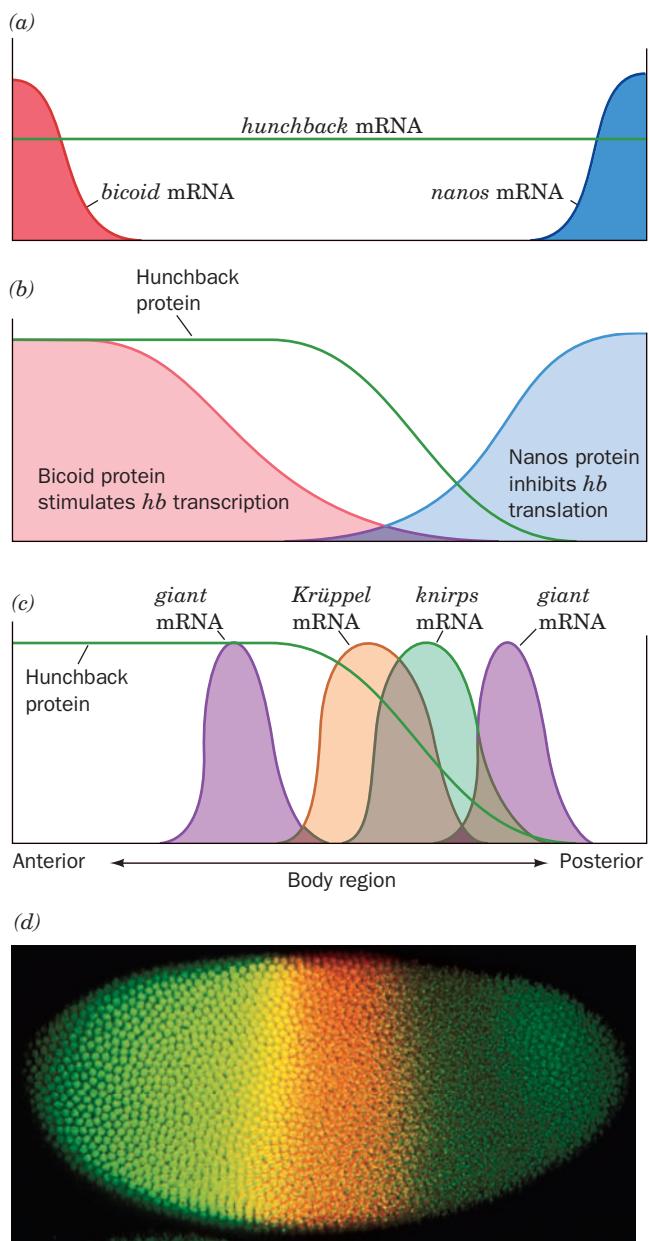
The mRNA of the gap gene *hunchback* (*hb*) is maternally deposited in a uniform distribution throughout the unfertilized egg (Fig. 34-97a). However, **Bicoid protein** activates the transcription of the embryonic *hb* gene, whereas **Nanos protein** inhibits the translation of *hb* mRNA. Consequently, **Hunchback protein** becomes distributed in a gradient that decreases from anterior to posterior (Fig. 34-97b).

DNase I footprinting studies have demonstrated that Bicoid protein binds to five homologous sites (consensus sequence TCTAATCCC) in the *hb* gene's upstream promoter region. Nüsslein-Volhard demonstrated the ability of Bicoid protein to activate the *hb* gene by fusing the *hb* promoter upstream of the CAT reporter gene (Section 34-3Bh) and injecting the resulting construct into early *Drosophila* embryos. CAT was produced in wild-type but not in *bcd*-deficient embryos. Moreover, the use of progressively shorter segments of the *hb*-derived promoter region demonstrated that at least three of the five Bicoid protein-binding sites must be present to obtain full CAT expression.

Hunchback protein, in turn, controls the expression of several other gap genes (Fig. 34-97c,d): High levels of Hunchback protein induce **giant** expression, **Krüppel** (German: cripple) is expressed where the level of Hunchback protein begins to decline, **knirps** (German: pygmy) is expressed at even lower levels of Hunchback protein, and **giant** is again activated in regions where Hunchback protein is undetectable.

Although the original positions of the proteins encoded by these latter gap genes are elicited by the appropriate concentrations of Hunchback protein, these positions are

Figure 34-97 Formation and effects of the Hunchback protein gradient in *Drosophila* embryos. (a) The unfertilized egg contains maternally supplied *bicoid* and *nanos* mRNAs placed at its anterior and posterior poles, together with a uniform distribution of *hunchback* mRNA. (b) On fertilization, the three mRNAs are translated. Bicoid and Nanos proteins are not bound to the cytoskeleton as are their mRNAs and hence their gradients are broader than those of the mRNAs. Bicoid protein stimulates the transcription of the *hunchback* gene, whereas Nanos protein inhibits its translation, resulting in a gradient of Hunchback protein that decreases nonlinearly from anterior to posterior. (c) Specific concentrations of Hunchback protein induce the transcription of the *giant*, *Krüppel*, and *knirps* genes. The gradient of Hunchback protein thereby specifies the positions at which these latter mRNAs are synthesized. (d) A photomicrograph of a *Drosophila* embryo (anterior end left) that has been immunofluorescently stained for both Hunchback (green) and Krüppel proteins (red). The region where these proteins overlap is yellow. [Parts a, b, and c after Gilbert, S.F., *Developmental Biology* (4th ed.), p. 543, Sinauer Associates (1994); Part d courtesy of Jim Langeland, Steve Paddock, and Sean Carroll, Howard Hughes Medical Institute, University of Wisconsin–Madison.]



stabilized and maintained through their mutual interactions. Thus **Krüppel protein** binds to the promoters of the *hb* gene, which it activates, and the *knirps* gene, which it represses. Conversely, **Knirps protein** represses the *Krüppel* gene. This mutual repression is thought to be responsible for the sharp boundaries between the various gap domains.

e. Pair-Rule Genes Are Expressed in “Zebra Stripes”

Pair-rule genes are expressed in sets of 7 stripes, each just a few nuclei wide, along the embryo's anterior–posterior axis. The embryo (which, at this stage, is just beginning to cellularize) is thereby divided into 15 domains (Fig. 34-98). These “zebra stripe” expression patterns for the various pair-rule genes are offset relative to one another.

The gap gene products are transcription factors for three **primary pair-rule genes**: *hairy*, *even-skipped* (*eve*), and *runt*. The striped pattern of expression arises because the control regions of most primary pair-rule genes comprise a

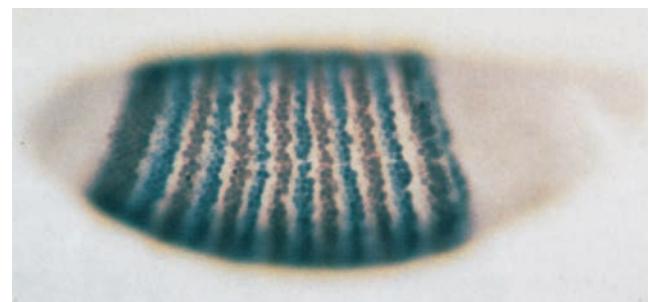
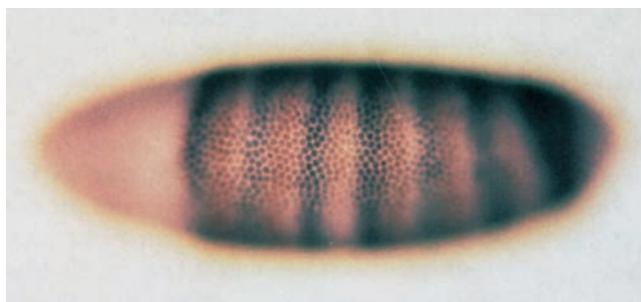


Figure 34-98 *Drosophila* embryos stained for *Ftz* (brown) and *Eve* (blue) proteins. These proteins are each expressed in seven stripes which, at first, are relatively blurred (left) but within

a short time become sharply defined (right). [Courtesy of Peter Lawrence, MRC Laboratory of Molecular Biology, U.K.]

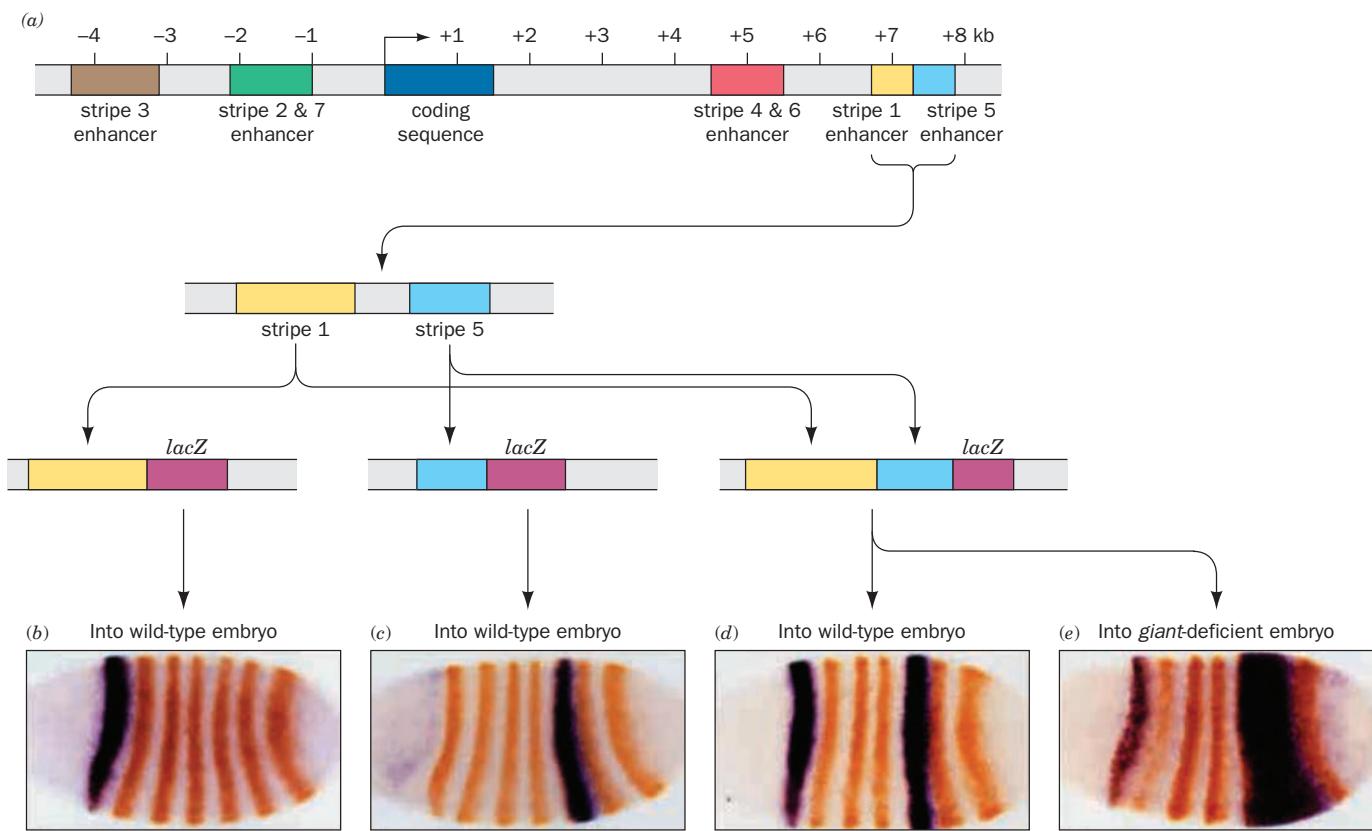


Figure 34-99 Expression of the even-skipped (eve) gene in a pattern of seven stripes in the *Drosophila* embryo. (a) Diagram of the eve gene, which contains a series of enhancer modules, some upstream of the coding sequence (blue) and others downstream, that on binding the particular combination of gap proteins present in their corresponding stripe, induce the expression of eve in that stripe. The positions of various elements in the gene (in kb) relative to the transcriptional start site (squared off arrow) are indicated. A transgenic lacZ reporter gene (magenta) under the control of (b) the stripe 1 enhancer (yellow), (c) the stripe 5 enhancer (cyan), or (d) both the stripe 1 and 5 enhancers were analyzed in otherwise wild-type

Drosophila embryos. The resulting embryos were hybridized *in situ* with dye-labeled lacZ antisense RNA to yield a black color where lacZ had been transcribed and then stained with anti-Eve antibodies (orange), thereby demonstrating that lacZ is expressed only in the corresponding stripe(s). (e) When the lacZ transgene under the control of the stripe 1 and 5 enhancers was analyzed in embryos deficient for the gap gene giant, stripe 1 was normal but stripe 5 lacked its posterior border, which indicates that Giant protein normally functions to inhibit eve expression past the end of stripe 5. [Part a based on a drawing by Scott Gilbert, Swarthmore College. Parts b, c, d, and e courtesy of James Jaynes and Miki Fujioka, Thomas Jefferson University.]

series of enhancer modules, each of which induce their gene's expression in a particular stripe (Fig. 34-99a). For example, the transformation of an embryo by the lacZ gene preceded by a specific enhancer module in the eve gene resulted in lacZ transcription in only stripe 1 (Fig. 34-99b), whereas a different module did so in only stripe 5 (Fig. 34-99c), and both of these modules together induced the production of the lacZ transcript in both stripes 1 and 5 (Fig. 34-99d). Each of these modules contains a particular arrangement of activating and inhibitory binding sites for the various gap gene proteins so as to enable the expression of the associated pair-rule gene under the particular combination of gap gene proteins present in the corresponding stripe. Thus, in giant-deficient embryos, the posterior border of stripe 5 is missing (Fig. 34-99e). As with the gap genes, the patterns of expression of the primary pair-rule genes become stabilized through interactions among themselves.

The primary pair-rule gene products similarly induce or inhibit the expression of five **secondary pair-rule genes** including ftz. Thus, as Walter Gehring demonstrated, ftz transcripts first appear in the nuclei lining the cortical cytoplasm during the embryo's 10th nuclear division cycle. The rate of ftz expression then increases as the embryo develops until the 14th division cycle, when the cellular blastoderm forms. At this stage, as immunochemical staining dramatically shows, ftz is expressed in a pattern of 7 belts around the cellular blastoderm, each 3 or 4 cells wide (Fig. 34-98), which correspond precisely to the missing regions in homozygous *ftz*⁻ embryos. Then, as the embryonic segments form, ftz expression subsides to undetectable levels (although it is later reactivated during the differentiation of specific nerve cells in which it is required to specify their correct "wiring" pattern). Evidently, the ftz gene must be expressed in alternate sections of the embryo for normal segmentation to occur.

f. Segment Polarity Genes Define Parasegment Boundaries

The expression of eight known segment polarity genes is initiated by pair-rule gene products. For example, by the 13th nuclear division cycle, as Thomas Kornberg demonstrated, *engrailed* (*en*) transcripts become detectable but are more or less evenly distributed throughout the embryonic cortex. However, since *en* is expressed in nuclei containing high concentrations of either **Eve** or **Ftz** proteins (Fig. 34-98), by the 14th cycle they form a striking pattern of 14 stripes around the cellular blastoderm (half the spacing of *ftz* expression). Continuing development reveals that these stripes are localized in the primordial posterior compartment of every segment (Fig. 34-100), just those compartments that are missing in homozygous *en*⁻ embryos. Thus, much like we saw for *ftz*, the *en* gene product induces the posterior half of each segment to develop in a different fashion from its anterior half.

Another segment polarity gene, *wingless* (*wg*), is expressed simultaneously with *en* but in narrow bands on the anterior side of most *en* bands (Fig. 34-101). Cells expressing *en* and *wg* genes thereby define the boundaries of the so-called **parasegments**, embryonic regions that consist of the posterior portion of one segment and the anterior portion of the segment behind it. Parasegments do not become morphological units in the larva or adult but, nevertheless, are thought to be the embryo's actual developmental units.

g. Homeotic Selector Genes Direct the Development of the Individual Body Segments

The structural components of developmentally analogous body parts, say, *Drosophila* antennae and legs, are nearly identical; only their organizations differ (Fig. 34-95d). Consequently, developmental genes must function to control the pattern of structural gene expression rather than simply turning these genes on or off. Thus, as we saw for the segmentation genes, the expression of the structural genes characteristic of any given tissue must be controlled by a

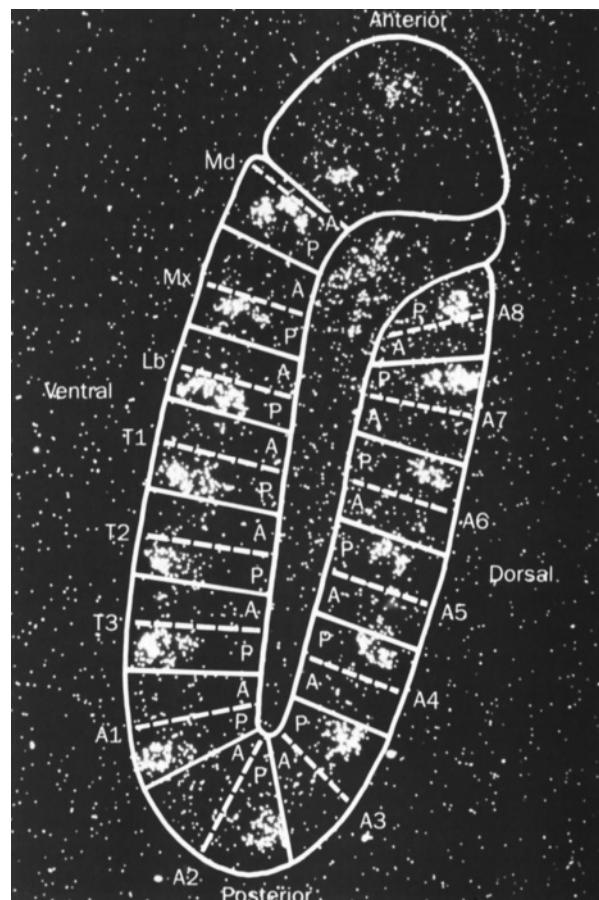


Figure 34-100 *In situ* hybridization demonstrates that the *Drosophila engrailed* gene is expressed in the posterior compartment of every embryonic segment. [Courtesy of Walter Gehring, University of Basel, Switzerland.]

complex network of regulatory genes. The homeotic selector genes, as we shall see, are the “master” genes in the control networks governing segmental differentiation.

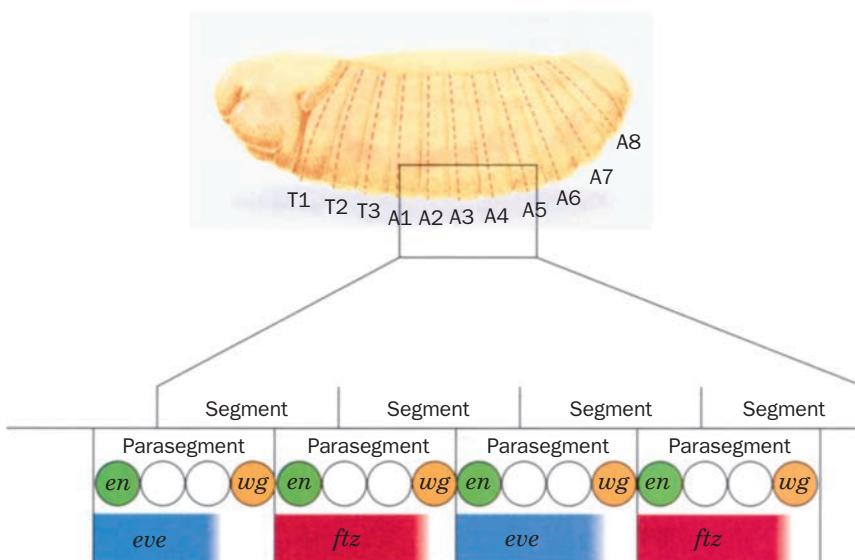


Figure 34-101 The pair-rule proteins **Eve** and **Ftz** regulate the expression of the segment polarity genes *engrailed* (*en*) and *wingless* (*wg*). When either Eve or Ftz is present, *en* is expressed, whereas when both proteins are absent, *wg* is expressed. The parasegment boundaries are thereby defined. Other pair-rule proteins are thought to inhibit *en* and *wg* expression in nuclei not at the parasegment boundaries.

Most homeotic mutations in *Drosophila* (which were first described in 1894 by William Bateson, who coined the name “homeosis” to indicate something that has been changed into the likeness of something else) map into eight related genes that are distributed in two clusters that function as a single cluster: the **bithorax complex (BX-C)**, which controls the development of the thoracic and abdominal segments, and the **antennapedia complex (ANT-C)**, which primarily affects head and thoracic segments. *Recessive mutations in BX-C, when homozygous, cause one or more segments to develop as if they were more anterior segments.* Thus, the combined ***bx***, ***abx***, and ***pbx*** mutations cause segment T3 to develop as if it were segment T2 (Fig. 34-95e). Similarly, the entire deletion of *BX-C* causes all segments posterior to T2 to resemble T2; apparently T2 is the developmental “ground state” of these 10 segments. The evolution of such gene families, it is thought, permitted arthropods (the phylum containing insects) to arise from the more primitive annelids (segmented worms) in which all segments are nearly alike.

Detailed genetic analysis of *BX-C* led Edward B. Lewis to formulate a model for segmental differentiation (Fig. 34-102). *BX-C*, Lewis proposed, contains at least one gene for each segment from T3 to A8, which for simplicity are numbered 0 to 8 in Fig. 34-102. These genes, for reasons that are not understood, are arranged in the same order, from “left” to “right,” as the segments whose development they influence. Starting with segment T3, progressively more posterior segments express successively more *BX-C* genes until, in segment A8, all of these genes are expressed. The developmental fate of a segment is thereby determined by its position in the embryo.

Sequence analysis of the *BX-C* region led to a difficulty with Lewis’ model: The *BX-C* contains only three protein-encoding genes, ***Ultrabithorax (Ubx)***, ***Abdominal-A (Abd-A)***, and ***Abdominal-B (Abd-B)***. However, further analysis indicated, for example, that mutations such as *bx*, *abx*, and *pdx*, which were previously assumed to occur on separate genes, are actually mutations of enhancer elements that enable the position-specific expression of the *Ubx* gene. Thus, the nine “genes” in Lewis’ model have turned out to be enhancer elements on the three *BX-C* genes.

h. Developmental Genes Have Common Sequences

In characterizing the ***Antennapedia (Antp)*** gene, Gehring and Matthew Scott independently discovered that *Antp* cDNA hybridizes to both the *Antp* and the *fitz* gene and that, therefore, *these genes share a common base sequence*. This startling observation rapidly led to the discovery that the *Drosophila* genome contains numerous such sequences, many of which occur in the homeotic gene complexes *ANT-C* and *BX-C*. DNA sequencing studies of these genes revealed that each contains a 180-bp sequence, the so-called **homeodomain** or **homeobox**, which are 70 to 90% identical to one another and which encode even more identical 60-residue polypeptide segments (Fig. 34-103).

Further hybridization studies using homeodomain probes led to the truly astonishing finding that *multiple copies of the homeodomain are also present in the genomes of segmented animals ranging from annelids to vertebrates*.

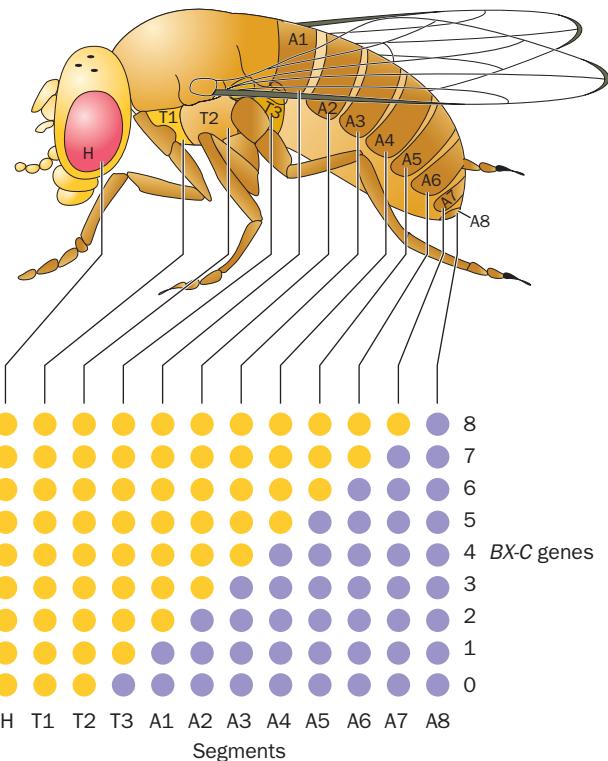


Figure 34-102 Model for the differentiation of embryonic segments in *Drosophila* as directed by the genes of the bithorax complex (BX-C). Segments T2, T3, and A1–8 in the embryo, as the lower drawing indicates, are each characterized by a unique combination of active (purple circles) and inactive (yellow circles) BX-C “genes.” These “genes” (which sequencing studies later demonstrated are really enhancer elements), here numbered 0 to 8, are thought to be sequentially activated from anterior to posterior in the embryo so that segment T2, the developmentally most primitive segment, has no active BX-C genes, while in segment A8, all of them are active. Such a pattern of gene expression may result from a gradient in the concentration of a BX-C repressor that decreases from the anterior to the posterior of the embryo. [After Ingham, P., *Trends Genet.* **1**, 113 (1985).]

such as *Xenopus*, *mice*, and *humans*. In some of these sequences the degree of homology is remarkably high; for example, the homeodomains of the *Drosophila Antp* gene and the *Xenopus MM3 gene* encode polypeptides that have 59 of their 60 amino acids in common (Fig. 34-103). The individuality of these homeodomain-containing proteins is presumably imparted by their other segments.

i. The Homeodomain’s DNA-Binding Motif Resembles a Helix-Turn-Helix Motif

Since vertebrates and invertebrates diverged over 600 million years ago, this strongly suggests that the gene product of the homeodomain has an essential function. What might this function be? The ~30% Arg + Lys content of homeodomain polypeptides suggests that they bind DNA. Sequence comparisons and NMR studies further suggest that these polypeptide segments form helix-turn-helix (HTH) motifs resembling those of prokaryotic gene regulators such as the *E. coli trp* repressor (Section 31-3Da) and

	1	10										20
Mouse <i>MO-10</i>	Ser	Lys	Arg	Gly	Arg	Thr	Ala	Tyr	Thr	Arg	Pro	Gln
Frog <i>MM3</i>	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln
<i>Antennapedia</i>	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln
<i>Fushi tarazu</i>	Ser	Lys	Arg	Thr	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln
<i>Ultrabithorax</i>	Arg	Arg	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln
	21	30										40
Mouse <i>MO-10</i>	His	Phe	Asn	Arg	Tyr	Leu	Met	Arg	Pro	Arg	Arg	Val
Frog <i>MM3</i>	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Glu
<i>Antennapedia</i>	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Ile	Ile
<i>Fushi tarazu</i>	His	Phe	Asn	Arg	Tyr	Ile	Thr	Arg	Arg	Arg	Ile	Glu
<i>Ultrabithorax</i>	His	Thr	Asn	His	Tyr	Leu	Thr	Arg	Arg	Arg	Ile	Asp
	41	50										60
Mouse <i>MO-10</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg
Frog <i>MM3</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg
<i>Antennapedia</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg
<i>Fushi tarazu</i>	Ser	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg
<i>Ultrabithorax</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg
	42										HTH Helix 3	
	50										HTH Helix 2	
	60										Homeodomain Helix 3	
	70										Homeodomain Helix 2	
	80										HTH Helix 3	
	90										HTH Helix 2	
	100										Homeodomain Helix 3	
	110										Homeodomain Helix 2	
	120										HTH Helix 3	
	130										HTH Helix 2	
	140										Homeodomain Helix 3	
	150										Homeodomain Helix 2	
	160										HTH Helix 3	
	170										HTH Helix 2	
	180										Homeodomain Helix 3	
	190										Homeodomain Helix 2	
	200										HTH Helix 3	
	210										HTH Helix 2	
	220										Homeodomain Helix 3	
	230										Homeodomain Helix 2	
	240										HTH Helix 3	
	250										HTH Helix 2	
	260										Homeodomain Helix 3	
	270										Homeodomain Helix 2	
	280										HTH Helix 3	
	290										HTH Helix 2	
	300										Homeodomain Helix 3	
	310										Homeodomain Helix 2	
	320										HTH Helix 3	
	330										HTH Helix 2	
	340										Homeodomain Helix 3	
	350										Homeodomain Helix 2	
	360										HTH Helix 3	
	370										HTH Helix 2	
	380										Homeodomain Helix 3	
	390										Homeodomain Helix 2	
	400										HTH Helix 3	
	410										HTH Helix 2	
	420										Homeodomain Helix 3	
	430										Homeodomain Helix 2	
	440										HTH Helix 3	
	450										HTH Helix 2	
	460										Homeodomain Helix 3	
	470										Homeodomain Helix 2	
	480										HTH Helix 3	
	490										HTH Helix 2	
	500										Homeodomain Helix 3	
	510										Homeodomain Helix 2	
	520										HTH Helix 3	
	530										HTH Helix 2	
	540										Homeodomain Helix 3	
	550										Homeodomain Helix 2	
	560										HTH Helix 3	
	570										HTH Helix 2	
	580										Homeodomain Helix 3	
	590										Homeodomain Helix 2	
	600										HTH Helix 3	
	610										HTH Helix 2	
	620										Homeodomain Helix 3	
	630										Homeodomain Helix 2	
	640										HTH Helix 3	
	650										HTH Helix 2	
	660										Homeodomain Helix 3	
	670										Homeodomain Helix 2	
	680										HTH Helix 3	
	690										HTH Helix 2	
	700										Homeodomain Helix 3	
	710										Homeodomain Helix 2	
	720										HTH Helix 3	
	730										HTH Helix 2	
	740										Homeodomain Helix 3	
	750										Homeodomain Helix 2	
	760										HTH Helix 3	

one end, where it also contacts a second DNA molecule that, in the crystal, forms a pseudocontinuous helix with the first DNA. The conformations of the two protein molecules, and the contacts they make with the DNA, are nearly identical. The two homeodomains are not in contact and so, in contrast to other DNA-binding motifs of known structure, *they bind to their target DNAs as monomers*. The X-ray structure is largely consistent with the NMR structure of the *Antennapedia* homeodomain in complex with a 14-bp DNA determined by Gehring and Kurt Wüthrich.

The homeodomain consists largely of three α helices, the last two of which, as sequence comparisons had previously suggested, form an HTH motif that is closely superimposable with the HTH motifs of prokaryotic repressors such as that of the λ repressor (Fig. 33-45a). However, although helix 3, the HTH motif's recognition helix, fits into the major groove of its corresponding DNA, it does so quite differently in the two complexes. In the λ repressor complex, for example, the N-terminal end of the recognition helix is inserted into the DNA's major groove, whereas in the homeodomain complex the DNA is shifted toward the C-terminal end of the helix, which is longer than that of the λ repressor (it extends from residues 42 to 58 in Fig. 34-103). As a consequence, the way in which the first helix of the HTH motif (helix 2; residues 28 to 37 in Fig. 34-103) contacts the DNA also differs between the two complexes.

Most homeodomain binding sites have the subsequence TAAT. The recognition helix in the X-ray structure makes base-specific hydrogen bonding contacts with this subsequence in the major groove through residues that are highly conserved in higher eukaryotic homeodomains. It therefore appears that these interactions function to align the homeodomain with the other bases that it contacts. In addition, two conserved Arg residues located in the N-terminal tail of the homeodomain make base-specific hydrogen bonding contacts with the TAAT subsequence in the minor groove of the DNA. The protein thereby grips the TAAT subsequence from two sides. Note that few other sequence-specific DNA-binding proteins contact bases in the minor groove. Finally, the homeodomain makes extensive contacts with the DNA backbone that, it is presumed, also play an important part in binding and recognition.

j. Homeodomain Genes Function Analogously in Vertebrates and *Drosophila*

Homeodomain-encoding genes have collectively become known as ***Hox genes***. In vertebrates, they are organized in four clusters of 9 to 11 genes, each located on a separate chromosome and spanning more than 100 kb. In contrast, *Drosophila*, as we saw, has a split *Hox* cluster, whereas in nematodes (roundworms), which are evolutionarily more primitive than insects, the single *Hox* cluster remains unsplit. The genes in the primordial *Hox* cluster presumably arose in some more primitive ancestral organism through a series of gene duplications, as did the four vertebrate *Hox* clusters. The genes in each vertebrate *Hox* cluster, as in *Drosophila*, are activated in the same order, left to right, as they are expressed from the anterior end of the embryo to its posterior end. Perhaps this arrangement is

necessary for the homeodomain genes to be activated in the proper order, although, at least in *Drosophila*, gap and pair-rule proteins can still act on *Hox* control regions that have been transplanted to other parts of the genome. Whatever the case, the various *Hox* clusters, as well as their component genes, almost certainly arose through a series of gene duplications and diversifications starting with a single *Hox* gene in a primitive ancestral organism.

Vertebrate *Hox* genes, like those of *Drosophila*, are expressed in specific patterns and at particular stages during embryogenesis. Most *Hox* genes are expressed at a gestational time when organogenesis prevails. That the *Hox* genes directly specify the identities and fates of embryonic cells, that is, are homeotic in character, was shown, for example by the following experiment. Mouse embryos were made transgenic for the *Hox-1.1* gene that had been placed under the control of a promoter that is active throughout the body even though *Hox-1.1* is normally expressed only below the neck. The resulting mice had severe craniofacial abnormalities such as a cleft palate and an extra vertebra and intervertebral disk at the base of the skull. Some also had an extra pair of ribs in the neck region. Thus, this *Hox* gene's "gain of function" resulted in a homeotic mutation, that is, a change in the development pattern, analogous to those observed in *Drosophila*.

Homozygotic mice resulting from the replacement of their *Hox-3.1* gene coding sequence in embryonic stem cells with that of *lacZ* are born alive but usually die within a few days. They exhibit skeletal deformities in their trunk regions in which several skeletal segments are transformed into the likenesses of more anterior segments. The pattern of β -galactosidase activity (Fig. 34-105), as colorimetrically detected through the use of X-gal (Section 5-5Ca), in both homozygotes and heterozygotes, indicates that *Hox-3.1* deletion modifies the properties but not the positions of the embryonic cells that normally express *Hox-3.1*.

k. Expression of the *Drosophila eyeless* Gene Induces the Ectopic Formation of Eyes

Mutations in the *Drosophila eyeless* (*ey*) gene, first described in 1915, result in flies whose compound eyes are reduced in size or completely absent. The expression of *ey*, which contains a homeodomain, is first detected in the embryonic nervous system and later in the embryonic primordia of the eye. In subsequent larval stages, it is expressed in the developing eye imaginal disks. Mutant forms of four other *Drosophila* genes that have similar phenotypes do not affect the expression of the *ey* gene, which indicates that *ey* acts before these other genes. These observations led to the suggestion that the *ey* gene is the master control gene for eye development.

Genetic engineering studies by Gehring have confirmed this hypothesis. Through the targeted expression of *ey* cDNA in various imaginal disk primordia of *Drosophila*, ectopic (inappropriately positioned) compound eyes were induced to form on the wings, legs, and antennae (Fig. 34-106) of various flies. Moreover, in many cases, these eyes appeared morphologically normal in that they consisted of fully differentiated ommatidia (the simple eye



Figure 34-105 Pattern of expression of the *Hox-3.1* gene in a 12.5-day postconception mouse embryo. The protein-encoding portion of the embryo's *Hox-3.1* gene was replaced by the *lacZ* gene. The regions of this transgenic embryo in which *Hox-3.1* is expressed are revealed by the blue color that develops on soaking the embryo in X-gal-containing buffer. [Courtesy of Phillippe Brûlet, Collège de France and the Pasteur Institute, Paris, France.]



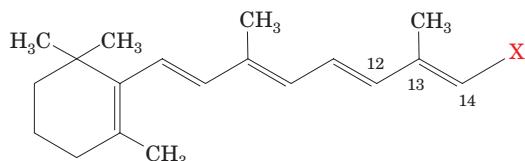
Figure 34-106 Ectopic eyes result from the targeted expression of the *Drosophila ey* gene in its imaginal disk primordia. Shown here is the cuticle of an adult *Drosophila* head in which both antennae have formed eye structures that exhibit the morphology and red pigmentation of normal eyes. Such eye structures have been similarly expressed on wings and legs. [Courtesy of Walter Gehring, University of Basel, Switzerland.]

elements that form a compound eye) with a complete set of photoreceptor cells that appear to be electrically active when illuminated (although it is unknown if the flies could see with these ectopic eyes, that is, whether these eyes made appropriate neural connections to the brain).

The mouse *Small eye* (*Sey* or *Pax-6*) gene and the human *Aniridia* gene are closely similar in sequence to the *Drosophila ey* gene and are similarly expressed during morphogenesis. Mice with mutations in one of their two *Sey* genes have underdeveloped eyes, whereas those with mutations in both *Sey* genes are eyeless. Similarly, humans that are heterozygotes for a defective *Aniridia* gene have defects in their iris, lens, cornea, and retina. Evidently, the *ey*, *Sey*, and *Aniridia* genes all function as master control genes for eye formation in their respective organisms, a surprising result considering the enormous morphological differences between insect and mammalian eyes. Thus, despite the 500 million years since the divergence of insects and mammals, their developmental control mechanisms appear to be closely related.

I. Retinoic Acid Is a Vertebrate Morphogen

Retinoic acid (RA), a derivative of vitamin A (retinol),



X = COOH: Retinoic acid (RA)

X = CH₂OH: Retinol (vitamin A)

has been found to have a graded distribution in developing chick limbs and is therefore thought to be a morphogen. The systematic administration of RA during mouse embryogenesis results in severe malformations, notably skeletal deformities that appear to arise from anterior or posterior shifts of their normal characteristics. A variety of evidence suggests that the expression of *Hox* genes mediates the positional information that RA disrupts. The *Hox* genes are differentially activated by RA according to their positions in their *Hox* clusters: Those toward the 3' end of a cluster are maximally induced by as little as $10^{-8} M$ RA, those toward the 5' end of the cluster require $10^{-5} M$ RA to do so, and those at the 5' ends are insensitive to RA. Moreover, $10^{-5} M$ RA sequentially activates the *Hox* genes from the 3' to the 5' end of a cluster, the same order as their expression patterns in developing axial systems such as the skeleton and the central nervous system.

The foregoing explains why the RA analog **13-cis-retinoic acid**, which, taken orally, has been invaluable in the treatment of severe **cystic acne**, induces birth defects if used by pregnant women. The characteristic pattern of cranial deformities in the resulting infants, whose analog is induced in mouse embryos that had been exposed to low concentrations ($2 \times 10^{-6} M$) of this drug, indicates that its presence alters the expression of *Hox* genes early

in gestation (~1 month postfertilization in humans, ~9 days in mice).

C. The Molecular Basis of Cancer

Cancer, being one of the major human health problems, has received enormous biomedical attention over the past several decades. Around 100 different types of human cancers are recognized, methods of cancer detection and treatment are highly developed, and cancer epidemiology has been extensively characterized. Nevertheless, we are far from fully understanding the biochemical basis of this collection of diseases. In Section 19-3B we discussed the general nature of cancer, its causes, and how tumor viruses cause cancer. In this section we outline how genetic alterations cause cancer.

a. Malignancies May Result from Specific Genetic Alterations

Although much of what we know concerning oncogenes stems from the study of retroviral oncogenes (Section 19-3B), few human cancers are caused by retroviruses. Nevertheless, *it seems likely that all cancers are caused by genetic alterations*. Robert Weinberg demonstrated this to be the case for mouse fibroblasts that had been transformed by a known carcinogen: Normal mouse fibroblasts in culture are transformed on transfection with DNA from the transformed cells. Moreover, these newly transformed cells, when inoculated into mice, form tumors. Similar investigations indicate that DNAs from a wide variety of malignant tumors likewise have transforming activity.

What sorts of genetic changes can give rise to cancer? Several types of changes have been observed:

1. Altered Proteins: *An oncogene, as we have seen, may give rise to a protein product with an anomalous activity relative to that of the corresponding proto-oncogene.* This may even result from a simple point mutation. For example, Weinberg, Michael Wigler, and Mariano Barbacid showed that the *ras* oncogene isolated from a human bladder **carcinoma** (a malignant tumor arising from epithelial tissue) differs from its corresponding proto-oncogene by the mutation of the Gly 12 codon (GGC) to a Val codon (GTC). The resulting amino acid change attenuates the GTPase activity of Ras protein (Section 19-3Cf) without affecting its ability to stimulate protein phosphorylation, thereby prolonging the time this G-protein remains in the “on” state. Indeed, comparison of the X-ray structures of normal human Ras and its oncogenic counterpart (G12V), both in complex with GDP, indicate that the mutation mainly alters the normal protein structure in the vicinity of its GTPase function. Most other *ras* oncogene-activating mutations also change residues close to this site. Ras, which plays a central role in MAP kinase cascades (Fig. 19-40), as might be expected, is one of the most commonly implicated proto-oncogenes in human cancers.

2. Altered Regulatory Sequences: *Malignant transformation can result from the inappropriately high expression*

of a normal cellular protein. For example, the proto-oncogene **c-fos**, which encodes the transcription factor **Fos** (which is activated by MAP kinase cascades; Fig. 19-40), differs from the retroviral oncogene **v-fos** mainly in regulatory sequences: *v-fos* has an efficient enhancer, whereas *c-fos* has a 67-nucleotide AT-rich segment in its unexpressed 3'-terminal end that, when transcribed, promotes rapid mRNA degradation (Section 31-4Av). Thus, *c-fos* can be converted to an oncogene by deleting its 3' end and adding the *v-fos* enhancer.

3. Loss of Degradation Signals: *An oncogene protein that is degraded more slowly than the corresponding normal cellular protein may cause malignant transformation through its consequent inappropriately high concentration in the cell.* For example, the transcription factor **c-Jun** (which is also activated by MAP kinase cascades; Fig. 19-40), but not **v-Jun**, is efficiently multiubiquinated and hence proteolytically degraded by the cell (Section 32-6Bb). This is because *v-Jun* lacks a 27-residue segment present in *c-Jun* that mutagenesis experiments indicate is essential for the efficient ubiquitination of *c-Jun* even though this segment does not contain the protein's principal ubiquitin attachment sites.

4. Chromosomal Rearrangements: *An oncogene may be inappropriately transcribed when brought under the control of a foreign regulatory sequence through chromosomal rearrangement (a position effect).* For example, Carlo Croce found that the human cancer **Burkitt's lymphoma** (a lymphoma is an immune system cell malignancy) is characterized by an exchange of chromosomal segments in which the proto-oncogene **c-myc** is translocated from its normal position at one end of chromosome 8 to the end of chromosome 14 adjacent to certain immunoglobulin genes. The misplaced *c-myc* gene is thereby brought under the transcriptional control of the highly active (in immune system cells) immunoglobulin regulatory sequences. The consequent overproduction of the normal *c-myc* gene product **Myc** (a transcription factor that is also activated by MAP kinase cascades and whose transient increase is normally correlated with the onset of cell division), or alternatively, its production at the wrong time in the cell cycle, is apparently a major factor in cell transformation.

5. Gene Amplification: *Oncogene overexpression can also occur when the oncogene is replicated multiple times, either as sequentially repeated chromosomal copies or as extrachromosomal particles.* The amplification of the *c-myc* gene, for example, has been observed in several types of human cancers. Gene amplification is usually an unstable genetic condition (Section 34-2De) that can only be maintained under strong selective pressure such as that conferred by cytotoxic drugs. It is not known how oncogene amplification is stably maintained.

6. Viral Insertion into a Chromosome: *Inappropriate oncogene expression may result from the insertion of a viral genome into a cellular chromosome such that the proto-oncogene is brought under the transcriptional control of a viral regulatory sequence.* For instance, **avian leukosis virus**,

a retrovirus that lacks an oncogene but that nevertheless induces lymphomas in chickens, has a chromosomal insertion site near *c-myc*. Some DNA tumor viruses also transform cells in this manner.

7. Inappropriate Inactivation or Activation of Chromatin Modification Enzymes: Heterozygotes for a defective *CPB* gene, whose gene product activates the PCAF HAT complex (Section 34-3Bbb), have **Rubinstein-Tabi syndrome**, a condition that predisposes to cancer. In a related example, the **retinoic acid receptor (RAR)**, which is important for myeloid (blood-forming) tissue differentiation, helps recruit HDAC complexes such as N-CoR and SMRT (Section 34-3Bdd) to **retinoic acid response elements (RAREs)**, but on binding ligand, releases them. However, in **promyelocytic leukemia**, a chromosomal translocation yields a defective RAR that binds to RAREs and recruits HDACs but is unresponsive to the presence of retinoids.

8. Loss or Inactivation of Tumor Suppressor Genes: The high incidence of particular cancers in certain families suggests that there are genetic predispositions toward these diseases. A particularly clear-cut example of this phenomenon occurs in **retinoblastoma**, a cancer of the developing retina that therefore afflicts only infants and young children. The offspring of surviving retinoblastoma victims also have a high incidence of this disease, as well as several other types of malignancies. In fact, retinoblastoma is associated with the inheritance of a copy of chromosome 13 from which a particular segment has been deleted. Retinoblastoma develops, as Alfred Knudson first explained, through a somatic mutation in a **retinoblast** (a retinal precursor cell) that alters the same segment of the second, heretofore normal copy of chromosome 13. This is because *the affected chromosomal segment contains a gene, the Rb gene, which specifies a factor that restrains uninhibited cell proliferation; that is, the Rb gene product, pRb, is a tumor suppressor (alternatively, an anti-oncogene protein)*. Indeed, the *Rb* gene is mutated in >70% of human cancers, which encompass a great variety of tumor types. The structure and function of pRb is further discussed below.

Several other tumor suppressors have been characterized including **p53**, which is encoded by another commonly altered gene in human cancers (~50% of cancers contain a mutation in p53 and many other oncogenic mutations occur in genes encoding proteins that directly or indirectly interact with p53; the structure and function of p53 is further discussed in Section 34-4D); **neurofibromatosis type 1 (NF1) protein**, whose defect causes benign tumors of the peripheral nerves, such as those of the famous “Elephant Man” of Victorian England, that occasionally become malignant; **BRCA1**, which forms a portion of a ubiquitin-protein ligase (E3) and whose defect predisposes to breast and ovarian cancers (Section 30-6Ag); **BRCA2**, a DNA-binding protein that participates in the repair of double-strand breaks and whose defect also predisposes to breast and ovarian cancers; and **PTEN**, an inositol polyphosphate 3-phosphatase, whose structure and function are discussed in Section 19-4Ec.

Mutations altering normal gene products, causing chromosomal rearrangements and deletions, and perhaps gene amplification can all result from the actions of carcinogens on cellular DNA. Thus, normal cells bear the seeds of their own cancers. To date, ~350 viral and cellular oncogenes and tumor suppressors have been identified. Nevertheless, at the cellular level, a malignant transformation is an extremely rare event. This is because it requires several oncogenic mutations, all low probability events, to cause malignant transformation (Section 19-3Ba) and because cells have highly effective defense mechanisms that guard against cancer.

b. pRb Functions by Binding to Certain Transcription Factors

pRb is a 928-residue, monomeric, DNA-binding protein that is localized in the nucleus of normal retinal cells but is absent in retinoblastoma cells. It is a phosphoprotein that is phosphorylated in a cell cycle-dependent manner, as is discussed in Section 34-4Di. Hypophosphorylated forms of pRb form complexes with certain transcription factors, including **E2F**, which regulates the expression of several cellular and viral genes. E2F was first identified as a cellular factor involved in the regulation of the adenovirus early **E2** gene by the adenovirus oncogene product **E1A**, although further investigation revealed that the adenovirus **E4** gene product also participates in this process. E1A protein, which does not bind DNA, promotes the dissociation of pRb from E2F by complexing pRb. It thereby frees E2F protein to combine with E4 protein on the adenovirus **E2 promoter** so as to stimulate the transcription of the **E2** gene. These observations suggest that *the interaction of pRb with E2F and other transcription factors to which it binds plays an important role in the suppression of cellular proliferation* and that the dissociation of this complex is, at least in part, the means by which E1A inactivates pRb function. Thus, *an additional way that oncogenes can cause cancer is by inactivating the products of normal cellular tumor suppressor genes*. We continue our discussion of pRb below.

D. The Regulation of the Cell Cycle

The cell cycle, as we discussed in Section 30-4A, is the sequence of major events that occur during the life of a eukaryotic cell. It is divided into four phases (Fig. 30-40): M phase, during which mitosis and cell division occur; G₁ phase, the main period of cell growth; S phase, during which DNA is synthesized; and G₂ phase, the interval in which the cell prepares for the next M phase. The progression through the cell cycle is regulated by external as well as internal signals. Thus, yeast have a regulatory point known as **START** that occurs late in G₁ beyond which they are committed to enter S phase, that is, replicate their DNA. However, if there are insufficient nutrients available or if the cell has not reached some minimum size, the cell cycle is arrested at START and the cell assumes a resting state until these criteria have been met. Animal cells have a similar decision point in G₁ named the **restriction point**,

but it responds mainly to the extracellular presence of the appropriate **mitogens**, protein growth factors that signal the cell to proliferate. The cell cycle also has a series of **checkpoints** that monitor its progress and/or the health of the cell and arrest the cell cycle if certain conditions have not been satisfied. Thus, G_2 has a checkpoint that prevents the initiation of M until all of the cell's DNA has been replicated, thereby ensuring that both daughter cells will receive a full complement of DNA. Similarly, a checkpoint in M prevents mitosis until all chromosomes have properly attached to the mitotic spindle. Checkpoints in G_1 and S, as well as that in G_2 , also arrest the cell cycle in response to damaged DNA so as to give the cell time to repair the damage (Section 30-5). In the cells of multicellular organisms, if after a time the checkpoint conditions have not been satisfied, the cell may be directed to commit suicide, a process named **apoptosis** (Section 34-4E), thereby preventing the proliferation of a genetically irreparably damaged and hence dangerous (e.g., cancerous) cell. However, single-celled eukaryotes such as yeast lack such a mechanism, presumably because, in their case, the survival of a genetically damaged cell is preferable in a Darwinian sense to its death.

a. The Activation of Cdk1 Triggers Mitosis

What are the molecular events that drive and coordinate the cell cycle? The first clues to this process came from studies of sea urchin embryos by Tim Hunt, which revealed that a class of proteins named **cyclins** accumulate steadily throughout the cell cycle and then abruptly disappear just before the anaphase portion of mitosis (Fig. 1-19). Homologs of these proteins have since been discovered in all eukaryotic cells examined. Indeed, humans encode at least 29 different cyclins (mostly named cyclin A, cyclin B, etc.), many but not all of which participate in cell cycle control or appear and disappear at specific times during the cell cycle. Each phase of the cell cycle has one or more associated cyclins. The cyclins form a diverse protein family that are 30 to 50% similar over an \sim 100-residue segment known as the **cyclin box**.

Further indications as to the way the cell cycle is controlled came from experiments in which human cells at different stages of the cell cycle were fused to yield a single cell with two nuclei. When a cell in G_1 phase was fused with one in S phase, the G_1 nucleus immediately entered S phase, whereas the S nucleus continued replicating its DNA. However, when S-phase and G_2 -phase cells were fused, the S nucleus continued replicating its DNA and the G_2 nucleus remained in G_2 . Similarly, when G_1 and G_2 cells were fused, the G_1 nucleus entered S phase according to its own schedule and the G_2 nucleus remained in G_2 . Evidently, S-phase cells contain a diffusible activator of DNA replication, only G_1 cells can initiate DNA replication, and cells that have transited S phase are unable to re-replicate their DNA until they have passed through M phase.

Many of the proteins that participate in cell cycle regulation were identified in the 1970s by Lee Hartwell through his study of temperature-sensitive mutants in *S. cerevisiae* (budding yeast) that were defective in cell cycle progres-

sion and by similar studies in *Schizosaccharomyces pombe* (fission yeast) by Paul Nurse. In what is perhaps the best characterized portion of the cell cycle, that inducing M phase, **cyclin B** combines with **Cdc2** (cell division cycle 2), whose sequence clearly indicates that it is a member of the Ser/Thr protein kinase family (Section 18-3Cb) and which is highly conserved from yeasts to humans. *It is Cdc2 that is the central cell cycle regulator in species ranging from yeasts to humans. It does so by phosphorylating a variety of nuclear proteins, among them histone H1, several oncogene proteins (see below), and proteins involved in nuclear disassembly, cytoskeletal rearrangement, spindle assembly, chromosome condensation, and Golgi fragmentation. This initiates a cascade of cellular events that culminates in mitosis.*

The binding of cyclin B to Cdc2 forms an activated complex that is alternatively called **cyclin-dependent protein kinase 1 (Cdk1)** and **maturation promoting factor (MPF)**. This, however, is by no means the entire activation story. Cdc2 is a phosphoprotein that can be phosphorylated on Tyr 15, Thr 161, and in higher eukaryotes, Thr 14. Cdk1 is active only when both Thr 14 and Tyr 15, which occupy the region of its ATP-binding site, are dephosphorylated and when Thr 161 is phosphorylated. Moreover, the phosphorylation of Tyr 15 requires that cyclin B be present. Thus, mitosis is triggered through the following series of events (Fig. 34-107):

1. The cell enters G_1 with cyclin B absent and with Cdc2, which is present at a constant level throughout the cell cycle, dephosphorylated. **Cdk-activating kinase (CAK)**, which is a complex of **cyclin H** and **Cdk7** then phosphorylates Cdc2's Thr 161. Curiously, CAK is also a component of the general transcription factor TFIH (Section 34-3Bb), where its function is unknown.

2. In S phase, newly synthesized cyclin B binds to Cdc2, whereupon its Tyr 15 is phosphorylated by **Wee1** (so named because, in fission yeast, its inactivation causes cells to enter mitosis prematurely and hence at an unusually small, that is wee, size) and Thr 14 is phosphorylated by the Wee1 homolog **Myt1**, which can also phosphorylate Tyr 15. The resulting triply phosphorylated cyclin B-Cdc2 complex is enzymatically inactive because Thr 14 and Tyr 15 prevent Cdc2 from binding ATP. Thus, the entire system appears designed to maintain Cdc2 in an inactive state while cyclin B gradually accumulates during S phase.

3. At the cell cycle's G_2/M boundary, Cdc2's Thr 14 and Tyr 15 are rapidly and specifically dephosphorylated by **Cdc25C**, a dual-specificity protein tyrosine phosphatase (PTP; Section 19-3Fa) that thereby activates Cdk1, which in turn triggers mitosis (M phase). This process is initiated by the activating phosphorylation of Cdc25C by **plk1** (for *polo-like kinase 1*; so named because it is a homolog of *Drosophila polo kinase*). The resulting activated Cdk1 also activates Cdc25C so that, through the intermediacy of Cdc25C, Cdk1 activates itself in a rapid burst. Moreover, Cdk1 inhibits its own inactivation by phosphorylating and thereby inactivating Wee1 and Myt1.

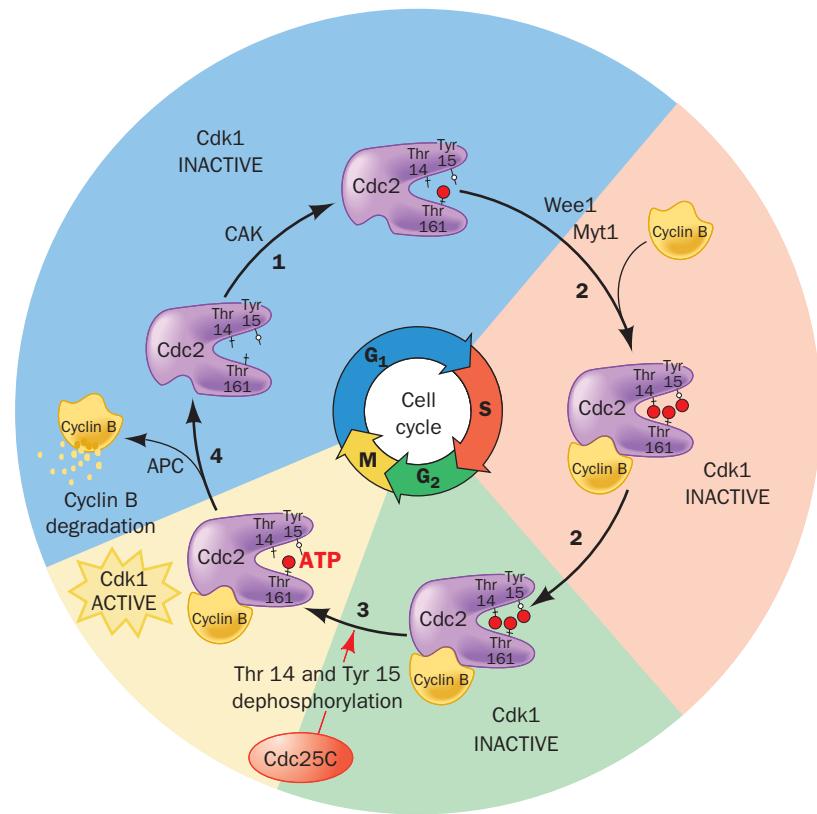


Figure 34-107 Regulation of Cdk1 in the animal cell cycle. Details are described in the text.

[After Norbury, C. and Nurse, P., *Annu. Rev. Biochem.* **61**, 451 (1992).]

4. Cyclin B is quickly proteolyzed by the proteasome in a ubiquitin-mediated pathway whose E3 is the multisubunit complex **anaphase-promoting complex (APC)** (Section 32-6Be), followed by the rapid dephosphorylation of Cdc2 Thr 161. This inactivates Cdk1, thereby returning the now divided cell to G₁. APC, which is inactive during S and G₂ phases, is activated, at least in part, by Cdk1, which thereby brings about the destruction of its own cyclin B component.

Other combinations of Cdks and cyclins that similarly mediate specific portions of the cell cycle include the following (Fig. 34-108): **Cdk4** and its close isoform **Cdk6**, which in complex with D-type cyclins (**cyclins D1, D2, and D3**) drive events in G₁; **Cdk2** and **cyclin E**, which are required for progression through the G₁/S boundary and for the initiation of DNA synthesis; and Cdk2 and **cyclin A**, which control passage through S phase. Thus, *Cdk-cyclin complexes form the engines that drive the various processes of the cell cycle as well as the clocks that time them*.

b. The X-Ray Structure of Cdk2 Resembles That of PKA

Cdk2, which is closely similar to Cdc2, is activated by the binding of a cyclin A, which is closely similar to cyclin B, followed by the CAK-catalyzed phosphorylation of Thr 160. Cdk2 is also negatively regulated by the phosphorylation of Tyr 15 and, to a lesser extent, the adjacent Thr 14.

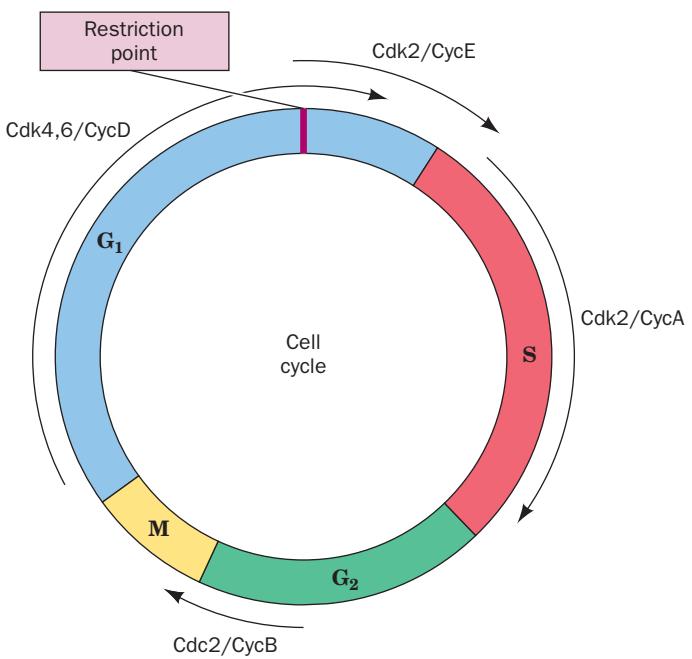


Figure 34-108 Complexes of cyclin-dependent kinases (Cdks) and cyclins (Cycs) that mediate passage through specific segments of the cell cycle.

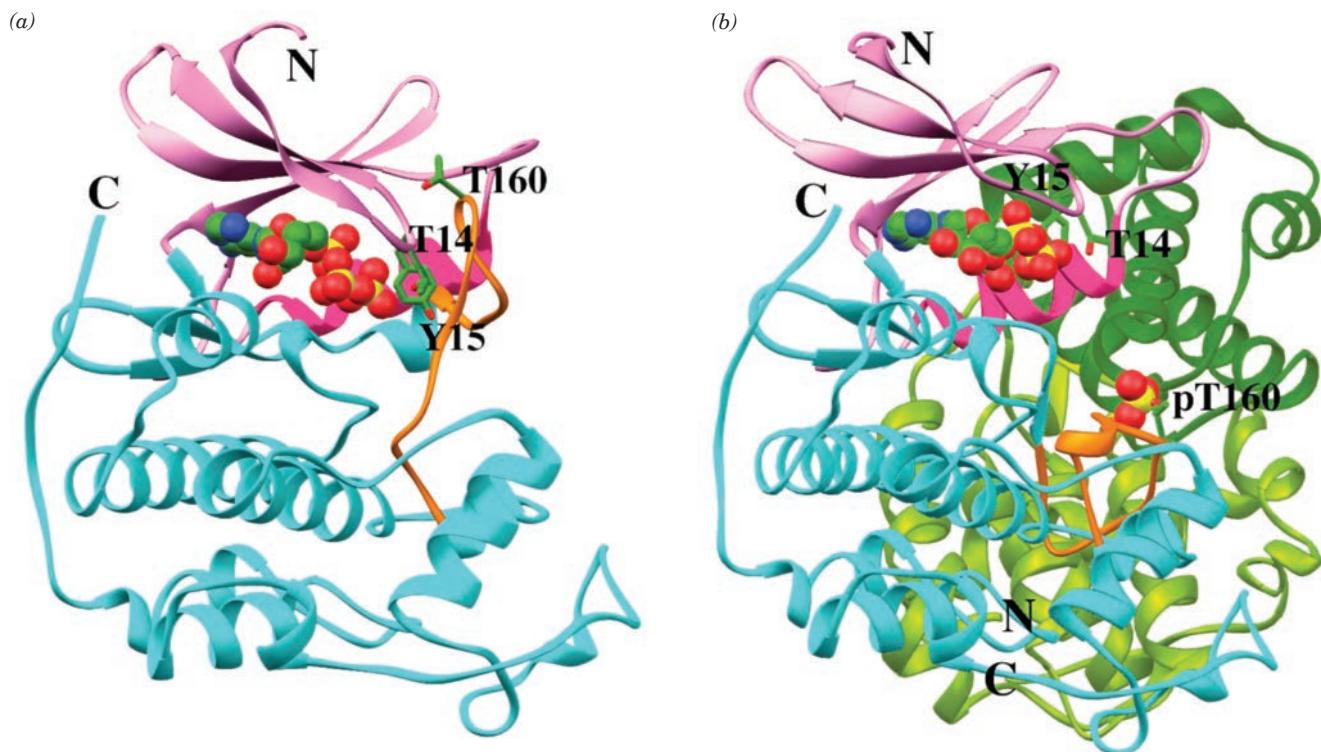


Figure 34-109 X-ray structure of human cyclin-dependent kinase 2 (Cdk2). (a) Cdk2 in complex with ATP. The protein is shown in the “standard” protein kinase orientation with its N-terminal lobe pink, its C-terminal lobe cyan, its PSTAIRE helix (residues 45–56) magenta, and its T loop (residues 152–170) orange. The ATP is shown in space-filling form and the phosphorylatable side chains of Thr 14, Tyr 15, and Thr 160 are shown in stick form, all colored according to atom type (C green, N blue, O red, and P yellow). Compare this structure with that of protein kinase A (PKA; Fig. 18-15). [Based on an X-ray structure by Sung-Hou Kim, University of California at Berkeley. PDBid 1HCK.] (b) The complex of T160-phosphorylated Cdk2 with cyclin A and ATP. The Cdk2 and ATP are represented as in

Part *a* and viewed similarly. The cyclin A is colored yellow-green with its cyclin box (residues 206–306) dark green. The Cdk2 phosphoThr 160 phosphoryl group is drawn in space-filling form. Note how the binding of cyclin A together with the phosphorylation of Cdc2 Thr 160 has caused a major structural reorganization of the T loop together with significant conformational adjustments of the Cdk2 N-terminal lobe, including its PSTAIRE helix. Also note the different conformations of the ATP triphosphate group in the two structures. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1JST.]  See the Interactive Exercises.

The X-ray structure of human Cdk2 in complex with ATP (Fig. 34-109*a*), determined by Sung-Hou Kim, indicates that this 298-residue monomeric Ser/Thr protein kinase closely resembles the catalytic subunit of protein kinase A (PKA; Section 18-3*Ca*), a protein whose sequence is 24% identical to that of Cdk2. However, there are functionally significant structural differences between these two kinases:

1. The relative arrangement of ATP’s β - and γ -phosphate groups in Cdk2 is likely to greatly reduce the reactivity of the γ -phosphate relative to that in the PKA-ATP complex (stereochemical control), thereby rationalizing, in part, why Cdk2 alone is catalytically inactive whereas the catalytic subunit of PKA alone is catalytically active.
2. Access to the γ -phosphate of Cdk2’s bound ATP by its protein substrates appears to be blocked by a 19-residue protein loop (residues 152–170) that has been named the “T loop” because it contains Thr 160.

The X-ray structure also explains why the phosphorylation of Thr 14 inactivates Cdk2: The hydroxyl group of this side chain is close to the ATP’s γ -phosphate so that phosphorylation of Thr 14 is likely to disrupt the conformation of the ATP’s phosphate groups. It is unclear, however, how the phosphorylation of Tyr 15 affects Cdk2 activity.

c. Cyclin A Binding and Thr 160 Phosphorylation Conformationally Reorganize Cdk2

The X-ray structure of human Cdk2 that is phosphorylated at its Thr 160 in complex with ATP and the C-terminal portion of human cyclin A (its residues 173–432), determined by Pavletich, indicates that cyclin A binds to Cdk2’s “back” side (Fig. 34-109*b*). There it interacts with both lobes of the Cdk2 to form an extensive and continuous protein–protein interface. Cyclin A consists mainly of a bundle of 12 α helices (and no β sheets) in which its cyclin box forms helices 2 to 6. Interestingly, helices 7 to 11 form a bundle that is nearly superimposable on that of the cyclin

box even though these two motifs exhibit little sequence similarity. Comparison of the X-ray structures of free cyclin A with that in its complex with Cdk2 indicates that cyclin A does not undergo significant conformational change on binding Cdk2. In contrast, cyclin A binding causes Cdk2 to undergo significant conformational shifts in the region around its catalytic cleft. In particular, the N-terminal α helix of Cdk2, which contains the PSTAIRE sequence motif characteristic of the Cdk family, rotates about its axis by 90° and moves several angstroms into the catalytic cleft relative to its position in free Cdk2, where it contacts the cyclin box segment of cyclin A. This movement brings Glu 51 (the E in PSTAIRE) from its solvent-exposed position outside the catalytic cleft of free Cdk2 to a position inside the catalytic cleft, where it forms a salt bridge with Lys 33, which in free Cdk2 instead forms a salt bridge to Asp 145. These three side chains (Lys 33, Glu 51, and Asp 145), which are conserved in all eukaryotic protein kinases, participate in ATP phosphate coordination and Mg²⁺ ion coordination. Their conformational reorientation on cyclin A binding apparently places them in a catalytically active arrangement.

The binding of cyclin A also induces Cdk2's T loop to undergo extensive conformational reorganization involving positional shifts of up to 21 Å, such that the T loop, which now also contacts the cyclin box, adopts a backbone conformation which closely resembles that of the analogous region of the catalytically active PKA. These movements greatly increase the access of a protein substrate to the ATP bound in the catalytic cleft, which has assumed a more reactive conformation than that in free Cdk2. The phosphate group on Thr 160 fits snugly into a positively charged pocket composed of three Arg residues that forms, in part, on cyclin A binding. Indeed, comparison of this structure with that in which Thr 160 is unphosphorylated indicates that Thr 160 phosphorylation induces activating conformational changes in the catalytic cleft of the cyclin A–Cdk2 complex as well as contributing to the reorganization of the T loop.

d. Cdk Inhibitors Function to Arrest the Cell Cycle

In addition to their control by phosphorylation/dephosphorylation and by the binding of the appropriate cyclin, Cdk activities are regulated by **cyclin-dependent kinase inhibitors (CKIs)**, which induce cell cycle arrest in response to such antiproliferative signals as contact with other cells, DNA damage, terminal differentiation, and senescence (in which cell cycle arrest is permanent). The known CKIs have been grouped, according to their sequence and functional similarities, into two families: (1) the **Kip/Cip family** (kinase interacting protein/cytokine-inducible protein), whose members inhibit most Cdk–cyclin complexes (but not Cdk4/6–cyclin D) and can bind to isolated Cdks or cyclins, although with lower affinity than for Cdk–cyclin complexes; and (2) the **INK4 family** (inhibitor of Cdk4 and Cdk6), whose members specifically inhibit Cdk4 and Cdk6 (which, together with cyclin D, mediate the cell's progression through G₁; Fig. 34-108) and can bind to either the isolated Cdk or its complex with cyclin D. The importance of

CKIs is indicated by their frequent alterations in cancer. For example, **p16^{INK4a}** is mutated in about one-third of all human cancers, **p21** (also called **p21^{Cip1}**) arrests the cell cycle on behalf of p53 (see below), and **p27^{Kip1}** (also known as **p27^{Cip2}**) may be degraded in several types of cancers and its low levels are correlated with a poor clinical prognosis. Moreover, certain herpes viruses, including **Kaposi's sarcoma-associated herpes virus**, express an oncogenic D-type cyclin known as **K-cyclin** that binds to and thereby activates Cdk4/6, thus contributing to the deregulation of the cell cycle.

The members of the Kip/Cip family have homologous N-terminal ~65-residue segments that are necessary and sufficient to bind to and inhibit Cdk–cyclin complexes, whereas their C-terminal segments are divergent in length, sequence, and function. **p27^{Kip1}** is an inhibitor of Cdk2 that occurs in cells that have been treated with the antimitotic protein named **transforming growth factor β (TGF β)**. The X-ray structure of the N-terminal inhibitory domain (residues 25–93) of the 198-residue human p27^{Kip1} bound to the human cyclin A–Cdk2 complex in which Cdk2 is phosphorylated at Thr 160 was determined by Pavletich (Fig. 34-110). The p27 inhibitory domain is draped across both Cdk2 and cyclin A, where it assumes an extended conformation that does not have a hydrophobic core of its own and whose secondary structural elements do not interact with each other. p27's N-terminal end interacts with cyclin A, whose conformation is essentially unaffected by this interaction. In contrast, the binding of the p27 inhibitory domain's C-terminal segment to Cdk2 causes extensive

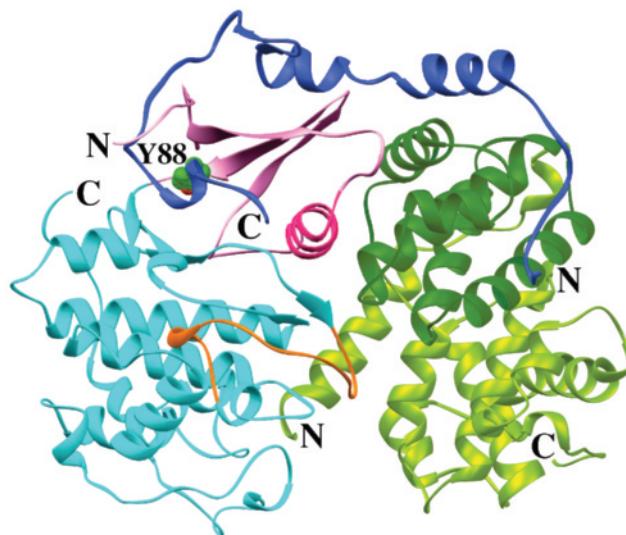


Figure 34-110 X-ray structure of the Cdk2–Cyclin A–p27^{Kip1} ternary complex. The Cdk2 and cyclin A are colored as in Fig. 34-109 and p27^{Kip1} is blue. The complex is viewed along the Cdk2 PSTAIRE helix (approximately from the right of Fig. 34-109). The p27^{Kip1} Tyr 88 side chain, which is drawn in space-filling form with C green and O red, occupies the binding site for the adenine moiety of ATP. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1JSU.]

conformational changes in Cdk2 that appear likely to destabilize its binding of ATP. More importantly, this C-terminal segment extends into Cdk2's active site cleft, where its conserved Tyr 88 side chain mimics the binding of ATP's adenine moiety in both its position and the contacts it makes to the active site groups, thereby eliminating any possibility of ATP binding. Last, p27's N-terminus occupies the peptide-binding groove on cyclin A's conserved cyclin box that is probably a docking site for a number of the cyclin A–Cdk2 complex's tight-binding substrates, thereby reducing the ability of substrate binding to reverse the effects of p27-induced conformational changes.

INK4 proteins can bind to monomeric Cdk4/6 so as to prevent its association with cyclin D or bind to a Cdk4/6–cyclin D binary complex to form a catalytically inactive ternary complex. *In vivo*, INK4–Cdk4/6 binary complexes are more abundant than INK4–Cdk4/6–cyclin D ternary complexes, which suggests that INK4 binding increases the rate of cyclin dissociation from the ternary complex. The X-ray structure of the ternary complex of human p18^{INK4c}, human Cdk6, and K-cyclin (Fig. 34-111), determined by Pavletich, reveals that p18^{INK4c} binds to the Cdk6–K-cyclin complex in an entirely different manner from the way p27^{Kip1} binds to the Cdk2–cyclin A complex (Fig. 34-110). The 160-residue p18^{INK4c}, which consists of five ankyrin repeats (which participate in protein–protein interactions; Section 12-3Db), binds to the 301-residue Cdk6 in the region of its ATP-binding site, where it interacts with Cdk6's N- and C-terminal lobes via the second and third ankyrin repeats of p18^{INK4c}. This rotates the N-terminal lobe by 13° with respect to the C-terminal lobe

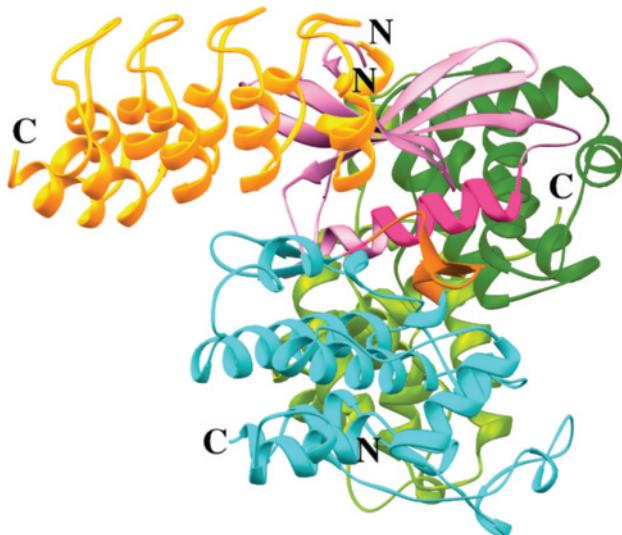


Figure 34-111 X-ray structure of the Cdk6–K-cyclin–p18^{INK4c} ternary complex. The Cdk6 and K-cyclin are colored as in the homologous Cdk2 and cyclin A in Figs. 34-109 and 34-110 and the p18^{INK4c} is gold. The structure is viewed with the Cdk6 in the “standard” protein kinase orientation as in Fig. 34-109. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1G3N.]

relative to their orientations in the Cdk2–cyclin A complex (Fig. 34-109b) and thereby distorts the Cdk6 ATP-binding site and misaligns its catalytic residues. Moreover, p18^{INK4c} binding distorts Cdk6's cyclin-binding site such that their interface is reduced in area by ~30%: The cyclin box of the 253-residue K-cyclin binds to the N-terminal lobe of Cdk6 centered on its PSTAIRE sequence motif (which has the sequence PLSTIRE) but, unlike in the structure of the Cdk2–cyclin A complex (Figs. 34-109b and 34-110), there are no significant contacts between the Cdk6 C-terminal lobe and K-cyclin. Apparently, INK4 binding reduces the stability of the Cdk–cyclin interface.

e. Cell Cycle Arrest at the G₂ Checkpoint Is Mediated by a Phosphorylation/Dephosphorylation Cascade

How does failure to satisfy a checkpoint cause cell cycle arrest? For the G₂ checkpoint, the process is initiated, as diagrammed in Fig. 34-112, by at least six poorly characterized **sensor proteins** that have been identified through mutations in the genes *rad1*, *rad3*, *rad9*, *rad17*, *rad26*, and *hus1* that are defective in repair and replication checkpoints. The sensor proteins bind to damaged or unreplicated DNA, which causes them to activate two related large (~3000 residue) protein kinases known as **ATM** and **ATR** [ATM for *ataxia telangiectasia mutated* (**ataxia telangiectasia** is a rare genetic disease characterized by a progressive loss of motor control, growth retardation, immune system deficiencies, premature aging, and a greatly increased risk of cancer); ATR for ATM and *Rad3*-related (**Rad3** is the ATR homolog in *S. pombe*)]. Activated ATM and ATR respectively phosphorylate and thereby activate **Chk2** and **Chk1** (Chk for *checkpoint kinase*). These latter protein kinases phosphorylate Cdc25C which, it will be recalled, functions to activate Cdk1 by dephosphorylating Cdc2's Thr 14 and Tyr 15. This phosphorylation, which occurs

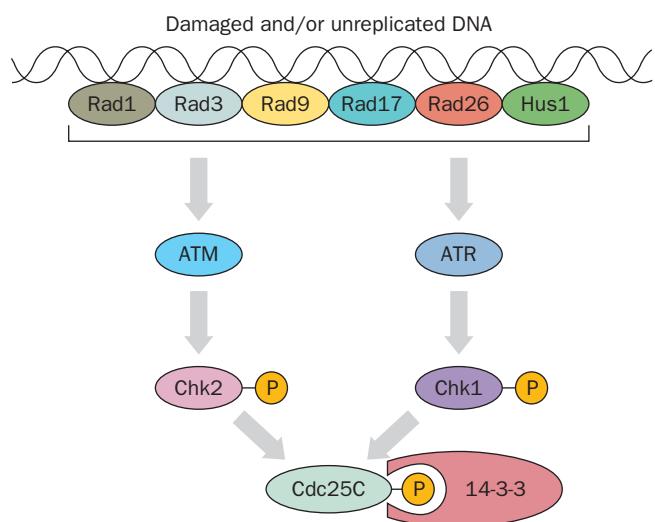


Figure 34-112 The G₂ checkpoint phosphorylation/dephosphorylation cascade that results in cell cycle arrest. Details are given in the text.

on Ser 216 of Cdc25C, does not inactivate this protein phosphatase. Rather, it provides a binding site for members of the **14-3-3 family** of adaptor proteins (which bind certain phosphorylated motifs in a wide variety of phosphorylated proteins; the name “14-3-3” is based on these proteins’ column fractionation and electrophoretic mobility properties), and the resulting complex is sequestered in the cytoplasm out of contact with the Cdk1 in the nucleus. Since Cdk1 is the protein kinase that activates mitosis, the cell remains in G_2 until its DNA is repaired and/or fully replicated.

f. p53 Is a Transcriptional Activator That Arrests the Cell Cycle in G_2

The idea that p53 is a tumor suppressor first arose from the discovery that germ line mutations in the *p53* gene often occur in individuals with the rare inherited condition known as **Li-Fraumeni syndrome** that renders them highly susceptible to a variety of malignant tumors, particularly breast cancer, which they often develop before their 30th birthdays. That p53 is indeed a tumor suppressor has been clearly demonstrated in mice in which the *p53* gene has been inactivated. These knockout mice (Section 5-5H) appear to be developmentally normal but spontaneously develop a variety of cancers by the age of 6 months. Indeed, p53 functions as a “molecular policeman” in monitoring genome integrity: On the detection of DNA damage, p53 arrests the cell cycle until the damage is repaired, or failing that, induces senescence (permanent cell cycle arrest in G_1 , which prevents cells with precancerous lesions, which are plentiful, from undergoing malignant progression to full blown cancer) or apoptosis. In doing so, *p53 directly or indirectly affects the expression of thousands of genes*.

Despite the central role of p53 in preventing tumor formation, the way it does so has only gradually come to light. This tumor suppressor is specifically bound in humans by the homolog of the mouse **Mdm2 protein**. The *mdm2* gene is the dominant transforming oncogene present on mouse double minute chromosomes (amplified extrachromosomal segments of DNA; Section 34-2Ec). Mdm2 protein is a ubiquitin-protein ligase (E3) that specifically ubiquitinates p53, thereby marking it for proteolytic degradation by the proteasome. Consequently, the amplification of the *mdm2* locus, which occurs in >35% of human **sarcomas** (none of which have a mutated *p53* gene; sarcomas are malignancies of connective tissues such as muscle, tendon, and bone), results in an increased rate of degradation of p53, thereby predisposing cells to malignant transformation. Similarly, as we have seen (Section 32-6Be), E6 protein from human papillomavirus, which causes the great majority of cervical cancers, functions to ubiquitinate p53. Certain DNA tumor virus oncoproteins, such as SV40 **large T antigen** and adenovirus **E1B protein**, inactivate p53 by specifically binding to it. Thus, *an additional way that oncogenes can cause cancer is by inactivating normal tumor suppressors*.

p53 protein is an efficient transcriptional activator. Indeed, all point mutated forms of p53 that are implicated in cancer have lost their sequence-specific DNA-binding properties. But then, how does p53 function as a tumor sup-

pressor? A clue to this riddle came from the observation that the treatment of cells with DNA-damaging ionizing radiation induces the accumulation of normal p53. This led to the discovery that both ATM and Chk2 phosphorylate p53, which prevents its binding by Mdm2 and hence increases the otherwise low level of this protein in the nucleus (the half-life of p53 increases from minutes to hours). *Although p53 does not initiate cell cycle arrest in G_2 , its presence is required to prolong this process. It does so by activating the transcription of the gene encoding the CKI p21, which binds to several Cdk-cyclin complexes so as to inhibit both the G_1/S and G_2/M transitions.* p21 also binds to PCNA, the homotrimeric sliding clamp in DNA replication (Section 30-4Ba), so as to prevent its participation in DNA replication but not in DNA repair. Thus, p21 has a dual role in cell cycle arrest in that it both blocks cell cycle progression and inhibits DNA replication in S-phase cells.

p53 also induces the transcription of the gene encoding the 14-3-3 family member **14-3-3 σ** , which binds to Cdk1, thereby confining it to the cytoplasm. Moreover, the 14-3-3 σ -Cdk1 complex binds the protein kinase Wee1 (which, as discussed above, inactivates Cdk1 by phosphorylating its Cdc2 component at its Tyr 15), thereby ensuring that Cdk1 remains in its inactive state. Thus, the disruption of the gene encoding 14-3-3 σ is fatal for cells that sustain DNA damage. Chk2 also phosphorylates Wee1, which inhibits its proteasomal degradation. Consequently, germ line mutations in the *chk2* gene are also associated with Li-Fraumeni syndrome. Excessive levels of p53 are toxic, which explains why loss of the *mdm2* gene in mice is lethal unless the *p53* gene is also knocked out. In the absence of p53 activation, cells control the level of p53 through a feedback loop in which p53 stimulates the transcription of the *mdm2* gene.

Cells that are irreparably damaged are induced by p53 to commit suicide via apoptosis (Section 34-4E), thereby preventing the proliferation of potentially cancerous cells. p53 does so by transactivating the expression of several of the proteins that participate in apoptosis (Section 34-4E) and repressing the expression of others that inhibit this process.

In addition to the foregoing, p53 represses the expression of numerous genes. One mechanism through which it does so is by directly or indirectly stimulating the transcription of certain microRNAs (miRNAs), which in turn repress the expression of a variety of pro-proliferative proteins (Section 32-4At). The most conspicuous of these miRNAs is **miR-34a**, which is implicated in inducing senescence and facilitating apoptosis.

A second repressive mechanism, which has only recently come to light, involves a **large intergenic noncoding RNA (lncRNA)** known as **lncRNA-p21** (so named because its gene neighbors that of p21, although the significance of this, if any, is unknown). LncRNAs are large (>5 kb), evolutionarily conserved RNAs that, like mRNAs, are transcribed by RNAP II, 5'-capped, and polyadenylated, but do not encode proteins. *Xist* and *Tsix* RNAs (Section 34-3Aa) are examples of lncRNAs that we have encountered. ChIP-seq-based methods (Section 34-3By) have identified over 3000 lncRNAs in the mammalian genome,

many of which appear to participate in transcriptional control. The transcription of lincRNA-p21, as John Rinn discovered, is induced by p53. LincRNA-p21 forms a complex with **heterogeneous nuclear ribonucleoprotein K (hnRNP-K; hnRNPs are proteins that bind hnRNAs)** that represses the transcription of hundreds of genes, many of which are pro-proliferative, although how this complex does so is poorly understood.

g. The X-Ray Structure of p53 Explains Its Oncogenic Mutations

p53 is a homotetramer of 393-residue subunits. Each subunit contains four domains: an N-terminal transactivation domain (residues 1–44), a sequence-specific, DNA-binding core domain (residues 99–289, which binds two half-site decamers, each with the consensus palindromic sequence $\text{RRRC}_{\text{TA}}^{\text{AT}}\text{GYYY}$, that are separated by 0–20 nt, with 0 being the most common, and with a p53 dimer binding to each such decamer), a tetramerization domain (residues 317–353), and a nonspecific DNA-binding domain (residues 353–390, which binds a wide variety of DNAs including short single strands, irradiated DNA, Holliday junctions, and insertions/deletions). Although the entire protein has so far resisted crystallization, Pavletich has determined the X-ray structure of the DNA-binding core (residues 102–313) in complex with a 21-bp DNA segment containing its 5-bp target sequence (AGACT). *The vast majority of the nearly 30,000 p53 mutations that have yet been reported in human tumors occur in this core.*

The structure of the p53 DNA-binding core domain (Fig. 34-113a) contains a sandwich of two antiparallel β pleated sheets, one with four strands and the other with five, and a loop–sheet–helix motif that packs against one edge of the β sandwich. This edge of the β sandwich also contains two large loops running between the two β sheets that are held together through their tetrahedral coordination of a Zn^{2+} ion via one His and three Cys side chains.

The p53 DNA binding domain does not resemble any other that has previously been characterized. The helix and loop from the loop–sheet–helix motif are inserted in the DNA's major groove, where they make sequence-specific contacts with the bases (lower right of Fig. 34-113a). One of the large loops provides a side chain (Arg 248) that fits in the minor groove (middle right of Fig. 34-113a). The protein also contacts the DNA backbone between the major and minor grooves in this region (notably with Arg 273).

The structure's most striking feature is that *its DNA-binding motif consists of conserved regions comprising the most frequently mutated residues in the p53 variants found in tumors.* Among them are one Gly and five Arg residues (highlighted in yellow in Fig. 34-113a) whose mutations collectively account for over 40% of the p53 variants in tumors. The two most frequently mutated residues, Arg 248 and Arg 273, as we saw, directly contact the DNA. The other four “mutational hotspot” residues appear to play a critical role in structurally stabilizing p53's DNA-binding

surface. The relatively sparse secondary structure in the polypeptide segments forming this surface (one helix and three loops) accounts for this high mutational sensitivity: Its structural integrity mostly relies on specific side chain–side chain and side chain–backbone interactions.

Marmorstein generated a stable tetramer of the p53 DNA-binding domain in complex with its target DNA by chemically cross-linking its protein subunits to the DNA at a position that the foregoing structure indicates closely approaches the DNA. The X-ray structure of this complex (Fig. 34-113b) reveals that it is a 2-fold symmetric dimer of 2-fold symmetric dimers. The dsDNA assumes the largely undistorted B form and the structures of the protein subunits are nearly identical to that of the monomer depicted in Fig. 34-113a, as are the contacts made by each subunit to the DNA.

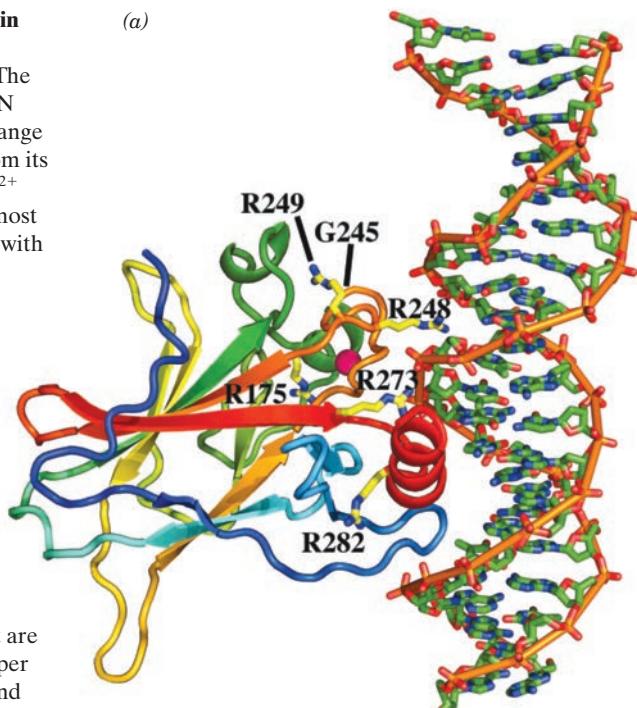
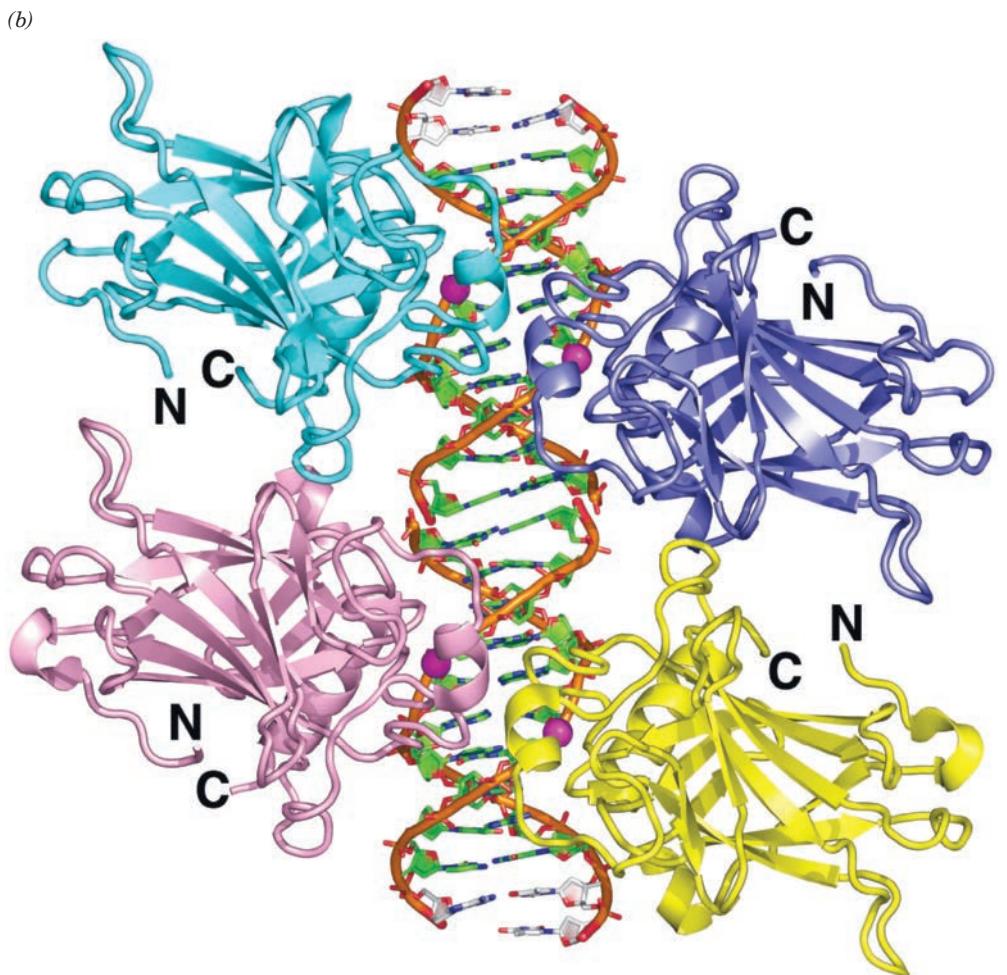
h. p53 Is a Sensor That Integrates Information from Several Pathways

p53 may be activated by several other pathways. For example, aberrant growth signals, such as those generated by oncogenic variants of MAP kinase cascade components such as Ras, stimulate the expression of a variety of transcription factors (Fig. 19-40), many of which are proto-oncogene products. One of them, **Myc**, activates the transcription of the gene encoding **p19^{ARF}** (in mice; **p14^{ARF}** in humans), which also encodes **p16^{INK4a}**. This is because these two proteins, which have no sequence similarity, are expressed through alternative splicing of their first exons and share second and third exons that are translated in different reading frames for the two proteins (ARF for *alternative reading frame protein*), an unprecedented economy of genomic resources in higher eukaryotes, although a common phenomenon in bacteriophage (Section 32-1D). p19^{ARF} binds to Mdm2 and inhibits its activity, thereby preventing the degradation of p53 and hence triggering the p53-dependent transcriptional programs leading to cell cycle arrest as well as apoptosis (Section 34-4E). Evidently, p19^{ARF} acts as part of a p53-dependent fail-safe system to counteract hyperproliferative signals.

A third activation pathway for p53 is induced by a wide variety of DNA-damaging chemotherapeutic agents, protein kinase inhibitors, and UV radiation. These activate ATR to phosphorylate p53 so as to reduce its affinity for Mdm2 in much the same way as do ATM and Chk2.

p53 is also subject to a rich variety of reversible post-translational modifications that markedly influence the expression of its target genes. These include acetylation at several Lys residues, glycosylation, ribosylation, and sumoylation (Section 32-6Bm), as well as phosphorylation at multiple Ser/Thr residues and ubiquitination. p53 does not bind to short DNA fragments that contain its target sites unless its C-terminal domain has either been deleted or modified by phosphorylation and/or acetylation. Yet NMR evidence indicates that the C-terminal domain does not interact with other p53 domains. This suggests that p53's C-terminal and core domains compete for DNA binding unless the C-terminal domain has been modified.

Figure 34-113 X-ray structures of the DNA-binding domain of p53 in complex with its target DNA. (a) A monomer of the human p53 DNA binding domain in complex with DNA containing its target sequence. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P orange) with its successive P atoms are joined by orange rods. The protein is shown in ribbon form colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). A tetrahedrally liganded Zn^{2+} ion is represented by a magenta sphere, and the side chains of the six most frequently mutated residues in human tumors are shown in stick form with C yellow. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1TSR.] (b) The X-ray structure of a tetramer of the mouse p53 DNA-binding domain whose subunits are each cross-linked to an 11-bp DNA containing p53's 10-bp palindromic target sequence. The complex contains two such dsDNAs stacked on each other to form a pseudocontinuous helix and related by 2-fold symmetry. Hence this complex is a 2-fold symmetric dimer of dimers. The p53 subunits are shown in ribbon form with those of the upper dimer cyan and lavender and those of the lower dimer pink and yellow. The DNA is drawn in paddle form colored as in Part *a* but with the C atoms of nucleotides that are not part of the target sequence white. Zn^{2+} ions are represented by magenta spheres. The view is along the complex's 2-fold axis with the DNA's helix axis vertical. The subunits on the right are viewed approximately from the rear of that in Part *a*. Note that the upper and lower dimers each have internal 2-fold symmetry whose axes extend toward the viewer. [Based on an X-ray structure by Ronen Marmorstein, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 3EXJ.]  See the Interactive Exercises.



p53, as we have only glimpsed, is the recipient of a vast number of intracellular signals and, in turn, controls the activities of a large number of downstream regulators. One way to understand the operation of this highly complex and interconnected network is in analogy with the Internet. In the Internet (cell), a small number of highly connected servers or hubs (“master” proteins) transmit information to/from a large number of computers or nodes (other proteins) that directly interact with only a few other nodes (proteins). In such a network, overall performance is largely unperturbed by the inactivation of one of the nodes (other proteins). However, the inactivation of a hub (master protein) will greatly impact system performance. *p53* is a master protein, that is, it is analogous to a hub. Inactivation of one of the many proteins that influences its performance or one of the many proteins whose activity it influences usually has little effect on cellular events due to the system’s redundant and highly interconnected components. However, the inactivation of *p53* or several of its most closely associated proteins (e.g., *Mdm2*) disrupts the cell’s responses to DNA damage and tumor-predisposing stresses. Nevertheless, a quantitative understanding of the functions and malfunctions of the *p53*-based network will require a complete description of all the proteins with which *p53* directly or indirectly interacts and how they do so under the conditions present in the cell, information that is lacking. Thus, for the foreseeable future, we will be limited to qualitative descriptions of the functioning of the *p53* network. Our understanding of other cellular signal transduction systems is similarly vague (Section 19-4F).

i. pRb Regulates the Cell Cycle’s G₁/S Transition

The tumor suppressor pRb (**retinoblastoma-associated protein**), which is localized in the nucleus of normal animal cells but is defective or absent in retinoblastoma cells (Section 34-4Ca), is an important regulator of the cell cycle’s G₁/S transition, that is, the cell’s passage through the restriction point. The effects of pRb are largely manifested through its interactions with the members of the **E2F** family of transcription factors, which, in the absence of pRb, activate their target promoters in complex with a member of the **DP** (for *E2F dimerization partner*) family. The mammalian E2F family consists of six, ~440-residue members, of which **E2F-1** through **E2F-4** interact with pRb via a conserved 18-residue polypeptide segment contained in their ~70-residue, C-terminal, transactivation domains. The mammalian DP family consists of two ~430-residue members, **DP-1** and **DP-2**. The E2F–DP heterodimers induce the transcription of a variety of genes that encode proteins required for S-phase entry (e.g., Cdc2, Cdk2, and cyclins A and E) and for DNA synthesis [e.g., DNA polymerase α (pol α ; Section 30-4Ba), Orc1 and several Mcm proteins (which participate in initiating DNA replication; Section 30-4Bc), ribonucleotide reductase (Section 28-3Aa), thymidylate synthase, and dihydrofolate reductase (Section 28-3Bd)].

How does pRb mediate cell cycle progression? pRb, which is synthesized throughout the cell cycle, is a phos-

phoprotein that is phosphorylated at as many as 16 of its Ser/Thr residues by Cdk4/6–cyclin D in mid to late G₁, by Cdk2–cyclin E in late G₁, by Cdk2–cyclin A in S, and by Cdk1 (Cdc2–cyclin B) in M, with different Cdk–cyclin complexes phosphorylating different sets of sites on pRb. *Hypophosphorylated but not hyperphosphorylated pRb binds to the transactivation domain of E2F so as to prevent it from activating transcription at the promoter to which it is bound.* In nonproliferating cells (those in early G₁), pRb remains hypophosphorylated because, unless such cells receive mitogenic signals, the highly unstable D-type cyclins (they have half-lives of ~10 min) do not accumulate to levels sufficient to generate significant amounts of Cdk4/6–cyclin D (mitogens trigger MAP kinase cascades that stimulate the expression of D-type cyclins). Moreover, since hypophosphorylated pRb prevents E2F–DP from activating the expression of Cdk2 and cyclins E and A, Cdk2–cyclin E and Cdk2–cyclin A do not accumulate to sufficient levels to hyperphosphorylate pRb. In fact, the small amounts of Cdk2–cyclin A and Cdk2–cyclin E that are present are inhibited by p27^{Kip1}, which occurs in high levels in pre-restriction point G₁ cells.

Mitogens also activate the expression of p27^{Kip1} and p21 which, contrary to what might be expected, do not inhibit Cdk4/6–cyclin D complexes but instead stimulate their activities by enhancing their assembly and promoting their nuclear import. Thus, mitogenic signals break the pRb-imposed blockade to cell cycle progression by inducing the formation of Cdk4/6–cyclin D–p27^{Kip1}/p21 complexes that begin the phosphorylation of pRb. This releases a small amount of E2F, which thereupon induces the expression of Cdk2 and cyclins E and A. The Cdk4/6–cyclin D complexes also sequester p27^{Kip1} and p21, which permits the resulting Cdk2–cyclin E complex to catalyze a second wave of pRb phosphorylation and import [although when large amounts of p21 are produced through the influence of activated p53 (see above), it inhibits Cdk2–cyclin E so as to arrest the G₁/S transition]. This frees large amounts of E2F, resulting in a surge in the transcription of the genes that promote cell cycle progression. As the cell cycle continues, pRb is increasingly phosphorylated, first by Cdk2–cyclin A and then by Cdk1, until the exit from M phase, whereupon pRb is abruptly dephosphorylated by the protein Ser/Thr phosphatase PP1 (Section 19-3Fd), permitting pRb to again arrest cell cycle progression by inhibiting E2F.

A variety of proteins that contain the LXCXE sequence motif bind to pRb. These comprise several cellular proteins, including the D-type cyclins, which may thereby be directed to pRb, and certain viral oncoproteins, whose binding to pRb prevents it from binding E2F (which lacks an LXCXE motif but, as we saw above, binds pRb via an 18-residue sequence). Indeed, E2F was first identified (and named) as a cellular factor involved in the regulation of the adenovirus early *E2* gene by the adenovirus oncogene product E1A, although further investigations revealed that the adenovirus *E4* gene product also participates in this process. E1A, which does not bind DNA, binds, via its LXCXE motif, to pRb. This causes the pRb to release

its bound E2F, which permits the E2F in combination with **E4**, to activate the transcription of the *E2* gene from the adenovirus *E2* promoter. The freed E2F also drives the infected cell into S phase, which facilitates adenovirus DNA replication. The SV40 large T antigen and the human papillomavirus **E7** protein, which also contain LXCXE motifs, similarly activate E2F. Over 100 proteins have been reported to bind to pRb, although in most cases by other means than via LXCXE motifs. The functions of these interactions are largely unknown.

j. E2F and the LXCXE Motif Bind to Separate Sites on the pRb Pocket Domain

pRb's so-called **pocket domain** forms the binding site for both E2F and the LXCXE motif and is the major site of genetic alterations in tumors. The pocket domain consists of its conserved A- and B-boxes (residues 379–572 and 646–772) linked by a poorly conserved spacer. However, when the spacer is excised, the A- and B-boxes nevertheless associate noncovalently.

The X-ray structure of the pRb pocket domain lacking its spacer in complex with the 18-residue pRb-binding peptide of E2F was independently determined by Mar-morstein and Gamblin and by Yunje Cho. It reveals that the 18-residue E2F peptide binds in a boomerang-shaped conformation at the highly conserved interface between the A- and B-boxes, both of which contain the five-helix cyclin fold (Fig. 34-114). However, the X-ray structure of the pRb pocket domain in complex with the 9-residue LXCXE-containing peptide from human papillomavirus E7 protein, determined by Pavletich, indicates that the E7 LXCXE peptide binds, in an extended conformation, in a shallow groove on the B-box that is ~30 Å distant from the

E2F-binding site (Fig. 34-114). This latter binding site, which is formed by highly conserved residues, closely resembles the primary Cdk2 binding site of cyclin A (Fig. 34-109b) and the TBP binding site of the 20% identical TFIIB (Fig. 34-54a). The corresponding portion of the A-box participates in forming the A–B interface.

k. pRb Also Represses Transcription by Recruiting HDACs and SWI/SNF Homologs

Binding experiments reveal that pRb associates with the histone deacetylases (HDACs; Section 34-3Bdd) HDAC1 and HDAC2, both of which contain an LXCXE sequence motif (actually IXCXE). Consequently, the presence of human papillomavirus E7 protein or mutations that disrupt the pRb pocket domain abolish the binding of these HDACs to pRb. These observations suggest that pRb also functions to recruit HDAC1 and HDAC2 to DNA-bound E2F so as to facilitate the histone deacetylation and hence the transcriptional inactivation of the chromatin containing E2F's target genes. This explains the observation that pRb can repress transcription from promoters to which it is artificially linked by a DNA-binding domain that differs from that of E2F. HDAC3 also associates with pRb, although it lacks an LXCXE motif.

The human SWI/SNF homologs **BRM** and **BRG1**, which both have LXCXE motifs, also bind to pRb. BRM and BRG1 are DNA-dependent ATPases that are components of chromatin-remodeling complexes. Thus, the observation that pRb can simultaneously bind BRG1 and an HDAC (despite both having LXCXE motifs) suggests that pRb recruits chromatin remodeling complexes to facilitate the action of HDACs at E2F promoters. pRb also interacts with condensin (which organizes the radial loops of metaphase chromosomes; Section 34-1Da), which suggests that pRb may participate in regulating the level of chromatin structure.

E. Apoptosis: Programmed Cell Death

The maxim that death is part of life is even more appropriate on the cellular level than it is on the organismal level. **Programmed cell death** or **apoptosis** (Greek: falling off, as leaves from a tree), which was first described by John Kerr in the late 1960s, is a normal part of development as well as the maintenance and defense of the adult animal body. For example, in the nematode worm *Caenorhabditis elegans*, a transparent organism whose cell lineages have been microscopically elucidated, precisely 131 of its 1090 somatic cells undergo apoptosis in forming the normal adult body. In many vertebrates, the digits of the developing hands and feet are initially connected by webbing that is eliminated by programmed cell death (Fig. 34-115), as are the tails of tadpoles and the larval tissues of insects during their metamorphoses into adults (Figs. 34-89 and 34-94). Apoptosis is particularly prevalent in the developing mammalian nervous system in which an approximately 3-fold excess of neurons is produced. However, only those neurons that make adequate

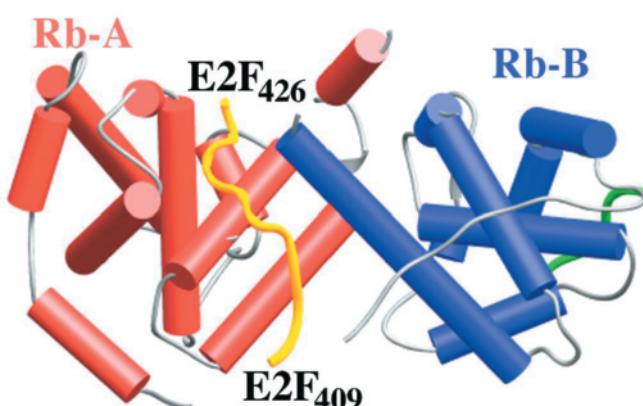


Figure 34-114 X-ray structure of the pRb pocket domain in complex with the 18-residue pRb-binding peptide of E2F. The helices of the A- and B-boxes are respectively drawn as red and blue cylinders and the main chain of the E2F peptide is shown as a gold worm. The superimposed structure of an LXCXE-containing nonapeptide segment of the human papillomavirus E7 protein in its complex with the pRb pocket domain is represented by a green worm. [Courtesy of Steven Gamblin, National Institute for Medical Research, London, U.K. PDBBids 1O9K and 1GUX.]

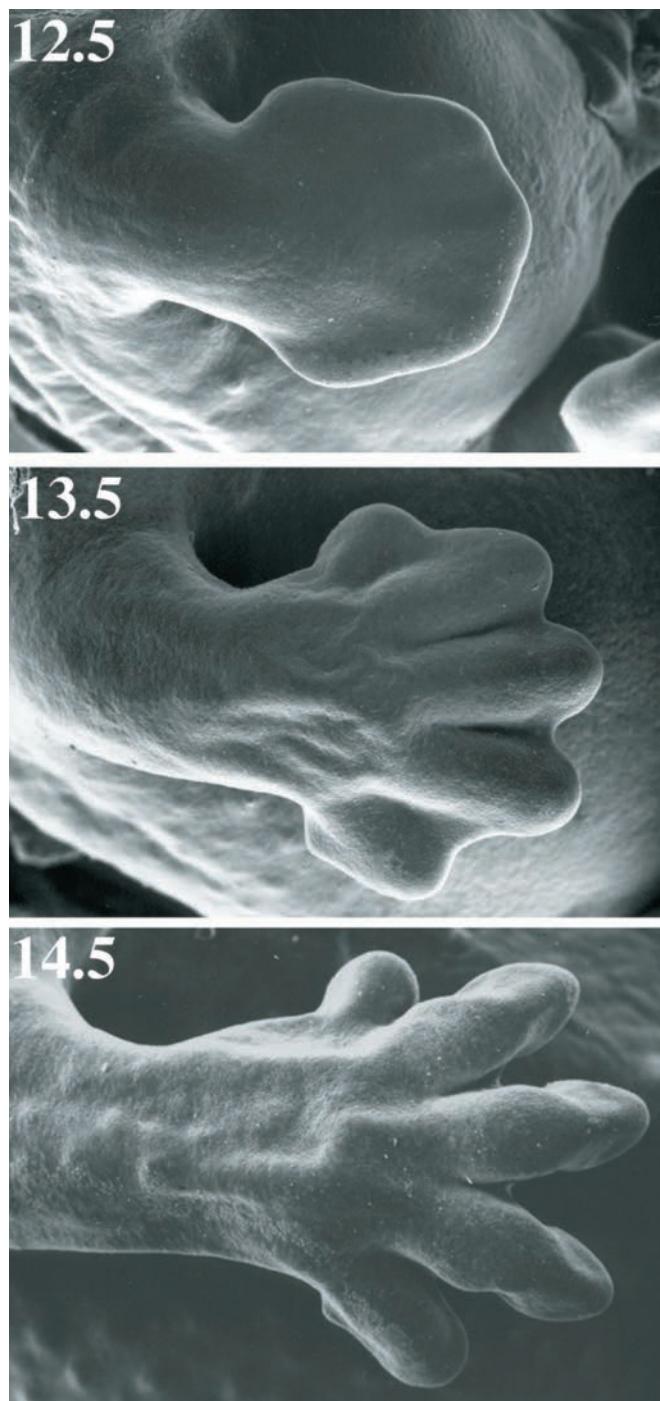


Figure 34-115 Programmed cell death in the embryonic mouse paw. At day 12.5 of development, its digits are fully connected by webbing. At day 13.5, the webbing has begun to die. By day 14.5, this apoptotic process is complete. [Courtesy of Paul Martin, University College of London, U.K.]

synaptic connections are retained; the remainder are eliminated via apoptosis (Fig. 34-116).

In the adult human body, which consists of nearly 10^{14} cells, an estimated 10^{11} cells are eliminated each day through programmed cell death (which closely matches

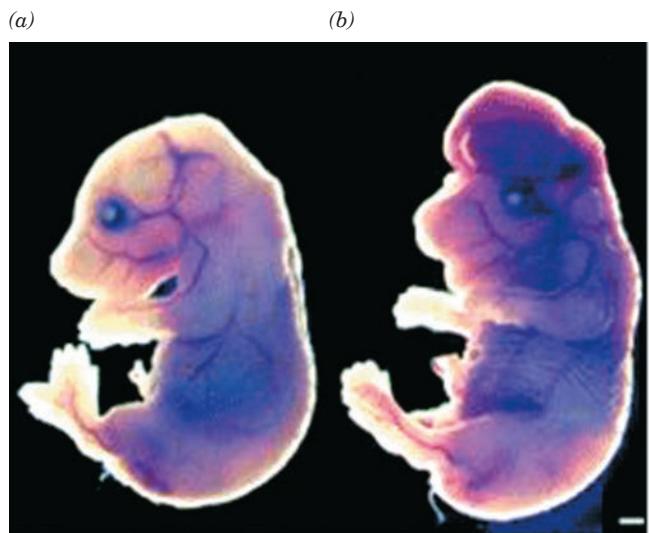


Figure 34-116 Brain development in 16.5-day-old mouse embryos. (a) A wild-type embryo. (b) An embryo in which caspase-9, an enzyme that mediates apoptosis (see below) has been knocked out. Note the protruding and morphologically abnormal brain in the knockout embryo due to the overproliferation of brain neurons. [Courtesy of Richard Flavell, Yale University Medical School.]

the number of new cells produced by mitosis). Indeed, the mass of the cells that we annually lose in this manner approaches that of our entire body. A particularly obvious manifestation of this phenomenon is the monthly sloughing off of the uterine lining in menstruation (Section 19-11). Similarly, the immune system cells known as T lymphocytes (T cells) undergo apoptosis in the thymus if the T cell receptors they produce recognize antigens that are normally present in the body or are improperly formed (Sections 35-2Aa and 35-2D); ~99.9% of immature T cells are eliminated in this way. Autoimmune diseases such as rheumatoid arthritis and insulin-dependent diabetes (Section 27-4Ba) arise when this process goes awry. Apoptosis is also an essential part of the body's defense systems. The immune system eliminates virus-infected cells, in part, by inducing them to undergo apoptosis, thereby preventing viral replication. Cells with irreparably damaged DNA and hence at risk for malignant transformation undergo p53-stimulated apoptosis, thereby protecting the entire organism from cancer. Indeed, *one of the defining characteristics of malignant cells is their ability to evade apoptosis*. Cells that become detached from their normal positions in the body likewise commit suicide. Indeed, as Martin Raff pointed out, *apoptosis appears to be the default option for metazoan cells: Unless they continually receive external hormonal and/or neuronal signals not to commit suicide, they will do so*. Thus, adult organs maintain their constant size by balancing cell proliferation with apoptosis. Not surprisingly, therefore, inappropriate apoptosis has been implicated in several neurodegenerative diseases including Alzheimer's disease (Section 9-5B), Parkinson's disease (Section 26-4B),

and Huntington's disease (Section 30-7f), as well as much of the damage caused by stroke and heart attacks. Consequently, the signaling systems that mediate apoptosis have become targets for therapeutic intervention. In fact, many of the chemotherapeutic agents in present use do not kill their target cancer cells outright but, rather, damage them so as to induce their apoptosis.

Apoptosis is qualitatively different from **necrosis**, which is cell death as caused by trauma (e.g., lack of oxygen, extremes of temperature, and mechanical injury). Cells undergoing necrosis essentially explode: They and their membrane-enclosed organelles swell as water rushes in through their compromised membranes, releasing lytic enzymes that digest the cell contents until the cell lyses, spilling its contents into the surrounding region (Section 22-4Ce). The cytokines that the cell releases often induce an inflammatory response (which can damage surrounding cells) that attracts **phagocytes** (white blood cells, such as **macrophages**, that ingest foreign particles and waste matter) to "mop up" the resulting cell debris. In contrast, apoptosis begins with the loss of intercellular contacts by an apparently healthy cell followed by its shrinkage, the condensation of its chromatin at the nuclear periphery, the collapse of its cytoskeleton, the dissolution of its nuclear envelope, the fragmentation of its DNA, the fragmentation of its ER, Golgi apparatus, and mitochondria, and the violent blebbing (blistering) of its plasma membrane due to the weakened cytoskeleton. Eventually, the cell disintegrates into numerous membrane-enclosed **apoptotic bodies** that are ingested by neighboring cells as well as by roving phagocytes without spilling the cell contents and hence not inducing an inflammatory response. The phagocytic cells recognize the apoptotic bodies through "eat me" signals. In mammals, this signal is the phospholipid phosphatidylserine, which in the plasma membrane of healthy cells occurs only in its inner leaflet (Fig. 12-35) but in apoptotic cells is translocated to the outer leaflet.

a. Apoptosis Is Induced by Signaling Cascades

The pathway for apoptosis, which was first elucidated in *C. elegans* through genetic studies by John Sulston and Robert Horvitz, involves three so-called *ced* (for *cell death abnormal*) gene products (Fig. 34-117): **CED-4** protein, a protease, activates the protease **CED-3**, which then initiates the destruction of the cell; **CED-9** functions to inactivate **CED-4**. In fact, mutations that inactivate **CED-9** result in numerous embryonic cells that would normally survive in the adult organism to inappropriately activate its **CED-4** and **CED-3** and hence die, thereby killing the embryo. Conversely, if **CED-9** is expressed at abnormally high levels or **CED-3** or **CED-4** is inactivated, cells that normally die will survive (which, curiously, has little apparent effect on the health of the adult organism). Later investigations revealed that a fourth protein, **EGL-1** (*egg-laying defective-1*), functions to inhibit **CED-9** and hence its overexpression induces apoptosis.

Apoptotic pathways in mammals are considerably more complex than that in *C. elegans*. Nevertheless, the above

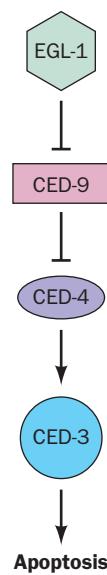


Figure 34-117 The pathway initiating apoptosis in *C. elegans*. Arrows indicate activation and blunted lines indicate inhibition.

CED proteins and **EGL-1** all have counterparts in mammalian pathways:

1. **CED-3** is the prototype of a family of proteases known as **caspases** (for *cysteinyl aspartate-specific proteases*) because they are **cysteine proteases** [whose mechanism resembles that of serine proteases (Section 15-3C) but with Cys replacing the active site Ser] that cleave after an Asp residue. Their target cleavage sites are specified mainly by this Asp and its three preceding residues.

2. **CED-4** is a scaffolding protein that plays an essential role in caspase activation. Its mammalian counterpart is called **Apaf-1** (for *apoptotic protease-activating factor-1*).

3. **CED-9** is a member of the **Bcl-2** family (so named because its founding member, Bcl-2, was initially characterized as a gene involved in *B cell lymphoma*). Some of the numerous members of this family, including **CED-9**, protect cells from death and hence are said to be **antiapoptotic**. Others promote cell death and are therefore said to be **proapoptotic**.

4. **EGL-1** is a proapoptotic member of the Bcl-2 family.

b. Caspases Have Closely Similar Structures

Caspases are $\alpha_2\beta_2$ heterotetramers that consist of two large α subunits (~ 300 residues) and two small β subunits (~ 100 residues). They are expressed as zymogens (**procaspsases**) that have three domains (Fig. 34-118): an N-terminal prodomain that is proteolytically excised on activation, followed by sequences comprising the active enzyme's α and β subunits that are proteolytically separated on activation. The activating cleavage sites all follow Asp residues and are, in fact, targets for caspases (the only other eukaryotic protease known to cleave after an Asp residue is **granzyme B**, a chymotrypsin-like serine protease expressed by

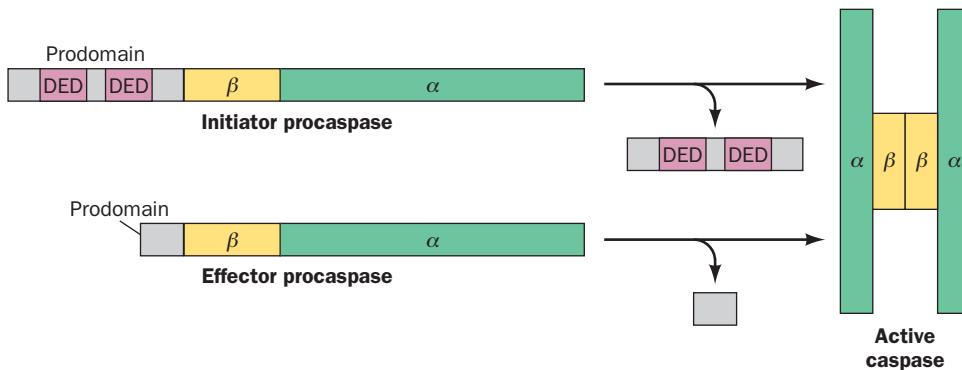


Figure 34-118 Caspase domain structure and activation. The zymogens of initiator caspases have long N-terminal prodomains, which in several cases contain two death effector domains (DEDs), whereas the zymogens of effector caspases have only

short prodomains. Pro caspases are activated by proteolytic cleavages that excise their prodomains and separate their α and β subunits to yield the active $\alpha_2\beta_2$ caspases.

cytotoxic T cells that functions to induce apoptosis in tumor- and virus-infected cells; Section 35-2Aa). Thus, caspase activation, as we shall see, may either be autocatalytic or be catalyzed by another caspase.

Humans express eleven caspases, six of which participate exclusively in apoptosis [with the remainder being involved mainly in cytokine activation and hence the control of inflammation; the founding member, **caspase-1**, is also known as **interleukin-1 β -converting enzyme (ICE)** because it proteolytically activates the cytokine **interleukin-1 β** (Section 35-2Aa)]. There are two classes of apoptotic caspases (Fig. 34-118):

1. Initiator caspases (caspases-8, -9, and -10) are characterized by long prodomains (129–219 residues) that target their zymogens to scaffolding proteins that promote their autoactivation. The prodomains of caspases-8 and -10 each contain two ~80-residue **death effector domains (DEDs)**, through which they bind to DEDs on their target adaptor proteins (see below). The prodomain of caspase-9 instead contains the structurally similar ~90-residue **caspase recruitment domain (CARD)** that promotes the interaction of this caspase with certain scaffolding and regulatory proteins.

2. Effector caspases (caspases -3, -6, and -7) have short (~25 residue) prodomains and are activated by initiator caspases. The activated effector caspases, which have been described as the cell's executioners, cleave a wide variety of cellular proteins (see below), thereby bringing about apoptosis.

The X-ray structure of caspase-7 (Fig. 34-119), determined by Keith Wilson and Paul Charifson, closely resembles those of the several other caspases of known X-ray structures. Each $\alpha\beta$ heterodimer of this 2-fold symmetric $\alpha_2\beta_2$ heterotetramer contains a six-stranded β sheet, five of whose strands are parallel, and which is flanked by five α helices, two on one side and three on the other, that are approximately parallel to the β strands. The β sheet is continued across the protein's 2-fold axis to form a twisted 12-stranded β sheet. Each $\alpha\beta$ heterodimer contains an active site that is located at the C-terminal ends of its parallel β strands and which recognizes a tetrapeptide on the N-terminal side of its Asp-X cleavage site. The structures of the

various caspases differ mainly in the conformations of the four loops forming their active sites. Comparison of the X-ray structure of caspase-7 with that of **procaspase-7** (with its active site Cys 186 mutated to Ser to prevent its autoactivation), independently determined by Weiguo Shi and Wolfram Bode, reveals that, although the two proteins are otherwise closely superimposable, the four active site loops

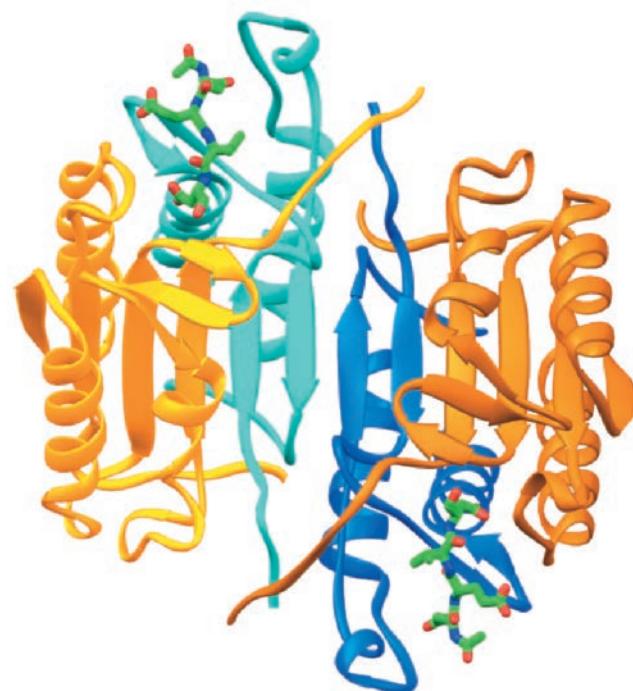


Figure 34-119 X-ray structure of caspase-7 in complex with the tetrapeptide aldehyde inhibitor acetyl-Asp-Glu-Val-Asp-CHO (Ac-DEVD-CHO). The $\alpha_2\beta_2$ heterotetrameric enzyme is viewed along its 2-fold axis with its large (α) subunits orange and gold and its small (β) subunits cyan and light blue. The Ac-DEVD-CHO inhibitor is drawn in stick form with C green, N blue, and O red. Note how the central 12-stranded β sheet is continued across all four subunits. [Based on an X-ray structure by Keith Wilson and Paul Charifson, Vertex Pharmaceuticals, Cambridge, Massachusetts. PDBid 1F1J.]

in procaspase-7 have undergone large conformational changes relative to those in caspase-7 that essentially obliterate the active site. In particular, the loop containing both the catalytically essential Cys residue and the activating cleavage site between the α and β subunits changes its orientation by 90° after this cleavage so as to expose and properly position the previously buried catalytic Cys residue.

c. Caspases Cleave a Wide Variety of Proteins and Activate the Degradation of Chromosomal DNA

Nearly 800 cellular proteins have been identified as caspase substrates. These include cytoskeletal proteins [e.g., actins (Section 35-3E) and **lamins** (intermediate filaments that form the meshwork lining the inner nuclear envelope)], proteins involved in cell cycle regulation (e.g., cyclin A, Wee1, p21, ATM, and pRb; Section 34-4D), proteins that participate in DNA replication [e.g., topoisomerase I (Section 29-3C) and Mcm3 (Section 30-4Bc)], transcription factors (e.g., Sp1 and NF- κ B; Section 34-3B), translational initiation factors (e.g., eIF4E and eIF4G; Section 32-3Cd), ribosomal proteins, proteins that participate in signal transduction [e.g., RasGAP (Section 19-3Ci) and protein kinase C (Section 19-4C)], and proteins that inhibit apoptosis (see below).

The induction of apoptosis also causes the rapid degradation of chromosomal DNA. Chromosomal DNA is attached to the chromosomal protein matrix at intervals of \sim 70 kb via AT-rich matrix-associated regions (MARs; Section 34-1Da). During apoptosis, **caspase-activated DNase (CAD)** cleaves the chromosomal DNA at these sites, which is often followed by its cleavage between nucleosomes to yield a series of DNA fragments that differ in their lengths by increments of \sim 200 bp. CAD is ubiquitously expressed in all tissues in complex with its inhibitor **ICAD** (inhibitor of *CAD*), which on induction of apoptosis is cleaved by caspases-3 and -7, thereby releasing active CAD. ICAD also functions as a chaperone that must be present when CAD is being ribosomally synthesized in order for CAD to fold to its native conformation. This ensures that native CAD can only form in complex with ICAD and hence prevents inappropriate DNA cleavage. Although the cleavage of a cell's chromosomal DNA would certainly cause its death, cells containing mutant ICAD undergo apoptosis even though their chromosomal DNA remains intact. This suggests that DNA cleavage during apoptosis functions to prevent the cells that have phagocytosed apoptotic bodies from being transformed by the intact viral or damaged chromosomal DNA that the apoptotic bodies might otherwise contain.

d. The Death-Inducing Signal Complex Activates Apoptosis

Apoptosis in a given cell may be induced either by externally supplied signals in the so-called extrinsic pathway (death by commission) or by the absence of external signals that inhibit apoptosis in the so-called intrinsic pathway (death by omission). The extrinsic pathway is initiated by the association of a cell destined to undergo apoptosis with a cell that has selected it to do so. In what is perhaps the best characterized such pathway (Fig. 34-120), a 281-residue, single-pass, transmembrane protein named **Fas ligand (FasL)** that projects from the plasma membrane of the

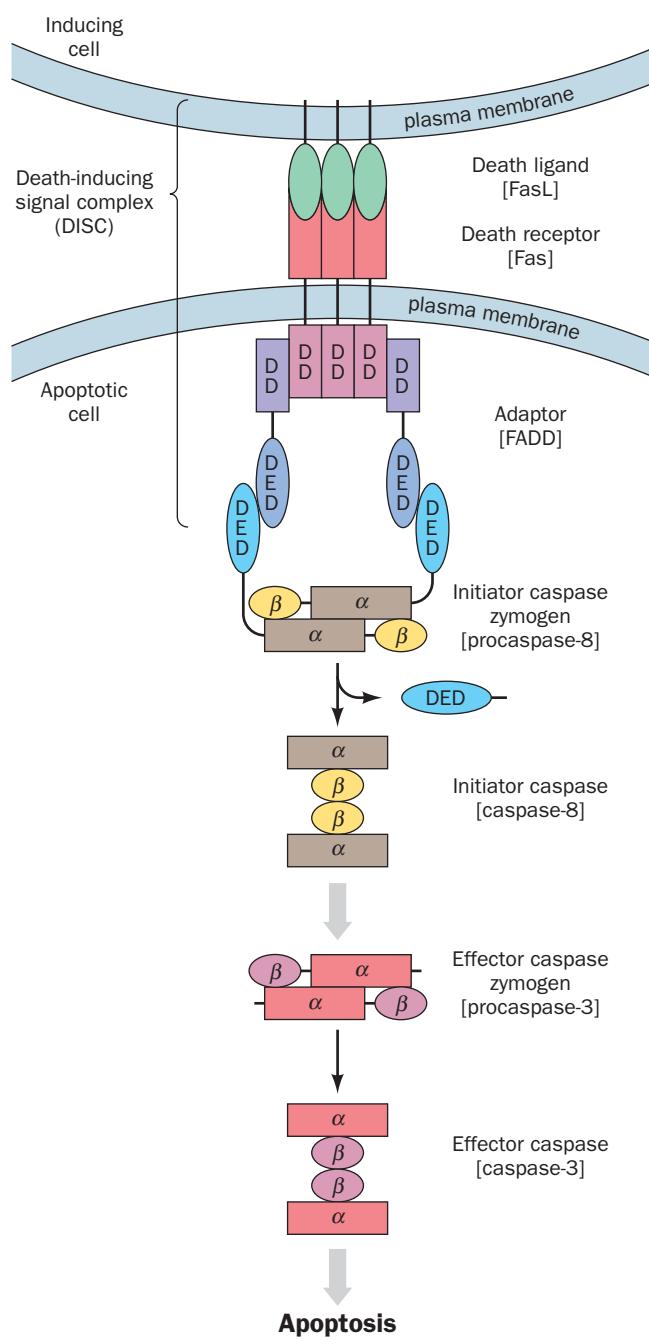


Figure 34-120 The extrinsic pathway of apoptosis. Flat arrows indicate activation. The binding of a trimeric death ligand (e.g., FasL) on the inducing cell to the death receptor (e.g., Fas) on the apoptotic cell causes the death receptor's cytoplasmic death domains (DDs) to trimerize. This recruits adaptors (e.g., FADD), which bind via their DDs to the DDs of the death receptor. The adaptors, in turn, recruit initiator procaspases (e.g., procaspase-8) via the interactions between the death effector domains (DEDs) on the adaptors and the initiator procaspases, which induces the autoactivation of the initiator procaspases to form the corresponding heterotetrameric initiator caspases (e.g., caspase-8). The initiator caspases then proteolytically activate effector procaspases (e.g., procaspase-3) to yield the heterotetrameric effector caspases (e.g., caspase-3), which catalyze the proteolytic cleavages, resulting in apoptosis.

inducing cell, a so-called **death ligand**, binds to a 335-residue, single-pass, transmembrane protein known as **Fas** (alternatively, **CD95** and **Apo1**) that projects from the plasma membrane of the apoptotic cell, a so-called **death receptor**. FasL is a cytokine that is predominantly expressed by certain immune system cells, including activated T cells (although the association between the apoptotic cell and the immune system cell is mainly mediated by antigen-containing complexes; Section 35-2Ed); it is a member of the **tumor necrosis factor (TNF)** family (so named because its founding member, **TNF α** , was originally characterized as a cytokine that kills tumor cells—but by inducing their apoptosis, not their necrosis).

FasL is a homotrimeric protein, whose extracellular C-terminal domains associate with the extracellular N-terminal domains of three Fas molecules to form a 3-fold symmetric complex, thereby causing the Fas cytoplasmic domains to trimerize. This is the triggering event of the extrinsic pathway; it can also be induced by cross-linking Fas molecules using antibodies. Fas, which is abundantly expressed in a variety of tissues, is a member of the **TNF receptor (TNFR)** family. Consequently, the arrangement of the FasL–Fas complex almost certainly resembles that observed in the X-ray structure of the **TNF β** trimer in its complex with the extracellular domains from three **TNF receptor 1 (TNFR1)** molecules that was determined by David Banner (Fig. 34-121). The cytoplasmic C-terminal domain of Fas consists mostly of an ~80-residue **death domain (DD)** that occurs in all of the six known mammalian death receptors (one of which is TNFR1), each of which are TNFR family members. The DD consists of six antiparallel, amphipathic α helices that have an unusual arrangement and whose structure resembles those of the death effector domain (DED) and the caspase recruitment domain (CARD), as Fig. 34-122 indicates.

Trimerized Fas recruits three molecules of the 208-residue adaptor protein known as **FADD** (for *Fas*-associating *death domain*-containing protein; alternatively **MORT1** for *mediator of receptor-induced toxicity 1*) via interactions between FADD's C-terminal DD and that on Fas. The remaining portion of FADD consists almost entirely of a DED that, in turn, recruits pro caspases-8 and -10 via the DEDs in their prodomains (Fig. 34-118) to form the **death-inducing signal complex (DISC)**. The consequent clustering of the pro caspases-8 and -10 molecules results in their proteolytic autoactivation, yielding caspases-8 and -10. These initiator caspases, in turn, activate the effector (executioner) caspase, caspase-3, whose actions cause the cell to undergo apoptosis.

Cells also express a protein named **c-FLIP** [for cellular *FLICE inhibitory protein*; **FLICE** (for *FADD-like ICE*) is an alternative name for pro caspase-8] that resembles caspase-8 but is catalytically inactive. It associates with FADD via its two DEDs and thereby inhibits the autoactivation of caspases-8 and -10. FLIP apparently functions to dampen the cell's response to Fas so as to prevent inappropriate apoptosis. Certain herpes viruses and poxviruses encode **v-FLIPs** that function similarly to c-FLIPs to prevent apoptosis, thereby permitting the virus to propagate in the infected cell.

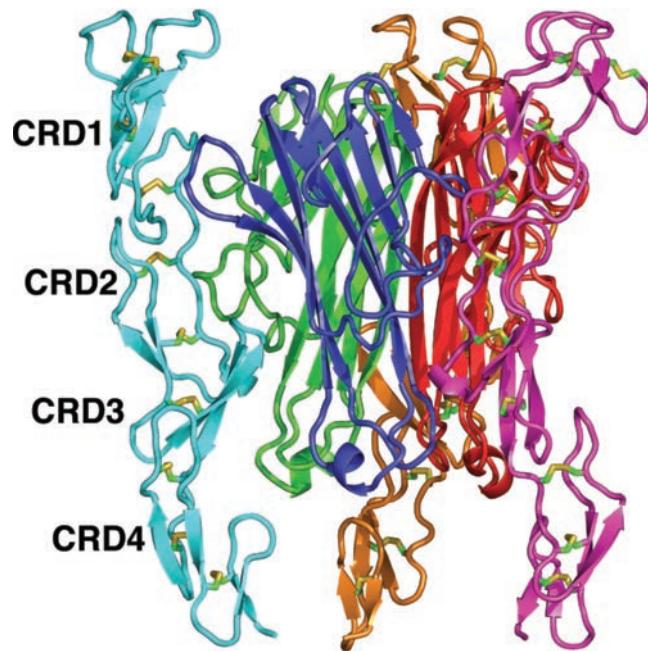


Figure 34-121 X-ray structure of the **TNF β** homotrimer in complex with the extracellular domains of three **TNFR1** molecules. The centrally located **TNF β** subunits are red, blue, and green, and the peripherally located **TNFR1** domains are cyan, orange, and magenta. The view is perpendicular to the complex's 3-fold axis with the apoptotic cell's plasma membrane below. The **TNFR1** domains each consist of four ~40-residue pseudorepeats known as **cysteine-rich domains (CRDs)** that each contain two disulfide bonds formed by four Cys side chains that are drawn here in stick form with C green and S yellow. The elongated **TNFR1** domains each bind at an interface between two **TNF β** subunits, out of contact with each other. This complex, whose formation also induces apoptosis, presumably resembles that between the extracellular domains of the homologous proteins **FasL** and **Fas** (whose extracellular domain contains only three CRDs). [Based on an X-ray structure by David Banner, F. Hoffmann-La Roche Ltd., Basel, Switzerland. PDBid 1TNR.]

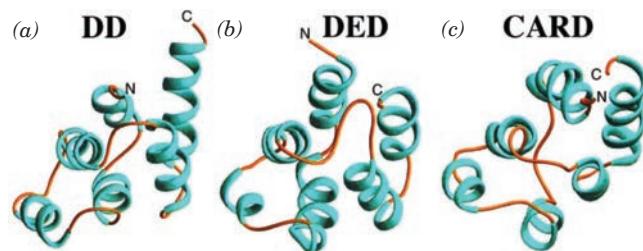


Figure 34-122 NMR structures of modules that transduce the death signal. (a) The death domain (DD) from Fas. (b) The death effector domain (DED) from FADD. (c) The caspase recruitment domain (CARD) from **RAIDD**, an adapter protein that is similar to FADD. Each of these domains consists of a bundle of six antiparallel α helices that associates with a domain of the same type but not with one of a different type. Nevertheless, the similarities of their structures and functions suggests that these domains are distantly related. [Courtesy of Stephen Fesik, Abbott Laboratories, Abbott Park, Illinois. Part c based on an NMR structure by Gerhard Wagner, Harvard University. PDBids 1DDF, 1A1Z, and 3CRD.]

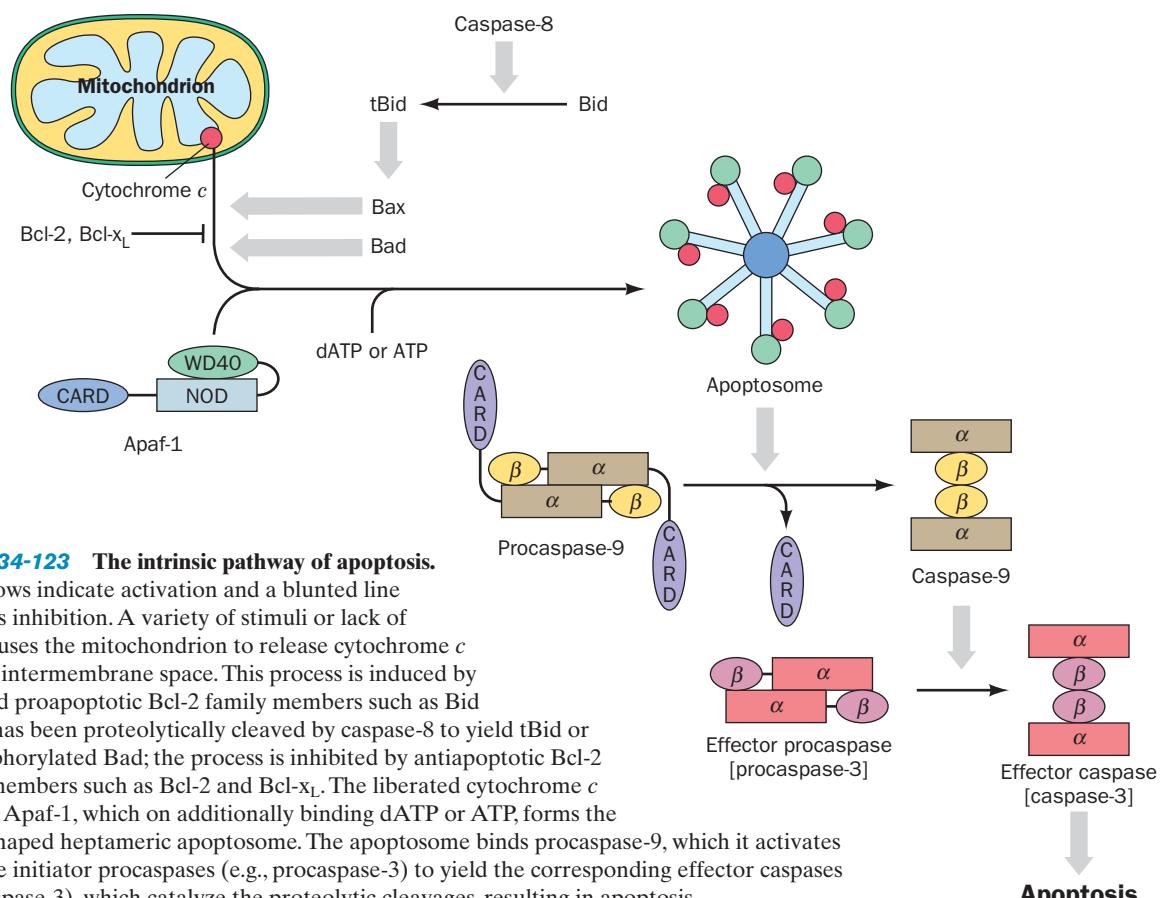


Figure 34-123 The intrinsic pathway of apoptosis.

Flat arrows indicate activation and a blunted line indicates inhibition. A variety of stimuli or lack of them causes the mitochondrion to release cytochrome *c* from its intermembrane space. This process is induced by activated proapoptotic Bcl-2 family members such as Bid after it has been proteolytically cleaved by caspase-8 to yield tBid or dephosphorylated Bad; the process is inhibited by antiapoptotic Bcl-2 family members such as Bcl-2 and Bcl-x_L. The liberated cytochrome *c* binds to Apaf-1, which on additionally binding dATP or ATP, forms the wheel-shaped heptameric apoptosome. The apoptosome binds procaspase-9, which it activates to cleave initiator procaspases (e.g., procaspase-3) to yield the corresponding effector caspases (e.g., caspase-3), which catalyze the proteolytic cleavages, resulting in apoptosis.

e. The Intrinsic Pathway Is Controlled by Bcl-2 Family Proteins

Most metazoan cells are continuously bathed in an extracellular soup, generated in part by neighboring cells, that contains a wide variety of cytokines that regulate the cell's growth, differentiation, activity, and survival. The withdrawal of this chemical support for its survival or the loss of direct cell-cell interactions induces a cell to undergo apoptosis via the intrinsic pathway. The initial step of this pathway (Fig. 34-123) appears to be the activation of one or more of the cell's several proapoptotic Bcl-2 family members.

The 15 known members of the ~180-residue mammalian Bcl-2 family have been classified into three groups (Fig. 34-124):

1. Group I members, which include Bcl-2 and **Bcl-x_L**, all have four short regions of homology, **BH1** to **BH4** (BH for *Bcl-2 homology region*), and a C-terminal hydrophobic segment that inserts into the outer mitochondrial membrane, or less frequently the endoplasmic reticulum, such that the bulk of these proteins face the cytosol. All Group I Bcl-2 family members are antiapoptotic.



Figure 34-124 Sequence comparisons of members of the Bcl-2 family of proteins. The BH1 through BH4 homology regions are blue, purple, red, and green, respectively, and the hydrophobic transmembrane (TM) region is yellow. Group I proteins are

antiapoptotic, whereas Group II and Group III proteins are proapoptotic. [After a drawing by Michael Hengartner, *Nature* 407, 770 (2000).]

2. Group II members, which include **Bax** and **Bak**, resemble Group I proteins but lack a BH4 region. All Group II members are proapoptotic.

3. Group III members, which include **Bad**, **Bid**, **Bik**, **Bim**, and **Blk** (and the *C. elegans* protein EGL-1), all possess only one BH region, the ~15-residue BH3, and have no other sequence resemblance to Bcl-2. These so-called **BH3-only proteins** are all proapoptotic.

The activities of many BH3-only proteins are controlled by specific post-translational modifications. For example, Bad is phosphorylated at two Ser residues by protein kinase A (PKA; Section 18-3Cb), mitogen-activated protein kinase (MAPK; Section 19-3D), and Akt (Section 19-4Dc) (which themselves are activated by complex signal transduction pathways), thereby generating a binding site for 14-3-3 proteins that then sequester Bad in the cytosol. On appropriate stimulation, calcineurin (Section 19-3Ff) and PP1 (Section 19-3Fd) dephosphorylate Bad, which permits it to interact with the mitochondrion, where it initiates apoptosis (see below). In contrast, Bid is activated by caspase-8-catalyzed proteolytic cleavage to **tBid** (truncated *Bid*), thereby providing a link between the extrinsic and intrinsic pathways of apoptosis. The transcription of two BH3-only proteins, **PUMA** (for *p53* upregulated modulator of apoptosis) and **Noxa** (Latin: damage), are transactivated by the tumor suppressor p53 (Section 34-4Df), thus accounting the ability of p53 to induce apoptosis. Moreover, p53 itself functions much like BH3-only proteins in that it binds antiapoptotic Bcl-2 family members and thereby inhibits their functions.

f. Cytochrome c Is an Essential Participant in the Intrinsic Pathway

The association of members of the proapoptotic Bcl-2 family with the mitochondrion causes it to release cytochrome *c* from its intermembrane space into the cytosol. There, as Xiaodong Wang unexpectedly discovered, this well-characterized component of the mitochondrial electron transport chain (Section 22-2C4) functions to induce apoptosis. It does so by combining with Apaf-1 and dATP or ATP to form an ~1100-kD complex named the **apoptosome** (Fig. 34-123). The apoptosome binds several molecules of procaspase-9 in a manner that induces their autoactivation to yield caspase-9, which remains bound to the apoptosome. This caspase-9 then activates procaspase-3 to instigate cell death.

g. The Apoptosome Has a Wheel-Like Structure

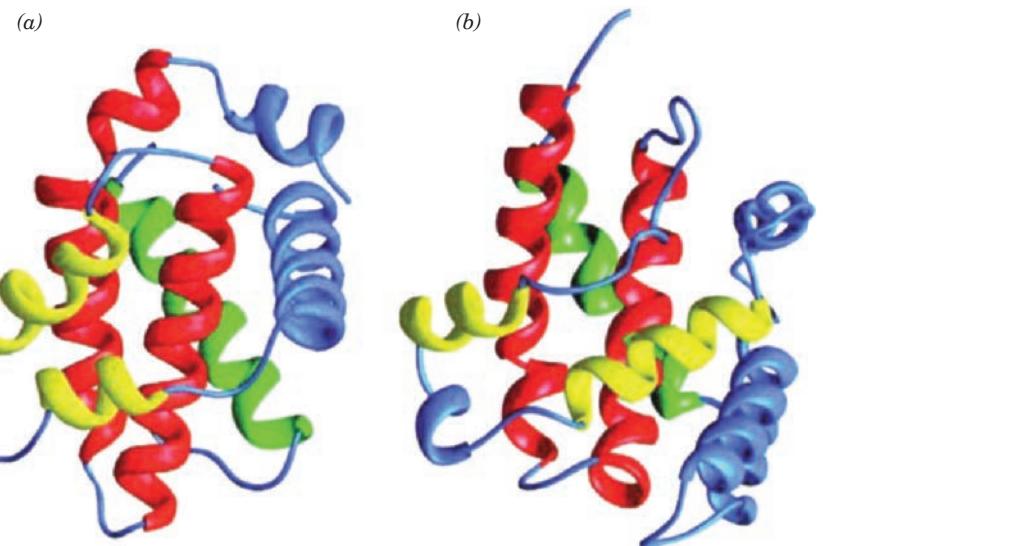
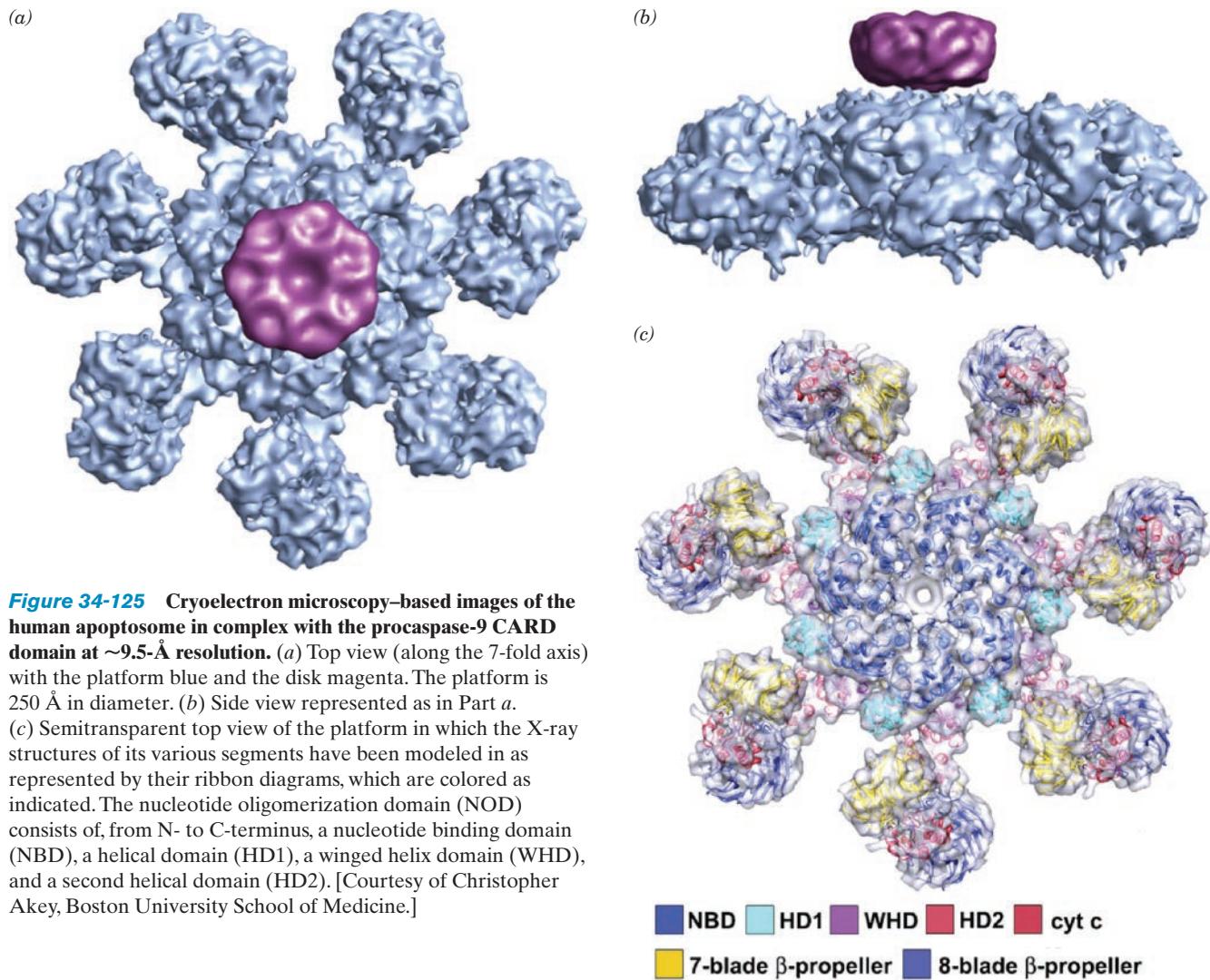
Apaf-1, the apoptosome's major component, is a 1248-residue scaffolding protein that consists of an N-terminal caspase recruitment domain (CARD), a central **nucleotide oligomerization domain (NOD)** that is homologous to CED-4 and a member of the AAA+ family of ATPases (Section 30-2Ca), and a C-terminal regulatory domain that consists of eight WD40 repeats (Section 25-6Bb), a short linker, and seven additional WD40 repeats. Procaspe-9 binds to Apaf-1's CARD in the apoptosome via its CARD, thereby placing several procaspase-9 molecules in close

proximity such that they proteolytically activate one another. More importantly, however, caspase-9's association with the apoptosome increases its catalytic activity by three orders of magnitude, presumably via an allosteric mechanism. Indeed, a procaspe-9 mutant (D315A) that is noncleavable between its α and β domains, in complex with the apoptosome, nevertheless efficiently activates procaspe-3. Apaf-1's WD40 repeats function to bind cytochrome *c* (WD40 repeats usually participate in protein–protein interactions); their excision from Apaf-1 permits it to bind and activate procaspe-9 in the absence of cytochrome *c*. This suggests that Apaf-1's WD40 repeats bind to its CARD so as to prevent it from binding procaspe-9 but preferentially bind cytochrome *c* and thereby release the CARD. Evolution presumably selected cytochrome *c* for this function because it is normally absent from the cytosol.

The ~9.5-Å resolution cryoEM-based structure of the human apoptosome in complex with the procaspe-9 CARD domain, determined by Christopher Akey, reveals a 7-fold symmetric wheel-like platform that has seven spokes ending in two lobes each that radiate from a central hub, and a disk of electron density floating above the central hub (Fig. 34-125a, b). Modeling of the X-ray structures of the various Apaf-1 components and cytochrome *c* into the platform (Fig. 34-125c) reveals that the larger and smaller of these lobes, respectively, consist of 8- and 7-bladed β propellers (WD40 repeats form β propellers of various sizes; e.g., Fig. 19-19b) that are bridged by a cytochrome *c* molecule and that the nucleotide oligomerization domain forms its hub and spokes. The platform has no additional space in which to fit the Apaf-1 CARD domains so that they must occupy the disk together with the procaspe-9 CARD domains. However, the disk is of lower resolution than the rest of the structure and too small to contain the seven CARD–CARD dimers. This suggests that the Apaf-1 CARD domains are flexibly linked to their respective NODs and hence partially disordered, which accounts for the apparent small size of the disk and the gap between it and the hub (Fig. 34-125b).

h. Several Mechanisms Have Been Proposed for Mitochondrial Cytochrome c Release

The way in which proapoptotic Bcl-2 family members cause the mitochondrion to release its cytochrome *c* is unclear. However, based on largely circumstantial evidence, three models for this process, which are not mutually exclusive, have been proposed. The first model is based on the structural resemblance of the antiapoptotic Bcl-2 family member Bcl-x_L to membrane-inserting bacterial toxins such as diphtheria toxin (Fig. 34-126; the mechanism of diphtheria toxin is discussed in Section 32-3Ge). This suggests that one or more molecules of Bcl-x_L and/or its homologs can insert into the outer mitochondrial membrane to form a pore. Indeed, Bcl-x_L, Bcl-2, and Bax have been shown to form pores in synthetic lipid bilayers. Moreover, tBid is targeted to mitochondria, where it triggers Bax oligomerization and its insertion into the outer mitochondrial membrane. However, it is unclear if the resulting pores are large enough to permit the passage of



amphipathic α helices. [Courtesy of Stephen Fesik, Abbott Laboratories, Abbott Park, Illinois. Part b based on an X-ray structure by David Eisenberg, UCLA. PDBIDs 1LXL and 1DDT.]

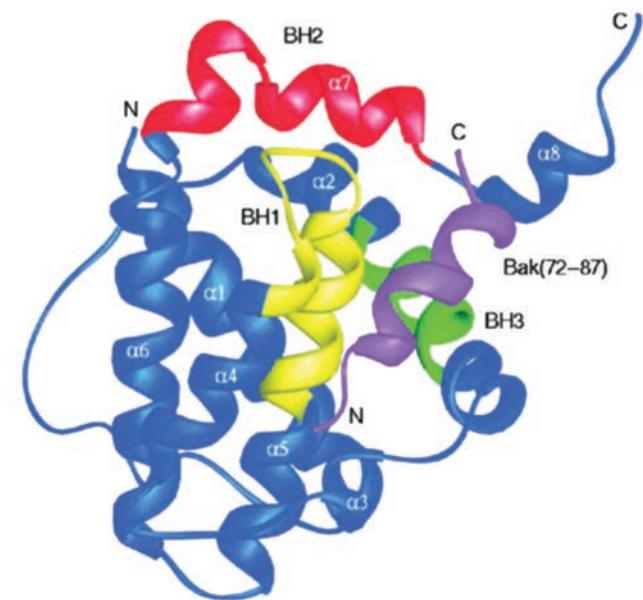


Figure 34-127 NMR structure of Bcl-x_L in complex with the 16-residue BH3 region of Bak. The BH1, BH2, and BH3 regions of Bcl-x_L are yellow, red, and green, respectively, and the Bak peptide is violet. [Courtesy of Stephen Fesik, Abbott Laboratories, Abbott Park, Illinois. PDBid 1BXL.]

cytochrome *c*; in addition, it is not known how pores formed by antiapoptotic proteins such as Bcl-x_L would inhibit apoptosis.

In the second model, Bcl-2 family members induce pre-existing mitochondrial outer membrane proteins to form channels through which cytochrome *c* is released. An attractive candidate for such a protein is the **voltage-dependent anion channel (VDAC; alternatively mitochondrial porin)** because several Bcl-2 family members can bind to it and alter its channel activity. However, the known size of the VDAC channel is too small to allow the passage of cytochrome *c*, so that this model requires that VDAC undergo significant conformational change on binding Bcl-2 family members.

In the third model, Bcl-2 family members perturb or stabilize the preexisting pores through which ATP and ADP are exchanged between the mitochondrial matrix and the cytosol. This exchange process is mediated by the ATP-ADP translocator in the inner mitochondrial membrane (Section 20-4C) and by VDAC in the outer mitochondrial membrane. The opening of these pores, it is postulated, would result in chemical equilibration between the cytosol and the matrix, which would cause the highly concentrated matrix to osmotically swell until it ruptured the outer mitochondrial membrane (which has considerably less surface area than the inner mitochondrial membrane due to the latter's cristae; Section 22-1A), thereby releasing cytochrome *c* into the cytosol. Consistent with this model, atractyloside and bongkrekic acid (Section 20-4C), inhibitors of the ATP-ADP translocator that respectively open and close its pore, respectively induce and inhibit

apoptosis. However, the rupture of the outer mitochondrial membrane has rarely been observed in cells undergoing apoptosis.

The way in which antiapoptotic Bcl-2 family members antagonize the functions of proapoptotic family members is better understood. The members of these opposing factions readily form homodimers in which the BH3 region of the proapoptotic protein, which forms an amphipathic α helix, binds in a hydrophobic groove on the antiapoptotic protein. Such an arrangement occurs in the NMR structure, determined by Fesik, of Bcl-x_L in its complex with the 16-residue segment from the BH3 region of the proapoptotic Bcl-2 protein Bak (Fig. 34-127). Since the BH3 regions of proapoptotic Bcl-2 proteins are necessary and probably even sufficient for their killing activity, the sequestering of these BH3 regions by antiapoptotic Bcl-2 proteins at least partially explains their antiapoptotic properties.

i. IAPs Regulate Apoptosis by Inhibiting Caspases

As might be expected, cells have elaborate systems that prevent their inadvertent apoptosis. We have previously discussed how the antiapoptotic Bcl-2 family proteins keep their proapoptotic cousins in check, and how c-FLIP inhibits the death-inducing signal complex (DISC)-mediated activation of caspases-8 and -10. In addition, the members of the **IAP** (for *inhibitors of apoptosis*) family of proteins, which are conserved from *Drosophila* to humans, regulate apoptosis by directly inhibiting caspases. Humans express eight IAPs, which vary in length from 236 to 4829 residues.

All IAPs have one to three ~70-residue **BIR domains** (for *baculovirus IAP repeat*; so named because they were discovered in the baculovirus protein **p35**, which functions to inhibit apoptosis in host cells during viral infection). BIRs contain a characteristic signature sequence, CX₂CX₁₆HX₆C, which forms a novel Zn²⁺-binding motif. In addition, many IAPs (five in humans) have a C-terminal RING finger domain [a ubiquitin-protein ligase (E3); Section 32-6Bb]. **BIR2** [the second BIR domain of the human protein **XIAP** (for *X-linked IAP*)] and its surrounding regions specifically bind and inhibit the effector caspases-3 and -7, whereas **BIR3** and its surrounding regions do so for the initiator caspase-9; the function of **BIR1** is unknown. An IAP that has a RING finger domain may also ubiquitinate its bound caspase, thereby condemning it to destruction in the proteasome.

The X-ray structure of caspase-3 in complex with the XIAP BIR2 domain and its 38-residue N-terminal extension, determined by Fesik, Robert Liddington, and Guy Salvesen, reveals, unexpectedly, that the BIR2 domain makes only limited contacts with caspase-3 (Fig. 34-128). Rather, most of the contacts to caspase-3 are made by the N-terminal extension, which spans the enzyme's active site so as to sterically block substrate binding. Curiously, the N-terminal extension extends across that active site in the reverse direction relative to that taken by polypeptide inhibitors of caspases such as Ac-DEVD-CHO (Fig. 34-119). The structures of both caspase-3 and the BIR2 domain in the complex are largely unperturbed relative to their uncomplexed structures.

The induction of apoptosis requires that the inhibitory effects of IAPs on caspases be relieved. This is the task of a homodimeric protein that is alternatively named **Smac** (for second mitochondria-derived activator of caspases) and **DIABLO** (for direct IAP-binding protein with low pI). Smac/DIABLO, which binds to the BIR domains of IAPs so as to prevent them from binding to caspases, is released from the mitochondrion together with cytochrome *c*, thereby ensuring that the intrinsic pathway of apoptosis will generate active caspases.

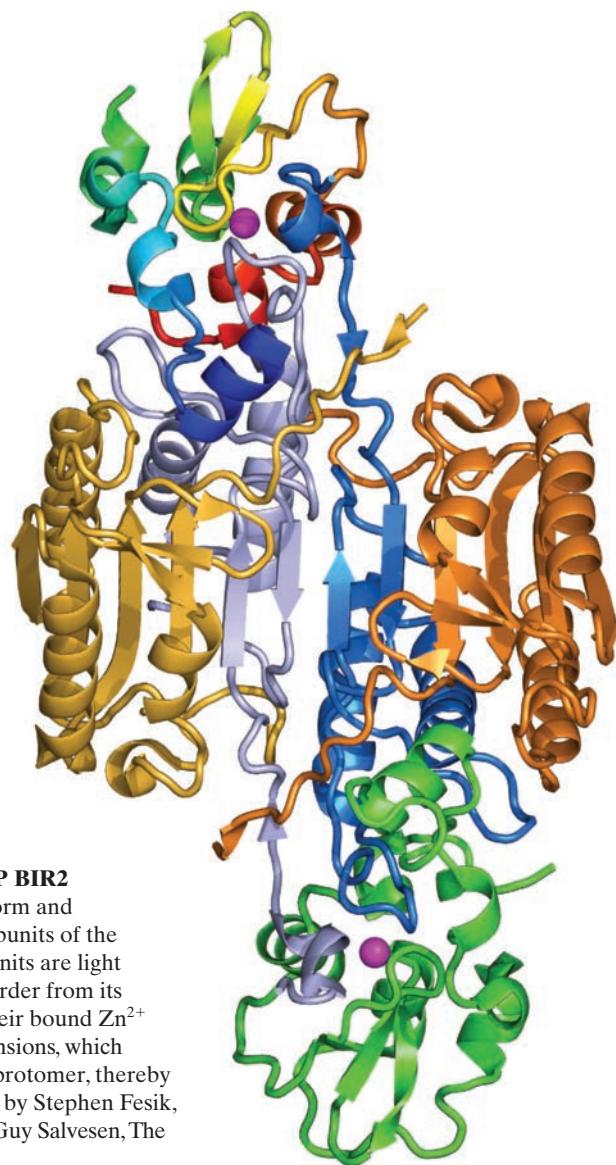


Figure 34-128 X-ray structure of caspase-3 in complex with the XIAP BIR2 domain and its N-terminal extension.

domain and its N-terminal extension. The complex is drawn in ribbon form and viewed along its 2-fold axis (as is caspase-7 in Fig. 34-119). The two α subunits of the heterotetrameric caspase-3 are light and dark orange and its two β subunits are light and dark blue. The upper globular BIR2 domain is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), the lower one is green, and their bound Zn^{2+} ions are represented by magenta spheres. Each of their N-terminal extensions, which begin with an α helix, are bound across the active site face of a caspase protomer, thereby blocking the binding of substrate proteins. [Based on an X-ray structure by Stephen Fesik, Abbott Laboratories, Abbott Park, Illinois, and Robert Liddington and Guy Salvesen, The Burnham Institute, La Jolla, California. PDBid 1I3O.]

CHAPTER SUMMARY

1 Chromosome Structure There are two types of eukaryotic chromatin: euchromatin, which can be transcriptionally active, and the more densely packed heterochromatin, which is transcriptionally inactive. Chromatin consists of DNA and proteins, the majority of which are the highly conserved histones. Chromatin is structurally organized in a hierarchical manner. In the first level of chromatin organization, ~ 200 bp of DNA are doubly wrapped around a histone octamer, $(H2A)_2(H2B)_2(H3)_2(H4)_2$, to form a nucleosome. Each nucleosome is associated with one molecule of histone H1 or H5. DNA replication causes the parental nucleosomes to be randomly distributed between the daughter duplexes. The assembly of nucleosomes from their components is mediated by the molecular chaperone nucleoplasm. The affinity of histone octamers for DNA varies with the DNA's sequence. In the second level of chromatin organization, the 10-nm-diameter nucleosome filaments coil into 30-nm-thick fibers. Then, in the

third and final level of chromatin organization, the 30-nm fibers form 15- to 30- μ m-long radial loops that project from the axis of the metaphase chromosome. This accounts for DNA's packing ratio of >8000 in the metaphase chromosome. Sister chromatids are held together by cohesin and the radial loops of metaphase proteins are organized by condensin, both of which are SMC proteins. The larvae of certain dipteran flies, including *Drosophila*, contain banded polytene chromosomes, which consist of up to 1024 identical DNA strands in parallel register. Chromosomes occupy largely nonoverlapping territories in the nucleus during interphase.

2 Genomic Organization The complexity of a DNA sample can be determined from its renaturation rate through C_0t curve analysis. Eukaryotic DNAs have complex C_0t curves that arise from the presence of unique, moderately repetitive, and highly repetitive sequences, as well as from inverted repeats. The function of inverted repeats, which form foldback

structures, is unknown. Highly repetitive sequences, which occur in the heterochromatic regions near the chromosomal centromeres, probably function to align homologous chromosomes during meiosis and/or to facilitate their recombination. Moderately repetitive DNAs, which consist largely of inactive retrotransposons, mainly LINEs and SINEs, make up ~42% of the human genome. For the most part, they have unknown functions; they may simply be selfish DNA. Expressed sequences comprise only 1.1 to 1.4% of the human genome, which has made the identification of genes an uncertain process. However, CpG islands are often associated with the 5' ends of genes.

Around 23,000 putative genes have been identified in the human genome. Some of them are transcribed to noncoding RNAs. The structural genes in the human genome have been classified according to function through sequence comparisons. Around 42% of them have unknown functions. The genes specifying rRNAs and tRNAs are organized into tandemly repeated clusters. The rDNA condenses to form nucleoli, the sites of rRNA transcription by RNA polymerase I and of partial ribosomal assembly. The 5S RNA and tRNAs are transcribed outside the nucleoli by RNA polymerase III. The genes specifying histones, which are required in large quantities only during S phase of the cell cycle, are the only repeated structural genes. The identity of a series of repeated genes is probably maintained through unequal crossing-over and/or gene conversion. Certain genes are amplified such as the *Xenopus* genes for rRNAs during oogenesis, chorion genes in *Drosophila*, and genes targeted by cancer chemotherapy.

Many families of genes specifying related proteins are clustered together. In mammals, the gene clusters encoding the α - and β -like hemoglobin subunits occur on separate chromosomes. Nevertheless, all vertebrate globin genes have the same exon–intron structure: three exons separated by two introns. The thalassemias are inherited diseases caused by the genetic impairment of hemoglobin synthesis. Many α -thalassemias are caused by the deletion of one or more of the α -globin genes, whereas many β -thalassemias arise from point mutations that affect the transcription or the post-transcriptional processing of the β -globin mRNAs.

3 Control of Expression Heterochromatin may be sub-classified as constitutive heterochromatin, which is never transcriptionally active, and facultative heterochromatin, whose activity varies in a tissue-specific manner. The Barr bodies in the cells of female mammals constitute a common form of facultative heterochromatin: One of each cell's two X chromosomes is permanently condensed via the binding of *Xist* RNA and epigenetically confers its state of inactivity on its progeny through histone modification and DNA methylation. Active chromatin has a relatively open structure that makes it available to the transcriptional machinery. Two well-characterized examples of transcriptionally active chromatin are the chromosome puffs that emanate from single bands in polytene chromosomes and the lampbrush chromosomes of amphibian oocytes.

The differential protein synthesis characteristic of the cells in a multicellular organism largely stems from the selective transcription of the expressed genes. The first step in the transcriptional initiation of RNAP II-transcribed genes is often the binding of TATA-box binding protein (TBP) to the promoter's TATA box, which is located around position -27. This is followed by the addition of TBP-associated factors (TAFs)

to form transcription factor IID (TFIID), together with general transcription factors (GTFs) and RNAP II to form the preinitiation complex (PIC), which is capable of a basal rate of transcription. Several TAFs assume the histone fold and associate with one another in TFIID, as do histones in the histone octamer. TBP, together with other GTFs, is also required for the transcriptional initiation of class I and III genes. Promoters of class II genes that lack a TATA box often contain an initiator (Inr) sequence that spans the transcription start site and may also have a downstream promoter element (DPE).

The cell-specific expression of a gene is mediated by the gene's promoter and enhancer elements. Consequently, cells contain specific upstream transcription factors that recognize these genetic elements. For example, Sp1 binds to the GC box that precedes many genes. Likewise, steroid hormones bind to their cognate receptors, which in turn bind to specific enhancers so as to modulate the transcriptional activity of the associated gene. The cooperative binding of several transcriptional factors to their target promoter and enhancer sites stimulates the associated PIC to increase the rate at which it initiates the transcription of the associated gene. The binding of transcription factors to a silencer represses the transcription of the associated gene. Several transcription factors may bind an enhancer and associate with architectural factors and coactivators to form an enhanceosome. Many transcription factors have two domains, a DNA-binding domain targeted to a specific sequence, and an activation domain, which interacts with the PIC in a largely nonspecific manner, often via a negatively charged surface region. Eukaryotic transcription factors have a great variety of DNA-binding motifs, including several types of zinc fingers, the bZIP motif, and the bHLH/Z motif. Many transcription factors, including those with the latter two types of motif, dimerize through the formation of a leucine zipper. Nuclear factor κ B (NF- κ B) is activated in the cytoplasm by the destruction of its bound inhibitor I κ B, whereon NF- κ B is translocated to the nucleus, where it binds to a κ B DNA segment so as to activate the associated gene. Mediator is an ~20-subunit yeast complex that binds to the RNAP II β' subunit's C-terminal domain (CTD), where it acts as a central scaffold around which the PIC assembles and influences RNAP II's activity through its binding of DNA-bound transcriptional regulators.

Transcriptionally poised or active genes contain nuclelease-hypersensitive sites that occur in nucleosome-free regions of DNA. Nuclease hypersensitivity is conferred on DNA by the binding of specific proteins that presumably make the genes accessible to the proteins mediating transcriptional initiation. This is largely due to the presence of the nonhistone, DNA-binding, high mobility group (HMG) proteins, which are architectural proteins that function to activate gene expression by decondensing chromatin and recruiting transcription factors. RNAP III transcribes through nucleosomes by inducing the histone octamers it encounters to step around it. In contrast, the elongation factor FACT extracts an H2A/H2B single dimers from nucleosomes associated with RNAP II-transcribed genes, which allows RNAP II to transcribe through the remaining hexosomes. Locus control regions (LCRs), which are DNase I-hypersensitive sites that function to suppress position effects, that is, the encroachment of heterochromatin, are activated by proteins that are expressed only in specific cell lineages such as erythroid cells. Insulators are DNA segments that, through the binding of specific proteins, inhibit heterochromatin from spreading into neighboring segments of

euchromatin and prevent regulatory elements outside the region controlled by the insulator from influencing the expression of the genes inside the region.

The transcriptional machinery gains access to the DNA packaged by chromatin through the post-transcriptional modification of the core histones' N-terminal tails and the ATP-driven remodeling of chromatin. The modifications to which histone N-terminal tails are subject include the acetylation/deacetylation of specific Lys side chains, the methylation/demethylation of specific Lys and Arg side chains, the phosphorylation/dephosphorylation of specific Ser side chains, and the ubiquitination of specific Lys side chains. There appears to be a histone code in which specific modifications, acting sequentially or in combination, evoke certain chromatin-based functions that result in unique biological outcomes such as transcriptional activation or silencing. In addition, some of these histone modifications may act as epigenetic markers through which cells confer their identities on their progeny. Histone acetylation is catalyzed by histone acetyltransferases (HATs) that are components of multisubunit transcriptional activators such as SAGA, PCAF, and TFIID, which all contain histonelike TAFs. Nearly all HAT-associated coactivators contain bromodomains that specifically bind acetylated histone Lys residues and hence are likely to recruit HATs to acetylate the N-terminal tails of nearby nucleosomes. Histone deacetylases (HDACs), many of which are also members of multisubunit complexes, serve as transcriptional corepressors. Histone methylation is catalyzed by histone methyltransferases (HMTs), which often function as transcriptional corepressors. Methylated histones are recognized by chromodomains such as that in heterochromatin protein 1 (HP1). The spreading of heterochromatin appears to be mediated, at least in part, by HP1's recruitment of the HMT Suv39, which methylates nearby nucleosomes such that additional HP1 can bind to them, etc. Methylated Lys residues are hydrolytically demethylated whereas methylated Arg residues are oxidatively demethylated. Histone monoubiquitination functions as an essential transcriptional regulator. Chromatin-remodeling complexes, such as yeast SWI/SNF and RSC, contain helicase-like ATPases that, it appears, "walk" up the DNA in a nucleosome so as to decrease its helical twist. The resulting DNA distortion, it is postulated, diffuses around the nucleosome in a wave that locally and transiently releases the DNA from the histone octamer, thereby permitting the nucleosome to slide along the DNA and providing transcriptional activators access to their target sequences that would otherwise be sequestered by the nucleosome.

Other forms of selective gene expression in eukaryotes include the use of alternative initiation sites in a single gene, the selection of alternative splice sites, the regulation of mRNA translocation across the nuclear membrane, the control of mRNA degradation, the control of translational initiation rates, and the selection of alternative post-translational processing pathways.

4 Cell Differentiation and Growth Embryogenesis occurs in four stages: cleavage, gastrulation, organogenesis, and maturation and growth. One of the most striking characteristics of embryological development is that cells become progressively and irreversibly committed to specific lines of development. The signals that trigger developmental changes, which are recognized over great evolutionary distances, may be transmitted through direct intercellular contacts or from

the gradients of substances, known as morphogens, released by other embryonic cells. Developmental signals act combinatorially; that is, the developmental fate of a specific tissue is determined by several not necessarily unique developmental stimuli. In *Drosophila*, early embryonic development is governed by maternal-effect genes whose distribution imposes the embryo's spatial coordinate system. These encode transcription factors that regulate the expression of gap genes, which in turn regulate the expression of pair-rule genes, which in turn regulate the expression of segment polarity genes. Sequentially finer domains of the embryonic body are thereby defined in a way that specifies the number and polarity of the larval and adult body segments. Homeotic selector or *Hox* genes, whose mutations transform one body part into another, then regulate the differentiation of the individual segments. These regulatory genes, which occur in two gene clusters, are, as the preceding genes, selectively expressed in the embryonic tissues whose development they control. They have closely related base sequences that encode ~60-residue polypeptide segments known as homeodomains, which bind their target DNA sequences in a manner similar to but distinct from that of the homologous HTH module. In vertebrates, *Hox* genes occur in four clusters and likewise control development.

Cancers result from specific genetic alterations to cells. The types of genetic changes that give rise to malignancies include the generation of altered proteins such as a Ras variant that lacks GTPase activity; altered regulatory sequences that, for example, result in the overexpression of key transcription factors; the loss of degradation signals that cause an oncogene protein such as Jun to be degraded abnormally slowly; chromosomal rearrangements that place proto-oncogenes such as *c-myc* under the control of inappropriately active regulatory sequences; gene amplification that results in the overexpression of a proto-oncogene; the insertion of viral genome into a chromosome such that a proto-oncogene is brought under the control of viral regulatory sequences; the inappropriate activation or inactivation of chromatin modification enzymes such as HAT and HDAC; and the loss or inactivation of tumor suppressor genes such as those encoding p53 and pRb or the proteins with which they interact. The mutations causing these gene alterations often arise from the actions of carcinogens on cellular DNA.

The progression of a cell through the cell cycle is regulated mainly by the presence of the appropriate mitogens together with a series of checkpoints that monitor the cell's health as well as its progress through the cell cycle. Checkpoints arrest the cell cycle until the proper conditions for its progression are met, for example, that the replication of DNA has been successfully completed and it is undamaged. The cell cycle is characterized by the accumulation of cyclins, which abruptly disappear at the end of mitosis. For example, M phase is induced when cyclin B combines with Cdc2 to form cyclin-dependent kinase 1 (Cdk1), which is preceded by Cdc2's phosphorylation at Thr 161 by Cdk-activating kinase (CAK) and succeeded by Cdc2's inactivating phosphorylation at Thr 14 and Tyr 15 by Wee1 and Myt1. At the G₂/M boundary, Thr 14 and Tyr 15 are rapidly dephosphorylated by Cdc25C, yielding active Cdk1, which phosphorylates a variety of nuclear proteins. The structures of Cdk1s resemble those of other protein kinases. However, cyclin binding to a Cdk and its phosphorylation at Thr 160 conformationally reorganize its active site. Members of the Kip/Cip family, such as p21, inhibit most Cdk-cyclin complexes

except Cdk4/6–cyclin D, whereas members of the INK4 family, such as p16^{INK4a}, inhibit Cdk4/6–cyclin D. Cell cycle arrest at the G₂ checkpoint is initiated by sensor proteins that bind to damaged and unreplicated DNA. These activate ATM and ATR to respectively phosphorylate Chk2 and Chk1, which phosphorylate Cdc25C, thereby providing a binding site for 14-3-3 proteins such that Cdc25C is sequestered in the cytoplasm, where it cannot dephosphorylate and hence activate Cdc2.

p53, a tumor suppressor that is implicated in ~50% of human cancers, is bound by Mdm2, which ubiquitinates it so as to mark it for destruction in the proteasome. Hence the cell normally has a low level of p53. However, when p53 is phosphorylated by ATM or Chk2, it no longer binds to Mdm2 and thereupon transactivates the expression of p21, which binds to several Cdk–cyclin complexes and to PCNA, thereby inhibiting both the G₁/S and G₂/M transitions and DNA replication. If the cell is irreparably damaged, p53 induces it to undergo apoptosis, thereby preventing the proliferation of potentially cancerous cells. p53 also inhibits gene expression by activating the transcription of miRNAs, which are targeted to the mRNAs of a variety of pro-proliferative proteins, as well as lincRNA-p21, which in complex with hnRNP-K, represses the transcription of numerous genes. The X-ray structure of p53's DNA-binding core in complex with its target DNA reveals that many of its residues that participate in DNA binding are frequently mutated in tumors. p53 is also activated by a variety of pathways such as MAP kinase cascades, the activation of ATR by DNA-damaging agents, and a variety of post-translational modifications. Thus, p53 is the recipient of numerous intracellular signals and activates a variety of downstream regulators.

The tumor suppressor pRb, a regulator of the cell cycle's G₁/S transition, functions by inhibiting E2F, a transcription factor for many proteins required for S-phase entry. pRb, a phosphoprotein that is phosphorylated at numerous Ser/Thr sites by various Cdk–cyclin complexes, must be in its hypophosphorylated form to bind E2F. A variety of proteins that have an LXCXE sequence motif bind to pRb at a separate site from E2F on pRb's pocket domain, a major site of genetic alteration in tumors. Viral proteins with the LXCXE motif, including adenovirus E1A and papillomavirus E7, cause pRb to release its bound E2F, thereby driving the infected cell into S phase, which facilitates viral DNA replication. The histone deacetylases HDAC1 and HDAC2 each contain an LXCXE motif, which suggests that pRb functions to recruit these proteins to E2F's target promoters, thereby deactivating them. The SWI/SNF homologs BRM and BRG1, which both have LXCXE motifs, can bind to pRb simultaneously with HDACs, which suggests that these chromatin-remodeling complexes are recruited to E2F promoters, where they facilitate the action of HDACs.

Apoptosis (programmed cell death) occurs normally during embryogenesis and in many adult processes. In fact, it is the default option for metazoan cells. Insufficient apoptosis can cause autoimmune diseases and cancer, whereas inappropriate apoptosis is responsible for several neurodegenerative diseases and much of the damage caused by stroke and heart attacks. In apoptosis, the cell dismantles itself in an orderly program to yield membrane-enclosed apoptotic bodies that are phagocytosed by surrounding cells without inducing an inflammatory response. The executioners in apoptosis are cysteine proteases known as caspases that specifically cleave polypeptides after Asp residues. Caspases are synthesized as zymogens called pro-caspases that are proteolytically activated via apoptotic pathways, ending in the activation of initiator caspases that activate effector caspases, which cleave a wide variety of cellular proteins. Among the latter is ICAD, which is an inhibitor of caspase-activated DNase (CAD) that, in the absence of ICAD, functions to fragment the cell's DNA.

In the extrinsic pathway of apoptosis (death by commission), a trimeric transmembrane cytokine of the tumor necrosis factor (TNF) family, such as Fas ligand (FasL), which is on the inducing cell, binds to a transmembrane so-called death receptor of the TNFR family, such as Fas, which is on the apoptotic cell. The binding of trimeric ligand to a death receptor causes its cytoplasmic death domain (DD) to form a trimer to which three molecules of the adaptor protein FADD then bind via their DDs. FADD, in turn, recruits pro-caspases-8 and -10 via interactions between the two proteins' death effector domains (DEDs) to form the death-inducing signal complex (DISC). This results in the proteolytic autoactivation of the bound pro-caspases-8 and -10, which then activate pro-caspase-3, an effector caspase. In the intrinsic pathway (death by omission), pro-apoptotic members of the Bcl-2 family are activated in various ways, including the withdrawal of cytokines and contact with other cells, to induce the release of cytochrome *c* from the mitochondrion. The cytochrome *c* binds to the scaffolding protein Apaf-1 to form a wheel-shaped heptameric complex called the apoptosome. The apoptosome binds several molecules of pro-caspase-9 through interactions between the two proteins' CARD domains, which activates pro-caspase-9 to activate pro-caspase-3. Pro-apoptotic Bcl-2 family members are kept in check through their heterodimerization with anti-apoptotic Bcl-2 family members. In addition, the members of the IAP family inhibit apoptosis by directly binding to caspases so as to block their active sites and, in some cases, also ubiquitinating them so as to mark them for destruction in the proteasome. Smac/DIABLO, which is released from the mitochondrion together with cytochrome *c*, reverses this inhibition by binding to IAPs, thereby permitting apoptosis to commence.

REFERENCES

General

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Molecular Biology of the Cell* (5th ed.), Chapters 4, 7, 17, 18, 20, and 22, Garland Publishing (2008).
- Brown, T.A., *Genomes* (3rd ed.), Chapters 7, 10, 11, and 14, Garland Science (2007).
- Latchman, D.S., *Gene Control*, Garland Science (2010).

- Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P., *Molecular Cell Biology* (6th ed.), Chapters 6–8, and 22 (2008).
- Morgan, D.O., *The Cell Cycle. Principles of Control*, Sinauer Associates (2007).
- Sumner, A.T., *Chromosomes, Organization and Function*, Blackwell Science (2003).

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M., and Losick, R., *Molecular Biology of the Gene* (6th ed.), Chapters 7, 12, 17, and 19, Cold Spring Harbor Laboratory Press (2008).

Chromosome Structure

Annunziato, A.T., Split decision: What happens to nucleosomes during DNA replication, *J. Biol. Chem.* **280**, 12065–12068 (2005).

Dehghani, H., Dellaire, G., and Bazett-Jones, D.P., Organization of chromatin in the interphase mammalian cell, *Micron* **36**, 95–108 (2005).

English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E.A., and Tyler, J.K., Structural basis for the histone chaperone activity of Asf1, *Cell* **127**, 495–508 (2006); and Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., and Senda, T., Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4, *Nature* **446**, 338–341 (2007).

Harp, J.M., Hanson, B.L., Timm, D.E., and Bunick, G.J., Asymmetries in the nucleosome core particle at 2.5 Å resolution, *Acta Cryst. D* **56**, 1513–1534 (2000).

Kornberg, R.D., Chromatin structure: a repeating unit of histones and DNA, *Science* **184**, 868–871 (1974). [The classic paper first indicating the constitution of nucleosomes.]

Kornberg, R.D. and Lorch, Y., Chromatin rules, *Nature Struct. Mol. Biol.* **14**, 986–988 (2007).

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J., Crystal structure of the nucleosome particle at 2.8 Å resolution, *Nature* **389**, 251–260 (1997); Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J., Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution, *J. Mol. Biol.* **319**, 1097–1113 (2002); and Richmond, T.J. and Davey, C.A., The structure of DNA in the nucleosome core, *Nature* **423**, 145–150 (2003).

Luger, K., Structure and dynamic behaviour of nucleosomes, *Curr. Opin. Genet. Dev.* **13**, 127–135 (2003); and Akey, C.W. and Luger, K., Histone chaperones and nucleosome assembly, *Curr. Opin. Struct. Biol.* **13**, 6–14 (2003).

Masse, J.E., Wong, B., Yen, Y.-M., Allain, F.H.-T., Johnson, R.C., and Feigon, J., The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding, *J. Mol. Biol.* **323**, 263–284 (2002).

Merika, M. and Thanos, D., Enhanceosomes, *Curr. Opin. Genet. Dev.* **11**, 205–208 (2001).

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M., Crystal structure of globular domain of histone H5 and its implications for nucleosome binding, *Nature* **362**, 219–223 (1993); and Ramakrishnan, V., Histone H1 and chromatin higher-order structure, *Crit. Rev. Euk. Gene Exp.* **7**, 215–230 (1997).

Robinson, P.J.J. and Rhodes, D., Structure of the ‘30 nm’ chromatin fibre: A key role for linker histone, *Curr. Opin. Struct. Biol.* **16**, 336–343 (2006), and Robinson, P.J.J., Fairall, L., Huynh, V.A.T., and Rhodes, D., EM measurements define the dimensions of the “30-nm” chromatin fiber: Evidence for a compact interdigitated structure, *Proc. Natl. Acad. Sci.* **103**, 6506–6511 (2006).

Sarma, K. and Reinberg, D., Histone variants meet their match, *Nature Rev. Mol. Cell Biol.* **6**, 139–149 (2005).

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J., X-ray structure of a tetranucleosome and its implications for the chromatin fibre, *Nature* **436**, 138–141 (2005).

Segal, E., Fondue-Mittendorf, Y., Chen, L., Thåström, A.C., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J., A genomic code for nucleosome positioning, *Nature* **442**, 772–778 (2006).

Segal, E. and Widom, J., Poly(dA:dT) tracts: major determinants

of nucleosome organization, *Curr. Opin. Struct. Biol.* **19**, 65–71 (2009).

Shintomi, K. and Hirano, T., Sister chromatid resolution: a cohesin releasing network and beyond, *Chromosoma* **119**, 459–467 (2010); and Hirano, T., At the heart of the chromosome: SMC proteins in action, *Nature Rev. Mol. Cell Biol.* **7**, 311–322 (2006).

Speicher, M.R. and Carter, N.G., The new cytogenetics: blurring the boundaries with molecular biology, *Nature Rev. Genet.* **6**, 782–792 (2005).

Talbert, P.B. and Henikoff, S., Histone variants—ancient wrap artists of the epigenome, *Nature Rev. Mol. Biol.* **11**, 264–275 (2010).

Thomas, J.O. and Travers, A.A., HMG1 and 2, and related ‘architectural’ DNA-binding proteins, *Trends Biochem. Sci.* **26**, 167–174 (2001).

Van Holde, K. and Zlatanova, J., Chromatin fiber structure: Where is the problem now? *Semin. Cell Dev. Biol.* **18**, 651–658 (2007).

Genomic Organization

Berry, M., Grosveld, F., and Dillon, N., A single point mutation is the cause of the Greek form of hereditary persistence of fetal hemoglobin, *Nature* **358**, 499–502 (1992).

Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M., *Mobile DNA II*, Chapters 35, 47, 48, and 49, ASM Press (2002). [Discussions of transposable elements in eukaryotic genomes.]

Hamlin, J.L., Leu, T.-H., Vaughn, J.P., Ma, C., and Dijkwel, P.A., Amplification of DNA sequences in mammalian cells, *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 203–239 (1991).

International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921 (2001); and Venter, J.C., et al., The sequence of the human genome, *Science* **291**, 1304–1351 (2001). [The landmark papers describing the base sequence of the human genome. They contain descriptions of repeating elements and gene distributions in the human genome.]

Kafatos, F.C., Orr, W., and Delidakis, C., Developmentally regulated gene amplification, *Trends Genet.* **1**, 301–306 (1985).

Li, W.-H., Gu, Z., Wang, H., and Nekrutenko, A., Evolutionary analysis of the human genome, *Nature* **409**, 847–849 (2001). [A survey of repetitive elements in the human genome.]

Mandal, R.K., The organization and transcription of eukaryotic ribosomal RNA genes, *Prog. Nucleic Acid Res. Mol. Biol.* **31**, 115–160 (1984).

Mattick, J.S., Taft, R.J., and Faulkner, G.J., A global view of genomic information—moving beyond the gene and the master regulator, *Trends Genet.* **26**, 21–28 (2009); and Mattick, J.S., The genetic signatures of noncoding RNAs, *PLoS Genet.* **5**, e1000459 (2009).

Maxson, R., Cohn, R., and Kedes, L., Expression and organization of histone genes, *Annu. Rev. Genet.* **17**, 239–277 (1983).

Nei, M. and Roychoudhury, A.K., Evolutionary relationships of human populations on a global scale, *Mol. Biol. Evol.* **10**, 927–943 (1993).

Orgel, L.E. and Crick, F.H.C., Selfish DNA: the ultimate parasite, *Nature* **284**, 604–607 (1980).

Orr-Weaver, T.L., *Drosophila* chorion genes: Cracking the eggshell’s secrets, *BioEssays* **13**, 97–105 (1991).

Saccone, C. and Pesole, G., *Handbook of Comparative Genomics. Principles and Methods*, Wiley-Liss (2003).

Schimke, R.T., Gene amplification in cultured cells, *J. Biol. Chem.* **263**, 5989–5992 (1988).

Stamatoyannopoulos, G., Majerus, P.W., Perlmutter, R.M., and Varma, H. (Eds.), *The Molecular Basis of Blood Diseases* (3rd

- ed.), Chapters 2–5, Elsevier (2001). [Discusses hemoglobin genes and their normal and thalassemic expression.]
- Südhof, T.C., Goldstein, J.L., Brown, M.S., and Russell, D.W., The LDL receptor gene: a mosaic of exons shared with different proteins, *Science* **228**, 815–828 (1985).
- Wainscoat, J.S., Hill, A.V.S., Boyce, A.L., Flint, J., Hernandez, M., Thein, S.L., Old, J.M., Lynch, J.R., Falusi, A.G., Weatherall, D.J., and Clegg, J.B., Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms, *Nature* **319**, 491–493 (1986).
- Weatherall, D.J., Clegg, J.B., Higgs, D.R., and Wood, W.G., The hemoglobinopathies, Chapter 181 in Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Contains a discussion of the thalassemias.]
- Willingham, A.T. and Gingeras, T., TUF love for “junk” DNA, *Cell* **125**, 1215–1220 (2006).

Control of Expression

- Andel, F., III, Ladurner, A.G., Inouye, C., Tjian, R., and Nogales, E., Three-dimensional structure of the human TFIID–IIA–IIB complex, *Science* **286**, 2153–2156 (1999).
- Andres, A.J. and Thummel, C.S., Hormones, puffs and flies: the molecular control of metamorphosis by ecdysone, *Trends Genet.* **8**, 132–138 (1992).
- Asturias, F.J., Chung, W.-H., Kornberg, R.D., and Lorch, Y., Structural analysis of the RSC chromatin-remodeling complex, *Proc. Natl. Acad. Sci.* **99**, 13477–13480 (2002).
- Bach, I. and Ostendorff, H.P., Orchestrating nuclear functions: Ubiquitin sets the rhythm, *Trends Biochem. Sci.* **28**, 189–195 (2003); and Muratani, M. and Tansey, W.P., How the ubiquitin–proteasome system controls transcription, *Nature Rev. Mol. Cell Biol.* **4**, 192–201 (2003).
- Bell, A.C., West, A.G., and Felsenfeld, G., Insulators and boundaries: Versatile regulatory elements in the eukaryotic genome, *Science* **291**, 447–450 (2001).
- Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D., FACT facilitates transcription-dependent nucleosome alteration, *Science* **301**, 1090–1093 (2003).
- Berger, S., Histone modifications in transcriptional regulation, *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).
- Berkowitz, B., Huang, D.-B., Chen-Park, F.E., Sigler, P.B., and Ghosh, G., The X-ray crystal structure of the NF-κB p50 · p65 heterodimer bound to the interferon β-κB site, *J. Biol. Chem.* **277**, 24694–24700 (2002).
- Bhaumik, S.R., Smith, E., and Shilatifard, A., Covalent modifications of histones during development and disease pathogenesis, *Nature Struct. Mol. Biol.* **14**, 1008–1016 (2007).
- Branden, C. and Tooze, J., *Introduction to Protein Structure* (2nd ed.), Chapters 9 and 10, Garland Publishing (1999).
- Burgess-Busse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G., The insulation of genes from external enhancers and silencing chromatin, *Proc. Natl. Acad. Sci.* **99**, 16433–16437 (2002).
- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D., Structural basis of transcription: An RNA polymerase II–TFIIB cocrystal at 4.5 angstroms, *Science* **303**, 983–988 (2004); Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D., Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism, *Science* **327**, 206–209 (2010); and Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P., RNA polymerase II–TFIIB structure and mechanism of transcription initiation, *Nature* **462**, 323–330 (2009).

- Cai, G., Imasaki, T., Takagi, Y., and Asturias, F.J., Mediator structural conservation and implications for the regulation mechanism, *Structure* **17**, 559–567 (2009).
- Campos, E.I. and Reinberg, D., Histones: annotating chromatin, *Annu. Rev. Genet.* **43**, 359–399 (2009).
- Chen, X., et al., Gene expression patterns in human liver cancers, *Mol. Biol. Cell* **13**, 1929–1939 (2002).
- Cheng, X., Collins, R.E., and Zhang, X., Structural and sequence motifs of protein (histone) methylation enzymes, *Annu. Rev. Biophys. Biomol. Struct.* **34**, 267–294 (2005).
- Clapier, C.R. and Cairns, B.R., The biology of chromatin remodeling complexes, *Annu. Rev. Biochem.* **78**, 273–304 (2009).
- Cohen, D.E. and Lee, J.T., X-Chromosome inactivation and the search for chromosome-wide silencers, *Curr. Opin. Genet. Dev.* **12**, 219–224 (2002).
- Conway, R.C. and Conway, J.W., General initiation factors for RNA polymerase II, *Annu. Rev. Biochem.* **62**, 161–190 (1993).
- Dotson, M.R., Yuan, C.X., Roeder, R.G., Myers, L.C., Gustafsson, C.M., Jiang, Y.W., Li, Y., Kornberg, R.D., and Asturias, F.J., Structural organization of yeast and mammalian mediator complexes, *Proc. Natl. Acad. Sci.* **97**, 14307–14310 (2000); and Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.J., Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction, *Mol. Cell* **10**, 409–415 (2002).
- Ellenberger, T.E., Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains, *Curr. Opin. Struct. Biol.* **4**, 12–21 (1994).
- Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C., The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: Crystal structure of the protein–DNA complex, *Cell* **71**, 1223–1237 (1992).
- Ferré-d’Amaré, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K., Recognition by Max of its cognate DNA through a dimeric bHLH/Z domain, *Nature* **363**, 38–45 (1993).
- Gangloff, Y.-G., Romier, C., Thuault, S., Werten, S., and Davidson, I., The histone fold is a key structural motif of transcription factor TFIID, *Trends Biochem. Sci.* **26**, 250–257 (2001).
- Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B., Crystal structure of the yeast TFIIA/TBP/DNA complex, *Science* **272**, 830–836 (1996); and Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J., Crystal structure of a yeast TFIIA/TBP/DNA complex, *Nature* **381**, 127–134 (1996).
- Grewal, S.I.S. and Moazed, D., Heterochromatin and epigenetic control of gene expression, *Science* **301**, 798–802 (2003).
- Huth, J.R., Bewley, C.A., Nissen, M.S., Evans, J.N.S., Reeves, R., Gronenborn, A.M., and Clore, G.M., The solution structure of an HMG-I(Y)–DNA complex defines a new architectural minor groove binding motif, *Nature Struct. Biol.* **4**, 657–665 (1997).
- Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R., Structure and function of the human TAF_{II}250 double bromodomain module, *Science* **288**, 1422–1425 (2000).
- Jin, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W., and Kusch, T., In and out: histone variant exchange in chromatin, *Trends Biochem. Sci.* **30**, 680–687 (2005).
- Juven-Gershon, T., Hsu, J.-Y., and Kadonaga, J.T., Perspectives on the RNA polymerase II core promoter, *Biochem. Soc. Trans.* **34**, 1047–1050 (2006).
- Khorasanizadeh, S., The nucleosome: from genomic organization to genomic regulation, *Cell* **116**, 259–272 (2004).
- Kim, T.H., Abdulleav, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B., Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome, *Cell* **128**, 1231–1245 (2007).

- Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B., Crystal structure of a yeast TBP/TATA-box complex; Kim, J.L., Nikolov, D.B., and Burley, S.K., Co-crystal structure of TBP recognizing the minor groove of a TATA element, *Nature* **365**, 512–520 and 520–527 (1993); and Nikolov, D.B. and Burley, S.K., 2.1 Å resolution refined structure of TATA box-binding protein (TBP), *Nature Struct. Biol.* **1**, 621–637 (1994).
- Klose, R.J. and Zhang, Y., Regulation of histone methylation by demethylimation and demethylation, *Nature Rev. Mol. Cell. Biol.* **8**, 307–318 (2007).
- Klug, A. and Rhodes, D., ‘Zinc fingers’: a novel protein motif for nucleic acid recognition, *Trends Biochem. Sci.* **12**, 464–469 (1987).
- Kornberg, R.D., Mediator and the mechanism of transcriptional activation, *Trends Biochem. Sci.* **30**, 235–239 (2005).
- Lee, K.K. and Workman, J.L., Histone acetyltransferase complexes: one size doesn’t fit all, *Nature Rev. Mol. Cell Biol.* **8**, 284–295 (2007).
- Leschziner, A.E., Saha, A., Wittmeyer, J., Zhang, Y., Bustamante, C., Cairns, B.R., and Nogales, E., RSC observed by electron microscopy and the orthogonal tilt reconstruction method, *Proc. Natl. Acad. Sci.* **104**, 4913–4918 (2007).
- Li, B., Carey, M., and Workman, J.L., The role of chromatin during transcription, *Cell* **128**, 707–719 (2007).
- Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G., Locus control regions, *Blood* **100**, 3077–3086 (2002).
- Liu, W.-L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., and Tjian, R., Structure of three distinct activator-TFIID complexes, *Genes Dev.* **23**, 1510–1521 (2009).
- Luisi, B.F., Xu, W.X., Otwowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B., Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA, *Nature* **352**, 497–505 (1991).
- Malik, S. and Roeder, R.G., Transcriptional regulation through mediator-like coactivators in yeast and metazoan cells, *Trends Biochem. Sci.* **25**, 277–283 (2000).
- Marmorstein, R., Structure of histone deacetylases: Insights into substrate recognition and catalysis, *Structure* **9**, 1127–1133 (2001).
- Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C., DNA recognition by GAL4: structure of a protein–DNA complex, *Nature* **356**, 408–414 (1992).
- Marmorstein, R. and Fitzgerald, M.X., Modulation of DNA-binding domains for sequence-specific DNA recognition, *Gene* **304**, 1–12 (2003).
- Marmorstein, R. and Trievel, R.C., Histone modifying enzymes: structures, mechanisms, and specificities, *Biochim. Biophys. Acta* **1789**, 58–68 (2009).
- Martin, G.M., X-Chromosome inactivation in mammals, *Cell* **29**, 721–724 (1982).
- Myers, L.C. and Kornberg, R.D., Mediator of transcriptional regulation, *Annu. Rev. Biochem.* **69**, 729–749 (2000).
- Ng, K., Pullirsch, D., Leeb, M., and Wutz, A., *Xist* and the order of silencing, *EMBO Rep.* **8**, 34–39 (2007).
- Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzin, N.V., and Laue, E.D., Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9, *Nature* **416**, 103–107 (2002).
- Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K., Crystal structure of a TFIIB–TBP–TATA-element ternary complex, *Nature* **377**, 119–128 (1995); and Tsai, F.T.F. and Sigler, P.B., Structural basis of preinitiation complex assembly on human Pol II promoters, *EMBO J.* **19**, 25–36 (2000).
- Orlando, V., Mapping chromosomal proteins *in vivo* by formaldehyde-crosslinked-chromatin immunoprecipitation, *Trends Biochem. Sci.* **25**, 99–104 (2000).
- O’Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T., X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science* **254**, 539–544 (1991).
- Panne, D., Maniatis, T., and Harrison, S.C., An atomic model of the interferon-β enhanceosome, *Cell* **129**, 1111–1123 (2007).
- Pavletich, N.P. and Pabo, C.O., Zinc finger–DNA recognition: Crystal structure of a Zif268–DNA complex at 2.1 Å, *Science* **252**, 809–817 (1991).
- Pelz, S.W., Brewer, G., Bernstein, P., Hart, P.A., and Ross, J., Regulation of mRNA turnover in eukaryotic cells, *Crit. Rev. Euk. Gene Express.* **1**, 99–126 (1991); and Atwater, J.A., Wisdom, R., and Verma, I.M., Regulated mRNA stability, *Annu. Rev. Genet.* **24**, 519–541 (1990).
- Poux, A.N., Cebrat, M., Kim, C.M., Cole, P.A., and Marmorstein, R., Structure of the GCN5 histone acetyltransferase bound to a bisubstrate inhibitor, *Proc. Natl. Acad. Sci.* **99**, 14065–14070 (2002).
- Ptashne, M. and Gann, A., *Genes & Signals*, Cold Spring Harbor Laboratory Press (2002). [Discusses mechanisms of genetic regulation.]
- Rando, O.J. and Chang, H.Y., Genome-wide views of chromatin structure, *Annu. Rev. Biochem.* **78**, 245–271 (2009).
- Reinberg, D. and Sims, R.J., III, deFACTo nucleosome dynamics, *J. Biol. Chem.* **281**, 23297–23301 (2006).
- Roth, S.Y., Denu, J.M., and Allis, C.D., Histone acetyltransferases, *Annu. Rev. Biochem.* **70**, 81–120 (2001).
- Sanders, B.D., Jackson, B., and Marmorstein, R., Structural basis for sirtuin function: What we know and what we don’t, *Biochim. Biophys. Acta* **1804**, 1604–1616 (2010).
- Schmiedeskamp, M. and Klevit, R.E., Zinc finger diversity, *Curr. Opin. Struct. Biol.* **4**, 28–35 (1994).
- Schwabe, J.W.R., Chapman, L., Finch, J.T., and Rhodes, D., The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements, *Cell* **75**, 567–578 (1993).
- Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q., Transcript elongation by RNA polymerase II, *Annu. Rev. Biochem.* **79**, 271–293 (2010).
- Shahbazian, M.D. and Grunstein, M., Functions of site-specific histone acetylation and deacetylation, *Annu. Rev. Biochem.* **76**, 75–100 (2007).
- Smale, S.T. and Kadonaga, J.T., The RNA polymerase II core promoter, *Annu. Rev. Biochem.* **72**, 449–479 (2003).
- Smith, B.C. and Denu, J.M., Chemical mechanisms of histone lysine and arginine modifications, *Biochim. Biophys. Acta* **1789**, 45–57 (2009).
- Stamatoyannopoulos, G. and Nienhuis, A.W. (Eds.), *The Regulation of Hemoglobin Switching*, The Johns Hopkins University Press (1991).
- Strahl, B.D. and Allis, C.D., The language of covalent histone modifications, *Nature* **403**, 41–45 (2000); and Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J., How chromatin-binding modules interpret histone modification: lessons from professional pocket pickers, *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2009).
- Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., and Felsenfeld, G., Chromatin remodeling by RNA polymerases, *Trends Biochem. Sci.* **29**, 127–136 (2004); and Studitsky, V.M., Kasavetis, G.A., Geiduschek, E.P., and Felsenfeld, G., Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase, *Science* **278**, 1960–1965 (1997).
- Taatjes, D.J., The human mediator complex: a versatile genome-wide regulator of transcription, *Trends Biochem. Sci.* **35**, 315–322 (2010).

- Tsai, M.J. and O'Malley, B., Molecular mechanisms of action of steroid/thyroid receptor superfamily members, *Annu. Rev. Biochem.* **63**, 451–486 (1994).
- Veenstra, G.J.C. and Wolffe, A.P., Gene-selective developmental roles of general transcription factors, *Trends Biochem. Sci.* **26**, 665–671 (2001).
- Weake, V.M. and Workman, J.L., Histone ubiquitination: triggering gene activity, *Mol. Cell* **29**, 653–663 (2008).
- Xiao, B., et al., Structure and catalytic mechanism of the human histone methyltransferase SET7/9, *Nature* **421**, 652–656 (2003).
- Xie, X., Kokubo, K., Cohen, S.L., Mirza, U.A., Hoffmann, A., Chait, B.T., Roeder, R.G., Nakatani, Y., and Burley, S.K., Structural similarities between TAFs and the heterotetrameric core of the histone octamer, *Nature* **380**, 316–322 (1996); Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.G., Buratowski, S., and Tan, S., A histone fold TAF octamer within the yeast TFIID transcriptional coactivator, *Nature Struct. Biol.* **8**, 695–700 (2001); and Werten, S., Mitschler, A., Romier, C., Gangloff, Y.-G., Thuault, S., Davidson, I., and Moras, D., Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF_{II}135) and hTAF12 (hTAF_{II}20), *J. Biol. Chem.* **277**, 45502–45509 (2002).
- Yang, X.-J. and Seto, E., Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression, *Curr. Opin. Genet. Dev.* **13**, 143–153 (2003).
- Zhao, K., Harshaw, R., Chai, X., and Marmorstein, R., Structural basis for nicotinamide cleavage and ADP-ribose transfer by NAD⁺-dependent Sir2 histone/protein deacetylases, *Proc. Natl. Acad. Sci.* **101**, 8563–8568 (2004).
- Development**
- Bate, M. and Arias, A.M. (Eds.), *The Development of Drosophila melanogaster*, Cold Spring Harbor Laboratory Press (1993).
- Davidson, E.H., *The Regulatory Genome. Gene Regulatory Networks in Development and Evolution*, Academic Press (2006).
- Ephrussi, A. and St Johnston, D., Seeing is believing: the bicoid morphogen gradient matures, *Cell* **116**, 143–152 (2004). [A historical account of the elucidation of *bicoid* gene function.]
- Fujioka, M., Emi-Sarker, Y., Yusibova, G.L., Goto, T., and Jaynes, J.B., Analysis of an *even-skipped* rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers and multi-stripe positioning by gap gene repressor gradients, *Development* **126**, 2527–2538 (1999).
- Gehring, W.J., Affolter, M., and Bürglin, T., Homeodomain proteins, *Annu. Rev. Biochem.* **63**, 487–526 (1994); and Gehring, W.J., Qian, Y.Q., Billeter, M., Furukobu-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K., Homeodomain–DNA recognition, *Cell* **78**, 211–223 (1994).
- Gilbert, S.F., *Developmental Biology* (9th ed.), Sinauer Associates (2010).
- Gurdon, J.B., The generation of diversity and pattern in animal development, *Cell* **68**, 185–199 (1992).
- Halder, G., Callaerts, P., and Gehring, W.J., Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*, *Science* **267**, 1788–1792 (1995).
- Kenyon, C., If birds can fly, why can't we? Homeotic genes and evolution, *Cell* **78**, 175–180 (1994).
- Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O., Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: A framework for understanding homeodomain–DNA interactions, *Cell* **63**, 579–590 (1990).
- Krumlauf, R., *Hox genes in development*, *Cell* **78**, 191–201 (1994).
- Lawrence, P.A., *The Making of a Fly*, Blackwell Scientific Publications (1992).
- Lawrence, P.A. and Morata, G., Homeobox genes: Their function in *Drosophila* segmentation and pattern formation, *Cell* **78**, 181–191 (1994).
- Lemons, D. and McGinnis, W., Genomic evolution of Hox gene clusters, *Science* **313**, 1918–1922 (2006).
- Le Mouellic, H., Lallemand, Y., and Brûlet, P., Homeosis in the mouse induced by a null mutation in the *Hox-3.1* gene, *Cell* **69**, 251–264 (1992).
- Maeda, R.K. and Karsch, F., The ABC of the BX-C: the bithorax complex explained, *Development* **133**, 1413–1422 (2006).
- Mann, R.S. and Morata, G., The developmental and molecular biology of genes that subdivide the body of *Drosophila*, *Annu. Rev. Cell Dev. Biol.* **16**, 243–271 (2000).
- Mavilio, F., Regulation of vertebrate homeobox-containing genes by morphogens, *Eur. J. Biochem.* **212**, 273–288 (1993).
- Nüsslein-Volhard, C., Axis determination in the *Drosophila* embryo, *Harvey Lect.* **86**, 129–148 (1992); and St. Johnston, D. and Nüsslein-Volhard, C., The origin of pattern and polarity in the *Drosophila* embryo, *Cell* **68**, 201–219 (1992). [Detailed reviews.]
- Nüsslein-Volhard, C., The identification of genes controlling development in flies and fishes (Nobel lecture); and Wieschaus, E., From molecular patterns to morphogenesis—The lessons from studies on the fruit fly *Drosophila* (Nobel lecture), *Angew. Chem. Int. Ed. Engl.* **35**, 2177–2187 and 2189–2194 (1996).
- Peel, A.D., Chipman, A.D., and Akam, M., Arthropod segmentation: beyond the *Drosophila* paradigm, *Nature Rev. Genet.* **6**, 905–916 (2005).
- Scott, M.P., Development: The natural history of genes, *Cell* **100**, 27–40 (2000).

Cancer and the Regulation of the Cell Cycle

- Adams, P.D., Regulation of the retinoblastoma tumor suppressor protein by cyclin/cdkks, *Biochim. Biophys. Acta* **1471**, M123–M133 (2001).
- Cho, Y., Gorina, S., Jeffrey, P.D., and Pavletich, N.P., Crystal structure of a p53 tumor suppressor–DNA complex: Understanding tumorigenic mutations, *Science* **265**, 346–355 (1994).
- Cooper, G.M. and Hausman, R.E., *The Cell. A Molecular Approach* (3rd ed.), Chapter 14, ASM Press (2004).
- De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O., and Kim, S.-H., Crystal structure of cyclin-dependent kinase 2, *Nature* **363**, 595–602 (1993); Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., and Pavletich, N.P., Mechanism of CDK activation revealed by the structure of a cyclin A–CDK2 complex, *Nature* **376**, 313–320 (1995); Russo, A.A., Jeffrey, P.D., Patten, A.K., Massagué, J., and Pavletich, N.P., Crystal structure of the p27^{Kip1} cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex, *Nature* **382**, 325–331 (1996); and Russo, A.A., Jeffrey, P.D., and Pavletich, N.P., Structural basis of cyclin-dependent kinase activation by phosphorylation, *Nature Struct. Biol.* **3**, 696–700 (1996).
- Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A., Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours, *Nature* **356**, 215–221 (1992).
- Green, D.R. and Kroemer, Cytoplasmic functions of the tumour suppressor p53, *Nature* **458**, 1127–1130 (2009).

- Harbour, J.W. and Dean, D.C., The Rb/E2F pathway: expanding roles and emerging paradigms, *Genes Dev.* **14**, 2393–2409 (2000).
- Harper, J.W. and Adams, P.D., Cyclin-dependent kinases, *Chem. Rev.* **101**, 2511–2526 (2001).
- Hickman, E.S., Moroni, M.C., and Helin, K., The role of p53 and pRB in apoptosis and cancer, *Curr. Opin. Genet. Dev.* **12**, 60–66 (2002).
- Huarte, M., et al., A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response, *Cell* **142**, 409–419 (2010).
- Jeffrey, P.D., Tong, L., and Pavletich, N.P., Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors, *Genes Dev.* **14**, 3115–3125 (2000).
- Johnson, D.G. and Walker, C.L., Cyclins and cell cycle checkpoints, *Annu. Rev. Pharmacol. Toxicol.* **39**, 295–312 (1999).
- Johnstone, R.W., Histone-deacetylase inhibitors: Novel drugs for the treatment of cancer, *Nature Rev. Drug Disc.* **1**, 287–299 (2002).
- Knudsen, E.S. and Knudsen, K.E., Retinoblastoma tumor suppressor: where cancer meets the cell cycle, *Exp. Biol. Med.* **231**, 1271–1278 (2006).
- Kuse, J.-P. and Gu, W., Modes of p53 regulation, *Cell* **137**, 609–622 (2009).
- Lee, E.Y.-H., Chang, C.-Y., Hu, N., Wang, Y.-C.J., Lai, C.-C., Herrup, K., Lee, W.-H., and Bradley, A., Mice deficient for Rb are nonviable and show defects in neurogenesis and hematopoiesis, *Nature* **359**, 288–294 (1992).
- Malecka, K.A., Ho, W.C., and Marmorstein, R., Crystal structure of a p53 core tetramer bound to DNA, *Oncogene* **28**, 325–333 (2009).
- Malumbres, M. and Barbacid, M., Cell cycle, CDKs and cancer: a changing paradigm, *Nature Rev. Cancer* **9**, 153–166 (2009).
- Morgan, D.O., Cyclin-dependent kinases: Engines, clocks, and microprocessors, *Annu. Rev. Cell Dev. Biol.* **13**, 261–291 (1997).
- Morris, E.J. and Dyson, N.J., Retinoblastoma protein partners, *Adv. Cancer Res.* **82**, 1–54 (2001).
- Murray, A.W., Recycling the cell cycle: cyclins revisited, *Cell* **116**, 221–234 (2004).
- Nigg, E.A., Mitotic kinases as regulators of cell division and its checkpoints, *Nature Rev. Mol. Biol.* **2**, 21–32 (2001).
- Norbury, C. and Nurse, P., Animal cell cycles and their control, *Annu. Rev. Biochem.* **61**, 441–470 (1992); and Forsburg, S.L. and Nurse, P., Cell cycle regulation in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, *Annu. Rev. Cell Biol.* **7**, 227–256 (1991).
- Pavletich, N.P., Mechanisms of cyclin-dependent kinase regulation: Structures of Cdkks, their cyclin activators, and Cip and INK4 inhibitors, *J. Mol. Biol.* **287**, 821–828 (1999).
- Pollard, T.D., and Earnshaw, W.C., *Cell Biology* (2nd ed.), Chapters 40–46, Elsevier (2008).
- Riley, R., Sontag, E., Chen, P., and Levine, A., Transcriptional control of human p53-regulated genes, *Nature Rev. Mol. Cell Biol.* **9**, 402–412 (2008).
- Russo, A.A., Jeffrey, P.D., Patten, A.K., Massagué, J., and Pavletich, N.P., Crystal structure of the p27^{Kip1} cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex, *Nature* **382**, 325–331 (1996).
- Sherr, C.J. and Weber, J.D., The ARF/p53 pathway, *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
- Vousden, K.H. and Prives, C., Blinded by the light: the growing complexity of p53, *Cell* **137**, 413–431 (2009); and Laptenko, O. and Prives, C., Transcriptional regulation by p53: one protein, many possibilities, *Cell Death Dif.* **13**, 951–961 (2006).
- Xiao, B., Spencer, J., Clements, A., Ali-Khan, N., Mittnacht, S., Broceño, C., Burghammer, M., Parrakis, A., Marmorstein, R., and Gamblin, S.J., Crystal structure of the retinoblastoma tumor suppressor protein bound to E2F and the molecular basis of its regulation, *Proc. Natl. Acad. Sci.* **100**, 2363–2368 (2003); and Lee, C., Chang, J.H., Lee, H.S., and Cho, Y., Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor, *Genes Dev.* **16**, 3199–3212 (2002).

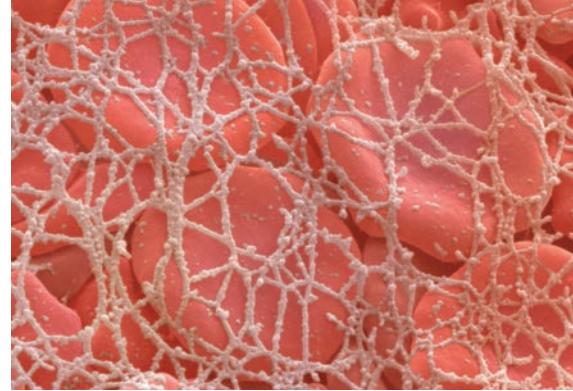
Apoptosis

- Adrain, C., Brumatti, G., and Martin, S.J., Apoptosomes: protease activation platforms to die from, *Trends Biochem. Sci.* **31**, 243–247 (2006).
- Chai, J., Wu, Q., Shiozaki, E., Srinivasula, S.M., Alnemri, E.S., and Shi, Y., Crystal structure of a pro-caspase-7 zymogen: Mechanisms of activation and substrate binding, *Cell* **107**, 399–407 (2001); and Riedl, S.J., Fuentes-Prior, P., Renatus, M., Kairies, N., Krapp, S., Huber, R., Salvesen, G.S., and Bode, W., Structural basis for the activation of human pro-caspase-7, *Proc. Natl. Acad. Sci.* **98**, 14790–14795 (2001).
- Earnshaw, W.C., Martins, L.M., and Kaufmann, S.H., Mammalian caspases: Structure, activation, substrates, and functions during apoptosis, *Annu. Rev. Biochem.* **68**, 383–424 (1999); and Grüter, M.G., Caspases: key players in programmed cell death, *Curr. Opin. Struct. Biol.* **10**, 649–655 (2000).
- Fesik, S.W., Insights into programmed cell death through structural biology, *Cell* **103**, 272–282 (2000).
- Jacobson, M.D. and McCarthy, N. (Eds.), *Apoptosis*, Oxford (2002).
- Jiang, X. and Wang, X., Cytochrome c-mediated apoptosis, *Annu. Rev. Biochem.* **73**, 87–106 (2004).
- Nagata, S., Fas ligand-induced apoptosis, *Annu. Rev. Genet.* **33**, 29–55 (1999).
- Riedl, S.J., Renatus, M., Schwarzenbacher, R., Zhou, Q., Sun, C., Fesik, S.W., Liddington, R.C., and Salvesen, G.S., Structural basis for the inhibition of caspase-3 by XIAP, *Cell* **104**, 791–800 (2001).
- Taylor, R.C., Cullen, S.P., and Martin, S.J., Apoptosis: controlled demolition at the cellular level, *Nature Rev. Mol. Cell Biol.* **9**, 231–241 (2008).
- Wei, Y., Fox, T., Chambers, S.P., Sinchak, J., Coll, J.T., Golec, J.M.C., Swenson, L., Wilson, K.P., and Charlifson, P., The structures of caspases-1, -3, -7, and -8 reveal the basis for substrate and inhibitor selectivity, *Chem. Biol.* **7**, 423–432 (2000).
- Yan, N. and Shi, Y., Mechanisms of apoptosis through structural biology, *Annu. Rev. Cell Dev. Biol.* **21**, 35–56 (2005).
- Yin, X.M. and Dong, Z. (Eds.), *Essentials of Apoptosis*, Humana Press (2003).
- Yuan, S., Yu, X., Topf, M., Ludtke, S.J., Wang, X., and Akey, C.W., Structure of an apoptosome–procaspase-9 CARD complex, *Structure* **18**, 571–583 (2010).

PROBLEMS

- 1.** What is the maximum possible packing ratio of a 10^6 -bp segment of DNA; of a 10^9 -bp segment of DNA? Assume the DNA is a 20-Å-diameter cylinder with a contour length of 3.4 Å/bp.
- 2.** The average length of linker DNA is ~ 18 bp in yeast and ~ 38 bp in humans. (a) What fraction of the DNA in each of these organisms is associated with nucleosome core particles? (b) How many nucleosomes does a diploid nucleus in each of these organisms contain (see Table 7-3).
- 3.** Explain why cleaving the DNA of a chromosome at a relatively few places causes it to fall apart, whereas subjecting it to light proteolysis has little apparent effect.
- 4.** When an SV40 minichromosome (a closed circular duplex DNA in complex with nucleosomes) is relaxed so that it forms an untwisted circle and is then deproteinized, the consequent closed circular DNA has about -1 superhelical turn for each of the nucleosomes that it originally had. Explain the discrepancy between this observation and the fact that the DNA in each nucleosome is wrapped nearly twice about its histone octamer in a left-handed superhelix.
- 5.** Explain why acidic polypeptides such as polyglutamate facilitate *in vitro* nucleosome assembly.
- *6.** Consider a 1 million-bp DNA molecule that has 1500 tandem repeats of a 400-bp sequence with the remainder of the DNA consisting of unique sequences. Sketch the $C_0 t$ curve of this DNA when it is sheared into pieces averaging 1000 bp long; when they are 100 bp long.
- 7.** Why do isolated foldback structures, when treated by an endonuclease that cleaves only single-stranded DNA and then denatured, yield complicated $C_0 t$ curves?
- 8.** During its 2-month period of maturation, the *Xenopus* oocyte synthesizes $\sim 10^{12}$ ribosomes. The consequent tremendous rate of rRNA synthesis is only possible because the normal genomic complement of rDNA has been amplified 1500-fold. (a) Why is it unnecessary to likewise amplify the genes encoding the ribosomal proteins? (b) Assuming that rRNA gene amplification occurs in a short time at the beginning of the maturation period, how long would oogenesis require if the rDNA were not amplified?
- 9. Hb Kenya** is a β -thalassemia in which the β -globin cluster is deleted between a point in the α -globin gene and the corresponding position in the β -globin gene. Describe the most probable mechanism for the generation of this mutation.
- 10.** Red-green color blindness is caused by an X-linked recessive genetic defect. Hence females rarely exhibit the red-green color-blind phenotype but may be carriers of the defective gene. When a narrow beam of red or green light is projected onto some areas of the retina of such a female carrier, she can readily differentiate the two colors but on other areas she has difficulty in doing so. Explain.
- 11.** Figure 34-58a contains a single band just above the bracketed region that increases in density as the Sp1 concentration increases. What is origin of this band?
- 12.** Is it possible for a transcription enhancer to be located within the protein-coding sequence of a gene? Explain.
- 13.** Why do the rare instances of male calico cats all have the abnormal XXY genotype?
- 14.** In *Drosophila*, an esc^- homozygote develops normally unless its mother is also an esc^- homozygote. Explain.
- 15.** The fusion of cancer cells with normal cells often suppresses the expression of the tumorigenic phenotype. Explain.

Molecular Physiology



CHAPTER 35

1 Blood Clotting

- A. Fibrinogen and Its Conversion to Fibrin
- B. Thrombin Activation and the Function of Vitamin K
- C. The Extrinsic Pathway
- D. The Intrinsic Pathway
- E. Control of Clotting
- F. Clot Lysis

2 Immunity

- A. The Adaptive Immune Response
- B. Antibody Structures
- C. Generation of Antibody Diversity
- D. T Cell Receptors
- E. The Major Histocompatibility Complex
- F. The Complement System

3 Motility: Muscles, Vesicle Transport, Cilia, and Flagella

- A. Structure of Striated Muscle
- B. Mechanism of Muscle Contraction
- C. Control of Muscle Contraction
- D. Smooth Muscle
- E. Actins in Nonmuscle Cells
- F. Myosins in Nonmuscle Cells
- G. Microtubules
- H. Microtubule-Associated Motors
- I. Bacterial Flagella

concern ourselves with “molecular physiology,” the study of the molecular basis of biological function. In particular, in what really are independent minichapters, we shall consider the workings of three of the better characterized complex biochemical systems, those mediating (1) blood clotting, (2) immunity, and (3) biological motility. Consider these sections to be this textbook’s “dessert,” an indication of modern biochemistry’s rapidly advancing capacity to explain and influence the workings of multicellular organisms. There are, of course, numerous other physiological systems that we might equally well have chosen to describe.

1 BLOOD CLOTTING

The circulatory system must be self-sealing; otherwise, continued blood loss from even the smallest injury would be life threatening. Normally, all but the most catastrophic bleeding is rapidly stopped, a process known as **hemostasis**, through several sequential but interrelated processes. Initially, an injury stimulates **platelets** (unpigmented enucleated blood cells—really fragments of giant bone marrow cells named **megakaryocytes**, each of which gives rise to $\sim 10^4$ platelets) to adhere to the walls of damaged blood vessels and then to aggregate with each other so as to form a plug that can stop minor bleeding. The adhesion is mediated by **von Willebrandt factor**, a large (up to $\sim 10^4$ kD) multimeric plasma glycoprotein of subunit mass 225 kD. This protein binds to both a specific receptor on the platelet membrane named **glycoprotein Ib (GPIb)** and to the collagen and possibly other components of the subendothelial membrane exposed by vascular injury. Adhesion induces platelet activation, which causes these cells to alter their shape, expose new molecules on their surfaces, and release several physiologically active substances, including serotonin (5-hydroxytryptamine; Section 26-4B), ADP, thrombin (Section 35-1B), and thromboxane A₂ (Section 25-7B). These substances induce further platelet activation and stimulate vasoconstriction, thereby reducing blood flow at the injury site.

Platelet aggregation is mediated by **integrin $\alpha_{IIb}\beta_3$** (integrins are cell-adhesion receptors that integrate internal and external signals). An integrin $\alpha_{IIb}\beta_3$ molecule from each of two activated platelets can bind to the same molecule of fibrinogen (a plasma protein that is the precursor of blood clots; Section 35-1A), thus causing platelet aggregation (platelet activation results in a conformation change in integrin $\alpha_{IIb}\beta_3$ that dramatically increases its affinity for fibrinogen). The aggregated platelets and the damaged tissue

Throughout this textbook we have been largely concerned with basic biochemical questions: What are the structures of biological molecules? How do enzymes work? How do cells extract free energy from nutrients? How do they synthesize biomolecules? How are these various processes controlled? How is genetic information expressed, and how is it transmitted to succeeding generations? Yet, to describe biological processes beyond the level of the single cell we must also reach an understanding of the various biochemical tasks that higher organisms apportion among their specialized tissues and how they coordinate the activities of these tissues. The elucidation of these complex processes has traditionally been the realm of physiology (the study of biological function), cell biology (the study of cell structure), and anatomy (the study of organismal structure); their molecular descriptions seemed out of reach. The enormous gains in biomedical knowledge and technology over the past few decades, however, have greatly blurred the boundaries among these disciplines and biochemistry as we have seen in our discussions of signal transduction (Chapter 19) and neurotransmission (Section 20-5). In this final chapter we shall further

initiate **blood clotting** or **coagulation**, the body's major defense against blood loss.

A blood clot (medically known as a *thrombus*) forms through the action of a bifurcated cascade of proteolytic reactions involving the participation of nearly 20 different substances, most of which are liver-synthesized plasma glycoproteins. This cascade is diagrammed in Fig. 35-1, and many of its components are listed in Table 35-1. All but two of these factors are designated by both a roman numeral and a common name, although, unfortunately, the order of the roman numerals has historical rather than mechanistic significance. Seven of the clotting factors are zymogens (inactive forms) of serine proteases that are proteolyti-

cally activated by serine proteases further up the cascade. Other clotting proteins, termed **accessory factors**, which are also activated by these serine proteases, form protease-accessory factor complexes that greatly enhance the activity of the protease. In both cases, the active form of a factor is designated by the subscript *a*.

The blood clotting system, which occurs in recognizable form in all vertebrates, contains a number of homologous serine proteases and therefore appears to have arisen through a series of gene duplications. The C-terminal ~250 residues of these proteases, which comprise their catalytically active domains, are also homologous to the pancreatic serine proteases trypsin, chymotrypsin, and elastase (Section 15-3). Like these digestive enzymes, the blood clotting proteases are activated by proteolytic cleavages (Section 15-3E). However, the clotting proteases differ from the digestive enzymes in that the zymogen-to-protease conversion only takes place in the presence of Ca^{2+} and on an appropriate phospholipid membrane (see below). The resulting N-terminal fragments are quite large (150–582 residues) and, with the

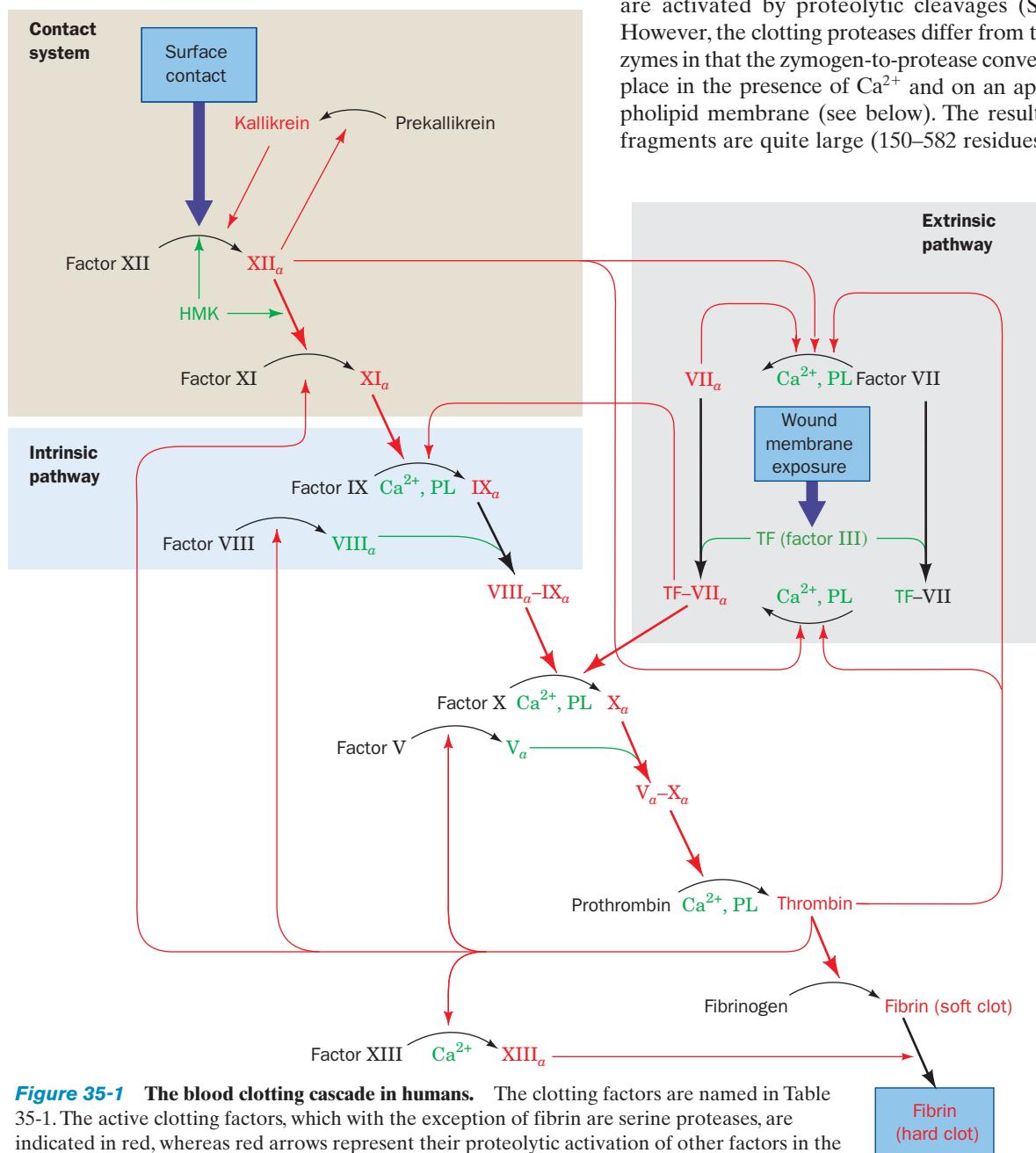


Figure 35-1 The blood clotting cascade in humans. The clotting factors are named in Table 35-1. The active clotting factors, which with the exception of fibrin are serine proteases, are indicated in red, whereas red arrows represent their proteolytic activation of other factors in the cascade. Similarly, active accessory factors, including Ca^{2+} and phospholipid membrane (PL), are indicated in green. Note the numerous feedback reactions that accelerate the clotting process and/or couple the intrinsic and extrinsic pathways.

Table 35-1 Human Blood Coagulation Factors

Factor Number	Common Name	Molecular Mass (kD)
I	Fibrinogen	340
II	Prothrombin	66
III	Tissue factor or thromboplastin	30
IV	Ca ²⁺	
V ^a	Proaccelerin	249
VII	Proconvertin	46
VIII	Antihemophilic factor	265
IX	Christmas factor	47
X	Stuart factor	50
XI	Plasma thromboplastin antecedent (PTA)	136
XII	Hageman factor	67
XIII	Fibrin-stabilizing factor (FSF)	301
	Prekallikrein	69
	High molecular weight kininogen (HMK)	70

^aFactor V was once called factor VI; consequently, there is no factor VI.

Source: Mainly Halkier, T., *Mechanisms in Blood Coagulation, Fibrinolysis, and the Complement System*, pp. 4–5, Cambridge University Press (1991).

exception of prothrombin (Section 35-1B), are linked to their C-terminal segments via disulfide bonds so that these segments do not separate upon activation. These N-terminal segments are thought to be responsible, at least in part, for the exquisite specificities of the proteolytic blood clotting factors: *Their substrates are limited, as we shall see, to the few inactive factors they function to activate.*

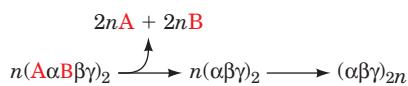
In this section we describe the chemical mechanism of blood clotting in humans from the bottom up, that is, starting with the formation of the clot itself and working backward through the sequence of activation steps leading to this process. We shall also discuss how clotting is inhibited in the absence of injury and how clots are dissolved as an injury heals.

**Figure 35-2** Scanning electron micrograph of a blood clot.

Shown are red cells enmeshed in a fibrin network. [© Andrew Syred/Photo Researchers.]

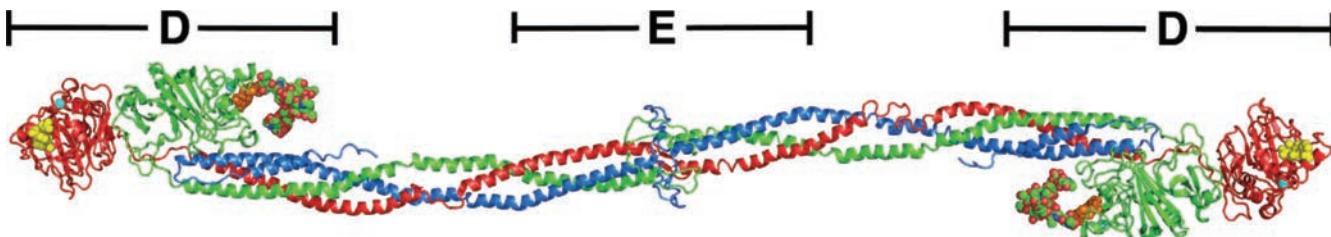
A. Fibrinogen and Its Conversion to Fibrin

Blood clots consist of arrays of cross-linked **fibrin** that form an insoluble fibrous network (Fig. 35-2). Fibrin is made from the soluble plasma protein **fibrinogen (factor I)** through a proteolytic reaction catalyzed by the serine protease **thrombin**. Fibrinogen comprises 2 to 3% of plasma protein. The 395-kD fibrinogen molecule consists of a dimer of trimeric protomers, with each protomer consisting of three nonidentical but homologous polypeptide chains, α (847 residues), β (461 residues), and γ (453 residues), and two *N*-linked 11-residue oligosaccharides each. Here A and B represent the 16- and 14-residue N-terminal **fibrinopeptides** that thrombin cleaves from fibrinogen, so a fibrin monomer is designated $(\alpha\beta\gamma)_2$. The reaction forming a blood clot from fibrinogen may therefore be represented



a. The Structure of Fibrinogen Explains How Fibrin Aggregates

The X-ray structure of human fibrinogen (Fig. 35-3), determined by Russell Doolittle, reveals that this 450-Å-long molecule forms an approximately 2-fold symmetric,

**Figure 35-3** X-ray structure of human fibrinogen as viewed along its molecular 2-fold axis. The α (blue), β (green), and γ (red) chains are drawn in ribbon form. Two tetrapeptides, Gly–His–Arg–Pro–amide and Gly–Pro–Arg–Pro–amide, which are analogs of fibrin’s N-terminal α and β knobs and which were cocrystallized with the fibrinogen, are shown in space-filling form in yellow and orange. These tetrapeptides occupy the holes (binding cavities) in the γ and β chains that the knobs occupy

when fibrinogen is converted to fibrin. A symmetry-related pair of *N*-linked carbohydrate groups are drawn in space-filling form with C green, N blue, and O red (another such pair is unobserved due to disorder). Ca²⁺ ions are represented by cyan spheres. The two fibrinogen molecules forming the dimer have somewhat different orientations, which is indicative of fibrinogen’s conformational flexibility. [Based on an X-ray structure by Russell Doolittle, University of California at San Diego. PDBid 3GHG.]

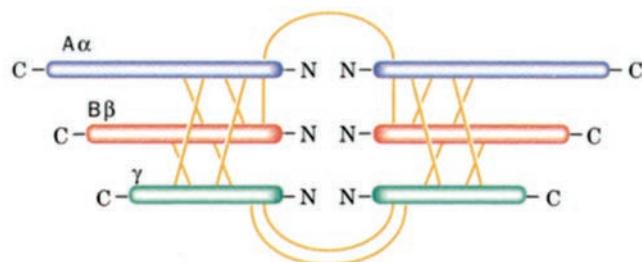


Figure 35-4 Schematic diagram of fibrinogen, $(A\alpha B\beta\gamma)_2$. The interchain disulfide bonds are represented by yellow lines.

three- to four-stranded α -helical coiled coil that has two globular nodules at each end (D fragments) and one in the middle (E fragment). The three-stranded coiled coil is formed by the central regions of the α , β , and γ subunits, and where a fourth strand exists, it is contributed by the folded back C-terminal end of the α subunit. Fibrinogen's six polypeptide chains are joined by 17 disulfide bonds, 7 within each protomer and 3 linking the two protomers (Fig. 35-4). Each D fragment contains two homologous and structurally nearly superimposable domains, one formed by the β chain and one formed by the γ chain, both of which contain deep indentations known as "holes" that

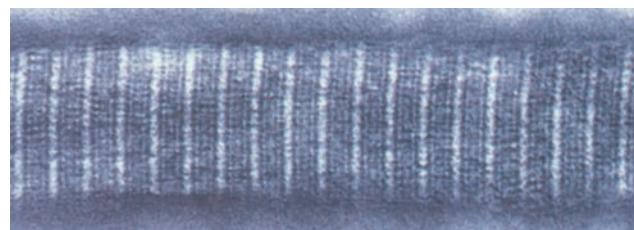


Figure 35-5 An electron micrograph of a fibrin fiber. The striations repeat every 225 Å, exactly half the length of a fibrin molecule. [Courtesy of John Weisel, University of Pennsylvania, and Carolyn Cohen, Brandeis University.]

are implicated in mediating the formation of fibrin clots (see below).

How does fibrin polymerize to form a clot? Thrombin specifically cleaves the Arg—X peptide bond (where X is Gly in most species) joining each fibrinopeptide to fibrin, sites that are located in the E region (although not visible in Fig. 35-3 due to disorder). Fibrin then spontaneously aggregates to form fibers that electron micrographs indicate have a banded structure that repeats every 225 Å (Fig. 35-5). This repeat distance is exactly half the 450-Å length of a fibrin molecule, suggesting that they associate as a half-staggered array (Fig. 35-6a). But why does fibrin aggregate

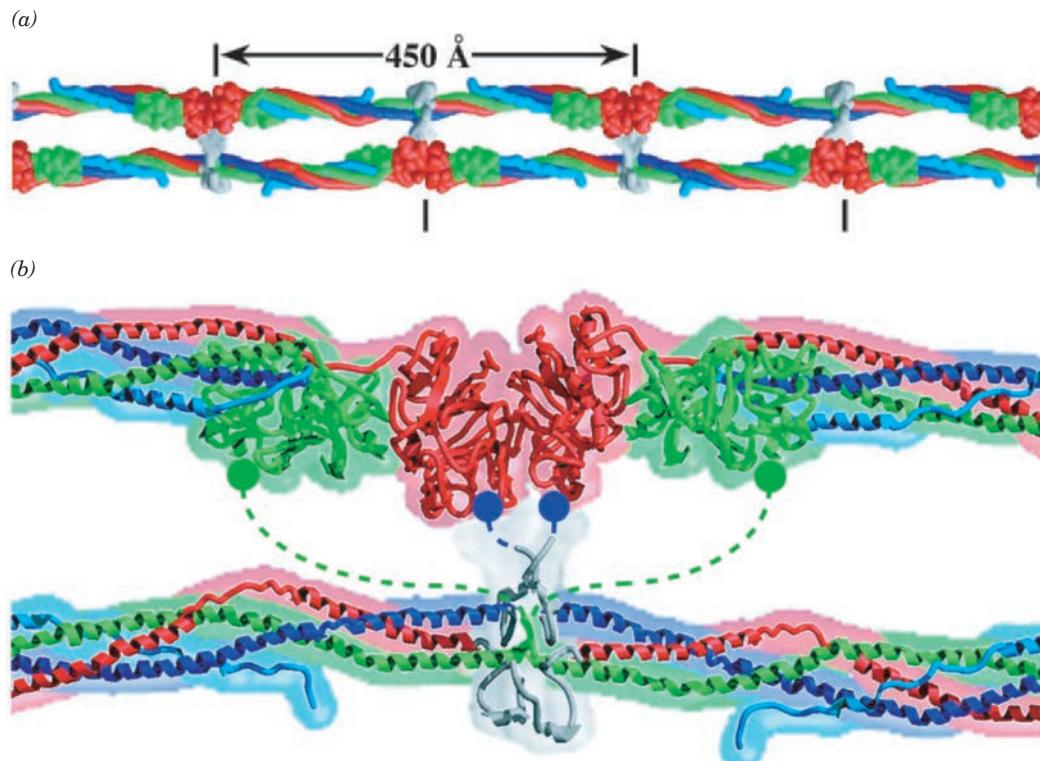


Figure 35-6 Model of the fibrin protofibril based on the X-ray structure of fibrinogen. (a) Fibrin monomers in a fibrin protofibril are aligned in a half-staggered array. Its component fibrin units, whose polypeptide chains are drawn in worm form, are colored as in Fig. 35-3. The black vertical bars mark the boundaries of the fibrin units. (b) Enlarged view of the head-to-head interaction of the γ chains and their interaction with the central portion of a second fibrin monomer. The chains are held

together predominantly by knob–hole interactions of γ and β holes and their complementary binding peptides (knobs) that are exposed by the excision of fibrinopeptides A and B. Since the α and β knobs extend from highly flexible chain segments and are therefore not visible in the X-ray structure, they are schematically drawn as blue and green balls extending from like-colored dashed lines. [Modified from drawings by Carolyn Cohen, Brandeis University. Derived from PDBid 1DEQ.]

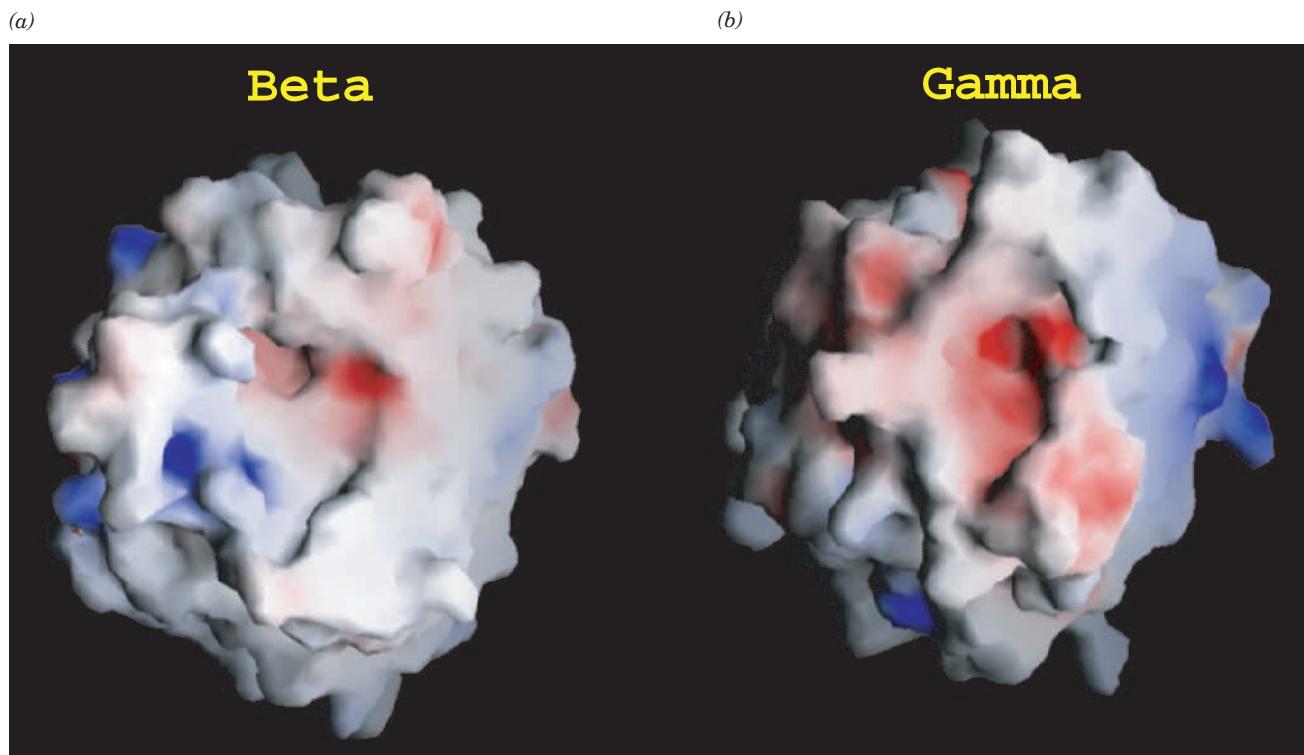


Figure 35-7 Charge distribution of the fibrinogen domains containing its binding cavities (holes). (a) That of the β chain. (b) That of the γ chain. The domains are represented by their surface diagrams colored according to their electrostatic charge

with red positively charged, white neutral, and blue negatively charged. [Courtesy of Russell Doolittle, University of California at San Diego. PDBid 1ZFA.]

while fibrinogen, which has an all but identical molecular structure, remains in solution? The main reason appears to be that the loss of the fibrinopeptides exposes otherwise masked “knobs,” formed by the new N-terminal sequences, Gly-Pro-Arg of the α chain (α knob) and Gly-His-Arg of the β chain (β knob), that mediate intermolecular associations (Fig. 35-6b).

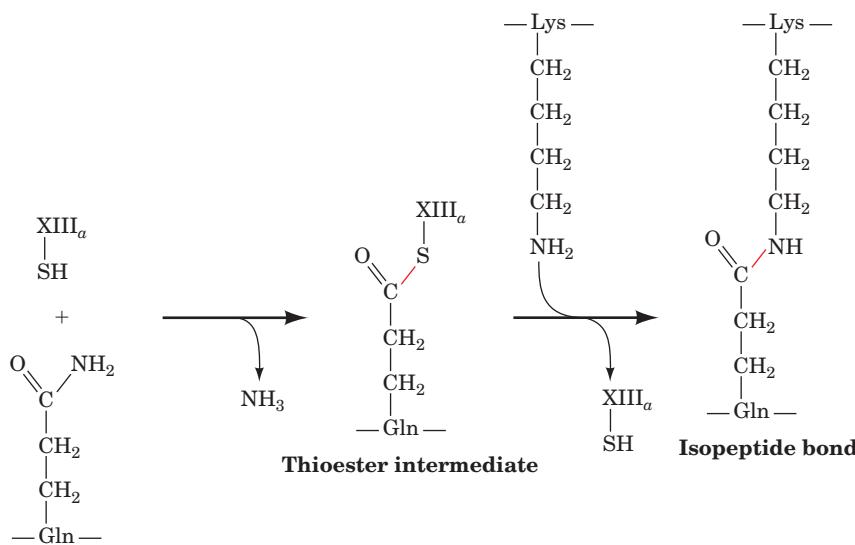
Fibrinogen aggregation is also inhibited by charge-charge repulsions: The fibrinopeptides are highly anionic, so much so that fibrinogen’s central region, where the fibrinopeptides reside, has a charge of -8 , whereas that of fibrin is $+5$. Fibrinogen’s C-terminal D segments each have a similar charge of -4 but maintain that charge in fibrin. The repulsions between fibrinogen’s like-charged segments help prevent this protein from aggregating, whereas the attractions between fibrin’s α and β knobs that extend from the central E domain and the γ and β holes in the D segments promote its specific association. Figure 35-7 shows the surface charge distribution of the D segment’s β and γ holes, indicating the locations of negative charges available to attract the positively charged knobs.

The diameters of fibrin fibers, which are fairly uniform (maximally ~ 50 nm), are important determinants of a clot’s physical properties. What controls this fiber diameter? John Weisel and Lee Makowski have shown through electron microscopy studies that fibrin fibers are uniformly

twisted. Consequently, the in-register molecules near the periphery of a twisted fiber must traverse a longer path than molecules near the fiber’s center. The degree to which a molecule can stretch therefore limits the diameter of the fiber: Molecules add to the outside of a growing fiber until the energy to stretch an added molecule exceeds its energy of binding. A similar mechanism may limit the diameters of other biological fibers such as those of collagen (Section 8-2Bb).

b. Factor XIII_a Cross-Links Fibrin Clots

The above “soft clot,” as this name implies, is rather fragile. Concomitantly with its formation, however, it is rapidly converted to a more stable “hard clot,” by the covalent cross-linking of neighboring D segments of fibrin molecules in a reaction catalyzed by **factor XIII_a (fibrin-stabilizing factor; FSF)**. This **transglutaminase** (a transamidase) covalently joins the C-terminal segments of adjacent γ chains by forming isopeptide bonds between the side chains of a Gln residue on one γ chain and a Lys residue on another (Fig. 35-8). Two such symmetrically equivalent bonds are rapidly formed between each neighboring pair of γ chains, which are brought together by the interactions of the α knobs with the γ holes (Fig. 35-7b) together with multiple noncovalent interactions. These interactions have been visualized by isolating cross-linked D segment dimers formed by the proteolytic digestion of XIII_a-treated fibrin.

**Figure 35-8** Reaction catalyzed by factor XIII_a

XIII_a The transamidation reaction forming the isopeptide bonds cross-linking fibrin monomers in hard clots involves an activated thiol group in a catalytic triad of the enzyme, producing a thioester intermediate. Except for a Cys thiol replacing the active site Ser, and a Lys residue replacing H₂O, the reaction resembles that of the serine proteases (Section 15-3).

The X-ray structures of these so-called D-dimers have been determined from several species (Fig. 35-9). In most cases, the cross-linked bridges are not visible in the X-ray structures, which indicates that they are flexible. The N-terminal domains of the γ chains of two fibrin protomers abut head to head in these dimers, but there are at least two different interaction surfaces that have been observed, further attesting to the flexibility of the isopeptide cross-fibrin-stabilizing link. The α chains are similarly cross-linked to one another but at a slower rate. The physiological importance of fibrin cross-linking is demonstrated by the observation that individuals deficient in XIII_a have a pronounced tendency to bleed.

XIII_a is present in both platelets and plasma. Platelet XIII_a consists of two 731-residue a chains, whereas plasma XIII_a additionally has two 641-residue b chains. Both species of XIII_a occur as zymogens that undergo thrombin-catalyzed cleavage of a specific Arg—Gly bond between positions 37 and 38 of each a chain with the

consequent release of a 37-residue propeptide. This treatment activates platelet XIII_a, designated a'_2 , but plasma XIII_a, a'_2b_2 , remains inactive until its b chains dissociate, a process that is triggered by the binding of Ca²⁺ to the a' subunits. We shall see below that Ca²⁺ is an essential factor in most stages of the blood clotting cascade. The b subunits are thought to prolong the survival of plasma XIII_a in the circulation.

B. Thrombin Activation and the Function of Vitamin K

Thrombin is a serine protease that consists of two disulfide-linked polypeptide chains: in humans, a 36-residue A chain and a 259-residue B chain. The thrombin B chain is homologous to trypsin and has similar specificity but is far more selective: It cleaves only certain Arg—X and, less frequently, Lys—X bonds with a clear preference for a Pro preceding the Arg or Lys.

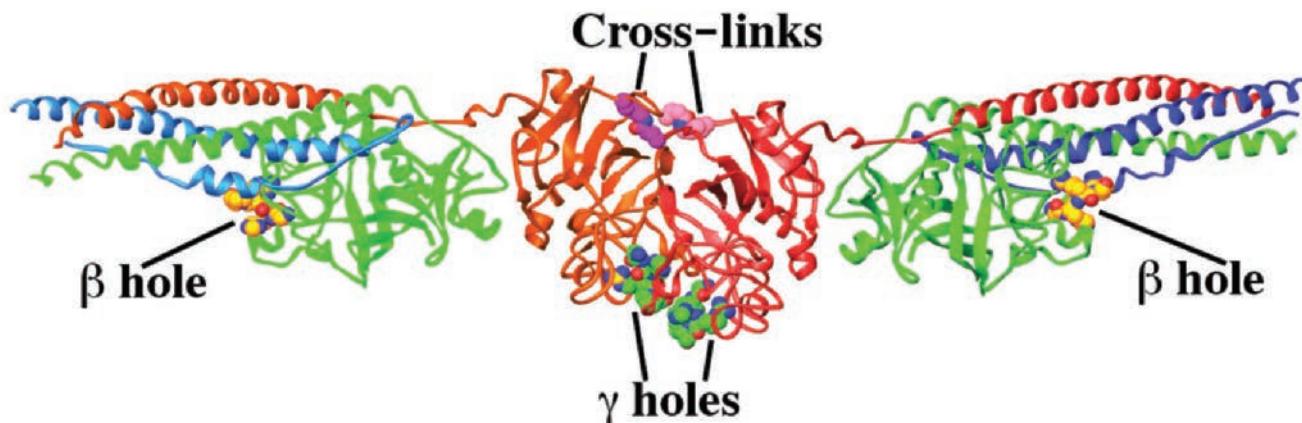


Figure 35-9 X-ray structure of a XIII_a-cross-linked D-D segment of lamprey fibrinogen. The α chains are blue or light blue, the β chains are green or light green, and the γ chains are red or orange. Cross-links between Q387 γ and K401 γ' and K401 γ and Q387 γ' are shown in space filling form with C magenta or pink,

N blue, and O red. The β and γ holes both bind the knob analog Gly-Pro-Arg-Pro-amide, which is drawn in space-filling form with C green or gold, N blue, and O red. [Based on an X-ray structure by Russell Doolittle, University of California at San Diego. PDBid 1N73.]

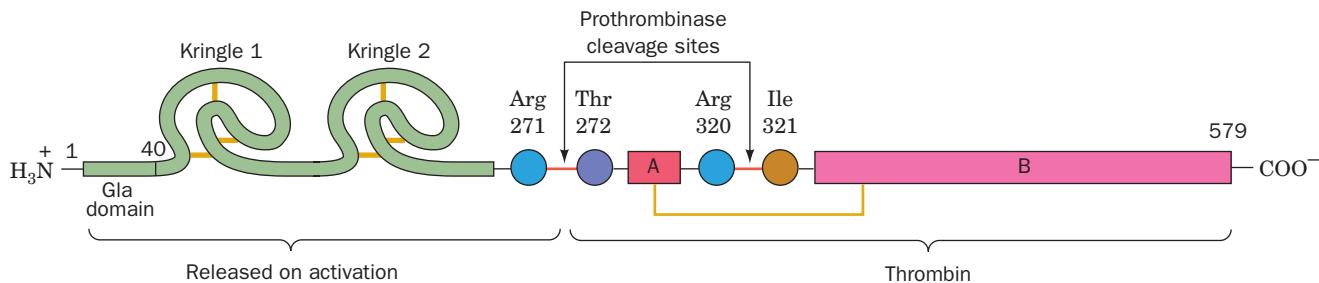


Figure 35-10 Human prothrombin. A schematic diagram indicating the two peptide bonds that are cleaved by prothrombinase (a V_a - X_a complex) to form thrombin. The N-terminal propeptide, which consists of a Gla domain and two

tandem kringle domains, is released in this activation process, whereas thrombin's A and B peptides remain linked by one of the protein's several disulfide bonds (yellow lines). Thrombin then autolytically excises the A peptide's N-terminal 13 residues.

Human thrombin is synthesized as a 579-residue zymogen, **prothrombin (factor II)**, which is activated by two proteolytic cleavages catalyzed by **prothrombinase**, a complex of activated **Stuart factor (X_a)** and activated **proaccelerin (V_a)**, products of the preceding steps of the clotting cascade. The cleavage of prothrombin's Arg 271—Thr 272 and Arg 320—Ile 321 bonds releases its N-terminal propeptide and separates the A and B chains (Fig. 35-10). The latter cleavage, which yields active enzyme, results in the formation of an ion pair between Ile 321 and Asp 524, much like that formed between chymotrypsinogen's homologous Ile 16 and Asp 194 in the activation of this zymogen (Section 15-3Ea). Thrombin then autolytically cleaves its Arg 285—Thr 286 bond, thereby trimming away the N-terminal 13 residues of the A chain to yield **α -thrombin**, hereinafter referred to as simply thrombin (recall that chymotrypsin also undergoes an autolytic cleavage that does not affect its catalytic activity; Section 15-3Ea).

Prothrombin's propeptide consists of three domains (Fig. 35-10; left): an N-terminal 40-residue **Gla domain** (so named for reasons indicated below) followed by two 40% identical ~115-residue **kringle** domains. Kringles are cross-linked by three characteristically located disulfide bonds that give these triple-looped sequence motifs a folded appearance as drawn in Fig. 35-10 reminiscent of a Scandinavian pastry of the same name. Gla domains and kringles occur in several of the proteins involved in the formation and breakdown of blood clots.

a. Vitamin K Is an Essential Cofactor in the Synthesis of γ -Carboxyglutamate

Prothrombin, as well as the homologous factors **VII, IX, and X** and the anticoagulant proteins **C and S**, is synthesized in the liver in a process that requires an adequate dietary intake of **vitamin K** (Fig. 35-11; *K* for the Danish *koagulation*). Lack of vitamin K or the presence of a competitive inhibitor such as **dicoumarol** (which was discovered in spoiled sweet clover because it causes fatal hemorrhaging in cattle) or **warfarin (Coumadin)**; a “blood thinner” that is used both therapeutically and as a rat poison) causes the production of an abnormal prothrombin that is activated by prothrombinase at only 1 to 2% of the normal rate. This observation was, at first, quite puzzling because normal and

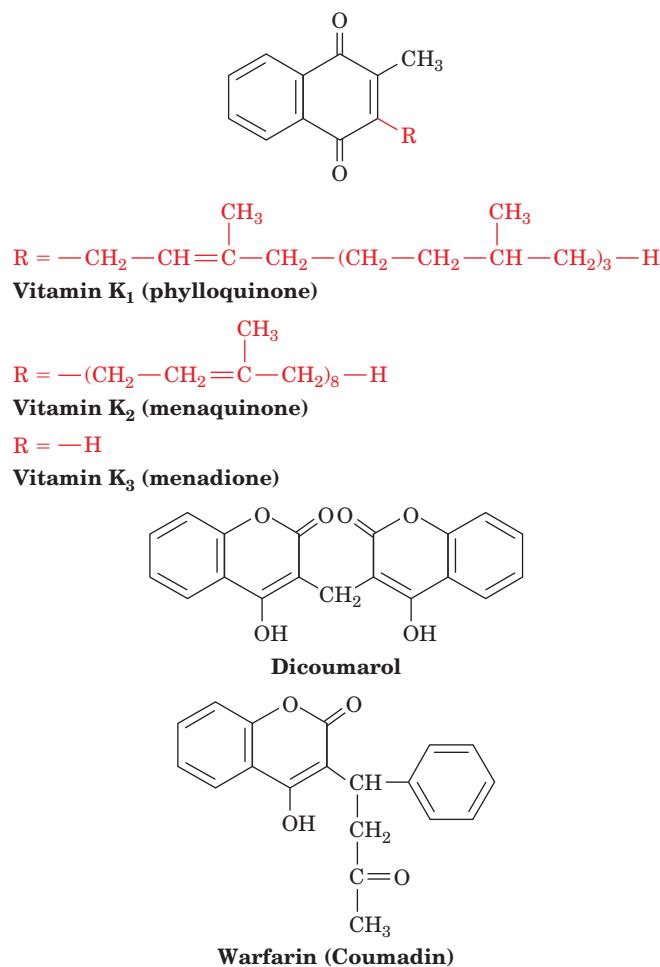
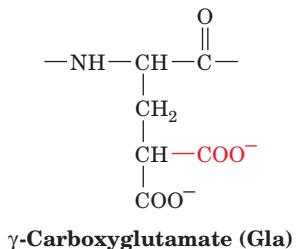


Figure 35-11 The molecular formulas of vitamin K and two of its competitive inhibitors, dicoumarol and warfarin. Vitamin K occurs in green leaves as **vitamin K₁ (phylloquinone)** and is synthesized by intestinal bacteria as **vitamin K₂ (menaquinone)**. Recall that these forms of the vitamin function as electron acceptors in chloroplast and bacterial photosynthesis (Section 24-2). The body converts the parent compound, **vitamin K₃ (menadione)**, to a vitamin-active form.



Figure 35-12 Sequence of the Gla domain of human prothrombin showing its 10 Gla residues.

abnormal prothrombins seemed to have identical amino acid compositions. NMR studies eventually established, however, that normal prothrombin contains **γ-carboxyglutamate (Gla)** residues,



10 of which occur between residues 6 and 32 in human prothrombin (in its Gla domain; Fig. 35-12). Abnormal prothrombin, in contrast, contains Glu in place of these Gla residues. *Vitamin K must therefore be a cofactor in the post-translational conversion of Glu to Gla.* The reason why prothrombin's Gla residues were not initially detected is because they decarboxylate to Glu under the conditions of acid hydrolysis normally used in amino acid composition determinations.

The liver reaction cycle that synthesizes Gla from Glu and then regenerates the vitamin K cofactor occurs in four reactions as catalyzed by three enzymes (Fig. 35-13):

1. In an O_2 -consuming reaction, **vitamin K-dependent carboxylase** converts the active hydroquinone form of vitamin K to an oxygenated intermediate that abstracts a γ proton from Glu to yield its γ -carbanion and **vitamin K-2,3-epoxide**. Thus vitamin K-dependent carboxylase is also an epoxidase. The oxygenated intermediate, whose identity has not been definitively established, apparently has the high basicity required to abstract the Glu proton as indicated in Fig. 35-13.
2. The Glu carbanion reacts with CO_2 to yield Gla in a reaction that is also catalyzed by vitamin K-dependent carboxylase.
3. **Vitamin K-epoxide reductase** reduces vitamin K-2,3-epoxide to vitamin K in its quinone form with the concomitant oxidation of thiols such as lipoic acid.
4. **Vitamin K reductase** completes the reaction cycle by reducing vitamin K quinone to its hydroquinone form in a reaction that also occurs with the concomitant oxidation of thiols.

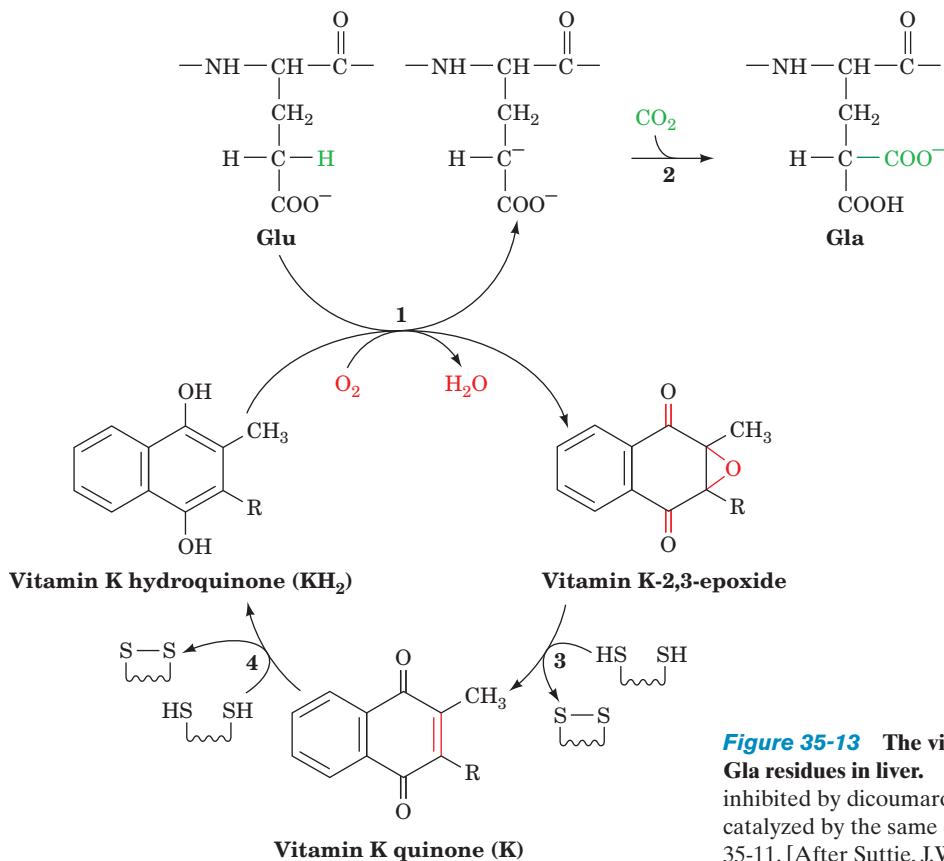


Figure 35-13 The vitamin K reaction cycle that produces **Gla residues in liver.** Reactions 3 and 4, which are both inhibited by dicoumarol and warfarin, are thought to be catalyzed by the same enzyme. The R group is indicated in Fig. 35-11. [After Suttie, J.W., *Annu. Rev. Biochem.* **54**, 472 (1985).]

Dicoumarol and warfarin act by inhibiting both vitamin K-epoxide reductase and vitamin K reductase.

The discovery of Gla residues in clotting factors led to their discovery in other tissues, which must therefore also contain vitamin K-dependent carboxylases. We shall see below that the Gla residues of clotting factors function to bind Ca^{2+} ion, which then forms a bridge binding the protein to negatively charged phospholipid membranes. Presumably they have similar roles in other tissues.

Vitamin K deficiency is uncommon in otherwise healthy adults because vitamin K_1 is widely available in green leafy vegetables and because the bacteria that normally inhabit the human colon synthesize around half the vitamin K requirement as vitamin K_2 . In contrast, newborn infants, particularly those born to poorly nourished mothers, are at risk for vitamin K deficiency with its attendant potentially fatal bleeding problems. This is because they are born with a limited supply of vitamin K (it does not readily cross the placenta), there is little of it in breast milk, and their initially sterile intestines are not immediately colonized by vitamin K_2 -producing bacteria. Consequently, newborns are routinely given an injection of vitamin K_1 .

b. Prothrombin Is Activated by Prothrombinase Complex in the Presence of Ca^{2+} Ion and Negatively Charged Phospholipid

Factor X_a , alone, is an extremely sluggish prothrombin activator. Yet, in the presence of activated proaccelerin (V_a), Ca^{2+} ion, and phospholipid membrane, its activity is enhanced 300,000-fold. The membrane surface in contact with the activation complex must contain anionic phospholipids such as phosphatidylserine in order to stimulate this rate enhancement. Such phospholipids occur almost exclusively on the cytoplasmic sides of cell membranes (Fig. 12-35), which of course, are normally not in contact with the blood plasma. Moreover, ~20% of the total factor V in blood is stored in the platelets and released only on platelet activation. Platelet activation also involves the exposure of phosphatidylserine on their external surfaces with the associated increase in binding of Gla-containing proteins. Consequently, *physiological prothrombin activation normally takes place at a significant rate only in the vicinity of an injury*.

Ca^{2+} is required for either prothrombin or factor X_a to bind to negatively charged phospholipid membranes; these proteins are anchored to the membrane via Ca^{2+} bridges. Prothrombin and factor X_a from vitamin K-deficient animals have greatly reduced membrane-binding affinities compared to the corresponding normal proteins. Evidently, the Gla side chains, which are much stronger Ca^{2+} ion chelators than Glu, form the proteins' Ca^{2+} -binding sites. In fact, the 9 to 12 conserved Gla residues that occur in each of the vitamin K-dependent clotting zymogens—prothrombin and factors VII, IX, and X—are contained in these proteins' homologous N-terminal segments (Fig. 35-12 for prothrombin). The excision of prothrombin's N-terminal propeptide releases the resulting thrombin from the phospholipid membrane so that it can activate fibrinogen in the plasma. Thrombin differs in this respect from the other vitamin K-dependent zymogens, which remain bound to the phospholipid membrane after their activation.

Active thrombin specifically cleaves prothrombin's propeptide at its Arg 155—Ser 156 bond to yield its 155-residue N-terminal segment, the so-called **prothrombin fragment 1**, which consists of prothrombin's Gla domain and its kringle 1 (Fig. 35-10). The X-ray structure of bovine prothrombin fragment 1 was determined by Alexander Tulinsky in both the presence and absence of Ca^{2+} ion. In the absence of Ca^{2+} , the Gla domain is disordered. However, when Ca^{2+} is present, the Gla domain folds to form two internal carboxylate surfaces that coordinate seven Ca^{2+} ions in a nearly linear array (Fig. 35-14). Kringle 1, which assumes the same conformation in the presence and

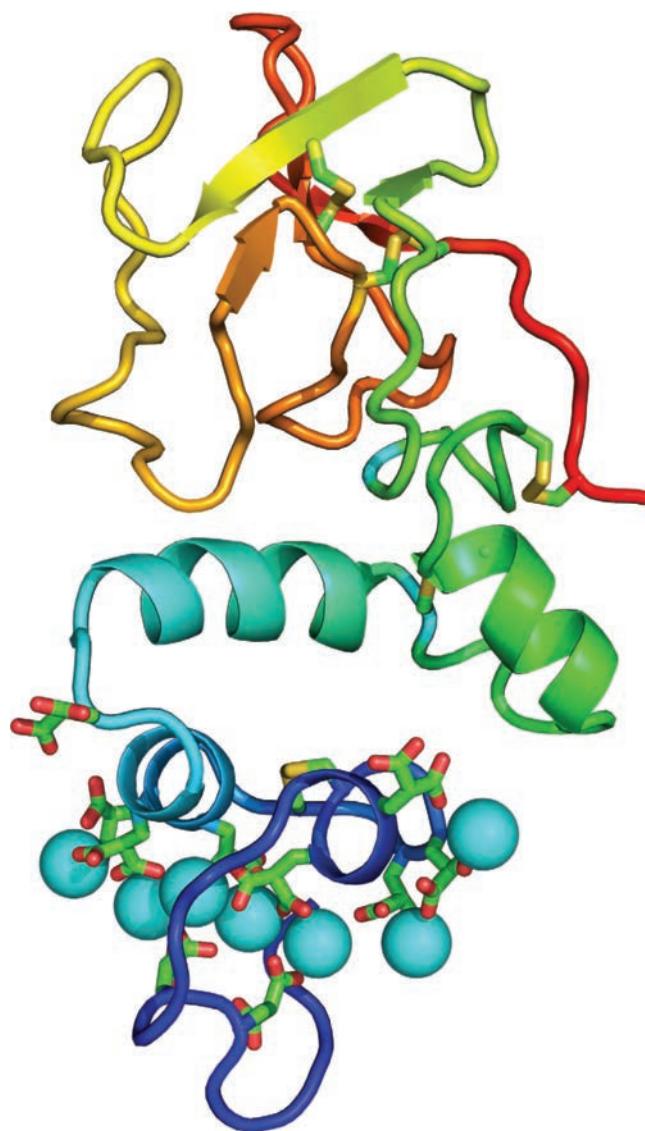


Figure 35-14 X-ray structure of bovine prothrombin fragment 1 in complex with Ca^{2+} . The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The protein's 10 Gla side chains and 5 disulfide linkages are drawn in stick form with C green, O red, and S yellow. Its 7 bound Ca^{2+} ions are represented by cyan spheres. The other Gla-containing proteins have between 6 and 12 Gla residues. [Based on an X-ray structure by Alexander Tulinsky, Michigan State University. PDBid 2PF2.]

absence of Ca^{2+} , folds into a compact globule that contains three short two-stranded antiparallel β sheets.

Activated proaccelerin (V_a), the accessory factor in prothrombin activation, is produced by a thrombin-catalyzed proteolytic cleavage. Prothrombin activation, in this indirect way, is thereby autocatalytic (thrombin, *in vitro*, can also directly activate prothrombin by cleaving its Arg 283—Thr 284 bond, but this reaction has been shown to be physiologically insignificant). However, under certain conditions (see below) thrombin also can activate **protein C**, a protease that inactivates V_a and therefore acts as an anticoagulant. Moreover, thrombin can proteolytically inactivate other thrombin molecules. Because of these, and other anticoagulant activities (Section 35-1Eb), *clot formation is self-limiting, a safeguard that helps prevent blood clots from propagating away from the site of an injury*.

c. Thrombin Structurally Resembles Trypsin

The X-ray structure of human thrombin, as inactivated by **D-Phe-Pro-Arg chloromethylketone [PPACK]** (which specifically alkylates thrombin's active site His (Section 15-3Ab) as well as forming a hemiketal (a tetrahedral intermediate analog) with its active site Ser], was independently determined by Tulinsky and by Wolfram Bode and Robert Huber (Fig. 35-15). The A chain and B chain, which are linked only by a disulfide bond (Fig. 35-10), form a nearly spherical molecule. The boomerang-shaped A chain, which is analogous to the propeptide of chymotrypsinogen (Section 15-3Ea), is nestled against the B chain globule opposite the substrate binding cleft.

The structure of the B chain closely resembles those of the pancreatic serine proteases (compare Figs. 15-19 and 35-15a) as was expected from the high degree of sequence similarity among these various proteins. However, thrombin's substrate-binding cleft is deeper and narrower than that of trypsin (Fig. 35-15b) due to the presence of several

(a)



(b)

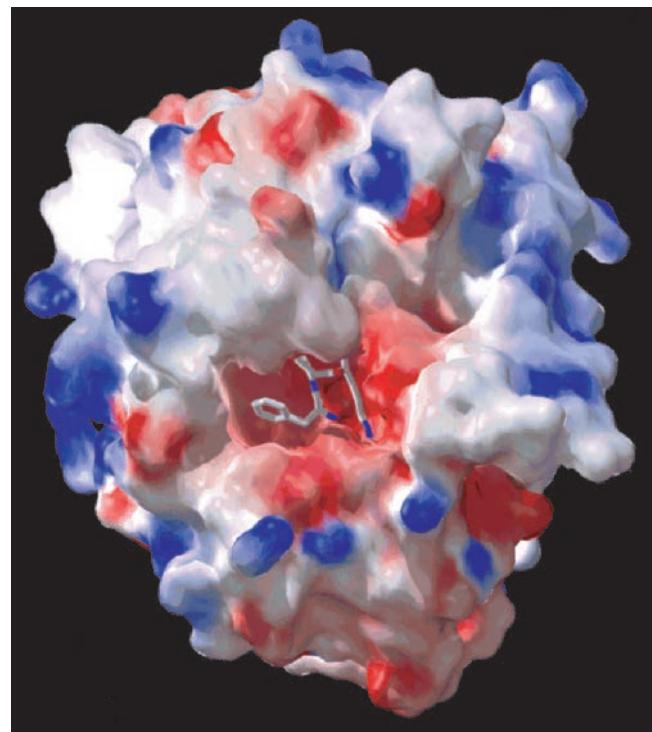


Figure 35-15 X-ray structure of human thrombin inactivated by D-Phe-Pro-Arg chloromethylketone (PPACK). (a) A ribbon diagram oriented similarly to the structurally homologous trypsin as drawn in Fig. 15-19. Those portions of the 253-residue B chain that have a structural analog in trypsin are magenta, whereas the remaining segments are cyan. Similarly, the portions of the 29-residue A chain that have structural analogs in chymotrypsinogen are red and other segments are blue. The side chains of the catalytic triad (His 57, Asp 102, and Ser 195 using chymotrypsinogen numbering), the Cys residues that disulfide-link the A and B chains, and PPACK are drawn in stick form with thrombin C green, PPACK C gold, N blue, O red, and S yellow. The enzyme's bound Na^+ ion is represented by a

yellow-green sphere, and the positions of exosites I and II and the 60-loop (which is not part of exosite II) are indicated. [Based on an X-ray structure by Enrico Di Cera and F. Scott Mathews, Washington University School of Medicine. PDBid 1SQF.]

(b) Surface diagram of thrombin colored according to its electrostatic surface potential (blue positive, white neutral, and red negative) shown in approximately the same orientation as in Part a. PPACK, which is partially occluded in this view, is drawn in stick form with C white, N blue, and O red. Note the deep substrate-binding cleft ending in the positively charged exosite I (*upper right*). [Courtesy of Wolfram Bode, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBid 1PPB.]

surface loops that are located around thrombin's substrate-binding cleft and which are not present in the pancreatic enzymes. Steric hindrance by these loops, particularly the 60-loop [so called because this 9-residue insert follows Val 60 (using chymotrypsinogen numbering)], greatly restricts access to the active site and presumably contributes to thrombin's high specificity and its poor binding of most natural serine protease inhibitors. Indeed, when a model of bovine pancreatic trypsin inhibitor (BPTI) is docked to that of thrombin in the same position and orientation that BPTI assumes in its complex with trypsin (Fig. 15-24), there is a significant steric clash between BPTI and the 60-loop.

Unlike the pancreatic serine proteases, thrombin has a binding site for Na^+ ion and exists in two conformational states, the so-called fast (F; Na^+ bound) and slow (S; Na^+ free) forms, which have high and low specificity for fibrinogen, respectively. Na^+ ion thus allosterically activates this monomeric enzyme to cleave fibrinogen and several other substrates. Interestingly, the slow form of thrombin is slightly more specific for protein C, so that a decrease in $[\text{Na}^+]$ causes a change in thrombin specificity from coagulant substrates to anticoagulant substrates.

The specificity of thrombin for fibrinogen is largely attributable to its so-called anion-binding exosite I, one of two positively charged exosites and an extension of thrombin's substrate-binding cleft, which contains numerous Arg and Lys residues. As we shall see (Section 35-1Ec), exosite I, together with Na^+ ion, is responsible for the switch in thrombin specificity regulating its coagulant and anticoagulant activity, whereas exosite II is involved in the binding of heparin.

d. Thrombin Acts as a Hormone to Activate Platelets

Platelet membranes contain embedded G protein-coupled receptors (GPCRs; Section 19-2B) called **protease-activated receptors (PARs)**. **PAR1** and **PAR4** are activated by their thrombin-catalyzed cleavage at a specific site near the N-terminus of the GPCR. This creates a new N-terminus that functions as a tethered ligand by binding to the GPCR

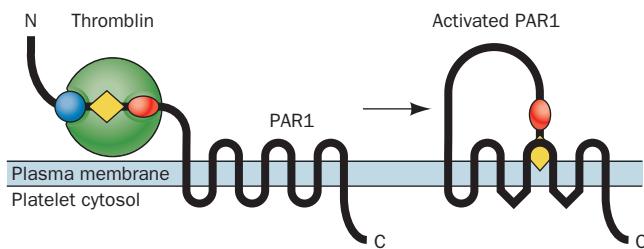


Figure 35-16 Activation mechanism of protease-activated receptor 1 (PAR1). Thrombin (green sphere) cleaves the Arg 41—Ser 42 bond of this heptahelical G protein-coupled receptor. Thrombin recognizes portions of PAR1's N- and C-terminal domains (blue sphere and pink oval). The latter domain resembles the C-terminal tail of the thrombin inhibitor hirudin and binds to thrombin in a similar way (Section 35-1Ed). The newly unmasked N-terminus (yellow diamond) functions as a tethered ligand, docking intramolecularly with the body of the receptor to initiate transmembrane signaling. [After a drawing by Shaun Coughlin, University of California at San Francisco.]

so as to trigger its response (Fig. 35-16). PARs are therefore peptide receptors that carry their own ligand, which is only active at signaling after a proteolytic cleavage. This signaling occurs via the GPCR-mediated activation of the heterotrimeric G proteins G_{ia} , G_{q} , $\text{G}_{12\alpha}$, and $\text{G}_{13\alpha}$, (Section 19-2C) and results in platelet activation and the release of granules containing serotonin, ADP, thrombin, and thromboxane A₂, which further activate platelets. Thus a small amount of platelet activation feeds back on itself to amplify the platelet activation process and hence blood clot formation.

C. The Extrinsic Pathway

Factor X may be activated by two different proteases, which are therefore known as **tenases** (Fig. 35-1):

1. By the **VIII_a–IX_a complex**, the product of the **intrinsic pathway** (so named because all of its protein components are contained in the blood).
2. By the **TF–VII_a complex**, the product of the **extrinsic pathway** [so called because one of its important components, **tissue factor (TF; III)**, occurs in the tissues].

Although the division of the blood clotting cascade into these two pathways is largely contrived, we shall discuss them separately beginning with the extrinsic pathway.

The extrinsic pathway (Fig. 35-1), the primary triggering pathway of the clotting cascade, is initiated when plasma comes in contact with cells expressing TF. This integral membrane glycoprotein occurs on cell surfaces in many tissues but not those lining blood vessels. TF binds **proconvertin (VII)**, which is then activated by any of several proteases, including thrombin and activated **Hageman factor (XII_a)**, to form the active serine protease complex **TF–VII_a**. Indeed, the protease activity of free VII_a is enhanced 16,000-fold when complexed to TF in the presence of negatively charged phospholipid membrane and Ca^{2+} ion, which interact with its Gla domain. TF–VII_a, in turn, proteolytically activates factors IX and X, which both contain Gla residues, in Ca^{2+} -requiring reactions that also occur on negatively charged phospholipid membrane surfaces.

Since the formation of the TF–VII_a complex activates the pathway that produces thrombin, which in turn generates VII_a, what is the origin of the VII_a that initiates this process? The answer is that $\sim 1\%$ of the factor VII in the bloodstream is present as VII_a due to the minimal proteolytic activity of VII itself. Evidently, the formation of VII_a can also precede its binding to TF.

TF is particularly abundant in brain, lung, blood vessel walls, and placenta. Consequently, an injury that exposes blood to these tissues rapidly initiates the extrinsic pathway. In fact, the addition of TF to the extrinsic system causes clot formation in ~ 12 s.

a. Heart Attacks and Strokes Can Result from Inappropriate Clotting Involving TF in Atherosclerotic Plaques

Tissue factor is usually not contained in vascular smooth muscle and endothelial cells, thus protecting blood normally exposed to blood vessel walls from inappropriate

clotting. The clotting cascade is only initiated when vascular injury releases blood from blood vessels onto TF-containing tissues. However, atherosclerotic plaques that develop in blood vessels contain TF because TF synthesis is induced in the monocytes and smooth muscle cells contained in these plaques by a variety of inflammatory mediators. A myocardial infarction (heart attack) or a stroke can occur when an atherosclerotic plaque ruptures, exposing TF and causing a clot (thrombus) to form that blocks a blood vessel in the heart or brain. Indeed, blood clots are the leading cause of heart attacks and strokes, the two major causes of human death in developed countries. In addition, venous thrombosis is the second leading cause of death in individuals with cancer, perhaps due to TF expression by malignant cells.

D. The Intrinsic Pathway

The intrinsic pathway involves two steps leading to the activation of factor X (Fig. 35-1). Factor XI_a catalyzes the proteolytic activation of factor IX, in a Ca^{2+} -requiring reaction that takes place on negatively charged phospholipid membrane surfaces. No accessory factor is required for this reaction. Factor IX is also activated by the TF-VII $_a$ complex, a product of the extrinsic pathway (Section 35-1C). The resulting activated **Christmas factor** (IX_a) forms a complex, with activated **antihemophilic factor** (VIII_a), that proteolytically cleaves factor X in a Ca^{2+} -dependent reaction that occurs on negatively charged phospholipid membrane surfaces. Factor VIII, as is proaccelerin (V), is proteolytically activated by thrombin in a second autocatalytic process leading to additional prothrombin activation (Fig. 35-1). Not surprisingly, factors V and VIII are homologous, Gla-containing proteins. Factor VIII circulates in the plasma in complex with von Willebrandt factor, the mediator of platelet adhesion; in fact, the activities of these two substances were initially attributed to a single protein.

a. Hemophilias Result from Clotting Factor Deficiencies

The discovery of antihemophilic factor (VIII) came about through its deficiency in individuals with the most common clotting disorder, **hemophilia A**, a sex-linked inherited deficiency (~1 per 10,000 male births). Indeed, several of the clotting factors were discovered through the diagnosis of their deficiencies in various clotting disorders [the existence of Christmas factor (IX) was discovered through its absence in Stephen Christmas, a hemophiliac whose deficiency, **hemophilia B**, is the second most common form of hemophilia]. Hemophiliacs may lose large amounts of blood from even the smallest injury and frequently hemorrhage without any apparent cause. However, the symptoms of their diseases may be alleviated by the intravenous administration of the deficient factor. In the past, this treatment was expensive and not without risk because large amounts of blood had to be fractionated to obtain therapeutic doses of most clotting factors. Hemophiliacs were therefore inordinately subject to a variety of dangerous bloodborne viral diseases including hepatitis

and AIDS. These risks have now been eliminated through the production of the required clotting factors by recombinant DNA techniques.

The severity of the hemophilias resulting from intrinsic pathway clotting factor deficiencies clearly establishes the importance of the intrinsic pathway in blood clotting. Indeed, the intrinsic and extrinsic pathways are really integrated, since they are coupled through a number of reactions (Fig. 35-1).

b. Clotting May Be Initiated by the Contact System

It has long been known that bringing blood into contact with negatively charged surfaces, such as those of glass or kaolin (a clay used to make porcelain), initiates clotting. This so-called **contact system** initiates the intrinsic pathway by activating factor IX.

The contact system consists of four glycoproteins: the serine protease zymogens Hageman factor (XII), **prekallikrein**, **plasma thromboplastin antecedent (PTA or XI)**, and **high molecular weight kininogen (HMK)**, an accessory factor that is also a precursor of the nonapeptide hormone **bradykinin** (a potent vasodilator and diuretic factor; Section 7-5B). Adsorption to a suitable surface is thought to somehow activate factor XII which, in the presence of HMK, proteolyzes prekallikrein to form the active protease **kallikrein**. Kallikrein, in turn, proteolytically activates factor XII so that these two proteins reciprocally activate each other.

The nature of contact-activated factor XII_a is enigmatic; it is by no means certain that physical adsorption to a surface cleaves the same bond as does kallikrein or, for that matter, cleaves any bond at all. Much of the experimental difficulty in resolving this issue is a consequence of the autocatalytic nature of the contact-activation process: Prekallikrein, contact-activated factor XII_a 's substrate, is the zymogen of the protease that activates factor XII. Consequently, in any measurement of its activity, the nature of contact-activated factor XII_a is immediately obscured by large amounts of rapidly generated kallikrein-activated factor XII_a .

The final reaction mediated by the contact system is the proteolytic activation of factor XI by factor XII_a in a process that also uses HMK as an accessory factor. Factor XI_a can then activate the intrinsic pathway by activating factor IX. Although the contact system is clearly effective in initiating *in vitro* clot formation, its *in vivo* importance in doing so is in doubt because individuals deficient in factor XII, prekallikrein, or HMK do not suffer from bleeding problems. However, these proteins participate in the initiation of the inflammatory response, angiogenesis (blood vessel growth), and fibrinolysis (see below).

E. Control of Clotting

The multilevel cascade of the blood clotting system permits enormous amplification of its triggering signals. Moving down the extrinsic pathway, for example, proconvertin (VII), Stuart factor (X), prothrombin, and fibrinogen are present in plasma in concentrations of <1, 8, 150, and up to

4000 $\mu\text{g} \cdot \text{mL}^{-1}$, respectively. Yet, clotting must be very strictly regulated since, as we have seen, even one inappropriate clot can have fatal consequences.

a. A Variety of Factors Limit Clot Growth

There are numerous physiological mechanisms that limit clot formation, so that *in vivo* blood clotting is self-limiting. We have seen that there are several interactions among the various clotting factors that inhibit blood coagulation (Fig. 35-1). The blood flow dilution of active clotting factors also does so as does their selective removal from the circulation by the liver. In addition, plasma contains several serine protease inhibitors whose presence prevents clots from spreading beyond the vicinity of an injury (most active clotting system serine proteases have half-lives in the plasma of only a few seconds). For example, the 432-residue **antithrombin III (ATIII or just antithrombin)** inhibits all active proteases of the clotting system except VII_a by binding to them in a tight 1:1 complex. It is a member of the class of compounds known as **serpins** (for *serine protease inhibitors*; they bind their target enzymes much like BPTI binds trypsin; Section 15-3Da). **Heparin**, a sulfated glycosaminoglycan and the most intensely negatively charged naturally occurring polymer (Section 11-2Eb), binds to thrombin's positively charged exosite II where it enhances the activity of antithrombin by several hundredfold. Heparin occurs almost exclusively in the intracellular secretory granules of the mast cells that line certain blood vessels. Its release, presumably by injury, activates antithrombin, thereby preventing runaway clot growth.

b. The Protein C Pathway Is a Major System for Controlling Clot Formation

The **protein C pathway** consists of thrombin, **thrombomodulin**, **endothelial cell protein C receptor (EPCR)**, **protein C**, and **protein S**. Protein C is plasma protein that limits clotting. This Gla residue-containing 419-residue zymogen is activated by thrombin to proteolytically inactivate V_a and VIII_a , the two clotting factors directly associated with prothrombin activation (Fig. 35-1). Activated protein C (**APC**) attacks the activated forms of these accessory factors more readily than their nonactive forms. The importance of protein C is demonstrated by the observation that individuals who lack it often die in infancy of massive thrombotic complications.

The activation of protein C is localized to the surface of endothelial cells by the action of two membrane-bound proteins, EPCR, and thrombomodulin. EPCR is a transmembrane protein that is homologous to class I MHC proteins (Section 35-2E). It recruits protein C to the surface of endothelial cells even in the absence of negatively charged phospholipid (the binding site for Ca^{2+} -Gla regions of Gla-containing proteins) by binding to its Ca^{2+} -Gla domain. Thrombomodulin, a 557-residue transmembrane glycoprotein that projects from the cell surface membranes of the vascular endothelium (inner lining), was discovered after the observation that blood clotting *in vitro* is not self-limiting. Thrombomodulin specifically binds

thrombin so as to recruit it to the endothelium, removing it from the circulation, and converting it to a form with decreased ability to catalyze clot formation but with >1000 -fold increased capacity to activate protein C. It does so by binding to thrombin exosite I competitively with fibrinogen, thereby inhibiting fibrin formation, while it creates a new binding surface to accommodate protein C. Thrombomodulin thus acts as a switch between thrombin's coagulant and anticoagulant properties.

APC, once dissociated from EPCR, binds to protein S on appropriate cell surfaces. Protein S (635 residues) is another vitamin K-dependent protein whose Ca^{2+} -Gla domain is responsible for its membrane association. Here, protein S acts as a cofactor for the APC-catalyzed proteolytic inactivation of factors V_a and VIII_a , thus inhibiting further thrombin formation and blood coagulation.

c. Thrombin Function Varies with $[\text{Na}^+]$

As we saw in Section 35-1Bc, Na^+ binding activates thrombin to cleave its clotting factor substrates (Fig. 35-1) and thereby function as a procoagulant. However, in the absence of bound Na^+ , thrombin functions as an anticoagulant because its fibrinogen-cleaving activity is greatly reduced whereas the rate at which it cleaves protein C is slightly increased. Thrombin's dissociation constant for binding Na^+ ion is 110 mM. Hence, under normal physiological conditions (plasma $[\text{Na}^+] = 140$ mM), thrombin's Na^+ binding site is 60% occupied. Thus changes in plasma $[\text{Na}^+]$ resulting in **hypernatremia** ($[\text{Na}^+] > 145$ mM) or **hyponatremia** ($[\text{Na}^+] < 135$ mM), which are among the most common electrolyte disorders, often result in thrombosis or bleeding, respectively. Moreover, under physiological conditions, the plasma $[\text{Na}^+]$ in the vicinity of aggregating platelets is significantly reduced, indicating that Na^+ regulates thrombin activity in the microenvironment of blood vessels.

d. Clotting Is Prevented by a Variety of Substances

The control of clotting is a major medical concern. Heparin, the most frequently used pharmaceutical anticoagulant, is administered before and after surgery to retard clot formation. For long-term control of hemostasis, warfarin (Coumadin; Section 35-1Ba) is often employed. In the design of artificial hearts the elimination of mechanically induced clots was the major problem (devising an adequate pump is a relatively simple task).

The prevention of clotting is also a concern of blood-sucking organisms. The leech *Hirudo medicinalis* solves this problem by secreting **hirudin**, a 65-residue protein, in its saliva. Hirudin, the most potent naturally occurring anticoagulant known, specifically binds to thrombin with an association constant of $\sim 5 \times 10^{13} \text{ M}^{-1}$ (greater than that of trypsin with BPTI; Section 15-3Da), thereby inactivating it. Consequently a leech bite, although a minor wound, bleeds quite freely.

The X-ray structure of the complex of hirudin with human thrombin, which was determined by Tulinsky, Bode, and Huber, reveals that hirudin binds to thrombin in a manner not previously observed for protease inhibitors

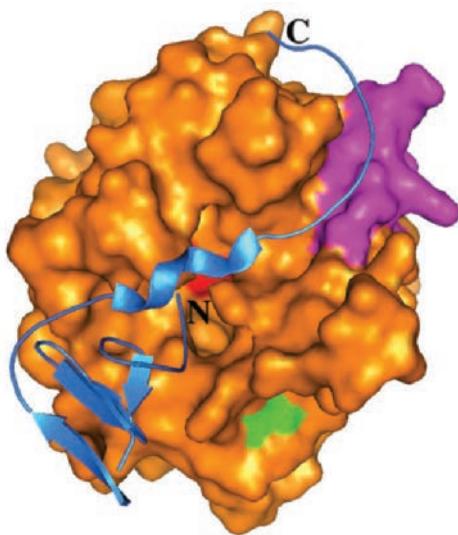


Figure 35-17 X-ray structure of hirudin in complex with

human thrombin. Hirudin is drawn in ribbon form in blue and thrombin is represented by its surface diagram in orange with its catalytic triad red, exosite I magenta, and its Na^+ -binding site green. The view of thrombin is similar to that in Fig. 35-15. Note hirudin's short segment of left-handed polyproline helix (Section 8-1Bb) that overlies the thrombin active site. [Based on an X-ray structure by Alexander Tulinsky, Michigan State University, and Wolfram Bode and Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBid 4HTC.]

(Fig. 35-17). Hirudin has a globular N-terminal domain and an extended C-terminal domain. The N-terminal domain binds in the region of thrombin's active site such that hirudin's N-terminal amino group forms a hydrogen bond with the active site Ser of thrombin. However, hirudin does not occupy thrombin's specificity pocket. The three N-terminal residues of hirudin form a parallel β strand with Ser 214 to Glu 217 of thrombin (using chymotrypsinogen numbering; Section 15-3B). Interestingly, this segment of thrombin forms an antiparallel β sheet in all other known natural serine protease inhibitor complexes (with which hirudin has no sequence homology) as well as in the above described D-Phe-Pro-Arg chloromethylketone (PPACK) complex. Thus, the hirudin chain runs in the direction opposite that expected for fibrinogen. Hirudin's highly anionic and extended 39-Å-long C-terminal tail, which comprises its residues 48 to 65, wraps around thrombin in its anion-binding exosite I (where highly anionic portions of fibrinogen also bind) making extensive ionic and hydrophobic contacts. These numerous interactions are almost certainly responsible for hirudin's tight binding to thrombin.

F. Clot Lysis

Blood clots are only temporary patches; they must be eliminated as wound repair progresses. This is a particularly urgent need when a clot has inappropriately formed or has broken free into the general circulation. Fibrin is a molecule that is “designed” to be easily dismantled in a process

termed **fibrinolysis**. The demolition agent is a plasma serine protease named **plasmin**, an enzyme that specifically cleaves fibrin's triple-stranded coiled coil segment and cuts away its covalently cross-linked α chain protuberances (Fig. 35-3). The open meshlike structure of a blood clot (Fig. 35-2) gives plasmin relatively free access to the polymerized fibrin molecules, thereby facilitating clot lysis.

Plasmin is formed through the proteolytic cleavage of the 791-residue zymogen **plasminogen**, a protein that is homologous to the zymogens of the blood clotting cascade. There are several serine proteases that activate plasminogen, most notably the 411-residue enzyme **urokinase**, which is synthesized by the kidney and occurs, as its name implies, in the urine, and the homologous 527-residue enzyme **tissue-type plasminogen activator (t-PA)**, which occurs in vascular tissues. In addition, activated Hageman factor (XII_a), in the presence of prekallikrein and HMK (the contact activation system) activates plasminogen although the physiological significance of the contact activation of the fibrinolytic system has not been determined. Nevertheless, the fibrinolysis system, as experience might have led us to expect, is not so simple as just a zymogen and its activators. It also employs several inhibitors, principally the 452-residue glycoprotein **α_2 -antiplasmin**, which forms an irreversible equimolar complex with plasmin that prevents it from binding to fibrin. The α_2 -antiplasmin cross-links to fibrin α chains through the action of activated FSF (XIII_a , the enzyme that also cross-links fibrin), thereby making “hard” clots less susceptible to fibrinolysis than “soft” clots. The importance of this serpin (which also inhibits chymotrypsin) is indicated by the observation that homozygotes for a defective α_2 -antiplasmin have a serious tendency to bleed.

a. Plasminogen Activators Have Medical Utility

Plasminogen activators have received considerable medical attention aimed at rapidly dissolving the blood clots responsible for heart attacks and strokes. **Streptokinase**, a 414-residue protein produced by certain streptococci, has shown considerable utility in this regard, particularly when administered together with aspirin (which inhibits platelet aggregation; Section 25-7Ba). In fact, the rapid administration of aspirin results in a 23% reduction in mortality in heart attack patients and is now routinely administered with other thrombolytic agents.

Despite its name, streptokinase exhibits no enzymatic activity. Rather, it acts by forming a tight 1:1 complex with plasminogen that undergoes a conformational change to expose a plasminogen active site that proteolytically activates other plasminogen molecules. The streptokinase-plasminogen complex is thus autoactivated to streptokinase-plasmin. When plasmin is in complex with streptokinase, it is not inhibited by α_2 -antiplasmin, further extending its clot-lysis activity. The use of streptokinase to dissolve clots has the apparent disadvantage that it activates circulating plasmin as well as the fibrin-associated enzyme. Circulating plasmin has a relatively low specificity and degrades fibrinogen as well as fibrin, thereby increasing the risk of bleeding problems, particularly strokes. The

therapeutic use of t-PA, which has been synthesized by recombinant DNA techniques, is thought to eliminate these problems (although their medical significance appears to be minimal) because this enzyme activates plasminogen only in the presence of a blood clot.

2 IMMUNITY

All organisms are continually subject to attack by other organisms. In response to predators, animals have developed an enormous variety of defensive strategies. An even more insidious threat, however, is attack by disease-causing microorganisms and viruses (pathogens). In order to deal with them, animals have evolved an elaborate protective array known as the **immune system** (Latin: *immunis*, exempt). Pathogens that manage to breach the physical and chemical barriers presented by the skin and mucous membranes (a vital first line of defense) are identified as foreign invaders and destroyed. At first, this process is carried out by the **innate immune system**, which recognizes pathogens according to their general characteristics, but after a few days it is largely taken over by the **adaptive immune system**, which recognizes pathogens according to their specific macromolecules. In Sections 35-2A–E we discuss how the adaptive immune system recognizes foreign invaders, how they are distinguished from normal components of self, and how they are destroyed. We discuss the innate immune system in Section 35-2F. Throughout these discussions, keep in mind that the immune system exhibits many of the qualities that are characteristic of the nervous system such as the ability to detect and react to stimuli and to remember. Indeed, the size and complexity of the vertebrate immune system rival that of the vertebrate nervous system.

A. The Adaptive Immune Response

Immunity in vertebrates is conferred by certain types of **leukocytes** (white blood cells) collectively known as **lymphocytes**. They arise, as do all blood cells, from common precursor cells (**stem cells**) in the bone marrow. Lymphocytes, however, in contrast to red blood cells, can leave the blood vessels and patrol the intercellular spaces for foreign intruders. They eventually return to the blood via the lymphatic vessels but not before interacting with specialized **lymphoid tissues** such as the thymus, the lymph nodes, and the spleen, the sites where much of the immune response occurs.

Two types of adaptive immunity have been distinguished:

1. Cellular immunity, which guards against virally infected cells, fungi, parasites, and foreign tissue, is mediated by **T lymphocytes** or **T cells**, so called because their development occurs in the thymus.

2. Humoral immunity (*humor* is an archaic term for fluid), which is most effective against bacterial infections and the extracellular phases of viral infections, is mediated by an enormously diverse collection of related proteins

known as **antibodies** or **immunoglobulins**. Antibodies are produced by **B lymphocytes** or **B cells**, which, in mammals, mature in the bone marrow.

We shall outline the operations and interactions of these systems as a prelude to discussing their biochemistry.

a. The Cellular Immune System

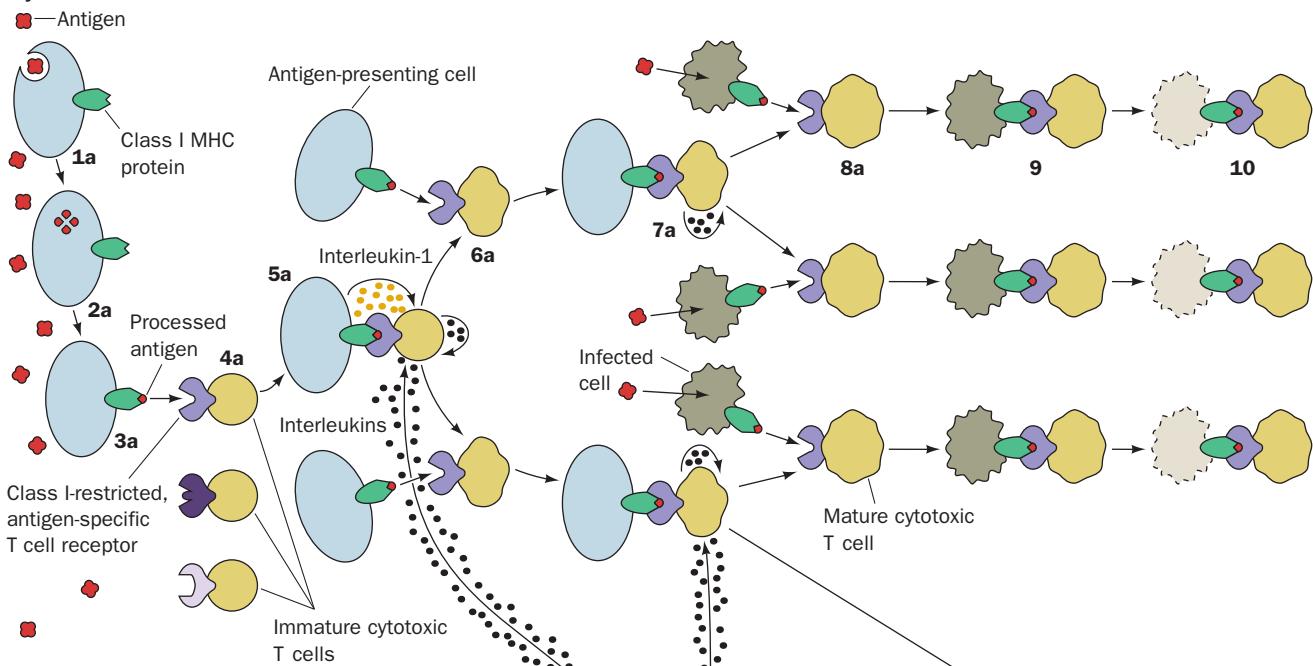
The adaptive immune response is triggered by the presence of foreign macromolecules, normally proteins, carbohydrates, and nucleic acids, known as **antigens** (antibody generators). This process occurs through a complex series of interactions among various types of T cells and B cells that specifically bind a particular antigen (Fig. 35-18). In the following paragraphs, italic numerals and letters refer to the corresponding drawing in Fig. 35-18.

The cellular immune response leads to the destruction of the offending cells. It begins when a **dendritic cell** or **macrophage** [types of white blood cells that are collectively known as **antigen-presenting cells (APCs)**] engulfs an invading pathogen via a process called **phagocytosis** (1a, 1b) and then digests it using proteasomes (Section 32-6Bf) whose three active protease subunits have been replaced by APC-specific subunits to form **immunoproteasomes** (2a, 2b). The resulting antigenic fragments are then displayed on the surface of the APC (3a, 3b) in complex with one of two types of transmembrane proteins known as **major histocompatibility complex (MHC) proteins** (so called because they are transcribed from a closely linked series of genes called the MHC; Section 35-2E). The MHC is remarkably polymorphic (has numerous alleles); so much so that any two unrelated individuals of the same species are highly unlikely to have an identical set of MHC proteins. *MHC proteins are therefore markers of individuality.*

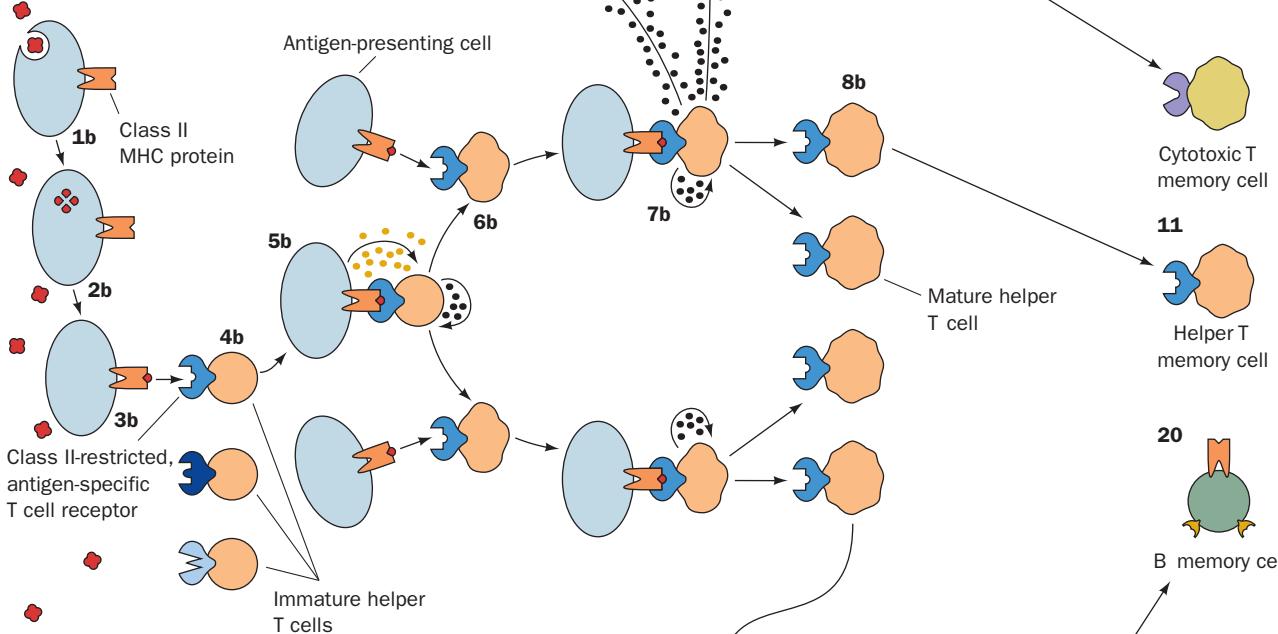
APCs expressing **Class I MHC proteins** are recognized by proteins known as **T cell receptors (TCRs)** that occur on the surfaces of immature **cytotoxic T (T_C) cells**. However, as Peter Doherty and Rolf Zinkernagel discovered, *in order to bind to an APC, TCRs must specifically complex the displayed antigen together with the Class I MHC protein (4a); neither molecule alone can do the job.* In the same way, APCs displaying antigenic fragments complexed to **Class II MHC proteins** are bound by immature **helper T (T_H) cells** bearing the cognate receptor (4b). This elaborate recognition system, as we shall see, focuses the attention of T cells on cell surfaces and thereby prevents the resources of the cellular immune system from being futilely squandered on noncellular targets.

T cells that bind to an APC-displayed antigen–MHC protein complex are induced to propagate, a process known as **clonal selection**, which was first recognized in the 1950s by Niels Kaj Jerne, Macfarlane Burnet, Joshua Lederberg, and David Talmage. Consequently, *only those T cells that specifically recognize the intruding antigen are produced in quantity.* Clonal selection occurs because an APC bound to a T cell releases a **cytokine** (cytokines are protein growth factors that regulate the differentiation, proliferation, and activities of the various types of blood cells; Section 19-3E_b) named **interleukin-1** (5a, 5b), which specifically stimulates

Cytotoxic T cell activation



Helper T cell activation



B cell activation for antibody production

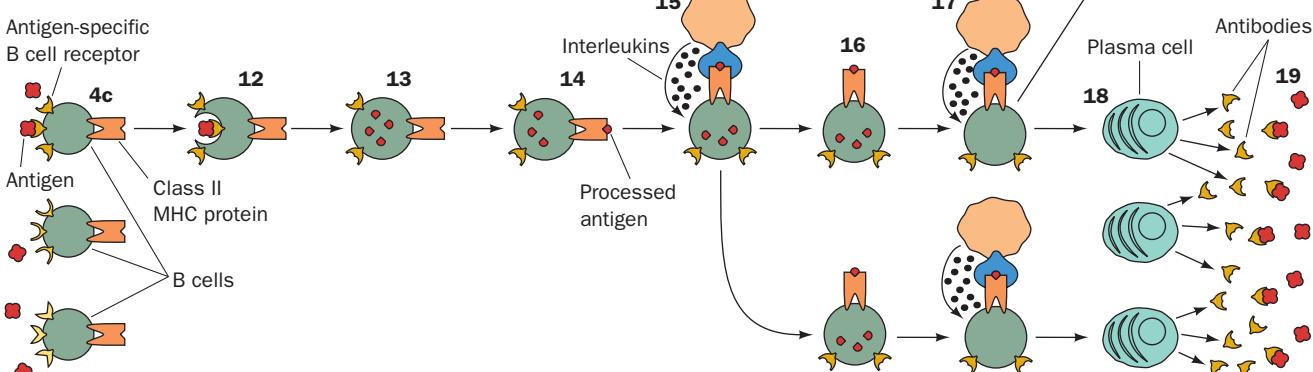


Figure 35-18 An outline of the adaptive immune response. See the text for an explanation. [After Marrack, P. and Kappler, J., *Sci. Am.* **254**(2), 38–39 (1986). Copyright © 1986 by Scientific American, Inc.]

T cells to proliferate and differentiate (6a, 6b). This process is enhanced by the T cells' autostimulatory secretion of **interleukin-2**. This helps ensure that T cell proliferation and differentiation occur only when a large number of T cells simultaneously respond to antigen in a confined space, such as a lymph node, because only then does the interleukin-2 concentration reach an effective level. *T cells only make interleukin-2 receptor so long as they remain bound to an APC* (7a, 7b), thereby preventing unlimited T cell proliferation. Nevertheless, a large number of mature cytotoxic T cells (8a) targeted to the specific foreign antigen are generated starting a few days after the antigen is first encountered.

Class I MHC proteins are expressed by nearly all nucleated vertebrate cells. These cells turn over their component proteins via their proteasomes (Section 32-6Bc) and display the resulting proteolytic fragments on their surfaces in complex with their Class I MHC proteins. Fragments of any foreign antigens that the cell contains are likewise displayed, thereby permitting the cytotoxic T cells that are targeted to these antigens to bind to the cell via their TCRs (9).

A cytotoxic T cell lives up to its name (10): At the point of contact with its target cell it releases the protein **perforin**, which lyses the target cell by aggregating to form pores in its plasma membrane (Section 35-2Fd). It simultaneously releases the serine proteases **granzymes B** and **A**, which enter the target cell through the perforin pores and induces its apoptosis by activating **caspases** (Section 34-4Ea), cleaving several chromatin-associated proteins, and destroying the nuclear envelope by cleaving its component intermediate filament proteins known as **lamins** (which are also caspase substrates). Cytotoxic T cells also display **Fas ligand (FasL)**, which signals the associated target cell to undergo apoptosis via the apoptotic extrinsic pathway (Section 34-4Ed).

The cellular immune system functions mainly to prevent the spread of a viral infection by killing virus-infected host cells. It is also effective against bacteria and parasites that inhabit the cellular interior (where they are protected from attack by antibodies) and certain types of cancers. Indeed, the cellular immune system's vital function has become painfully evident since 1981 through the tragic spread of **acquired immune deficiency syndrome (AIDS**; Section 15-4C), whose causative agent, **human immunodeficiency virus (HIV)**, acts by specifically attacking helper T cells. The cellular immune system is also responsible for various difficulties elicited by modern medicine that do not occur in nature such as the rejection of tissue and organ grafts from foreign donors. Such grafts, which are recognized as foreign because they almost always bear MHC proteins that differ from those of the host, have only been made possible by the development of drugs known as **immunosuppressants**, such as cyclosporin A and FK506 (Sections 9-2Ba and 19-3Ff), that suppress the immune response (but not so much as to leave the body defenseless against pathogens).

b. The Humoral Immune System

B cells display both immunoglobulins and Class II MHC proteins (4c) on their surfaces. If a B cell encounters an antigen that binds to its particular immunoglobulin, it engulfs the complex (12), partially digests the antigen (13), and displays

the fragments on its surface in complex with the Class II MHC protein (14; B cells are a type of APC). Mature helper T cells (8b) bearing receptors specific for this complex bind to the B cell (15) and, in response, release interleukins that stimulate the B cell to proliferate and differentiate (16). Cell division continues so long as the B cells are stimulated by the helper T cells (17) which, in turn, depends on the continuing presence of antigen (1b-8b). *Most of the terminally differentiated B cell progeny are plasma cells (18) that are specialized to secrete large amounts of the antigen-specific antibodies (several thousand per second; they contain large amounts of rough endoplasmic reticulum; Fig. 34-2). The antibodies bind to the available antigen (19), thereby marking it for destruction either by phagocytosis by white cells known as phagocytes or by activating the complement system (a series of interacting proteins of the innate immune system that lyse cells and trigger local inflammatory reactions; Section 35-2F).*

Most T cells and B cells live only a few days unless stimulated by their corresponding antigen. Moreover, the proliferation of T and B cells is limited by their interactions with **regulatory T cells** (also called **suppressor T cells**), an additional type of T lymphocyte progeny which have essentially the opposite function of helper T cells. Yet, one of the hallmarks of the immune system is that an animal is rarely infected twice by exactly the same type of pathogen; that is, *recovery from an infection by a pathogen renders an animal immune from that pathogen*. This so-called **secondary immune response** is mediated by long-lived **memory T cells (11)** and **memory B cells (20)** which, upon reencountering their cognate antigen, perhaps decades after its previous appearance, proliferate much faster and more massively than do "virgin" T and B cells (those that have never encountered their corresponding antigen) as is indicated in Fig. 35-19. This characteristic

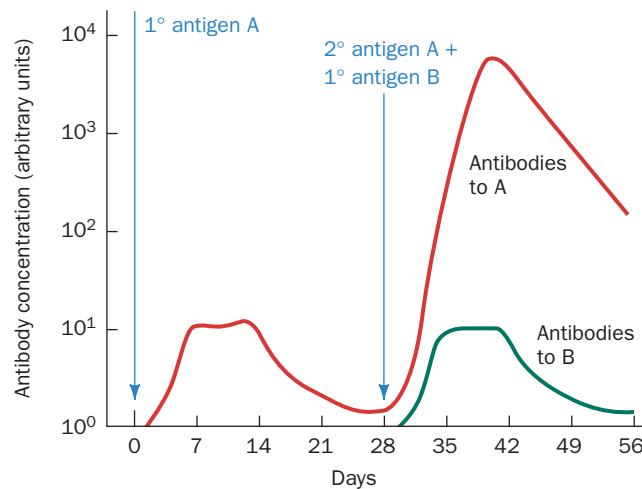


Figure 35-19 Primary and secondary immune responses. The rates of appearance of antibodies in the blood serum following primary (1°) immunization on day 0 with antigen A and secondary (2°) immunization on day 28 with antigens A and B. Antigen B is included in the secondary immunization to demonstrate the specificity of immunological memory for antigen A. Note that the secondary response to antigen A is both faster and greater than the primary response. T cell-mediated responses exhibit similar immunological memory.

of the immune system has been recognized since ancient times: The Greek historian Thucydides noted over 2400 years ago that the sick could be treated by those who had recovered, for a man was never attacked twice by the same disease. **Vaccines**, which present antigens characteristic of a particular pathogen to the immune system in a harmless way, exploit the secondary immune response.

C. The Immune System Is Self-Tolerant

Nearly all biological macromolecules are antigenic. *To prevent self-destruction, an animal's immune system must therefore discriminate between self-antigens and foreign antigens.* Such a process must be exquisitely selective. After all, a vertebrate, for example, has tens of thousands of different macromolecules, each with numerous distinctive antigenic sites.

What is the mechanism of immunological **self-tolerance**? The immune system in mammals becomes active around the time of birth. If a foreign antigen is implanted in an embryo before this time, the resulting animal is unable to mount an immune attack against that antigen. This is mainly because the immune system eliminates the clones of B and T cells that recognize the antigens that were present during the critical period when the immune system became active by inducing their apoptosis (**clonal deletion**). Yet, because new clones of lymphocytes, each with a nearly unique set of antigenic determinants, arise throughout an animal's lifetime (Section 35-2C; antibodies and TCRs are themselves antigenic), self-tolerance must be an ongoing process. This process, although of dimly perceived mechanism, has been shown to occur in the bone marrow and thymus where *only those virgin B and T cells that have no affinity for self-antigens are selected for further propagation*. Indeed, only $\sim 0.1\%$ of the lymphocytes that are processed by the bone marrow and thymus ever leave those organs.

Occasionally, the immune system loses tolerance to some of its self-antigens, resulting in an **autoimmune disease**. For example, **myasthenia gravis**, an autoimmune disease in which individuals make antibodies against the acetylcholine receptors of their own skeletal muscles

(acetylcholine is a neurotransmitter that triggers muscle contraction; Section 20-5Cb), results in a progressive and often fatal muscular weakness. Similarly, individuals with **systemic lupus erythematosus**, an often fatal inflammatory disease, produce antibodies against many of their own cellular components including certain ribonuclear proteins (Section 31-4Ah) and DNA. Other common autoimmune diseases include **rheumatoid arthritis** (in which the immune system attacks the connective tissue in the joints), **insulin-dependent diabetes mellitus** (Section 27-4Ba), **multiple sclerosis** [in which the immune system destroys the myelin insulation surrounding nerve fibers in the brain and spinal cord (Section 20-5Bc), thereby causing paralysis], **psoriasis** (in which the immune system attacks the skin), **Crohn's disease** and **ulcerative colitis** (in which the immune system attacks the intestinal lining), and **Addison's disease** [in which the immune system attacks the adrenal cortex (Section 19-1Ga)].

B. Antibody Structures

The immunoglobulins form a related but yet enormously diverse group of proteins. In this section, we consider the structures of these essential molecules. How their diversity is generated is the subject of the following section.

a. There Are Five Isotypes of Immunoglobulins

Most immunoglobulins, and the basic building blocks of all of them, consist, as Gerald Edelman and Rodney Porter showed, of four subunits: two identical $\sim 23\text{-kD}$ **light chains (L)** and two identical 53- to 75-kD **heavy chains (H)**. These subunits associate via disulfide bonds as well as by noncovalent interactions to form, as electron micrographs indicate, a Y-shaped symmetric dimer, $(\text{L}-\text{H})_2$ (Fig. 35-20). Immunoglobulins are glycoproteins; each heavy chain has an N-linked oligosaccharide.

Humans have five **isotypes** (alternatively, **classes**) of secreted immunoglobulins, designated **IgA** (for **immunoglobulin A**), **IgD**, **IgE**, **IgG**, and **IgM**, which differ in their corresponding heavy chains, designated α , δ , ϵ , γ , and μ , respectively (Table 35-2). There are also two types of light chain,

Table 35-2 Isotypes of Human Immunoglobulins

Class	Heavy Chain	Light Chain	Subunit Structure	Molecular Mass (kD)
IgA ^b	α	κ or λ	$(\alpha_2\kappa_2)_n\text{J}^a$ $(\alpha_2\lambda_2)_n\text{J}^a$	360–720
IgD	δ	κ or λ	$\delta_2\kappa_2$ $\delta_2\lambda_2$	160
IgE	ϵ	κ or λ	$\epsilon_2\kappa_2$ $\epsilon_2\lambda_2$	190
IgG ^b	γ	κ or λ	$\gamma_2\kappa_2$ $\gamma_2\lambda_2$	150
IgM	μ	κ or λ	$(\mu_2\kappa_2)_5\text{J}$ $(\mu_2\lambda_2)_5\text{J}$	950

^a $n = 1$ or 2 .

^bIgA has two subtypes, IgA1 and IgA2, which differ in their α chains; and IgG has four subtypes, IgG1, IgG2, IgG3, and IgG4, which differ in their γ chains.

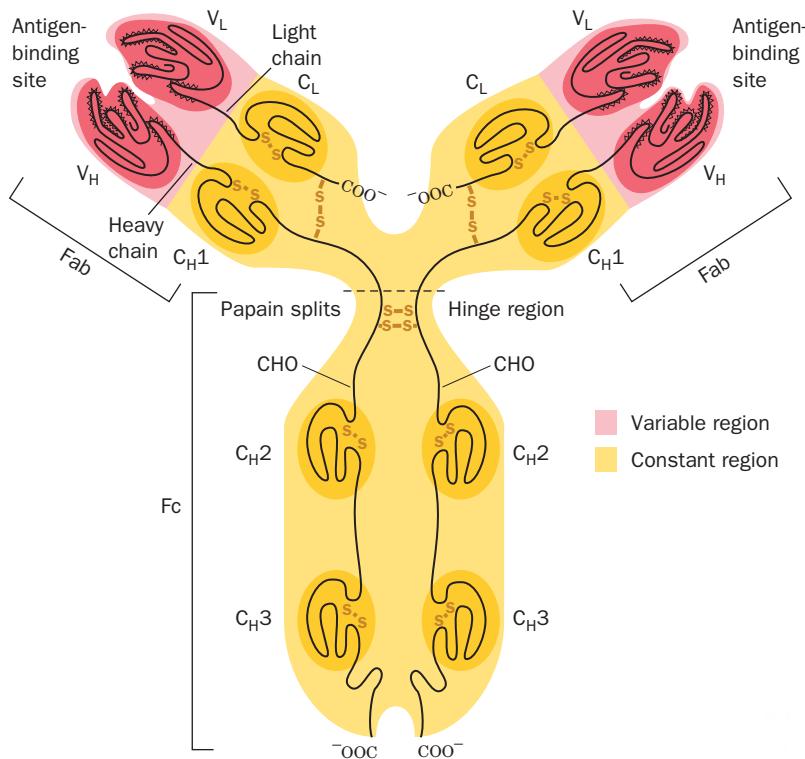
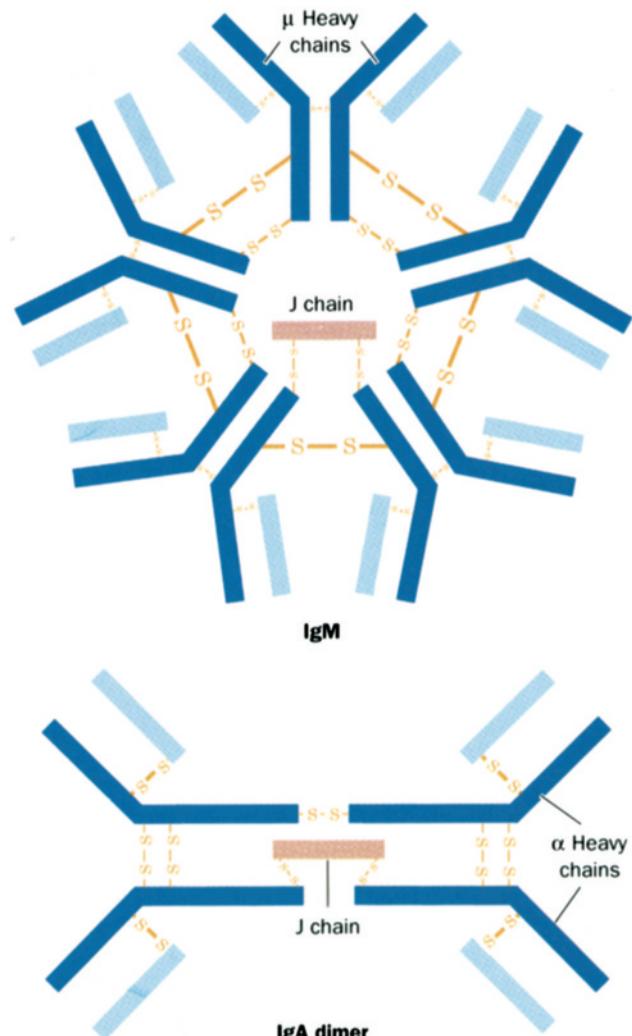


Figure 35-20 Diagram of the human IgG molecule. Each light (L) chain consists of two homologous units, V_L and C_L, where V and C indicate the polypeptide chain's variable and constant regions. Each heavy (H) chain is composed of four such units, V_H, C_{H1}, C_{H2}, and C_{H3}. Treatment of IgG by the proteolytic enzyme papain results in the cleavage of this immunoglobulin molecule in its "hinge" region yielding two Fab fragments and one Fc fragment. CHO represents carbohydrate chains.

[Illustration, Irving Geis/Geis Archives Trust. Copyright © Howard Hughes Medical Institute. Reproduced with permission.]



κ and λ , but these occur in immunoglobulins of all isotypes. IgD, IgE, and IgG exist only as (L-H)₂ dimers. IgM, however, consists of pentamers of its respective dimers, and IgA occurs mainly as a dimer of its dimers (Fig. 35-21) but may also be a monomer. The dimeric units of these multimers are linked by disulfide bonds to each other and to an ~20-kD protein termed the **joining chain (J)**. All of these isotypes also occur in B cell–displayed monomeric membrane-bound forms. It is antigen binding by this latter form of IgM that triggers the humoral immune response.

The various isotypes of secreted immunoglobulins have different physiological functions. IgM, which is largely confined to the blood, is most effective against invading microorganisms. It is the first immunoglobulin to be secreted in response to an antigen; its production begins 2 to 3 days after antigen is first encountered. IgG, the most common immunoglobulin in internal body fluids, is equally distributed between the blood and the interstitial fluid. It is the only antibody that can cross the placenta (via receptor-mediated endocytosis) and thus provide the fetus with immunity. IgG production begins 2 to 3 days after IgM first appears. IgA, which is made in greater quantities than all other immunoglobulin isotypes combined, occurs predominantly in the mucosal areas of the intestinal, respiratory, and urogenital tracts and in such secretions as saliva, sweat, and tears; it defends against invading pathogens by binding to their antigenic sites so as to block their attachment to epithelial (outer) surfaces (and hence are called **neutralizing antibodies**). IgA is also the major antibody of milk and colostrum (the first milk secreted after pregnancy) and thereby protects nursing infants from gastrointestinal invasion by pathogens. IgE, which is normally present in the blood in minute concentrations, protects against parasites

Figure 35-21 IgM and IgA. The five dimeric subunits of IgM (top) are held together by disulfide bonds. A single J chain joins two of the pentamer's μ heavy chains. However, IgM can also form hexamers that lack a J chain. The J chain also participates in joining IgA chains to form dimers (bottom).

and is implicated in allergic reactions. IgD, which is also present in blood in very small amounts, is of unknown function. Indeed, transgenic mice that cannot synthesize IgD appear to have normal immune systems.

b. Immunoglobulin's Functional Segments May Be Proteolytically Separated

In 1959, Porter showed that IgG, the most common isotype of immunoglobulin, is cleaved, through limited proteolysis with the papaya protease papain, into three ~ 50 -kD fragments: two identical **Fab fragments** and one **Fc fragment**. The Fab fragments, which form the arms of the Y-shaped IgG molecule and which each consist of an entire L chain and the N-terminal half of an H chain (Fig. 35-20), contain IgG's antigen-binding sites ("ab" stands for *antigen binding*). Immunoglobulin's consequent divalent (or, for IgA and IgM, multivalent) antigen-binding character forms the basis of the **precipitin reaction**, a sensitive test that has long been used for determining the presence of antibody or antigen: A mixture of antibody and the antigen against which it is directed combine as an extended cross-linked lattice (most antigens have multiple antigenic determinants) that yields an easily detected precipitate (Fig. 35-22). *The formation of these cross-linked lattices enhances antibody–antigen binding through cooperative interactions and is required to trigger B cell proliferation.*

The Fc fragment (so named because it is readily crystallized) derives from the stem of the Y and consists of the identical C-terminal segments of two H chains (Fig. 35-20). *Fc fragments contain the effector sites that mediate the functions common to a particular immunoglobulin isotype such as inducing phagocytosis, triggering the complement system (Section 35-2F), and directing the transport of immunoglobulins to their sites of action.* They do so by binding to the

corresponding **Fc receptors** that are displayed on the surfaces of many types of immune system cells.

c. IgG's Heavy and Light Chains Both Have Constant and Variable Regions

In order to characterize a molecule, it is necessary to obtain it in reasonably pure form. This requirement, at first, presented immunologists with a seemingly insurmountable obstacle. Exposing an animal to a particular antigen elicits the formation of numerous clones of plasma cells, each of which synthesizes a slightly different immunoglobulin molecule that binds the antigen. The resulting antibodies are therefore quite heterogeneous. This obstacle was largely removed by the demonstration, in the early 1960s, that an individual with **multiple myeloma**, a plasma cell cancer, synthesizes large amounts of a single species of immunoglobulin termed a **myeloma protein**. Some myelomas make excess light chains, which, when they are excreted in the urine, are known as **Bence Jones proteins** (after Henry Bence Jones who first described them in 1847).

The amino acid sequences of several different Bence Jones proteins, which each have 214 residues, revealed that *the sequence differences among light chains are largely confined to their N-terminal halves.* Light chains are therefore said to have a **variable region**, **V_L**, spanning residues 1 to 108, and a **constant region**, **C_L**, comprising residues 109 to 214 (Fig. 35-20). Similarly, comparisons of myeloma heavy chains, which have 446 residues, revealed that all the sequence differences among them occur between residues 1 to 125. *Thus heavy chains also have a variable region, V_H, and a constant region, C_H* (Fig. 35-20). Complexes of only the **V_H** and **V_L** regions, which are usually generated by genetic engineering, are known as **Fv fragments**.

Additional sequence comparisons indicated that the **C_H** region of IgG consists of three ~ 110 -residue segments, **C_H1**, **C_H2**, and **C_H3** (Fig. 35-20), which are homologous to each other and to **C_L**. In fact, even the constant and variable sequences are related albeit not as closely as the members of each of these groups are related to each other. These homologies, together with the observation that each **C_H** homology unit is cross-linked by a disulfide bond, correctly suggest (see below) that *an IgG molecule's 12 homology units each fold into an independent domain.* Evidently, modern light chain and heavy chain genes evolved through duplications of a primordial gene encoding an ~ 110 -residue protein. Indeed, each **C_H** domain is encoded by a separate exon.

The other immunoglobulins, IgA, IgD, IgE, and IgM, have similar structural organizations to that of IgG. All immunoglobulins secreted by a particular clone have identical light chains and **V_H** regions. However, their **C_H** regions differ in the numbers and positions of their intersubunit disulfide bonds and carbohydrate units. Furthermore, the heavy chains of IgE and IgM each have an additional constant domain that replaces the so-called **hinge region**, the polypeptide segment joining each Fab region to its Fc region in IgG (Fig. 35-20) as well as in IgA and IgD.

The **V_L** and **V_H** regions are not uniformly variable. Rather, most of their amino acid variations are concentrated

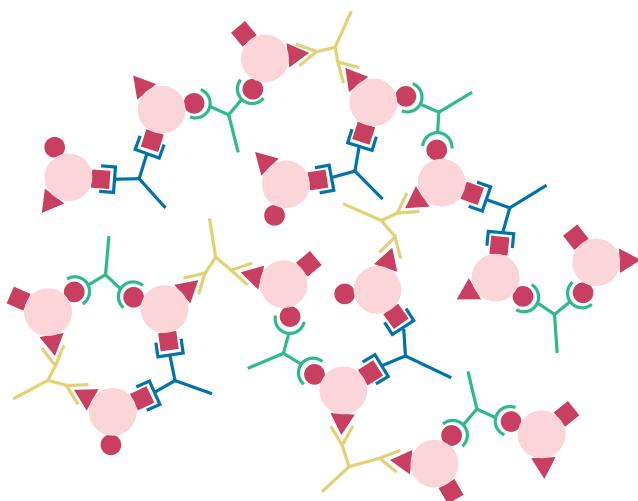


Figure 35-22 Antigen cross-linking by antibodies. A mixture of divalent (or multivalent) antibodies that recognize the several different epitopes (antigenic regions; red) of an intruding particle (pink) such as a protein or bacterium can form a lattice of antigen and antibody molecules.

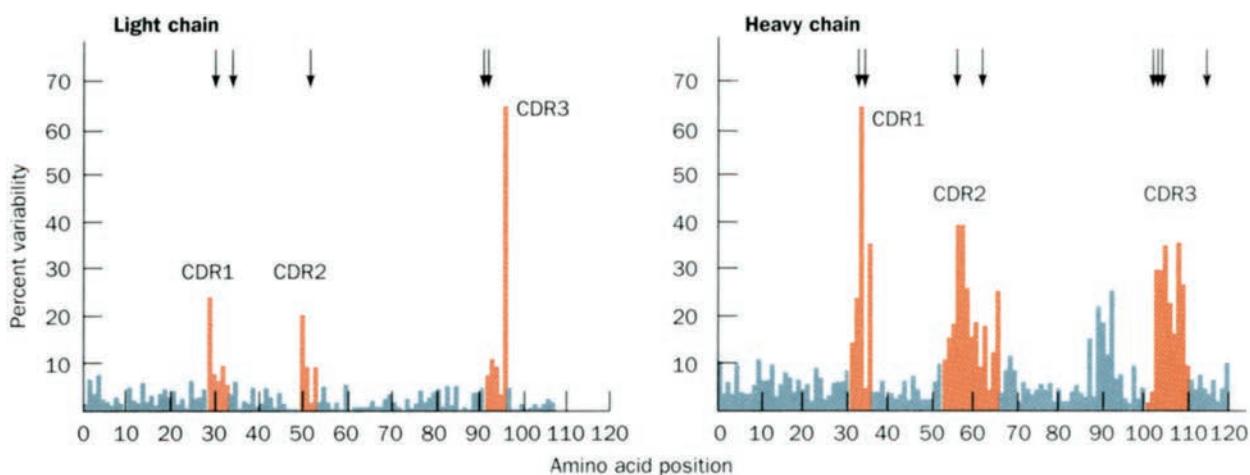
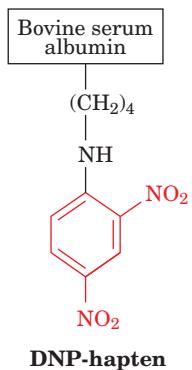


Figure 35-23 Sequence comparisons of a number of immunoglobulins. The hypervariable segments (orange bars) are responsible for most of the sequence variation in the variable

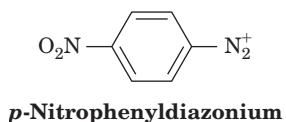
regions of both the light and heavy chains. The arrows mark the sites on anti-DNP antibodies that are derivatized by the affinity label *p*-nitrophenyldiazonium.

into three short **hypervariable sequences** (Fig. 35-23). Elvin Kabat therefore predicted that *the hypervariable sequences line the immunoglobulin's antigen-binding site and that their amino acids determine its binding specificity*.

Kabat's hypothesis was supported by affinity labeling experiments. Molecules of <5 kD are rarely antigenic. Yet, when small organic groups termed **haptens**, such as the **2,4-dinitrophenyl (DNP)** group, are covalently attached to a carrier protein such as bovine serum albumin (by reaction of fluorodinitrobenzene with its Lys residues)



and then injected into an animal, the animal produces antibodies that bind to the hapten in the absence of the carrier. If the DNP analog ***p*-nitrophenyldiazonium**



is combined with anti-DNP antibodies, the hapten's highly reactive diazonium group will form diazo bonds with the His, Lys, and Tyr side chains in the vicinity of the antibodies' DNP-binding sites (affinity labeling; Section 15-3Ab). Most of the side chains so derivatized are, in fact, members of the antibodies' hypervariable sequences (Fig. 35-23),

thereby indicating that *antigen-binding sites are lined with hypervariable residues*. Immunoglobulins' hypervariable segments are therefore also called **complementarity-determining regions (CDRs)**.

d. Monoclonal Antibodies Are Both Biomedical Tools and Therapeutic Agents

One might expect that homogeneous immunoglobulins could be obtained in quantity by simply cloning a single lymphocyte and harvesting the immunoglobulin the clone produced. Unfortunately, lymphocytes do not grow continuously in culture. In the late 1970s, however, César Milstein and Georges Köhler developed a technique for immortalizing such clones (Fig. 35-24). **Monoclonal antibodies** can now be obtained in virtually unlimited quantities and specific for almost any antigen by fusing myeloma cells with lymphocytes raised against that antigen (that is, isolated from an animal that has been immunized with the antigen). A clone of the resulting **hybridoma** (hybrid myeloma) cell synthesizes the lymphocyte's immunoglobulin but has the myeloma cell's immortality. Monoclonal antibodies have become indispensable biomedical tools; they can be used to assay for and to isolate extremely small amounts of nearly any specific biological substance. For example, they have made possible the routine testing of blood for the presence of HIV (AIDS virus), thereby protecting the public blood supply. More recently, **monoclonal antibodies that act as antitumor and anti-inflammatory agents have come into clinical use**. For example, the monoclonal antibody named **ce-tuximab** (trade name **Erbitux**) targets colorectal cancer cells that express epidermal growth factor receptor (EGFR; Section 19-3A), thereby blocking (neutralizing) the action of this oncogene product, whereas **infliximab** (trade name **Remicade**) binds the cytokine **tumor necrosis factor α (TNF α)**, which inhibits the progression of the autoimmune diseases rheumatoid arthritis, Crohn's disease, and ulcerative colitis.

Figure 35-24 Procedure for producing monoclonal antibodies against an antigen, X. HAT medium, so called because it contains hypoxanthine, amethopterin (methotrexate, an antifolate; Section 28-3Be), and thymine, prevents the growth of mutant cell lines lacking hypoxanthine–guanine phosphoribosyl transferase (HGPRT, a purine salvage enzyme that catalyzes the formation of the AMP and GMP precursor IMP; Section 28-1D). The amethopterin blocks the *de novo* synthesis of purines, which *HGPRT*[−] cells cannot replace through salvage pathways. Thymine, whose synthesis is also inhibited by the amethopterin, is available from the HAT medium. *HGPRT*[−] myeloma cells are fused with spleen-derived lymphocytes from a mouse immunized against X, and the resultant preparation is transferred to a HAT medium. This treatment selects for fused cells (hybridomas): The *HGPRT*[−] myeloma cells cannot grow in HAT medium; lymphocytes, which make HGPRT, do not grow in culture; but the hybridoma cells, which have the lymphocytes' HGPRT and the myeloma cells' immortality, proliferate. Individual hybridoma cells are then cloned and screened for the production of anti-X antibody. A satisfactory clone can be grown in virtually unlimited quantities, either in culture or as a mouse tumor, so as to synthesize the desired amounts of monoclonal antibody.

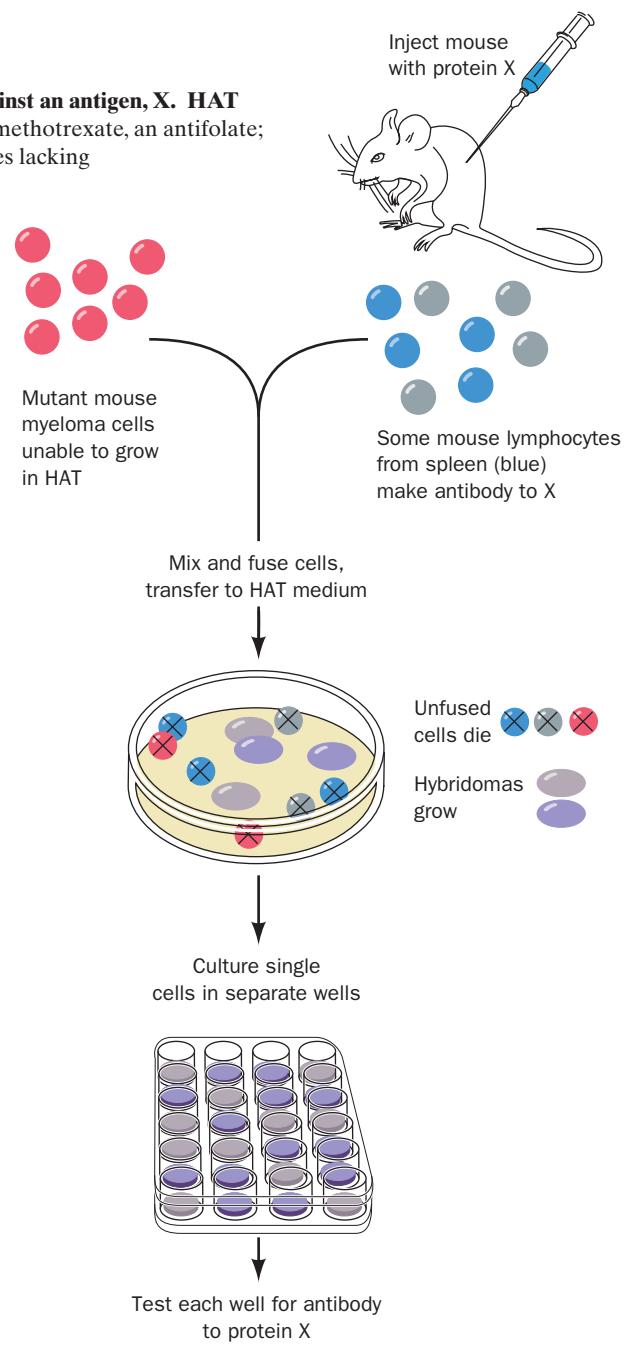


Figure 35-25 X-ray structure of a murine antibody against canine lymphoma (a type of cancer against which this antibody is therapeutically useful). The antibody is shown in ribbon form with its two heavy chains gold and cyan and its two light chains both magenta. Its two identical nonasaccharide chains, which are sandwiched between the protein's C_H2 domains, are drawn in space-filling form with C green, N blue, and O red. The antigen-combining sites are located at the ends of the two approximately horizontal Fab arms formed by the association of the light chains with the heavy chains. Compare this figure with Fig. 35-20. [Based on an X-ray structure by Alexander McPherson, University of California at Irvine. PDBid 1IGT.]

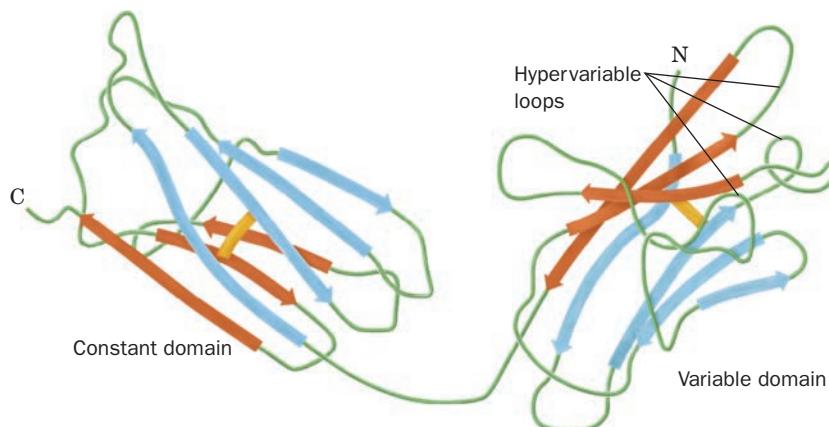


Figure 35-26 The chain folding of a myeloma protein light chain. Both its constant and variable domains assume the immunoglobulin fold: a barrel-shaped sandwich of a four-stranded antiparallel β sheet (blue arrows) and a three-stranded

antiparallel β sheet (brown arrows) that are linked by a disulfide bond (yellow). The positions of its three CDRs are indicated. [After Schiffer, M., Girling, R.L., Ely, K.R., and Edmundson, A.B., *Biochemistry* **12**, 4628 (1973).]

a three- and a four-stranded antiparallel β sheet that are linked by a disulfide bond (Figs. 35-26 and 8-48). The V domains differ from the C domains mainly by an additional polypeptide loop flanking each V domain's three-stranded β sheet.

Immunoglobulins, as both physicochemical studies and X-ray structural analyses indicate, exhibit considerable intersegmental flexibility (Fig. 35-25). This is particularly evident in IgG's hinge region (Fig. 35-20), although these regions are more like tethers than hinges as Fig. 35-25 (in which the two hinge angles are $\sim 65^\circ$ and $\sim 115^\circ$) strikingly indicates.

Since the basic IgG structure must accommodate an enormous variety of antigens, its flexibility presumably facilitates antigen binding by permitting an optimal fit between the antigen and its combining site. The IgG's carbohydrate moiety extends from between the C_{H1} and C_{H2} homology units and therefore also modulates the interactions between the Fab and Fc regions.

f. Antigen-Binding Sites Are Complementary to Their Corresponding Antigens

Myeloma proteins, on which much of our structural knowledge of immunoglobulins is based, are produced by cancer cells that originally proliferated in response to unknown, if any, antigens. Nevertheless, haptens that bind to particular myeloma proteins have been identified by screening many different compounds.

The X-ray structures of several hapten–myeloma protein complexes indicate that an immunoglobulin's antigen-binding site is located at the tip of each Fab region in a crevice between its V_L and V_H units (Fig. 35-20). *The size and shape of this crevice depend on the amino acid sequences of the V_L and V_H units, and its walls are formed, as predicted, by the six hypervariable segments (CDRs; Fig. 35-27).* Antibody–hapten complexes, not surprisingly, resemble enzyme–substrate complexes; both types of associations involve van der Waals interactions, hydrophobic forces,

hydrogen bonding, and ionic interactions. Indeed antibody–hapten complexes and enzyme–substrate complexes have similar ranges of dissociation constants, from 10^{-4}



Figure 35-27 The X-ray structures of seven different V_H units superimposed on their conserved framework residues. The C_α backbone is colored light gray for framework (non-CDR) residues, red for CDR1, yellow for CDR2, and green for CDR3. Note that the conformational variation among these structures resides almost entirely in their CDRs. The CDRs of the V_L units are similarly varied. [Courtesy of Elizabeth Getzoff, Victoria Roberts, Michael Pique, and John Tainer, The Scripps Research Institute, La Jolla, California.]

to 10^{-10} M, which correspond to binding energies of 25 to 65 $\text{kJ} \cdot \text{mol}^{-1}$.

Antibody–hapten complexes are imperfect models of antibody–antigen complexes because a hapten only partially fills its corresponding antigen-binding site. However, the advent of monoclonal antibodies made it possible to determine the X-ray structures of protein antigens in complex with Fab's derived from the monoclonal antibodies raised against them. For example, the comparison of the X-ray structures of three complexes of hen egg white (HEW) lysozyme with Fab's derived from different anti-HEW lysozyme monoclonal antibodies indicates that each Fab binds to a largely independent, irregularly shaped, $\sim 750\text{-}\text{\AA}^2$ surface patch on lysozyme such that one molecule's protruding side chains fit neatly into depressions on its mate (Fig. 35-28a). In each of these associations, all six Fab CDRs participate in lysozyme binding. These complexes, much like other known protein–protein associations, are cemented by highly complementary and thus solvent-excluding sets of van der Waals interactions, salt bridges, and hydrogen bonds. In several of these interactions, the lysozyme backbone and side chains maintain conformations identical to those in isolated lysozyme (Section 15-2A), but in others, there are significant local conformational variations.

The exquisite specificity of anti-lysozyme antibodies for their antigenic sites is demonstrated by the effect of a single amino acid change on the lysozyme contact surface. The dissociation constant of the anti-HEW Fab named D1.3 with HEW lysozyme is 2.2×10^{-8} M. Yet the dissociation constant of this monoclonal antibody with those of the nearly identical egg white lysozymes from partridge, California quail, and turkey are all $>10^{-5}$ M. In all these latter lysozymes, Gln 121, which conspicuously protrudes from

the HEW lysozyme surface into its Fab antigen-binding site (Fig. 35-28b), is replaced by His.

What are the special characteristics, if any, of the **epitopes** (antigenic sites) to which antibodies bind? All of the above lysozyme epitopes consist of 14 to 16 surface residues from two or more polypeptide segments. Some of these residues exhibit high mobility (Section 9-4), but others do not. Thus, the observation that our sample of three antibody–lysozyme complexes covers nearly half of lysozyme's surface strongly suggests that *a protein's entire accessible surface is potentially antigenic*.

How do Fab's respond to the binding of antigen? The answer, determined in part through X-ray studies by Ian Wilson, is that it depends on the Fab and antigen in question. Comparison of the X-ray structures of an Fab from a monoclonal antibody directed against a 36-residue segment of influenza hemagglutinin (Section 33-4B), alone and in complex with a 9-residue fragment of this antigen that the Fab binds with high affinity (Fig. 35-29), reveals that antigen binding causes the Fab to undergo a major structural rearrangement, most notably through a conformational shift of its heavy chain CDR3. Indeed, had this conformational change not occurred, the Fab would have been sterically unable to bind the antigen, at least in the conformation it assumes in the complex. On the other hand, in a similar X-ray study by Wilson on an Fab directed against a 19-residue segment of **myohemerythrin** (a non-heme Fe-containing oxygen-binding protein from a marine worm), both alone and in complex with this antigen, the Fab underwent only a few small main chain and side chain conformational adjustments on binding this antigen.

In both of the foregoing complexes, the conformation of Fab-bound antigen is substantially different from its conformation in its intact parent protein. This raises the intriguing and as yet unanswered question of how an antibody can bind to both a peptide and to the protein from which the

(a)

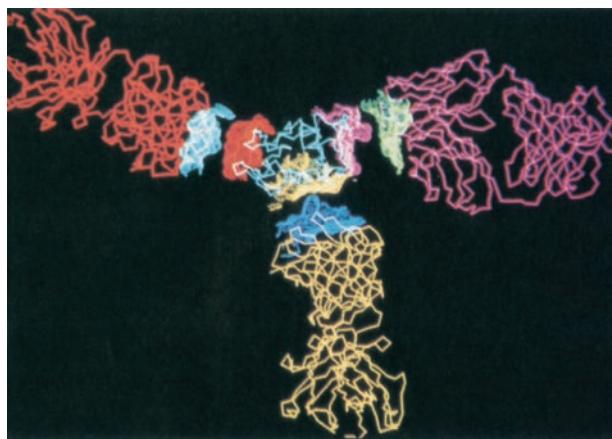
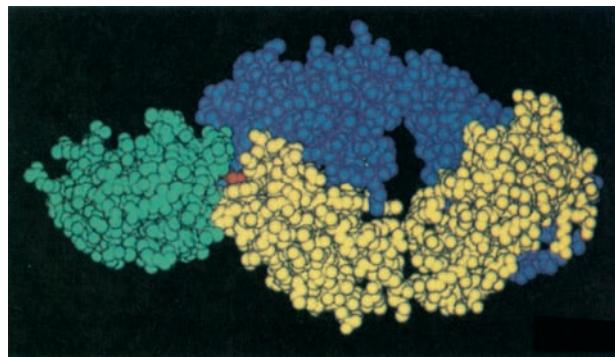


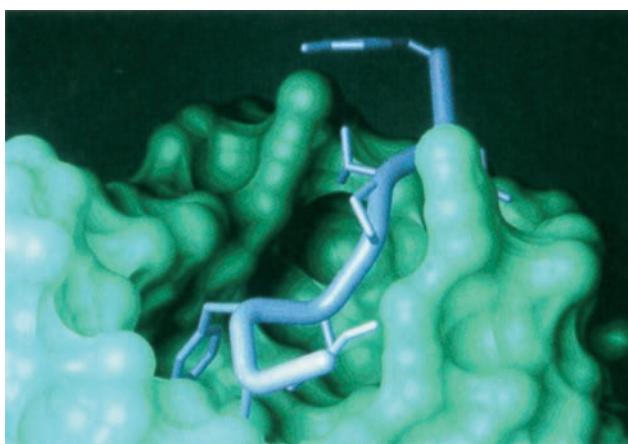
Figure 35-28 Hen egg white lysozyme in complex with the Fab fragments of monoclonal antibodies raised against it. (a) An exploded-view collage indicating how three different anti-lysozyme Fab's interact with lysozyme (center) in the X-ray structures of their respective complexes. The proteins are represented by their C_α chains, and their interacting surfaces are outlined by juxtaposed dot surfaces. Note that the three crystal structures on which this diagram is based each contain only one

(b)



Fab species; the Fab's do not crystallize together. [Courtesy of Steven Sheriff and David Davies, NIH. PDBids 1FDL, 2IFF, and 3HFM.] (b) The X-ray structure of HEW lysozyme in complex with the anti-lysozyme Fab named D1.3 (Part a, upper right). In this space-filling representation, the Fab's L chain is yellow, its H chain is blue, the lysozyme molecule is green, and lysozyme Gln 121 is red. [From Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., and Poljak, R.J., *Science* **233**, 749 (1986). PDBid 1FDL.]

(a)



(b)

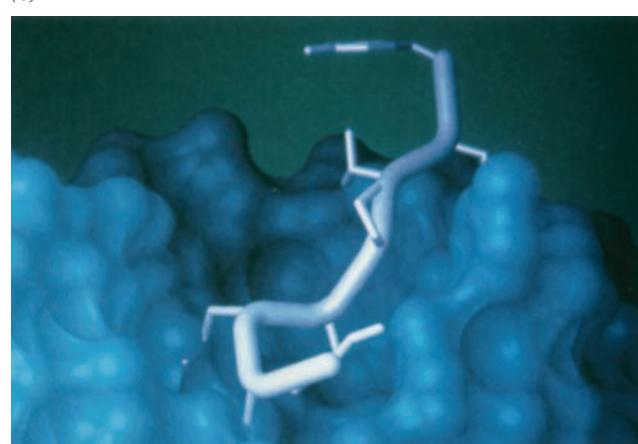


Figure 35-29 Comparison of the shapes of the antigen-binding pockets in the X-ray structures of the liganded and unliganded forms of an Fab derived from a monoclonal antibody directed against a 36-residue segment of influenza hemagglutinin. (a) The solvent-accessible surface of the Fab antigen-binding pocket (green) with its bound 9-residue peptide fragment of the antigen shown in stick form (light blue; sequence YDVPDYASL–amide). (b) The solvent-accessible surface of the Fab antigen-binding pocket in the absence of the peptide ligand (blue) with the

peptide positioned as in the complex to illustrate the extensive conformational changes that the Fab makes on peptide binding. Note that the peptide's Tyr 6 side chain (lower left of Part a) would collide with a portion of the binding pocket of the unliganded Fab (Part b) formed by its heavy chain CDR3. Consequently, the complex could not form without these conformational changes. [Courtesy of Ian Wilson, The Scripps Research Institute, La Jolla, California. PDBIDs 1HIL and 1HIM.]

peptide is derived. The answer to this question will almost certainly have important consequences for the generation of peptide-based vaccines.

C. Generation of Antibody Diversity

The immune system has the capacity to generate antibodies against almost any antigen that it encounters; it can produce a virtually unlimited variety of antigen-binding sites. What is the origin of this enormous diversity? One might reasonably expect that immunoglobulin gene expression resembles that of other proteins in that every distinct H and L chain is encoded by a separate germline gene. If this were true, then to encode the billions of different antibodies each vertebrate appears capable of producing would require huge numbers of these genes. For example, it would require 10^3 H and L chain genes each to encode $10^3 \times 10^3 = 10^6$ different immunoglobulins. However, the human genome contains far fewer than this number of immunoglobulin genes. Consequently, this so-called **germline hypothesis** must be rejected.

Two other models for the origin of antibody diversity have been seriously considered:

1. The somatic recombination hypothesis, which was originally formulated in 1965 by William Dreyer and Claude Bennett, proposes that *antibody diversity is generated by genetic recombination among a relatively few gene segments encoding the variable region of an immunoglobulin chain*. This process occurs via intrachromosomal recombination during B cell differentiation so that each B cell clone expresses an all but unique immunoglobulin.

2. The somatic hypermutation hypothesis proposes that *antibody diversity arises through an extraordinarily high rate of immunoglobulin gene mutation during B cell differentiation*.

We shall see below that both of these mechanisms contribute to antibody diversity.

a. κ Light Chain Genes Are Assembled from Three Sets of Gene Segments

DNA sequencing studies by Leroy Hood, Philip Leder, and Susumu Tonegawa revealed that κ light chains are each encoded by four exons (Fig. 35-30):

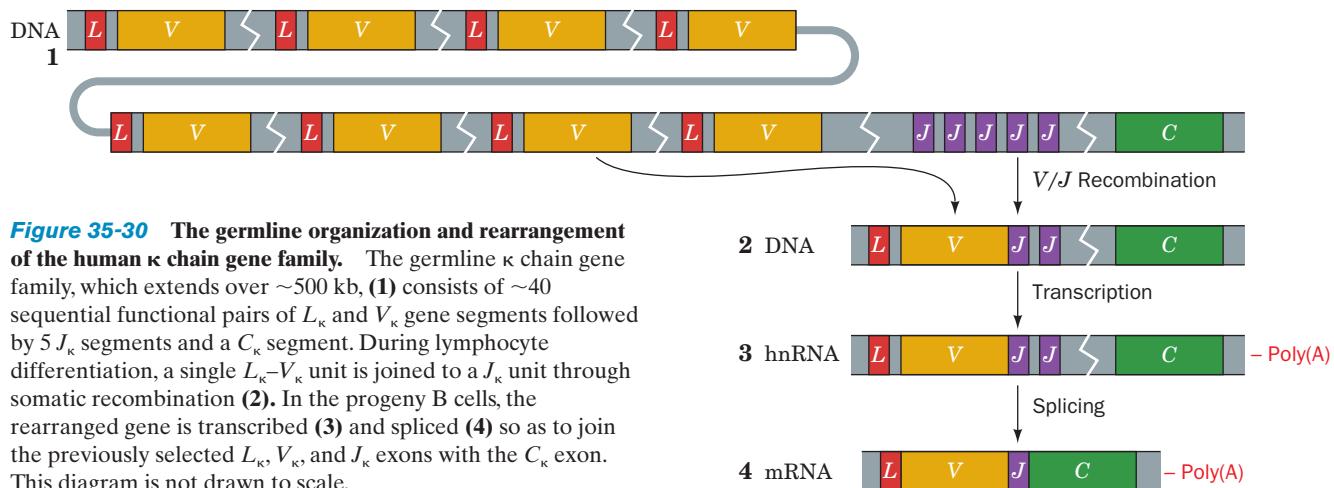
1. A leader or L_κ segment, which encodes a 17- to 20-residue hydrophobic signal peptide. This polypeptide directs newly synthesized κ chains to the endoplasmic reticulum and is then excised (Section 12-4Ba).

2. A V_κ segment, which encodes the first 95 residues of the κ chain's 108-residue variable region.

3. A joining or J_κ segment (not to be confused with the J chain of IgA and IgM), which encodes the variable region's remaining 13 residues.

4. The C_κ segment, which encodes the κ chain's constant region.

The arrangement of these exons in human embryonic tissues (which do not make antibodies) and nonlymphoid tissues differs strikingly from those in gene families we have previously encountered. The L_κ and V_κ segments are separated by an intron as occurs in other split genes. However, the κ chain gene family contains an array of ~ 40 of these ~ 400 -bp L_κ – V_κ units separated from each other by ~ 7 -kb



spacers. This series of exon pairs is followed, well downstream, by 5 J_{κ} segments at intervals of ~ 300 bp, a 2.4-kb spacer, and a single C_{κ} segment.

The assembly of a κ chain mRNA is a complex process involving both somatic recombination and selective gene splicing (Fig. 35-30) that occurs over several cell generations. The first step of this process, which occurs in a progenitor of each B cell clone, is an intrachromosomal recombination that joins an L_{κ} - V_{κ} unit to a J_{κ} segment and deletes the intervening sequences. Then, in later cell generations, the entire modified gene is transcribed and selectively spliced so as to join the L_{κ} - V_{κ} - J_{κ} unit to the C_{κ} segment. The L_{κ} and V_{κ} segments are also spliced together in this step, yielding an mRNA that encodes one of each of the elements of a κ chain gene.

Highly conserved sequences known as **recombination signal sequences (RSSs)** on the 3' side of each V_{κ} segment and on the 5' side of each J_{κ} segment suggest how the somatic recombination sites are selected. The V_{κ} sequence is immediately followed by the heptamer of consensus sequence CACAGTG, a nonconserved 12 ± 1 nucleotide spacer, and an AT-rich nonamer of consensus sequence ACAAAAAACC. The J_{κ} chain is preceded by the complementary heptamer, a nonconserved 23 ± 1 nucleotide spacer, and the complementary AT-rich nonamer. These sequences combine under the influence of a system of recombinatory enzymes, discussed below, to form a stem-and-loop structure (Fig. 35-31).

b. Recombinational Flexibility Contributes to Antibody Diversity

The joining of one of 40 V_{κ} segments to one of 5 J_{κ} segments can generate only around $40 \times 5 = 200$ different κ chains, far less than the number observed. However, studies of many joining events involving the same V_{κ} and J_{κ} segments revealed that the V/J recombination site is not precisely defined; these two gene segments can join at different crossover points (Fig. 35-32). Consequently, the amino acids specified by the codons in the vicinity of the V/J recombination site depend on what part of the sequence is supplied by the germline V_{κ} segment and what part is supplied by the

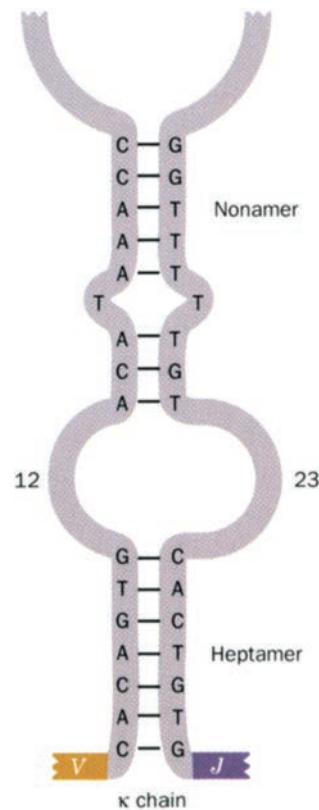


Figure 35-31 The germline κ gene family contains complementary heptamers and nonamers succeeding each V_{κ} segment and preceding each J_{κ} segment. These sequences mediate somatic recombination by forming the indicated stem-and-loop structure.

germline J_{κ} segment. Indeed, the amino acids specified by the codons surrounding the recombination junction form the light chain hypervariable region in the vicinity of residue 96 (CDR3; Fig. 35-23). Assuming that this recombinational flexibility increases the possible κ chain diversity 10-fold, the expected number of possible different κ chains is increased to around $40 \times 5 \times 10 = 2000$.

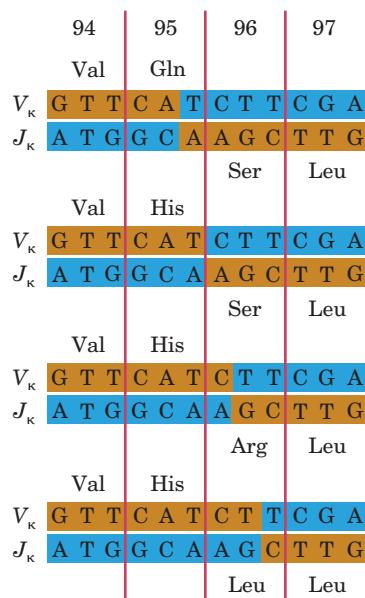


Figure 35-32 Flexibility in the V/J recombination site. The cross-over point at which the V_{κ} and J_{κ} sequences somatically recombine varies by several nucleotides, thereby giving rise to different nucleotide sequences (brown bands) in the active κ gene. For example, as is indicated here, amino acid 96, which occurs in the κ chain's third hypervariable region, can be Ser, Arg, or Leu.

The imprecision of the V/J joining often results in the random loss of a few nucleotides from the ends of the V_{κ} and J_{κ} segments (see below). Consequently, up to two-thirds of the recombination products have an out-of-phase reading frame downstream from the recombination joint, so the resulting gene encodes a nonsense protein. Such proteins are not expressed. Moreover, some gene segments are pseudogenes in that they have accumulated mutations that prevent them from forming functional proteins (although the numbers given here count only functional gene segments). *A cell in which a nonproductive recombination event has occurred will attempt further κ gene rearrangements between its remaining $L_{\kappa}-V_{\kappa}$ and J_{κ} units and, if all of these fail, will rearrange its λ genes (see below).* This phenomenon accounts for the observation that κ expressing cells rarely have their λ genes rearranged, whereas λ expressing cells nearly always have their κ genes rearranged. Cells that have exhausted their possibilities of forming a functional light chain are

eliminated. The mechanism by which a cell detects a productive recombination event is unclear.

c. λ Light Chains Derive from Multiple Constant Regions

The κ and λ chain gene families, which occur on different chromosomes, have different germline arrangements of their L , V , J , and C segments. Mice have only two $L_{\lambda}-V_{\lambda}$ segments, each followed by a pair of $J_{\lambda}-C_{\lambda}$ units. Mice therefore have relatively little λ chain diversity compared to that of their κ chains. This is probably why murine light chains are 95% κ and only 5% λ . Humans, in contrast, have $\sim 30 L_{\lambda}-V_{\lambda}$ followed by $4 J_{\lambda}-C_{\lambda}$ units (Fig. 35-33), which can form around $30 \times 4 \times 10 = 1200$ different λ chains. Human immunoglobulins contain κ and λ chains in approximately 2:1 ratio.

d. Heavy Chain Genes Are Assembled from Four Sets of Gene Segments

Heavy chain genes are assembled in much the same way as are light chain genes but with the additional inclusion of an ~13-bp diversity or D_H segment between their V_H and J_H segments. The human heavy chain gene family, which occurs on a different chromosome from either of the light chain gene families, consists of clusters of ~ 40 different L_H-V_H units, $\sim 25 D_H$ segments, 6 J_H segments, and 9 C_H segments, 1 for each of the 9 immunoglobulin isotypes and subtypes (Fig. 35-34). The D_H segments encode the core of the heavy chain's third hypervariable region (Fig. 35-23).

Germline V_H , D_H , and J_H segments are flanked by heptamer–nonamer recombination signals similar to those that occur in light chain genes (Fig. 35-35). Moreover, heavy chain V/D and D/J joining sites are subject to the same recombinational flexibility as are light chain V/J sites. This **$V(D)J$ joining** process is tightly regulated in that it occurs in a particular temporal order; that is, D_H is joined to J_H in an earlier cell generation than that in which V_H is joined to $D_H J_H$. Similarly, light chain assembly occurs in later cell generations from those in which heavy chain assembly occurs.

Assuming that recombinational flexibility contributes a factor of 100 toward heavy chain diversity, somatic recombination can generate some $40 \times 25 \times 6 \times 100 = 6 \times 10^5$ different heavy chains of a given isotype. Then, taking into account κ and λ chain diversity (and ignoring the fact that not every light chain can combine with every heavy chain to form a functional immunoglobulin), *there can be as many as $(2000 + 1200) \times 6 \times 10^5 = 1.9 \times 10^9$ different*

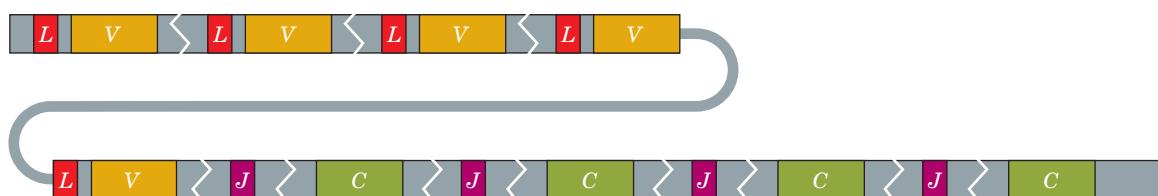


Figure 35-33 The germline organization of the human λ gene family. This gene family, which extends over ~ 900 kb, consists of ~ 30 sequential pairs of functional L_{λ} and V_{λ} segments

followed by four functional pairs of J_{λ} and C_{λ} segments. This diagram is not drawn to scale.

Germline heavy chain DNA

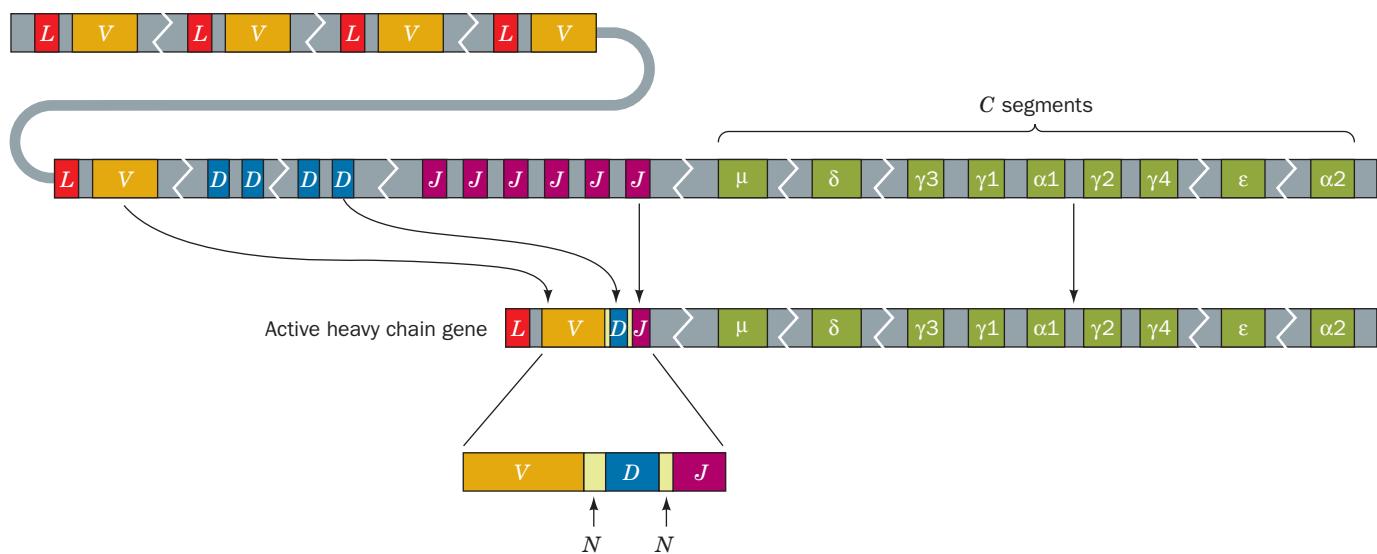


Figure 35-34 The germline organization and rearrangement of the human heavy chain gene family. The germline family, which extends over ~2 Mb, consists of ~40 sequential pairs of functional L_H and V_H gene segments followed by 25 D_H segments, 6 J_H segments, and 9 C_H segments (one for each isotype and subtype of heavy chain). For simplicity, the introns in the C_H segments are not shown (but see Fig. 35-37). During

lymphocyte differentiation, an L_H-V_H unit is recombinantly joined to a D_H segment and a J_H segment. In this process, the D_H segment becomes flanked by short segments of random sequence called N regions. In the B cell and its progeny, transcription and splicing join the $L_H-V_H-N-D_H-N-J_H$ unit to one of the eight C_H gene segments. This diagram is not drawn to scale.

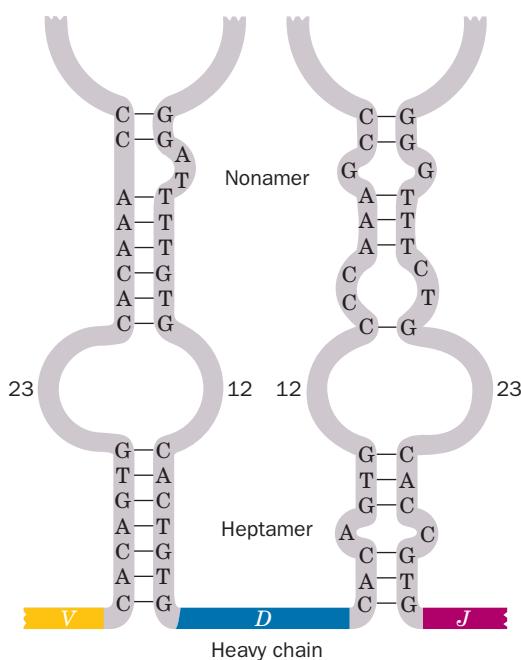


Figure 35-35 The stem-and-loop recombination sites in the germline heavy chain gene family. These mediate somatic recombination between its V_H and D_H segments (left) and between its D_H and J_H segments (right). Compare them to the κ chain recombination signal (Fig. 35-31). The recombination system's requirement for both 12- and 23-bp spacers, the so-called **12/23 rule**, prevents it from skipping the D_H segments by directly joining V_H and J_H segments.

types of human immunoglobulins of each isotype formed by somatic recombination among ~160 different gene segments.

e. RAG1 and RAG2 Proteins Are Essential Components of the $V(D)J$ Recombinase

The recombination signal sequences (RSSs) that mediate $V(D)J$ joining (Figs. 35-31 and 35-35), which are necessary and sufficient to direct this process, are conserved among the different loci and species that carry it out, and are functionally interchangeable. This latter observation suggests that all $V(D)J$ joining reactions are catalyzed by a single type of **$V(D)J$ recombinase**.

Through the use of recombinant DNA techniques, David Baltimore identified two highly conserved **recombination activating genes**, **RAG1** and **RAG2**, which are expressed only in developing lymphocytes engaged in assembling their antigen receptors. **RAG1** encodes the 1043-residue **RAG1** protein, whereas **RAG2**, which is chromosomally adjacent to **RAG1**, encodes the 527-residue **RAG2** protein that is unrelated to RAG1 or any other protein of known sequence. Neither RAG1 nor RAG2 alone exhibit any catalytic activity. However, together they form a heterodimeric endonuclease known as **RAG** that cleaves dsDNA between a $V(D)J$ coding segment and an RSS. Other components of the $V(D)J$ recombinase include the ubiquitously expressed proteins involved in the nonhomologous end-joining (NHEJ) of double-strand DNA breaks, Ku70, Ku80, DNA ligase IV,

and Xrcc4 (Section 30-5E), the DNA-bending proteins HMGB1 and HMGB2 (Section 34-3Bv), and two additional proteins known as **DNA-dependent protein kinase (DNA-PK)** and **Artemis**. Individuals with defects in their *V(D)J* recombinase cannot complete lymphocyte development and hence have severe combined immunodeficiency disease (SCID).

The *V(D)J* recombinase mediates the following series of reactions (Fig. 35-36):

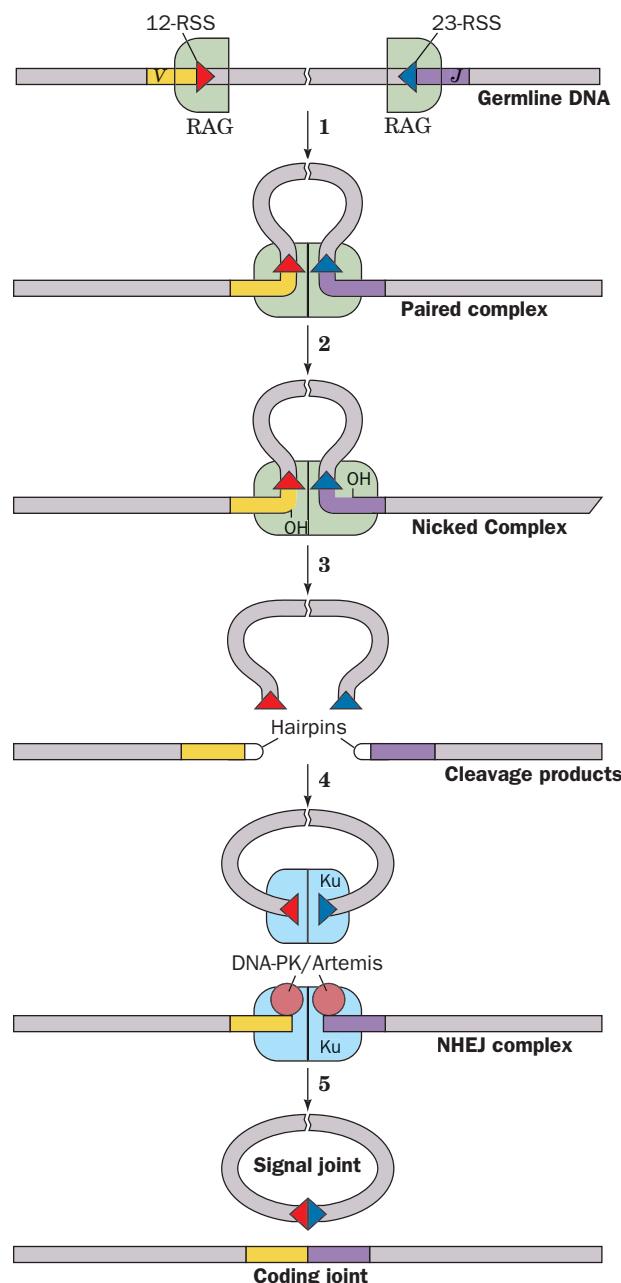


Figure 35-36 The mechanism of *V(D)J* joining. See the text for details. Note that the region between the selected *V* and *J* segments may be of considerable length and contain multiple *V* and several *J* segments.

1. RAG, together with HMGB1 and HMGB2, binds to the RSS that contains the 12-bp stem (the 12-RSS), another such complex binds to the RSS that contains the 23-bp stem (the 23-RSS), and the two complexes come together to form a paired complex that follows the 12/23 rule. A light chain *V-J* joining (Fig. 35-31) is drawn in Fig. 35-36, but it could also be a heavy chain *V-D* or *D-J* joining (Fig. 35-35). Presumably, the two RSSs that are chosen are selected at random.

2. RAG catalyzes a single-strand nick on each dsDNA at the 5' border of each RSS to yield a free 3'-OH group past each coding segment.

3. RAG mediates the nucleophilic attack of each 3'-OH group on the phosphodiester bond of its opposite strand to yield a closed hairpin on each coding segment and a flush double-strand break preceding each RSS.

4. Ku protein (a heterodimer of Ku70 and Ku80) binds to all four DNA ends, following which complexes of DNA-PK and Artemis open up each hairpin with a single-strand nick at or near the end of the hairpin.

5. Ku brings together the two coding segments, end-to-end, so as to permit nucleases to trim inappropriate ends at the variably positioned nicks and/or DNA polymerases to fill in gaps (Section 30-5E). This results in the sequence variability at the eventual recombination joint that is indicated in Fig. 35-32. The coding segments are then joined through the agency of DNA ligase IV and Xrcc4 (Section 30-5E) to yield a coding joint. Simultaneously, the NHEJ machinery precisely joins the heptamer sequences of the two RSSs to form a covalently closed circle (a signal joint) that is eventually degraded.

The similarities between the mechanisms of RAG-mediated DNA cleavage and transposon excision strongly suggest that the *V(D)J* recombination system evolved from a DNA transposon (Section 30-6B). Indeed, **RAG1** protein has an active site that resembles those of transposases and retroviral integrases. Moreover, under suitable conditions, RAG catalyzes the insertion of an RSS-containing DNA fragment into DNA in a reaction much like that catalyzed by DNA transposases.

f. Somatic Mutation Is a Further Source of Antibody Diversity

Despite the enormous antibody diversity generated by somatic recombination, *immunoglobulins are subject to even more variation through somatic mutations of two types:*

1. During *V/D* and *D/J* joining, a few nucleotides may be added at the recombination joints before they are sealed. The added nucleotides, which form so-called ***N* regions** (*N* for nontemplated), yield ND_H units of up to 30 bp that encode enormously variable heavy chain segments of 0 to 10 amino acid residues (Fig. 35-34). The *N* regions arise through the action of **terminal deoxynucleotidyl transferase (TdT)**, a template-independent DNA polymerase that is present in the B cell progenitors that make

heavy chain joints but is largely absent in later cell generations when the light chain joints are formed (N regions occur in only one-quarter of human light chain joints).

2. The variable regions of both heavy and light chains are more diverse than is expected on the basis of comparisons of their amino acid sequences with their corresponding germline nucleotide sequences. Indeed, these regions mutate at rates of up to 10^{-3} base changes per nucleotide per cell generation (i.e., about one base change per V coding region per cell generation), rates that are at least a millionfold higher than the rates of spontaneous mutation in other genes. This **somatic hypermutation** process, which occurs in B cells and/or their progenitors, is initiated by signals, mainly provided by T cells, that induce the expression of **activation-induced deaminase (AID)**, which deaminates deoxycytosine residues to yield uracil. AID only deaminates single-stranded DNA so that the DNA segment being deaminated must be simultaneously undergoing transcription, which generates a transcription bubble (Section 31-2Ab). The uracil may then participate in normal DNA replication, yielding a C · G to T · A transition (Section 32-1A). More frequently, however, the uracil is excised by uracil-DNA glycosylase (Section 30-5Bd) and replaced by a base excision repair (BER; Section 30-5Bc) process that employs error-prone translesion DNA polymerases (Sections 30-5Db and 32-6B1). Many of the resulting mutations decrease the affinity of an antibody for its antigen but some increase this affinity. Since the rate at which B cells are activated for proliferation increases with the affinity of their surface-displayed antibodies for antigen, *somatic hypermutation acts over many cell generations to tailor antibodies to a particular antigen.*

These somatic mutation processes increase the possible number of different antibodies that humans can produce by many orders of magnitude beyond the 1.9×10^9 we estimated on the basis of somatic recombination alone. The final number is so large, perhaps $>10^{18}$, that an individual synthesizes only a small fraction of its potential immunoglobulin repertoire over its lifetime. Somatic diversification arising from both recombination and mutation thereby permits an individual organism to cope, in a kind of Darwinian struggle, with the rapid mutational rates of pathogenic microorganisms.

g. Allelic Exclusion Ensures That Antibodies Are Monospecific

The immunoglobulins synthesized by a given B cell, as we have seen, consist of two identical heavy chains and two identical light chains. Such homogeneity is essential for the immune system's proper functioning because immunoglobulins consisting of two types of heavy and/or light chains would have two different antigen-combining sites and therefore could not form lattices of cross-linked antigens. Yet B cells, which like other somatic cells are diploid, contain two gene families specifying heavy chains (one maternal allele and one paternal allele) and four gene families encoding light chains (two κ 's and two λ 's). Apparently, B cells suppress the expression of all but one heavy chain allele and one light chain allele, a process known as **allelic exclusion**, by inhibiting further somatic recombination of

heavy and light chain genes after a productive recombination has occurred. Allelic exclusion was experimentally demonstrated by microinjecting plasmids containing an already recombined heavy chain gene into fertilized mouse ova. The resulting transgenic mice suppress the somatic recombination of their endogenous heavy chain genes. Analogous results were obtained for light chain genes. What is the mechanism of allelic exclusion?

A productive $V(D)J$ joining of a heavy chain gene results in the expression of a μ_m heavy chain, whose C-terminal 41 residues form a hydrophobic membrane anchor. At this stage in B cell differentiation, the **pre-B cell**, the μ_m chain combines with the so-called **surrogate light chain**, which is composed of two proteins, **VpreB**, which resembles a light chain V domain with an additional N-terminal region, and **$\lambda 5$** , which resembles the C_λ domain (in pre-B cells, light chains have yet to be assembled). On binding its corresponding antigen, this **pre-B cell receptor** recruits two transmembrane proteins, **Ig α** and **Ig β** , both of whose cytoplasmic tails contain an **immunoreceptor tyrosine-based activation motif (ITAM)**. This induces the associated Src-like nonreceptor tyrosine kinases **Fyn** and **Lyn** to phosphorylate the two Tyr residues on each ITAM, thereby initiating a kinase cascade that acts to inhibit the expression of RAG, thus preventing all further $V(D)J$ joining of heavy chain genes. This is consistent with the observation that in mice in which the $\lambda 5$ gene has been knocked out, the heavy chains on both chromosomes are rearranged with $\sim 10\%$ of pre-B cells having two productive rearrangements.

h. The Switch from the Membrane Bound to the Secreted Form of an Antibody Involves a Change in the Polyadenylation Site of Its Heavy Chain Transcript

The clonal selection model of antibody generation requires that the antibody displayed on the surface of a virgin B cell have the same specificity for antigen as the antibody secreted by its mature B cell progeny. Membrane-bound IgM (the antibody synthesized by virgin B cells) is an $(L-H)_2$ dimer that contains μ_m heavy chains. In the pentameric secreted form of IgM (Fig. 35-21; the first antibody secreted by mature B cells), the heavy chain (μ_s) has a C-terminal segment that differs from μ_m 's hydrophobic membrane anchor but is otherwise identical. How does the B cell alter the synthesis of this heavy chain?

Somatically recombined heavy chain genes consist of eight exons (Fig. 35-37): an L segment that encodes a signal peptide leader; a VDJ unit that encodes the V_H domain; four exons that encode the C_{H1} domain, the hinge region, the C_{H2} domain, and the C_{H3} domain; and two exons that collectively encode the transmembrane tail of μ_m . In forming μ_m mRNA (Fig. 35-37, *left*), the primary transcript is cleaved and polyadenylated (Section 31-4Ab) at a site following the exons encoding the transmembrane tail and the splicing system then excises the segment at the end of the C_{H3} exon that specifies the μ_s tail (i.e., it is an intron under these circumstances). However, in forming μ_s mRNA (Fig. 35-37, *right*), cleavage and polyadenylation occur at a site that precedes these latter exons and the splicing system retains the segment specifying the μ_s tail. Note that all of the

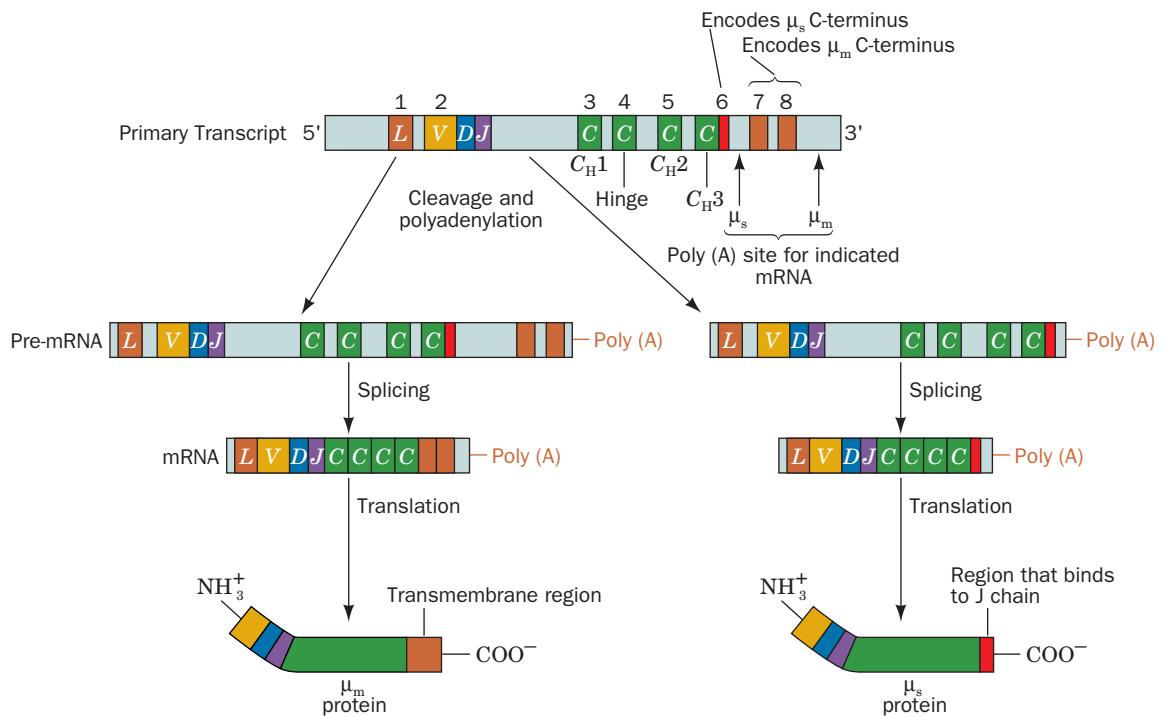


Figure 35-37 The C_μ gene specifies both the μ_m and the μ_s proteins through the selection of alternative polyadenylation and splice sites. In forming μ_m mRNA (left), the primary transcript is cleaved and polyadenylated after the two exons specifying its transmembrane segment (7 + 8) and the segment at the end of

the C_{H3} exon (6) that specifies the μ_s tail is then spliced away. μ_s mRNA (right), on the other hand, is generated by cleavage and polyadenylation just past its μ_s tail segment, which is then retained by the splicing machinery.

different immunoglobulin isotypes have both membrane-bound and secreted forms that are similarly generated.

i. Mature B Cells Can Switch the Immunoglobulin Isotype They Synthesize

Virgin B cells synthesize membrane-bound IgM and IgD. Yet, the progeny of B cells that have been antigenically stimulated to proliferate may eventually synthesize immunoglobulins of different isotypes that have the same variable regions as the original IgM (recall that these different immunoglobulin isotypes have distinct physiological roles; Section 35-2Ba). The nucleic acid sequences specify-

ing the variable region of the heavy chain must therefore become linked to the sequences specifying the constant regions of its various isotypes. How does this **isotype switching** (alternatively, **class switching**) occur?

The downstream regions of the human heavy chain gene family consist, as we have seen, of nine segments encoding the constant regions for the various immunoglobulin isotypes and subtypes (Fig. 35-38). Isotype switching might occur either through RNA processing or through DNA processing. In fact, both mechanisms are employed. In the RNA processing mechanism, it is uncertain whether the switching event is a change in transcriptional termina-

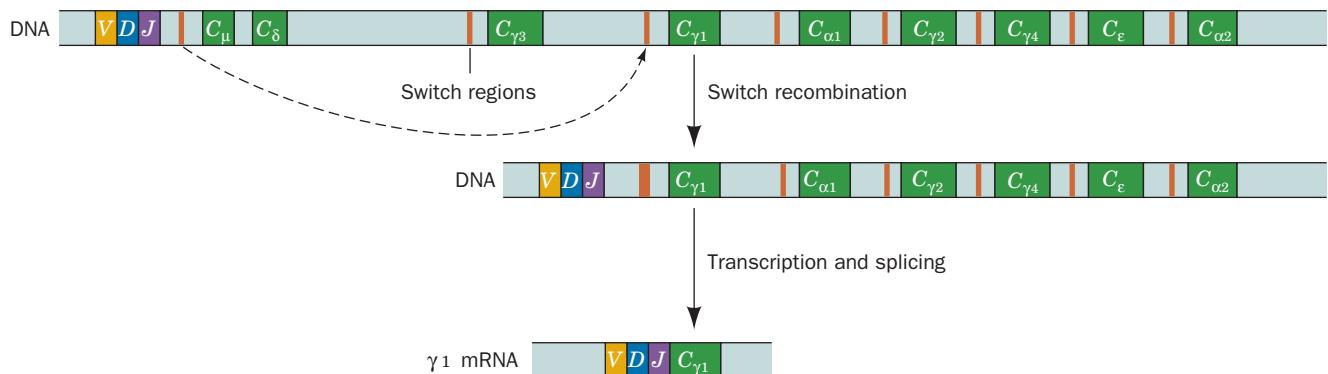


Figure 35-38 DNA-mediated isotype switching. An irreversible change in heavy chain synthesis from μ to a downstream constant region (shown here as γ_1) occurs through somatic recombination between the switch regions located

upstream of each of the constant regions but C_δ . Each constant region consists of multiple exons and encodes alternative secreted and membrane-binding C-termini (see Fig. 35-37).

tion, polyadenylation, and/or splicing, but, in any case, the result is the synthesis of heavy chain mRNAs with identical variable regions but different constant regions. B cells can therefore simultaneously synthesize two or more immunoglobulin isotypes with identical antigen-binding sites.

The DNA processing mechanism of isotype switching, which is stimulated by external signals such as antigen binding and cytokines secreted by helper T cells, occurs through somatic recombination between the *VDJ* unit assembled in previous cell generations and the selected C_H segment. In doing so, the intervening segment of DNA is deleted so that this process is progressive and irreversible. For example, in recombinational switching from making IgM to making IgG1 (Fig. 35-38), a B cell loses its C_{μ} , C_{δ} , and $C_{\lambda 3}$ segments so that its progeny can no longer synthesize IgM, IgD, or IgG3. Yet, the progeny still have the potential to switch to IgA1, IgG2, IgG4, IgE, and IgA2 synthesis via RNA processing since the recombination does not eliminate the $C_{\alpha 1}$, $C_{\gamma 2}$, $C_{\gamma 4}$, C_{ϵ} , and $C_{\alpha 2}$ segments.

Each C_H segment, with the exception of C_{δ} (which is only expressed through RNA processing), is preceded by a **switch or S region** that consists of multiply repeated, short

(20–80 bp), G-rich, complementary elements. Recombination occurs between the switch region preceding the C_{μ} segment and that preceding the selected C_H segment ($C_{\gamma 1}$ in Fig. 35-38). Isotype switching, which occurs concurrently with somatic hypermutation, is likewise initiated by the action of activation-induced deaminase (AID; Section 35-2Cf), which deaminates several cytosine residues in the switch regions preceding C_{μ} and the selected C_H segment. The single-strand nicks generated during the subsequent base excision repair process are converted to double-strand breaks by a poorly understood process. The breaks in the two switch regions are then joined by nonhomologous end-joining (NHEJ; Section 30-5E), thereby eliminating the intervening DNA. Isotype switching events are always productive because the recombined switch regions lie in introns and therefore cannot cause frameshift mutations.

D. T Cell Receptors

T cell receptors (TCRs), as we have seen, are in many ways the cellular immunity system's analog of immunoglobulins. Like immunoglobulins, TCRs exhibit enormous specificity

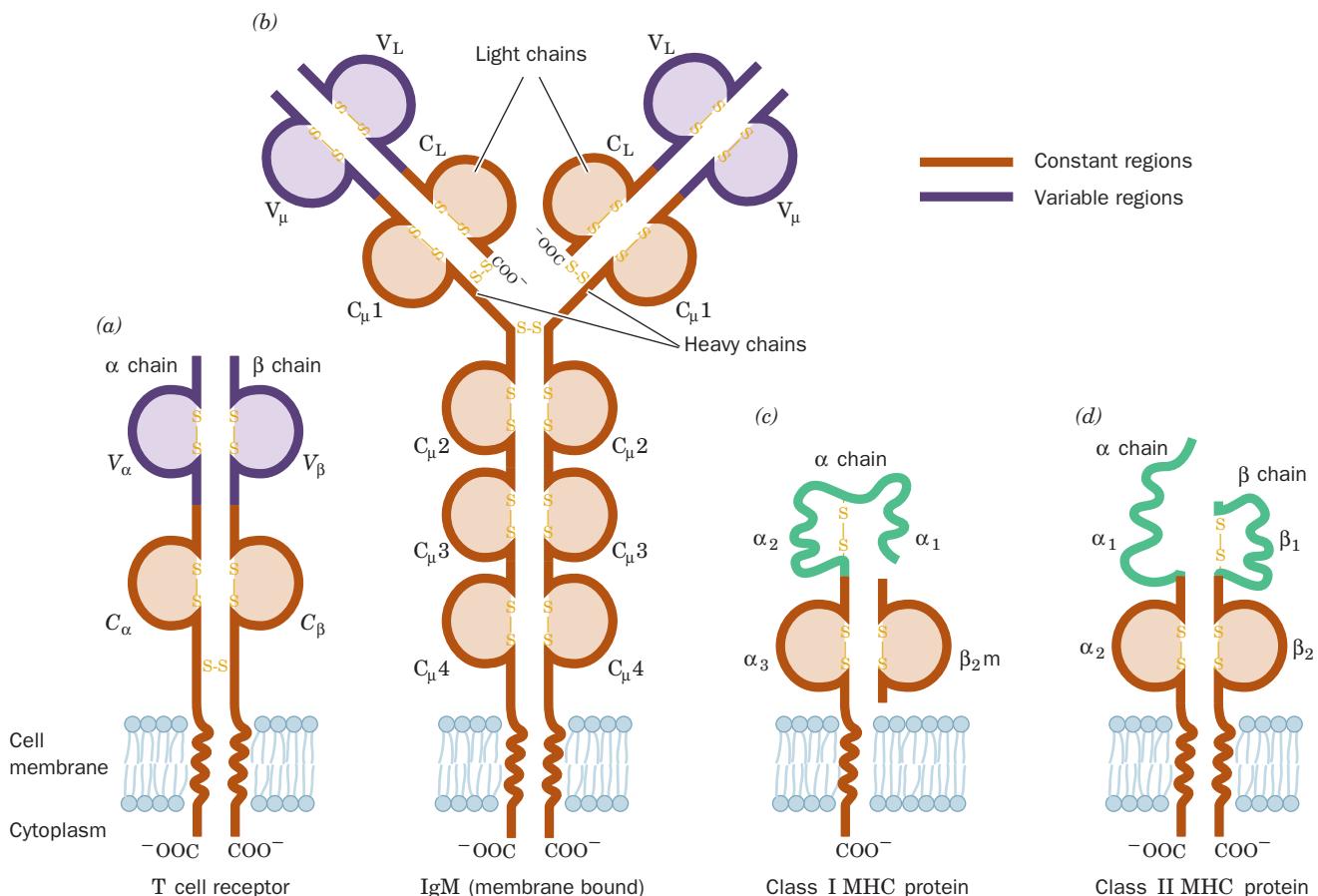


Figure 35-39 Schematic diagrams of members of the immunoglobulin gene superfamily. These proteins, such as (a) the T cell receptor, (b) membrane-bound IgM, (c) Class I MHC protein, and (d) Class II MHC protein, all have similar domain structures. Each protein contains multiple immunoglobulin domains (brown and purple regions). Note that the IgM heavy chain has four constant regions compared with three in IgG (Fig. 35-20). There are numerous other immune system proteins that

contain one or more immunoglobulin domains. However, the members of the immunoglobulin gene superfamily are specialized for molecular recognition rather than limited to the immune system. For example, the immunoglobulin domain-containing protein named **neuronal cell-adhesion molecule (N-CAM)** participates in cell–cell recognition but is not an immune system component.

in binding antigens. Yet, the characterization of TCRs lagged far behind that of immunoglobulins because TCRs occur only as cell surface proteins and are therefore present in small quantities. They were finally isolated in 1983 through the use of monoclonal antibodies directed against them and characterized through reverse genetics (Section 34-3).

A TCR consists of two glycosylated subunits, α and β , which in humans contain ~ 250 and ~ 290 residues (there is also a minor class of TCRs that consist of γ and δ subunits). Each chain has a constant domain and an N-terminal variable domain of approximately equal size as well as a C-terminal transmembrane segment (Fig. 35-39a). Not surprisingly, considering their similar functions, the TCR subunits are sufficiently similar to the immunoglobulin subunits (Fig. 35-39b) that their constant and variable domains are each predicted to assume the immunoglobulin fold and hence all of these subunits must have arisen from a common ancestor. Moreover, the α and β subunit gene families are each organized into clusters of V and J regions, with the α family having an additional D region. The somatic recombination of these regions, which is mediated by the $V(D)J$ recombinase acting on 12- and 23-RSSs, similar to the analogous processes in immunoglobulin genes (Section 35-2Ce), is the major source of TCR diversity. Nevertheless, only the CDR3's in both subunits are variable; the CDR1s and CDR2's are constant, that is, they are germline derived. In addition, TCRs are not subject to somatic hypermutation so that their affinities for antigens remain relatively low. Consequently, as we discuss in Section 35-3E, an effective interaction between a TCR and its target antigen requires the assistance of both an MHC protein and certain coreceptors.

The X-ray structure of a TCR (Fig. 35-40), which was first determined by Ian Wilson, reveals, as expected, that the TCR structurally resembles an Fab fragment, although the similarity is closer for the variable (V_α and V_β) segments than for the constant (C_α and C_β) segments. The TCR's combining sites (CDRs) are relatively flat except for a deep hydrophobic crevice between the CDR3's of its α and β subunits.

E. The Major Histocompatibility Complex

The membrane-bound proteins encoded by the major histocompatibility complex (MHC; *histo-* refers to tissue), as we have seen (Section 35-2Aa), are the antigen-presenting markers through which the immune system distinguishes cells harboring invading antigens from normal body cells (Class I MHC proteins) and immune system cells from other cells (Class II MHC proteins). *This guides cytotoxic T cells to attack only infected host cells, while permitting helper T cells to interact only with immune system cells.* Note that antigen-presenting cells (APCs) display both Class I and Class II MHC proteins, thereby enabling them

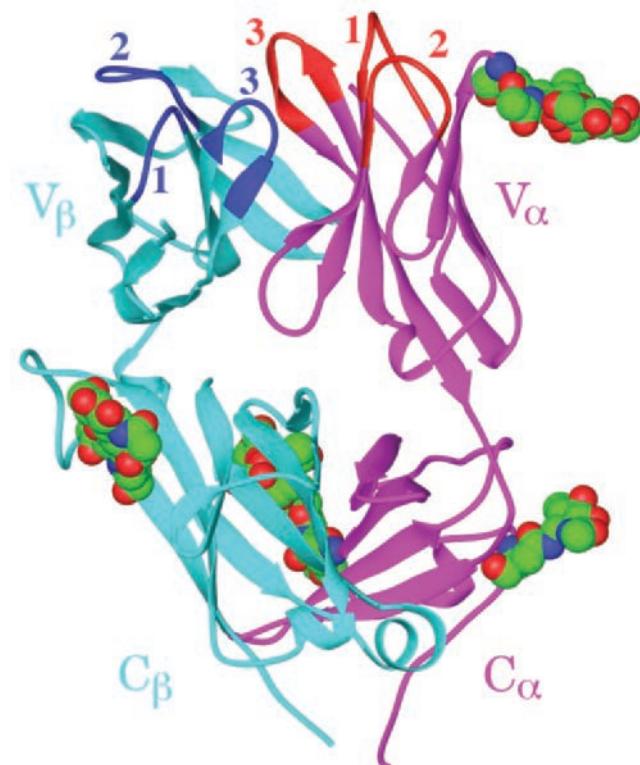


Figure 35-40 X-ray structure of the extracellular portion of a mouse $\alpha\beta$ T cell receptor. The α subunit is magenta with its CDRs red, and the β subunit is cyan with its CDRs blue. The CDRs in both chains are numerically labeled. The carbohydrate moieties, together with the Asn residues to which they are N-linked, are drawn in space-filling form with C green, N blue, and O red. The complex is oriented with its N-terminal variable (V) domains above and its C-terminal constant (C) domains below. In the intact protein, ~ 45 -residue transmembrane segments would extend from bottom (C-terminus) of each subunit. [Based on an X-ray structure by Ian Wilson, The Scripps Research Institute, La Jolla, California. PDBid 1TCR.]

to stimulate the development of both cytotoxic and helper T cells (Fig. 35-18). In the following paragraphs, we outline the structures and properties of the MHC proteins.

a. MHC Proteins Are Highly Polymorphic

The MHC has been extensively studied in both humans and mice. In humans, the Class I MHC proteins are encoded by three separate although homologous genetic loci, **HLA-A**, **HLA-B**, and **HLA-C** (Fig. 35-41; *HLA* stands for *human-leukocyte-associated antigen* because these proteins were first observed on leukocytes), so each individual synthesizes up to six different Class I MHC proteins (see below). There are also three human Class II MHC proteins whose α and β

Chromosome 6

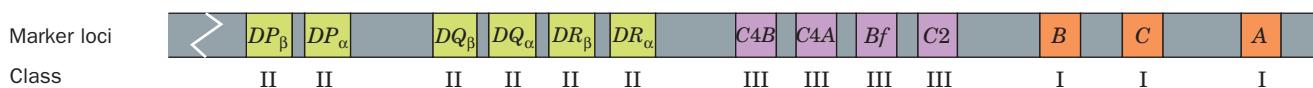


Figure 35-41 Genetic map of the human MHC, which encodes the HLA proteins.

The Class III genes encode several complement system proteins (Section 35-2F).

chains are encoded by genes designated DP_α , DP_β , DQ_α , DQ_β , DR_α , and DR_β (Fig. 35-41). Mouse MHC genes, which occupy the so-called **H-2** loci, are similarly arranged.

The most striking feature of the Class I and Class II MHC genes is their high level of polymorphism among individuals of the same species; *they are, by far, the most polymorphic genes known in higher vertebrates*. For example, in humans, the *HLA-A*, *HLA-B*, and *HLA-C* genes, respectively, have around 1000, 1600, and 870 known alleles. Class II MHC proteins are similarly polymorphic. Since both paternally and maternally supplied MHC genes are expressed in a cell (although not all Class II MHC α and β chains can form stable dimers), *it is highly unlikely that two individuals, including siblings (but not identical twins), have the same set of MHC genes*. Many of the HLA alleles are

quite ancient; in many cases a human allele is more closely related to one from a chimpanzee than to a different human allele from the same gene.

b. Class I MHC Proteins

Tissues can be readily transplanted from one part of an individual's body to another or between genetically identical individuals (e.g., identical twins). Yet, when tissues are transplanted between even closely related individuals, the graft is generally destroyed by the recipient's immune system (a phenomenon that is a major impediment to the transplantation of organs such as hearts and kidneys). Studies of such **graft rejection** in the 1950s led to the discovery of the Class I MHC proteins, which were therefore also known as **transplantation antigens**.

The Class I MHC proteins are ~44-kD transmembrane glycoproteins that are displayed on the surfaces of nearly all nucleated vertebrate cells. [In contrast, nonnucleated cells, such as red blood cells, lack MHC proteins. This is of little import for viral infection because viral replication requires the use of a cell's genetic machinery. However, it is

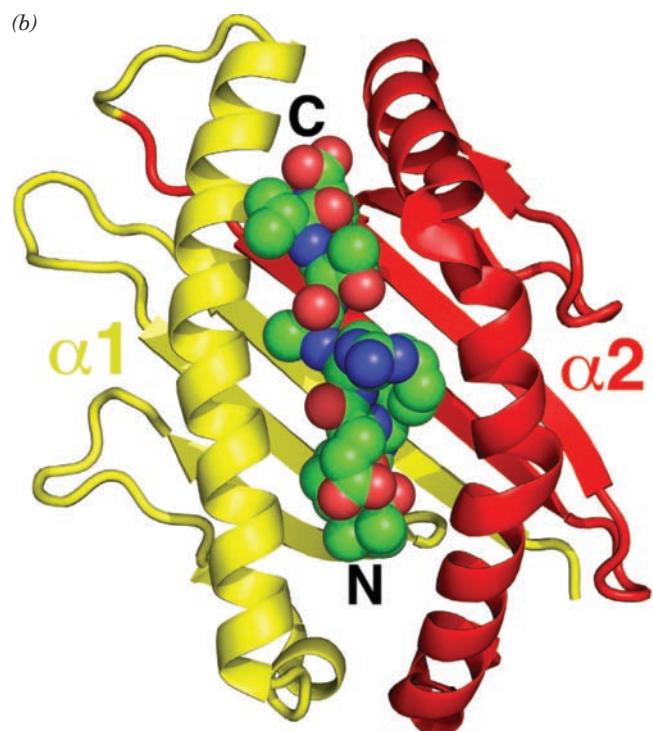
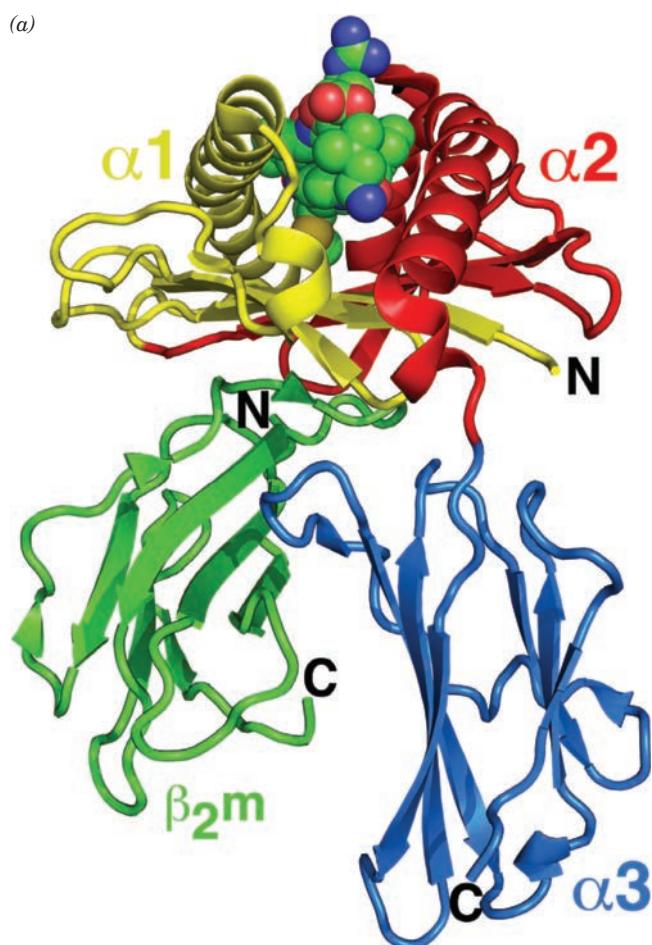


Figure 35-42 X-ray structure of the human Class I MHC protein HLA-B*3501 in complex with an epitope from the 2009 H1N1 influenza virus. (a) A ribbon diagram of the protein, which lacks its C-terminal membrane anchor, oriented such that the plasma membrane from which it would otherwise project would be horizontal at the bottom of the drawing. Its polymorphic α_1 and α_2 domains are yellow and red, and its immunoglobulin-like α_3 domain and β_2m subunit are blue and green. The nonapeptide (sequence LPFERATVM) from the influenza virus nucleocapsid protein (Section 33-4A), which occupies the protein's antigen recognition site, is drawn in space-filling form with C green, N blue, and O red. (b) The α_1 and α_2 domains as

viewed from above in Part a. These domains, which each consist of four antiparallel β strands followed by a long helix, pair with pseudo-2-fold symmetry to form a relatively flat 8-stranded antiparallel β sheet flanked by the two helices. The resulting groove forms the binding site for the peptide that the Class I MHC protein presents to the TCR. Most of the protein's polymorphic residues line the surface of this groove as do many residues critical for TCR recognition. [Based on an X-ray structure by Jamie Rossjohn, Monash University, Clayton, Victoria, Australia, and Katherine Kedzierska, University of Melbourne, Parkville, Victoria, Australia. PDBid 3LKR.]

why malarial parasites (*Plasmodia*), which occupy red blood cells during much of their life cycle (Section 7-3Ab), are shielded from the immune system during this period.] Class I MHC proteins' amino acid sequences suggest that they are folded into five domains that are, from N-terminus to C-terminus, three external domains of ~90 residues each, designated α_1 , α_2 , and α_3 , a transmembrane segment of ~40 residues, and an ~30-residue cytoplasmic domain (Fig. 35-39c). The Class I MHC proteins are invariably non-covalently associated in a 1:1 ratio with **β_2 -microglobulin** (**β_2 m**; Fig. 35-39c), an ~120-residue protein whose gene resides on a different chromosome from that containing the MHC. X-ray structures of the extracellular portion of Class I MHC proteins indicate that its α_3 domain as well as β_2 -microglobulin, both of which are homologous to immunoglobulins, assume the immunoglobulin fold (Fig. 35-42a). *Evidently, all of these proteins, together with TCRs, are evolutionarily related and therefore form a gene superfamily (a set of evolutionarily related genes with divergent functions).*

The homologous α_1 and α_2 domains of class I MHC proteins form a relatively flat eight-stranded antiparallel β sheet that is parallel to the cell membrane and which is flanked by two α helices (Fig. 35-42b). *The resulting deep groove or cleft, which is of sufficient size and convoluted shape to bind a largely extended 8- to 10-residue polypeptide, forms the binding site of a cell-processed epitope that, together with the Class I MHC protein itself, is recognized by a TCR* (Section 35-2Aa). Indeed, the amino acid residues that differ among the numerous human Class I MHC proteins

whose X-ray structures have been determined are concentrated in and around this antigen-binding cleft.

The X-ray structures of both human and murine Class I MHC proteins complexed either to endogenous peptides or to specific exogenous octa- and nonapeptides have revealed how these proteins bind their cognate peptides and present them to TCRs. Peptides associate with Class I MHC proteins, mostly via hydrogen bonds involving the peptide's backbone, such that the peptides assume nearly extended but twisted conformations resembling that of the polyproline II helix (Section 8-1Bb). As a consequence, consecutive peptide side chains protrude in largely opposite directions, somewhat like the side chains of a strand of β sheet (Fig. 8-17). In the complex of murine H-2K^b protein with a nonapeptide, for example (Fig. 35-43), the side chains of residues P2, P3, P6, and P9 (where P_n represents the n th residue of the peptide) face inward to contact the protein in pockets that appear designed to bind them. The remaining side chains are at least partially in contact with solvent and presumably could interact with a TCR. In addition, the peptide's N- and C-termini bind to deep and conserved pockets at each end of the MHC protein's binding cleft through hydrogen bonding contacts to conserved residues that thereby dictate the peptide's orientation. Thus, an octapeptide bound to H-2K^b (Fig. 35-43) maintains essentially the same contacts as does the nonapeptide because the nonapeptide's P5 residue is accommodated through the formation of a central bulge (i.e., residues P6 to P9 in the nonamer correspond to residues P5 to P8 in the octamer).

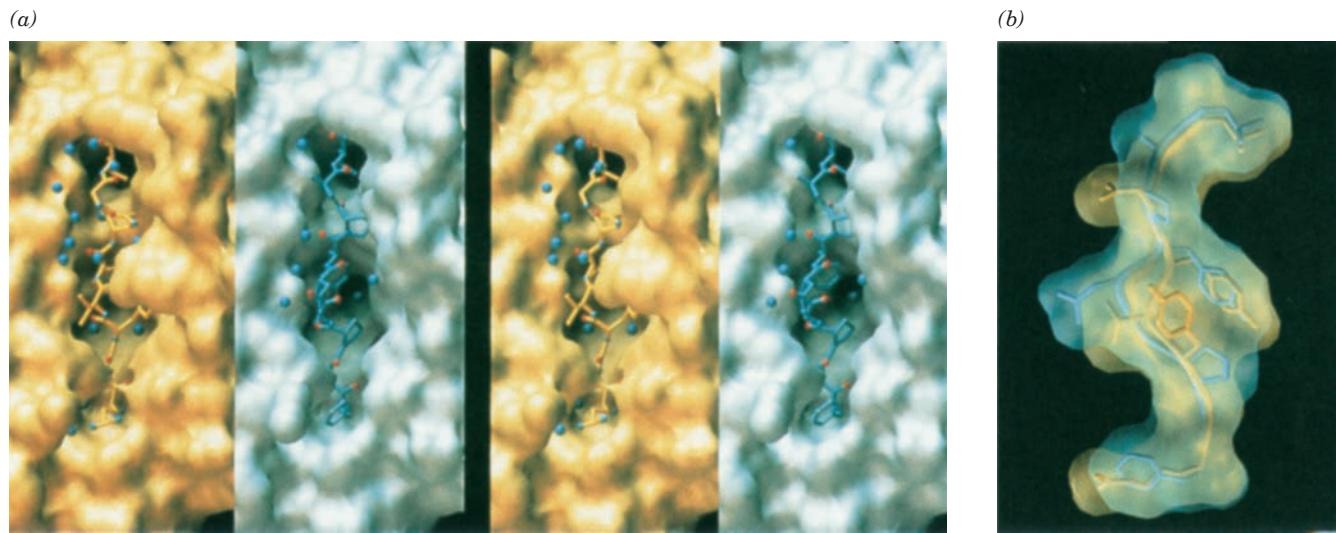


Figure 35-43 Comparison of the X-ray structures of two peptide complexes of murine H-2K^b Class I MHC protein. The protein forms complexes with a nonapeptide fragment of **Sendai virus** nucleoprotein (SEV-9; sequence FAPGNYPAL) and with an octapeptide fragment of vesicular stomatitis virus nucleoprotein (VSV-8; sequence RGYVYQCL). (a) A double stereo drawing of H-2K^b displayed as its solvent-accessible surface and viewed into its antigen-binding cleft with the SEV-9 complex (blue) on the right and the VSV-8 complex (yellow) on the left. The peptides are shown in ball-and-stick form with N

blue and O red, and with bound water molecules represented by blue spheres. Note the depth of the cleft and the relative inaccessibility of the peptides. Directions for viewing stereo drawings are given in the appendix to Chapter 8. (b) The SEV-9 (blue) and the VSV-8 (yellow) polypeptides shown superimposed in their H-2K^b binding clefts. The polypeptide backbones are each represented by a rod with their side chains shown in skeletal form. The view is from the right in Part a such that the top of the cleft is on the left. [Courtesy of Ian Wilson, The Scripps Research Institute, La Jolla, California. PDBIDs 2VAB and 2VAA.]

The sizes and amino acid compositions of the pockets containing the occluded side chains suggest that any particular Class I MHC protein can only bind a limited selection of peptides. For example, in the complex of HLA-B27 with endogenous peptides, the P2 side chain binds in a hydrophobic pocket that ends near Cys 67 and the negatively charged Glu 45, which suggests that this site preferentially binds a long, positively charged side chain. In fact, all 11 peptides that were eluted from HLA-B27 contain Arg at P2. The identities of its other occluded side chains, those at P3, P7, and P9, although not as restricted as that of P2, are more or less consistent with the properties of the pockets in which they bind, whereas the solvent-exposed side chains have the expected wide range of identities.

The peptides eluted from other class I MHC proteins also exhibit distinct allele-specific sequence motifs. In particular, each sequence motif contains two so-called anchor positions that are occupied by only one or, at most, a few residues with closely related side chains. The anchor positions vary with the MHC protein. In contrast to the anchor positions, side chains that are solvent-exposed or bound in pockets that are tolerant of sequence variation show considerable diversity. This system presumably permits TCRs to distinguish between self and foreign antigens. Immunoproteasomes (Section 35-2Aa) appear to be specialized to produce peptide fragments that bind to class I MHC proteins.

c. Class II MHC Proteins

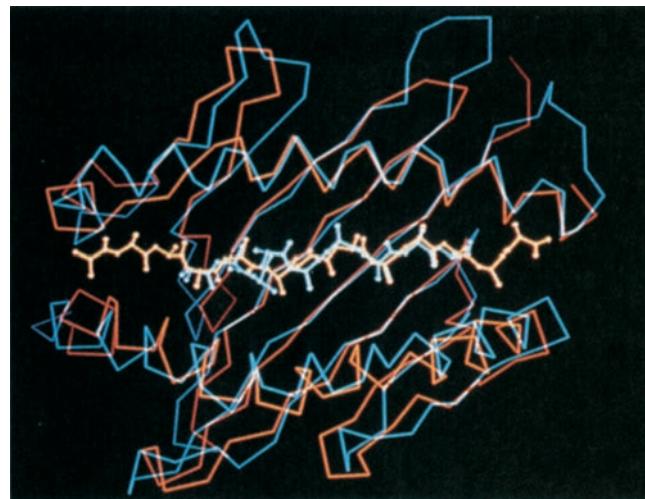
The discovery of Class II MHC proteins came about through Baruj Benacerraf's observation that certain T cell-dependent immune responses are mediated by gene products that are not antibodies. For example, when guinea pigs are inoculated with a simple antigen such as polylysine, some individuals mount a vigorous immunological response to the antigen, whereas others fail to respond. Immunological responsiveness to a given simple antigen is a dominant genetic trait. *A small number of so-called immune response (Ir) genes apparently govern how an individual responds to all simple antigens.* [A naturally occurring antigen, such as a protein, is complex, that is, it has numerous different epitopes (antigenic determinants; Section 35-2Bf). Consequently, the several different MHC proteins in every individual make it highly likely that an individual will be able to mount an immune response against a naturally occurring antigen.]

The *Ir* genes map into the MHC so that they are now known as Class II MHC genes. They encode the two subunits of a heterodimeric transmembrane glycoprotein composed of a 33-kD α chain and a 28-kD β chain, each of which consists of two domains (Fig. 35-39d). The amino acid sequences of these subunits indicate that the C-terminal α_2 and β_2 domains are members of the immunoglobulin gene superfamily. Moreover, their α_1 and β_1 domains can be convincingly aligned on the known structures of the Class I MHC proteins' α_1 and α_2 domains, respectively. This indicates that Class I and Class II MHC proteins are structurally as well as functionally similar.

This prediction was largely confirmed by the X-ray structures of the extracellular portion of a Class II MHC

protein, HLA-DR1, in its complexes with both a mixture of endogenous peptides and with a 13-residue fragment of influenza hemagglutinin protein (HA; Section 33-4Bb), both determined by Jack Strominger and Don Wiley. However, the peptide-binding site of HLA-DR1 is an open-ended groove (Fig. 35-44), whereas those of Class I MHC proteins

(a)



(b)

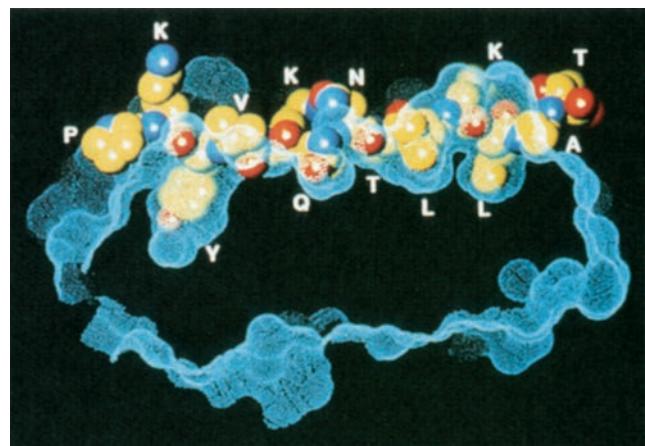


Figure 35-44 X-ray structure of the human Class II MHC protein HLA-DR1. (a) Superposition of the peptide-binding sites, as seen in "top" view, of HLA-DR1 in complex with a 13-residue polypeptide fragment from HA (orange; sequence PKYVKQNTLKLAT) and the Class I MHC protein HLA-B27 in complex with an endogenous mixture of nonapeptides (cyan). The proteins are represented by their C_α backbones and their bound peptides are shown in ball-and-stick form with their N-termini to the left and their side chains removed for clarity. Note the close structural similarity of these two classes of MHC proteins and that the HLA-DR1-bound peptide extends beyond the edges of its binding site, whereas peptides bound to Class I MHC proteins are entirely embedded in protein (Fig. 35-43). (b) A "vertical" section through HLA-DR1's molecular surface (cyan) with its bound 13-residue HA peptide shown in space-filling form (C yellow, N blue, and O red). The peptide's side chains are identified by their one-letter codes. [Courtesy of Don Wiley, Harvard University. PDBIDs 1DLH and HSA.]

are elongated but closed-ended clefts (Fig. 35-43). This explains why Class II MHC proteins bind peptides of arbitrary length, whereas Class I MHC proteins bind mostly extended but bulged-out nonapeptides. Indeed, the X-ray structure of HLA-DR1 in complex with the 13-residue HA peptide reveals that the peptide extends out from both ends of its binding groove (Fig. 35-44b).

The X-ray structure of a tertiary complex of a Class II MHC protein, an $\alpha\beta$ TCR, and their bound polypeptide antigen (Fig. 35-45a) reveals, as expected, that only the CDRs of the TCR contact the MHC-peptide complex.

Moreover, only CDR1 α , CDR3 α , and CDR3 β interact directly with the peptide (Fig. 35-45b); CDR1 β , CDR2 α , and CDR2 β , all of which are constant (as is CDR1 α ; Section 35-2D), contact only the MHC helices. Similar arrangements are observed in other X-ray structures of TCR-peptide-MHC protein complexes, although CDR1 α does not always contact the peptide.

d. T Cell Receptors Require the Assistance of CD4 and CD8 Coreceptors

Since cytotoxic and helper T cells express similar TCRs, what restricts cytotoxic T cells from binding only to cells that display Class I MHC proteins, and helper T cells from binding only to cells that express Class II MHC proteins? The answer is that helper T cells express a cell surface glycoprotein called **CD4** that binds only to Class II MHC proteins (so that helper T cells are also called **CD4⁺ T cells**), whereas cytotoxic T cells express a cell surface glycoprotein called **CD8** that binds only to Class I MHC proteins (so that cytotoxic T cells are also called **CD8⁺ T cells**). This

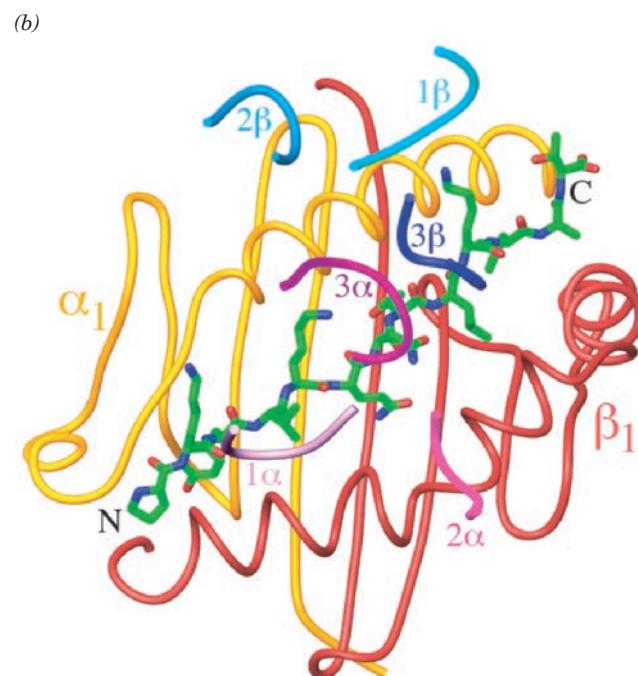
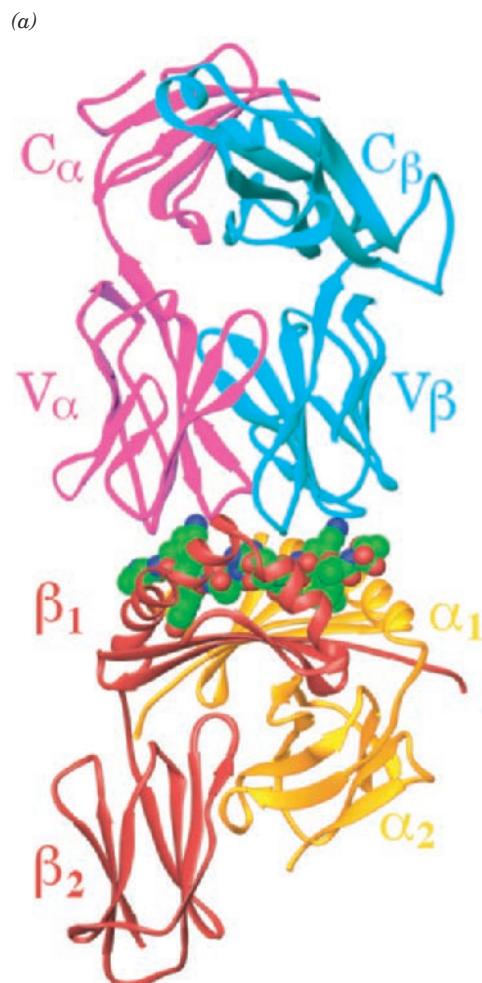


Figure 35-45 X-ray structure of an MHC-antigen-TCR complex. Only the extracellular portions of the human Class II MHC protein, HLA-DR1, and the influenza hemagglutinin (HA)-binding $\alpha\beta$ TCR, HA1.7, are present. The antigen is a 13-residue polypeptide fragment of HA (sequence PKYVKQNTLKLAT). (a) The complex as viewed in parallel to the plasma membranes of the antigen-presenting cell (*below*) and the helper T cell (*above*) from which the MHC protein and TCR would normally project. The α and β chains of the MHC protein (*below*) are gold and red, and those of the TCR (*above*) are magenta and cyan (viewed upside down from the orientation in Fig. 35-40). The peptide is drawn in space-filling form (C green, N blue, O red). (b) The disposition of the TCR's six CDRs in

relation to the MHC protein's antigen-presenting α_1 and β_1 domains and their bound peptide as viewed from the top of Part a. The α_1 and β_1 domains of the MHC protein (gold and red) and the CDRs of the TCR (different shades of purple and blue) are drawn in worm form, and the peptide antigen is drawn in stick form (C green, N blue, O red). The TCR's CDRs are labeled, for example, 1α for CDR1 α . Note that only CDR1 α , CDR3 α , and CDR3 β contact the bound peptide antigen. The remaining CDRs contact only the MHC protein. The TCR is oriented more or less diagonally across the MHC protein's peptide binding groove, although this orientation varies somewhat in the X-ray structures of other MHC-antigen-TCR complexes. [Based on an X-ray structure by Don Wiley, Harvard University. PDBid 1J8H.]

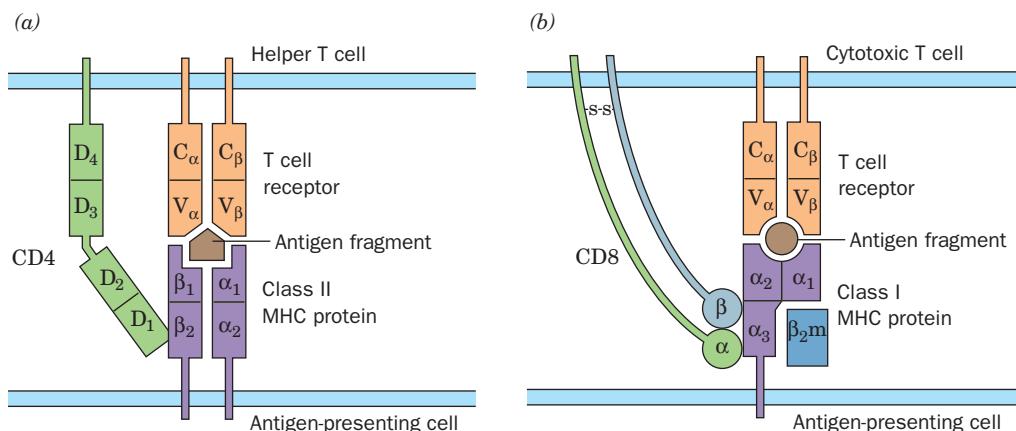


Figure 35-46 CD4 and CD8 coreceptors and their interactions with MHC proteins. (a) CD4, which is expressed only by helper T cells, is a transmembrane glycoprotein whose N-terminal extracellular portion consists of four immunoglobulin domains, D₁ through D₄. X-ray studies indicate that D₁ and D₂ form an ~60-Å-long rigid rod, which is connected via a flexible hinge to D₃, which forms a similar rigid rod with D₄. CD4, which is anchored in the helper T cell's plasma membrane, associates with a Class II MHC protein, which is anchored in an antigen-presenting cell's plasma membrane. This buttresses the interaction between the MHC–antigen complex and its bound TCR, which is also anchored in the helper T cell's plasma membrane. Since CD4 does not bind to Class I MHC proteins, this prevents helper T cells from interacting with antigen-

presenting cells bearing only Class I MHC proteins. (b) CD8, which is expressed only by cytotoxic T cells, is an αβ heterodimer of disulfide-linked transmembrane glycoprotein subunits, each of which consists of an N-terminal extracellular immunoglobulin domain, an extended region, and a transmembrane segment, which is anchored in the cytotoxic T cell's plasma membrane. CD8 binds to a Class I MHC protein that is anchored in an antigen-presenting cell's plasma membrane, thereby reinforcing the MHC–antigen–TCR interaction. However, CD8 does not associate with Class II MHC proteins, which prevents the association of cytotoxic T cells with antigen-presenting cells bearing only Class II MHC proteins. Note that neither CD4 nor CD8 associate with a T cell receptor.

binding is required for effective T cell responses so that CD4 and CD8 function as **coreceptors**.

CD4 is an ~55-kD glycoprotein monomer that consists of four immunoglobulin domains, D₁ through D₄, followed by a transmembrane anchor (Fig. 35-46a). It binds to a Class II MHC protein projecting from an apposing antigen-presenting cell without contacting the TCR that is also bound to the MHC–antigen complex.

The X-ray structure of the CD4 domains D₁ and D₂ in complex with the extracellular portion of an MHC Class II protein and a 16-residue polypeptide antigen (Fig. 35-47a) reveals that the two proteins associate via an interaction between the tip of the CD4 D₁ domain and nonpolymorphic regions of the MHC protein's β₂ domain. This contact is well clear of the MHC protein's antigen-binding cleft, thereby permitting a quaternary CD4–MHC–antigen–TCR complex to form. CD4 is also the HIV (AIDS virus) receptor, which is why this virus specifically infects helper T cells (Section 15-4C).

CD8 is an ~34-kD disulfide-linked αβ heterodimer, whose subunits both consist of a single immunoglobulin domain followed by an extended and extensively glycosylated polypeptide segment (the glycosylation probably maintains this segment in an extended conformation and protects it from proteases) and then a transmembrane anchor (Fig. 35-46b). Much like CD4, CD8 binds to a Class I MHC protein projecting from an apposing antigen-presenting cell without contacting the TCR that is also bound to the MHC–antigen complex.

The X-ray structure of the extracellular portion of a Class I MHC protein in complex with the CD8αβ immunoglobulin domains and a 10-residue peptide reveals that both the α and β subunits of CD8 contacts only the α₃ domain of the MHC protein (Fig. 35-47b). These interactions involve only nonpolymorphic regions of the MHC protein. Note that CD4 and CD8 interact with analogous regions of their corresponding MHC proteins, although these interactions differ in character.

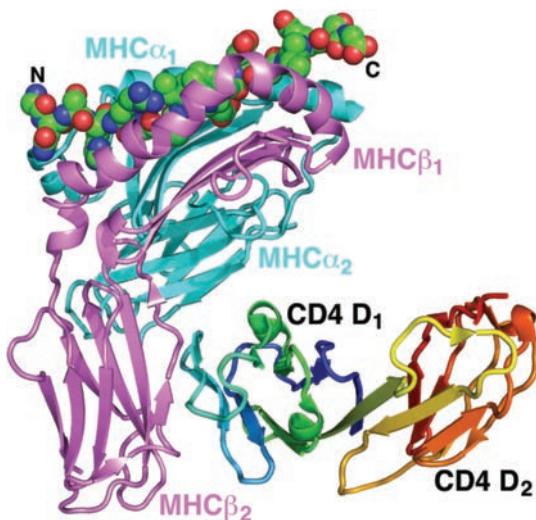
e. T Cell Receptors Act in Concert with CD3, Lck, and ZAP-70 to Induce T Cell Proliferation

How does the binding of a TCR and its coreceptor to an MHC–antigen complex induce a T cell to proliferate? This, it has been found, is mediated by another T cell protein, **CD3**, which associates with TCRs. CD3 is complex of three dimers, γε, δε, and ζζ, in which γ, δ, and ε are transmembrane subunits that each have an N-terminal extracellular immunoglobulin domain and ζ is a transmembrane subunit with a cytoplasmic domain. The γ, δ, and ε subunits each have a cytoplasmic immunoreceptor tyrosine-based activation motif (ITAM; Section 35-2Cg) and the ζ subunit has three such motifs.

T cell proliferation is induced as follows (Fig. 35-48):

1. The cytoplasmic domains of the CD4 and CD8 coreceptors both strongly bind the Src-like nonreceptor tyrosine kinase known as **Lck** (Section 19-3Ea). The formation of an MHC–antigen–TCR complex brings the CD4 or CD8 coreceptor into close proximity to the ITAMs on the

(a)



(b)

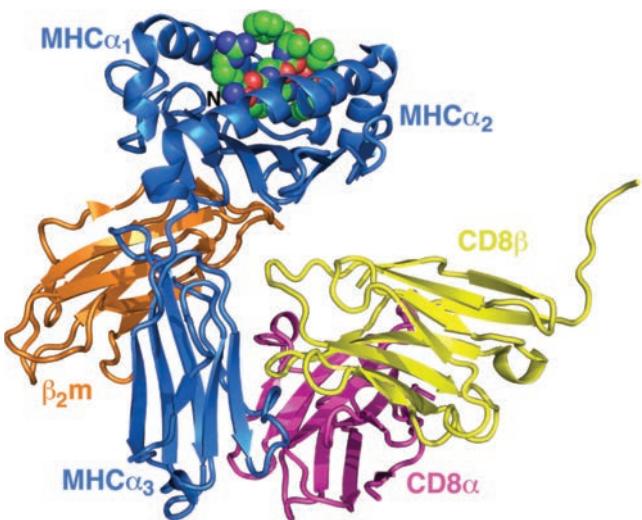


Figure 35-47 X-ray structures of MHC proteins in complex with their coreceptors. (a) X-ray structure of the tertiary complex of the human CD4 N-terminal two domains (D₁ and D₂), the extracellular portion of the murine Class II MHC protein I-A^k, and a 16-residue polypeptide fragment of hen egg **conalbumin** (sequence GNSHRGRAIEWEGIESG). The MHC α subunit is cyan, the MHC β subunit is pink, the CD4 is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), and the conalbumin peptide is drawn in space-filling form with C green, N blue, and O red. The view is approximately parallel to the plasma membrane of the antigen-presenting cell, which would be below, while the helper T cell from which CD4 extends would be above (Fig. 35-46a). Note that the only contact between

the two proteins is between the MHC β₂ domain and the end of the CD4 D₁ domain. [Based on an X-ray structure by Ellis Reinherz, Harvard Medical School. PDBid 1JL4.] (b) The X-ray structure of a complex of the immunoglobulin domains of the murine CD8 αβ heterodimer, the extracellular portion of the murine Class I MHC protein H-2D^d, and a 10-residue peptide fragment from HIV **envelope glycoprotein 120 (gp120)** (sequence RGPGRAFVTI). The CD8 α subunit is magenta, CD8 β subunit is yellow, the MHC α subunit is blue, β₂-microglobulin (β₂m) is orange, and the HIV peptide is drawn in space-filling form with C green, N blue, and O red. The view is analogous to that in Part a. [Based on an X-ray structure by Kannan Natarajan and David Margulies, NIH, Bethesda, Maryland. PDBid 3DMM.]

TCR's bound CD3 protein, which permits the associated Lck to phosphorylate the now adjacent CD3 ITAMs on both their Tyr residues. The phosphorylated ITAMs are

then bound by another nonreceptor tyrosine kinase, named **ZAP70** (for zeta-associated protein, 70 kD; Fig. 19-44), via the latter's SH2 domains (Section 19-3Cb).

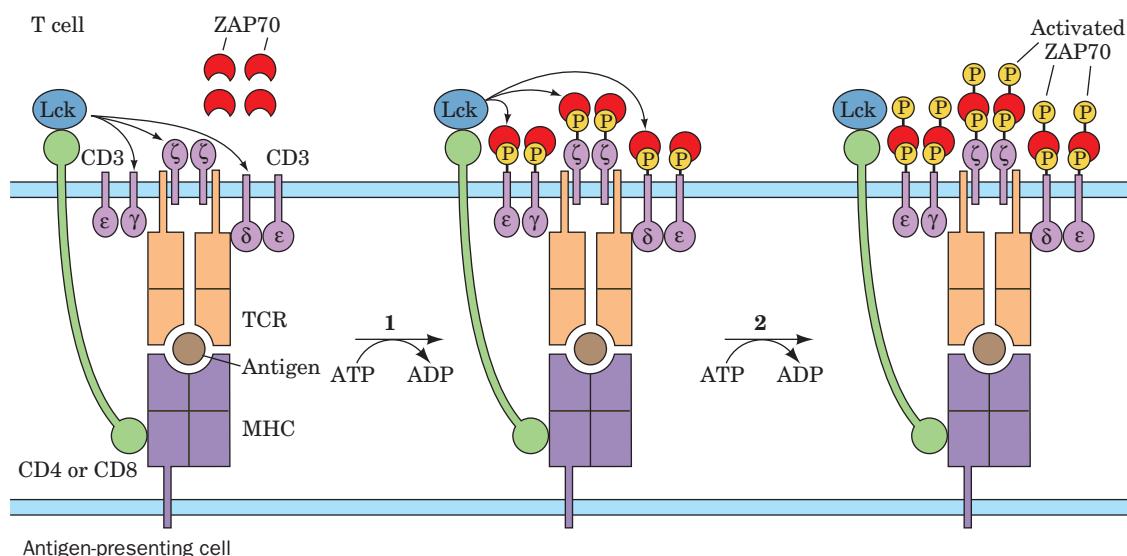


Figure 35-48 The induction of T cell proliferation by CD3, Lck, and ZAP70. CD3 binds to the MHC-antigen-TCR complex as does CD4 or CD8. (1) The nonreceptor tyrosine kinase named Lck, which is strongly bound to both CD4 and

CD8, phosphorylates CD3's several subunits at their ITAMs. This recruits the nonreceptor tyrosine kinase named ZAP-70. (2) Lck phosphorylates ZAP-70, thereby activating it to initiate the cascade of events that induces T cells to proliferate.

2. Lck phosphorylates the now nearby ZAP-70, thereby activating it to initiate the cascade of signaling events that causes T cells to proliferate. The formation of 10 to 100 MHC–antigen–TCR complexes is required for this to occur (a T cell displays ~100,000 TCRs on its surface).

f. MHC Polymorphism Has an Important Protective Function

Most of the polymorphic residues in MHC proteins are clustered, as we saw, in their antigen-binding grooves so that each polymorph binds a given antigenic fragment with characteristic affinity (e.g., it is estimated that any Class I MHC polymorph can bind <1% of the octa- and nonapeptides it encounters). The above-described observations on the variation of immunological responses with Class II MHC (*Ir*) genes therefore suggest that some Class II MHC protein polymorphs are less effective than others in associating with a given epitope. Indeed, *epidemiological studies indicate that certain polymorphs of MHC genes are associated with increased or decreased susceptibilities to particular infectious and/or autoimmune diseases*. For example, 95% of individuals with insulin-dependent diabetes mellitus (Section 27-4Ba) carry at least one *DR2* or *DR3* allele of the *DR* gene in comparison to 50% of normal individuals, whereas **celiac disease** (a violent intestinal upset resulting from eating wheat gluten) is 100% linked to the *DQw2* allele of the *DQ* gene. Conversely, a study of the distribution of MHC alleles in West African children with severe (probably fatal in the absence of treatment) malaria compared to that in infected but largely unaffected children (only a small fraction of those infected with malaria have life-threatening illness) showed that the Class I MHC protein HLA-Bw53 and the Class II MHC protein DRB1*1302-DQB1*0501 are independently associated with protection from severe malaria. These alleles are quite common in West African populations but rare or even nonexistent in other regions. Indeed, in West African populations (in which ~1% of the children under 5 years of age die from malaria), these MHC alleles provide greater protection against malaria than does the sickle-cell trait (Section 7-3Aa).

What is the function of MHC protein polymorphism? Recall that TCRs recognize antigens only when they are presented by MHC proteins (Section 35-2Aa). If every member of a single species had an identical set of MHC proteins, a pathogen whose epitopes interacted poorly with these MHC proteins would obliterate that species. MHC gene polymorphism apparently prevents pathogens from evolving the capacity to do so. Natural selection therefore tends to maintain a large variety of MHC proteins in a population.

If a variety of MHC proteins protects a population against pathogens, why hasn't evolution endowed individuals with more than the small number of different MHC proteins that they express? The answer appears to be that T cells whose TCRs bind strongly to self-antigen–MHC protein complexes must be eliminated to prevent autoimmunity. The consequent loss of T cells would tend to counteract any benefit gained by an additional MHC protein. Thus, the number of MHC proteins expressed by an individual appears to strike a balance between the advantage

of increasing of MHC protein diversity and the disadvantage of decreasing TCR diversity.

g. Why Did So Many Amerinds Die of Introduced Infectious Diseases?

The majority of the 56 million Amerinds (indigenous people of North and South America) estimated to have died as a consequence of European occupation of the New World succumbed to a variety of common infectious diseases that originated in the Old World (e.g., measles and smallpox). Moreover, this excess mortality, which constituted as much as 90% of some populations, has persisted into modern times through continued episodes of the same diseases, even with modern medical treatment. Yet, Amerinds do not have unusual genetic susceptibilities to these diseases or immune systems that are somehow deficient. Rather, it appears that the biochemical basis of this calamity stems from the fact that Amerind populations are unusually homogeneous.

A clue to this enigma came from the observation that children who catch measles from a family member have twice the fatality rate of children who catch this disease from an unrelated individual. The measles virus, as is true of many other RNA viruses, replicates with low fidelity (because reverse transcriptase lacks a proofreading exonuclease function; Section 30-4Ca) and hence rapidly evolves to better adapt to its particular host. Thus, a virus that has adapted its peptides to bind less effectively to the particular MHC proteins of its host will, on average, more readily and severely infect the relatives of this host than unrelated individuals, because the relatives are more likely to express some of the same MHC protein alleles. Conversely, the greater the allelic diversity of the MHC proteins within a population, the less readily a virus can spread through that population.

The MHC genes of Amerinds have been found to have less than half the allelic variation at their *HLA-A* and *HLA-B* loci (and, by inference, other MHC loci) than do most other ethnic populations. This effect is magnified by the fact that the probability of a virus encountering the same MHC allele in two successive hosts is proportional to the square of the frequency with which that allele occurs in the population. Thus, there is a 32% chance that a virus passing between two South Amerinds will encounter the same MHC type at either the *A* or *B* locus but only a 0.5% chance of it doing so when passing between Africans. Evidently, the relatively low MHC polymorphism among Amerinds is responsible for the severity of infectious diseases within their populations.

h. Why Are Tissue and Organ Transplants So Readily Rejected?

Tissues and organs that are transplanted between members of the same vertebrate species, a process known as **allograft** (Greek: *allos*, other), are strongly rejected within 1 to 2 weeks. When such **allograft** rejection was first observed, it was quite puzzling because there seemed to have been no opportunity for nature to have evolved such a mechanism. However, investigations as to the roles of MHC proteins and TCRs led to the realization that allograft

rejection occurs because TCRs and MHC proteins have evolved to interact with one another. Mature T cells have survived a process that selects those that express TCRs that bind to self-MHC proteins presenting only foreign antigens. However, some of these TCRs also bind nonself-MHC proteins (those of a different polymorph) presenting self-antigens (fragments of proteins that are identical in the two individuals), as well as foreign antigens (fragments of proteins that differ between the allograft donor and recipient). Indeed, experiments have shown that from 1 to 10% of an individual's T cells will respond to cells from another, unrelated, individual, a far greater rate of response than the ~0.1% for a conventional antigen. T cell binding to a foreign cell strongly induces the T cell to proliferate because of the high concentration of nonself-MHC proteins on the cell surfaces of the allograft. Moreover, the allograft recipient's antigen-presenting cells take up and present antigens from the allograft to the recipient's T cells. Thus, in the absence of immunosuppressive therapy (Sections 9-2Ba and 19-3Ff), the transplant is rapidly destroyed.

In humans, the Class II MHC proteins encoded by the *DR* locus (Fig. 35-41) play the most important role in allograft rejection. Thus, the closer the match between the HLA-DR proteins of the donor and recipient, the greater the probability of allograft survival. In fact, if the donor and recipient do not share any *HLA-DR* alleles, the allograft is all but certain to be rejected.

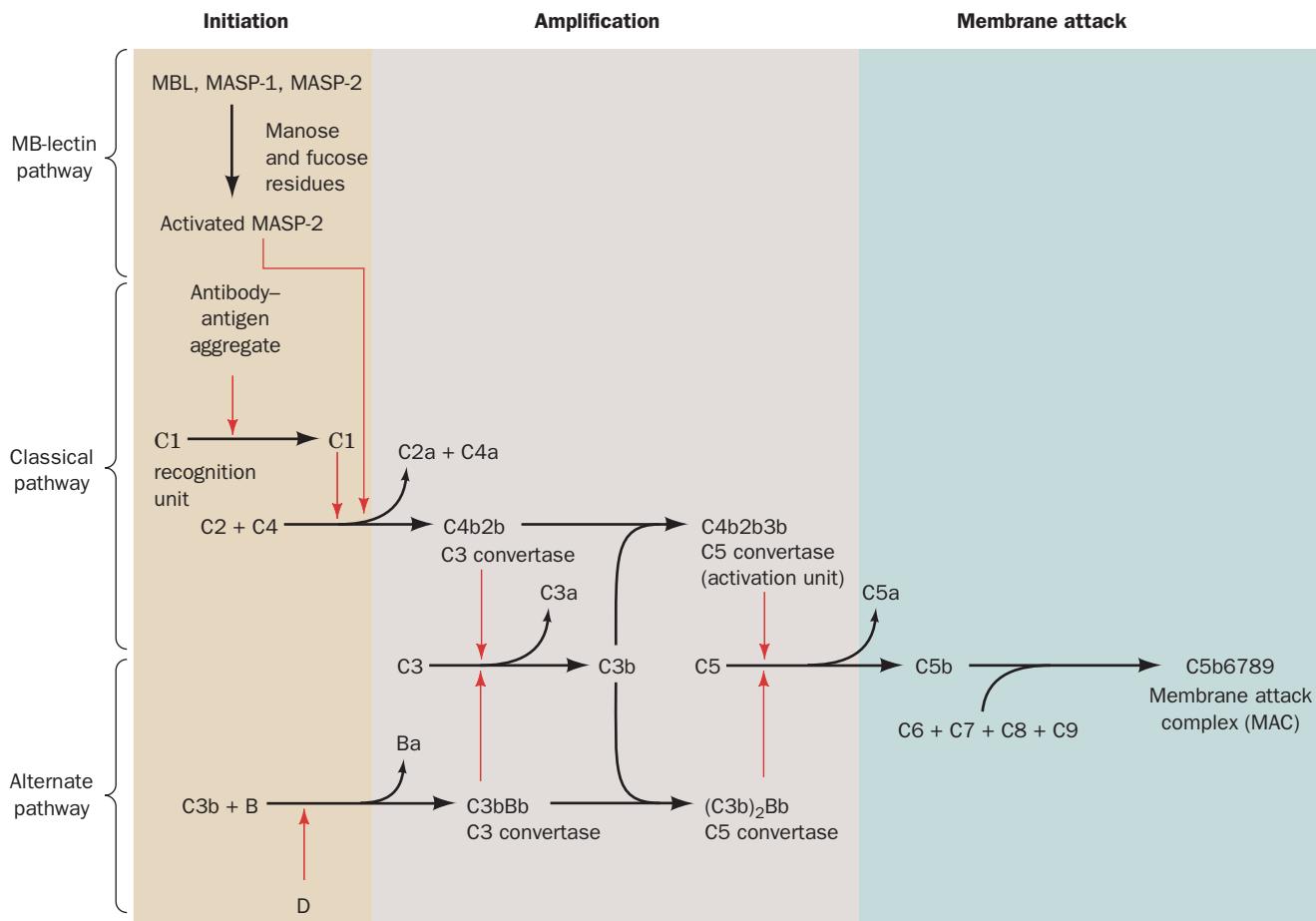


Figure 35-49 Schematic diagram of the complement system activation pathways. Red arrows indicate proteolytic activations.

F. The Complement System

Antibodies, for all their complications, only serve to identify foreign antigens. Other biological systems must then inactivate and dispose of the intruders. The **complement system**, a complex series of interacting plasma proteins of the innate immune system, is one of these essential defensive systems. Indeed, it was named to indicate that it “complements” the function of antibodies in eliminating antigens. It does so in three ways:

1. It kills foreign cells by disrupting their cell membranes so as to lyse them, a process called **complement fixation**.
2. It stimulates the phagocytosis of foreign particles by marking them with certain proteins, a process named **opsonization**.
3. It triggers a local acute inflammatory response that walls off the area and attracts phagocytic cells.

In the remainder of this section, we describe the organization and function of the complement system.

The complement system consists of over 30 plasma and cell-surface proteins (Table 35-3) that interact in three related sets of reactions (Fig. 35-49): the **classical pathway**, the **mannose-binding lectin (MB-lectin) pathway**, and the **alternative pathway**. All three pathways largely consist of the sequential activation of a series of serine proteases, much like the blood clotting pathway (Section 35-1).

Table 35-3 Protein Components of the Human Complement System

Protein	Subunit Structure	Molecular Mass (kD)
Recognition Unit (C1)		
C1q	$\text{A}_6\text{B}_6\text{C}_6$	410
C1r	Monomer	85
C1s	Monomer	85
Activation Unit		
C2	Monomer	95
C3	$\alpha\beta$	190
C4	$\alpha\beta\gamma$	210
Membrane Attack Complex		
C5	$\alpha\beta$	180
C6	Monomer	128
C7	Monomer	121
C8	$\alpha\beta\gamma$	150
C9	Monomer	79
MB-Lectin Pathway		
MBL	$\text{A}_6\text{B}_6\text{C}_6$	600
MASP-1	α_2	94
MASP-2	α_2	91
Alternative Pathway		
Factor B	Monomer	100
Factor D	Monomer	25
Properdin (P)	α_4	220
Regulatory Proteins		
Factor H	Monomer	150
Factor I	$\alpha\beta$	105
C4b-binding protein	$\alpha_6\beta$	560
C1 inhibitor (C1INH)	Monomer	105
S protein (vitronectin)	Monomer	83
Decay-accelerating factor (DAF)	Monomer	70
Membrane cofactor of proteolysis (MCP)	Monomer	60
Receptors		
Complement receptor-1 (CR1)	Monomer	~240
Complement receptor-2 (CR2)	Monomer	140
C3a receptor	Monomer	42
C5a receptor	Monomer	50

Source: Haviland, D.L. and Wetsel, R.A., Complement system, in Creighton, T.E. (ed.), *Encyclopedia of Molecular Biology*, Vol. 1, p. 527, Wiley-Interscience (1999).

The complement system has its own peculiar nomenclature. Many complement protein names consist of the uppercase letter “C” followed by a component number (assigned in the order they were discovered) and, if the protein is a fragment of a larger protein, the lowercase letters “a” or “b,” with “b” being assigned to the bigger fragment. However, some components of the alternative pathway are assigned uppercase letters such as B and D.

a. The Classical Pathway Is Triggered by Antibody-Antigen Complexes

In the classical pathway, the complement proteins form three sequentially activated membrane-bound complexes (Fig. 35-49, *middle*):

1. The **recognition unit**, which assembles on cell surface-bound antibody–antigen complexes.

2. The **activation unit**, which amplifies the recognition event through a proteolytic cascade.

3. The **membrane attack complex (MAC)**, which punctures the antibody-marked cell’s plasma membrane, causing cell lysis and death.

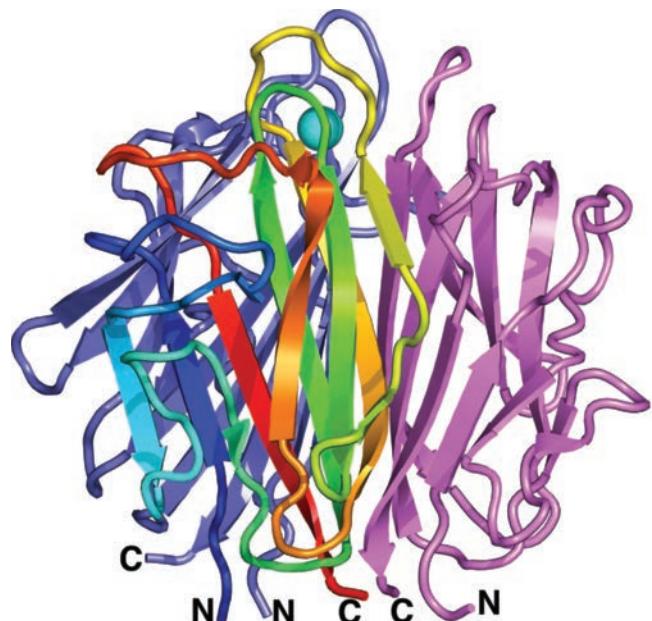
b. The Recognition Unit

The classical pathway is initiated when **C1**, the recognition unit, specifically binds to a cell-surface antigen–antibody aggregate. C1 occurs in the plasma as a loosely bound complex of **C1q**, **C1r**, and **C1s**. C1q is a remarkable

(a)



(c)



(b)

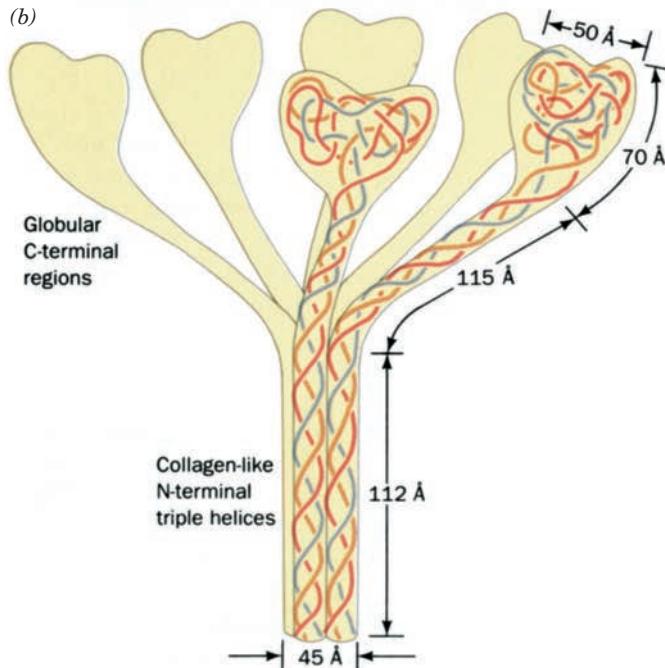


Figure 35-50 Structure of the complement protein C1q.

(a) Electron micrograph of C1q seen in “side” view showing its six C-terminal domains attached to a central stalk. [Courtesy of Tibor Borsos, USPHS-NIH.] (b) Schematic diagram of C1q. Its central stalk consists of a bundle of six collagenlike triple helices. The 80-residue triple helices are ~ 227 Å in length. [After Porter, R.R. and Reid, K.B.M., *Nature* **275**, 701 (1978).] (c) X-ray structure of the C1q globular head. The A subunit is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), the B subunit is purple, and the C subunit is pink. Their mutually bound Ca^{2+} ion is represented by a cyan sphere. [Based on an X-ray structure by Juan Fontecilla-Camps and Gérard Arlaud, Université Joseph Fourier, Grenoble Cedex, France. PDBid 1PK6.]

mutually bind a Ca^{2+} ion near the top of the nearly spherical globule.

The C1q globular domains bind antigen-bound antibodies through their recognition of the Fc region C_H domains of IgM, IgG1, IgG2, and IgG3 (the other immunoglobulins do not bind C1q). Moreover, C1 is only activated if at least two of its C1q heads are simultaneously bound to antibody, a process that requires the participation of at least two IgGs but only one IgM (recall that IgM is pentameric). IgM, the first immunoglobulin to be synthesized in response to an antigen, is therefore far more effective in activating the complement system than are IgGs. C1 is also activated by a wide variety of foreign substances produced by pathogens, including the lipopolysaccharides of gram-negative bacteria known as **endotoxins**, cell wall teichoic acids from gram-positive bacteria (Section 11-3Bc), and viral membranes. These various ligands share no obvious structural similarities, but since many of them are polyanionic, it may be that C1q recognizes their charge distribution.

The remaining C1 components, C1r and C1s, are homologous serine protease zymogens, two each of which are

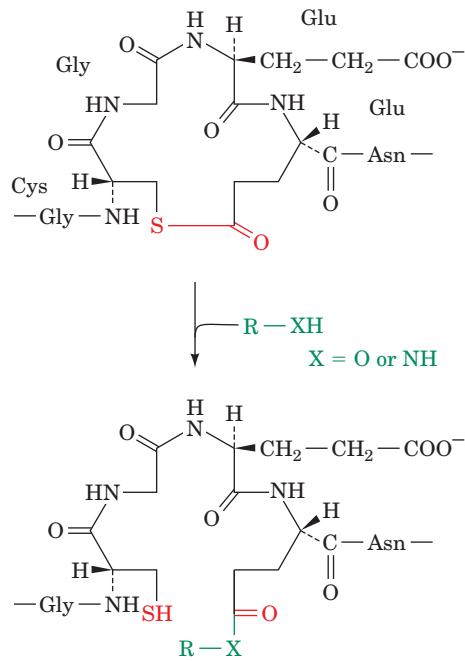
18-polypeptide chain protein, $\text{A}_6\text{B}_6\text{C}_6$, in which the ~ 80 N-terminal residues of each chain have the repeating sequence Gly-X-Y characteristic of collagen, where X is often Pro and Y is often 4-hydroxyproline or 5-hydroxylysine (Section 8-2B). C1q is therefore a bundle of six collagenlike triple helices that each end in a C-terminal globular domain so as to form an assembly that resembles a bunch of six tulips (Fig. 35-50a, b). The X-ray structure of a C1q globular domain (Fig. 35-50c) reveals that its A, B, and C subunits all form closely superimposable β sandwiches with jellyroll topology (Section 8-3Bg), each consisting of two 5-stranded antiparallel β sheets. The three subunits

bound to C1q between its collagenlike tails. C1r and C1s, like most of the blood clotting zymogens, are each activated by a single proteolytic cleavage that yields two disulfide-linked chains. The binding of antibody–antigen complex stimulates C1q to bind C1r and C1s more tightly, which, in a Ca^{2+} -dependent process, results in the autoactivation of C1r through the cleavage of an Arg—Ile bond. Activated C1r in turn activates C1s, its only known substrate, by specifically cleaving it, also at an Arg—Ile bond.

c. The Activation Unit

The activation unit consists of components derived from **C2**, **C3**, and **C4**. In the initial step forming the activation unit, activated C1s cleaves C4 at an Arg—Ala bond and the larger of the resulting fragments, **C4b**, covalently binds to the cell membrane (as described below) in the vicinity of the recognition unit. Activated C1s in association with membrane-bound C4b then specifically cleaves C2. The resulting larger fragment, the serine protease **C2b**, combines with C4b to yield C4b2b, a protease named **C3 convertase**, that excises the N-terminal 77-residue **C3a** from the α chain of the C3 $\alpha\beta$ heterodimer, leaving **C3b** [before its secretion, the 1641-residue **pro-C3** is cleaved by the excision of an (Arg)₄ segment yielding a 645-residue β chain and a 992-residue α chain that remain associated]. Finally, C3b combines with C3 convertase to yield the activation unit, C4b2b3b, also known as **C5 convertase**, which functions to activate **C5** by cleaving an Arg—Leu bond.

Both **C4** and **C3**, which are homologs, have buried hyperreactive thioester groups that, when exposed, can covalently link these proteins to the cell membrane. In C3, the thioester consists of a Cys thiol and a Glu γ -carboxyl group forming a macrocyclic ring of sequence Gly-Cys-Gly-Glu-Glu-Asn:



On cleavage of C3, the product C3b undergoes a massive conformational rearrangement that exposes its thioester

group (see below). The thioester then reacts with a nearby cell surface amine or hydroxyl group to yield the corresponding amide or ester bond together with a free Cys sulphydryl group. *This confines complement activation to the pathogen surface*, a process that is further discussed below. C4 behaves in a similar manner on activation. However, C5, which is homologous to C3 and C4, lacks the Cys and Glu residues that form the above thioester bond.

The X-ray structure of human C3, determined by Piet Gros, reveals that it folds into thirteen domains (Fig. 35-51a) in which the hyperreactive thioester group is buried in a hydrophobic pocket that shields it from reacting with water ($t_{1/2} \approx 230$ h). The X-ray structure of human C3b, also determined by Gros, reveals that the excision of C3a results in a dramatic and complex conformational change in which the domain containing the hyperreactive thioester group has swung around in a movement of >85 Å that exposes the thioester for covalent attachment to its target surfaces (Fig. 35-51b).

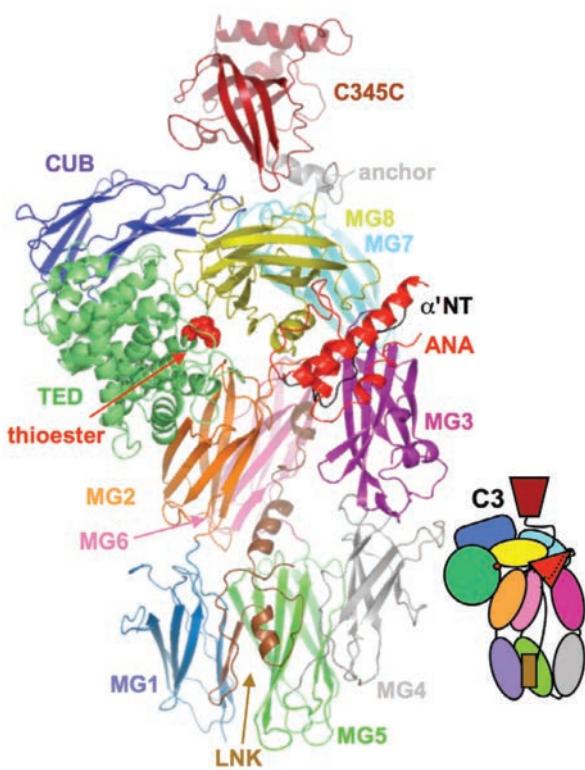
The activation of C3, C4, and C5 also triggers other immune system functions. C3b and, to a lesser extent, C4b are **opsonins**, substances that stimulate phagocytosis (opsonization), whereas C3a, **C4a**, and **C5a** are **anaphylatoxins**, substances that trigger local acute inflammatory reactions and smooth muscle contraction. When anaphylatoxins are produced in large amounts or injected systematically, they induce an often fatal general circulatory collapse like that seen in systematic allergic reactions involving IgE known as **anaphylactic shock**.

d. The Membrane Attack Complex

C5b, the larger product of the C5 convertase reaction, exhibits no proteolytic activity. Rather, it sequentially binds **C6** and **C7** to form a lipophilic complex that binds to the surface of cell membranes. *This C5b67 complex then binds one molecule of C8, whose α subunit spontaneously inserts into the membrane. Around 12 to 18 molecules of C9 then join this complex to form the MAC, in which, the C9 molecules form a tubular membrane-embedded structure to which the C5b678 complex is firmly attached (Fig. 35-52).* Cell lysis ensues because the MAC forms an ~ 100 -Å-diameter aqueous channel that pierces the membrane, which permits the cell's small molecules, but not its macromolecules, to exchange with the surrounding medium. Water is therefore osmotically drawn into the cell, causing it to swell and burst. MACs are efficient cell killers; very few, possibly only one, can lyse a cell. Nevertheless, deficiencies in C5, C6, C7, C8, and/or C9 are associated with susceptibilities to infection by only a few species of bacteria, most notably those causing gonorrhea and a common form of bacterial meningitis. Evidently, the other functions of the complement system, opsonization and the induction of inflammatory responses (see below), are more important for host defense against pathogenic microorganisms.

Complement components C6 to C9 and perforin (the pore-forming protein released by cytotoxic T cells; Section 35-2Aa) all contain an ~ 40 -kD MACPF domain. Not surprisingly, therefore, perforin pores have similar structures to MAC pores, although they lack an analog of the C5b678

(a)



(b)

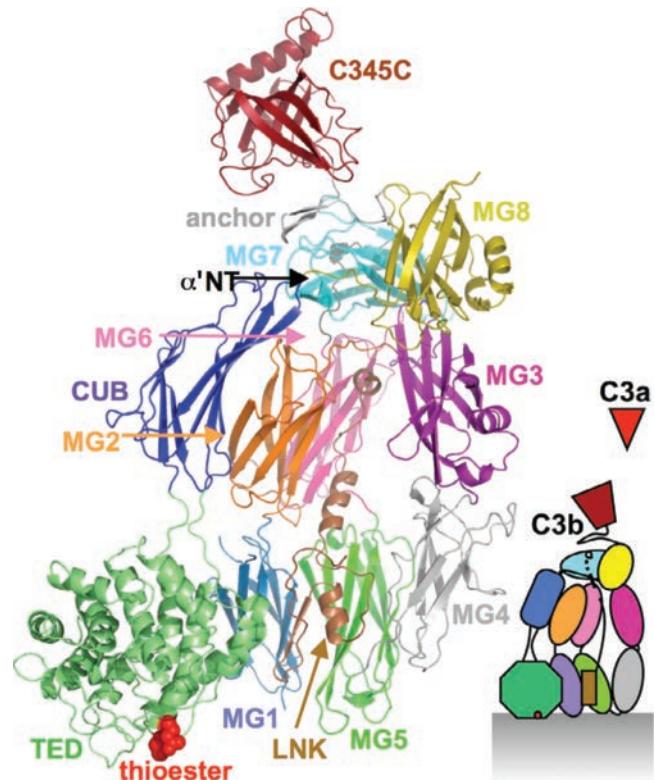


Figure 35-51 X-ray structures of human complement components C3 and C3b. (a) Complement C3 drawn in ribbon form with each of its 13 domains differently colored and with its hyperreactive thioester group shown in space-filling form in red. The β chain (residues 1–645) contains, from N- to C-terminus, domains MG1 through MG5 (MG for *macroglobulin-like*), the N-terminal half of MG6, and LNK (for *linker*). The α chain (residues 650–1641) contains the domains ANA (for *anaphylatoxin*; which when excised by C3 convertase becomes C3a), the C-terminal half of MG6, MG7, CUB (named for the complement subcomponents in which it was initially found: C1r/C1s, Uegf, and bone morphogenetic protein-1), TED (for

thioester-containing domain), MG8, and C345C (for C3, C4, and C5, in which it occurs). Note how the thioester group is occluded by the protein. The inset, which schematically shows this domain arrangement, has the same coloring scheme. (b) Complement C3b represented and viewed as in Part a. Note how the removal of C3a, which in C3 binds to MG8, causes MG7 and MG8 to swivel and CUB and TED to swing around relative to their positions and orientations in C3 so as to expose and position the thioester “warhead” to covalently attach to the surface of a pathogen. The gray bar at the bottom of the inset represents the pathogen surface. [Based on X-ray structures and drawings by Piet Gros, Utrecht University, The Netherlands. PDBids 2A73 and 2I07.]

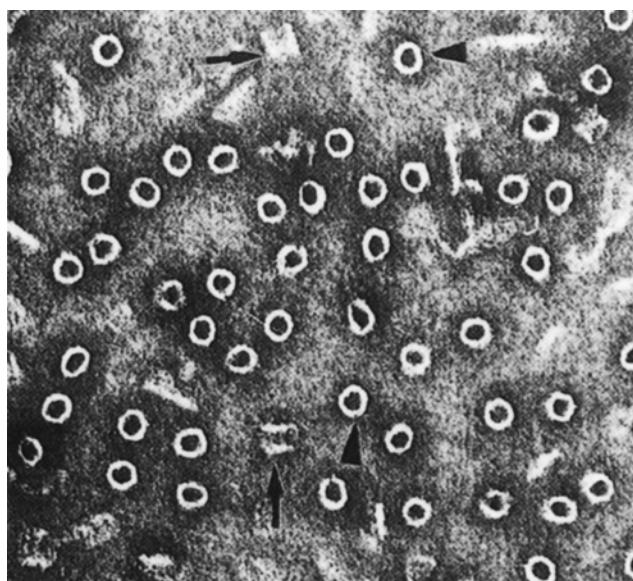


Figure 35-52 Electron micrograph of C9 ring complexes. Complexes are seen both in side view (arrows) and in top view (arrowheads). [Courtesy of Zanvil Cohn, The Rockefeller University.]

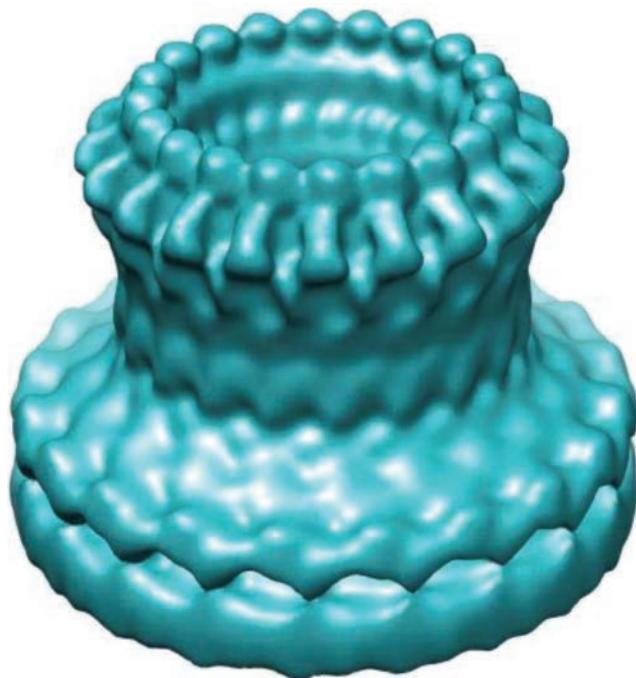


Figure 35-53 Cryoelectron microscopy-based image of a mouse perforin pore with 20-fold symmetry. The resolution is 28.5 Å. [Courtesy of Helen Saibil, Birkbeck College, London, U.K..]

complex. A cryo-EM-based image of a perforin pore is shown in Fig. 35-53.

e. The MB-Lectin Pathway Closely Resembles the Classical Pathway

The MB-lectin pathway of complement fixation (Fig. 35-49, *top*) is initiated by a complex of three proteins: **mannose-binding lectin (MBL)**, **MBL-associated serine protease 1 (MASP-1)**, and **MASP-2**. MBL is a six-headed protein that is a homolog of and whose electron micrographs closely resemble those of C1q (Fig. 35-50a). However, MBL is activated by its binding to two or more mannose and/or fucose residues that have the relatively short spacing that occurs on the surfaces of bacteria but not the longer spacing that is characteristic of host cell surfaces. MASP-1 and MASP-2 are both homologs of C1r and C1s. However, when MBL binds to the surface of a pathogen, MASP-2 is activated, whereas the role of MASP-1 is uncertain. In any case, activated MASP-2 proteolytically cleaves C2 and C4 in the same way as does activated C1s. After this point, the MB-lectin pathway follows the classical pathway.

The MB-lectin pathway is antibody independent. It therefore appears that the MB-lectin pathway functions to defend against invading microorganisms before an adaptive immune response against them can be mounted (although the classical pathway can also function in an antibody-dependent manner). Once sufficient antibody has been synthesized, the MB-lectin pathway assumes a secondary role relative to that of the classical pathway. Nevertheless, individuals deficient in MBL have a greatly

increased susceptibility to bacterial infections, particularly in early childhood when the maternal antibodies transferred across the placenta and obtained from colostrum have largely disappeared but the adaptive immune system is not yet fully matured.

f. The Alternative Pathway Is Also Antibody Independent

The alternative pathway of complement fixation (Fig. 35-49, *bottom*) uses many of the same components as the classical pathway and likewise causes formation of a C5 convertase that triggers MAC assembly. The alternative pathway, like the MBL-lectin pathway, is antibody independent, but unlike the MBL-lectin and classical pathways, does not require binding to a pathogen surface protein for activation.

The alternative pathway always operates at a low level (see below) so as to continually produce small amounts of C3b, the same protein produced by C3 convertase of the classical pathway. In the alternative pathway, however, C3b combines with the plasma protein **factor B** in a Mg^{2+} -dependent reaction (the conversion of C3 to C3b exposes a putative binding site for factor B). The resultant complex, C3bB, is the only known substrate for the active plasma serine protease, **factor D**, which cleaves the B subunit of C3bB to yield C3bBb. This latter complex is a C3 convertase that is equivalent to but distinct from that of the classical pathway (C3b is homologous to C4b and Bb is homologous to C2b). It cleaves C3 to C3b, which participates in the formation of more C3 convertase in a cyclically amplified process. The additional C3b also binds to C3 convertase to yield (C3b)₂Bb, a C5 convertase distinct from that of the classical pathway but which likewise catalyzes the formation of the MAC.

What is the origin of the C3b that initiates the alternative pathway? It may, of course, be generated by the classical pathway in which case the alternative pathway acts as an amplification mechanism for antibody-induced complement activation. In the absence of this process, however, the reactive but unexposed thioester bond of native C3 undergoes slow spontaneous hydrolysis to yield a C3b-like protein in that it binds factor B and mediates its D-catalyzed activation. The resulting C3 convertase, in turn, generates authentic C3b.

How does the alternative pathway target only invading microorganisms? The C3b concentration in solution is limited by a plasma protein named **factor I** which, together with a second protein, **factor H**, forms a complex (IH) that proteolytically degrades C3b in solution, first to **iC3b** and then to **C3dg**, neither of which is a component of C5 convertase. When C3b is covalently bound to a surface, however, its degradation rate is greatly reduced. Moreover, the surface-bound C3 convertase complex is stabilized by the binding of the plasma protein **properdin (factor P)**, which further protects C3b from IH-mediated degradation as well as retards the dissociation of Bb from C3 convertase. Consequently, the faster C3b covalently attaches to a surface, the more slowly it is degraded. *Substances to which C3b efficiently attaches are therefore alternative pathway*

activators. These include endotoxins, teichoic acids, certain whole bacteria, fungi, and cells infected by certain viruses. The alternative pathway therefore provides an effective defense against invading microorganisms.

g. C3b Is Also a Potent Opsonin

Since C3 is the most abundant complement protein in the blood plasma, the majority of the C3b serves to opsonize the pathogen to which it is attached, that is, to signal phagocytes to ingest the pathogen. Indeed, up to 1000 C3b molecules may bind in the vicinity of a single C3 convertase. Phagocytes recognize C3b through **complement receptor-1 (CR1)** on their plasma membrane surfaces. However, the binding of C3b to CR1 alone is insufficient to induce phagocytosis. For this to occur, phagocytes must also bind other immune mediators, such as C3a and C5a, which they bind via the homologous G protein-coupled receptors (GPCRs; Section 19-2B) known as **C3a** and **C5a receptors**. Moreover, C3dg, the IH-mediated cleavage product of C3b, binds only to **complement receptor-2 (CR2)** on B cells, thereby augmenting the signal generated when a B cell's membrane-bound immunoglobulin specifically binds to an epitope on the pathogen surface. Thus, *the activation of complement also helps produce a strong antibody response*.

h. The Complement System Is Strictly Regulated

The inability of many complement components to discriminate between normal tissues and foreign substances requires that the complement system be maintained under tight control. Otherwise the complement system would destroy host cells. Indeed, *the actual damage in many autoimmune diseases is caused by the complement system*.

The complement system is regulated by the inactivation of its activated components. This occurs in three ways:

1. Complement components are inactivated through their spontaneous decay. For example, the hyperreactive thioesters of newly activated C3b and C4b react with water with half-lives of $\sim 60 \mu\text{s}$. These proteins are therefore lost to the classical pathway unless they immediately attach to a membrane in the immediate vicinity of their activating recognition unit, that is, to the membranes of the invading microorganisms that triggered their activation (rather than to those of host cells). Similarly, classical pathway C3 convertase, C4b2b, is but transiently active; its C2b component readily dissociates with the consequent loss of enzymatic activity.

2. Complement components associated with host cells are inactivated through their degradation by specific proteases. For instance, factor I forms a complex with the membrane-bound **C4b-binding protein** that proteolytically degrades C4b, much like, as we have seen, the IH complex degrades C3b. **Membrane cofactor of proteolysis (MCP)** functions similarly to C4b-binding protein.

3. Complement components are inactivated through their association with specific binding proteins. For example, **C1 inhibitor (C1INH)**, a serine protease inhibitor (Sec-

tion 15-3Da), tightly binds C1r and C1s and causes them to dissociate from C1q, thereby inactivating the recognition unit. Similarly, **S protein** (also called **vitronectin**) attaches to C5b67 assembling in the plasma so as to prevent MAC insertion into cell membranes. MAC insertion is consequently limited to the site of complement activation. **Decay-accelerating factor (DAF)**, a homolog of C4b-binding protein that is GPI-anchored to host cell plasma membranes (Section 12-3Bc), competes with factor B for binding to C3b on host cell surfaces and displaces Bb from C3 convertase, thereby inhibiting complement activation on host cells.

The regulation of the complement system therefore functions to target foreign invaders while minimizing host cell damage.

i. Some Pathogens Evasive or Co-opt the Complement System

Pathogens have evolved a variety of mechanisms to elude the complement system or to utilize it for their own purposes. For example, the bacterium *Borrelia burgdorferi*, which causes Lyme disease, hijacks factor H from its host by binding it to its **complement regulator-acquiring surface protein (CRASP)** and thereby inhibits complement amplification on its surface. *Staphylococcus aureus* **extracellular fibrinogen-binding protein (Etf)** binds to and changes the conformation of C3 in a way that prevents the excision of C3a. **Vaccinia virus complement control protein (VCP)** is a homolog of complement regulators that inhibits complement activation. HIV, which causes AIDS, incorporates opsonized C3b in its membranous envelope (Fig. 15-33), which helps target it to complement regulator-expressing cells such as macrophages and dendritic cells. Thus, the development of strategies that antagonize these complement-evasive processes have therapeutic promise.

3 MOTILITY: MUSCLES, VESICLE TRANSPORT, CILIA, AND FLAGELLA

Perhaps the most striking characteristic of living things is their capacity for organized movement. Such phenomena occur at all structural levels and include such diverse vectorial processes as active transport through membranes, the translocation of DNA polymerase along DNA, the separation of replicated chromosomes during cell division, the intracellular transport of vesicles and organelles, the beating of flagella and cilia, and, most obviously, the contraction of muscles. In this section we consider the structural and chemical basis of biological motility. In doing so we shall be mainly concerned with **striated muscle** since it is the most familiar and best understood motility system. We shall also consider several other types of biological motors and end with discussions of eukaryotic cilia and prokaryotic flagella.

A. Structure of Striated Muscle

The voluntary muscles, which include the skeletal muscles, have a striated appearance when viewed under the light

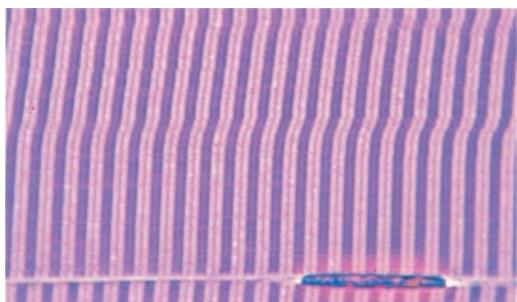


Figure 35-54 Photomicrograph of a muscle fiber in longitudinal section. Its transverse dark A bands and light I bands are clearly visible. [J.C. Revy/C.N.R.I/Photo Researchers]

microscope (Fig. 35-54). Such muscles consist of long parallel bundles of 10- to 100- μm -diameter **muscle fibers** (Fig. 35-55), which may span the entire length of the muscle and which are actually giant multinucleated cells that arise during muscle development by the end-to-end fusion of numerous precursor cells. Muscle fibers are, in turn, composed of in-register parallel bundles of around one thousand 2- to 5- μm diameter **myofibrils** (Greek: *myos*, muscle), which may extend the full length of a fiber.

Electron micrographs show that muscle fiber striations arise from an underlying banded structure of multiple in-register myofibrils (Fig. 35-56). The bands are formed by alternating regions of greater and lesser electron density, respectively, named **A bands** and **I bands** (Fig. 35-57). The myofibril's repeating unit, the **sarcomere** (Greek: *sarkos*, flesh), which is 2.5 to 3.0 μm long in relaxed muscle but progressively shortens as the muscle contracts, is bounded by dark (electron dense) **Z disks** or **lines** at the center of each I band. The A band is centered on the lighter **H zone** which, in turn, is centered on the dark **M disk** or **line**.

Cross sections through the sarcomere reveal the origin of its banded pattern. The H zones contain an array of parallel, hexagonally packed \sim 150- \AA -diameter and 1.6- μm -long **thick filaments** that are anchored to the M disk, whereas the lighter I bands consist of twice as many hexagonally arranged \sim 100- \AA -diameter and 1- μm -long **thin filaments** that are anchored to the Z disk. The darker areas at the ends of each A band mark the regions where the two sets of fibers interdigitate. The thick and thin filaments associate in this region by means of regularly spaced **cross-bridges** (Fig. 35-58). We shall see below that it is these associations that are responsible for the generation of muscular tension.

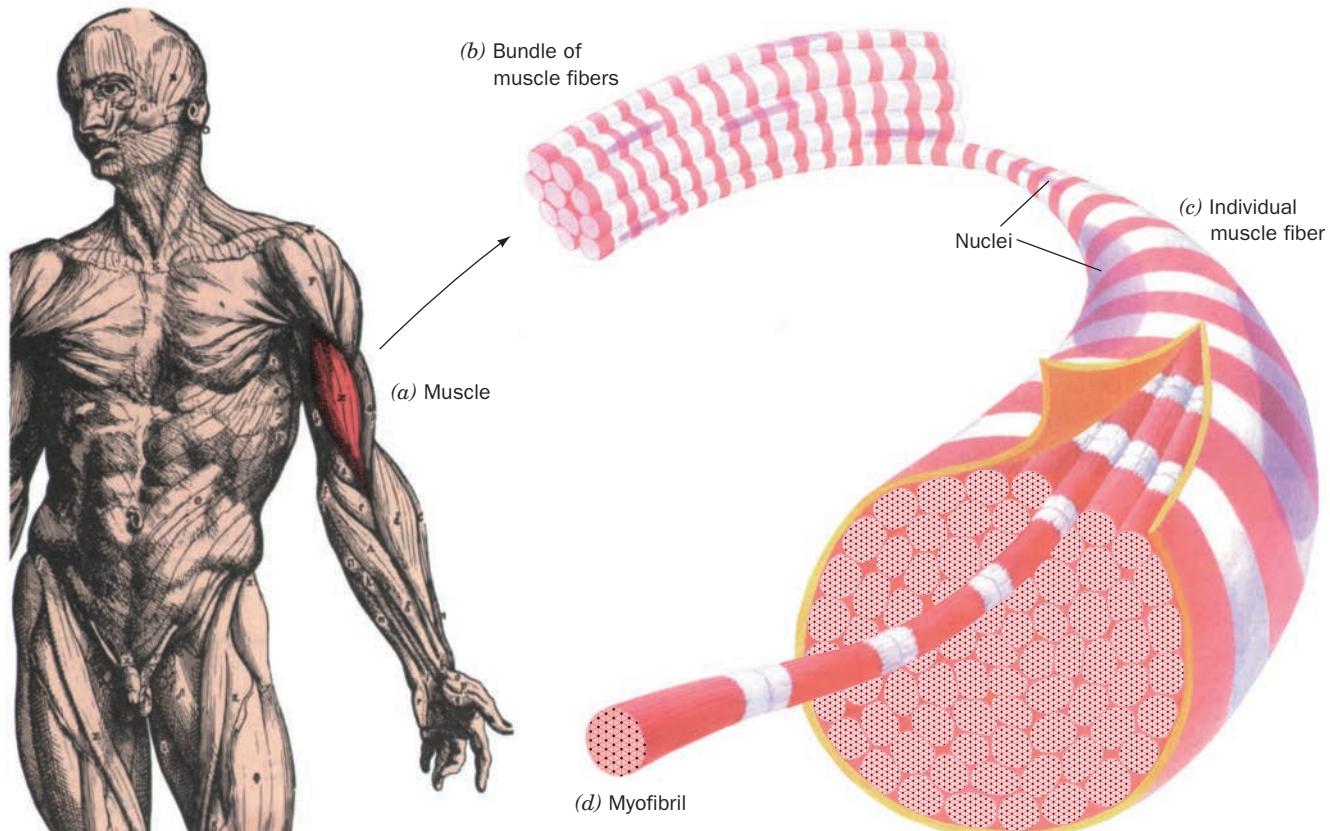


Figure 35-55 Skeletal muscle organization. A muscle (a) consists of bundles of muscle fibers (b), each of which is a long, thin, multinucleated cell (c) that may run the length of the

muscle. Muscle fibers contain bundles of laterally aligned myofibrils (d), which consist of bundles of alternating thick and thin filaments.

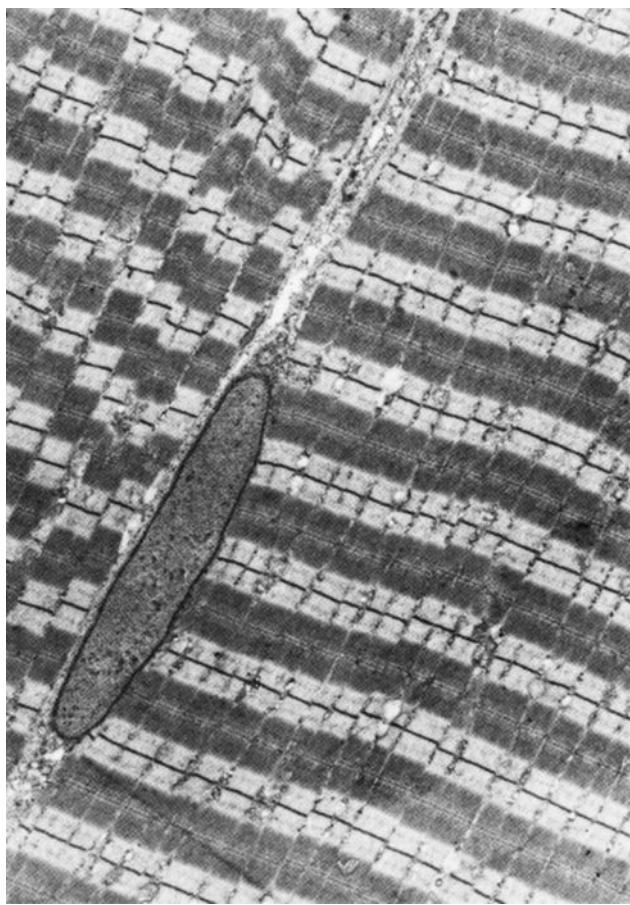


Figure 35-56 Electron micrograph of skeletal muscle showing that the myofibrils in a muscle fiber are in register. The out-of-register portion of the myofibril at the upper left is probably an artifact of preparation. The ovoid object near the center is a nucleus. Note its peripheral location in its fiber. [Don Fawcett/Photo Researchers, Inc.]

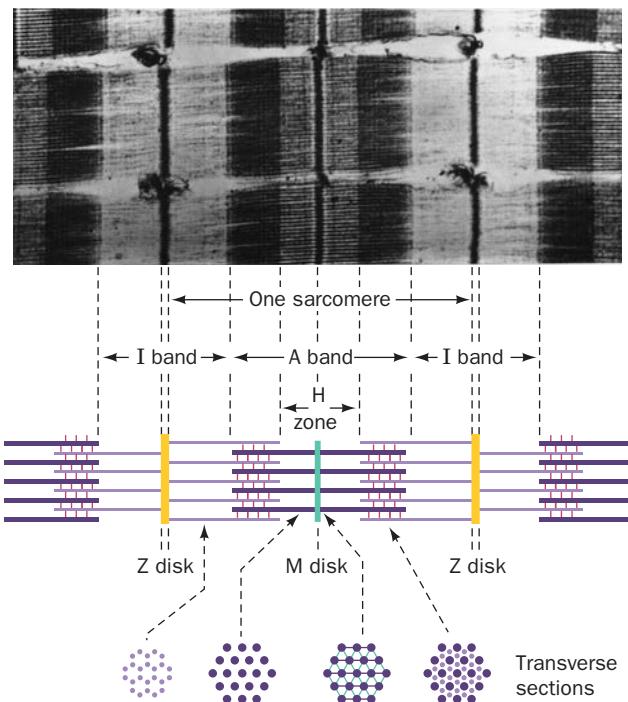


Figure 35-57 Anatomy of a myofibril. The electron micrograph shows parts of three myofibrils that are separated by horizontal gaps. A myofibril's major features, as indicated in the accompanying interpretive drawings, are the light I band, which contains only hexagonally arranged thin filaments; the A band, whose dark H zone contains only hexagonally packed thick filaments, and whose even darker outer segments contain overlapping thick and thin filaments; the Z disk, to which the thin filaments are anchored; and the M disk, which arises from a bulge at the center of each thick filament. The myofibril's functional unit, the sarcomere, is the region between two successive Z disks. [Courtesy of Hugh Huxley, Brandeis University.]

a. Thick Filaments Consist of Myosin

The major protein components of striated muscle are listed in Table 35-4. Vertebrate thick filaments are composed almost entirely of a single type of protein, **myosin**,

which occurs in virtually every vertebrate cell. **Myosin molecules consist of six highly conserved polypeptide chains: two 230-kD **heavy chains** and two pairs of different **light chains**, the so-called **essential** and **regulatory light chains****

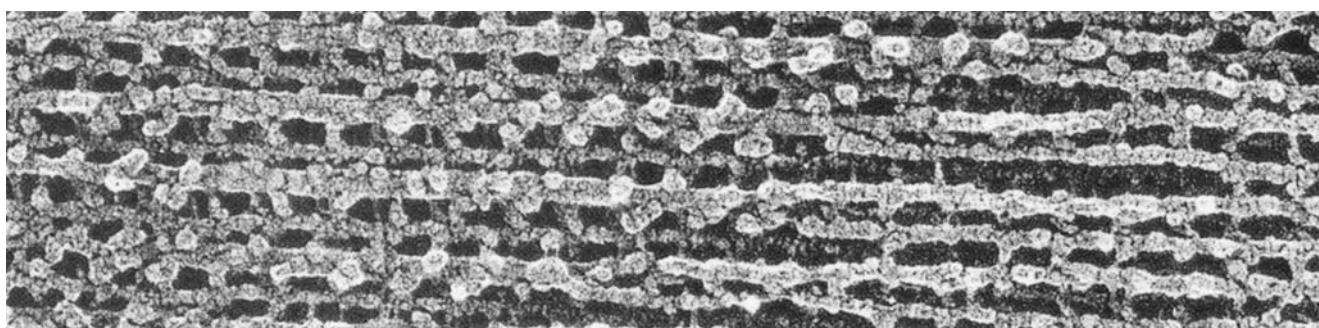


Figure 35-58 Electron micrograph of deep-etched, freeze-fractured myofibril showing its alternating thick and thin filaments. The knobs (cross-bridges) projecting from the thick

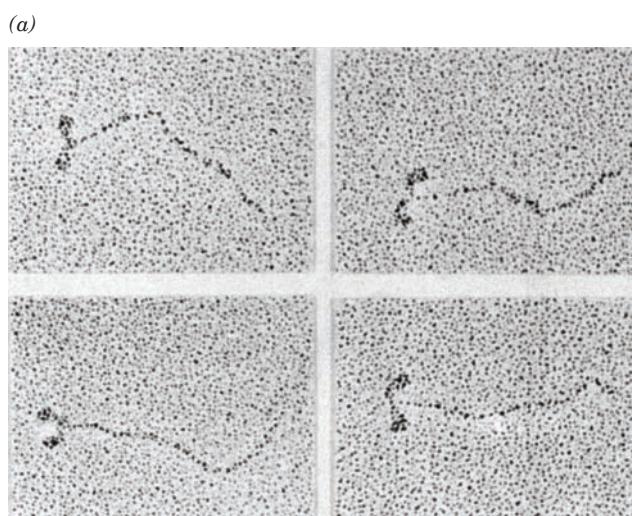
filaments are helically arrayed. [Courtesy of John Heuser, Washington University School of Medicine.]

Table 35-4 Proteins of Striated Muscle

Protein	Molecular Mass (kD)
Myosin	540
Heavy chain	230
Essential light chain (ELC)	15–22
Regulatory light chain (RLC)	15–22
G-Actin	42
Tropomyosin	33
Troponin	72
TnC	18
TnI	23
TnT	31
α -Actinin	200
Desmin	50
Vimentin	52
CapZ (β -actinin)	65
Tropomodulin	39
Titin	3816
Nebulin	773
Dystrophin	427
Myomesin	185
M-Protein	165
Myosin-binding protein (MyBP-C, C-protein)	128

(ELC and RLC), that vary in size between 15 and 22 kD, depending on their source. Myosin is an unusual protein in that it has both fibrous and globular properties (Fig. 35-59). The N-terminal half of its heavy chain folds into an elongated globular head, around $55 \times 200 \text{ \AA}$, whereas its C-terminal half forms a long fibrous α -helical tail. Two of these α -helical tails associate to form a left-handed parallel coiled coil, yielding an $\sim 1600\text{-\AA}$ -long rodlike segment with two globular heads. The amino acid sequence of myosin's α -helical tail is characteristic of coiled coils: It has a seven-residue pseudorepeat, *a-b-c-d-e-f-g*, with nonpolar residues concentrated at positions *a* and *d*. Thus, much like in the coiled coils of the fibrous protein keratin (Section 8-2A) and leucine zippers (Section 34-3Bp), the myosin helix has a hydrophobic strip along one side that promotes its lengthwise association with another such helix. One of each type of light chain is associated with each of the heavy chain dimer's globular heads.

Myosin exists as single molecules only at low ionic strengths. Under physiological conditions, these proteins form aggregates that resemble thick filaments. *Vertebrate thick filaments consist of several hundred myosin molecules with their rodlike tails forming a bipolar staggered array* (Fig. 35-60). The globular myosin heads project from both ends of the thick filament in a quasi-helical arrangement, leaving a bare central region (Fig. 35-60b) in which the myosin tails associate in an antiparallel manner. It is the myosin heads that form the cross-bridges that interact with the thin filaments in intact myofibrils.



(b)

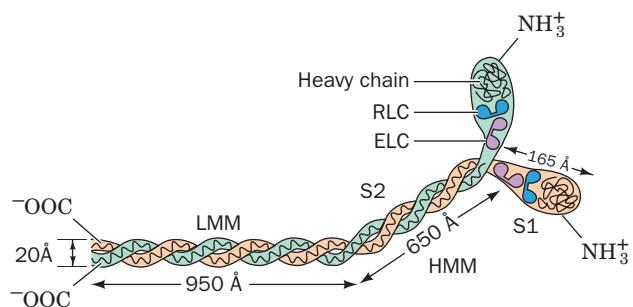


Figure 35-59 The myosin molecule. (a) Electron micrographs showing that the myosin molecule is a fibrous entity with two globular heads. [Courtesy of Paula Flicker, Theo Wallimann, and Peter Vibert, NIH.] (b) Its rod-shaped tail is formed by the two extended α helices, one from each of its two identical heavy chains, that wrap around each other to form a parallel coiled coil. One of each type of myosin light chain, an essential light chain (ELC) and a regulatory light chain (RLC), is associated with each of myosin's identical globular heads.

In addition to its structural function, the myosin heavy chain is an ATPase: It hydrolyzes ATP to ADP and P_i in a reaction that powers muscle contraction. *Muscle is therefore a device for transducing the chemical free energy of ATP hydrolysis to mechanical energy*. The myosin light chains, through their level of phosphorylation, are thought to modulate the ATPase activity of their associated heavy chains.

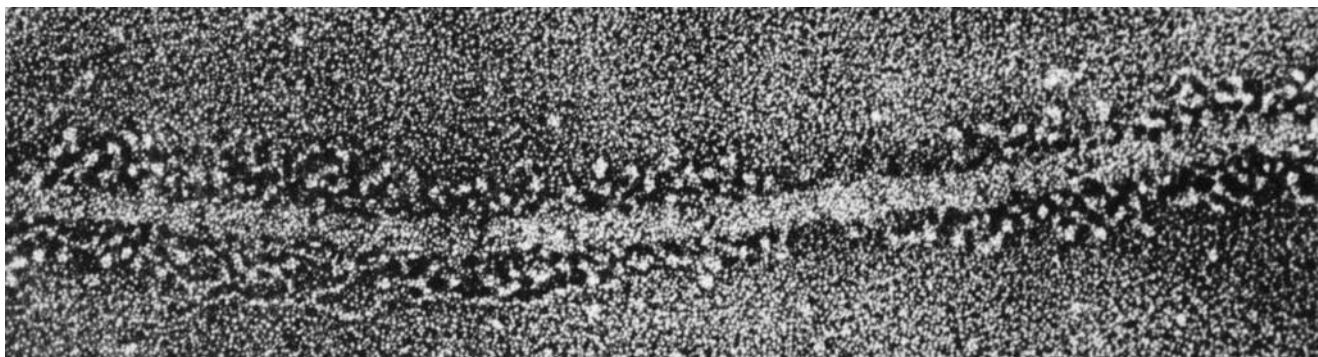
In 1953, Andrew Szent-Györgyi demonstrated that limited trypsin digestion cleaves myosin into two fragments (Fig. 35-61):

1. Light meromyosin (LMM), a 950- \AA -long α -helical rod that aggregates to form filaments but lacks both ATPase activity and the ability to associate with light chains.

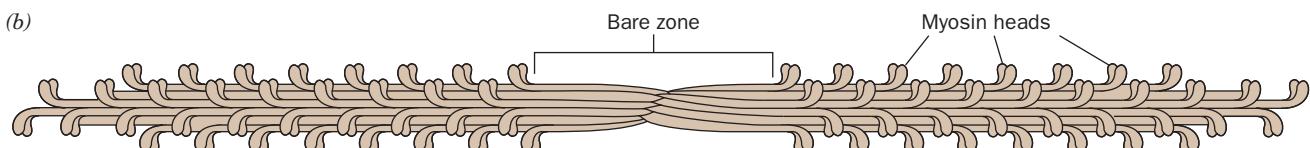
2. Heavy meromyosin (HMM), which has a rodlike tail and two globular heads, does not aggregate, but has ATPase activity and binds to light chains.

HMM can be further split by treatment with papain to yield two identical molecules of **subfragment-1 (S1)** and one of

(a)



(b)

**Figure 35-60** The thick filament of striated muscle.

(a) Electron micrograph showing the myosin heads projecting from the thick filament's outer segments and its bare central zone. [From Trinick, J. and Elliott, A., *J. Mol. Biol.* **131**, 135

(1977).] (b) A thick filament typically contains several hundred myosin molecules organized in a repeating staggered array such that the myosin molecules are oriented with their globular heads pointing away from the filament's center.

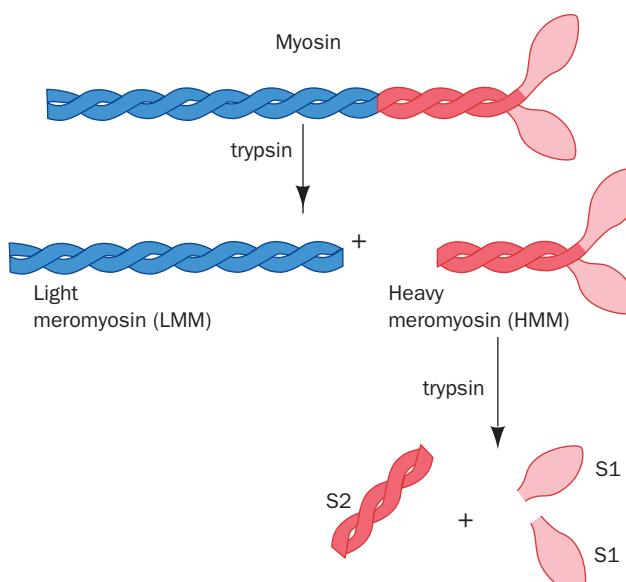
the rod-shaped **subfragment-2 (S2)**. The 130-kD S1, which contains myosin's ATPase activity and its thin filament-binding site, consists of a 95-kD (843-residue) heavy chain fragment and one molecule each of ELC and RLC. S1 is a pear-shaped molecule that electron microscopy studies indicate is 190 Å long and 50 Å wide at its widest point. Further trypsic digestion of S1 yields three fragments: a 25-kD N-terminal segment that binds nucleotide (ATP or ADP), a central 50-kD segment, and a 20-kD C-terminal segment.

b. X-Ray Structure of Myosin Subfragment-1

The X-ray structure of chicken muscle S1, determined by Ivan Rayment and Hazel Holden, reveals this tadpole-shaped protein to be nearly 50% α helical (Fig. 35-62). However, the core of the molecule consists of a mostly parallel 7-stranded β sheet whose strands are contributed by all three trypsic segments. Perhaps the most conspicuous structural feature of S1 is an ~85-Å-long helix that extends from the thick part of the head, the so-called motor domain, to the protein's C-terminal region at its narrow end. The motor domain contains the binding site for ATP as well as that for the protein **actin**, the thin filament's major component (see below).

The ATP-binding site, which is located in a 13-Å-deep V-shaped pocket at the point where the three trypsic segments come together, was identified from its resemblance to such sites in G proteins such as Ras (Section 19-3Cf) and from the positions of amino acids previously shown to be at myosin's ATP-binding site. The X-ray structures of myosins in their complexes with ATP analogs reveal that this pocket closes around ATP, as the tight binding of ATP to myosin (association constant $K = 3 \times 10^{11} M^{-1}$) had previously suggested. The site that is implicated in binding actin is located on the opposite side of the S1 globule from the nucleotide-binding site and is formed by the 50-kD trypsic fragment. The nucleotide and actin binding sites are connected by a deep cleft that divides the 50-kD fragment into two domains, the so-called actin cleft. The way in which myosin and actin interact is considered below.

The two myosin light chains share both sequence and structural homology with calmodulin (CaM; Section 18-3Cd). However, RLC contains only one of CaM's four Ca^{2+} -binding motifs and ELC has none. ELC embraces the

**Figure 35-61** The enzymatic cleavage pattern of myosin.

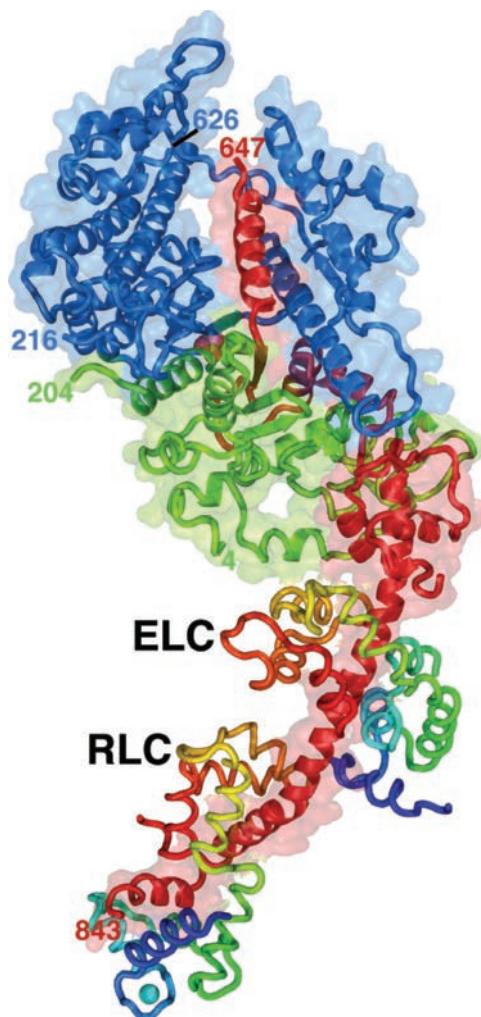


Figure 35-62 X-ray structure of chicken muscle myosin subfragment-1 (S1). The heavy chain is represented by a ribbon diagram embedded in its semitransparent molecular surface in which its 25-, 50-, and 20-kD segments are green, blue, and red, respectively, and with the numbers indicating the N- and C-terminal residues of each of these segments. The essential and regulatory light chains, RLC and ELC, are drawn in worm form with each colored in rainbow order from its N-terminus (blue) to its C-terminus (red). A sulfate ion, shown in space-filling form (C red and S yellow), is bound near the confluence of the three heavy chain fragments, where it occupies the binding site of ATP's β -phosphate group (sulfate is a competitive inhibitor of myosin's ATPase function). An RLC-bound Ca^{2+} ion (lower left) is represented by a cyan sphere. Note the prominent vertical cleft that divides the 50-kD fragment, which serves as a communications route between the active site and the actin-binding site. [Courtesy of Ivan Rayment and Hazel Holden, University of Wisconsin. PDBid 2MYS.]

middle region of myosin's long helix (Fig. 35-62) in a manner resembling the way that CaM interacts with the CaM-binding helix of myosin light chain kinase (Fig. 18-19). RLC also clasps the long helix, but near its C-terminus and in a manner different from that of ELC (Fig. 35-62). Nevertheless, both light chains bind to similar sequences on myosin, the so-called **IQ motif**, which has the core consensus sequence IQXXXRGXXXR.

c. Thin Filaments Consist of Actin, Tropomyosin, and Troponin

Actin, the most abundant cytosolic protein in eukaryotes (comprising \sim 20% of the protein in muscle cells and up to 15% of the protein in nonmuscle cells), is the major constituent of thin filaments. At low ionic strengths, actin occurs as \sim 375-residue bilobal globular monomers called **G-actin** (G for globular) that each bind one molecule of ATP. Under physiological conditions, however, G-actin polymerizes to form fibers known as **F-actin** (Fig. 35-63; F for fibrous), a process that hydrolyzes the ATP to ADP, which remains bound to the F-actin monomer unit. *F-actin forms the core of the thin filament, which typically contains \sim 360 actin monomers.*

Actin has a single high-affinity metal ion binding site that has a greater affinity for Ca^{2+} than for Mg^{2+} . However, because the *in vivo* concentration of Mg^{2+} is much higher than that of Ca^{2+} , this site is largely occupied by Mg^{2+} *in vivo*. Actin is one of the most highly conserved eukaryotic proteins (human and chicken muscle actins are identical and 88% identical to yeast cytoplasmic actin), with the actins expressed in the corresponding tissues of different species more closely related than the several actin isoforms expressed in different tissues within an organism. In fact,

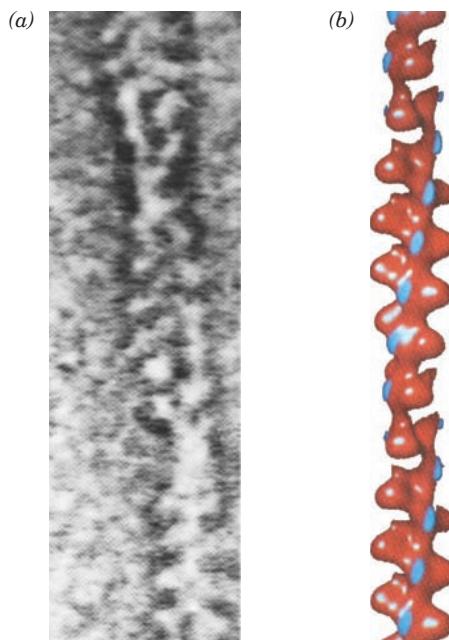


Figure 35-63 F-actin. (a) Electron micrograph of a thin filament from striated muscle. [Courtesy of Hugh Huxley, Brandeis University.] (b) Cryo-EM-based image of an actin fiber (red). Note the bilobal appearance of each monomeric (repeating) unit. The tropomyosin binding sites (see text) are blue. The F-actin helix has a maximum diameter of \sim 100 Å, 2.17 actin monomers per left-handed helical turn (13 subunits in 6 turns), and a rise per turn of \sim 60 Å. [Alternatively, F-actin may be described as a double (two-start) helix with 13 subunits per right-handed turn of each strand and a pitch of 720 Å.] [Courtesy of Daniel Safer, University of Pennsylvania, and Ronald Milligan, The Scripps Research Institute.]

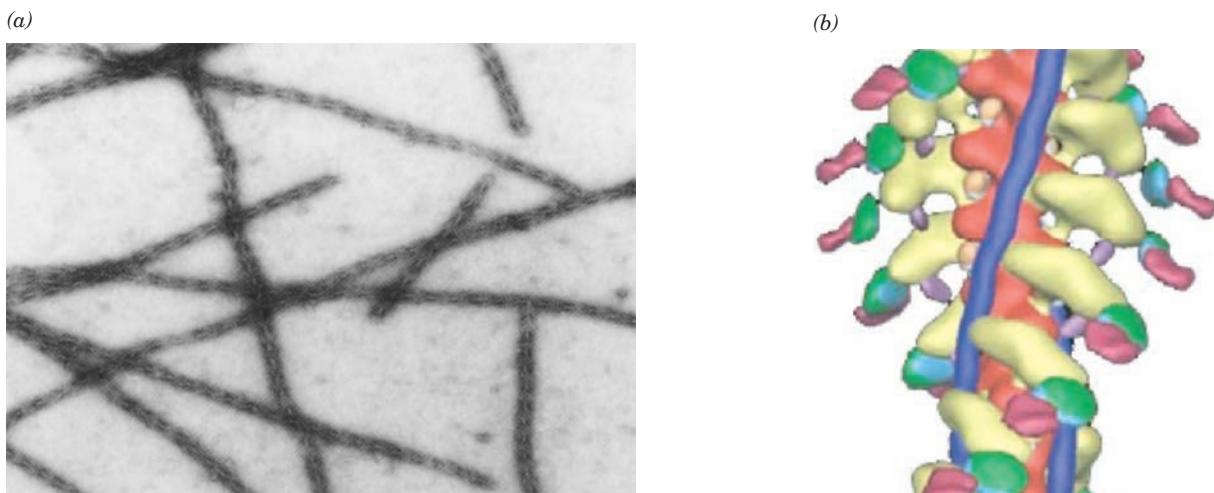


Figure 35-64 Thin filament decorated with myosin S1

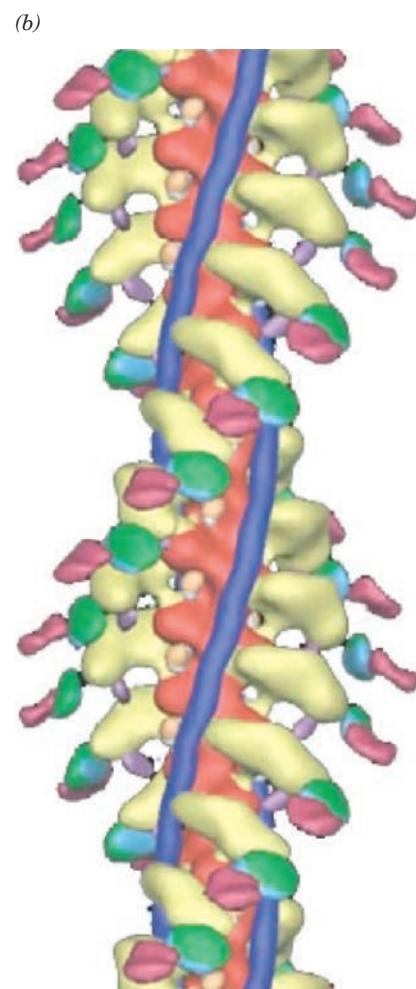
fragments. (a) Electron micrograph. Note the resemblance to a series of arrowheads all pointing in the same direction along the filament. [Courtesy of Hugh Huxley, Brandeis University.] (b) Cryo-EM-based image at $\sim 25\text{-}\text{\AA}$ resolution. F-actin is red, tropomyosin is purple, the myosin motor domain is yellow, the essential light chain (ELC) is cyan, the regulatory light chain (RLC) is pink, and the SH3 domain (whose functional role is unclear) at the C-terminus of the myosin heavy chain is lavender. The helical filament has a pitch of 370 \AA . [Courtesy of Ronald Milligan, The Scripps Research Institute, La Jolla, California.]

human cytoplasmic actin is more closely related to yeast cytoplasmic actin than it is to human muscle actin.

Each of muscle F-actin's monomeric units is capable of binding a single myosin S1 head. Electron micrographs of S1-decorated F-actin have the appearance of a series of head-to-tail arrowheads (Fig. 35-64a). F-actin must therefore be a polar entity; that is, all of its monomer units have the same orientation with respect to the fiber axis (Fig. 35-64b). The “arrowheads” in S1-decorated thin filaments that are still attached to their Z disk all point away from the Z disk, indicating that *the thin filament bundles extending from the two sides of the Z disk have opposite orientations*.

Myosin and actin, the major components of muscle, account for 60 to 70% and 20 to 25% of total muscle protein, respectively. Of the remainder, two proteins that are associated with the thin filaments are particularly prominent:

1. Tropomyosin, a homodimer whose two 284-residue α helical subunits wrap around each other to form a parallel



coiled coil (Fig. 35-65). These $\sim 400\text{-}\text{\AA}$ long rod-shaped molecules are joined head-to-tail with ~ 9 -residue overlap to form cables wound in the grooves of the F-actin helix such that each tropomyosin molecule contacts seven consecutive actin monomers in a quasi-equivalent manner (Fig. 35-64b).

2. Troponin, which consists of three subunits: **TnC**, a Ca^{2+} -binding protein that is 70% identical to calmodulin; **TnI**, which binds to actin; and **TnT**, an elongated molecule, which binds to tropomyosin at its head-to-tail junctions. The X-ray structure of troponin in complex with four Ca^{2+}



Figure 35-65 Low-resolution X-ray structure of rabbit cardiac muscle tropomyosin. Its two identical 284-residue polypeptides, which are colored in rainbow order from N-terminus (blue) to C-terminus (red), form a parallel coiled coil of α helices. Only the

residues at the ends of the chains are not in α helices. [Based on an X-ray structure by George N. Phillips, Jr. and Carolyn Cohen, Brandeis University. PDBid 2TMA.]

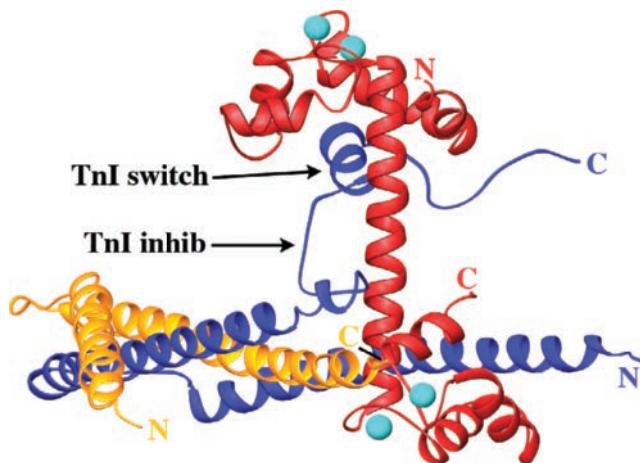


Figure 35-66 X-ray structure of chicken skeletal muscle troponin. TnC is red, TnI is blue, and TnT is gold. The four Ca^{2+} ions bound by TnC are represented by cyan spheres. Note the structural resemblance of TnC to the homologous Ca^{2+} -binding regulatory protein calmodulin (Fig. 18-17). [Based on an X-ray structure by Robert Fletterick, University of California at San Francisco. PDBid 1YTZ.]

ions (Fig. 35-66), determined by Robert Fletterick, reveals that the structure of TnI closely resembles that of calmodulin (Fig. 18-17) and the myosin light chains and that the inhibitory segment of TnI binds to TnC's rigid central helix in this Ca^{2+} -activated state.

The tropomyosin-troponin complex, as we shall see in Section 35-3B, regulates muscle contraction by controlling the access of the myosin S1 cross-bridges to their actin-binding sites.

d. Structure of Actin

The X-ray structure of rabbit muscle G-actin in complex with ATP and Ca^{2+} ion was determined by Leslie Burtnick (Fig. 35-67). G-actin is a rather flat molecule that consists of two domains, which for historical reasons, are referred to as the small and large domains, even though the former is only slightly smaller than the latter. Each domain has extensive secondary structure and is divided into two subdomains. The ATP binds at the bottom of a deep cleft between the two domains, where the Ca^{2+} ion is liganded in part by ATP's β - and γ -phosphates. Subdomains 1 and 3 (magenta and orange in Fig. 35-67) each contain a 5-stranded β sheet consisting of a β hairpin motif followed by a right-handed $\beta\alpha\beta$ motif (Fig. 8-46a), suggesting that these subdomains arose by gene duplication even though their amino acid sequences exhibit no significant similarity.

The fibrous nature of F-actin has thwarted its crystallization. Consequently, our current understanding of the atomic structure of F-actin is based on lower resolution images generated by cryo-EM-based techniques and the X-ray analysis of oriented F-actin gels into which the high resolution X-ray structure of G-actin had been fitted and adjusted (Fig. 35-68). Two such models, independently generated by Keiichi Namba and by Toshiro Oda, are in close

agreement. Both indicate that in forming F-actin, the propeller twist between G-actin's two domains is reduced by $\sim 20^\circ$, which partially closes the cleft between them and flattens the molecule, and that the 2-turn α helix in subdomain 2 of G-actin (upper right of Fig. 35-67) is converted to an extended loop in F-actin. The actin polymer is a double helix of subunits in which each actin subunit has the same head-to-tail orientation (e.g., all the nucleotide-binding clefts open upward in Fig. 35-68), so the assembled fiber has a distinct polarity. Successive actin subunits within a chain are rotated $\sim 167^\circ$ relative to each other about the helix axis, with the small domain of each actin subunit at higher radius from the F-actin helix axis than the large domain. The main contacts between protomers are between subdomain 1 and subdomain 2 below it, and between subdomain 3 and subdomain 4 below it, whereas the contacts between horizontally adjacent protomers in the double helix are surprisingly modest.

The end of the fiber toward which the nucleotide-binding sites open is known as the **(-)** end (also known as the **pointed end** from its appearance in the electron microscope when decorated with myosin S1 fragments; Fig. 35-64), and the opposite end is the **(+)** end (also known as the **barbed end**). The **(+)** ends of the thin filaments bind to the Z disk.

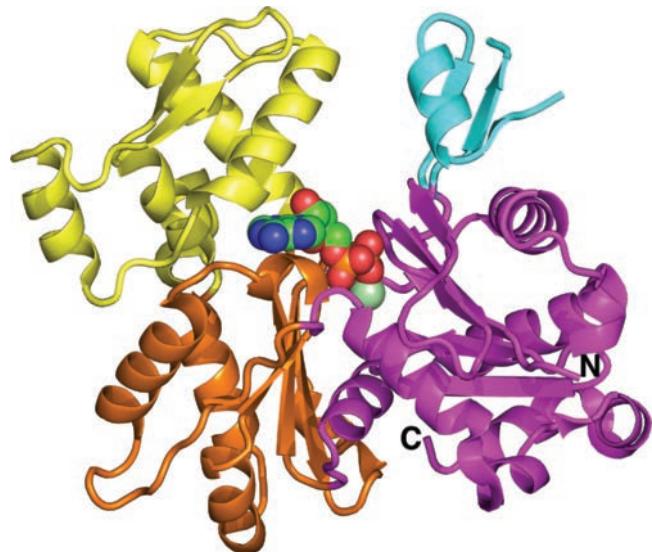


Figure 35-67 X-ray structure of rabbit muscle actin in complex with ATP and a Ca^{2+} ion. Subdomains 1 (magenta) and 2 (cyan) form the so-called small domain, whereas subdomains 3 (orange) and 4 (yellow) form the so-called large domain. The ATP, which is drawn in space-filling form with C green, N blue, O red, and P orange, binds at the bottom of a deep cleft between the domains. The Ca^{2+} ion, which is represented by a light green sphere, is liganded by the β - and γ -phosphates of the ATP. Note that both the N- and C-termini occur in subdomain 1. The so-called D-loop, which consists of residues 40 to 50 in subdomain 2, is disordered. [Based on an X-ray structure by Leslie Burtnick, University of British Columbia, Vancouver, British Columbia, Canada. PDBid 3HBT.]

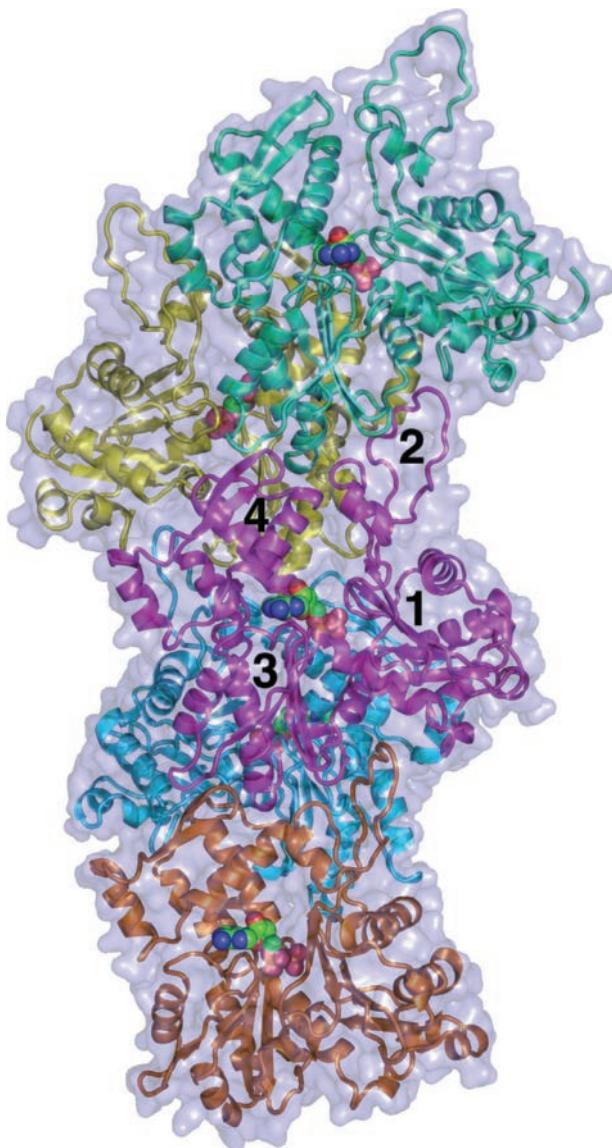


Figure 35-68 Structure of the F-actin filament. This model, which was generated by fitting the X-ray structure of G-actin into a 6.6-Å-resolution cryo-EM-based image of rabbit muscle F-actin, is represented by five consecutive actin · ADP subunits drawn in ribbon form, each with a different color, embedded in their semitransparent molecular surface. The ADPs are drawn in space-filling form with C green, N blue, O red, and P orange. The central subunit (magenta) is oriented as is the G-actin in Fig. 35-67 with the numbers indicating its subdomains. [Based on a structure by Keiichi Namba, Osaka University, Japan. PDBid 3MFP.]

e. Minor Muscle Proteins Control Myofibril Assembly and Integrity

The myofibril contains numerous minor proteins, several of which we discuss here. The Z disk, which anchors two sets of oppositely oriented thin filaments (Fig. 35-57), is an amorphous entity that contains several fibrous proteins. For instance, **α-actinin**, a rodlike homodimeric protein that cross-links F-actin filaments, is localized in the

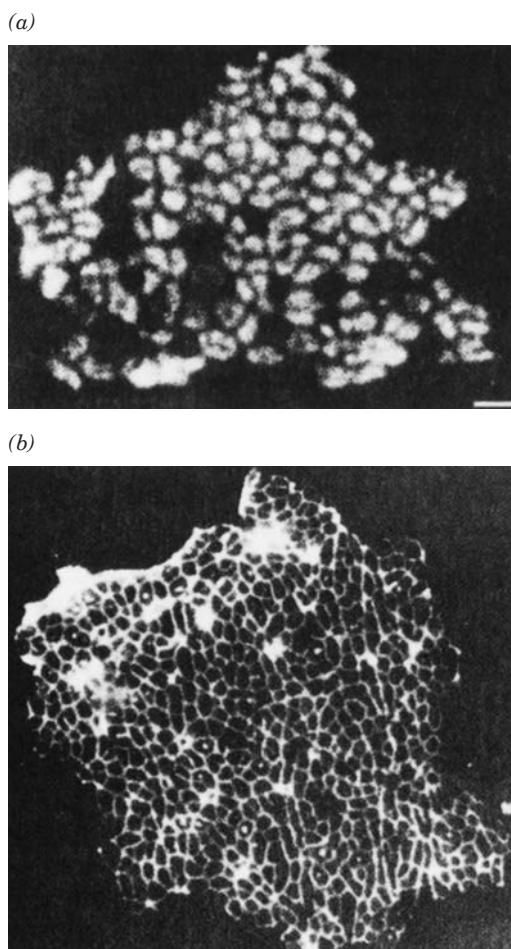


Figure 35-69 Indirect immunofluorescence micrographs of isolated sheets of skeletal muscle Z disks. (a) Using antibodies to α -actinin, indicating that α -actinin occurs at the interior of the Z disk. The bar represents 2.5 μ m. (b) Using antibodies to desmin, showing that desmin is distributed about the Z-disk periphery. Antibodies to vimentin exhibit the same distribution. In the indirect immunofluorescence technique, proteins are labeled with rabbit antibodies raised against them. The bound rabbit antibodies are subsequently labeled with goat anti-rabbit immunoglobulin antibodies to which fluorescent molecules such as fluorescein are covalently linked. The proteins are then observed under UV light. This indirect approach of using two types of antibodies rather than directly using fluorescently tagged rabbit antibodies increases the sensitivity of the method because several fluorescently labeled goat antibodies can bind to each rabbit antibody. [From Lazarides, E., *Nature* **283**, 251 (1980).]

Z disk's interior (Fig. 35-69a) and is therefore thought to attach oppositely oriented thin filaments to the Z disk. Two intermediate filament proteins, **desmin** and **vimentin** (Fig. 1-8d) largely occur at the Z disk periphery (Fig. 35-69b), where they apparently act to keep adjacent myofibrils in lateral register, a requirement for their proper function (Section 35-3B). **CapZ** (also called β -actinin), is an α -actinin-associated heterodimer that binds to the (+) end of F-actin so as to inhibit its further polymerization and its depolymerization, thus helping maintain the precise thin filament length

necessary for proper muscle function (actin polymerization and depolymerization are discussed in Section 35-3Eb). **Tropomodulin** similarly caps the (–) end of F-actin, but only when it is coated with tropomyosin.

Titin, the third most abundant muscle protein, is the product of a gene that in humans encodes 34,350 residues (3816 kD) and hence expresses the largest known polypeptide (it was discovered because it does not enter polyacrylamide gels). It extends from the M disk, where its N-terminal ends from adjacent sarcomeres overlap, to the Z disk, where its C-terminal ends likewise overlap, thereby forming a third filament system in striated muscle and one that runs the entire length of the myofibril. Six molecules of titin are associated with each half of a thick filament. Titin consists largely of 244 domains, of which 132 are immunoglobulin-like (Ig) and 112 are fibronectin type III-like (Fn-3; **fibronectin** is a multidomain protein that serves as a linker in the extracellular matrix). It also contains a region that has 31 ~26-residue PEVK repeats, so called because they are 70% Pro, Glu, Val, and Lys. The Fn-3 domains are located only in titin's A-band region, the PEVK region occurs only in the I band, and the Ig domains are distributed along the entire length of the protein. At low stretching forces, the randomly aligned Ig domains in titin's I-band region straighten (but do not unfold), whereas at higher forces, the PEVK region lengthens from a semifolded state (it lacks a stable conformation due to the charges on its Glu and Lys residues) to an extended state. Thus, *titin functions as a spring that keeps the thick filament centered in the sarcomere and permits the myofibril to recover when it is overstretched*. In addition, titin contains specific attachment sites for several sarcomeric proteins and thus appears to direct the organization of the thick filament.

Nebulin, which is also extremely large (6669 residues, 773 kD), consists, over ~97% of its length, of 185 repeats of an ~35-residue actin-binding motif that is predicted to be α helical (although it lacks the heptad repeats of α helices that form coiled coils). The observations that the aggregate length of these putative α helices is approximately that of a thin filament, that nebulin antibodies label the thin filament, and that this labeling pattern is fixed with respect to the Z disk suggest that nebulin winds along the ~1- μ m full length of a thin filament, thereby controlling this length and forming a fourth filament system in striated muscle. This proposal is consistent with the observation that the N-terminal end of nebulin binds to tropomodulin. The A-band region of titin, which binds to myosin at multiple sites, appears to similarly control the length of thick filaments.

The M disk (Fig. 35-57) arises from the local enlargement of in-register thick filaments. Two proteins that are associated with this structure, **myomensin** and **M-protein**, bind to titin and are therefore likely to participate in thick filament assembly, as does the thick filament-associated **myosin-binding protein C (MyBP-C; alternatively, C-protein)**. Invertebrate thick filaments contain a core of **paramyosin** which, in some muscles, is the dominant component.

Duchenne muscular dystrophy (DMD) and the less severe **Becker muscular dystrophy (BMD)** are both X-linked

muscle-wasting diseases. In DMD, which occurs once in ~3500 male births and has an onset age of 2 to 5 years, muscle degeneration exceeds muscle regeneration, causing progressive muscle weakness and ultimately death, typically due to respiratory disorders or heart failure, usually by age 25. In BMD, which occurs once in ~30,000 male births, the onset age is 5 to 10 years and there is an overall less progressive course of muscle degeneration and a longer (sometimes nearly normal) lifespan than in individuals with DMD.

The protein whose defects are responsible for DMD/BMD is named **dystrophin**. This 3685-residue polypeptide is encoded by an ~2400-kb gene (the largest known human gene) that contains 79 exons (which comprise only 0.44% of the gene sequence). However, dystrophin has numerous isoforms that differ at their C-termini through alternative mRNA splicing, as well as at their N-termini through alternative transcriptional initiation sites (Section 31-4Am). Dystrophin appears to be a member of the family of flexible rod-shaped proteins that includes α -actinin and the actin-binding cytoskeletal component spectrin (Section 12-3Db), each of which contains segments homologous to portions of dystrophin. Subcellular fractionation and immunofluorescence studies reveal that dystrophin, which has a normal abundance in muscle tissue of 0.002%, is associated with the inner surface of the muscle plasma membrane where it helps anchor F-actin to the extracellular matrix, much as do spectrin and ankyrin in the erythrocyte (Fig. 12-38d). Dystrophin thereby protects the plasma membrane from being torn by the mechanical stress of muscle contraction. Although such small tears are common in muscle cells, they occur much more frequently in dystrophic cells, leading to a greatly increased rate of cell death.

The dystrophin gene in most individuals with DMD/BMD contains deletions or, less frequently, duplications of one or more exons. Individuals with DMD usually have no detectable dystrophin in their muscles, whereas those with BMD mostly have dystrophins of altered sizes. Evidently, the dystrophins of individuals with DMD are rapidly degraded, whereas those of individuals with BMD are semifunctional.

B. Mechanism of Muscle Contraction

So far we have only described the components of striated muscle. Now, like good engineers, we must ask how those components fit together and how they interact. In other words, how does muscle work?

a. Thick and Thin Filaments Slide Past Each Other during Muscle Contraction

Physiologists have long known that a contracted muscle is as much as one-third shorter than its fully extended length. Electron micrographs have demonstrated that this shortening is a consequence of a decrease in the length of the sarcomere (Fig. 35-70). Yet, during muscle contraction, the thick and the thin filaments maintain constant lengths as is indicated by the observations that the width of the A band as well as the distance between the Z disk and the

(a)

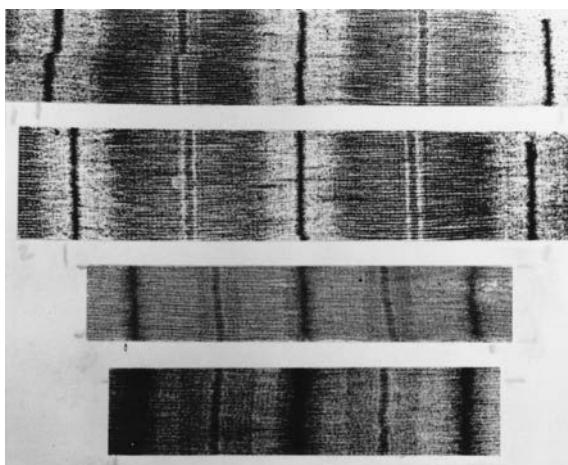
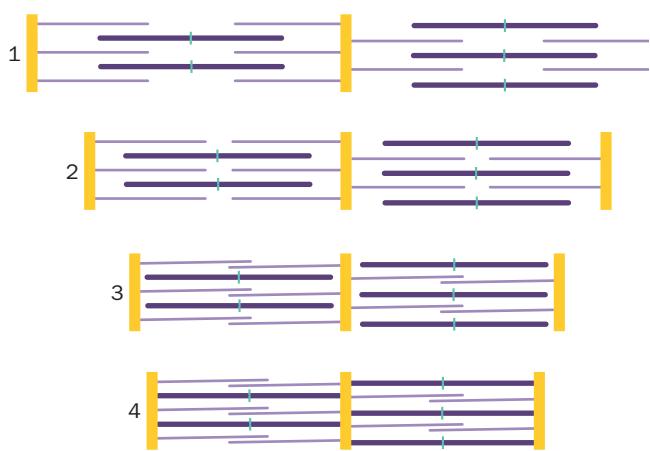


Figure 35-70 Myofibril contraction. (a) Electron micrographs showing myofibrils in progressively more contracted states. The lengths of the I band and H zone decrease on contraction, whereas the lengths of the thick and thin

(b)



filaments remain constant. (b) Interpretive drawings showing interpenetrating sets of thick and thin filaments sliding past each other. [Courtesy of Hugh Huxley, Brandeis University.]

edge of the adjacent H zone do not change. Rather, sarcomere contraction is accompanied by equal reductions in the widths of the I band and the H zone. These observations were independently explained by Hugh Huxley and Jean Hanson and by Andrew Huxley and Rolf Niedergerke who, in 1954, proposed the **sliding filament model**: *The force of muscle contraction is generated by a process in which interdigitated sets of thick and thin filaments slide past each other (Fig. 35-70).*

b. Actin Stimulates Myosin's ATPase Activity

The sliding filament model partially explains the mechanics of muscle contraction but not the origin of the contractile force. Albert Szent-Györgyi's work in the 1940s pointed the way toward the elucidation of the contraction mechanism. The mixing of solutions of actin and myosin to form a complex known as **actomyosin** is accompanied by a large increase in the solution's viscosity. This viscosity increase is reversed, however, when ATP is added to the actomyosin solution. *Evidently, ATP reduces myosin's affinity for actin.*

Further insight into the role of ATP in muscle contraction was provided by kinetic studies. Isolated myosin's ATPase function has a turnover number of $\sim 0.05 \text{ s}^{-1}$, far less than that in contracting muscle. Paradoxically, however, the presence of actin increases myosin's ATP hydrolysis rate to the physiologically more realistic turnover number of $\sim 10 \text{ s}^{-1}$, a rate enhancement of ~ 200 (indeed, actin was so named because it *activates* myosin). This is because isolated myosin rapidly hydrolyzes ATP



but only slowly releases the products ADP + P_i as is indicated by the observation that myosin-catalyzed ATP hydrolysis begins with a rapid burst of H^+ , whereas free ADP and P_i appear much more slowly. Actin enhances myosin's

ATPase activity by binding to the myosin-ADP- P_i complex and stimulating it to sequentially release P_i followed by ADP. The myosin-ADP- P_i complex cannot be formed by simply mixing myosin, ADP, and P_i , which suggests that this complex is a "high-energy" intermediate in which the free energy of ATP hydrolysis has somehow been conserved.

The foregoing observations led Edwin Taylor to formulate a model for actomyosin-mediated ATP hydrolysis (Fig. 35-71):

Step 1 ATP binding to the myosin component of actomyosin results in the dissociation of actin and myosin.

Step 2 The myosin-bound ATP is rapidly hydrolyzed to form a stable "high-energy" myosin-ADP- P_i complex.

Step 3 Actin binds to the myosin-ADP- P_i complex.

Step 4 In a process accompanied by a conformational relaxation to its resting state, the actin-myosin-ADP- P_i complex sequentially releases P_i followed by ADP, yielding actomyosin that can undergo another round of ATP hydrolysis.

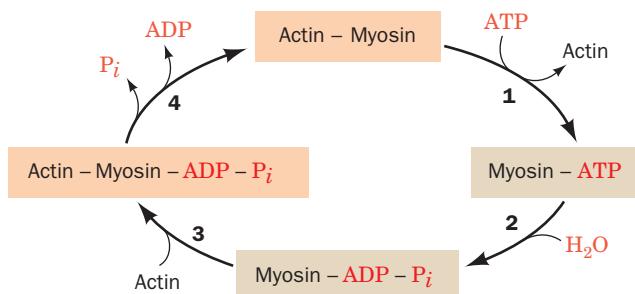


Figure 35-71 Reaction sequence in actomyosin-catalyzed ATP hydrolysis.

This ATP-driven alternate binding and release of actin by myosin provides, as we shall see, the vectorial force of muscle contraction.

c. A Structure-Based Model for the Interaction of Actin and Myosin

Rayment, Holden, and Ronald Milligan have formulated a model for the so-called **rigor complex** of the myosin S1 head and F-actin (Fig. 35-72; the rigor complex is that taken up by ATP-deprived muscle, which occurs in **rigor mortis**, the temporary rigidity of muscles after death). This was done, with an estimated accuracy of 5 to 8 Å, by fitting the X-ray structure of the myosin S1 head (Fig. 35-62) and the structure of F-actin as derived from the X-ray structure of G-actin (Fig. 35-68) to the electron density map obtained from the electron microscopy-based image of this complex (Fig. 35-64). The bulky motor domain of S1 binds tangentially to the actin filament at an $\sim 45^\circ$ angle to the filament axis. Its extension, the narrow S1 tail, which contains the two light chain binding regions, projects tangentially away from the filament axis at $\sim 90^\circ$, an orientation that permits S1 to impose tension on the rodlike myosin tail that associates with other such tails to form the thick filament. The myosin head appears to interact with actin via ion pairing involving several Lys residues on myosin and several Asp and Glu residues on actin. These interactions are bolstered by what appears to be a stereospecific association between juxtaposed surface-exposed patches of hydrophobic residues on actin and myosin.

d. Myosin Heads “Walk” along Actin Filaments

In order to complete our description of muscle contraction, we must determine how ATP hydrolysis is coupled to the sliding filament model. If the sliding filament model is correct, then it would be impossible for a myosin cross-bridge to remain attached to the same point on a thin filament during muscle contraction. Rather, it must repeatedly detach and then reattach itself at a new site further along the thin filament toward the Z disk. This, in turn, suggests that *muscular tension is generated through the interaction of myosin cross-bridges with thin filaments*.

The X-ray structure of myosin S1 (Fig. 35-62) together with the model of its rigor complex with actin (Fig. 35-72) suggests how ATP hydrolysis is coupled to myosin’s conformational change. Despite the excellent fit of the structures of actin and myosin S1 to the image of their rigor complex (Fig. 35-64b), the resulting atomic model of the rigor complex (Fig. 35-72) contains a steric clash between residues at the actin–myosin contact region. However, it seems quite plausible that, on ATP binding, this clash is relieved by the opening of the actin cleft, the cleft that divides the 50-kD segment into two domains (Fig. 35-62). Rayment, Holden, and Milligan therefore postulated that the closure of the actin cleft under the impetus of the release of ADP from the nucleotide binding site is responsible for the conformational change that produces myosin’s “power stroke” in muscle’s contractile cycle. This has led to the following variation of the widely accepted “rowboat” model for the contractile cycle (Fig. 35-73):

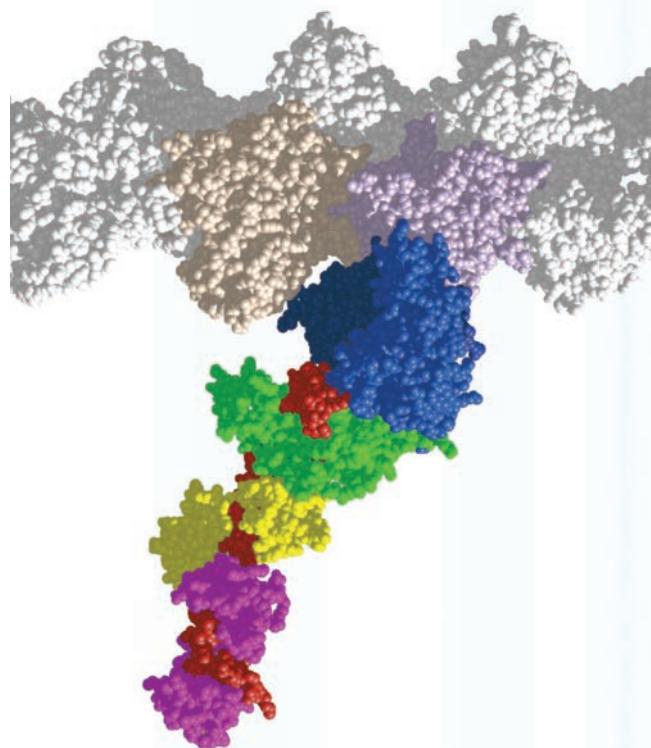


Figure 35-72 Model of the myosin–actin interaction. This space-filling atomic model was constructed by fitting atomic models of myosin S1 (Fig. 35-62) and the actin filament (Fig. 35-68) to the electron microscopy-based image of S1-decorated actin filaments (Fig. 35-64). The myosin S1 is rotated about its long axis relative to the view in Fig. 35-62, its ELC is yellow, and its RLC is magenta. Myosin’s coiled-coil tail is not shown. In a myofibril, every actin monomer has the potential to bind a myosin head, and the thick filament has many myosin heads projecting from it. [Modified from a drawing by Ivan Rayment and Hazel Holden, University of Wisconsin.]

1. ATP binds to the S1 head in a manner that opens up the actin cleft. This, in turn, causes S1 to release its bound actin.

2. The active site cleft (distinct from the actin cleft) closes about the ATP in a manner that catalyzes its hydrolysis. This process “cocks” the myosin molecule; that is, puts it into its “high energy” state in which its S1 head is approximately perpendicular to the thick filament.

3. The S1 head binds weakly to an actin monomer that is closer to the Z disk than the one to which it had been bound previously.

4. S1 releases P_i , which causes its actin cleft to close, thereby increasing S1’s binding affinity for actin.

5. The resulting transient state is immediately followed by the power stroke, a conformational shift that sweeps S1’s C-terminal tail by ~ 100 Å toward the Z disk relative to the motor domain, thus translating the attached thin filament by this distance toward the M disk. It seems likely that the ~ 85 -Å-long helix that connects S1’s motor domain

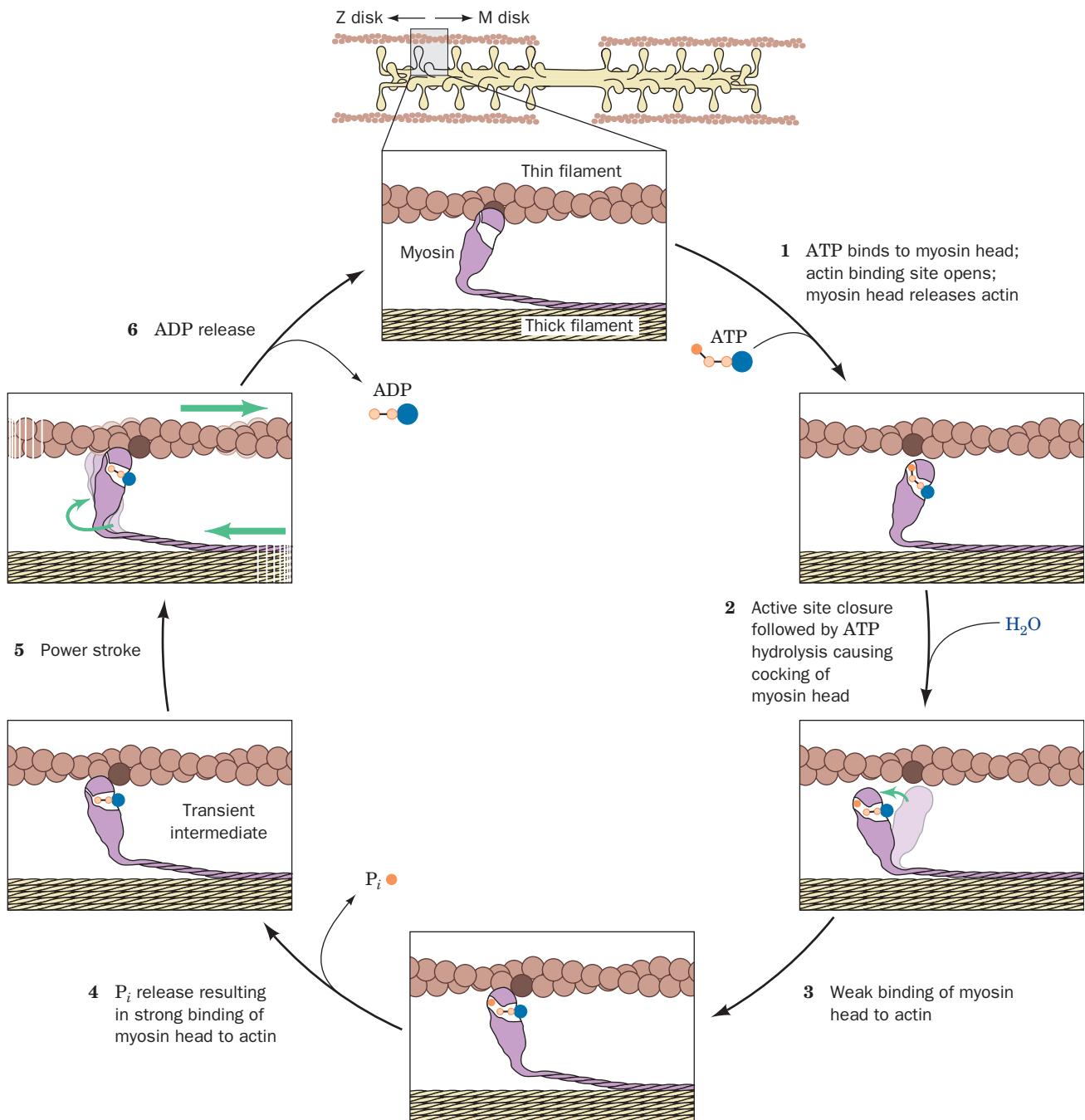


Figure 35-73 Mechanism of force generation in muscle. The myosin head “walks” up the actin thin filament through a vectorial process that is driven by ATP hydrolysis. Only one of myosin’s two independent S1 heads is shown. The narrow cleft that splits the 50-kD segment of the S1 head into two domains (Fig. 35-62) is represented by a horizontal gap perpendicular to

the thin filament (although in the actual atomic model, Fig. 35-72, this gap is inclined by $\sim 30^\circ$ to the thin filament, thereby obscuring it in that figure). The actin monomer to which S1 was bound at the beginning of the cycle is more darkly colored for reference. [After Rayment, I. and Holden, M., *Curr. Opin. Struct. Biol.* **3**, 949 (1993).]

to its C-terminal tail is the conformational coupler in this energy transduction step.

6. ADP is released, thereby completing the cycle. Note that actin functions here as a nucleotide exchange factor, thereby strongly enhancing myosin’s ATPase function.

The cyclic nature of this process is necessary to prevent this molecular motor from reversing its power stroke while still bound to the actin which, if it occurred, would result in no net movement of the myosin head relative to the actin filament. The ATP-driven active transport of ions across a

membrane is a similarly cyclic vectorial process (Sections 20-3A and 20-3B).

The ~500 S1 heads on every thick filament asynchronously cycle through this reaction sequence about five times per second each during a strong muscular contraction. The S1 heads thereby “walk” or “row” up adjacent thin filaments toward the Z disk with the concomitant contraction of the muscle. Nevertheless, myosin does not walk along an actin filament in an anthropomorphic manner. Rather, a detached myosin head undergoes a diffusional search ending in its reattachment to the actin filament, with the lever arm position of the attached head making it more likely that the detached head rebinds toward the (+) end of the actin filament.

The foregoing rowboat model is supported by the X-ray structures of scallop S1 in the absence of bound nucleotide and in complex with ADP and vanadate ion (VO_4^{3-}), an ADP + P_i mimic (VO_4^{3-} structurally resembles P_i), both determined by Anne Houdusse, Andrew Szent-Györgyi, and Cohen. The superposition of these two structures (Fig. 35-74)

indicates that ATP binding causes S1’s C-terminal “lever arm” to reorient relative to its motor domain as the model predicts.

e. Myosin Light Chains Function to Increase the Rate of Muscle Contraction

A motility assay has demonstrated that the rate at which myosin heavy chains slide along actin filaments is reduced 10-fold when the light chains are removed, even though removing the light chains does not significantly reduce myosin’s ATPase activity. It therefore appears that the 85-Å-long α helix to which both light chains bind (Fig. 35-62), and which forms the lever arm that amplifies the conformation change in myosin’s motor domain to produce the power stroke, is stabilized by its bound light chains. Indeed, a bare α helix is rarely observed in proteins in aqueous solution. Thus, in the absence of light chains, the long α helix is likely to collapse, leading to a smaller power stroke.

C. Control of Muscle Contraction

Striated muscles are, for the most part, under voluntary control; that is, their contraction is triggered by motor nerve impulses. How do these nerve impulses trigger muscle contraction? To answer this question, let us begin at the level of the myofibril and work up.

a. Ca^{2+} Regulates Muscle Contraction in a Process Mediated by Troponin and Tropomyosin

The involvement of Ca^{2+} ion in muscle contraction has been known since 1882, when the English physician Sydney Ringer found that isolated frog hearts would contract in saline solutions made with London tap water but not in saline solutions made with distilled water. Ringer determined that it was the Ca^{2+} ions in the tap water that triggered muscle contraction. It was not until the early 1960s, however, that Setsuro Ebashi demonstrated that the effect of Ca^{2+} is mediated by troponin and tropomyosin. He did so by showing that actomyosin extracted directly from muscle, and therefore bound to troponin and tropomyosin, contracts in the presence of ATP only when Ca^{2+} is also present, whereas actomyosin prepared from purified actin and myosin contracts in the presence of ATP regardless of the Ca^{2+} concentration. The addition of tropomyosin and troponin to the purified actomyosin system restored its sensitivity to Ca^{2+} . Indeed, it was through these experiments that troponin was discovered.

The TnC subunit of troponin (Fig. 35-66) is the only Ca^{2+} -binding component of the tropomyosin-troponin complex. In fact, the Ca^{2+} -binding affinity of TnC’s C-terminal lobe (C-lobe) is sufficiently high that it binds two Ca^{2+} ions under all physiological conditions. However, TnC’s N-terminal lobe (N-lobe) only binds Ca^{2+} ions when the $[\text{Ca}^{2+}]$ reaches that in a contracting myofibril (see below). Comparison of the X-ray structures of TnC in the presence and absence of Ca^{2+} ion indicates that Ca^{2+} binding induces the opening of a hydrophobic pocket in the N-lobe such that it binds the so-called switch segment of TnI (Fig. 35-66; note that the C-lobe similarly binds a portion of

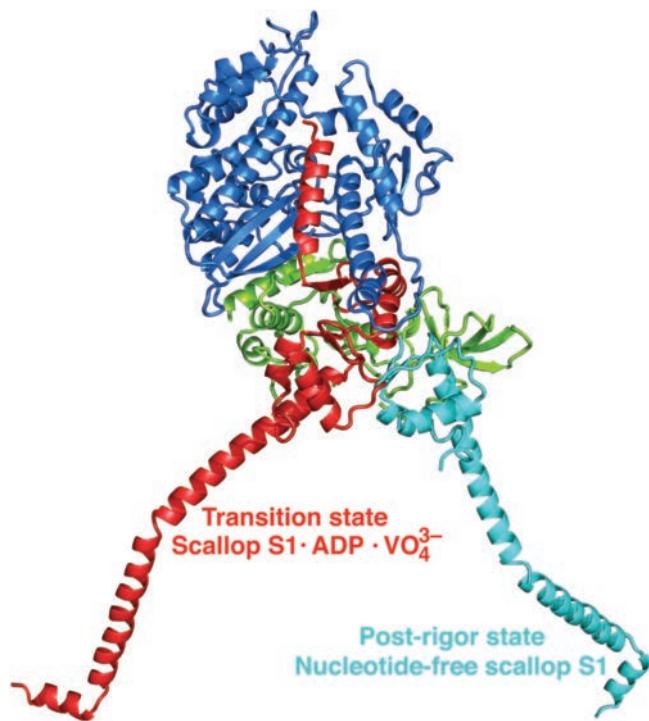


Figure 35-74 Conformational changes in scallop myosin S1.

Here the motor domains in the X-ray structures of S1 in its nucleotide-free state (the post-rigor state; previously called the near-rigor state) and in complex with $\text{ADP} \cdot \text{VO}_4^{3-}$ (the transition state) have been superimposed. The motor domain is oriented and colored similarly to that in Fig. 35-62 (which is also in the post-rigor state), the C-terminal lever arm in nucleotide-free S1 is cyan, and that in the $\text{ADP} \cdot \text{VO}_4^{3-}$ complex is red. The light chains bound to the lever arms in both structures have been deleted for clarity (but are drawn in Fig. 35-62). The orientations of the lever arms in the two states differ by ~60°. [Based on X-ray structures by Anne Houdusse, Andrew Szent-Györgyi, and Carolyn Cohen, Brandeis University. PDBIDs 1DFK and 1DFL.]

the N-terminal helix of TnI). This binding, it is postulated, extracts TnI's so-called inhibitory segment from its binding site on actin.

Tropomyosin, as we saw, binds along the thin filament groove (Fig. 35-64b). Comparisons of cryoelectron microscopy-based images of thin filaments in the absence and presence of Ca^{2+} ion reveal that Ca^{2+} ion causes tropomyosin to move $\sim 10 \text{ \AA}$ deeper into the thin filament groove (Fig. 35-75). X-ray diffraction studies indicate that this motion is triggered by an allosteric interaction between Ca^{2+} -troponin and tropomyosin, presumably the removal of the TnI inhibitory segment from its binding site on actin. *In the absence of Ca^{2+} , the sites on actin to which the myosin S1 heads bind are occluded by tropomyosin (Fig. 35-76a). However, the Ca^{2+} -induced movement of tropomyosin exposes these sites, thereby switching on muscle contraction (Fig. 35-76b).* This switching mechanism appears to be cooperative in that the binding of a myosin

head to an actin subunit pushes tropomyosin away from neighboring myosin-binding sites in a conformational change that could also increase the Ca^{2+} -binding affinity of its associated TnC N-lobe.

b. Nerve Impulses Release Ca^{2+} from the Sarcoplasmic Reticulum

In order to understand how a nerve impulse affects the $[\text{Ca}^{2+}]$ in a myofibril we must further consider the anatomy of striated muscle fibers. A nerve impulse arriving at a **neuromuscular junction** is transmitted directly to each sarcomere by a system of **transverse** or **T tubules**, nervelike infoldings of the muscle fiber's plasma membrane that surround each myofibril at its Z disk (Fig. 35-77; neurotransmission is discussed Section 20-5). All of a muscle's sarcomeres therefore receive the signal to contract within a few milliseconds of each other so that the muscle contracts as a unit. The electrical signal is transferred, in a poorly understood manner, to the **sarcoplasmic reticulum (SR)**, a system of flattened membranous vesicles derived from the endoplasmic reticulum that

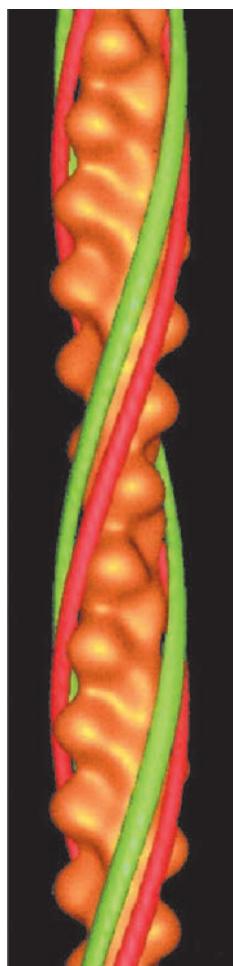


Figure 35-75 Comparison of the positions of tropomyosin on the thin filament in the absence and presence of Ca^{2+} . In this superposition of cryo-EM-based images, the F-actin filament is gold, the tropomyosin in the absence of Ca^{2+} is red, and that in the presence of Ca^{2+} is green. [Courtesy of William Lehman, Boston University School of Medicine.]

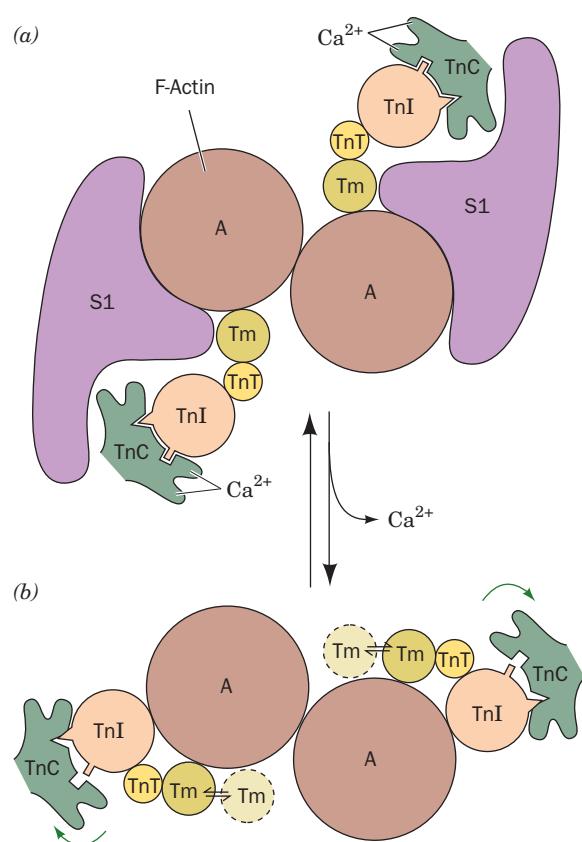


Figure 35-76 Control of skeletal muscle contraction by troponin and tropomyosin. (a) In contracting muscle, here diagrammed in cross section, the myosin S1 heads freely interact with and thereby “walk” up the F-actin filaments (A). (b) Muscle relaxes when Ca^{2+} dissociates from troponin's TnC subunit, thereby allosterically moving the tropomyosin (Tm) molecules to positions which sterically block myosin–actin interactions. [After Zot, A.S. and Potter, J.D., *Annu. Rev. Biophys. Biophys. Chem.* **16**, 555 (1987).]

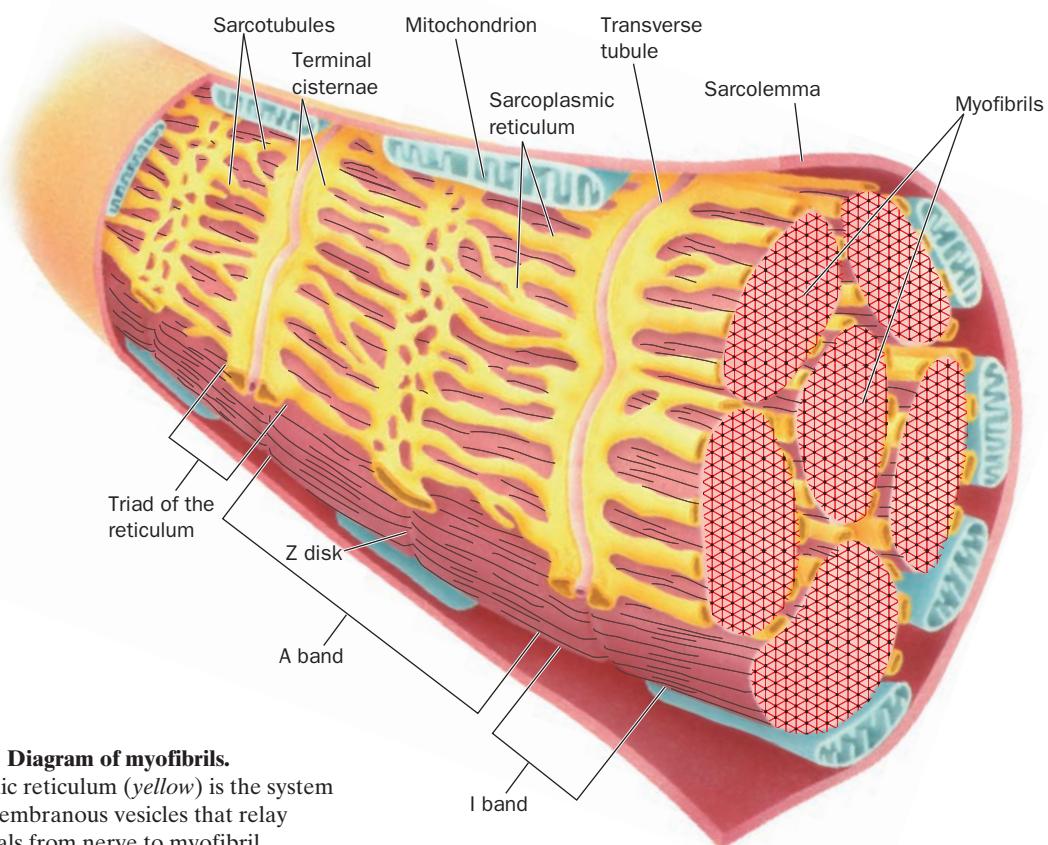
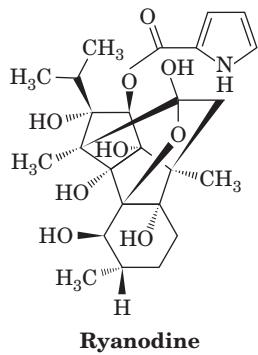


Figure 35-77 Diagram of myofibrils.

The sarcoplasmic reticulum (yellow) is the system of connected membranous vesicles that relay contractile signals from nerve to myofibril.

surround each myofibril rather like a net stocking. Around 80% of the protein in the SR membrane is a transmembrane Ca^{2+} -ATPase (Section 20-3B) that continually pumps Ca^{2+} into the SR so as to maintain the cytosolic $[\text{Ca}^{2+}]$ of resting muscle below 10^{-7} M , whereas that in the SR is over 10^{-3} M . The SR's ability to store Ca^{2+} is enhanced by the presence of a highly acidic (37% Asp + Glu) 55-kD protein named **calsequestrin**, which has $>40 \text{ Ca}^{2+}$ -binding sites.

The arriving nerve impulse opens voltage-gated Ca^{2+} channels known as **ryanodine receptors** (so named because they specifically bind ryanodine),



an alkaloid produced by the plant *Ryania speciosa*) in the otherwise Ca^{2+} -impermeable SR membrane. This causes the myofibril's internal $[\text{Ca}^{2+}]$ to increase to $\sim 10^{-5} \text{ M}$ within a few milliseconds, which in turn triggers the conformational change in troponin–tropomyosin that induces muscle contraction. Once nerve excitation has subsided,

the ryanodine receptors close so that the Ca^{2+} inside the myofibril is pumped back into the SR (the Ca^{2+} -ATPase has a greater affinity for Ca^{2+} ion than does TnC's N-lobe). Tropomyosin therefore resumes its resting conformation, causing the muscle to relax.

D. Smooth Muscle

Vertebrates have two major types of muscle besides skeletal muscle: **cardiac muscle** and **smooth muscle**. Cardiac muscle, which is responsible for the heart's pumping action, is striated, indicating the similarity of its organization to that of skeletal muscle. However, cardiac muscle cells are much shorter than those of skeletal muscle and contain only a single nucleus. In addition, cardiac and skeletal muscle differ in their metabolism, with cardiac muscle, which must function continuously for a lifetime, being much more dependent on aerobic metabolism than is skeletal muscle. Vertebrate heart muscle contraction is spontaneously initiated by the heart muscle itself rather than through external nervous stimuli, although the nervous system can influence this contractile response. Smooth muscle, which is responsible for the slow, long-lasting, and involuntary contractions of large blood vessels and hollow organs such as the intestines, bronchi, and uterus, has a quite different organization from that of striated muscle. Smooth muscle consists of spindle-shaped, mononucleated cells whose thick and thin filaments are more or less aligned along the cells' long axes but which do not form myofibrils.

Smooth muscle myosin, a genetically distinct protein, is functionally distinct from striated muscle myosin in several ways:

1. Its maximum ATPase activity is only $\sim 10\%$ of that of striated muscle.
2. It interacts with actin only when its regulatory light chain (RLC) is phosphorylated at a specific Ser residue.
3. It forms thick filaments whose cross-bridges lack the regular repeating pattern of striated muscle and are distributed along the thick filament's entire length.

a. Smooth Muscle Contraction Is Triggered by Ca^{2+}

The thin filaments of smooth muscle contain actin and tropomyosin but lack troponin. *Smooth muscle contraction is nevertheless triggered by Ca^{2+} because myosin*

light chain kinase (MLCK), the enzyme that phosphorylates the RLC and thereby stimulates smooth muscle to contract, is enzymatically active only when it is associated with Ca^{2+} -calmodulin (Fig. 35-78, bottom; the phosphorylation of skeletal muscle RLC, a different isoform from that in smooth muscle, appears to modulate the degree of tension produced by contraction). The phosphorylation of the RLC presumably causes a conformational change that, in a poorly understood way, is conveyed to the motor domain of the myosin S1 head. The mechanism whereby Ca^{2+} -CaM activates MLCK is discussed in Section 18-3Ce.

The intracellular $[\text{Ca}^{2+}]$ in a smooth muscle cell varies with the permeability of its plasma membrane to Ca^{2+} , which in turn is under the control of the autonomic (involuntary) nervous system. When the $[\text{Ca}^{2+}]$ rises to $\sim 10^{-5} \text{ M}$,

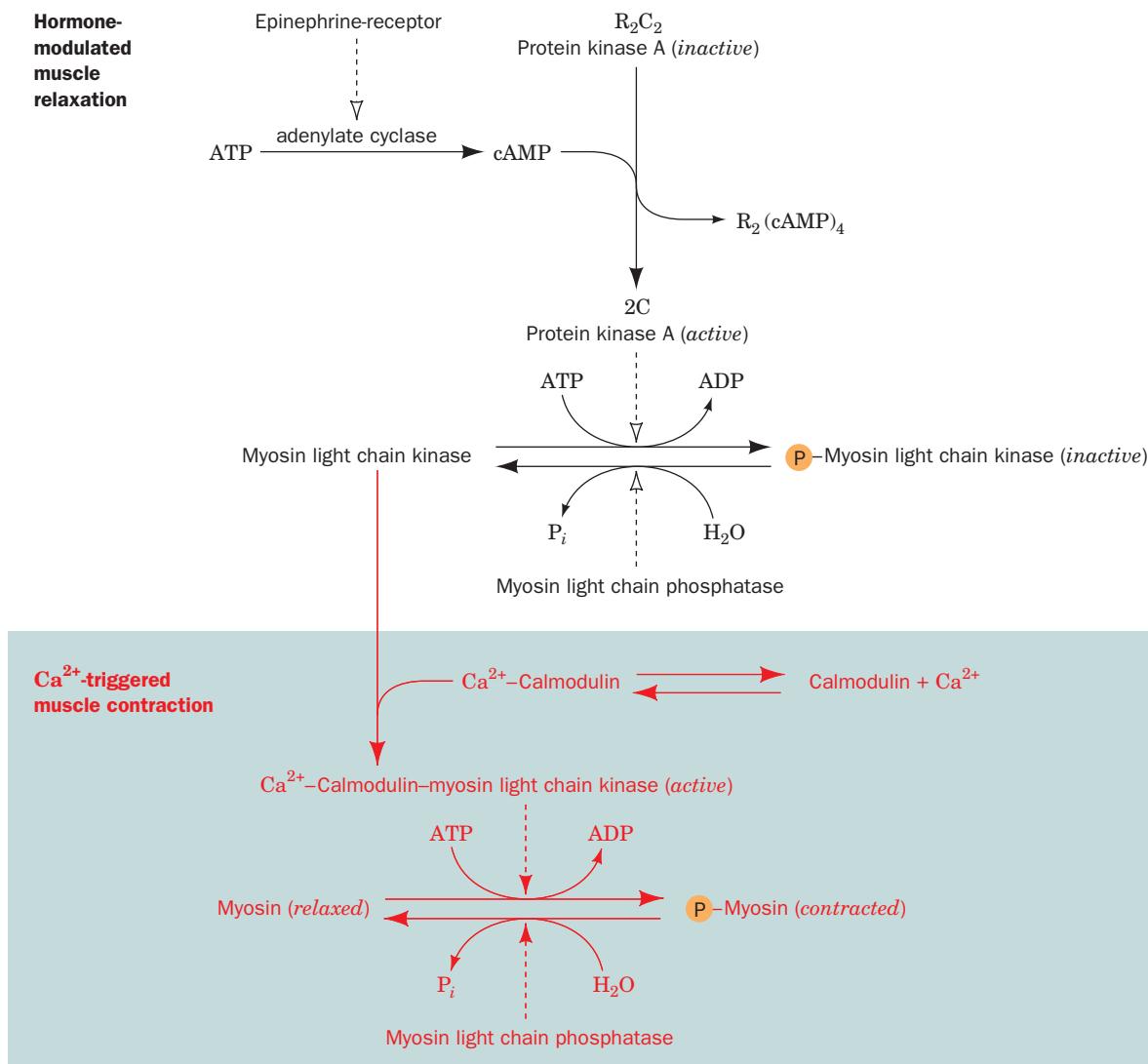


Figure 35-78 Control of smooth muscle contraction. Dashed arrows indicate stimulation or catalysis of a modification or demodification reaction. The lower part of the diagram (shaded) indicates how Ca^{2+} ion, whose intracellular concentration increases in response to nerve impulses, triggers muscle

contraction. Activated myosin light chain kinase (MLCK) phosphorylates myosin's RLC subunit, thereby triggering smooth muscle contraction. The upper part of the diagram indicates how hormones such as epinephrine inhibit the contractile response, causing smooth muscle relaxation.

smooth muscle contraction is initiated as described. When the $[Ca^{2+}]$ falls to $\sim 10^{-7}$ M through the action of the plasma membrane's Ca^{2+} -ATPase, the MLCK is deactivated, the RLC is dephosphorylated by **myosin light chain phosphatase**, and muscle relaxation ensues. *Thus, Ca^{2+} , like cAMP, is a second messenger that transmits extracellular signals within the interior of a cell. In the many situations in which Ca^{2+} is a second messenger, calmodulin or a calmodulin-like protein is the intracellular signal receiver.*

b. Smooth Muscle Activity Is Hormonally Modulated

Smooth muscles also respond to hormones such as epinephrine (Fig. 35-78, top). The binding of epinephrine to its plasma membrane-bound α -adrenergic receptor (Section 19-1F) activates adenylate cyclase. The cytosolic cAMP that is thereby generated binds to and thereby activates protein kinase A (PKA; Section 18-3Cb), which then phosphorylates MLCK. Phosphorylated MLCK binds Ca^{2+} -calmodulin only weakly, so the extracellular presence of epinephrine causes smooth muscles to relax. Note the resemblance of this system to that controlling glycogen metabolism in skeletal muscle (Section 18-3C). **Asthma**, a breathing disorder caused by the inappropriate contraction of bronchial smooth muscle, is often treated by the inhalation of an aerosol containing epinephrine, thereby relaxing the contracted bronchi.

Nitric oxide (NO), as we have seen (Section 19-1L), also induces the relaxation of smooth muscle. The binding of NO to guanylate cyclase stimulates this enzyme to synthesize cGMP which, in turn, binds to and thereby activates cGMP-dependent protein kinase to phosphorylate and thereby activate myosin light chain phosphatase.

The sequence of events culminating in smooth muscle contraction is inherently much slower than those leading to skeletal muscle contraction. Indeed, the structure and regulatory apparatus of smooth muscle suits it to its function: the maintenance of tension for prolonged periods while consuming ATP at a much lower rate than skeletal muscle performing the same task. The structural and functional resemblance of TnC to CaM therefore suggests that TnC is

a CaM variant that has evolved in skeletal muscle to provide a rapid response to the presence of Ca^{2+} .

E. Actins in Nonmuscle Cells

Although actin and myosin are most prominent in muscle, they also occur in other tissues. In fact, actin is ubiquitous and is usually the most abundant cytoplasmic protein in eukaryotic cells, typically comprising 5 to 10% of their total protein. Myosin, in contrast, is usually present in only about one-tenth the quantity of actin. This ratio reflects the fact that actin, in addition to its role in actomyosin-based contractile systems, participates in several myosin-independent motility systems as well as being a principal cytoskeleton component.

a. Actin Forms Microfilaments

Actin in muscles is entirely in the form of thin filaments. Nonmuscle actin, however, is about equally partitioned between soluble G-actin and F-actin fibers known as **microfilaments**. The actin content of microfilaments was established both through the immunofluorescence microscopy of living cells (e.g., Fig. 35-79) and because microfilaments can be decorated with S1 myosin heads to form arrowhead structures that are visually indistinguishable from those formed by muscle thin filaments (Fig. 35-64a). Such decoration occurs because the various actin isoforms are highly conserved (Section 35-3Ac).

Actin *in vitro* is monomeric at low temperatures, low ionic strengths, and alkaline pH's. However, under physiological conditions, G-actin polymerizes in a process that is accelerated by the presence of ATP. *In vitro*, the initiation of actin polymerization is a relatively slow process because actin dimers and trimers have relatively few subunit–subunit contacts and hence low stability. However, once a stable nucleus of ~ 4 subunits has formed, additional subunits can form the full complement of interactions and hence rapidly associate with both ends of a filament. As the aggregate mass of the F-actin filaments increases, the concentration of G-actin monomers decreases until equilibrium is

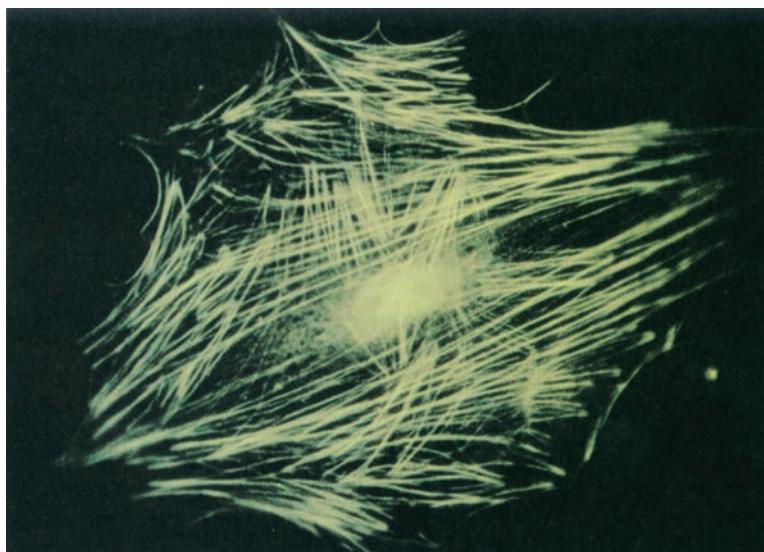
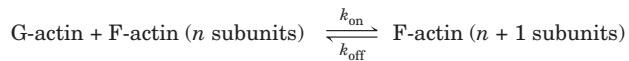


Figure 35-79 Microfilaments in a fibroblast resting on the surface of a culture dish as revealed by indirect immunofluorescence microscopy using anti-actin antibody. When the cell begins to move, the filaments disassemble to form a diffuse mesh, suggesting that actin plays a central role in cellular movement. [Courtesy of Elias Lazarides, California Institute of Technology.]

reached. At this stage, the F-actin subunits at the two ends of each filament exchange with G-actin monomers in solution but there is no change in the aggregate mass of F-actin filaments; that is, the filament lengths have reached a steady state.

The reaction for microfilament formation is expressed



where k_{on} is the rate constant for subunit association with filaments and k_{off} is the rate constant for subunit dissociation from filaments. The G-actin concentration at the steady state is called the **critical concentration**, C_c , which is typically $\sim 0.1 \mu\text{M}$. Thus, at equilibrium,

$$k_{\text{on}}C_c[\text{F-actin ends}] = k_{\text{off}}[\text{F-actin ends}] \quad [35-1]$$

so that

$$C_c = \frac{k_{\text{off}}}{k_{\text{on}}} \quad [35-2]$$

(here the F-actin ends are reactants for both the forward and reverse reactions and hence their concentration is the same for both reactions). Above the critical concentration, G-actin polymerizes; below it, F-actin depolymerizes.

The value of k_{on} is 5- to 10-fold greater for subunit addition to the (+) end (hence its name) than for subunit addition to the (−) end (the pointed end of S1-decorated filaments). Thus, according to Equation 35-2, k_{off} for subunit dissociation from the (+) end must be larger than that from the (−) end by the same factor so that at the steady state neither end experiences a net change in length. This makes physical sense because subunits added to either end

ultimately must have identical interactions with other subunits in the filament.

b. Microfilaments Undergo Treadmilling

The foregoing is predicated on the assumption that the G-actin subunits that associate with F-actin are identical to those that dissociate from F-actin. However, this is not necessarily the case. Polymerization activates F-actin subunits to hydrolyze their bound ATP to ADP + P_i with the subsequent dissociation of P_i . The resulting conformation change (Section 35-3Ad) reduces the affinity of an ADP-F-actin subunit for its neighboring subunits relative to that of ATP-F-actin; that is, the value of C_c for ADP-F-actin is greater than that for ATP-F-actin. Since F-actin-catalyzed ATP hydrolysis occurs more slowly than actin polymerization and F-actin's bound nucleotide does not exchange with those in solution, F-actin's more recently polymerized and hence predominantly ATP-containing subunits occur largely at its (+) end, whereas its (−) end consists mainly of less recently polymerized and hence predominantly ADP-containing subunits.

Under these conditions, the steady state (when the microfilament maintains a constant length) occurs when the net rate of addition of subunits at the (+) end matches the net rate of dissociation of subunits at the (−) end; that is, the G-actin concentration falls between the C_c values for the two ends. Then, *subunits that have added to the (+) end move toward the (−) end where they dissociate* (Fig. 35-80). This process, called **treadmilling**, is driven by the free energy of ATP hydrolysis and hence is not at equilibrium. Treadmilling has been directly observed *in vivo* by loading cells with a

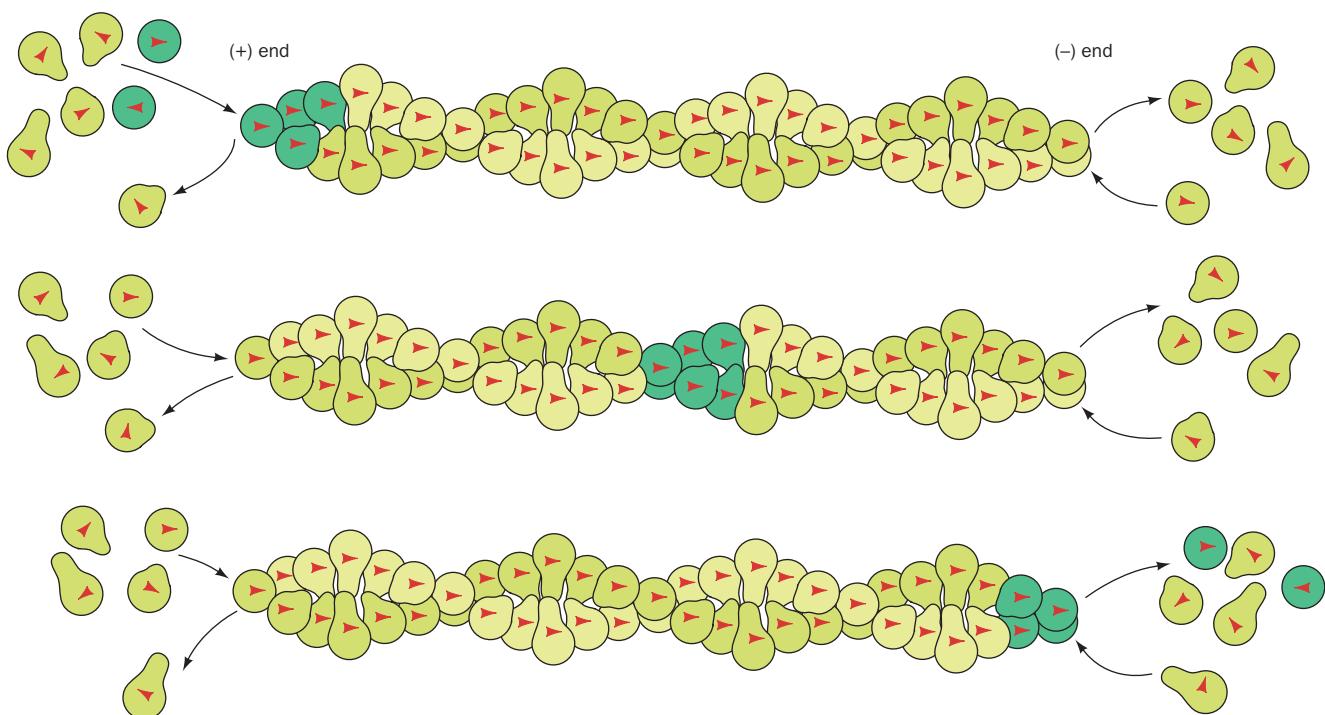


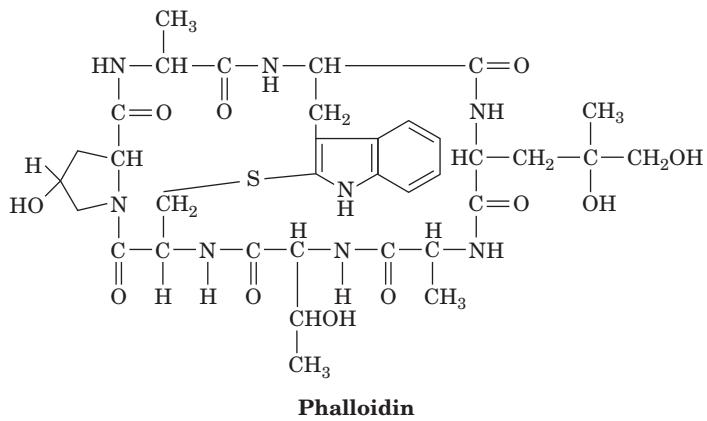
Figure 35-80 Microfilament treadmilling. In the steady state, actin monomers continually add to the (+) end of the filament (left), with eventual ATP hydrolysis, but dissociate at the same

rate from the (−) end (right). The filament thereby maintains a constant length while its component monomers translocate from left to right.

small amount of fluorescently labeled G-actin so that the newly formed F-actin contains just a few randomly located fluorescent subunits. Each microfilament is uniquely speckled and can therefore be distinguished from other microfilaments as it undergoes structural changes over time.

c. Microfilament Dynamics Is Responsible for Cell Motility Processes

There is clear evidence that the assembly and disassembly of actin filaments play an important role in such cellular motility processes as ameboid locomotion, phagocytosis, **cytokinesis** (the separation of daughter cells in the last stage of mitosis), and the extension and retraction of various cellular protuberances, such as **microvilli** (fingerlike projections of cell surfaces) and neuronal axons. This evidence was obtained largely through the use of drugs that interfere with actin aggregation. For example, the fungal alkaloid **cytochalasin B** (which we have seen inhibits Na^+ -independent glucose transport; Section 20-4Aa) blocks actin polymerization by specifically binding to the (+) end of a growing F-actin filament so as to inhibit actin polymerization from that end. In contrast, **phalloidin**,



a bicyclic heptapeptide produced by the poisonous mushroom *Amanita phalloides* (which also synthesizes the chemically similar eukaryotic RNA polymerase inhibitor α -amanitin; Section 31-2Eb), blocks microfilament depolymerization by specifically binding to its actin units.

In a resting cell, the rate of actin polymerization is low, and the cell maintains a steady state as described above. Following stimulation, however, the rates of actin polymerization and branching (see below) increase dramatically, allowing the cell to extend its cytoplasm in one direction. In this event, the (+) ends of growing microfilaments actually push the plasma membrane outward. If the cytoplasmic protrusion anchors itself to the underlying surface, then the cell can use the adhesion point for traction in order to advance in the direction of microfilament growth. In order to crawl, however, the trailing edge of the cell must release its contacts with the surface while newer contacts are being made at the leading edge (Fig. 35-81). In addition, as microfilament polymerization proceeds at the leading edge, depolymerization must occur elsewhere in the cell, because the pool of G-actin is limited.

Actin-mediated cell locomotion, that is, ameboid motion, is the most primitive mechanism of cell movement.

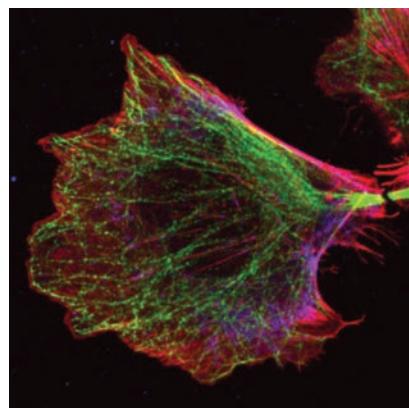


Figure 35-81 Fluorescence micrograph of a crawling cell.

The leading edge of the cell (left) is ruffled where it has become detached from the surface and is in the process of extending. The cell's trailing edge or tail (right), which is still attached to the surface, is gradually pulled toward the leading edge. Here actin microfilaments are red, microtubules (Section 35-3G) are green, and phosphorylated myosin light chains, which are implicated in cell migration, are blue. Note that the rate of actin polymerization is greatest at the leading edge. [Photo by Miguel Vicente-Manzanares and Rick Horwitz, University of Virginia.]

Nevertheless, virtually all eukaryotic cells undertake some version of it, even if it involves just a small patch of actin near the cell surface. More extensive microfilament rearrangements are essential for cells such as neutrophils (a type of white blood cell) that travel relatively long distances to sites of infection or inflammation.

d. Actin-Binding Proteins Control Microfilament Dynamics

At the pH and salt concentrations that typically occur in the cytosol, actin by itself would be almost entirely polymerized (mono- and divalent cations bind to low affinity sites on actin so as to promote its polymerization). This is not the case, however, because the rates of actin polymerization and depolymerization at each end of a microfilament are controlled by numerous other proteins. For example, **β -thymosin** (so named because it was originally isolated from mammalian thymus), a small (43 residues) and abundant protein, binds to G-actin so as to inhibit both its polymerization and its nucleotide exchange. In contrast, **profilin** (139 residues) binds to ADP-G-actin (i.e., actin that has just been released from a depolymerizing microfilament) so as to increase the rate of exchange of its bound ADP for ATP by over 1000-fold, thereby increasing the concentration of ATP-G-actin (the substrate for repolymerization). Moreover, profilin binds to the end of ATP-G-actin opposite its ATP-binding cleft, thereby blocking actin subunits from binding to the microfilament's (-) end but permitting them to associate with its (+) end. Since on binding to the (+) end, an actin subunit undergoes a conformational change that releases profilin, this greatly enhances the rate of microfilament elongation. The activity of profilin is regulated by several mechanisms including its phosphorylation and its binding to the membrane component phosphatidylinositol-4,5-bisphosphate

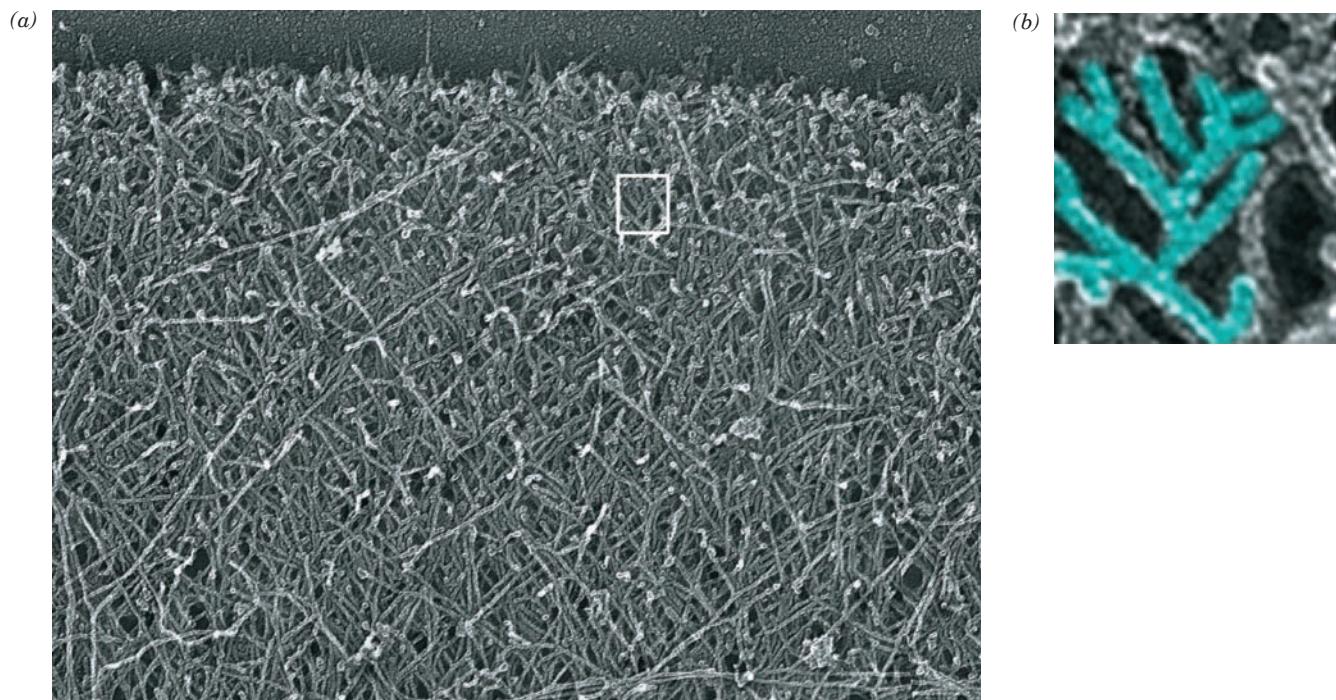


Figure 35-82 Electron micrograph of branched microfilaments. (a) The leading edge of a migrating *Xenopus* keratinocyte (a type of skin cell) prepared by detergent extraction. (b) Magnified view of the boxed region in Part a in

which tightly spaced multiple branches are highlighted in cyan. Note that new microfilaments branch from the parent filament at a constant angle of 70°. [Courtesy of Tatyana M. Svitkina, University of Pennsylvania.]

(PIP₂), a central participant in the phosphoinositide signal transduction pathway (Section 19-4A).

ADF/cofilin (ADF for actin depolymerizing factor) accelerates depolymerization at the (−) end of a microfilament by preferentially binding to ADP-P_i-actin and promoting the dissociation of P_i. However, the binding of ADF/cofilin to F-actin is prevented by the phosphorylation of ADF/cofilin by a specific protein kinase, **LIM kinase**, which is itself activated by phosphorylation signaling pathways involving Rho family GTPases (Section 19-3Cf). *In vitro*, ADF/cofilin and profilin together accelerate microfilament treadmilling by a factor of ~125, which approaches the rate observed in crawling cells.

e. Microfilament Branching Is Mediated by Arp2/3

Signals that lead to actin polymerization also activate the 7-subunit, ~220-kD **Arp2/3 complex** (Arp for actin related protein). This protein complex initiates microfilament polymerization at random sites along a microfilament so as to form branches (Fig. 35-82). Indeed, actin filaments may be very densely branched, particularly near the periphery of that portion of a crawling cell that is being extended.

The X-ray structure of the Arp2/3 complex, determined by Thomas Pollard and Senyon Choe, reveals that **Arp2** and **Arp3**, which are both ~45% identical to actin, structurally resemble actin (Fig. 35-83). In the inactive Arp2/3 complex, the two Arp subunits are arranged head to tail but are not aligned as in an actin filament. The binding of the Arp2/3 complex to the side of an actin filament, as mediated by members of the **WASP/WAVE** protein

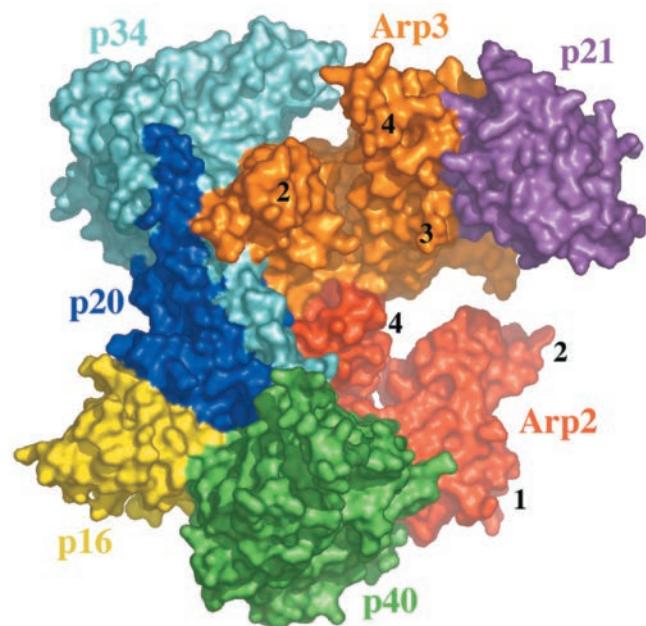


Figure 35-83 X-ray structure of the bovine Arp2/3 complex. In this surface diagram, the subunits are drawn in different colors. Arp2 (red) and Arp3 (orange) structurally resemble actin (their subdomains are numbered; compare with Fig. 35-67) but are not aligned for polymerization. Activation of the complex causes it to bind to an existing microfilament and aligns the Arp2 and Arp3 subunits so that they can initiate the growth of a branch. [Courtesy of Thomas Pollard, Yale University. PDBid 1K8K.]

family (WASP for *Wiskott–Aldrich syndrome protein*; **Wiskott–Aldrich syndrome** is an X-linked immunodeficiency and bleeding disorder arising from the expression of defective WASP in lymphocytes and platelets), induces conformational changes in the Arp2/3 complex that rotate its Arp subunits so that they more closely resemble an actin dimer. The Arp subunits thereby nucleate (initiate) the growth of an F-actin branch by providing a start site for the polymerization of actin subunits. The Rho family GTPases **Cdc42** and **Rac** are implicated in activating WASP/WAVE proteins, although how they do so is unclear.

f. Capping and Severing Proteins Limit Microfilament Growth

In a crawling cell, actin polymerization occurs almost exclusively at microfilament (+) ends near the cell's leading edge because the ends of microfilaments elsewhere in the cell are blocked by **capping proteins**. In effect, all the available G-actin is funneled to the uncapped ends, resulting in rapid microfilament growth. The balance between G-actin and F-actin strongly favors polymerization, so the presence of capping proteins is essential for maintaining normal

microfilament dynamics. CapZ and tropomodulin, as we have seen (Section 35-3Ae), are capping proteins that help control the lengths of thin filaments in skeletal muscle.

In some cases, capping proteins also act to sever microfilaments. One such protein is **gelsolin**, which is activated by increases in intracellular Ca^{2+} concentrations. Gelsolin consists of six homologous 125- to 150-residue segments, three of which can bind to actin. The X-ray structure of an actin trimer in which each monomer is bound to a gelsolin segment 1 molecule, determined by James Spudich and Robert Fletterick, has the same general shape as F-actin but is untwisted and extended (Fig. 35-84). This form of actin has been dubbed **X-actin** (X for extended). The structure suggests that gelsolin severs a microfilament by inserting between polymerized actin subunits and altering the shape of the polymer so that intersubunit interactions are weakened. Actin subunits then dissociate from the (–) end of the newly formed microfilament ends, with gelsolin remaining associated with the (+) end as a cap. However, a local rise in the PIP_2 concentration induces gelsolin to dissociate from the filament end. ADF/cofilin may act in a similar fashion as a capping and severing protein.

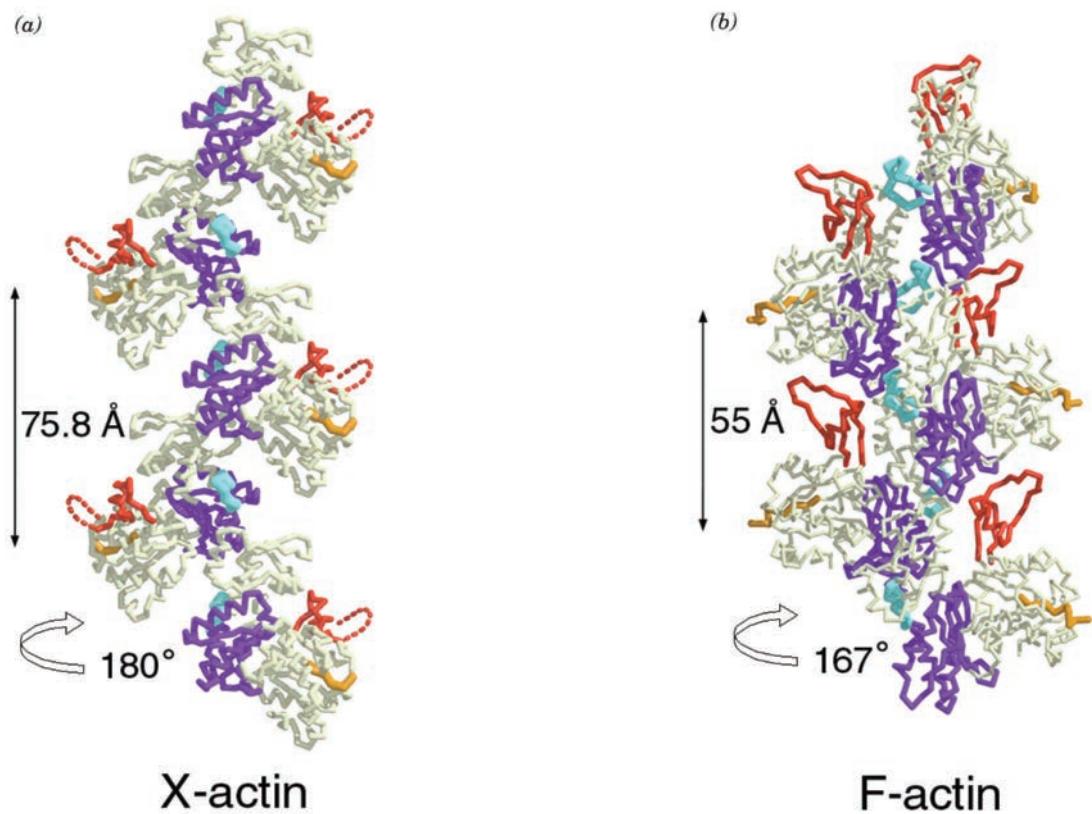


Figure 35-84 Structures of X-actin and F-actin. (a) The model of X-actin is based on the X-ray structure of an actin trimer bound to the 125-residue segment 1 of gelsolin (not shown; PDBid 1MDU). Subdomain 2 is red, subdomain 3 is purple, a hydrophobic loop (residues 263–275) is cyan, and the N-terminus is yellow. The gelsolin segment inserts between actin subunits so as to extend and untwist the actin filament, thereby

promoting its breakage. (b) Model of F-actin colored as in Part a and viewed similarly to Fig. 35-68. The rotational symmetry and axial translation per subunit are indicated. Note that each model shows only one strand of the double-stranded actin polymer (Fig. 35-68). [Courtesy of Robert Fletterick, University of California at San Francisco.]

F. Myosins in Nonmuscle Cells

Myosins have been classified, according to their sequences, into two broad categories, **conventional myosins** (or **type II myosins**), which were first identified in muscle tissue (and which we have heretofore referred to as just “myosin”), and **unconventional myosins**. There are 17 classes of unconventional myosins (types I and III–XVIII), some of which are widely expressed in eukaryotes and others that are expressed only in certain organisms. For example, **myosin X** occurs only in vertebrates, whereas **myosins VIII, XI, and XIII** occur only in plants. The human genome contains the genes for 40 different myosins, which are collectively members of 12 of the 18 myosin classes.

The myosins of most classes are homodimers in which the two ATP-hydrolyzing heads are held together by a coiled coil stalk, as we saw to be the case for myosin II (Fig. 35-59). However, the myosins of several classes function as monomers and therefore have only one head. Indeed, **myosin I**, the first of the monomeric myosins to be discovered, was so named because it has one head. Conventional myosin was then renamed myosin II to indicate that it has two heads (the remaining classes of myosins were named in the approximate order of their discovery).

The most conserved feature among the members of the myosin superfamily is the motor domain, which always occurs at the proteins’ N-termini. All two-headed myosins have a light chain-binding region followed by a coiled coil region responsible for their dimerization. However, only class II myosins have a coiled coil that promotes filament formation. The number of light chain binding sites, which are all IQ motifs (Section 35-3Ab), varies from one to six per myosin heavy chain among the various myosin classes and even varies within classes. In unconventional myosins, most of these sites bind Ca^{2+} –calmodulin (Ca^{2+} –CaM) but there are several instances of their binding myosin II light chains (which, recall, are homologs of CaM). For example, each chicken **myosin V** heavy chain binds two myosin light chains and four CaMs and hence the entire homodimeric assembly consists of 14 subunits. As we saw to be the case for myosin II (Fig. 35-62), the IQ motifs with their associated Ca^{2+} –CaM and light chains function as lever arms for the associated motor domain and hence altering the length of a lever arm will likely change the step size of the myosin movement.

Nearly all members of the myosin superfamily move toward the (+) ends of their associated actin filaments. However, the two-headed **myosin VI** travels in the opposite direction. This occurs because the lever arms of myosin VI rotate in the direction opposite to those of other myosins, apparently due to an insert in its motor domain that is unique to myosin VI.

a. Myosin II in Nonmuscle Cells Has Contractile Functions

The myosin II in nonmuscle cells forms thick filaments that participate in contractile processes with microfilaments, much as occurs in muscle. One of the best characterized such processes occurs during cytokinesis (cytoplasmic division) in animal cells and protozoa. In the final stages of mitosis, a

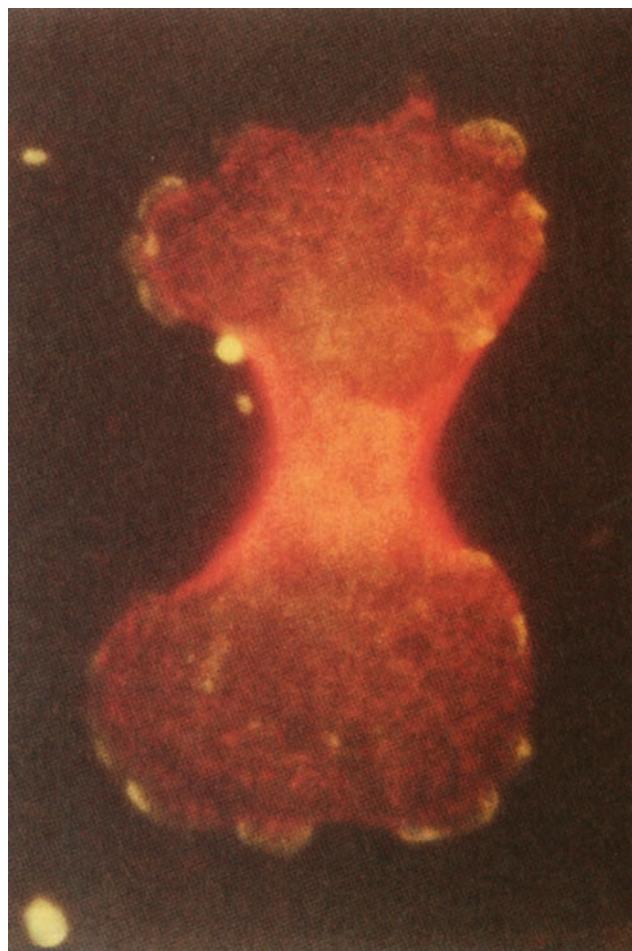


Figure 35-85 Indirect immunofluorescence micrograph of a dividing *Dictyostelium* (slime mold) amoeba. The cell was stained with anti-myosin II antibodies (red), thus demonstrating that its contractile ring contains myosin II. The cell was also treated with anti-myosin I antibodies (green), thereby showing that myosin I is localized at the leading edges of the daughter cells’ **lamellipodia** (sheetlike extensions of the cell surface that participate in cell locomotion). [Courtesy of Edward Korn, National Institutes of Health, and Yoshio Fukui, Northwestern University Medical School.]

cleavage furrow forms around the equator of the dividing cell in the plane perpendicular to the long axis of the mitotic spindle. Immunofluorescence microscopy demonstrates that the cleavage furrow is lined with an actomyosin belt (Fig. 35-85). Cytokinesis is accomplished through the tightening of this so-called **contractile ring**, which disperses once cleavage has occurred. Blood platelets also contain actomyosin, which on blood clot formation (Section 35-1), contracts so as to strengthen the clot. The contraction is initiated by the Ca^{2+} –CaM activation of MLCK as occurs in smooth muscle (Section 35-3Da).

b. Myosin V Moves via a Hand-over-Hand Mechanism

Many two-headed unconventional myosins function to transport cargo such as vesicles along microfilament tracks

from one site in a cell to another. For example, in mammals, a myosin V isoform named **myosin Va** acts to transport **melanosomes** (pigment granules) into the spiny protrusions of **melanocytes** (pigment cells), from which they are transferred to hair follicles for incorporation into a growing hair. Indeed, myosin Va was discovered as the product of a mutant gene in mice with reduced hair color. Humans with mutant myosin Va suffer from **Griscelli syndrome**, a rare autosomal recessive condition that is characterized by partial albinism (reduced skin coloration and silvery gray hair) together with severe neurological impairment, both resulting from defective vesicle transport.

The ATP-dependent movement of myosin V along a microfilament track is highly processive; that is, myosin V rarely dissociates from the microfilament during the transport process. Evidently, at least one of myosin V's two heads is bound to F-actin at any time. *High processivity is essential for a transport engine such as myosin V, because its cargo must be moved long distances to specific destinations.* In contrast, it is estimated that the myosin II heads in contracting striated muscle are associated with thin filaments only $\sim 5\%$ of the time. In a muscle cell, low processivity is tolerable because the numerous myosin–actin interactions

that occur more or less simultaneously keep thin and thick filaments in contact.

Electron micrographs of myosin V bound to F-actin reveal that its globular heads are separated by 13 actin subunits (Fig. 35-86a) corresponding to the $\sim 370\text{ \AA}$ length of the microfilament's helical repeat (Fig. 35-63b). How might myosin V move along a microfilament track? Two mechanisms have been proposed: (1) the hand-over-hand mechanism in which the two heads alternate in taking the lead, much like the feet of a person walking; and (2) the inchworm mechanism, in which one of the heads is always in the lead so that their movement resembles the feet of a person walking sideways. To distinguish between these possibilities, Yale Goldman and Paul Selvin observed the movement of a myosin V head domain bearing a fluorescently labeled light chain via fluorescence microscopy. The myosin V head moved $\sim 740\text{ \AA}$ for every ATP it hydrolyzed (Fig. 35-86a), which is consistent with the hand-over-hand mechanism, although since the other head remained attached to the microfilament during a single step, the cargo moved half this distance, $\sim 370\text{ \AA}$, per step (Fig. 35-86b). The inchworm model, which requires that the heads move $\sim 370\text{ \AA}$ per step, is thereby ruled out.

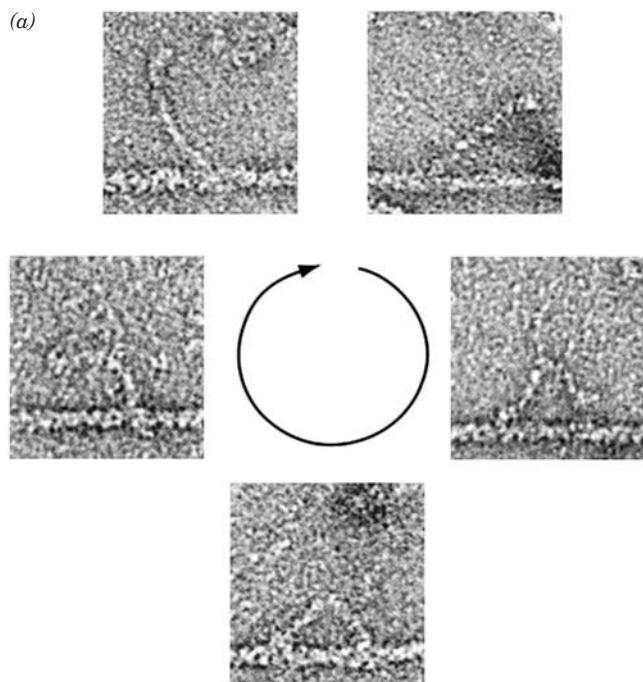
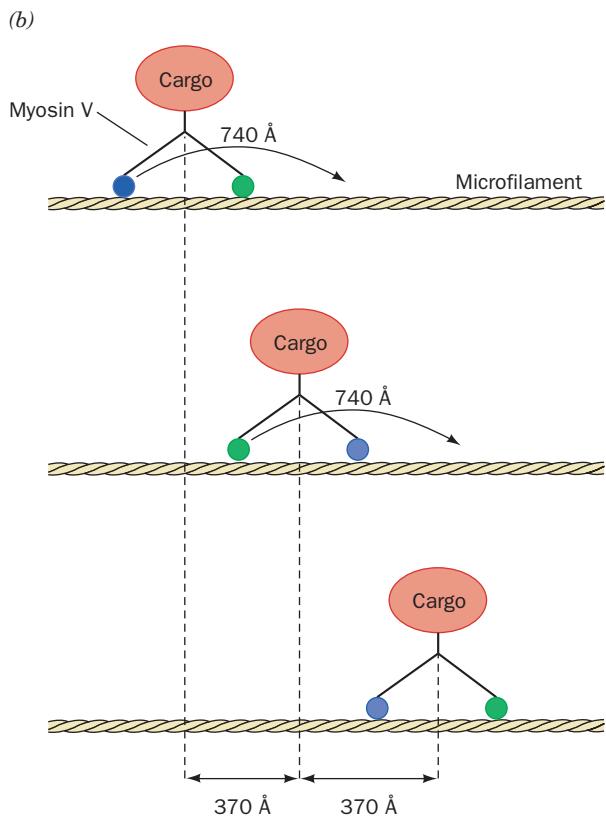
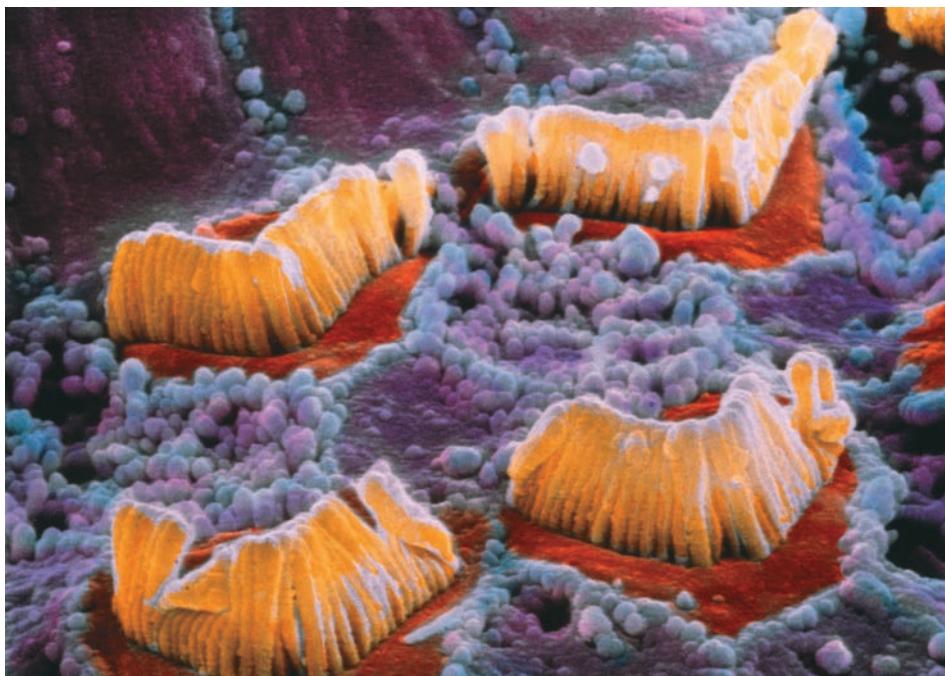


Figure 35-86 Hand-over-hand mechanism of myosin V movement. (a) Series of electron micrographs of individual myosin V molecules at different stages in their mechanical cycle as each was “walking” along an actin filament. The circular arrow is indicative of the series of events as the protein moves from left to right along the actin filament towards its (+) end. In the lower micrograph, both heads are attached to the actin filament with a spacing of $\sim 370\text{ \AA}$. [Courtesy of Peter Knight, University of Leeds, U.K.] (b) In this schematic diagram, the trailing head (purple in the top panel) advances in steps of $\sim 740\text{ \AA}$. However,



because the leading head (green in the top panel) remains in place during this step (and thereby becomes the trailing head), the cargo advances $\sim 370\text{ \AA}$ per step. [After Yildiz, A., Forkey, J.N., McKinney, S.A., Ha, T., Goldman, Y.E., and Selvin, P.R., *Science* **300**, 2061 (2003).]

(a)



(b)

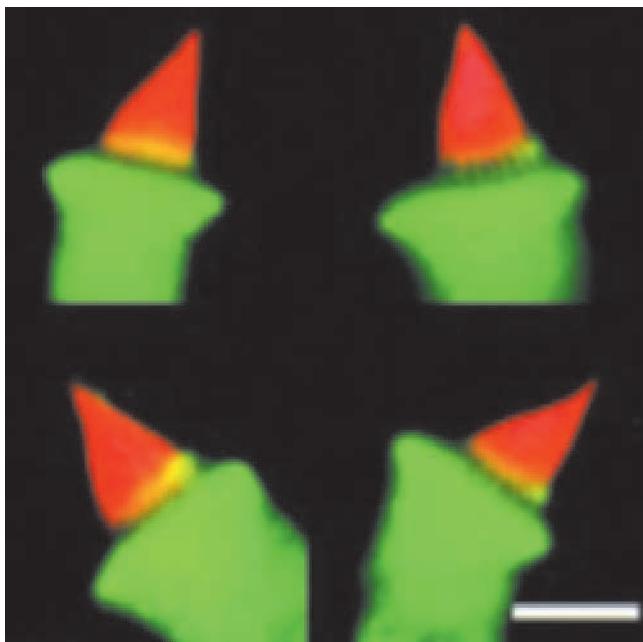


Figure 35-87 Stereocilia of auditory hair cells. (a) Scanning electron micrograph showing the stereocilia on four adjacent auditory hair cells. [Photo by P. Motta, University La Sapienza, Rome/Science Photo Library/Photo Researchers.] (b) Fluorescence micrographs of four separate auditory hair cells. The actin filaments are labeled with a red fluorescent dye and myosin VIIa is labeled with a green fluorescent dye. The yellow bands near the bases of the hair bundles are indicative of high concentrations of myosin VIIa that may participate in linking adjacent stereocilia. The length of the scale bar is 10 μm . [Courtesy of Peter G. Gillespie, Oregon Health Science University.]

gated ion channels, thereby triggering nerve impulses that are perceived as sound.

Usher 1B syndrome, the most common form of deaf-blindness in industrialized nations, is characterized by profound hearing loss from birth, **retinitis pigmentosa** (which causes a progressive loss of vision ending in blindness in the second or third decade), and in some cases, difficulties with balance due to vestibular problems. Usher 1B syndrome is an autosomal recessive condition arising from defects in the gene encoding myosin VIIa. The deafness is caused by the consequent disorganization of the otherwise precisely arranged stereocilia. The inability of the defective stereocilia to respond to sound waves is probably also responsible for their inability to detect fluid movement in the vestibular apparatus, which is required for sensing balance. Myosin VIIa is also present in the retina where it normally functions to transport pigment particles in both **rod cells** (the photoreceptors of night vision) and their associated **pigmented epithelial cells**. Hence, in individuals with Usher 1B syndrome, the retina progressively loses its ability to respond to light, beginning with the loss of night vision.

c. Usher 1B Syndrome Is Caused by Defective Myosin VIIa

The **cochlea**, the spiral-shaped organ of the inner ear, contains ~ 3500 **auditory hair cells**, which are so named because of the bundles of stiff, hairlike **stereocilia** that project from their apical surfaces and which are arranged in a manner reminiscent of organ pipes (Fig. 35-87a). Each stereocilium consists of a rigid bundle of actin filaments that are associated with **myosin VIIa** near their base (Fig. 35-87b). The tilting of stereocilia by sound waves opens mechanically

G. Microtubules

Eukaryotes have two nearly ubiquitous but unrelated types of motility systems: (1) microfilament-based systems, such as muscle, which contain actin, and (2) **microtubule-based systems**, which consist mainly of the protein **tubulin**. Microtubules (Fig. 35-88), as their name implies, are tubular structures, ~ 250 Å in diameter, which form a class of cytoskeletal components distinct from the ~ 90 Å in diameter microfilaments and the 100- to 150-Å-diameter **intermediate filaments** (the cytoskeleton's third major component, which apparently has only a structural role; Section 1-2Ae). *Microtubules comprise the major components of such cellular organelles as the mitotic spindle and cilia, and form the framework that organizes the cell.*

a. Tubulin Is a Heterodimer

Tubulin consists of two 40% identical, ~ 450 -residue subunits named **α -tubulin** and **β -tubulin**, both of which are highly conserved among eukaryotes. The structure of tubulin, determined by Kenneth Downing and Eva Nogales using electron crystallography (Section 12-3A), reveals that the core of each tubulin subunit consists of a 4-stranded and a 6-stranded β sheet surrounded by 12 α helices. The two subunits associate in head-to-tail fashion to form the heterodimer (Fig. 35-89). α -Tubulin and β -tubulin differ primarily in the lengths and conformations of certain loops and in the sequences at their C-termini (which are absent in the electron crystal structure).

The contacts between α - and β -tubulin in the dimer consist of numerous hydrophobic residues. Indeed, the

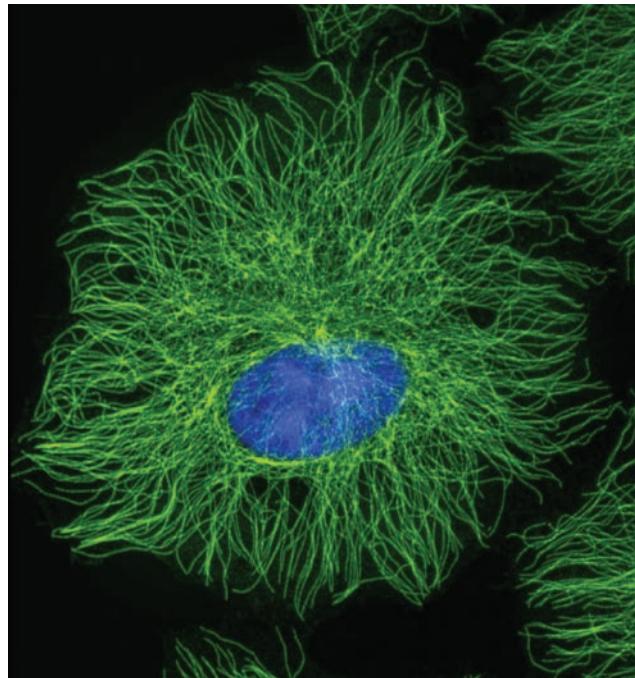


Figure 35-88 The network of microtubules in a fibroblast as revealed by indirect immunofluorescence microscopy using anti-tubulin antibodies. The microtubules are green and the nucleus is blue. [Thomas Deerinck, NCMIR/Photo Researchers, Inc.]

heterodimer dissociates only under denaturing conditions. Tubulin monomers, with their exposed hydrophobic surfaces, are therefore prone to aggregation so that their assembly into dimers requires the participation of several molecular chaperones.

b. Tubulin Binds GTP

The N-terminal ~ 200 residues of both α - and β -tubulin form dinucleotide-binding folds typical of nucleotide-binding proteins (Section 8-3Bi). Unlike actin and myosin, however, α - and β -tubulin each bind a guanine nucleotide, either GTP or GDP. Specificity for guanine is conferred by

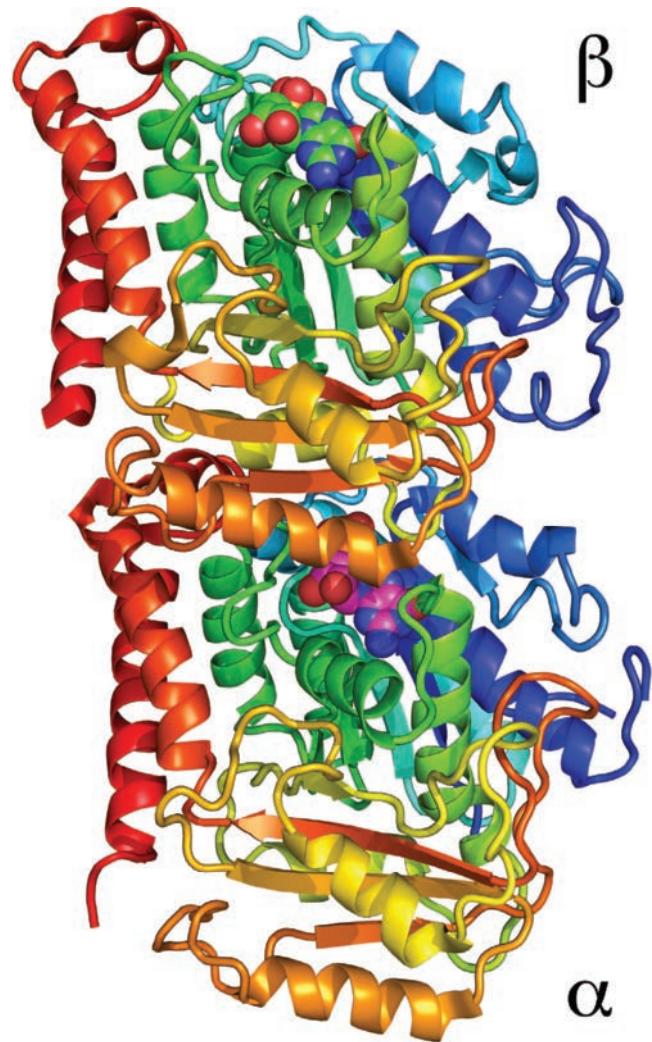


Figure 35-89 Electron crystal structure of the tubulin dimer.

This 3.5-Å-resolution model of bovine tubulin reveals the structural similarities between its α subunit (below) and its β subunit (above), both colored in rainbow order from N-terminus (blue) to C-terminus (red). The GTP bound to α -tubulin and the GDP bound to β -tubulin are drawn in space-filling form with GTP C magenta, GDP C green, N blue, O red, and P magenta. Note that the GTP is inaccessible in the dimer, whereas the GDP is more exposed to the solvent. [Based on a structure by Kenneth Downing and Eva Nogales, Lawrence Berkeley National Laboratory, Berkeley, California. PDBid 1JFF.]

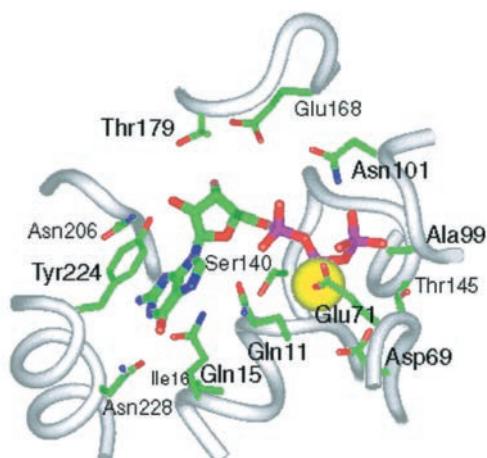


Figure 35-90 The nucleotide-binding site of α -tubulin. The protein backbone is gray and its bound GTP and its associated amino acid side chains are shown in stick form with C green, N blue, O red, and P magenta. Note how Tyr 224 stacks on the guanine base. An Mg^{2+} ion (yellow sphere) is coordinated by the nucleotide phosphate groups and by Asp 69 and Glu 71. [Courtesy of Kenneth Downing and Eva Nogales, Lawrence Berkeley National Laboratory, Berkeley, California. PDBid 1JFF.]

two Asn residues that hydrogen-bond with guanine's 2-amino group (which adenine lacks). Guanine fits into a hydrophobic pocket defined by nonpolar residues, including a Tyr side chain that stacks on the base (Fig. 35-90). The phosphate groups form hydrogen bonds with backbone groups and ligand a bound Mg^{2+} ion.

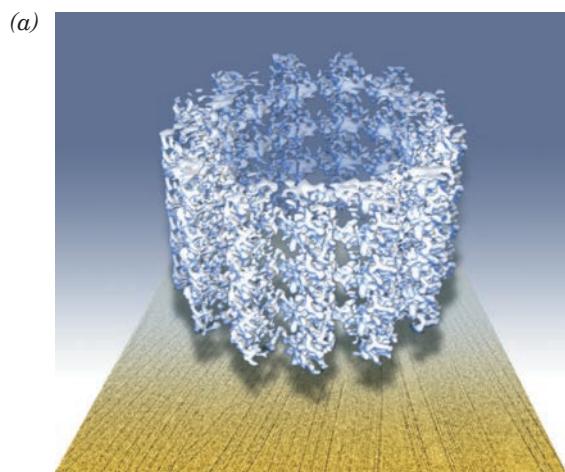
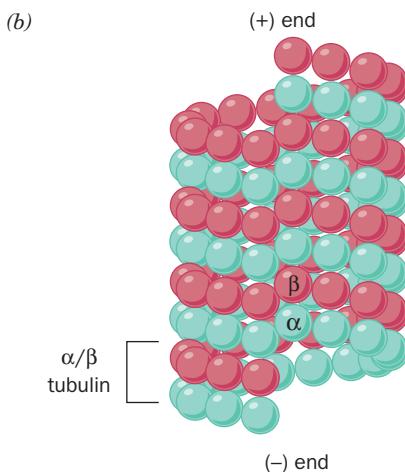


Figure 35-91 Structure of microtubules. (a) Cryoelectron microscopy-based surface view of a microtubule at 8-Å resolution. It consists of 13 adjacent protofilaments that are slightly staggered so that the subunits are arranged in a shallow left-handed helix with a rise of three subunits per turn. Some of the electron micrographs that were used to generate the microtubule image are shown below it. [Courtesy of Kenneth Downing and Eva Nogales, Lawrence Berkeley National

In the tubulin dimer, α -tubulin's bound GTP is buried at the subunit interface (Fig. 35-89) and is therefore nonexchangeable. A Lys residue (Lys 254) from β -tubulin interacts with the terminal phosphate group of the GTP bound to α -tubulin. However, the GTP bound to β -tubulin is not protected in this way and hence is exposed to solvent. When tubulin dimers polymerize to form a microtubule (see below), the GTP in β -tubulin is hydrolyzed to GDP + P_i . This reaction occurs only when α -tubulin from an adjacent dimer closes off the β -tubulin nucleotide-binding site. α -Tubulin has Glu rather than Lys at position 254, and this acidic residue helps catalyze GTP hydrolysis in β -tubulin. Thus, β -tubulin cannot catalyze the hydrolysis of its bound GTP until it is part of a tubulin polymer, at which point the resulting GDP becomes nonexchangeable until the subunit dissociates from the microtubule. The GTP bound to the α subunit is not hydrolyzed.

c. Microtubule Structure and Assembly

Cryoelectron microscopy-based images of microtubules (Fig. 35-91a) reveal that the subunits are arranged in a left-handed helical tube consisting of 13 parallel **protofilaments** (the linear strips of alternating α and β subunits that are parallel to the helix axis; Fig. 35-91b). The helix has a rise per turn of three subunits in which subunits of a given type contact subunits of the same type in neighboring protofilaments to form shallowly spiraling rows of α and β subunits. This results in a seam running the length of the helix across which subunits of the opposite type are in contact. Microtubules are polar, as are microfilaments (Section 35-3E), with the so-called (+) end terminating in β -tubulin subunits and the (-) end terminating in α -tubulin subunits.



Laboratory, Berkeley, California.] (b) Model of microtubule structure. Here, the α -tubulin subunits are green and the β -tubulin subunits are red. Note that although the subunits in most protofibrils contact subunits of the same type in adjacent protofibrils, there is a seam running the length of the microtubule across which the contacting subunits have the opposite type.

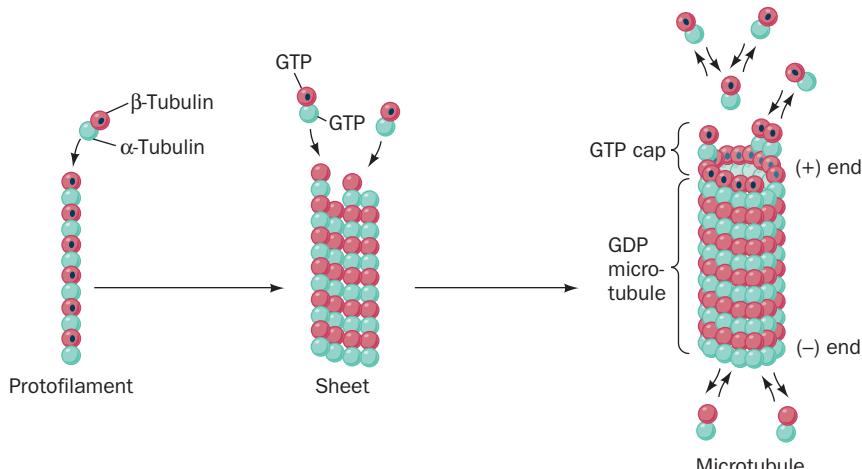


Figure 35-92 Assembly of a microtubule. $\alpha\beta$ Dimers of tubulin initially form linear protofilaments. These associate side by side to form a curved sheet of ~ 1000 dimers that wraps around to form a microtubule which, in most cases, has 13 protofilaments. The microtubule then grows by addition of tubulin dimers to either end, but growth is about twice as fast at the (+) end, which terminates in the β subunits. A tubulin dimer

initially has a GTP bound to both its α and β subunits, but, on incorporation of a dimer into a microtubule, its β subunit (but not its α subunit) hydrolyzes its bound GTP (black dots) to GDP. When the rate of polymerization is greater than that of GTP hydrolysis, the (+) end will have a cap of GTP-containing β subunits, although the remaining β subunits contain GDP.

The assembly of a microtubule begins with the end-to-end association of tubulin dimers to form protofilaments (Fig. 35-92, left). The protofilaments then align side by side in a curved sheet (Fig. 35-92, middle), which closes on itself to form a microtubule (Fig. 35-92, right). It seems likely that the seam running the length of the microtubule marks its site of closure. Most microtubules contain 13 protofilaments, although this number ranges from 9 to 16.

The microtubule lengthens as tubulin dimers add to both ends. However, *tubulin dimers bind preferentially to the (+) end of the microtubules so that the (+) end grows about twice as fast as the (-) end*. The higher affinity of tubulin dimers for the (+) end permits polymerization to occur even when opposed by a compressive force. Consequently, a growing microtubule can perform work such as centering the nucleus or moving chromosomes during mitosis.

d. GTP Hydrolysis Destabilizes Microtubules

GTP hydrolysis is not required for microtubule assembly, as demonstrated by the polymerization of tubulin subunits that have bound the nonhydrolyzable GTP analog GMPPNP (Section 19-3Ch). *After each tubulin dimer adds to the (+) end of a growing microtubule, residues from an incoming α -tubulin help catalyze the hydrolysis of the GTP bound to β -tubulin.* The resulting GDP destabilizes the microtubule. Nevertheless, the microtubule does not disassemble as long as the (+) end is capped by a ring of β -tubulin subunits binding GTP. If this “GTP cap” dissociates or exchanges its bound GTP for GDP, as occasionally occurs, the microtubule depolymerizes at a rate that is ~ 100 -fold greater than if the cap were present (microtubules binding GMPPNP are essentially unable to depolymerize).

GTP hydrolysis and the departure of the phosphate group apparently trigger a conformational change in β -tubulin such

that the GDP- β -tubulin complex bends slightly away from its α -tubulin partner in the dimer. Consequently, protofilaments containing only GTP are straight, whereas protofilaments with GDP bound to the β subunit are curved. In fact, electron micrographs of depolymerizing microtubules show protofilaments splaying out from the ends of the microtubule (Fig. 35-93). Depolymerization through the fraying of protofilaments is faster than the dissociation of tubulin dimers from the microtubule end.

e. Microtubules Exhibit Dynamic Instability

In vivo, the (+) end of a microtubule undergoes random changes in length by polymerization (when it adds tubulin) and depolymerization (when it loses its GTP cap). Thus, in a phenomenon known as **dynamic instability**, some microtubules in a given population will grow while others shrink. This continual activity at the (+) end, a form of substrate cycling (Section 17-4E), probably facilitates rapid microtubule growth when needed, just as a car with an idling motor can make a faster getaway than a car with its ignition off. The balance between microtubule polymerization and depolymerization depends on **microtubule-associated proteins (MAPs)** that cap microtubule ends or bind preferentially to bent tubulin dimers to promote protofilament fraying.

Under the conditions that tubulin subunits add to the (+) end as fast as they dissociate from the (-) end, the microtubule undergoes treadmilling. However, *in vivo*, the (-) ends are often anchored to some sort of organizing center. Hence, most microtubule growth and regression occur at the (+) end, which is usually near the cell periphery.

Subunits of **γ -tubulin**, which is $\sim 30\%$ identical in sequence to α - and β -tubulins but does not by itself assemble into microtubules, form rings or lockwasherlike assemblies

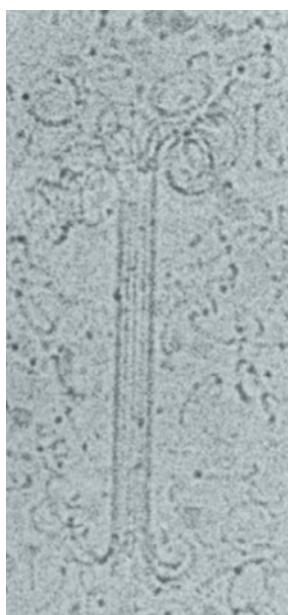


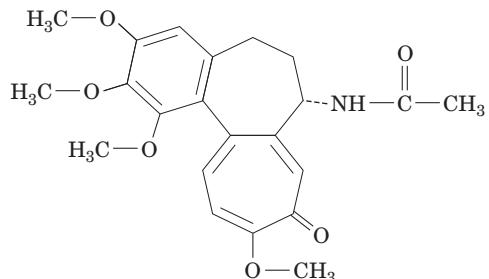
Figure 35-93 Electron micrograph of a depolymerizing microtubule with an accompanying interpretive drawing.

Protofilaments curve and separate from the ends of the microtubule, giving it a frayed appearance. This presumably promotes the dissociation of tubulin and segments of protofilaments from the ends. [Electron micrograph courtesy of Ronald Milligan, The Scripps Research Institute, La Jolla, California.]

at the (–) ends of certain microtubules where they apparently function to nucleate microtubule assembly. The microtubules that form the mitotic spindle in animal cells emanate from structures known as **centrioles** (Fig. 1-5), which contain the variant **δ-tubulin**. By regulating the rate of tubulin polymerization, cells presumably vary their shapes and induce the formation and dissolution of such cellular apparatus as the mitotic spindle.

f. Antimitotic Drugs Inhibit Microtubule Formation

Microtubules participate in essential cell activities and hence compounds that interfere with microtubule dynamics can drastically affect cell structure and function. Thus, **colchicine**,



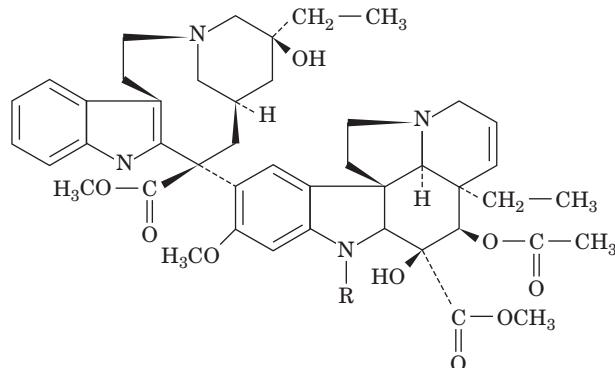
Colchicine

a product of the autumn crocus (*Colchicum autumnale*), inhibits microtubule-dependent cellular processes by inhibiting the polymerization of tubulin protomers. For example, colchicine arrests mitosis in both plant and animal cells at

metaphase (when the condensed and replicated chromosomes line up on the cell's equator; Section 1-4Aa) by preventing the formation of the mitotic spindle. It also inhibits cell motility.

Colchicine has been used for centuries to treat acute attacks of gout (which result from elevated uric acid levels in body fluids; Section 28-4Ba). The lysosomes of the white cells that engulf urate microcrystals are ruptured by these needle-shaped crystals, causing cell lysis and triggering the local acute inflammatory reaction responsible for the exquisite pain characteristic of gout attacks. Colchicine slows the ameboid movements of white cells by inhibiting their microtubule-based systems, thus limiting their ability to migrate to the site of inflammation.

The **vinca alkaloids**, **vinblastine** and **vincristine**,

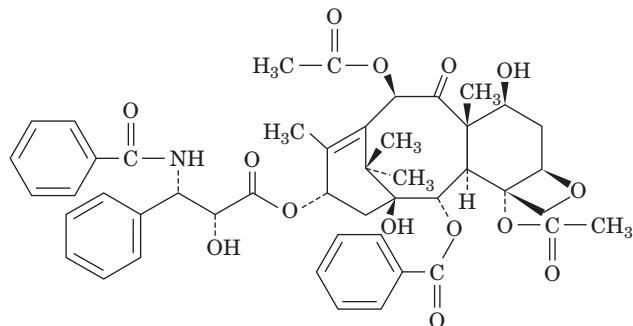


Vinblastine: R = CH₃

Vincristine: R = CHO

products of the Madagascan periwinkle (*Catharanthus roseus*; formerly named *Vinca rosea*), also inhibit microtubule polymerization by binding to tubulin. These substances are widely used in cancer chemotherapy since blocking mitosis preferentially kills fast growing cells. Curiously, colchicine is not selectively toxic to cancer cells.

Taxol



Taxol

binds to the β-tubulin subunits in microtubules, but not to free tubulin. It therefore stabilizes microtubules by preventing their depolymerization. This inhibits the dissolution of the mitotic spindle, thereby preventing cell division. The taxol–tubulin interaction appears to include close contacts between taxol's phenyl groups and hydrophobic residues such as Phe, Val, and Leu. *Taxol blocks cell division and is therefore particularly toxic to rapidly dividing cells such as tumor cells.* Indeed, it is a frequently used

chemotherapeutic agent against ovarian, breast, and certain types of lung cancer. Taxol was originally extracted from the bark of the slow-growing and endangered Pacific yew tree (*Taxus brevifolia*), but is now obtained from other more common species of yew trees combined with chemical synthesis.

H. Microtubule-Associated Motors

Microtubules participate in several different forms of cell movement: The growth and rearrangement of microtubules can alter cell shape; microtubules can push or pull cellular constituents such as chromosomes during mitosis; the motions of cilia and flagella are driven by the sliding of microtubules past one another; and microtubules serve as tracks for vesicle- and organelle-carrying transport motors, much as do microfilaments with many unconventional myosins (Section 35-3F). In this section we examine the two microtubule-associated motor proteins, kinesins and dyneins.

a. Kinesins

Although **kinesins** were only discovered in 1985, they are nearly as well characterized as myosins, with which they share important structural and functional properties. The kinesin superfamily consists of 16 classes, which vary as to the position of a conserved ATP-binding motor domain in the polypeptide chain (N-terminal, middle, or C-terminal), the nature of the remaining sequences, the number of

identical and different subunits with which they associate, and their direction of transport along a microtubule [(-) to (+) or (+) to (-)]. In fact, the human genome encodes as many as 45 kinesins. Even though the size of kinesin's motor domain is less than half that of myosin, the two proteins have several short sequences in common. Evidently, myosin and kinesin have a common ancestor. The various classes of kinesins function to transport different types of cargos [e.g., Golgi-derived vesicles (Section 12-4C), organelles (lysosomes, mitochondria, peroxisomes), cytosolic components (e.g., mRNAs, proteins, and cytoskeletal elements such as intermediate filaments), chromosomes, and certain viruses and viral components] in different portions of the cell and in different types of cells. Here we describe the first kinesin to be characterized, **kinesin-1**, which was previously known as **conventional kinesin**.

Kinesin-1 (Fig. 35-94a) is an ~1000-Å-long protein that consists of two identical heavy chains (~125 kD) and two identical light chains (~65 kD) with the heavy chains forming two large globular heads and a coiled-coil tail (as do myosin heavy chains). It functions to transport vesicles and organelles in the (-) to (+) direction along microtubules (Fig. 35-94b).

The X-ray structure of the N-terminal portion of rat kinesin-1, determined by Eva-Maria Mandelkow and Eckhardt Mandelkow, reveals that each 100-Å-long head consists of an 8-stranded β sheet flanked by three α helices on each side and includes a tubulin-binding site and a nucleotide-binding site (Fig. 35-95). A so-called neck linker

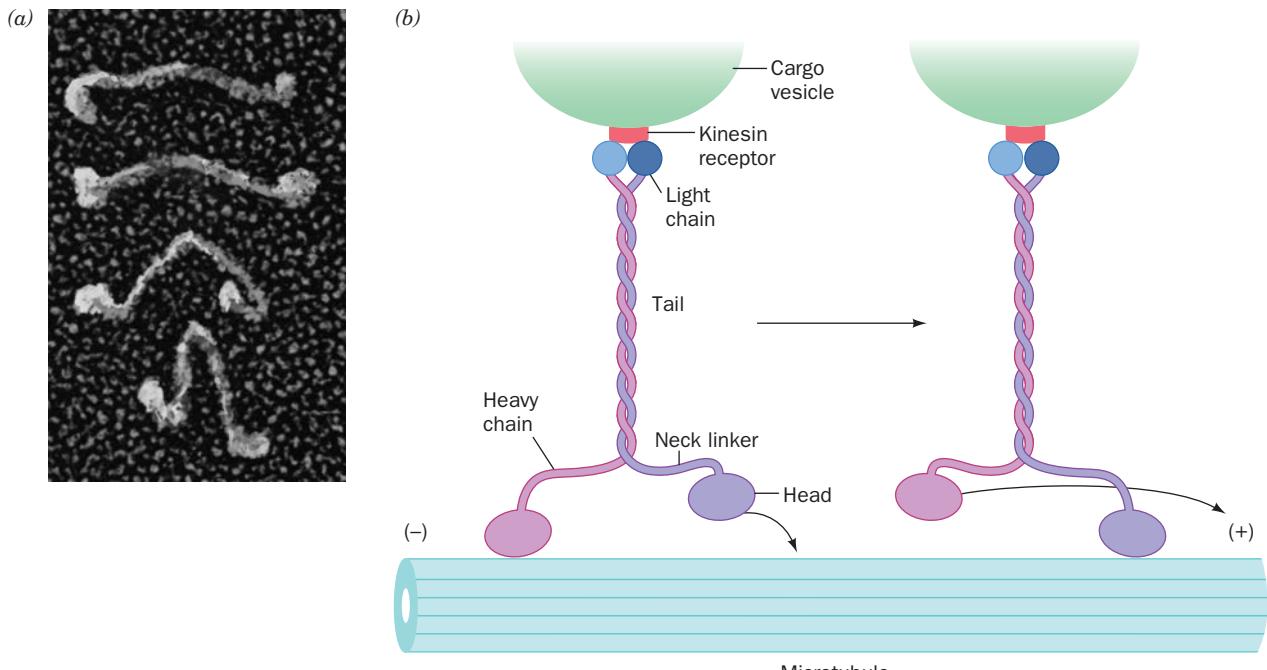


Figure 35-94 Kinesin-1. (a) Electron micrograph of kinesin-1 molecules. [Courtesy of John Heuser, Washington University School of Medicine.] (b) Schematic diagram of vesicle transport along microtubule tracks by kinesin-1. Kinesin may

bind to a kinesin receptor on the vesicle surface via its light or heavy chains, depending on the vesicle identity, thus permitting kinesin to “walk” the vesicle along the microtubule from its (-) end to its (+) end.

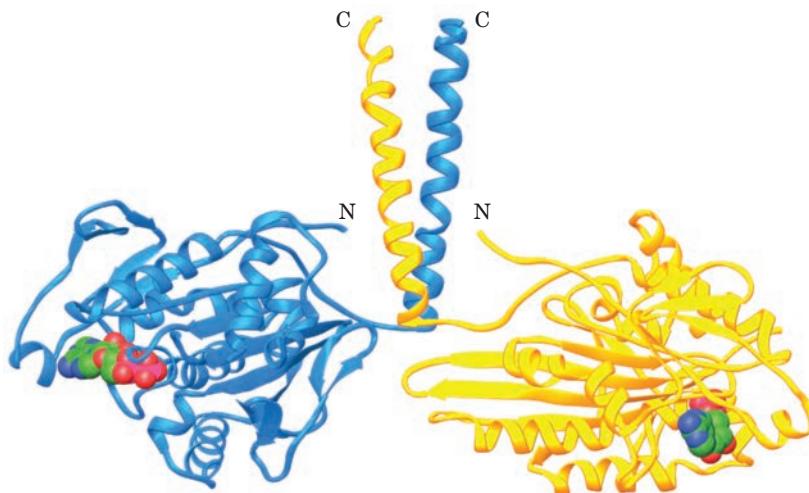


Figure 35-95 X-ray structure of the head and neck region of rat kinesin-1 in complex with ADP. The two identical but asymmetrically arranged heavy chain fragments are drawn in ribbon form in blue and gold. The ADP is shown in space-filling form with C green, N blue, O red, and P gold. Note that each globular head terminates in a flexible neck linker. The linker connects into an α helix that winds around its counterpart in the

other subunit to form a coiled coil. Two light chains at the end of the coiled-coil tail (not present in this X-ray structure) can interact with a membranous vesicle, the cargo. [Based on an X-ray structure by Eva-Maria Mandelkow and Eckhardt Mandelkow, Max-Planck Unit for Structural Molecular Biology, Hamburg, Germany. PDBid 3KIN.]

consisting of 10 to 15 residues is located between kinesin's globular core and its tail. It is followed by an α -helical stalk that leads into the coiled coil that forms through the dimerization of the two heavy chains. The light chains bind to both the end of the coiled-coil tail and to proteins in the membrane of a vesicle or organelle, which then becomes kinesin's cargo. Although homodimeric, kinesin is asymmetric, with its two chemically identical globular heads separated and rotated by $\sim 120^\circ$ relative to one another. This asymmetry is due to the fact that the two heads operate in an alternating fashion so that they never have the same conformation (see below).

b. ATP Hydrolysis Triggers Conformational Changes in the Neck Linker

Kinesin, as do other motor proteins, couples the steps of ATP hydrolysis to conformational changes that cause it to translocate along its associated microtubule track. However, *kinesin cannot follow the myosin lever mechanism because its head domains are not rigidly linked to its neck regions*. Thus kinesin's relatively flexible neck linker connects its ATPase domain to its coiled-coil tail in contrast to the stiff α -helical lever arm that does so in myosin. In kinesin-1, the neck linker takes the form of two short consecutive β strands. The flexibility of kinesin's neck linker is critical for its function, because each kinesin head must swing through an arc much longer than itself.

ATP hydrolysis, followed by the dissociation of P_i , triggers a conformational change in kinesin's catalytic domain that is largely localized to two short polypeptide segments known as **switch I** and **switch II** (Fig. 35-96). Structurally analogous switch regions occur in myosin and G proteins (Section 19-2Cb). Movement of the switch

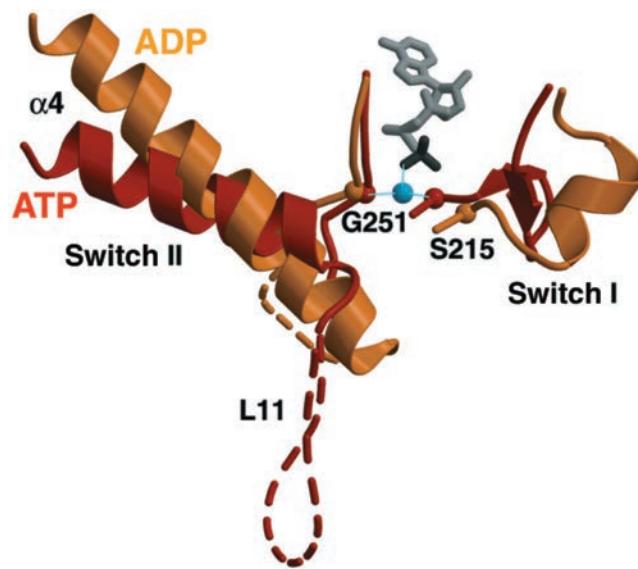


Figure 35-96 Conformational changes in switch I and switch II of the monomeric mouse kinesin KIF1A. The structures known as switch I and switch II undergo conformational changes when ATP is hydrolyzed to ADP. The ATP-bound conformation is drawn in brown and the ADP-bound conformation is drawn in tan. Disordered loops are indicated by dashed lines. ATP is shown as a gray stick model. Residues Gly 251 and Ser 215 are marked by small spheres. The blue sphere is a bridging water molecule. [Courtesy of Nobutaka Hirokawa, University of Tokyo, Japan. PDBids 1I6I and 1I5S.]

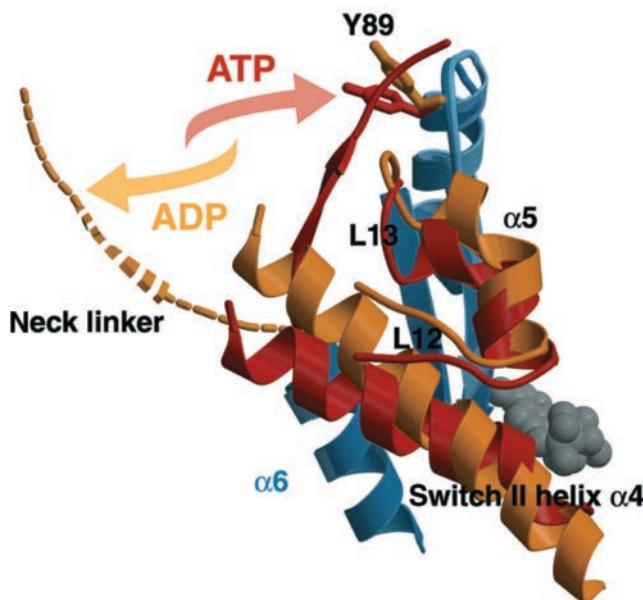


Figure 35-97 Relationship between the switch regions and the neck linker in the monomeric mouse kinesin KIF1A. The ATP-bound conformation is drawn in brown and the ADP-bound conformation is drawn in tan. Portions of the globular core are shown in blue, and ATP is represented by a gray space-filling model. The movement of the neck linker is indicated by arrows. When ATP is bound to kinesin, the neck linker docks with the head domain. Following ATP hydrolysis, the neck linker moves away. [Courtesy of Nobutaka Hirokawa, University of Tokyo, Japan. PDBIDs 1I6I and 1I5S.]

regions triggers additional conformational changes in kinesin's neck linker.

When ATP is bound to kinesin, the neck linker docks with the catalytic core. Following ATP hydrolysis, the neck linker “unzips” so that it moves away from the globular head. The relationship between the switch regions and the neck linker is shown in Fig. 35-97. A kinesin head containing ADP (or no nucleotide at all) is thereby free to bind to a position farther along the microtubule track.

c. Kinesin Follows a Hand-over-Hand Mechanism

Like myosin V (Section 35-3Fb), kinesin moves by a stepping mechanism (Fig. 35-94b). Its two heads work in a coordinated fashion so that at least one is bound to the microtubule track at all times. According to one model for kinesin activity (Fig. 35-98), ATP binds to the leading kinesin head, which is already bound to the microtubule. When the neck linker “zips up” against the catalytic core, the trailing kinesin head is thrown forward. The ADP-binding trailing head has a reduced affinity for the microtubule and is therefore readily detached, which helps keep kinesin moving unidirectionally. *Each kinesin head swings in turn past its partner by $\sim 160 \text{ \AA}$ so that the net movement of the kinesin tail and its cargo is $\sim 80 \text{ \AA}$ for each ATP hydrolyzed*, the length of a tubulin dimer in a protofilament. There is mounting experimental evidence that the two heads coordinate their alternating conformational states through intramolecular tension. Most of the kinesin–tubulin contacts involve the negatively charged C-terminus of β -tubulin, which faces the outside of the microtubule.

Although kinesin and myosin both respond to the events of ATP hydrolysis with conformational changes, the two proteins differ in how they do so. For example, ATP binding triggers the power stroke in kinesin, but causes myosin to dissociate from actin and reclock its lever arm (the recovery stroke; Fig. 35-73). The release of P_i weakens

kinesin's grip on the microtubule and loosens the neck linker, but causes myosin to bind tightly to actin and swing its lever arm forward.

d. Kinesin Is a Processive Motor

A free kinesin head quickly rebinds to its associated microtubule, so the heads spend most of their reaction cycle bound to the microtubule track. Consequently, kinesin is highly processive; it can advance hundreds of steps before dissociating from its microtubule track. Because kinesin maintains its grip on the microtubule at all times, it does not slide backward; it advances despite the viscous forces in the cell that oppose the movement of the bulky vesicle or organelle bound to the kinesin tail. Kinesin processivity is enhanced by the fact that, *in vivo*, cargos are propelled by multiple motors.

Most kinesins, including kinesin-1, move toward the (+) end of a microtubule, away from the center of a cell. However, members of the **kinesin-14** family, which appear to follow the same ATP-triggered mechanism, move in the opposite direction. In kinesin-14, the neck linker is located on the N-terminal side of the catalytic domain (it is on the C-terminal side in kinesin-1) where it swings in the opposite direction to that of kinesin-1 with each ATP hydrolysis cycle. (Kinesins whose catalytic domains are in the middle of the polypeptide are not strictly motors; they participate in the disassembly of microtubules, which is important for remodeling them during mitosis.)

e. Dyneins

Eukaryotic cells contain a third type of motor protein, **dynein**, in addition to myosin and kinesin. The 12 known mammalian dyneins have been grouped into **cytoplasmic dyneins**, which, like kinesin-1, are two-headed proteins that participate in microtubule-based transport but toward the (−) end (toward the cell center), and **axonemal dyneins**,

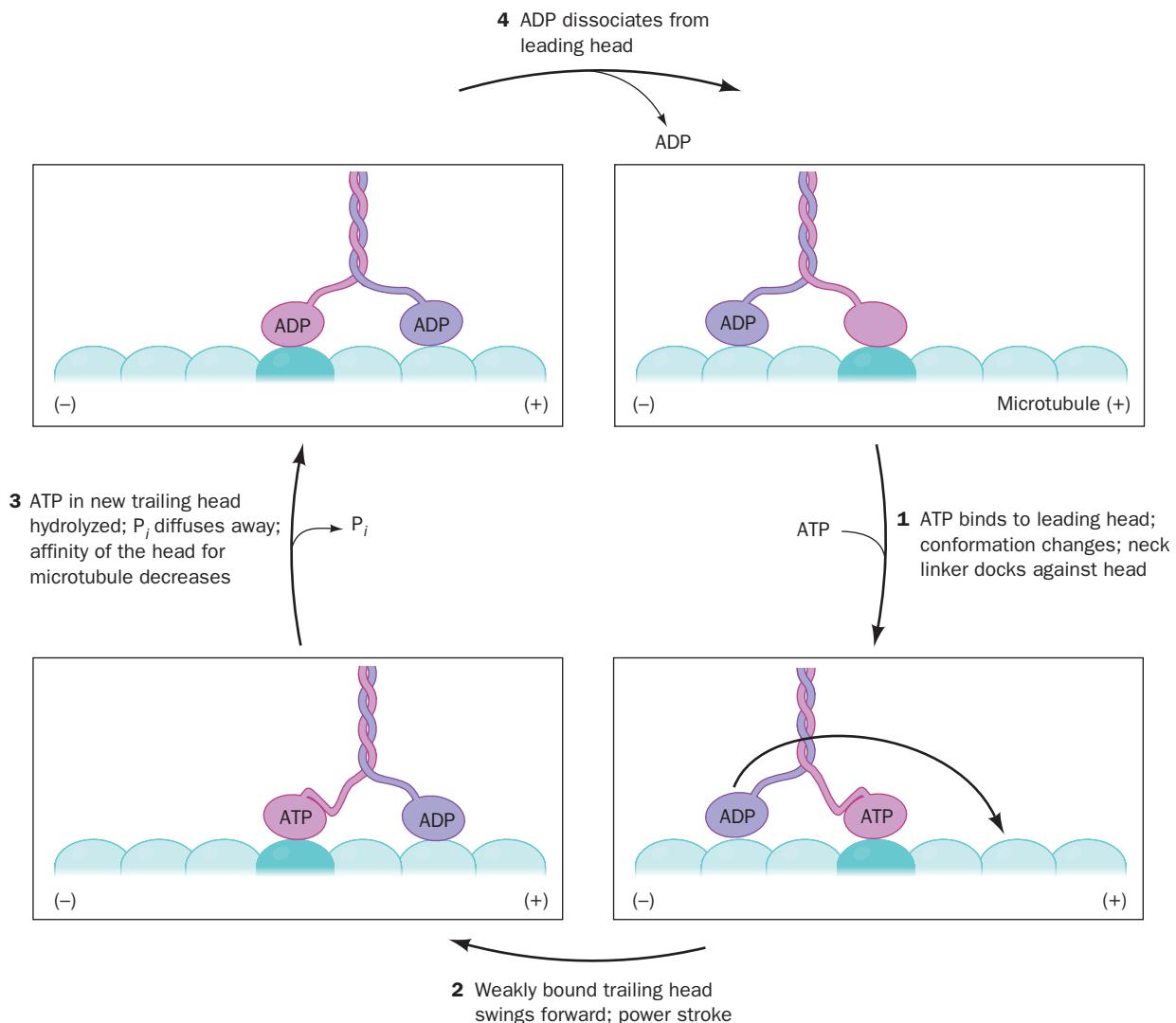


Figure 35-98 The kinesin reaction cycle. A portion of the microtubule track is drawn with one tubulin subunit shaded more darkly for reference. The (+) end of the microtubule is on

the right. The vesicular cargo attached to kinesin's coiled-coil tail is not shown.

which power the motion of cilia and eukaryotic flagella (see below). Dyneins are members of the AAA+ superfamily (Section 30-3Cc) and hence are unrelated to myosins and kinesins.

Dyneins are large (1–2 million D) proteins that consist of one or more ~520-kD heavy chains and a large and variable number of intermediate chains (50–140 kD) and light chains (8–45 kD). A dynein heavy chain has an ~350-kD globular head that contains dynein's motor unit and its ATP-hydrolyzing unit. An ~100-Å-long stalk that protrudes from the head forms an antiparallel coiled coil that ends in an ~120-residue globular microtubule-binding domain. The N-terminal third of the heavy chain forms a long stem that extends from the head and which sequence analysis predicts contains many short stretches of coiled coils. Two or three heavy chains are connected via their stems to a common base that contains the intermediate and

light chains and which functions to bind dynein to its cargo or to the core of a cilium (Fig. 35-99).

f. Dynein and Kinesin Motivate the Intracellular Transport of Vesicles and Organelles along Microtubule Tracks

Eukaryotic cells, as we have discussed, transfer proteins and lipids between their various organelles via membranous vesicles (Section 12-4C). But how do these vesicles find their way to their proper destinations at a reasonable rate? The answer to this question was determined, in part, through the study of vesicle transport in the **axons** of neurons (cellular projections that extend from the cell body by up to 1 m; Fig. 1-10d). The use of **video-enhanced contrast microscopy** (subcellular components are generally smaller than the resolution limit of light) revealed that vesicles and even entire organelles such as mitochondria are unidirectionally

transported within axons at rates of 1 to $5 \mu\text{m} \cdot \text{s}^{-1}$, so they can traverse the length of even the longest axon in several days. This apparently purposeful traffic, which simultaneously occurs in both directions (Fig. 35-100), moves along filamentous tracks that have been identified as microtubules through their binding of specific antibodies.

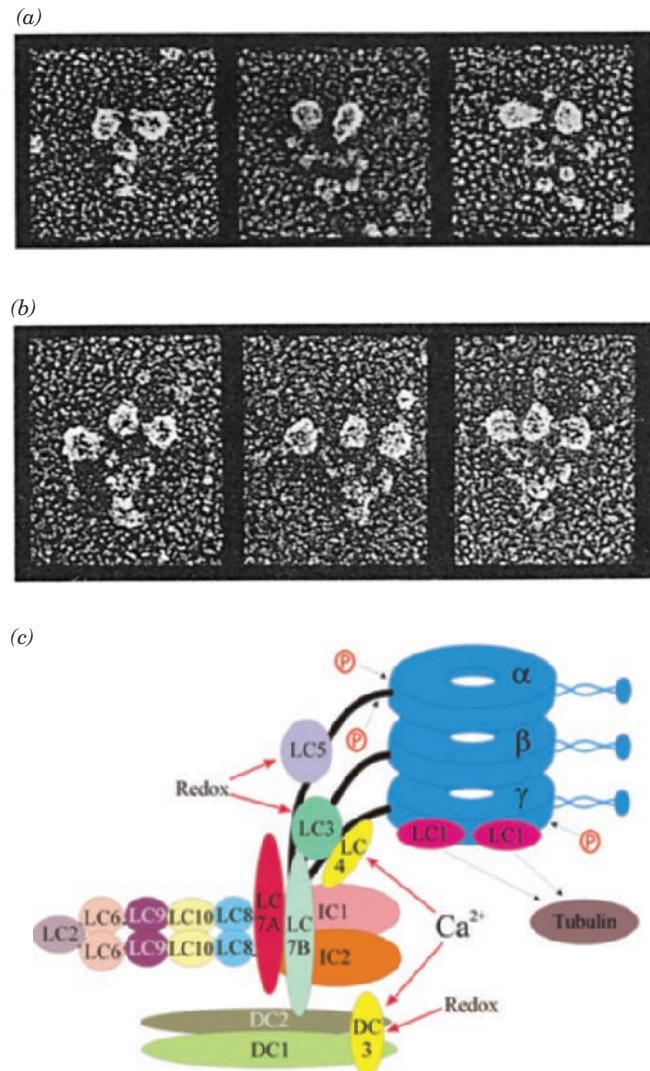


Figure 35-99 Structures of dyneins. (a) Electron micrographs of two-headed dynein from the outer dynein arm of sperm from the sea urchin *Strongylocentrotus purpuratus*. (b) Electron micrographs of three-headed dynein from the flagellar outer dynein arm of the unicellular green alga *Chlamydomonas reinhardtii*. (c) Schematic diagram of the *Chlamydomonas* outer arm dynein indicating the arrangement of its subunits. IC, LC, and DC are abbreviations for intermediate chain, light chain and docking complex. The red arrows point to components that modulate activity in response to Ca^{2+} ion or redox changes. The black arrows pointing from the red circled P's indicate phosphorylation sites. [Parts a and b courtesy of John Heuser and Ursula Goodenough, Washington University School of Medicine. Part c courtesy of Stephen King, University of Connecticut Health Center, and Ritsch Kamiya, University of Tokyo, Japan.]

We have seen that vesicle trafficking also occurs over microfilament tracks as powered by unconventional myosin motors (Section 35-3Fb). In general, it appears that microtubules function in the transport of vesicles over relatively long distances in the cell. However, when vesicles approach the end of a microtubule, they are often transferred

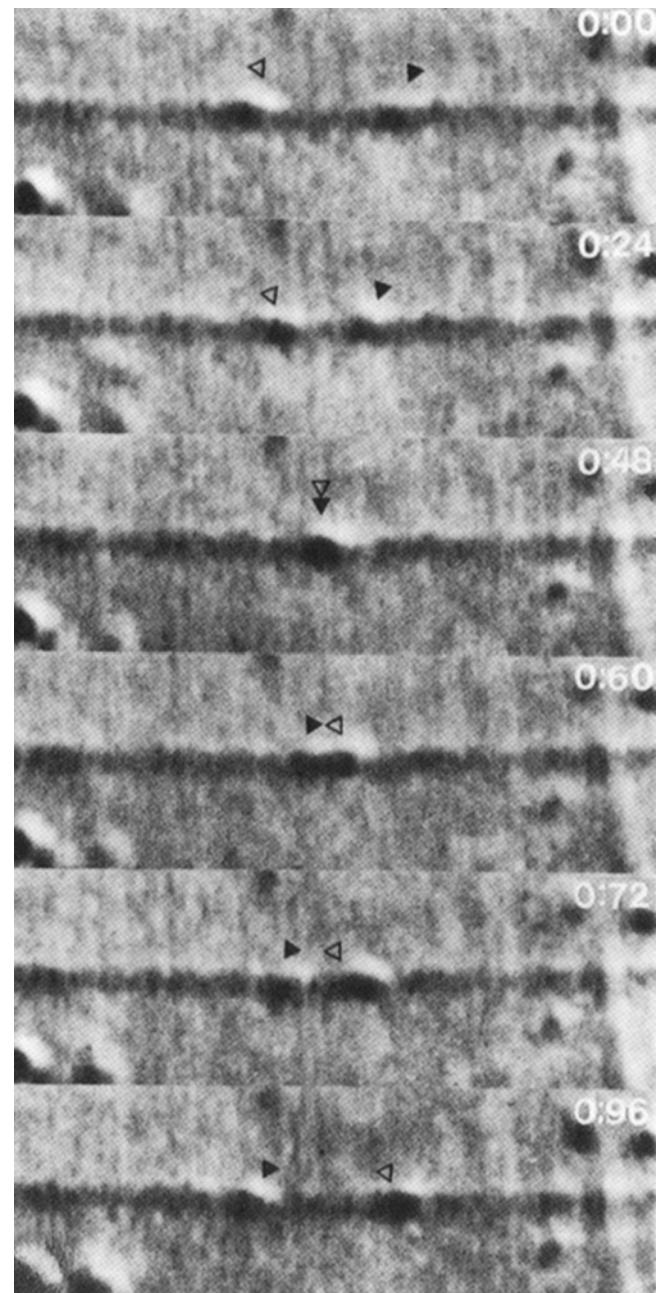


Figure 35-100 Two organelles moving in opposite directions along a microtubule track. This series of successive video-enhanced contrast micrographs shows the two organelles (triangles) passing one another without colliding. The number in the upper right corner of each frame is the elapsed time in seconds from the top frame. [Courtesy of Bruce Schnapp, Boston Medical Center, and Thomas S. Reese, NIH.]

to microfilaments for local transport through the actin-rich cell periphery (e.g., Fig. 35-82). Indeed, in vesicles that bear both microtubule-based motors (kinesins and/or cytoplasmic dyneins) and microfilament-based motors (unconventional myosins), the two types of motors may be physically linked.

g. Cilia and Eukaryotic Flagella Contain Organized Sheaves of Microtubules

Cilia are the hairlike organelles on the surfaces of many animal and lower plant cells that function to move fluid over the cell's surface or to propel single cells through a fluid. In humans, for example, epithelial cells lining the respiratory tract each bear ~ 200 cilia that beat in synchrony to sweep mucus-entrained foreign particles toward the throat for elimination (Fig. 35-101). Cilia are relatively short (as little as a few microns), operate with a whiplike motion, and occur in large numbers on a single cell. **Eukaryotic flagella** (as distinct from prokaryotic flagella, which have entirely different components; Section 35-31), which occur on certain protozoa and comprise sperm tails, are much longer (up to 2 mm), carry out their propulsive function via undulatory motions, and occur in quantities of only one or a few per cell. Nevertheless, both types of organelles have the same basic architecture.

A cilium or flagellum consists of a plasma membrane-coated bundle of microtubules called an **axoneme**. Electron micrographs indicate that an axoneme contains a ring of nine double microtubules surrounding two single microtubules to form a common biological motif known as a **9 + 2 array** (Fig. 35-102a). Each outer doublet consists of a ring of 13-protofilaments, **subfiber A**, fused



Figure 35-101 Scanning electron micrograph of cilia lining the epithelial surface of a mammalian trachea. The rounded surfaces of numerous mucus-secreting goblet cells (yellow) are also visible. [Courtesy of Bruce Schapp, Boston Medical Center, and Thomas S. Reese, NIH.]

to a C-shaped assembly of 10, or in some cases 11, protofilaments, **subfiber B** (Fig. 35-102b), all with their (+) ends extending away from the cell. The 11 microtubules forming

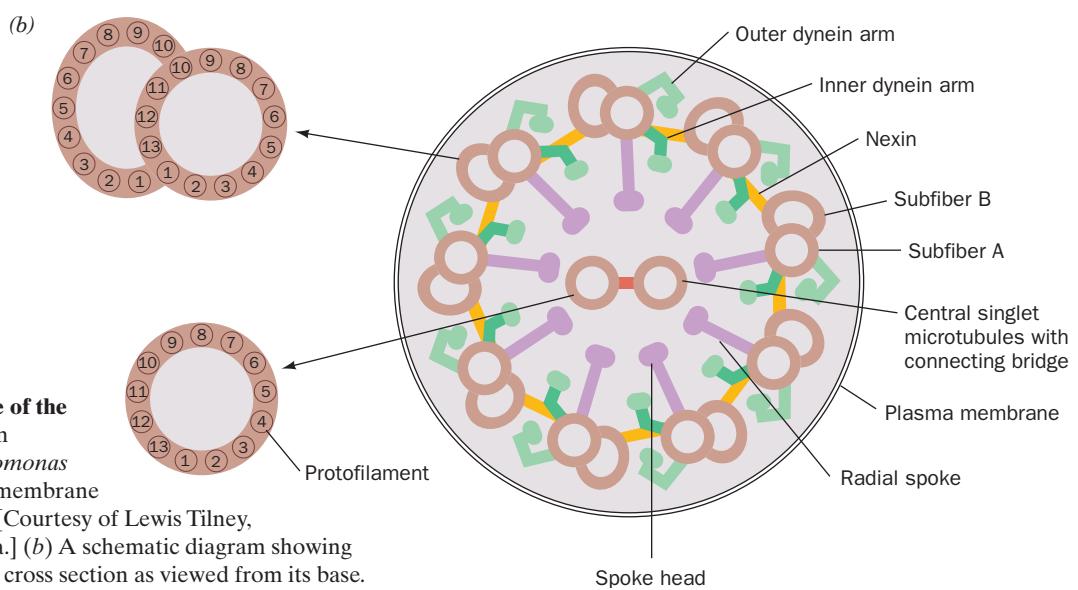
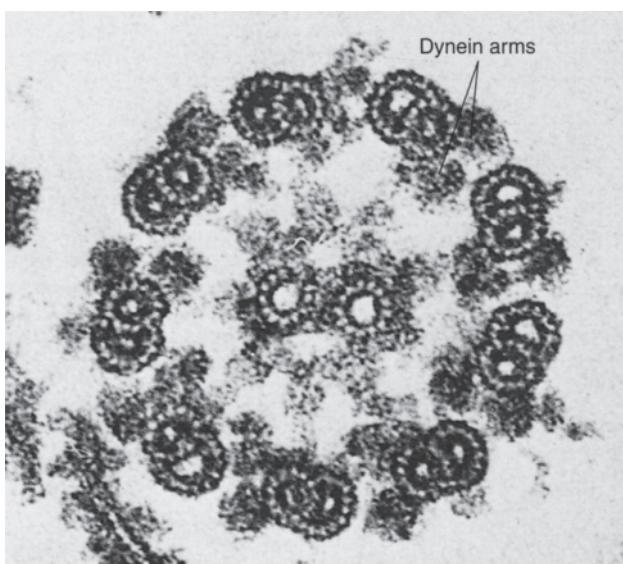


Figure 35-102 Structure of the axoneme. (a) An electron micrograph of a *Chlamydomonas* flagellum, with its plasma membrane removed, in cross section. [Courtesy of Lewis Tilney, University of Pennsylvania.] (b) A schematic diagram showing the structure of a cilium in cross section as viewed from its base.

an axoneme are held together by three types of connectors (Fig. 35-102b):

1. Subfibers A are joined to the central microtubules by radial **spokes**, which each terminate in a knoblike feature termed a **spoke head**.
2. Adjacent outer doublets are joined by circumferential linkers that, in part, consist of a highly elastic protein named **nexin**.
3. The central microtubules are joined by a connecting bridge.

Like microfilaments in the sarcomere, axonemal microtubules have a more or less fixed length and are not subject to dynamic instability, as are cytoplasmic microtubules (Section 35-3Ge). Each type of connector is repeated along the length of the axoneme with its own characteristic periodicity. Finally, every subfiber A bears two arms, an **inner dynein arm** and an **outer dynein arm** (Fig. 35-103), which both point clockwise when viewed from the base of the cilium (Fig. 35-102b).

h. Ciliary Motion Results from the ATP-Powered “Walking” of Dynein Arms along an Adjacent Subfiber B

An isolated eukaryotic flagellum (excised by a laser microbeam) whose plasma membrane has been removed by treatment with a nonionic detergent will continue to beat when supplied with ATP. Evidently, the eukaryotic flagellar “motor” is contained within the axoneme itself rather than at its base, as occurs in bacterial flagella (Section 35-3Ib). What is the site of the eukaryotic flagellar motor? Several observations point to the dynein arms:

1. The dynein arms can be selectively extracted from naked axonemes by solutions containing high salt concentrations. This treatment immobilizes the axonemes while it solubilizes their ATPase activity (although its activity is much lower in solution than in intact axonemes). The addition of purified dynein to the salt-extracted axonemes restores their ATPase activity and their ability to beat.

2. In the absence of ATP, eukaryotic flagella become rigid. Electron micrographs indicate that the dynein arms

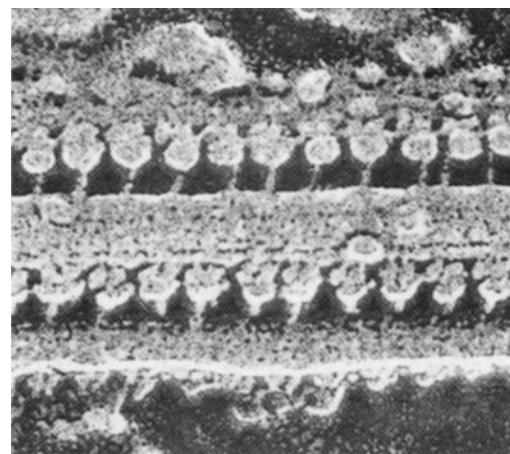


Figure 35-103 Freeze-etch electron micrograph of a flagellar microtubule from the unicellular alga *C. reinhardtii*. The microtubule is shown in transverse view, with its dynein arms projecting like “lollipops.” The outer dynein arms, which are spaced every 240 Å along the microtubule, have 100-Å-diameter heads attached to stalks that are <30 Å wide (an arrangement reminiscent of myosin’s S1 heads). The spacing of the inner dynein arms is less regular. [Courtesy of John Heuser and Ursula W. Goodenough, Washington University School of Medicine.]

in such ATP-deprived flagella are attached to their adjacent subfiber B.

3. Dynein resembles the S1 heads of myosin in both appearance and function. The outer dynein arms consist of either two-headed (~1200-kD) or three-headed (~1900-kD) entities, depending on the species, in which the globular heads are joined to a common base by flexible stems (Fig. 35-103). The inner dynein arms consist of either one- or two-headed structures and hence contain one or two heavy chains. Dynein’s ATPase functions are located in these heads.

4. Brief trypsin treatment, which selectively cleaves the radial spokes and nexin circumferential linkers, followed by the addition of ATP, causes axonemes to elongate up to nine times their original length (Fig. 35-104). The elongation results from the telescoping of the axoneme’s component

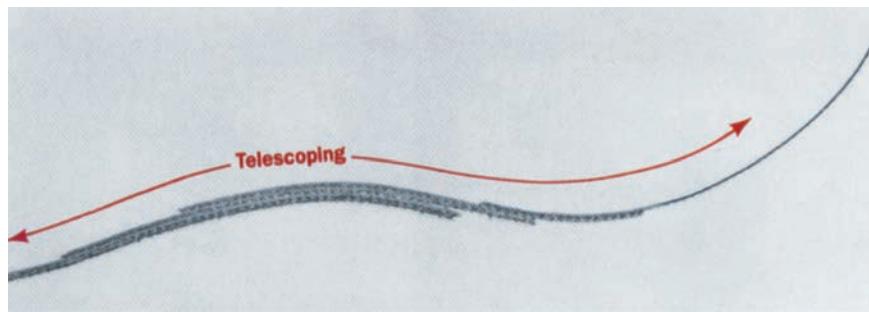


Figure 35-104 Electron micrograph of a telescoped axoneme. It shows an isolated axoneme from a *Tetrahymena* cilium that has been briefly treated with trypsin to degrade its protein connectors and then exposed to ATP. The individual microtubule

doublets telescope out from each other so that the axoneme elongates by up to a factor of 9. [From Warner, F.D. and Mitchell, D.R., *J. Cell Biol.* **89**, 36 (1981).]

microtubules out of the disrupted structure; the individual microtubules do not change in length.

These observations indicate that ciliary motion results from an ATP-driven process reminiscent of the sliding filament model of muscle contraction: *The dynein arms on one microtubule “walk” up the neighboring subfiber B so that these two microfilaments slide past each other.* However, the cross-links between microtubules in an intact cilium prevent neighboring microtubules from sliding past each other by more than a short distance. *These cross-links therefore convert the dynein-induced sliding motion to a bending motion of the entire axoneme.* This model is supported by electron microscopy studies showing that in straight flagella all the outer doublets have the same length and terminate at the same level but, in bent flagella, the doublets at the inside of the bend extend further than those on the outside of the bend (Fig. 35-105).

In order for a cilium or eukaryotic flagellum to bend first in one direction and then in the other, sliding of microtubules on one side of the axoneme must alternate with sliding on the other side (Fig. 35-105). It is unclear how dynein molecules on the two sides of an axoneme are activated in the required reciprocating fashion. In addition, there must be a mechanism to coordinate the movements of the arrays of the thousands of cilia in some cells that beat in synchrony, up to 40 times per second.

i. How Does Dynein Work?

The mechanism whereby dynein transduces the chemical energy of ATP into mechanical force has been elusive because dynein's motor domain is much larger than the motor domains of myosin and kinesin. Its globular head

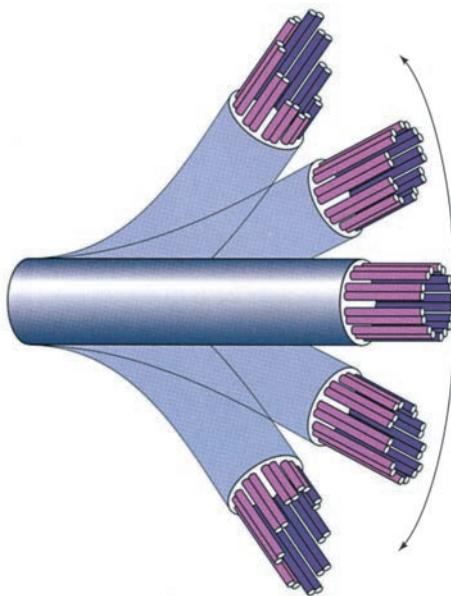


Figure 35-105 The sliding microtubule model of ciliary motion. When the cilium is straight (center), the ends of all the microtubules end at the same level. The cilium bends when microtubules on the inner side of the bend slide past microtubules on the outer side (top and bottom).

consists of six AAA+ domains followed by a C-terminal domain of unknown fold but essential for motor function. Electron micrographs (Fig. 35-106) indicate that the AAA+ domains form a 135-Å-diameter ringlike structure

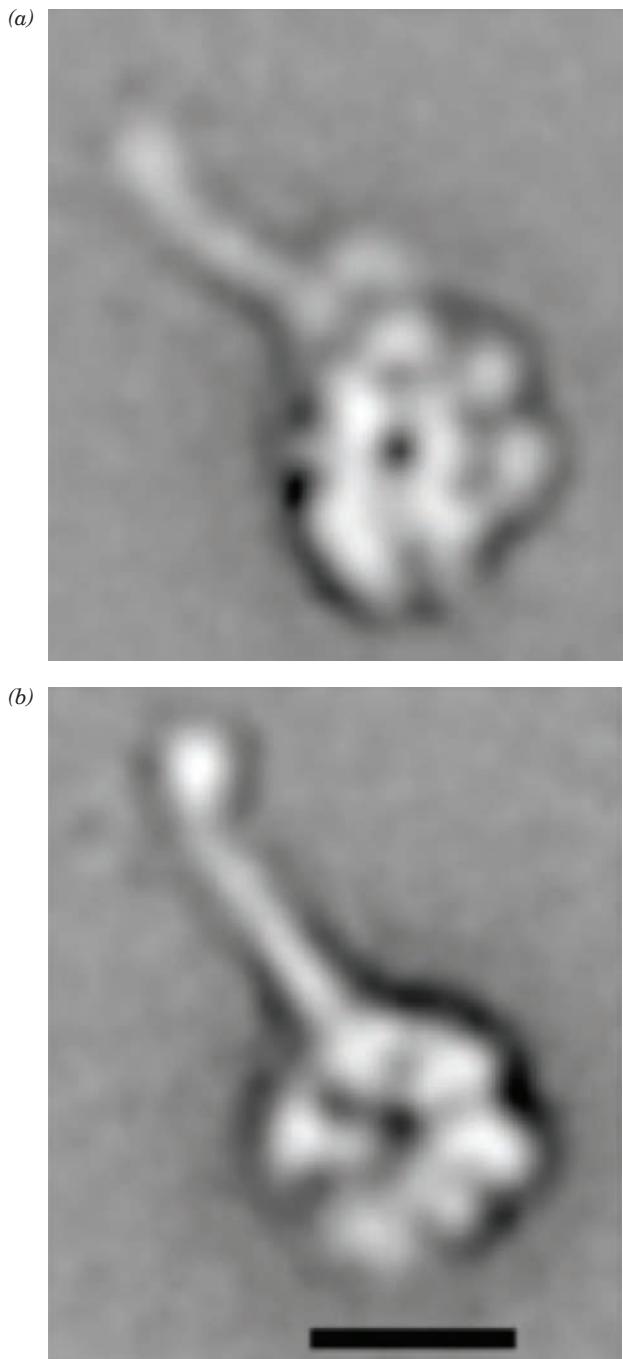


Figure 35-106 Structure of a dynein heavy chain head from the slime mold *Dictyostelium discoideum*. (a) An electron micrograph showing its domain structure. The stalk, which extends diagonally from near the top of the ring, ends in the microtubule-binding domain. (b) Electron micrograph of a head construct in which the N- and C-terminal domains have been deleted. This clearly shows that the six AAA+ domains form a hexagonal ring. The scale bar represents 10 nm. [Courtesy of Stan Burgess, University of Leeds, U.K.]

characteristic of oligomeric AAA+ proteins (e.g., Figs. 30-16 and 30-34), although dynein is unusual in that its AAA+ modules are covalently linked. In addition, the C-terminal domain is closely associated with the ring. AAA+ domains 5 and 6 lack the residues associated with ATP binding and only the ATPase activity of AAA+ domains 1 and 3 is required for motility. The stalk is located between AAA+ domains 4 and 5.

The structure of the dynein heavy chain differs in the presence and absence of ATP or ADP. During the power stroke, which corresponds to the release of ADP, the angle between the stem and stalk decreases, and a linker segment between the head (ring of six subdomains) and the stem is repositioned (Fig. 35-107). Thus, as in other motor proteins, conformational changes in the nucleotide-binding domain are transmitted to an attached elongated structure, here the stalk, which thereby functions as a lever arm to move the attached cargo (microtubule) relative to the rest of the motor protein.

j. Cilia Serve Important Functions

A genetic defect in axonemal dynein causes **primary ciliary dyskinesia (PCD)**; also called **immotile cilia syndrome**. This disorder is characterized by recurring respiratory infections resulting from the inability of the ciliated cells lining the airways (Fig. 35-101) to clear bacteria-laden mucus. Moreover, males with this condition are infertile because their sperm are immotile (females with PCD may also have fertility problems due to the impairment of the cilia that sweep the ovum through the oviduct to the uterus, although this is a matter of debate). About half of all cases of PCD are accompanied by **situs inversus**, a largely benign condition in which the right-left positions of the body's organs are inverted. In normal (**situs solitus**) individuals, cells in the gastrula (an early embryonic development stage; Section 34-4A) have cilia that rotate in the counterclockwise direction. This movement affects the distribution of morphogens that direct organ development so that, for example, the heart and stomach develop on the left and the liver on the right. If ciliary movement is absent or impaired, the right-left axis is established at random, accounting for the 50% incidence of situs inversus.

I. Bacterial Flagella

The final aspect of biological motility that we shall consider is the nature of bacterial flagella. These remarkable propulsive organelles operate via true rotary motion; *they are propellers rather than bending or contractile devices*.

Many species of bacteria, *E. coli*, for example (Fig. 1-3b), propel themselves through solution via the action of one or a few flagella in order to reach favorable environments or avoid unfavorable environments. Bacterial flagella, however, are entirely different from those of eukaryotes in both structure and chemistry. To begin with, bacterial flagella are only ~ 240 Å in diameter, less than the width of a single microtubule, and contain no tubulin. They consist of at least 24 different proteins, all of which

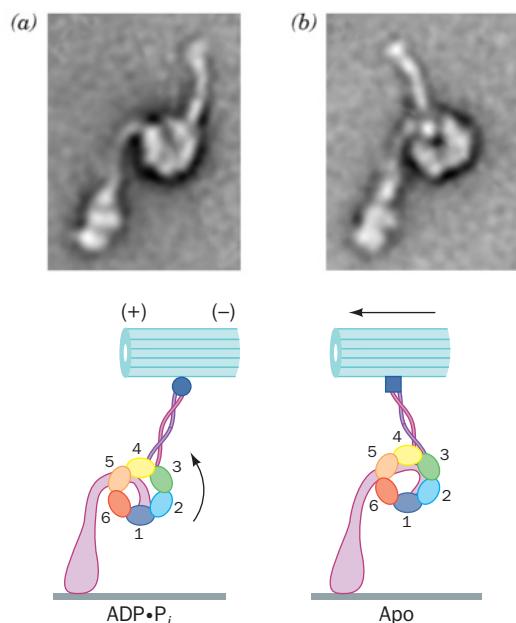


Figure 35-107 Conformational changes in dynein. The images are based on electron micrographs of a dynein heavy chain from the flagellum of the alga *C. reinhardtii* and are accompanied by interpretive drawings. The heavy chains are oriented with their tails pointing down and their stalks pointing up from the central ring of six AAA+ subdomains. (a) Dynein in the presence of ADP and VO_4^{3-} , which together mimic ADP + P_i . This conformational state corresponds to that following ATP hydrolysis, just before the power stroke. (b) Dynein in the absence of a nucleotide. This conformational state corresponds to that immediately following the power stroke. Comparison of Parts a and b indicates that the conformational change has rotated the ringlike head and stalk domains relative to the tail, thereby moving the stalk to the left by ~ 150 Å. [Courtesy of Stan Burgess, University of Leeds, U.K.]

are present in multiple copies ranging from a few to tens of thousands. Electron micrographs (Fig. 35-108) indicate that bacterial flagella consist of three major segments (Fig. 35-109):

1. The **flagellar filament**, their most prominent segment, is a tightly coiled helix up to 15 μm long (several times the length of the cell from which it emanates) that consists of tens of thousands of ~ 490 -residue subunits of the protein **flagellin** (also called **FliC**) ending in a pentameric cap of **FliD**.

2. The **flagellar hook**, which is assembled from ~ 120 subunits of the ~ 400 -residue **hook protein (FlgE)**, forms a short, curved structure to which the flagellar filament is attached.

3. The **basal body**, which in gram-negative bacteria penetrates the outer membrane, the peptidoglycan cell wall, and the inner (plasma) membrane, consists of a rod connecting several ringlike structures that anchor the flagellar hook to the bacterium. It is composed of ~ 20 different proteins totaling several hundred subunits.

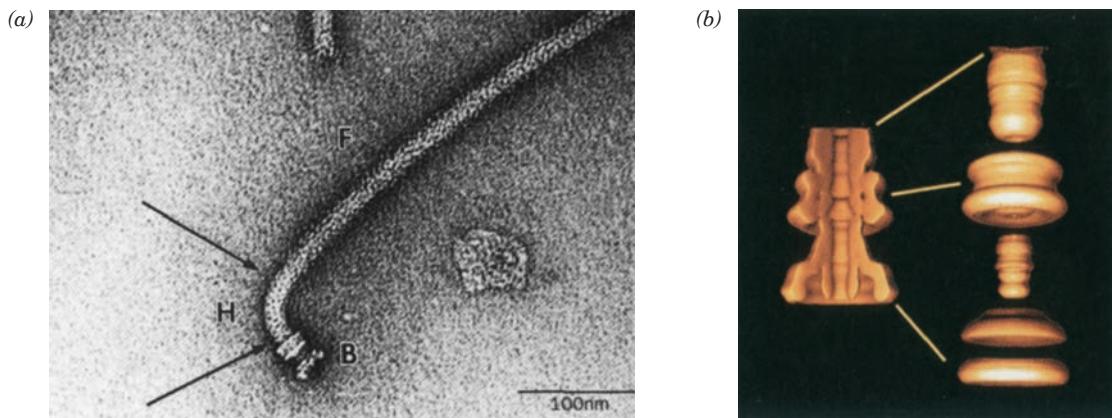


Figure 35-108 Electron micrographs of *Salmonella* flagella. (a) An intact flagellum. Its helical filament (F), which extends to the upper right, is connected via its hook (H) to its four-ringed basal body (B). [Courtesy of Robert McNab, Yale University.] (b) A close-up of the basal body, as generated by image reconstruction from electron micrographs, showing it in cutaway

(left) and exploded (right) views. The top line points at the hook–filament junction, the middle line points at the L–P ring complex, and the bottom line points at the MS ring in the center of the motor. [Courtesy of Noreen Francis and David DeRosier, Brandeis University.]

a. The Flagellar Filament Is a Hollow Helix

The structure of the *Salmonella typhimurium* flagellar filament, as Namba determined by cryo-EM to a resolution of ~ 4.5 Å, reveals that it is 240 Å in diameter with a 20-Å-diameter hollow core. Each flagellin subunit consists of four linearly connected domains, D0, D1, D2, and D3, that are arranged from the inside to the outside of the flagellar filament to form a Γ-shaped molecule that is 140 Å in height with a 100-Å-long arm (Fig. 35-110a). Flagellin subunits

assemble to form a left-handed helix with ~ 5.5 subunits per turn and a pitch of 25.6 Å in which domains D0 and D1 form a double-layered tube from which domains D2 and D3 project radially (Fig. 35-110b, c).

Flagella grow by adding flagellin subunits at the end away from the cell body (the distal end). The subunits, which of course are ribosomally synthesized in the bacterial cytosol, diffuse through the flagellum's hollow core, thereby permitting them to reach the tip without being lost

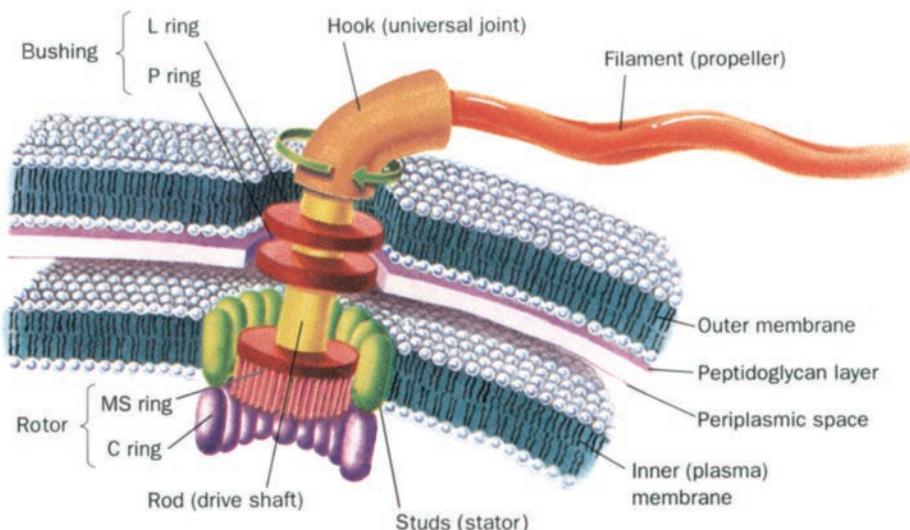


Figure 35-109 Structure of the gram-negative bacterial flagellum. This rotational motor consists of the basal body's studs and the C and MS rings. The studs are anchored to the plasma membrane, whereas the C and MS rings, which are attached via the rod and hook to the flagellar filament, rotate freely in the plasma membrane. Torque is generated through the

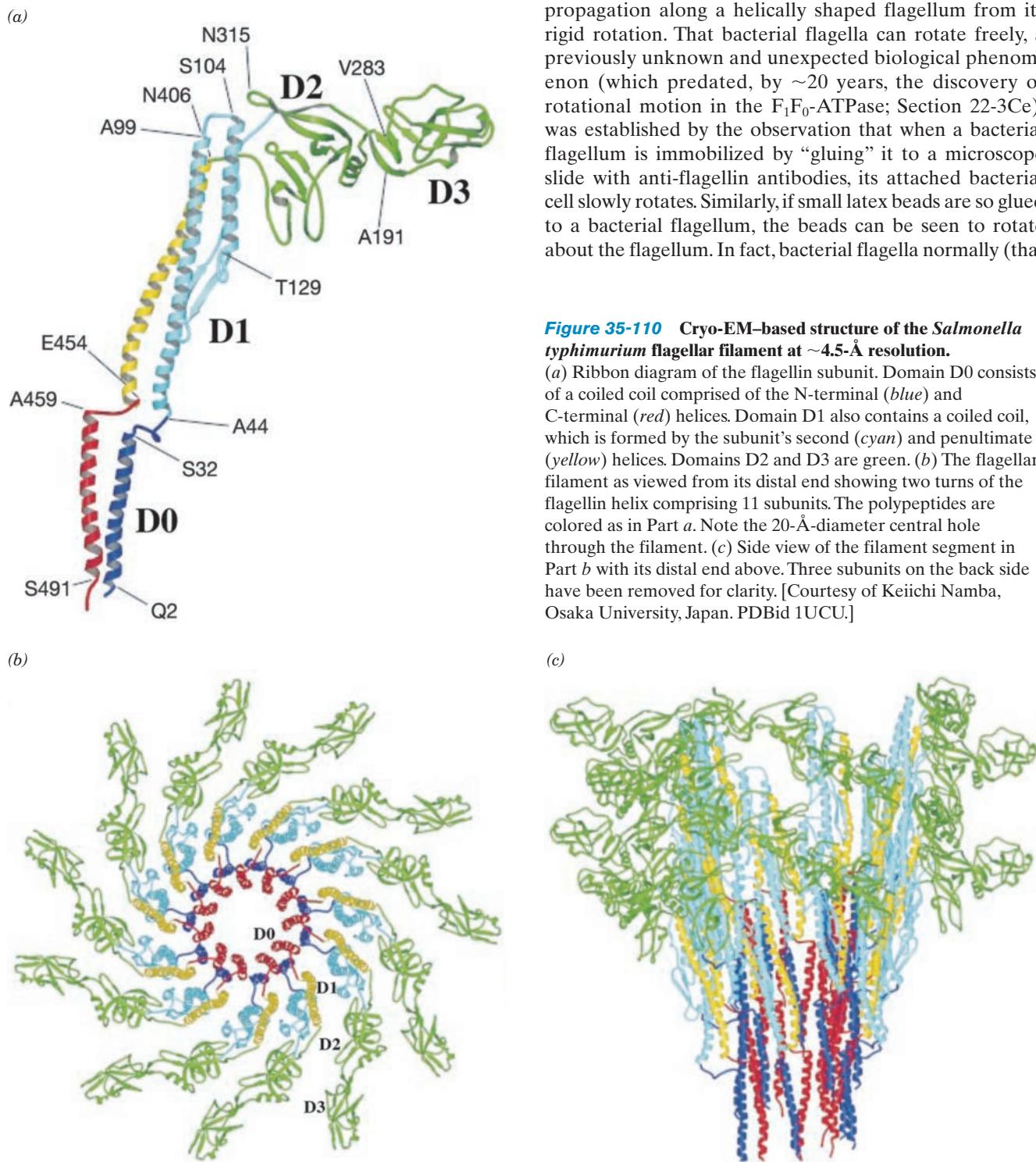
interaction of this motor with the electrochemical proton gradient across the plasma membrane. Gram-positive bacteria, which lack an outer membrane, also lack the bushing formed by the L and P rings. [After Schuster, S.C. and Khan, S., *Annu. Rev. Biophys. Biomol. Struct.* **23**, 524 (1994).]

to the medium. The flagellum's hollow core is too narrow to contain fully folded flagellin subunits but it seems likely that it can accommodate partially unfolded flagellin subunits. The inner surface of the flagellar core is lined with mainly polar side chains, which would facilitate the fast diffusion of partially unfolded subunits with their exposed hydrophobic residues. This further suggests that when the subunits reach the flagellum's distal end, the wider space

formed by its outer tube (Fig. 35-110c) together with its pentameric FliD cap provides a protected chamber (an Anfinsen cage; Section 9-2Cd) in which the subunits complete their folding.

b. Bacterial Flagella Rotate

Microscopic observations of swimming bacteria indicate that they are driven by what appear to be flagellar undulations. These undulations are an illusion; one cannot, by only watching through a light microscope, distinguish wave propagation along a helically shaped flagellum from its rigid rotation. That bacterial flagella can rotate freely, a previously unknown and unexpected biological phenomenon (which predated, by ~20 years, the discovery of rotational motion in the F_1F_0 -ATPase; Section 22-3Ce), was established by the observation that when a bacterial flagellum is immobilized by “gluing” it to a microscope slide with anti-flagellin antibodies, its attached bacterial cell slowly rotates. Similarly, if small latex beads are so glued to a bacterial flagellum, the beads can be seen to rotate about the flagellum. In fact, bacterial flagella normally (that



is, with a small viscous load) rotate at \sim 1000 revolutions per second. Moreover, the direction of rotation can be reversed nearly instantaneously.

What is the nature of the flagellar motor? It cannot be located in either the flagellar filament or the flagellar hook since both flagellin and hook protein have no demonstrable enzymatic activity. The basal body must therefore form the rotary motor. If this is so, then it must have the same mechanical elements as other rotary devices: a rotor (the rotating element) and a stator (the stationary element). Indeed, as Figs. 35-108b and 35-109 indicate, the ringlike structures of the basal body form just such elements. The rotor is comprised of the C and MS rings (it was originally thought that the MS ring was formed by two separate rings, M and S, but is now known to be formed mainly from two domains of the same protein, **FliF**); the stator is formed by the **studs** (so called because they appear as a ring of \sim 10 particles in freeze-fracture preparations), which are anchored to the plasma membrane and are comprised of the proteins **MotA** and **MotB** (Mot for *motility*); and the L and P rings form a bushing (sleeve) through which the rotating rod penetrates the bacterial outer membrane. The flagellar hook and filament are chemically passive elements that mechanically convert rotary motion to linear thrust in the manner of a propeller.

A final question we shall ask is what makes the flagellar rotor rotate? One might guess, in analogy with muscles and cilia, that the MS ring contains an ATPase that acts as a

mechanical transducer. However, the observation that bacterial swimming is unaffected by drastic reductions of the bacterial ATP pool requires that this hypothesis be abandoned. Rather, *the driving force behind flagellar rotation is the electrochemical proton gradient across the plasma membrane* (some flagellar motors are similarly driven by a Na^+ ion concentration gradient, notably those from certain marine bacteria and bacteria that live at high pH). This phenomenon was first demonstrated in *E. coli* mutants that lack an active F_1F_0 -ATPase (which catalyzes the proton gradient-driven synthesis of ATP; Section 22-3C) and are therefore unable to generate a proton gradient under anaerobic conditions (when a functional F_1F_0 -ATPase works in reverse). These mutant bacteria can only swim under aerobic conditions, whereas wild-type *E. coli* can also swim under anaerobic conditions.

The rotation of a flagellum requires the downhill transport of at least 1000 protons/turn (a small fraction of a bacterium's total proton circulation). Consequently the flagellar motor has a power output that is two to three orders of magnitude greater than other known biological motors. The transduction of the proton gradient's free energy to mechanical work occurs via the passage of protons through the channel-forming studs' subunit MotA in a way that allows the protons to interact with suitably disposed charges on the **FliG** protein, which is located at the base of the MS ring. This mechanism is similar to that proposed for the F_1F_0 -ATPase (Section 22-3Cf).

CHAPTER SUMMARY

1 Blood Clotting Blood coagulation occurs through a bifurcated cascade of proteolytic reactions, which result in the formation of cross-linked fibrin clots. Fibrin, which is formed from fibrinogen by the thrombin-catalyzed proteolytic excision of this soluble precursor's fibrinopeptides, self-associates to form a fibrous meshwork of half-staggered elongated molecules. The clot is later cross-linked by the formation of intermolecular Gln-Lys isopeptide bonds in a reaction catalyzed by the thrombin-activated fibrin-stabilizing factor (FSF or XIII_a).

Thrombin, which like all the proteolytic clotting factors is a serine protease homologous to trypsin, is synthesized as a zymogen that is proteolytically activated by a complex of activated Stuart factor (X_a) with activated proaccelerin (V_a). Thrombin, as do four other clotting factors as well as proteins C and S, contains several γ -carboxyglutamate (Gla) residues whose posttranslational synthesis from Glu residues requires a vitamin K cofactor. The Gla residues chelate Ca^{2+} ions whose presence is required for prothrombin activation together with a phosphatidylserine-containing membrane surface. The Gla side chains function to anchor the proteins of which they are components to the head groups of anionic membranes via Ca^{2+} bridges. Although thrombin resembles trypsin, it has a much deeper active site cleft, thereby partially accounting for its high substrate specificity. Thrombin has a Na^+ -binding site which, when unoccupied, converts thrombin to an anticoagulant state that has reduced specificity for

fibrinogen but has slightly greater affinity for protein C, which proteolytically cleaves V_a . Thrombin also proteolytically activates protease-activated receptors (PARs) embedded in platelet membranes, thereby activating the platelets.

Factor X may be activated through either the intrinsic pathway or the extrinsic pathway. The extrinsic pathway begins with the proteolytic activation of proconvertin (VII) in its complex with tissue factor (TF, III) by either Hageman factor (XII_a) or thrombin. The resulting TF- VII_a complex proteolytically activates X in the presence of Ca^{2+} ion and anionic phospholipid membrane. TF is a membrane glycoprotein that occurs in most tissues but not in those lining blood vessels. Hence an injury that exposes blood plasma to TF initiates blood clotting. Atherosclerotic plaques contain TF, and hence their rupture precipitates the inappropriate formation of clots, which can cause heart attacks and strokes.

In the intrinsic pathway, XI_a proteolytically activates Christmas factor (IX). The resulting XI_a , in complex with thrombin-activated antihemophilic factor (VIII_a), proteolytically activates X. Both of these reactions only occur on an anionic phospholipid membrane surface in the presence of Ca^{2+} ion. Hemophilias are hereditary bleeding disorders caused by defects in certain clotting factors, notably VIII (hemophilia A) and Christmas factor (hemophilia B). Clotting may also be initiated by the contact system in which XII, in the presence of high molecular weight kininogen (HMK), is activated by adsorption to a negatively charged surface such as glass to proteolyze

prekallikrein to kallikrein. Kallikrein then reciprocally proteolyses XII, which, in the presence of HMK, activates plasma thromboplastin antecedent (PTA or XI), which in turn, activates the intrinsic pathway by activating IX.

Clot formation is inhibited by numerous physiological mechanisms, including the inactivation of all of the clotting system proteases but VII_a by the binding of antithrombin III (ATIII). The presence of heparin, a highly anionic sulfated glycosaminoglycan that occurs in the mast cells lining certain blood vessels, greatly enhances this complexation. Similarly, protein C, which is proteolytically activated at endothelial cell surfaces by thrombin, inactivates V_a and VIII_a, thus inhibiting further prothrombin activation. Protein C is recruited to the endothelial cell surface by endothelial cell protein C receptor (EPCR), whereas thrombomodulin recruits thrombin to the cell surface and in doing so converts it to its anticoagulant form. The resulting activated protein C (APC) binds to its cofactor protein S on appropriate cell surfaces.

Once wound healing is underway, clots are dismantled by the serine protease plasmin in a process termed fibrinolysis. Plasmin is formed by the proteolysis of plasminogen by several serine proteases, notably urokinase and tissue-type plasminogen activator (t-PA). This process is limited through the inhibition of plasmin by α_2 -antiplasmin.

2 Immunity The adaptive immune response, which is conferred by lymphocytes in association with lymphoid tissues such as the thymus and the lymph nodes, results from cellular and humoral immunity. Cellular immunity, which guards against virally infected cells, fungi, parasites, cancers, and foreign tissue, is mediated by T cells, whereas humoral immunity, which is most effective against bacterial infections and extracellular phases of viral infections, is mediated by antibodies produced by B cells. A T cell is selected for proliferation if its T cell receptor (TCR) simultaneously binds a foreign antigen and the host's Class I MHC protein. Some of the progeny, cytotoxic T cells, bind to antigen-bearing host cells and kill them by inserting pore-forming proteins in their plasma membranes and inducing them to undergo apoptosis. A B cell displaying an antibody that binds to a foreign antigen is similarly selected for proliferation, and its plasma cell progeny secrete large amounts of that antibody. The immune system must be self-tolerant; failure to prevent the proliferation of T cells bearing TCRs and B cells bearing antibodies that bind self-antigens results in autoimmune diseases.

Antibodies (immunoglobulins) are glycoproteins that consist of two identical light (L) chains and two identical heavy (H) chains which, in turn, each have a constant (C) region and a variable (V) region. The five classes of secreted immunoglobulins, IgA, IgD, IgE, IgG, and IgM, have different physiological functions and vary only in the identities of their H chains, whose differences occur in the antibodies' Fc segments. The L chain in all of these immunoglobulins is either a κ or a λ chain. An immunoglobulin's two identical Fab segments, which each consist of one C and one V domain from both the H and the L chains, contain the antibody's antigen-binding sites. The antigen-binding specificity of an immunoglobulin is largely dependent on the sequences of the hypervariable segments from both its H and its L chains that line its antigen-binding site, thereby forming its six complementarity determining regions (CDRs). X-ray studies of IgGs reveal that their L and H chains are folded, respectively, into

two and four domains that each have the characteristic immunoglobulin fold.

The immune system generates a virtually unlimited variety of antigen-binding sites through both somatic recombination and somatic mutation. The κ light chain is encoded by four gene segments known as the leader (L_{κ}), variable (V_{κ}), joining (J_{κ}), and constant (C_{κ}), segments. B cell differentiation requires several cell generations. Early in this process, one of the ~ 40 $L_{\kappa} + V_{\kappa}$ units contained in the embryonic human κ chain gene recombines with one of the gene's five J_{κ} segments. Then, in later cell generations, the resulting $L_{\kappa}-V_{\kappa}-J_{\kappa}$ unit is transcribed and selectively spliced to the κ chain pre-mRNA's single C_{κ} segment yielding κ chain mRNA. The V/J recombination joint in this process is not precisely defined leading to further variation in the κ chain. Heavy chain genes are similarly assembled but, in addition, have a D segment between their V_H and J_H segments that leads to even greater heavy chain diversity. This $V(D)J$ joining process is initiated by RAG protein, an endonuclease that recognizes the recombination signal sequences (RSSs) between two randomly selected segments to be joined and excises the intervening DNA. The resulting two DNA segments are then joined, in an error prone process, through the mediation of Ku protein and DNA ligase IV together with several other proteins. Somatic mutation provides yet more diversity: Nucleotides may be added at random at the V_H/D_H and D_H/J_H joints through the action of terminal deoxynucleotidyl transferase and, furthermore, both heavy and light chain genes in B cell progenitors are subject to somatic hypermutation. B cells express only one heavy chain and one light chain allele, a phenomenon named allelic exclusion, thereby ensuring that each cell expresses only one species of immunoglobulin. Differentiating B cells switch from the synthesis of membrane-bound IgM, whose heavy chain has a hydrophobic transmembrane tail, to a secreted IgM with the same antigen specificity but lacking this C-terminal polypeptide. This switch occurs via the selection of alternative polyadenylation sites located before or after the exons specifying the transmembrane tail. B cell progeny also progressively switch from the synthesis of IgM to other classes of immunoglobulins, a process known as isotype switching, through either selective splicing or somatic recombination leading to the expression of alternative C_H segments.

T cell receptors (TCRs) resemble immunoglobulins in that they consist of α and β chains that have constant and variable domains, each of which assumes the immunoglobulin fold. TCR diversity is generated by the $V(D)J$ joining of different V , J , and, for α chains, D regions, thereby yielding six CDRs, much as occurs in immunoglobulins.

The major histocompatibility complex (MHC) encodes a highly polymorphic group of membrane-bound proteins that act as individuality markers among members of the same species (Class I MHC proteins) and differentiate immune system cells from other body cells (Class II MHC proteins). MHC proteins have domains that structurally resemble those in immunoglobulins and TCRs, and therefore the genes encoding all these proteins form a gene superfamily. An 8- or 9-residue polypeptide fragment of the processed antigen is presented to TCRs in complex with the Class I MHC protein through its binding in the cleft formed by the Class I MHC protein's α_1 and α_2 domains. Class II MHC proteins have very similar structures but their bound peptides extend out from both ends

of the binding groove. TCRs contact the MHC-peptide complexes only via their CDRs at the ends of their V_{α} and V_{β} domains. However, helper T cells also contact Class I MHC proteins displayed by antigen-presenting cells via a cell surface coreceptor known as CD4, and cytotoxic T cells similarly interact with Class II MHC proteins via the coreceptor CD8. TCRs and their coreceptors that are binding an MHC-antigen complex activate the T cell to proliferate through the mediation of CD3 and the nonreceptor tyrosine kinases Lck and ZAP70. The MHC gene's polymorphism is thought to prevent pathogens from evolving antigens that interact poorly with a particular MHC protein during the antigen recognition process. The strong interactions between TCRs and nonself MHC proteins is largely responsible for allograft rejection.

The complement system, an essential component of the innate immune system, defends against foreign invaders by killing foreign cells through complement fixation, inducing the phagocytosis of foreign particles (opsonization), and triggering local acute inflammatory reactions. The complement system consists of ~ 30 plasma proteins that interact in the antibody-dependent classical pathway, and in the antibody-independent MB-lectin and alternative pathways, which defend against invading microorganisms before an effective adaptive immune response can be mounted. The classical pathway contains three sequentially activated protein complexes: The recognition unit, which assembles on cell surface-bound antibody-antigen complexes; the activation unit, which amplifies the recognition process through a proteolytic cascade involving a series of serine proteases; and the membrane attack complex (MAC), which punctures the antibody-marked cell's plasma membrane, causing cell lysis and death. The MB-lectin pathway functions similarly but recognizes the mannose and fucose residues on the surfaces of bacteria. The alternative pathway also induces the assembly of MACs but in a series of reactions that are activated by the presence of certain bacterially synthesized polymers, whole bacteria, and host cells that are infected by certain viruses. The complement system is tightly regulated by the structural instabilities of several activated complement proteins, by the degradation of complement components through the actions of specific proteases such as C4b-binding protein, and by the sequestering of complement components by their specific binding of proteins such as C1 inhibitor and S protein.

3 Motility: Muscles, Vesicle Transport, Cilia, and Flagella

Skeletal muscle fibers consist of banded myofibrils which are, in turn, comprised of interdigitated thick and thin filaments. The thick filaments are composed almost entirely of myosin, a dimeric protein with two globular heads and an elongated rod-like segment comprising two α helices in a coiled coil. The myosin molecules aggregate end to end in a regular staggered array to form the bipolar thick filament. Myosin's elongated S1 head contains an ATPase. The thin filaments consist mainly of actin, a globular protein (G-actin) that polymerizes to form a helical filament (F-actin) in which each monomer unit is capable of binding a single myosin head. The thin filament also contains two other major proteins, tropomyosin and troponin. Tropomyosin is a heterodimeric protein that consists entirely of a coiled coil of two α helices that is wound in the grooves of the F-actin helix. Troponin consists of three subunits: TnC, a Ca^{2+} -binding calmodulin homolog; TnI, which binds actin; and TnT, which binds tropomyosin. The troponin-tropomyosin

complex regulates muscle contraction by varying the access of myosin heads to their actin-binding sites in response to the concentration of Ca^{2+} . Other, less-abundant proteins function in muscle assembly and stabilization including α -actinin, desmin, vimentin, CapZ, tropomodulin, titin, nebulin, and dystrophin.

Structural studies indicate that the thick and thin filaments slide past each other during muscle contraction. Tension is generated through a six-part ATP-driven reaction cycle in which myosin heads on thick filaments attach to actin monomers on thin filaments, change conformation, and then release the thin filament. Repeated such cycles cause the numerous myosin heads on a thick filament to "walk" up the adjacent thin filaments, resulting in muscle contraction.

Muscle contraction is triggered by an increase in $[\text{Ca}^{2+}]$. The Ca^{2+} binds to the TnC subunit of troponin, with the resulting conformational change causing tropomyosin to move deeper into the thin filament groove, thereby exposing actin's myosin head-binding sites, and thus switching on muscle contraction. The Ca^{2+} is released from the sarcoplasmic reticulum (SR) in response to a nerve impulse, which renders the SR membrane permeable to Ca^{2+} via voltage-gated Ca^{2+} channels. The cytosolic $[\text{Ca}^{2+}]$ is otherwise maintained at a very low level through the action of SR membrane-bound Ca^{2+} -ATPases, which pump the Ca^{2+} into the SR thus terminating muscle contraction.

Smooth muscle, which is responsible for long-lasting and involuntary contractions, lacks the banded pattern of skeletal muscles. Its myosin heads only interact with actin when their regulatory light chains are phosphorylated at a specific Ser residue. Smooth muscle contraction is nevertheless triggered by Ca^{2+} because myosin light chain kinase, the enzyme that catalyzes the phosphorylation of myosin's regulatory light chain, is active only when associated with Ca^{2+} -calmodulin. Myosin light chain phosphatase hydrolyzes myosin's activating phosphate group, so in the absence of active myosin light chain kinase, smooth muscle relaxation ensues. Nerve impulses increase the permeability of the smooth muscle cell plasma membrane to Ca^{2+} , which acts as an intracellular second messenger in stimulating smooth muscle contraction. Smooth muscles also respond to hormones such as epinephrine through the intermediacy of cAMP, whose presence activates protein kinase A to phosphorylate myosin light chain kinase, which reduces its affinity for binding Ca^{2+} -calmodulin.

Actin is also prominent in nonmuscle cells, where it has both structural and functional roles. Nonmuscle actin is generally in a steady state involving its monomeric G-actin form and polymeric F-actin microfilaments and hence undergoes treadmilling. The assembly, disassembly, and branching of microfilaments, as influenced by the presence of actin-binding proteins such as profilin, ADF/cofilin, and Arp2/3, play important roles in cellular motility, phagocytosis, and the extension and retraction of cellular protuberances.

There are 18 known classes of myosins. Conventional myosins (type II myosins), in addition to being a major component of striated muscle, form thick filaments that, in concert with microfilaments, participate in intracellular contractile processes such as the tightening of the contractile ring during cytokinesis. Many two-headed unconventional myosins, such as myosin V, function to transport vesicles along microfilament tracks toward the (+) end of the microfilament via a highly

processive hand-over-hand mechanism. However, myosin VI travels in the opposite direction.

Microtubules are formed from the protein tubulin, an $\alpha\beta$ dimer, that polymerizes to form shallowly spiraling rows of α and β subunits in which the β subunits then hydrolyze their bound GTP. Microtubules undergo treadmilling and dynamic instability as controlled by microtubule-associated proteins. Several antimitotic drugs, including colchicine, inhibit microtubule formation, whereas taxol prevents tubulin polymerization.

Kinesins and dyneins are microtubule-associated motor proteins. Kinesin motor domains have sequences in common with those of myosins. Kinesin-1 is a two-headed protein that functions to processively transport cargo along microtubule tracks in the (+) direction via a hand-over-hand mechanism. Although both myosin and kinesin motors are powered by ATP hydrolysis, they do so via different mechanisms. Dyneins also transport cargo along microtubule tracks but in the (-) direction. They are very large proteins that consist of one to three heavy chains and numerous smaller sized chains. In cilia and eukaryotic

flagella, microtubules are arranged in a 9 + 2 array in which 9 double microtubules surround 2 single microtubules in an assembly that is cross-linked by three types of proteins. Subfibers A of the outer fibers each bear two dynein arms that “walk” up neighboring subfibers B in an ATP-powered process. However, the cross-links between neighboring fibers prevent these fibers from sliding past each other; rather, the cilia bend, which accounts for their whiplike or undulatory motion.

Bacterial flagella, which are responsible for bacterial propulsion, are entirely different from eukaryotic flagella. Bacterial flagella consist of a flagellin filament comprised of tens of thousands of subunits of flagellin, a flagellar hook made of hook protein, and a complex basal body that is embedded in the bacterial plasma membrane. The basal body is a true rotary motor. In contrast, the flagellar filament and hook are passive elements that, like a propeller, convert the rotary motion of the basal body to linear thrust. The basal body’s rotary motion is directly powered by the discharge of the metabolically generated electrochemical proton gradient across the plasma membrane.

REFERENCES

Blood Clotting

- Arni, R.K., Padmanabhan, K., Padmanabhan, K.P., Wu, T.-P., and Tulinsky, A., Structures of the noncovalent complexes of human and bovine prothrombin fragment 2 with human PPACK-thrombin, *Biochemistry* **32**, 4727–4737 (1993).
- Berkner, K.L., The vitamin K-dependent carboxylase, *Annu. Rev. Nutr.* **25**, 127–149 (2005).
- Bode, W., Turk, D., and Karshikov, A., The refined 1.9-Å crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human α -thrombin: Structure analysis, overall structure, electrostatic properties, detailed active site geometry, and structure-function relationships, *Protein Sci.* **1**, 426–471 (1992).
- Colman, R.W., Marder, V.J., Clowes, A.W., George, J.N., and Goldhaber, S.Z., (Eds.), *Hemostasis and Thrombosis* (5th ed.), Lippincott Williams & Wilkins (2006). [An authoritative compendium of articles on many aspects of blood clotting and clot lysis.]
- Coughlin, S.R., Thrombin signaling and protease-activated receptors, *Nature* **407**, 258–264 (2000).
- Esmon, C.T., Regulation of blood coagulation, *Biochim. Biophys. Acta* **1477**, 349–360 (2000).
- Esmon, C.T., The protein C pathway, *Chest* **124**, 26–32 (2003).
- Furie, B. and Furie, B.C., Mechanisms of thrombus formation, *New Engl. J. Med.* **359**, 938–949 (2008).
- Kollman, J.M., Pandi, L., Sawaya, M.R., Riley, M., and Doolittle, R.F., Crystal structure of human fibrinogen, *Biochemistry* **48**, 3877–3886 (2009).
- Mengwasser, K.E., Bush, L.A., Shih, P., Cantwell, A.M., and Di Cera, E., Hirudin binding reveals key determinants of thrombin allostery, *J. Biol. Chem.* **280**, 26997–27003 (2005).
- Pineda, A.O., Carrell, C.J., Bush, L.A., Prasad, S., Caccia, S., Chen, Z.-W., Mathews, F.S., and Di Cera, E., Molecular dissection of Na^+ binding to thrombin, *J. Biol. Chem.* **279**, 31842–31853 (2004).
- Rydel, T.J., Tulinsky, A., Bode, W., and Huber, R., Refined structure of the hirudin-thrombin complex, *J. Mol. Biol.* **221**, 583–604 (1991).
- Seshadri, T.P., Tulinsky, A., Skrzypczak-Jankun, E., and Park, C.H., Structure of bovine prothrombin fragment I refined at 2.25 Å resolution, *J. Mol. Biol.* **220**, 481–494 (1991); and Soriano-

Garcia, M., Padmanabhan, K., de Vos, A.M., and Tulinsky, A., The Ca^{2+} ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1, *Biochemistry* **31**, 2554–2566 (1992).

Spraggon, G., Everse, S.J., and Doolittle, R.F., Crystal structures of fragment D from human fibrinogen and its crosslinked counterpart from fibrin, *Nature* **389**, 455–462 (1997).

Stamatoyannopoulos, G., Majerus, P.W., Perlmuter, R.M., and Varmus, H. (Eds.), *The Molecular Basis of Blood Diseases* (3rd ed.), Chapters 18–24, Saunders (2001). [Authoritative discussions of hemostasis, including vitamin K-dependent proteins, protein C, platelets, hemophilia, fibrin, and fibrinolysis.]

Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>, Chapters 169–178.

Weisel, J.W., Fibrinogen and fibrin, *Adv. Prot. Chem.* **70**, 247–299 (2005).

Weisel, J.W., Nagaswami, C., and Makowski, L., Twisting of fibrin fibrils limits their radial growth, *Proc. Natl. Acad. Sci.* **84**, 8991–8995 (1987).

Yang, Z., Pandi, L., and Doolittle, R.F., The crystal structure of fragment double-D from cross-linked lamprey fibrin reveals isopeptide linkages across an unexpected D–D interface, *Biochemistry* **41**, 15610–15617 (2002).

Immunity

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Molecular Biology of the Cell* (5th ed.), Chapter 25, Garland Publishing (2008).
- Barber, L.D. and Parham, P., Peptide binding to major histocompatibility complex molecules, *Annu. Rev. Cell Biol.* **9**, 163–206 (1993).
- Bassing, C.H., Swat, W., and Alt, F.W., The mechanism and regulation of chromosomal V(D)J recombination, *Cell* **109**, S45–S55 (2002); and Jung, J. and Alt, F.W., Unraveling V(D)J recombination: insights into gene regulation, *Cell* **116**, 299–311 (2004).
- Bjorkman, P.J. and Parham, P., Structure, function, and diversity of Class I major histocompatibility complex molecules, *Annu. Rev. Biochem.* **59**, 253–288 (1990).
- Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger,

- J.L., and Wiley, D.C., Structure of human class I histocompatibility antigen, HLA-A2; and The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens, *Nature* **329**, 506–512 and 512–518 (1987).
- Black, F.L., Why did they die? *Science* **258**, 1739–1740 (1992). [Discusses why so many Amerinds died of infectious diseases.]
- Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C., Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1, *Nature* **364**, 33–39 (1993).
- Cochran, J.R., Aivazian, D., Cameron, T.O., and Stern, L.J., Receptor clustering and transmembrane signaling in T cells, *Trends Biochem. Sci.* **26**, 304–310 (2001).
- Davies, D.R. and Chacko, S., Antibody structure, *Acc. Chem. Res.* **26**, 421–427 (1993).
- Deng, L. and Mariuzza, R.A., Recognition of self-peptide–MHC complexes by autoimmune T-cell receptors, *Trends Biochem. Sci.* **32**, 500–508 (2007).
- Di Noia, J.M. and Neuberger, M.S., Molecular mechanisms of antibody somatic hypermutation, *Annu. Rev. Biochem.* **76**, 1–22 (2007).
- Englehard, V.H., Structure of peptides associated with Class I and Class II MHC molecules, *Annu. Rev. Immunol.* **12**, 181–207 (1994).
- Gaboriaud, C., Juanhuix, J., Gruez, A., Lacroix, M., Darnault, C., Pignol, D., Verger, D., Fontecilla-Camps, J.C., and Arlaud, G., The crystal structure of the globular head of complement protein C1q provides a basis for its versatile recognition properties, *J. Biol. Chem.* **278**, 46974–46982 (2003).
- Garcia, K.C. and Adams, E.J., How the T cell receptor sees antigen—a structural view, *Cell* **122**, 333–336 (2005).
- Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., and Wilson, I.A., An $\alpha\beta$ cell receptor structure and its orientation in the TCR–MHC complex, *Science* **274**, 209–219 (1996).
- Gellert, M., V(D)J recombination, in Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (Eds.), *Mobile DNA II*, Chapter 29, ASM Press (2002).
- Harris, L.J., Larson, S.B., Hasel, K.W., and McPherson, A., Refined structure of an intact IgG2a monoclonal antibody, *Biochemistry* **36**, 1581–1597 (1997).
- Hennecke, J. and Wiley, D.C., T cell receptor–MHC interactions up close, *Cell* **104**, 1–4 (2001).
- Hennecke, J. and Wiley, D.C., Structure of a complex of the human $\alpha\beta$ T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLS-DR4 (DRA*0101 and DRB1*0401): Insight into TCR cross-restriction and alloreactivity, *J. Exp. Med.* **195**, 571–581 (2002).
- Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twu-masi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J., and Greenwood, B.M., Common West African HLA antigens associated with protection from severe malaria, *Nature* **352**, 595–600 (1991).
- Janssen, B.J.C., Huijiga, E.G., Raaijmakers, H.C.A., Roos, A., Daha, M.R., Nilsson-Ekdahl, K., Nilsson, B., and Gros, P., Structures of complement component C3 provides insights into the function and evolution of immunity, *Nature* **437**, 505–511 (2005); and Janssen, B.J.C., Christodoulidou, A., McCarthy, A., Lambris, J.D., and Gros, P., Structure of C3b reveals conformational changes that underlie complement activity, *Nature* **444**, 213–216 (2006).
- Jung, D. and Alt, F.W., Unraveling V(D)J recombination: insights into gene regulation, *Cell* **116**, 299–311 (2004).
- Law, R.H.P., et al., The structural basis for membrane binding and pore formation by lymphocyte perforin, *Nature* **468**, 447–451 (2010).
- Lewis, S.M., The mechanism of V(D)J joining: Lessons from molecular, immunological, and comparative analyses, *Adv. Immunol.* **56**, 27–150 (1994).
- Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P., *Molecular Cell Biology* (6th ed.), Chapter 24 (2008).
- Mak, T.W. and Saunders, M.E., *Primer to the Immune Response*, Academic Press–Elsevier (2008).
- Mariuzza, R.A., Phillips, S.E.V., and Poljak, R.J., The structural basis of antigen–antibody recognition, *Annu. Rev. Biophys. Biophys. Chem.* **16**, 139–159 (1987). [The structure of a lysozyme–Fab complex.]
- Marrack, P., Scott-Browne, J.P., Dai, S., Gapin, L., and Kappler, J., Evolutionarily conserved amino acids that control TCR–MHC interaction, *Annu. Rev. Immunol.* **26**, 171–203 (2008).
- Mostoslavsky, R., Alt, F.W., and Rajewsky, J., The lingering enigma of the allelic exclusion mechanism, *Cell* **118**, 539–544 (2004).
- Murphy, P., Travers, P., and Walport, M., *Janeway's Immunobiology* (7th ed.), Garland Science (2008).
- Papavasiliou, F.N. and Schatz, D.G., Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity, *Cell* **109**, S35–S44 (2002).
- Parham, P., *The Immune System* (3rd ed.), Garland Science (2009).
- Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D., The biochemistry of somatic hypermutation, *Annu. Rev. Immunol.* **26**, 481–511 (2008).
- Rudolph, M.G., Stanfield, R.L. and Wilson, I.A., How TCRs bind MHCs, peptides, and coreceptors, *Annu. Rev. Immunol.* **24**, 419–466 (2006).
- Stavnezer, J., Guikema, J.E.J., and Schrader, C.E., Mechanism and regulation of class switch recombination, *Annu. Rev. Immunol.* **26**, 261–292 (2008).
- Tonegawa, S., Somatic generation of immune diversity, *Angew. Chem. Int. Ed. Engl.* **27**, 1028–1039 (1988). [A Nobel lecture.]
- Wang, J., Meijers, R., Yi, X., Liu, J., Sakihama, T., Zhang, R., Joachimiak, A., and Reinherz, E.L., Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule, *Proc. Natl. Acad. Sci.* **98**, 10799–10804 (2001).
- Wang, R., Natarajan, K., and Margulies, D.H., Structural basis of the CD8 $\alpha\beta$ /MHC class I interaction: focused recognition orients CD8 β to a T cell proximal position, *J. Immunol.* **183**, 2554–2564 (2009).
- Werlen, G., Hausmann, B., Naehe, D., and Palmer, E., Signaling life and death in the thymus: Timing is everything, *Science* **299**, 1859–1863 (2003).

Motility

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Molecular Biology of the Cell* (5th ed.), Chapter 16, Garland Publishing (2008).
- Berg, H.C., The rotary motor of bacterial flagella, *Annu. Rev. Biochem.* **72**, 19–54 (2003).
- Bugyi, B. and Carlier, M.-F., Control of actin filament treadmilling in cell motility, *Annu. Rev. Biophys.* **39**, 449–470 (2010).
- Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J., and Oiwa, K., Dynein structure and power stroke, *Nature* **421**, 715–718 (2003); and Burgess, S.A., and Knight, P.J., Is the dynein motor a winch? *Curr. Opin. Struct. Biol.* **14**, 138–146 (2004).
- Clark, K.A., McElhinny, A.S., Beckerle, M.C., and Gregorio, C.C.,

- Striated muscle cyoarchitecture: An intricate web of form and function, *Annu. Rev. Cell. Dev. Biol.* **18**, 637–706 (2002).
- Craig, R. and Woodhead, J.L., Structure and function of myosin filaments, *Curr. Opin. Struct. Biol.* **16**, 204–212 (2006).
- Dawson, J.F., Sablin, E.P., Spudich, J.A., and Fletterick, R.J., Structure of an F-actin trimer disrupted by gelsolin and implications for the mechanism of severing, *J. Biol. Chem.* **278**, 1229–1238 (2003).
- Downing, K.H. and Nogales, E., Cryoelectron microscopy applications in the study of tubulin structures, microtubule architecture, dynamics and assemblies, and interaction of microtubules with motors, *Methods Enzymol.* **483**, 121–142 (2010).
- Endow, S.A. and Barker, D.S., Processive and nonprocessive models of kinesin movement, *Annu. Rev. Physiol.* **65**, 161–175 (2003).
- Engle, A.G. and Franzini-Armstrong, C. (Eds.), *Myology*, 3rd ed., Chapters 7, 8, 9, and 13 (2004).
- Fujii, T., Iwane, A.H., Yanagida, A., and Namba, A., Direct visualization of secondary structures of F-actin by electron cryomicroscopy, *Nature* **467**, 724–728 (2010); and Murakami, K., Yasunaga, T., Noguchi, T.Q.P., Gomibuchi, Y., Ngo, K.X., Uyeda, T.Q.P., and Wakabayashi, T., Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release, *Cell* **143**, 275–287 (2010).
- Gennerich, A. and Vale, R.D., Walking the walk: how kinesin and dynein coordinate their steps, *Curr. Opin. Cell Biol.* **21**, 59–67 (2009).
- Houdusse, A., Szent-Györgyi, A.G., and Cohen, C., Three conformational states of scallop myosin S1, *Proc. Natl. Acad. Sci.* **97**, 11238–11243 (2000).
- Kikkawa, M., The role of microtubules in processive kinesin movement, *Trends Cell Biol.* **18**, 128–135 (2008).
- Kikkawa, M., Sablin, E.P., Okada, Y., Yajima, H., Fletterick, R.J., and Hirokawa, N., Switch-based mechanism of kinesin motors, *Nature* **411**, 439–445 (2001).
- Kojima, S. and Blair, D.F., The bacterial flagellar motor: Structure and function of a complex molecular machine, *Int. Rev. Cyt.* **233**, 93–134 (2004).
- Kozielski, F., Sack, S., Marx, A., Thormählen, M., Schönbrunn, E., Biou, V., Thompson, A., Mandelkow, E.-M., and Mandelkow, E., The crystal structure of dimeric kinesin and implications for microtubule-dependent motility, *Cell* **91**, 985–994 (1997).
- Lee, L.K., Ginsburg, M.A., Crovace, C., Donohoe, M., and Stock, D., Structure of the torque ring of the flagellar motor and the molecular basis for rotational switching, *Nature* **466**, 996–1000 (2010).
- Li, H., DeRosier, D.J., Nicholson, W.V., Nogales, E., and Downing, K.H., Microtubule structure at 8 Å resolution, *Structure* **10**, 1317–1328 (2002).
- Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., and Matsudaira, P., *Molecular Cell Biology* (6th ed.), Chapters 17 and 18 (2008).
- Löwe, J., Li, H., Downing, K.H., and Nogales, E., Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution, *J. Mol. Biol.* **313**, 1045–1057 (2001).
- Mandelkow, E. and Johnson, K.A., The structural and mechanochemical cycle of kinesin, *Trends Biochem. Sci.* **23**, 429–433 (1998). [Summarizes the motor mechanism of kinesin.]
- Nogales, E., Structural insights into microtubule function, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 397–420 (2001).
- Oda, T., Iwasa, M., Aihara, T., Maeda, Y., and Narita, A., The nature of the globular- to fibrous-actin transition, *Nature* **457**, 441–445 (2009).
- Pollard, T.D., Regulation of actin filament assembly by Arp2/3 complex and formins, *Annu. Rev. Biophys.* **36**, 451–477 (2007).
- Rayment, I., Rypniewski, W.R., Schmidt-Bäse, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., and Holden, H.M., Three-dimensional structure of myosin subfragment-1: A molecular motor, *Science* **261**, 50–58 (1993); and Rayment, I., Holden, H.M., Whittaker, M., Yohn, C.B., Lorenz, M., Holmes, K.C., and Milligan, R.A., Structure of the actin–myosin complex and its implications for muscle contraction, *Science* **261**, 58–65 (1993).
- Reisler, E. and Egelman, E.H., Actin structure and function: what we still do not understand, *J. Biol. Chem.* **282**, 36133–36137 (2007).
- Roberts, A. J., et al., AAA+ ring and linker swing mechanism in the dynein motor, *Cell* **136**, 485–495 (2009).
- Robinson, R.C., Turbedsky, K., Kaiser, D.A., Marchand, J.-B., Higgs, H.N., Choe, S., and Pollard, T.D., Crystal structure of Arp2/3 complex, *Science* **294**, 1679–1684 (2001).
- Sablin, E.P. and Fletterick, R.J., Coordination between motor domains in processive kinesins, *J. Biol. Chem.* **279**, 15707–15710 (2004).
- Schliwa, M. (Ed.), *Molecular Motors*, Wiley–VCH (2003); and Schliwa, M. and Woehlke, G., Molecular motors, *Nature* **422**, 759–765 (2003). [Reviews of myosin, kinesin, and dynein.]
- Schröder, D.J., Jahn, W., Holden, H., Rayment, I., Holmes, K.C., and Spudich, J.A., Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1, *Nature* **364**, 171–174 (1993).
- Sindelar, C.V. and Downing, K.H., An atomic-level mechanism for activation of the kinesin molecular motors, *Proc. Natl. Acad. Sci.* **107**, 4111–4116 (2010).
- Spudich, J.A., The myosin swinging cross-bridge model, *Nature Rev. Mol. Cell Biol.* **2**, 387–392 (2001).
- Squire, J.M. and Parry, D.A.D. (Eds.), *Fibrous Proteins: Muscle and Molecular Motors*, *Adv. Prot. Chem.* **71**, Elsevier (2005). [A collection of detailed reviews.]
- Svitkina, T.M. and Borisy, G.G., Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia, *J. Cell Biol.* **145**, 1009–1026 (1999).
- Sweeney, H.L. and Houdesse, A., Structural and functional insights into the myosin motor mechanism, *Annu. Rev. Biophys.* **39**, 539–557 (2010).
- Terashima, H., Kojima, S., and Homma, M., Flagellar motility in bacteria: structure and function of the flagellar motor, *Int. Rev. Cell Mol. Biol.* **270**, 39–85 (2008).
- Thomas, D.R., Francis, N.R., Xu, C., and DeRosier, D.J., The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of *Salmonella enterica* serovar typhimurium, *J. Bacteriol.* **188**, 7039–7048 (2006).
- Trinick, J., Titin and nebulin: Protein rulers in muscle? *Trends Biochem. Sci.* **19**, 405–409 (1994).
- Vinogradova, M.V., Stone, D.B., Malanina, G.G., Karatzaferi, C., Cooke, R., Mendelson, R.A., and Fletterick, R.J., Ca^{2+} -regulated structural changes in troponin, *Proc. Natl. Acad. Sci.* **102**, 5038–5043 (2005).
- Walker, M.L., Burgess, S.A., Sellers, J.R., Wang, F., Hammer, J.A., III, Trinick, J., and Knight, P.J., Two-headed binding of a processive myosin to F-actin, *Nature* **405**, 804–807 (2000).
- Wang, H., Robinson, R.C., and Burtnick, L.D., The structure of native G-actin, *Cytoskeleton* **87**, 456–465 (2010).
- Whitby, F.G., Kent, H., Stewart, F., Stewart, M., Xie, X., Hatch, V., Cohen, C., and Phillips, G.N., Jr., Structure of tropomyosin at 9 angstroms resolution, *J. Biol. Chem.* **227**, 441–452 (1992).
- Xu, C., Craig, R., Tobacman, L., Horowitz, R., and Lehman, W., Tropomyosin positions in regulated thin filaments revealed by cryoelectron microscopy, *Biophys. J.* **77**, 985–992 (1999).

Yildiz, A., Forkey, J.N., McKinney, S.A., Ha, T., Goldman, Y.E., and Selvin, P.R., Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization, *Science* **300**, 2061–2065 (2003); and Yildiz, A., Tomishige, M., Vale, R.D., and Selvin, P.R., Kinesin walks hand over hand, *Science* **303**, 676–678 (2004).

Yonekura, K., Maki-Yonekura, S., and Namba, K., Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature* **424**, 643–650 (2003).

PROBLEMS

1. There is only one known symptom, in humans, of vitamin K deficiency. What is it?
2. Clotting is prevented in stored whole blood by mixing it with citrate. What is the function of the citrate?
3. Blood samples taken from patients with either hemophilia A or hemophilia B exhibit little tendency to clot. Yet, when these two types of blood are mixed, the resulting mixture has nearly normal clotting properties. Explain.
4. Explain why the following mixtures do not form a precipitate. (a) An Fab fragment with its corresponding antigen. (b) A hapten with the antibodies raised against it. (c) An antigen and its corresponding antibody when either component is in great excess.
5. Why do antibodies raised against a native protein sometimes fail to bind to the corresponding denatured protein?
6. Explain why: (a) Antibodies to antibodies raised against a particular enzyme occasionally specifically bind the enzyme's substrate. (b) Antibodies raised against a transition state analog of a particular reaction occasionally catalyze that reaction (such antibodies have been named **abzymes**).
7. **Gamma globulin** is the pooled globulin fraction from the blood plasma of a large, diverse human population. Explain how an injection of gamma globulin protects an individual against infectious diseases. Why does this protection last only several weeks?
8. The injection of bacterial cell wall constituents into an animal can trigger many of the symptoms caused by an infection including fever and inflammation. Explain how these symptoms are elicited.
- *9. When you hold a weight at arm's length, you are not doing any thermodynamic work but the muscles supporting the weight are nevertheless consuming energy. Describe, on the molecular level, how muscles might maintain such state of constant tension without contracting. Why does this state consume ATP?
10. When deprived of ATP, muscles assume a rigid and inextensible form known as the **rigor state** (after rigor mortis, the stiffening of the body after death). What is the molecular basis of rigor?
11. In nonmuscle cells, tropomyosin is associated with microfilaments that have a structural function but not with those that participate in contractile processes. Rationalize this observation.
12. Explain why the treatment of cells with colchicine results in the disappearance of their previously existing microtubules, even though colchicine does not cause microtubule dissociation *in vitro*.

INDEX

Page references in **bold face** refer to a major discussion of the entry. Positional and configurational designations in chemical names (e.g., 3-, α -, N -, *p*-, *trans*, *D*-, *sn*-) are ignored in alphabetizing. Numbers and Greek letters are otherwise alphabetical as if they were spelled out.

Note: Chapters 33–35, which are only available on our book companion website, www.wiley.com/college/voet, are separately indexed at this website.

- A**
A77, 1116
 AAA^+ domain, 1184–1185
 AAA^+ family of ATPases, 1197, 1206
 AAA^+ proteins, 1197–1198
 AADP^+ (3-aminopyridine adenine dinucleotide phosphate), 1125
A antigen, 415
aaRSs, *see* Aminoacyl-tRNA synthetase
Abasic sites, 1218
Abbé refractometer, 156
AB blood type, 22
ABC_A1 (ATP-cassette binding protein A1), 458
ABC transporters, **766–768**, 767F
Abe (abequose), 379F
Abelson, J., 1312, 1351
Abequose (Abe), 379F
 $\text{A}\beta$ (amyloid- β protein), 311–312
 $\text{A}\beta$ precursor protein (β PP), 311–312
Ab initio methods, 305
Abl, 708
A blood type, 22
ABO blood group system, 22, **414–415**
Abortive initiation, 1267–1268
Abrin, 1395T
Absolute configuration, 74–75
Absolute rate theory, 484–487
Absorbance:
 in Beer–Lambert law, 90
 of DNA and nucleic acids, 90, 92, 92F
Absorbance spectrum, 92
AC, *see* Adenylate cyclase
 ac^4C (N^4 -acetylcytidine), 1347F
ACAT, *see* Acyl-CoA:cholesterol acyltransferase
ACC (acetyl-CoA carboxylase), **962–964**
Accepted names, enzymes, 479
Acceptor, tRNA, 1346
Acceptor control, 862
Accessory pigments, 908–909
Accessory proteins:
 G proteins, 692
 GroEL/ES system, **294–302**
 peptidyl prolyl cis-trans isomerases, 292
 protein disulfide isomerases, **290–292**, 290F, 291F
Accommodation, 1388
ACE (angiotensin converting enzyme), 547F
ACE inhibitors, 547F
Acetal, 520F
Acetaldehyde, *re* and *si* faces, 77F
Acetaminophen, 544–545, 544F, 1000
Acetate, as cholesterol precursor, 975F, 976–977
Acetate ion, 45, 45T, 48F
Acetic acid:
 creation in Miller–Urey experiments, 32T
 distribution curve, 48, 48F
 titration curve, 47, 47F
Acetimidoquinine, 544–545, 544F
Acetoacetate, 959, 960F
 decarboxylation, 510–511
 and degradation of leucine/lysine, 1040–1041
 and degradation of phenylalanine/tyrosine, 1043–1047
 degradation of tryptophan, 1041–1042
Acetoacetyl-CoA, 960
 α -Aceto- α -hydroxybutyrate, 1075
Acetolactate, 1075
Acetone, 959
 and cell wall lysis, 130
 dielectric constant and dipole moment, 43T
 protein solubility in, 134
Acetonel, 679F
Acetyl-CoA, 806–807
Acetylcholine (ACh), 527, **778–780**
Acetylcholine receptor (AChR), **780–781**, 780F
Acetylcholinesterase, **781–782**, 782F
 diffusion-limited, 490, 491
 inactivation by DIPF, 527
 Michaelis–Menten kinetic constants, 489T
Acetyl-CoA, 561, 561F
 in citric acid cycle, 789, 792–796, 792F, 795F, 816–819
 and degradation of leucine/lysine, 1040–1041
 and integration of pathways, 1090
 in ketogenesis, 960–961, 960F
 pyruvate carboxylase regulation, 873–874
Acetyl-CoA acetyltransferase, 960–961
Acetyl-CoA carboxylase (ACC), **962–964**
Acetyl-coenzyme A, *see* Acetyl-CoA
 N^4 -Acetylcytidine (ac^4C), 1347F
 N -Acetyl-D-galactosamine, 392F
Acetyl-dihydrolipamide-E2, 795
 N -Acetylglucosamine, *see* NAG
 N -Acetylglutamate, 1025, 1029, 1070F, 1071
 N -Acetylglutamate-5-semialdehyde, 1070F, 1071
 N -Acetylglutamate synthase, 1029, 1070F, 1071
 N -Acetylimidazolium, 513
 ϵ -N-Acetyllysine, 79F
 N -Acetylmuramic acid, *see* NAM
 N -Acetylneurameric acid, *see* NANA
 N -Acetylneuraminate, 392F
Acetyl phosphate, 580, 607
Acetylsalicylic acid, 998
 N -Acetylsine, 79F
 O -Acetylsine, 1071
Acetyl transferase (AT), 967, 968
 N -Acetylxylosamine (XylNAc) residue, 523
ACh, *see* Acetylcholine
AChR, *see* Acetylcholine receptor
Acids, **45**
 dissociation constants, 45–47, 46T
 polyprotic, **48–50**, 49F
Acid-base buffers, 48
Acid-base catalysis, **506–510**
Acid-base indicators, 51
Acid-base reactions, **45–47**
 and amino acids, **72–73**
 for protein assays, 131
 and standard state conventions, 60
Acid chain hypothesis, 446
Acidic solutions, 47
Acidosis, 955
Ackee fruit, 948
Ackers, G., 353
Aconitase, 789, 790F, **808–809**, 808F
cis-Aconitate, 789
ACP (acyl-carrier protein), 961
ACP synthase, 961
Acquired immune deficiency syndrome, *see* AIDS
Acridine orange, 157
Acromegaly, 685
Acrosomal protease, 525T
Acrylamide, 147, 147F
ACTH, *see* Adrenocorticotropic hormone
Actin, 10, 10F, 301, 857
 in erythrocyte membranes, 413F, 414
 helical symmetry, 269
Actinomycin D, 79, 1272, 1272F
Action potentials, **775–777**, 775F, 776F
Activated complex, 485
Activation barrier, 486
Activators, 96, 351
Active, rolling mechanism (DNA unwinding), 1185–1186, 1186F
Active transport, 420, **747**
 ATP-driven, **758–768**
 ion-driven, **768–771**
Activity, 59–61
Activity coefficient, 61
Acute intermittent porphyria, 1056
Acute lymphoblastic leukemia, 1034
Acute pancreatitis, 537
Acyl-carnitine, 946
Acyl-carrier protein (ACP), 961
Acyl-CoA:cholesterol acyltransferase (ACAT), 455, 984
Acyl-CoA dehydrogenase, 948F

- Acyl-CoA dehydrogenase (AD), 947, 948
 Acyl-CoA oxidase, 958
 Acyl-CoA synthases, 945–946
 Acyl-dihydroxyacetone phosphate, 971
 1-Acylidihydroxyacetone phosphate, 1007
 Acyl-enzyme intermediate, 533
 2-Acylglycerol, 941
 1-Acylglycerol-3-phosphate acyltransferase, 971
 Acyl group transfer, 565
 Acyl phosphates, 580
 N-Acylsphingosine phosphocholine, 1008
 Acyltransferase (AT), 968
 AD, *see* Acyl-CoA dehydrogenase; Alzheimer's disease
 ADA, *see* Adenosine deaminase
ada gene, 1216
 Adair, G., 348
 Adair constants, 349
 Adair equation, **348–349**
 Ada protein, 1216
 Adapter proteins, *see* APs
 Adaptive enzymes, 1261
 Adaptor, **705–711**, 709
 Adaptor hypothesis, 1345
 ADAR1, 1150–1151, 1150F
 and RNA editing, 1322–1323, 1323F
 ADAR2:
 and RNA editing, 1322, 1323F
 Addison's disease, 681
 Adducin, 412F–413F
 Ade, *see* Adenine
 Adenine, 18, **83**, 83T. *See also* Watson-Crick
 base pairs
 and genetic code, 100T, 1343T
 IR spectra of derivatives, 1155F
 modified nucleosides of, 1347F
 and point mutations, 1339–1340
 Adenine phosphoribosyltransferase
 (APRT), 1114
 Adenohypophysis, 682
 Adenosine, 83T, 578, 1130F
 Adenosine-3',5'-cyclic monophosphate,
 see cAMP
 Adenosine-5'-(β-amino) diphosphate
 (AMPPN), 763, 763F
 Adenosine-5'-(β,γ-imido) triphosphate,
 see AMPPNP
 Adenosine-5'-(β,γ-methylene) triphosphate
 (AMPPCP), 763, 763F, 764
 Adenosine-5'-phosphosulfate, 183
 Adenosine-5'-phosphosulfate (APS)
 kinase, 1072
 Adenosine deaminase (ADA), 1130F, 1131
 Adenosine diphosphate, *see* ADP
 Adenosine monophosphate, *see* AMP
 Adenosine triphosphate, *see* ATP
 S-Adenosylhomocysteine, 1034
 S-Adenosylmethionine, *see* SAM
 Adenyl-3',5'-uridyl-3',5'-cytidyl-3',5'-
 guanylyl-3'-phosphate, 84F
 Adenylate cyclase (AC), 653, **697–698**, 1286
 Adenylate kinase (AK), 582
 and glycolysis control in muscle,
 627–628, 628F
 X-ray structure of porcine, 254, 255F
 Adenylic acid, *see* AMP
 Adenylosuccinate, 1112
 Adenylosuccinate lyase (PurB), 1109F,
 1111, 1112
 Adenylosuccinate synthetase, 1112F
 Adenyltransferase, 1069, 1071
 ADH, *see* Alcohol dehydrogenase;
 Antidiuretic hormone
 Adhesins, 382
 Adipocytes, 388–389, 388F, 1096, 1097F
 Adiponectin, 1096, 1097F
 Adiponectin receptors, 1096
 Adipose tissue, 389
 brown, 860
 glucose uptake regulation in, 751F
 and other organs, 1091F, 1093
 A-DNA, 1146F, 1147F, **1148–1149**, 1148T, 1151
 AdoCbl, *see* 5'-Deoxyadenosylcobalamin
 AdoMet, *see* SAM
 ADP (adenosine diphosphate), 17, 82
 in Calvin cycle, 931
 in catabolism, 561, 561F
 in citric acid cycle, 816–817
 in glycolysis, 594F, 595
 in GroEL/ES system, 295–302
 RuvB binding of, 1231
 and thermodynamics of life, 578–583
 ADP/ATP carrier, 771
 ADP-glucose, 645
 ADP-glucose pyrophosphorylase, 931
 ADPPNP, *see* AMPPNP
 ADP-ribosylated diphthalamide residue,
 eEF2, 1398
 ADP-ribosylation, 1398
 ADP-ribosylation factor, *see* ARF1
 Adrenal glands, 672F
 Adrenaline, *see* Epinephrine
 β-Adrenergic receptors, 664, 680
 Adrenocortical steroids, 680
 Adrenocorticotropic hormone (ACTH),
 673T, 682–683
 Adrenoleukodystrophy (ALD), 767, 958
 β-Adrenoreceptors, 660, 680
 Adriamycin, 1169
 Adsorption chromatography, 132T, 135,
 143–144
 Adult-onset diabetes, 1102–1103
Aequorea victoria, 120
 Aerobes, 6. *See also* specific organisms
 Aerobic fermentation, 594F
 Aerobic metabolism, 864–865
 Aerobic oxidation, 594F
 Affinity chromatography, 132T
 ion exchange, 136
 metal chelation, 145
 mRNA, 158
 for proteins, 118, 141–143, 141F
 Affinity labeling, 527
 Aflatoxin B₁, 1225
 AFM (atomic force microscopy), 376
 Africa, sickle-cell anemia in, 188, 343
 African clawed toad, 1202
 African sleeping sickness, 799
 Agard, D., 456
 Agarose gel electrophoretogram, 106F
 Agarose gels:
 in affinity chromatography, 142, 142F
 in gel electrophoresis, 147–149
 in gel filtration chromatography,
 139, 140T, 141
 in ion exchange chromatography, 137
 AGC family of protein kinases, 730
 A⁺ gene, 1261–1262, 1264
 a gene (fruit fly), 25F
 Agglutinin, 366
 Aggrecan, 373–375, 373T, 374F
 Aging, telomere length and, 1210–1211
 Aglycone, 366
 AGO, *see* Argonaute
 Agonist, 539, 680
 Agouti related peptide (AgRP), 1099
 Agrin, 373T
Agrobacterium radiobacter, 106F
 AgRP (agouti related peptide), 1099
 AICAR, *see* 5-Aminomidazole-4-
 carboxamide ribotide
 AICAR transformylase (PurH), 1109F, 1111
 AIDS (acquired immune deficiency
 syndrome), 123, **545–546**. *See also* HIV
 AIR, *see* 5-Aminomidazole ribotide
 AIR carboxylase, 1109F, 1111
 AIR synthetase (PurM), 1109F, 1110
 AK, *see* Adenylate kinase
 AKAPs (A-kinase anchoring proteins), 714
 Akey, C., 424
 A-kinase anchoring proteins (AKAPs), 714
 Akt, *see* Protein kinase B
 Akt1/PKB α , 734
 Akt2/PKB β , 734
 Akt3/PKB γ , 734
 Ala, *see* Alanine
 ALA (δ-aminolevulinic acid), 1048
 ALAD (δ-aminolevulinic acid
 dehydratase), 1048
 Alanine:
 biosynthesis, 1067–1073
 degradation of, 1030–1034
 and degradation of tryptophan,
 1041–1042
 genetic codes for, 100T, 1343T
 half-life, 1413T
 α helix/β sheet propensity, 302T, 304
 hydropathy scale for, 264T
 in Miller-Urey experiments, 32T
 in native unfolded proteins, 283
 nonessential amino acid, 1065T
 nonpolar side chain, 70
 Ramachandran diagram, 224F, 225
 structure and general properties, 68T, 71F
 (S)-Alanine, 76, 77F
 β-Alanine:
 creation in Miller-Urey experiments, 32T
 in pyrimidine catabolism, 1136
 L-Alanine, 76, 77F
 Alanine tRNA, 176, 1345
 AlaRS, 1351
 ALAS-1, 1056
 ALAS-2, 1056
 ALA synthase, 1056
 Ala-tRNA^{Cys}, 1359
 Albumin, 154F, 542, 677. *See also*
 Serum albumin
 Alcaptonuria, 569
 Alcohols:
 effect on protein denaturation, 265
 reactions to form hemiacetals and
 hemiketals, 361F
 Alcohol dehydrogenase (ADH), 479
 in fermentation, 618–619
 and methanol oxidation, 505
 molecular mass of, 140F
 stereospecificity of, 470–473
 Alcoholic fermentation, 470, 593–594,
 616–619, 616F
 Alcohol:NAD⁺ oxidoreductase, 479
 ALD, *see* Adrenoleukodystrophy
 Aldehydes, hemiacetal formation of, 361F
 Aldehyde dehydrogenase, 447F
 Alditols, 364
 Aldohexoses, 360F
 Aldolase, 568
 binding to erythrocyte membrane, 412
 in glycolysis, 596F, **600–603**, 600F, 602F
 half-life, 1408T
 molecular mass determination by gel
 filtration chromatography, 140F
 sedimentation coefficient, 154F
 Aldol condensation, 568
 Aldol histidine, 239F

- Aldonic acids, 364
 Aldopentoses, 360F
 Aldoses, 360F, 361
 Aldose-ketose interconversion, 567, 567F
 Aldosterone, 681
 Aldotetroses, 360F
 Aldotrioses, 360F
 ALD protein, 958
 Aledronate, 679F
al gene (fruit fly), 25F
 Alignment score (AS), 196
 Alkaptonuria, 25–26, **1045–1047**
 1-Alkyl-2-acyl-
 sn-glycerophosphoethanolamine, 1007, 1007F
 Alkylacylglycerophospholipids, 574, 1005, 1007–1008
 O⁶-Alkylguanine-DNA alkyltransferase, 1216
 O⁶-Alkylguanine residues, 1215, 1216
 1-Alkyl-*sn*-glycerol-3-phosphate, 1007
 Allantoic acid, 1134
 Allantoin, 1134
 Allantoinase, 1135F
 Alleles, **21**
 Allison, A., 187
 Allo forms (enantiomers), 75–76
 Allolactose, 97, 117
 1,6-Allolactose, 1261
 Allopurinol, 1135
 D-Allose, 360F
 Allosteric control:
 ATCase activity, 476–479
 gluconeogenesis, 878–879, 879F, 879T
 glycogen phosphorylase and glycogen synthase, **647–650**
 hemoglobin, **347–354**
 and metabolic flux, 620, 624
 D-*allo*-Threonine, 76, 76F
 L-*allo*-Threonine, 76, 76F
 Allotransplantation, 121
 Alloxanthine, 1135
 Allysine, 238
 Allysine aldol, 239F
 α_1 Adrenergic receptors, 680
 α_2 Adrenergic receptors, 680
 α/α Barrels, 982
 $\alpha\alpha$ Motifs, 249–250, 249F
 α/β Barrels, 250, 252, 254, 300
 α/β Domains, 250, 300
 α Configuration, of nucleotides, 83
 α Domains, 249F
 α Helix, 226, 226F, 227F
 globular proteins, 246
 helix forms vs., 228F
 physical basis of, 304
 Ramachandran diagram, 224F
 and space filling, 281
 α -Ketoglutarate:ferredoxin reductase, 815
 α -MSH (α -melanocyte stimulating hormone), 1099
 α Oxidation, of fatty acids, 958–959, 959F
 α Particles, 572
 α Solenoid, 433
 α -Synuclein, 720
 α -Tocopherol, 866
 Alternative splicing, 1317–1318
 Altman, S., 1330
 D-Altrose, 360F
 Altschul, S., 201
 AluI, 105T, 106, 109
 Alumina, 144
 Alzheimer's disease (AD), 309, **311–312**, 458–459, 720
Amanita phalloides (death cap), 1280
 α -Amanitin, 1280
amber codon, 1343
amber mutants, half-lives of, 1408
amber suppressor, 1362
 Ambros, V., 1326
 Ames, B., 1224
 Ames test, 1224–1225, 1225F
 Amethopterin, 492, 1129
 Amido black, 149F
 Amidophosphoribosyl transferase, 1108, 1109F, 1114
 Amines, biosynthesis of, **1058–1062**
 α -Amino acids, 67, 67F, 70F
 D-Amino acids, 73–75, 79
 L- α -Amino acids, 75, 75F, 226
 L-Amino acids, 73–75, 75F, 78
 Amino acids, 15, 67–80. *See also*
 Protein sequencing
 acid-base properties, **72**
 as biosynthetic precursors, 1047–1064
 chemical synthesis of polypeptides, 206–209
 classification and characteristics, **70–71**
 conformation angle distribution, 224F
 creation in Miller-Urey experiments, 32T, 33
 defined, 67
 energy recovery by citric acid cycle, 789
 essential and nonessential, 1019, 1065T
 genetic code, 99, 100T, 1338–1345, 1343T
 hydropathy scale for side chains, 264T
 as metabolic fuels, 1095–4–1095
 as neurotransmitters, 782–783, 783F
 nomenclature, **72–73**
 nonstandard, **78–80**, 79F, 80F
 optical activity, **73–78**
 peptide bonds, **70**
 and phylogenetic tree, 6
 of proteins, 67–73, 68T–69T
 in protein synthesis, 98–100
 racemization, 116
 side chains as covalent catalysts, 511
 transport in blood, 973
 Amino acid biosynthesis:
 alanine, 1067–1073
 arginine, 1071
 asparagine, 1067–1073
 aspartate, 1067–1073
 citric acid cycle intermediates, 818
 cysteine, 1071–1072
 essential amino acids, **1072–1078**
 GCN2 regulates, 1400–1401
 glutamate, 1065–1071
 glutamine, 1067–1073
 glycine, 1071–1072
 histidine, 1078
 integration and interrelationships of, 1090
 isoleucine, 1075
 leucine, 1075
 lysine, 1072–1073, 1075
 methionine, 1072–1073, 1075
 nonessential amino acids, **1064–1072**
 ornithine, 1071
 phenylalanine, 1075–1078
 proline, 1071
 serine, 1071–1072
 threonine, 1072–1073, 1075
 tryptophan, 1075–1078
 tyrosine, 1075–1078
 valine, 1075
 Amino acid deamination, **1019–1020**
 oxidative, **1023–1025**
 transamination, **1020–1023**
 Amino acid metabolism, 561F. *See also*
 Urea cycle
 alanine, cysteine, glycine, serine, and threonine degradation, 1030–1034
 amino acids as metabolic fuels, 1094–1095
 arginine, glutamate, glutamine, histidine, and proline degradation, 1034
 asparagine and aspartate degradation, 1034
 biosynthesis of psychologically active amines, **1058–1062**
 deamination of amino acids, 1019–1025
 glucogenic and ketogenic amino acids, 1029–1030
 heme biosynthesis, **1047–1058**
 heme degradation, **1056**
 integration and interrelationships with other pathways, 1090
 isoleucine, methionine, and valine degradation, 1034–1039
 leucine and lysine degradation, 1040–1041
 overview of common metabolic intermediates, 1029F
 phenylalanine and tyrosine degradation, 1043–1047
 tryptophan degradation, 1041–1042
 and urea cycle, 1025–1029
 D-Amino acid oxidase, 1025
 L-Amino acid oxidase, 1025
 Amino acid residues, 70, 188, 302T
 Amino acid sequencing, *see* Protein sequencing
 Amino acid stem, tRNA, 1346
 Aminoacrylate, 1031
 Aminoacrylate-PLP Schiff base, 1078
 Aminoacetylation, tRNA, 1345–1362
 Aminoacyl peptide, 170
 Aminoacyl site, *see* A site
 Aminoacyl-tRNA, 98, 1349, 1349F
 Aminoacyl-tRNA synthetase (aaRS), 99, **1349–1359**
 characteristics of, 1350F
 classes of, 1350–1351
 Class I, 1350–1354, 1350T, 1355F, 1358, 1358F
 Class II, 1350–1351, 1350T, 1354–1355, 1355F, 1358, 1358F
 and CysRS in archaeabacteria, 1359
 and Gln-tRNA^{Gln} formation, 1358–1359
 proofreading by, 1355–1358
 recognition by, 1351–1352, 1351F, 1352F
 γ -Aminobutyric acid (GABA), 79–80, 80F, **783**, 1049
 biosynthesis, 1058–1060
 α -Amino- β -chlorobutyrate, 1408
 Aminoglycosides, 1395
 Amino groups, 43F
 5-Aminoimidazole-4-carboxamide ribotide (AICAR), 1078, 1104, 1109F, 1111
 5-Aminoimidazole-4-(*N*-succinylcarboxamide) ribotide (SAICAR), 1109F, 1111
 5-Aminoimidazole ribotide (AIR), 1109F, 1110
 β -Aminoisobutyrate, 1136
 α -Aminoisobutyric acid, 32T
 δ -Aminolevulinate synthase, 1048, 1049
 δ -Aminolevulinic acid (ALA), 1048
 δ -Aminolevulinic acid dehydratase (ALAD), 1048
 Aminomethyltransferase (AMT), 1032
 α -Amino-*n*-butyric acid, 32T
 Aminopeptidases, 166, 168, 168T
 Aminopeptidase M, specificity of, 168T
 Aminoprocollagen peptidase, 1403–1404
 β -Aminopropionitrile, 239, 276
 Aminopterin, 1129
 2-Aminopurine, *see* 2AP

- 3-Aminopyridine adenine dinucleotide phosphate (AADP⁺), 1125
- Amino sugars, 365
- Amino terminus, *see N-terminus*
- Aminotransferases, 1020, 1136F
- Ammonia, 32, 33, 43T, 1025
- Ammonification, 1083
- Ammonium ion (NH₄⁺), 45, 45T, 47, 47F
- Ammonium persulfate, 146F
- Ammonium sulfate, 133F, 134
- Ammonotelic organisms, 1025
- Amoeba, 13F, 398T
- AMP (adenosine monophosphate), 83T
- in animal catabolism, 1130F
- and DNA ligase, 1187
- in glycogen metabolism, 648–650
- from IMP, 1112F
- synthesis of, 1111–1113
- and thermodynamics of life, 578
- AMP deaminase, 1131
- AMP-dependent protein kinase (AMPK), 963
- and cholesterol biosynthesis, 989
- in fatty acid metabolism, 973, 975
- and glycogen synthase cascade, 660
- and hypoxia, 864
- and metabolic homeostasis, 1095–1097, 1097F
- and noninsulin-dependent diabetes, 1103–1104
- Amphibolic pathways, 789
- Amphipathic materials, 44
- Amphiphilic materials, 43–44
- Ampholytes, 67, 151, 151F
- Amphoteric substances, 67
- Ampicillin, 26, 107F, 111, 377F, 378, 621
- AMPK, *see AMP-dependent protein kinase*
- AMPPCP, *see Adenosine-5'-*
- (β , γ -methylene) triphosphate
- AMPPNP, *see Adenosine-5'-(β -amino)*
- diphosphate
- AMPPNP (adenosine-5'-(β , γ -imido) triphosphate), 609, 633, 768, 1231
- Amprenavir, 550F
- amp^R gene, 107F, 111, 621
- AMT (aminomethyltransferase), 1032
- α -Amylase, 370, 469
- Amylin, 310
- Amylo-(1,4 → 1,6)-transglycosylase, 646, 667
- Amylo-1,6-glucosidase, 667
- β -Amyloid, 312
- Amyloids, 309–310
- Amyloid diseases, 310–311
- Amyloid fibrils, 309, 309F
- Amyloidoses, 309–310
- Amyloid plaques, 311–312, 458–459
- Amyloid- β protein (A β), 311–312
- Amylopectin, 369–370, 370F
- Amyloplast, 11F
- α -Amylose, 369, 369F
- Amytal, 831
- Anabaena, 4F
- Anabolism, 17, 561–562
- Anaerobes, 6. *See also specific organisms*
- Anaerobic autotrophs, 814
- Anaerobic fermentation, 594F, 614–619
- Anaerobic metabolism, 864–865
- Analbuminemia, 944
- Analytical ultracentrifuge, 154
- Anaphase, 20F, 21F
- Anaphase-promoting complex (APC), 1414
- Anaphylaxis, 1000
- Anaplerotic reactions, 818–819
- Anchoring proteins, 714
- Andersen's disease, 667
- Anderson, W. F., 122
- Androgens, 673T, 680, 991
- Anencephaly, 1035
- Anesthetics, neuronal membranes and, 398–399
- Anfinsen, C., 278
- Anfinsen cage, 298, 300
- Angelman syndrome (AS), 1251
- Angina pectoris, 686
- Angiosperms, 13F
- Angiotensin converting enzyme (ACE), 547F
- Angiotensin I, 547F
- Angiotensin II, 547F
- Angiotensinogen, 547F
- Anhydrides, formation of, 515T
- 8-Anilino-1-naphthalenesulfonate (ANS), 287
- Animals, 7F, 12, 13F. *See also specific organisms*
- in carcinogenesis tests, 1224
- divergence, 192
- DNA polymerases, 1202–1205, 1202T, 1203T
- FAS-I, 965–967, 966F
- metabolic homeostasis, 1096–1101
- organ specialization, 1090–1095
- Animal cells, 7F
- Animalia, 12
- Anion channel, 412, 413F
- Anion exchangers, 135, 989–990
- Ankyrin, 413F, 414
- Ankyrin repeats, 414, 414F
- Annealing conditions, 93
- Anomeric carbon, 361
- Anomeric forms, cyclic sugars, 361–363
- ANS (8-anilino-1-naphthalenesulfonate), 287
- Antagonist, 539, 680
- Antenna chlorophylls, 905–906, 906F
- Anterograde transport, 428
- Anthrax, 714–715
- Anthrax toxin, 714–715
- Antibiotics, protein synthesis inhibition and, 1395–1398
- Antibiotic-resistant transposons, 1242
- Antibodies, 131, 150. *See also Monoclonal antibodies*
- Anticodons, 98, 1330, 1346, 1389F
- Anticodon arm, 1346
- Anticodon stem-loop (ASL), 1394, 1394F
- Anti conformation, 1152F
- Antidiuretic hormone (ADH), 683
- Antifolates, 1129–1130
- Antigens, 376
- Antigenic groups, in bacterial cell walls, 378–379
- Antimycin, 831, 849
- Antioxidants, 865–866
- Antipain, 1409
- Antiparallel β pleated sheets, 224F, 229–230, 232
- Antiport, 758F
- Antirrhinum, 22
- Antisense oligodeoxynucleotide, 1402–1403
- Antisense oligonucleotides, 1402–1403
- Antisense RNA, 1323, 1402
- Antisense strand, 1266
- Antisnorkeeling, 406
- 2AP (2-aminopurine), 1339, 1339F
- AP1, 435–436
- AP2, 435–436, 436F
- AP3, 436
- AP4, 436
- Ap₅A, 628
- ApA, melting curve of, 1157F
- APC (anaphase-promoting complex), 1414
- AP endonuclease, 1218
- Aphidicolin, 1203
- Apical domain, 410
- ApoA-I, *see Apolipoprotein A-I*
- ApoA-II, *see Apolipoprotein A-II*
- Apocytochrome c, 448
- ApoE, *see Apolipoprotein E*
- apoE4 gene, 312
- Apoenzyme, 473–474, 1023
- Apoferitin, 140F
- Apolipoproteins, 449–451, 986
- Apolipoprotein A-I (apoA-I), 450–451, 450F, 450T, 451F, 458
- Apolipoprotein A-II (apoA-II), 450T, 458
- Apolipoprotein B-48, 450T, 452, 1322
- Apolipoprotein B-100, 449F, 450T, 452, 455, 1322
- Apolipoprotein C-I, 450T
- Apolipoprotein C-II, 450T, 451
- Apolipoprotein C-III, 450T
- Apolipoprotein D, 450T
- Apolipoprotein E (apoE), 312, 450T, 453, 455–456, 456F
- Apolipoprotein E2, 458–459
- Apolipoprotein E3, 458, 459
- Apolipoprotein E4, 458, 459
- Apomyoglobin, 173F, 284, 289
- Apoproteins, 449–451
- Apoptosis, 717
- Apparent dissociation constants, 348
- Appetite control, 1098, 1099, 1100F
- Applied Biosystems, 184
- A β precursor protein, 311–312
- APRT (adenine phosphoribosyltransferase), 1114
- AP (adapter protein), 435–436, 436F
- APS/Cbl complex, 738
- AP sites, 1218
- APS (adenosine-5'-phosphosulfate) reductase, 1072
- Aptamers, 214, 1300
- Apurinic sites, 1218
- Apyrase, 183
- Apyrimidinic sites, 1218
- AQP0, 756
- AQP1, 756–757, 757F, 787
- AQP12, 756
- Aquaglyceroporins, 756
- Aquaporins, 399F, 756–757, 757F, 787
- Aqueous solutions, 40–50
- acids, bases, and buffers, 45–50
- hydrocarbon transfer in, 262–264, 263T
- nucleotide base stacking in, 1156–1157
- polyprotic acids, 48–50
- and properties of water, 40–45
- araBAD operon, 1287, 1291–1294, 1292F, 1293F
- Arabidopsis thaliana*, 177T, 182, 1410F
- Arabinose, glucose and, 1286
- d-Arabinose, 360F
- L-Arabinose, 1291–1294
- AraC, 1292–1294
- Arachidic acid, 387T
- Arachidonate, 994
- Arachidonic acid, 387T, 994–997, 995F
- araI₁, 1292
- araO₂, 1292
- Arber, W., 104
- Archaea, 6, 7F, 13F, 192
- Archaeal introns, 1307T
- Archaeabacteria, 6, 1359
- Archaeoglobus fulgidus*, 177T
- Arcuate nucleus, 1099
- AREs (AU-rich elements), 1327
- ARF1 (ADP-ribosylation factor), 434–435, 435F, 695–696

- ARF GAP, 437
 ARF nucleotide-binding site opener (ARNO), 434
 Arg, *see* Arginine
 Arginine:
 biosynthesis, 571F, 1071
 degradation, 1034
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 α helix/ β sheet propensity, 302T
 isoelectric point, 72
 location in globular proteins, 246
 side chains, 71, 264T
 specific rotation, 74
 structure and general properties, 69T
 trypsin-catalyzed cleavage, 170
 Argininosuccinase, 1028
 Argininosuccinate synthetase, 1028
 Argonaute (AGO), 1324, 1325, 1325F
 ArgRS, 1352
 A-RNA, 1151
 ARNO (ARF nucleotide-binding site opener), 434
 Arnold, E., 1208
 Arnold, W., 906
 aroH operon, 1297
 Aromatic amino acid decarboxylase, 1060
 ar/R constriction, 756
 Arrhenius, A., 45
 Arsenate, 636
 Arsenic, 799, 822
 Arsenicals, 799
 Arsenite, 799
 ARSs, *see* Autonomously replicating sequences
 Arteriosclerosis, 456–458
 Arthropods, 13F
 AS, *see* Alignment score; Angelman syndrome
 Ascobolus, 13F
 Ascorbic acid (vitamin C), 236, 364–365, 364F
 A-side transfer, 472
 A (aminoacyl) site, 98F, 1373, 1385, 1387
 ASL, *see* Anticodon stem-loop
 Asn, *see* Asparagine
 Asp, *see* Aspartic acid
 L-Asparaginase, 1034
 Asparagine:
 biosynthesis, 1067–1073
 degradation, 1034
 genetic codes for, 100T, 1343T
 in globular proteins, 246–247
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 as nonessential amino acid, 1065T
 side chains, 71, 264T
 structure and general properties, 69T
 Asparagine synthetase, 1068
 Aspartame, 1087
 Aspartate, 71. *See also* Aspartic acid
 biosynthesis, 1067–1073
 degradation, 1034
 half-life, 1413T
 as nonessential amino acid, 1065T
 Aspartate aminotransferase, 876, 877F
 Aspartate transcarbamoylase, *see* ATCase
 Aspartic acid:
 degradation, 1034
 genetic codes for, 100T, 1343T
 in globular proteins, 246
 α helix/ β sheet propensity, 302T
 isoelectric point, 72
 in Miller-Urey experiments, 32T
 side chains, 71, 264T
 structure and general properties, 69T
 Aspartic proteases, 546–549, 548F
 Aspartokinase, 1072, 1075T
 β -Aspartyl-AMP, 1068
 Aspartyl phosphate, 759
 Aspartyl- β -phosphate, 1075
 Aspirin, 994, 998, 999F
 Aspirin-triggered *epi*-lipoxins (ATLs), 1004
 AspRS, tRNA^{Asp} and, 1354–1355, 1354F, 1355F
 Asp-tRNA^{Asn}, 1358
 Assembly factors, 1365
 Astbury, W., 233
 Asthma, 680
 Asx, 68T, 73. *See also* Asparagine; Aspartic acid
 Asymmetric centers, 74–76
 Asymmetric PCR, 116
 AT, *see* Acetyl transferase
 AT-AC introns, 1319
 AT-AC spliceosome, 1319
 Ataxin-1, 1252
 ATCase (aspartate transcarbamoylase), 475F, 478F
 allosteric control, 476–479
 feedback inhibition, 474–479
 in UMP synthesis, 1115F, 1116
 X-ray structure, 476F
 Atheromas, 456
 Atherosclerosis, 456–458
 Atherosclerotic plaques, 240, 457F
 ATLs (aspirin-triggered *epi*-lipoxins), 1004
 Atomic fluctuations, 307, 308
 Atomic force microscopy (AFM), 376
 Atorvastatin, 990, 991F
 ATP (adenosine triphosphate), 82, 559, 586.
 See also Electron transport; Oxidative phosphorylation; Photosynthesis and aerobic vs. anaerobic metabolism, 864–866
 in ATCase feedback inhibition, 475
 in Calvin cycle, 931
 in catabolism, 561, 561F
 in citric acid cycle, 789, 816–817, 819
 consumption of, 582–583
 coupled reactions with, 579F
 and DNA ligase, 1188
 effect on hemoglobin oxygen binding, 330
 in electron transport chain, 823–824, 824F, 828, 829
 formation, 582–583
 in gluconeogenesis, 877–878
 in glycogen metabolism, 648–649
 in glycolysis, 593, 594F, 595, 600, 608–610, 613
 in GroEL/ES system, 294, 298–300
 in mitochondria, 9
 in nitrogen fixation, 1082
 overview of roles, 17
 in oxidative phosphorylation, 845–863, 846F
 photorespiration dissipates, 935
 in photosynthesis, 902, 926
 production control for, 863–864, 863F
 in protein folding, 300
 source of free energy for metabolic pathways, 561F
 sources of, during exercise, 1092F
 structure, 578F
 and thermodynamics of life, 578–583
 in transcription, 95–96, 1265
 turnover rate, 583
 and vesicle fusion, 444–445
 ATP/ADP carrier, 448
 ATP-ADP translocator, 448, 768, 771
 ATPases:
 A-, F-, P-, and V-types, 758
 (Ca^{2+})-ATPase, 762–764
 (H^+ – K^+)-ATPase, 764–765
 (Na^+ – K^+)-ATPase, 758–762
 and molecular chaperones, 293
 ATP-cassette binding protein A1 (ABCA1), 458
 ATP-citrate lyase, 818, 968
 ATP-driven active transport, 758–768
 ATP mass action ratio, 862
 ATP_oS, 698
 ATP_yS, 294, 444–445, 1204F
 ATP sulfurylase, 183, 1072
 Atractyloside, 771
 Attenuation, 1296–1299
 Attenuator, 1297–1299
 A-type ATPases, 758
 AU-AC introns, 1307T, 1319
 AU-AC spliceosome, 1319
 AUG (codon), 1343
 AU-rich elements (AREs), 1327
 Aurora, 1402
 Autocrine hormones, 671
 Autoimmune disorders, 688
 cyclosporin A for, 292
 multiple sclerosis, 777
 and type 1 diabetes, 1102
 Autolysis, 131
 Autonomously replicating sequences (ARSs), 109, 1205
 Autophagic vacuoles, 1409
 Autophosphorylation, 670, 700
 Autoradiography, 94–95, 95F, 148, 572
 of sequencing gels, 177–178, 178F
 Autosomes, 23
 Autotrophs, 5
 Auxilin, 436
 Auxotrophic mutants, 570
 Avery, O., 86
 Avidin, 577, 873
 Avogadro's number, 53T
 Axial groups, in sugars, 363
 Axon, 771
 5-Azacytosine (5-AzaC) residue, 1249
 Azaserine, 80, 80F, 1142
 3'-Azido-3'-deoxythymidine (AZT; zidovudine), 546, 1207
 Azidomet, 135F
Azotobacter vinelandii, 1080F–1081, 1341
 AZT, *see* 3'-Azido-3'-deoxythymidine
 Azurin, 135F

B

- Babcock, G., 844
 Baboons, xenotransplantation with, 121
 BAC-ends (sequence-tagged connectors), 181
 Bacilli, 4
Bacillus, 4F
Bacillus subtilis, 376, 377F, 1284
Bacillus thuringiensis (Bt), 122
 Bacitracin, 888–889, 889F
 Backside attack, 514F
 BACs (bacterial artificial chromosomes), 108–109, 113, 114, 181
 Bacteria. *See also* specific organisms
 conjugation of, 1262–1264, 1262F
 divergence of, 192
 DNA base composition, 84
 DNA size, 94T
 and eukaryotic protein production, 118
 fatty acids of, 387
 genetic regulation of sugar transport, 765T, 766, 767F

- Bacteria (*Contd.*)
 genetics of, 26
 glutamine synthetase regulation, 1069F
 glycogen synthase, 645
 glycoproteins of, 375–379
 homologous recombination in, 1225
 microfossil, 29F
 phylogenetic tree for, 7F
 in prokaryotic classification, 6
 RNAP, 96, 1265–1266
 self-compartmentalized proteases of, 1419–1420
 spores of, *see* Spores
 virulence of, 375
- Bacterial artificial chromosomes, *see* BACs
- Bacteriochlorophyll *a* (BChl *a*), 135F, 903, 904F
- Bacteriochlorophyll *b* (BChl *b*), 903, 904F, 916
- Bacteriophages, 27, 1190
 DNA, 86–88
 RNAP, 1265
- Bacteriophage λ :
 cloning vector, 108, 108F, 111F
 DNA size, 94T
 electron micrograph, 108F
 helix–turn–helix motif, 1288
 template binding, 1264
- Bacteriophage 434, helix–turn–helix motif, 1288, 1289F
- Bacteriophage M13:
 cloning vector, 108, 108F
 replication, 1190, 1190F
- Bacteriophage MS2, sequencing, 176
- Bacteriophage P1, Cre recombinase, 1243–1244
- Bacteriophage P22, 1288
- Bacteriophage Φ X174:
 genetic map, 1344, 1344F
 Pol I, 1177
 replication, 1190–1193, 1192F
 template binding, 1266
- Bacteriophage SP01, 1284
- Bacteriophage T2, 1264, 1265F
 attachment to *E. coli*, 87F
 DNA, 94F, 94T
 in Hershey–Chase experiment, 87, 87F
- Bacteriophage T4, 27F, 28, 1264, 1265F
 cloning vector, 109
 DNA size, 94T
 FCO mutation, 1340
 template binding, 1266
- Bacteriophage T5, attachment of, 86F
- Bacteriophage T6, 94T, 1263F
- Bacteriophage T7, in chain-terminator method, 179
- Bacteriopheophytin *b* (BPheo *b*), 910–911
- Bacteriorhodopsin (BR), 850–851, 851F
 membrane structure, 402–403, 403F
- Baculovirus vectors, 108
- Baker, D., 305, 306, 310
- Baker's yeast, *see* *Saccharomyces cerevisiae*
- Balch, W., 437
- Baldwin, R., 285, 287
- Baltimore, D., 545, 1207
- Bamakan, 373T
- BamHI, 105T
- Ban, N., 966–968
- B antigen, 415
- Banting, F., 575, 1102
- Barber, J., 916
- Barford, D., 722, 723
- Barnett, J., 750
- Barrels:
 α/α , 982
- α/β , 250, 252, 254, 300
 β , 231F, 448–449
 Swiss roll, 252, 253F
 TIM, 252, 254F, 300
 up-and-down β , 251–252
- Barré-Sinoussi, F., 545
- Basal cells, 16F
- Basal laminae, 373
- Basal level, 1261
- Bases, 45–47. *See also* Nucleotide bases and buffers, 48
 nucleic acid, *see* Nucleic acid bases
- Base-catalyzed hydrolysis, RNA, 85, 85F
- Base excision repair (BER), 1216, 1218
- Base flipping, 1215
- Basement membranes, 373
- Base pairs, 18, 84, 1154–1156. *See also* Watson–Crick base pairs
 association constants for formation, 1155T
 complementary, 89
 non-Watson–Crick, 1154–1156, 1154F
 and recombination, 28
- Base stacking, in B-DNA, 89
- Basic helix–loop–helix/leucine zipper (bHLH/Zip), 987
- Basic Local Alignment Search Tool, *see* BLAST
- Basic solutions, 47
- Basolateral domain, 410
- Bassham, J., 927
- Batrachotoxin, 777
- Battersby, A., 1053
- Baumeister, W., 1414
- Bax, A., 656
- B blood type, 22
- BBP (branch point-binding protein), 1312, 1314
- BChl *a*, *see* Bacteriochlorophyll *a*
- BChl *b*, *see* Bacteriochlorophyll *b*
- BCKDH kinase, 1039
- BCKDH phosphatase, 1039
- BCM7, 534–535
- Bcr, 718
- B-DNA, 88F, 1145–1148, 1146F, 1147F
 conversion to Z-DNA, 1149F
 structure of ideal, 1148T
 three-dimensional structure, 89F
 Watson–Crick structure, 88–90
- Beadle, G., 26, 570
- Becker, J., 1002
- Beckmann, R., 422, 424
- Bed volume, in gel filtration chromatography, 138
- Beer–Lambert law, 90
- Begg, G., 171
- Behenic acid, 387T
- Beinert, H., 834
- Benign tumors, 703
- Benner, S., 472
- Benson, A., 927
- Benzamidine, 556
- Benzene, 43, 43T
- Benzer, S., 28
- Benzoinic acid, 945
- BER, *see* Base excision repair
- Berg, P., 109, 1350
- Berger, J., 1167–1168, 1188, 1195, 1275
- Bergström, S., 994
- Beriberi, 474T, 617–618
- Berman, H., 236, 1287
- Bernal, J. D., 241, 331
- Berry, E., 838, 840
- Berzelius, J., 469
- Best, C., 575
- Best, G., 1102
- β_1 -Adrenergic receptors, 690
- β_2 -Adrenergic receptors, 690
- β -Adrenergic receptor kinase (β ARK), 697
- $\beta\alpha\beta$ Motif, 249, 249F
- β ARK (β -Adrenergic receptor kinase), 697
- β -Arrestins, 697
- β Barrel, 231F, 448–449
- β Bends, 232, 233F
- β_c , 717–718
- β Cells, pancreatic, *see* Pancreatic β cells
- β Clamp, Pol III, 1182–1183, 1183F
- β Configuration, nucleotides, 83
- β Domains, 249–254, 249F
- β Hairpin motif, 249, 249F
- β Helix, 314
- β -Horseshoe, GyrA intein, 1407
- β -Lactamase, 378
- β Oxidation, fatty acids, 945, 947–950, 958
- β Particles, 572
- β Pleated sheets, 229–232, 229F–232F
 globular proteins, 246
- propensities and classifications of amino acid residues for, 302T
- Ramachandran diagram, 224F
 and space filling, 281
- β PP (β β precursor protein), 311–312
- β -Propeller, 692
- β Ribbons, 1290–1291
- β Sandwich, 251
- Beta structures, proteins, 229–232, 229F–232F
- β Subunit, Pol III, 1182–1183
- b* gene (fruit fly), 25F
- BglII, 105T
- BglIII, 105T
- BH₄ (5,6,7,8-tetrahydrobiopterin), 686, 1043–1044
- bHLH/Zip (basic helix–loop–helix/leucine zipper), 987
- Bi Bi reactions, 498–499
- Bibliome, 576
- Bicyclic cascade, 650, 651F
- Bidirectional replication, 1174, 1176
- Bienert, H., 808
- Bifurcated hydrogen bond, 261
- Bilayers, *see* Lipid bilayers
- Bile acids, 941, 992, 993F
- Bile pigments, 1056
- Bile salts, 941, 943–944, 989–990, 992–993
- Bilins, 909
- Bilirubin, 1056
- Bilirubin diglucuronide, 1056, 1057
- Bilirubin UDP-glucuronosyltransferase, 1056
- Biliverdin, 1056
- Bimolecular reactions, 483
- Binding modules, 705–711
- Binding specificity, protein purification and, 132
- Bioavailability, 542
- Biochemical literature, 34–36
- Biochemical review publications, 35T
- Biochemistry, 14–19
 biological structures, 14–17, 16F
 defined, 3
 expression and transmission of genetic information, 18–19
 genetics review, 19–28
 metabolic processes, 17
- Bioethics, 123
- Bio-Gel A, 139
- Bio-Gel P, 139
- Biogenic amines, 782
- Bioinformatics, 194
 globular proteins, 256–259
 phylogenetic tree construction, 203–205, 203F

- sequence alignment, 195–203
 sequence databases, 194–195
 Biological evolution, 29, 30F
 Biological membranes, *see* Membranes
 Biological structures, 14–17, 16F
 Biotin, 1043
 BIOSIS Previews, 35
 Biosynthesis, 561
 Biotin, 473T, 577, 857, 873
 Biotin carboxylase, 963
 Biotin carboxyl-carrier protein, 963
 Biotinyllysine, 873
 Bipolar disorder, 736
 Bishop, M., 705
 1,3-Bisphosphoglycerate, *see* 1,3-BPG
 2,3-Bisphosphoglycerate, *see* 2,3-BPG
 Bisphosphoglycerate mutase, 596F, 611, 611F
 2,3-Bisphosphoglycerate phosphatase, 596F, 611
 Bisphosphonates, 679, 679F
 Bisubstrate enzyme reactions, 497–501, 498F
 Bisulfite ion, 1249
 Bisulfite sequencing, 1249
 Black, J., 764
 Blackburn, E. H., 1210
 Black mouth, 342
 Blaese, M., 122
 BLAST (Basic Local Alignment Search Tool), 201, 202F
 Blastocyst, 1250
 Blobel, G., 420, 424, 1371
 Bloch, K., 961, 975, 976
 Block, S., 1273
 Blood, 973
 ABO blood groups system, 22, 414–415
 buffering, 48
 universal donors/recipients, 466
 Blood–brain barrier, 542, 783
 Blood clotting factors, recombinant, 119
 Blood glucose, 638, 661–664, 1094
 Blood groups, 22, 414–415
 Blood group determinants, 414–415
 Blood plasma, 323–324
 Blood vessels, 235
 BLOSUM45 substitution matrix, 201
 BLOSUM62 substitution matrix, 201, 203
 Blow, D., 527
 Blue-green algae, *see* Cyanobacteria
 Blundell, T., 549
 Blunt ends, 106
 Blunt end ligation, 1188
 Boat conformation, 363, 363F
 BOC-amino acid, 206
 BOC (*tert*-Butyloxycarbonyl) group, 206–208
 Bohm, A., 1303
 Bohr, C., 328
 Bohr, N., 329
 Bohr effect, hemoglobin, 328–329, 328F, 329, 340
 Boltzmann, Ludwig, 55
 Boltzmann constant (k_B), 53T, 55, 486
 Bonaparte, N., 799
 Bone, 235, 540
 Bongrekic acid, 771
 Boniva, 679F
 BoNT/A, 442
 BoNT/G, 442
 Borhani, D., 451
 Boron, 31
 Borsook, H., 1408
 Botulinus toxin, 780
 Botulism, 442, 1340
 Bovine 20S proteasome, 1417
 Bovine pancreatic ribonuclease A, 470.
See also RNase A
 acid-base catalytic mechanism, 508–510, 509F
 physical constants, 153T
 Bovine pancreatic trypsin inhibitor, *see* BPTI
 Bovine papillomavirus, E1 protein, 1184–1185, 1185F
 Bovine somatotropin (bST), 119
 Bovine spongiform encephalopathy (BSE), 312, 315
 Box C/D snoRNAs, 1329
 Box H/ACA snoRNAs, 1329
 Boyer, H., 107
 Boyer, P., 612, 845, 856, 857
 1,3-BPG (1,3-bisphosphoglycerate), 580, 596F, 607, 609
 2,3-BPG (2,3-bisphosphoglycerate):
 in glycolysis, 610
 and hemoglobin oxygen binding, 329–331, 329F–331F, 340–341
 BPheo b (bacteriopheophytin b), 910–911
 B protein, 1314
 BPTI (Bovine pancreatic trypsin inhibitor), 308–309, 533, 533F
 BR, *see* Bacteriorhodopsin
 Brachet, J., 1260
 Bradykinin, 208
 Bragg, L., 331
 Bragg, W., 331
 Brain, metabolic relationships, 1090–1091, 1091F
 Branch, phylogenetic trees, 203–204, 203F
 Branched-chain α -keto acid dehydrogenase, 1039
 Branching enzyme, 646, 667
 Branch migration, 1226, 1230–1234
 Branch points, phylogenetic trees, 203–204, 203F
 Branch point-binding protein, *see* BBP
 Brändén, C.-I., 256, 929
 Branton, D., 408
 Braunstein, A., 1021
 BRCA1, 1236
 BRCA2, 1236
 Breaker, R. R., 1300, 1318
 Breast cancer, 119–120, 1236
 Breathing motions, 306
 Breathnach, R., 1305
 Brenner, S., 1264, 1340, 1341
 Brewer's yeast, 617F
 Briggs, G. E., 488
 Brij 30, 399F
 Brij 35, 399F
 Brinster, R., 86
 British anti-lewisite, 822
 Brittle bone disease, 240
 Brodsky, B., 236
 Bromide ion, 45T
 5-Bromo-4-chloro-3-hydroxyindole, 110
 5-Bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal), 110
 Bromohydroxyacetone phosphate, 604
 5-Bromouracil, 1339, 1339F
 Bronsted, J., 45
 Bronsted acid, 45
 Bronsted base, 45
 Brouillette, C., 451
 Brown, A., 323, 488
 Brown, D., 1283
 Brown, M., 452, 453, 987
 Brown adipose tissue, 860
 Brown algae, 13F
 Browner, M., 1001
 Brown fat, 860
 Brownian ratchet, 428, 1280
 Bruice, T., 514, 515
 Brünger, A., 441, 444
 Brush border cells, 38, 768
 BSE, *see* Bovine spongiform encephalopathy
 B-side transfer, 472
 bST (bovine somatotropin), 119
 Bt (*Bacillus thuringiensis*), 122
 Bt corn, 122
 5-BU, *see* 5-Bromouracil
 Bubonic plague, 722
 Buchanan, J., 1107
 Buchner, E., 470, 593
 Budding yeast, *see* *Saccharomyces cerevisiae*
 Buffers, 47–48
 Buffer capacity, 51
 Bugg, C., 656
 Bundle-sheath cells, 936
 α -Bungarotoxin, 781
 Burley, S., 1303, 1378
 Burns, J., 620
 Burst phase, in protein folding, 286
 Butryl-ACP, 965
tert-Butyloxycarbonyl (BOC) group, 206–208
 Butyrylcholinesterase, 782
 bw gene (fruit fly), 25, 25F
 Bypass DNA polymerases, 1222
 C
 C1 domain, 730–731
 C2 domain, 727, 730–731
 C3G, 737F
 C₄ cycle, 934–937, 936F
 CA1P (2-carboxyarabinitol-1-phosphate), 930
 Ca²⁺, *see* Calcium ion
 Ca²⁺-binding protein, 678
 (Ca²⁺)-ATPase, 762–764
 CABP (2-carboxyarabinitol-1,5-bisphosphate), 930, 930F
 CACCC box, 1282
 cacophony pre-mRNA (*Drosophila melanogaster*), 1322
Caenorhabditis elegans:
 genome sequencing, 177T, 182
 miRNA, 1326
 sense RNA injection for, 1323
 trans-splicing in, 1320
 Cahn, R., 76
 Cahn-Ingold-Prelog system, 76–77
 N³-CAIR, 1111
 CAIR (carboxyaminoimidazole ribotide), 1109F, 1111
 Cairns, J., 1173, 1181
 Calcineurin (CaN), 724–725
 Calcitonin, 673T, 677, 678
 Calcium ion (Ca²⁺), 45T, 805, 816–817
 Calcium metabolism, 677–678
 Calmodulin (CaM), 655–658, 688, 764
 Calmodulin-dependent protein kinase I (CaMKI), 658
 Calmodulin-dependent protein kinase II (CaMKII), 699
 Calnexin (CNX), 427, 885–887, 886F, 887F
 Caloric restriction, 1101
 Calorie (unit), 53, 53T
 Calpain, 736
 Calreticulin (CRT), 427, 885–887, 886F
 Calvin, M., 927, 930
 Calvin cycle:
 control of, 932–934
 free energy changes from, 933T
 net reaction, 931

- Calvin (*Contd.*)
 in photosynthesis, 927–934, 929F
 stoichiometry of, 931
- CaM, *see* Calmodulin
- CAM (Crassulacean acid metabolism), 937
- Cambillau, C., 941
- Camels, erythrocyte membranes, 414
- CaMKI (calmodulin-dependent protein kinase I), 658
- CaMKII (calmodulin-dependent protein kinase II), 699
- cAMP (adenosine-3',5'-cyclic monophosphate), 653
 and AraC, 1292
 and blood glucose control, 661
 and catabolite repression, 1286
 controlling fatty acids with, 861
 and eicosanoid mediation, 993
 and integration of metabolic cycles, 1089
- cAMP-dependent protein kinase (CAPK), 653
- cAMP-phosphodiesterases (cAMP-PDEs), 698
- cAMP receptor protein (CRP), 1286
- cAMP response element (CRE), 879
- Camptothecin, 1170
- CaN (calcineurin), 724–725
- Cancer. *See also* related topics, e.g.:
 Carcinogenesis
 and cell cycle, 1202
 cell metabolism, 864–865
 and chaperone proteins, 720
 and DNA methylation, 1251
 and gene therapy, 123
 role of glycoproteins in, 381
 and telomeres, 1211–1212
 thymidylate synthesis inhibition, 1129–1130
 and tyrosine kinase-based signaling, **703–705**
- Canis familiaris* (dog), genome sequencing, 177T
- Cantor, C., 158
- CAP, *see* Catabolite gene activator protein; Cbl-associated protein
- cap-0, 1302
- cap-1, 1302
- cap-2, 1302
- Cap binding protein, eIF4E, 1378
- Capillary electrophoresis (CE), 152
- CAPK (cAMP-dependent protein kinase), 653
- Capping, of telomeres, 1210
- Capping enzyme, 1302, 1302F
- 19S Caps, 1411F
- Capsids, 26
- Caps structure, 1302, 1302F
- Capsule, bacterial, 4
- Carbamates, 329, 1025
- N-Carbamoylaspartate, 475
- Carbamoyl aspartate, in UMP synthesis, 1115F, 1116
- Carbamoylation, hemoglobin, 340
- Carbamoyl phosphate, 475, 1025–1028, 1115F
- Carbamoyl phosphate synthetase (CPS), **1025–1028**
- Carbamoyl phosphate synthetase I (CPS I), 1116
- Carbamoyl phosphate synthetase II (CPS II), 1115F–1116
- Carbanions, 563, 563F, 568–569, 569F
- Carbocation, 563, 563F
- Carbodiimides, 207
- Carbohydrates, **359**. *See also* Glycoproteins; Monosaccharides; Oligosaccharides; Photosynthesis; Polysaccharides
- analysis of, **366–367**
 in collagen, 238
 energy recovery from, 789
 Fischer convention for, 360–361
 functional groups of, 43
 glycoprotein, 380–381
 low carbohydrate diets, 1106
 from photosynthesis, 5
- Carbohydrate metabolism, 561F
- Carbohydrate microarrays, 383
- Carbohydrate recognition markers, 439
- Carbon, properties of, 29, 31
- Carbon–carbon bonds, breaking and making, 562–563, **567–569**, 568F
- Carbon dioxide:
 in Calvin cycle, 927, 931
 compensation point in plants, 935–936
 in photosynthesis, 901, 903, 935–937
 transport by hemoglobin, **328–329**
- Carbonic anhydrase:
 diffusion-limited, 490
 metal ion catalysis, 511–513, 512F
 Michaelis–Menten kinetic constants, 489T
 tertiary structure, 246F
- Carbon monoxide, heme group and, 324
- Carbon monoxide dehydrogenase, 954
- Carbonylcyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP), 860
- Carboxyaminoimidazole ribotide, *see* CAIR
- 2-Carboxyarabinitol-1,5-bisphosphate, *see* CABP
- 2-Carboxyarabinitol-1-phosphate (CA1P), 930
- Carboxyatractyloside (CATR), 771
- γ-Carboxyglutamate, 79F
- γ-Carboxyglutamic acid, 79
- Carboxy-hemoglobin, solubility and ionic strength of, 133F
- Carboxyl groups, 43F
- Carboxyl terminus, *see* C-terminus
- Carboxyltransferase, 963
- Carboxymethyl-cellulose ion exchangers, *see* CM-cellulose ion exchangers
- Carboxymethyl-CoA, 806
- Carboxymethylcysteine, 607
- Carboxymyoglobin, 1057
- Carboxypeptidases, 165–166, 168T
- Carboxypeptidase A, 168T, 231F, 254F, 473
- Carboxypeptidase B, 168T
- Carboxypeptidase C, 168T
- Carboxypeptidase Y, 168T
- Carboxyphosphate, 873, 1025
- Carboxyprocollagen peptidase, 1403–1404
- Carboxypropyl-CoA, 955
- Carcinogens, 1202, 1224–1225
- Carcinogenesis, 1224
- Cardiac glycosides, 761–762, 761F
- Cardiolipin, 389T, 398T, 1005, 1006F
- Cardiotonic steroids, 761–762
- Carell, T., 1215
- Carnitine, 946
- Carnitine palmitoyltransferase I, 946, 975
- Carnitine palmitoyltransferase II, 946
- β-Carotene, 122, 908, 976
- Carotenoids, 905F, 907–909
- Carrager, B., 437
- Carriers, 745, 748
- Carsonella ruddii*, 177T
- CART (cocaine and amphetamine-regulated transcript), 1099
- Cartilage, 235, 240T, **375**
- Caruthers, M., 211
- Casein, 150
- β-Casein, 534
- κ-Casein, 547
- Casein kinase 1, 660
- Casein kinase 2, 660
- CASP (Critical Assessment of Structure Prediction), 305
- Caspersson, T., 1260
- Cassette exons, 1318
- Cassette mutagenesis, 119–120
- Catabolism, 17, 561–562, 561F
 purine ribonucleotides, **1130–1134**
 pyrimidine ribonucleotides, **1136**
- Catabolite gene activator protein (CAP), 654
 and AraC, 1292
 and catabolite repression, 1286–1288, 1287F
- Catabolite repression, 766, **1285–1288**
- Catalase, 10, 866, 935
 heme group in, 324
 horse liver, 153T
 Michaelis–Menten kinetic constants, 489T
 molecular mass of, 140F
 sedimentation coefficient, 154F
- Catalysis, free energy of reaction for, 487, 487F
- Catalysts, 469
- Catalytic constant, 490
- Catalytic triad, 529
- Cataplerosis, 1090
- Cataplerotic reactions, 818
- Cataracts, and galactose, 633
- Cate, J., 1366
- Catechol, 680, 1059
- Catecholamines, 680, 1058–1060
- Catenation, 1163, 1163F
- CATH (Class, Architecture, Topology, and Homologous superfamily), 256T, 258
- Cathepsins, 1408–1409
- Cathepsin D, 439, 547
- Cathepsin E, 547
- Cathepsin K, 540
- Cation exchangers, 135–136
- CATR (carboxyatractyloside), 771
- Cattle, *see* Cow
- Caveolae, 411
- Caveolins, 411
- C_α backbone, 244–245
- Cbl-associated protein (CAP), 737F
- CBP, *see* CREB-binding protein
- CCA-adding polymerase, 1331
- CCAAT box, 1282
- CCAAT/enhancer-binding protein α (C/EBPα), 572
- c-Cbl, 1412–1413, 1412F
- CCdA-p-Puro, 1383, 1383F
- C chain, proinsulin, 280–281, 280F
- CCHL (cytochrome *c* heme lyase), 448
- CCK, *see* Cholecystokinin
- CCT chaperonins, 301–302, 301F
- CCVs, *see* Clathrin-coated vesicles
- CD, *see* Circular dichromism
- Cdc6, 1206–1207
- Cdc7, 1205
- Cdc18, 1206–1207
- Cdc45, 1205
- CDKs (cyclin-dependent protein kinases), 1202, 1206–1207, 1414
- cDNA:
 depositing on DNA chips, 212–213
 and PCR, 116
 reverse transcriptase, 1207–1208
- CDP-diacylglycerol, 1004–1005
- CD spectroscopy, 284–285, 285F
- Cdt1, 1206, 1207
- CE (capillary electrophoresis), 152

- C/EBP α (CCAAT/enhancer-binding protein α), 572
- Cech, T., 1306–1308
- Celebrex, 543, 1000
- Celecoxib, 999–1000
- Celera Genomics, 181–182
- Cells, 3, 16F. *See also* Prokaryotes
- eukaryote, 7F
 - human, 12F
 - plant, 11F
 - prokaryotic, 4F
- Cell cycle, 1202, 1202F. *See also* Apoptosis
- Cell division:
- meiosis, 21F
 - mitosis, 20F
 - in mitosis, 1202
- Cell growth, 1202, 1211
- Cell membranes, 4, 4F, 7F. *See also* Plasma membranes
- Cellobiose, 367, 367F
 - Cellophane, 141
 - Cell proliferation, 1202. *See also* Cancer
 - Cell sorters, 575
 - Cellular immunity, 292
 - Cellulase, 369
 - Cellulose, 11, 359, 367–369, 368F
 - Cellulose-based ion exchangers, 137–138, 137F
 - Cell walls, 4
 - plant cells, 11, 11F
 - prokaryotes, 4F
- Central dogma of molecular biology, 95, 95F, 1260
- Centrifugation:
- differential, 130
 - ultra-, 132T, 152–156, 159
- Centrioles, 7F
- Centromeres, 20, 24F, 109
- Ceramides, 390, 1008–1009
- Cerebrosides, 391, 1008–1010, 1009F, 1010F
- Ceruloplasmin, 140F
- Cesium chloride, 156, 156F
- CETP (cholesteryl ester transfer protein), 458
- Cetuximab, 720
- Cetyltrimethylammonium bromide (CTAB), 399F
- CE website, *see* Combinatorial extension of optimal pathway
- C-extein, 140T
- CF₁CF₀ complex, 926
- cf gene (fruit fly), 25, 25F
- CFI (cleavage factor I), 1302
- CFII (cleavage factor II), 1302
- c-fos proto-oncogene, 705
- CFTR (cystic fibrosis transmembrane regulator) protein, 427–428
- CFTs (channel-forming toxins), 416–418
- CG, *see* Chorionic gonadotropin
- c gene (fruit fly), 25F
- cGMP, 687
- cGMP-dependent protein kinase, 687
- cGMP-phosphodiesterase (cGMP-PDE), 692
- CGN, *see* Cis Golgi network
- Chagas-Cruz disease, 799
- α Chain, hemoglobin, 193, 194
- β Chain, hemoglobin, 193, 194
- δ Chain, hemoglobin, 194
- ϵ Chain, hemoglobin, 194
- γ Chain, hemoglobin, 194
- ζ Chain, hemoglobin, 194
- Chain elongation:
- protein synthesis, 1372–1373, 1379–1388
 - RNA, 1270–1272, 1270F, 1271F
- Chain initiation:
- codons for, 1343
 - protein synthesis, 1373–1379
 - RNA, 1267–1269
- Chain termination:
- protein synthesis, 1391–1395
 - RNA, 1271F, 1273–1275
- Chain-terminator method, 176. *See also* Sanger method
- Chair conformation, 363, 363F, 519F
- Chalfie, M., 120
- Chamberlin, M., 1273
- Chambon, P., 1305
- Chang, G., 767
- Changeux, J.-P., 349
- Channel formers (ionophores), 748
- Channel-forming membrane proteins, 416–418
- Channel-forming toxins (CFTs), 416–418
- Channeling, 794, 1028, 1075
- Ca²⁺ channels, 772
- Chaotropic ions, 266
- Chaperone proteins, 293–302, 720
 - role in secretory pathway, 421F, 427–428
- Chaperonins, 293
- CHAPS (detergent), 399F
- Chargaff, E., 84, 86, 88
- Chargaff's rules, 84, 85, 88
- Charge relay system, 536
- Charges, of amino acid side chains, 71
- Charge shielding, 513
- Charge transfer complex, 803–804, 804F
- Chase, M., 87
- CHCRs, *see* Clathrin heavy chain repeats
- Checkpoint, in cell cycle, 1202
- C Helix, 242F
- Chemical carcinogens, 704–705
- Chemical coupling hypothesis, 845
- Chemical damage, DNA susceptibility to, 1214F
- Chemical equilibria, *see* Equilibrium
- Chemical evolution, 29, 30F, 31–33, 185–194
 - and gene duplication, 192–194
 - and neutral drift, 188–192
 - rates for proteins, 191F, 192
 - and sickle-cell anemia, 185–188
- Chemical kinetics:
- elementary reactions, 483
 - reaction rates, 483–484
 - and spontaneous processes, 54
 - transition state theory, 484–487
- Chemical logic, 563–565
- Chemical mutagenesis, 1339–1340
- Chemical mutagens, 1181, 1339–1340
- Chemical potential, 58, 744
- Chemical protons, 843
- Chemical synapses, 778
- Chemical synthesis:
- of genes, 118
 - of oligonucleotides, 209–214
 - of polypeptides, 205–209
- Chemiluminescence, 183
- Chemiosmotic hypothesis, 846
- Chemokines, 717
- Chemolithotrophs, 5
- Chemotaxis, 717
- Chemotrophs, 559
- Cheng, X., 1248
- Chenodeoxycholate, 992
- Chickens:
- cancer viruses, 704–705
 - embryonic development, 14F
 - lactate dehydrogenase H, 153T
 - ovalbumin, 1304, 1304F, 1305F
 - as source of proteins for purification, 130
 - triose phosphate isomerase, 231F, 252, 254F
- Chicken foot (Holliday junction), 1234
- Chimeric vector, 104
- Chimpanzees, 116, 177T
- Chipman, D., 523
- Chiral centers, 74–76
- Chirality, 74, 78, 225–226
- Chiral organic synthesis, 78
- Chi sequences, 1230
- Chi structures, 1228
- Chitin, 369, 369F
- Chiu, W., 301
- Chl, *see* Chlorophyll
- Chl *a*, *see* Chlorophyll *a*
- Chl *b*, *see* Chlorophyll *b*
- Chloramphenicol, 107, 1395T, 1396F, 1397
- Chloride ion, 45T
- Chlorocruorins, 325
- Chloroform, 43T, 144
- Chlorophyll (Chl), 12, 403, 903–912. *See also* Photosynthesis
- absorption spectrum, 905F
 - electronic states, 905F
- Chlorophyll *a* (Chl *a*), 903, 904F, 905F, 908, 916, 917
- Chlorophyll *b* (Chl *b*), 903, 904F, 905F, 908
- Chloroplasts, 11F, 13F
 - photorespiration in, 935
 - photosynthesis site, 11–12, 901–903

Chloroquine, 1058, 1409

Cholate, 992

Cholecalciferol-25-hydroxylase, 678

Cholecystokinin (CCK), 673T, 675

Cholera toxin (CT), 694–697, 1398

Cholesterol, 392–393, 393F

 - and dietary EPA, 1002–1003
 - excretion of, 993
 - fluidity of, 398
 - and LDL, 449F
 - in membranes, 398T, 400F
 - metabolism of, 975, 984–987
 - synthesis of, 1096
 - takeup by LDL, 453–455
 - transport of, 452F, 456
 - utilization of, 991–993

Cholesterol 7 α -hydroxylase, 993

Cholesterol biosynthesis, 975–976

 - citric acid cycle intermediates, 818
 - control of, 989–991
 - HMG-CoA in, 976–977
 - HMG-CoA reductase in, 987–989
 - lanosterol in, 982–984, 985F
 - and LDL receptor activity, 989–991
 - pyrophosphomevalonate decarboxylase in, 977–978
 - squalene formation, 978–982, 979F

Cholesteryl esters, 393, 449F, 984, 986

Cholesteryl ester hydrolase, 466

Cholesteryl ester transfer protein (CETP), 458

Cholesteryl stearate, 393

Cholestyramine, 990

Choline, 389T, 390, 779

Choline acetyltransferase, 779

Cholinergic synapses, 778–780

Chondroitin-4-sulfate, 371F, 372

Chondroitin-6-sulfate, 371F, 372

Chorionic gonadotropin (CG), 132, 673T, 684

Chorthippus parallelus, 24F

Chou, P., 302, 303

Chou–Fasman method, 302–303

Chromate, 1075, 1076F, 1077F

Chromatids, 20, 24, 24F

Chromatin, 7F, 8

Chromatogram, 144

- Chromatographic separations, 135
 affinity chromatography, 118, 141–143
 gel filtration chromatography, 138–141
 ion exchange chromatography, 135–138
 nucleic acids, 157
 proteins, 135–146
- Chromatophores, 902
- Chromogenic substrates, 110
- Chromophore, 402, 908
- Chromosomal theory of inheritance, 22–26
- Chromosomes, 4, 19–20, 19F
 bacterial artificial, *see* BACs [bacterial artificial chromosomes]
 in meiosis, 21F
 in mitosis, 20F
 multiforked, 1195, 1195F
 number of in eukaryotes, 19T
 of prokaryotes, 4
 recombination of, *see* Recombination
 replication of, *see* DNA replication
 sex, 22–23
 and telomere capping, 1210
 yeast artificial, *see* YACs (yeast artificial chromosomes)
- Chromosome walking, 114, 114F, 180
- Chronic myelogenous leukemia (CML), 718
- Chyle, 451
- Chylomicrons, 449, 451–452, 944
 and cholesterol biosynthesis, 986
 properties, 449T
 transport in blood, 973
- Chylomicron remnants, 451–452, 986
- Chymosin, 547
- Chymotrypsin, 470, 525, 525T, 529T, 565
 active sites, 529F, 531F
 catalytic mechanism, 531–537
 gene duplication, 194
 kinetics and catalytic group, 525–527
 Michaelis–Menten kinetic constants, 489T
 proenzymes, 538
 X-ray structure, 527–529, 528F
- α-Chymotrypsin, 153T, 538, 538F
- π-Chymotrypsin, 538, 538F
- Chymotrypsinogen, 140F, 264F, 527, 538, 538F
- CIC channels, 755–756, 755F
- Ciechanover, A., 1410
- Cilia, 10
- Ciliates, 7F, 13F
- Cimetidine, 764
- Ciprofloxacin (Cipro), 1167, 1170, 1395
- Circular dichromism (CD), 284–285, 285F
- cI repressor, 1288
- Cis-acting elements, 1264
- Cis cisternae, 428, 429F
- Cis configuration, 28, 388
- Cis Golgi network (CGN), 428, 429F
- Cis-SNARE complex, 444–445
- Cis-splicing, 1320
- Cisternae, 428
- Cisternal maturation, 428
- Cisternal progression, 428
- Cis-trans test, 28, 28F
- Cistrons, 28, 1264, 1340
- Citrate:
 in citric acid cycle, 789, 790F, 806–808, 806F, 807F, 815–816
 glycolysis inhibition, 863–864
 and metabolic cycles, 1088–1089
- Citrate condensing enzyme, 806
- Citrate synthase, 568, 621
 in citric acid cycle, 789, 790F, 806–808, 806F, 807F, 815–816
- Citric acid cycle, 561F, 789–820
 acetyl-CoA synthesis, 792–805
 aconitase, 789, 790F, 808–809, 808F
- amphibolic functions, 789, 817–819, 818F
 citrate synthase, 789, 790F, 806–808, 806F, 807F, 815–816
 coupling to glycolysis, 594F
 electron-transport chain, 824F
 energy-producing capacity, 813
 enzymes of, 806–815
 evolution of, 814–815
 fumarase, 790F, 791, 812–813, 812F
 integration of, 813–814, 1090
 isocitrate dehydrogenase, 789, 809–810, 809F, 815–816
 α-ketoglutarate dehydrogenase, 789, 810, 815–816
 malate dehydrogenase, 790F, 791, 813
 net reaction, 791
 overview, 789–792, 790F
 oxaloacetate in, 873F
 pathways using intermediates, 818, 872F
 pyruvate dehydrogenase complex, 792–800, 804–805, 805F
 rate-controlling reactions of, 815–816
 regulation, 815–817, 816F
 replenishing intermediates of, 818–819
 succinate dehydrogenase, 791, 811–812, 812F
 succinyl-CoA synthetase, 791, 810–811, 811F
- Citrulline, 80, 80F, 571F, 1028
 L-Citrulline, 686
- Citryl-CoA, 807
- CJD, *see* Creutzfeldt–Jakob disease
- Claisen ester cleavage, 949–950
- Claisen ester condensation, 568
- Clamp loader, 1196–1198, 1197F
- Clans, of Pfam families, 258
- Clardy, J., 1116
- Clark, B., 1380
- Class, Architecture, Topology, and Homologous superfamily, *see* CATH
- Class I aaRSs (aminoacyl-tRNA synthetases), 1350–1351, 1350T, 1355F, 1358, 1358F
 X-ray structure of, 1352–1354
- Class I promoters, 1286
- Class I release factors, 1391–1392
- Class II aaRSs (aminoacyl-tRNA synthetases), 1350–1351, 1350T, 1355F, 1358, 1358F
 X-ray structure of, 1354–1355
- Class II major histocompatibility complex (MHC) proteins, 1102
- Class II promoters, 1286–1287
- Class III promoters, 1287
- Classical thermodynamics, 587
- Classification number, 470
- Clathrates, 262F, 263
- Clathrin, 430–439
- Clathrin barrel, 437F
- Clathrin-coated vesicles (CCVs), 431–439, 431F, 433F, 434F
- Clathrin heavy chain repeats (CHCRs), 432F, 433
- Clathrin heavy chains, 430, 432, 432F
- Clathrin light chains, 430, 432, 432F, 433F
- Claude, A., 1362
- Claudius, R., 56
- Cl[–] channels, 755–756, 755F
- Cleavage and polyadenylation specifying factor (CPSF), 1302–1303
- Cleavage factor I, 1302
- Cleavage factor II, 1302
- Cleavage sites, type II restriction enzymes, 105, 105T
- Cleavage stimulating factor (CstF), 1302–1303
- Cleland, W. W., 498, 536
- Clinical trials, 542–543
- Clones, 104, 113–114
- Cloning, *See also* Molecular cloning
 directional, 117, 117F
 molecular, *see* Molecular cloning
 shotgun, 113
- Cloning vectors (vehicles), 104, 106–109
- Clore, M., 656
- Closed complexes, 1268
- Closed conformations:
 Klentaq1, 1179–1180, 1179F
 Rep helicase, 1186
- Closed systems, 52
- Clostridium botulinum*, 442, 1340
- Clotting, *see* Blood clotting
- Cloverleaf structure, 1329–1330, 1329F, 1346F
- ClpA, 1419
- ClpP, 531, 531F, 1419, 1419F
- Clp protease, 1419
- ClpX, 1419
- CLUSTAL, 201–202, 202F, 205, 304
- Cmc (critical micelle concentration), 394
- CM-cellulose ion exchangers, 137–138, 141
- CML (chronic myelogenous leukemia), 718
- CMP (cytidine monophosphate), 83T, 1136F
- CMV (cytomegalovirus), 1403
- Cn3D, 256T, 258
- CNBr, *see* Cyanogen bromide
- CNX, *see* Calnexin
- CO₂ compensation point, 935–936
- CoA, *see* Coenzyme A
- Coagulation, *see* Blood clotting
- Coated pits, 454, 454F
- Coated vesicles, 428–439, 429F
- Coatomer, 430
- Cobalamin coenzymes, 473T, 474T
- Cobalt ion, 1158
- Cobra toxin, 781
- Cocaine and amphetamine-regulated transcript (CART), 1099
- Cocci, 4
- Cochaperone proteins, 293
- Cockayne syndrome (CS), 1218
- Cocoonase, 525T
- Coding strand, 1266
- Codominant traits, 22, 23F
- Codons, 95F, 98–101, 1338, 1343T
 anticodon interactions with, 1389F
 frequently used, 1360
 nonsense suppressors, 1362, 1362T
 triplet, 1340–1341
- Coelenterates, 13F
- Coenzymes, 473–474, 473T, 474T
 biosynthesis of flavin, 1138
 biosynthesis of nicotinamide, 1136–1138
 as covalent catalysts, 511
- Coenzyme A (CoA), 82, 473T, 795T. *See also* Acetyl-CoA
 biosynthesis, 1138–1139
 discovery, 791–792
- Coenzyme B₁₂, 954–955, 954F
- Coenzyme Q (CoQ; ubiquinone):
 electron-transport chain coenzyme, 829–830, 836F
 in oxidative phosphorylation, 848–849
- Coenzyme Q: Cytochrome c oxidoreductase (Complex III), 840–841, 840F
- Coenzyme QH*, 836F
- Coenzyme QH₂, 836F, 848–849
- Coenzyme Q reductase, 834–835, 837–840
- Cofactors, 473

- Cohen, P., 1025
 Cohen, S., 107
 Cohesive ends, 105. *See also* Sticky ends
 Coil conformation, **230, 232**
 Coiled coil structure (keratin), 234, 235F
 Cointegrate, 1239–1240, 1239F–1241F
 Cold denatured proteins, 284
 Cold-labile proteins, 130
 Colicins, 418
 Colicin E3, 1375
 Colipase, 941–942, 941F
 Coliphage, 1190
 Collagen, 78, 163, **235–240**
 assembly, 1405
 diseases associated with, 240
 isoelectric point, 134T
 Ramachandran diagram, 224F
 sedimentation coefficient, 154F
 Collagen α 1(I), 236F
 Collagen fibrils, 237–240, 237F, 239F, 240T
 Collective motions, 307, 308
 Collins, F., 181
 Collman, J., 328
 Collodion, 141
 Colony hybridization, 113–114, 114F
 Colony-stimulating factors, 119
 Combinatorial chemistry, 541
 Combinatorial extension of optimal pathway, 256T, 258
 Comparative modeling, 304–305
 Compare 3D, 258
 Competitive-binding studies, 674–675
 Competitive inhibition, **493–494**, 494F, 747
 Complementary base pairing, 18, 89
 Complementation group, 27
 Complementation tests, 24–25, 25F, 27
 Complement C1, 525T
 Complement C9, 1317
 Complement system-mediated reactions, 121
 Complex I (NADH dehydrogenase, NADH: coenzyme Q oxidoreductase), 832T, **834–837**, 836F–837F, 845
 Complex II (Succinate dehydrogenase, succinate: coenzyme Q oxidoreductase) 493, 947
 in citric acid cycle, 791, **811–812**, 812F
 and electron transport, 832T, **837–840**, 839F, 845
 Complex III (Coenzyme Q:Cytochrome *c* oxidoreductase, cytochrome *bc*₁), 832T, **840–841**, 840F, **847–850**, 849F, 912
 Complex IV (Cytochrome *c* oxidase, COX), 153T, 188, 584, 832T, **841–845**, 842F, **851–852**, 851F, 852F
 Complex V (proton-translocating ATP synthase), *see* F₁F₀-ATPase
 Complex oligosaccharides, 888
 Complex systems, 736–738
 Complimentary DNA, *see* cDNA
 Composite transposons, 1237, 1237F
 α -Conarachin, 140F
 Concanavalin A, 230, 231F, 246F, 249F, 366
 Concanavalin B, 153T
 Concentration, free energy and, 58, 61
 Concentration cells, 586
 Concerted general acid-base catalyzed reaction, 507
 Condensation reactions, for peptide bonds, 70F
 Condensing enzyme, 965
 Conformational coupling hypothesis, 845
 Conformation maps, 223–225
 Congenital erythropoietic porphyria, 1056
 Conjugate acid, 45
 Conjugate base, 45
 Conjugate redox pair, 584
 Conjugation, 1225
 Connective tissues, 235
 Connexins, 415, 416
 Connexin 26 (Cx26), 416, 417F
 Connexons, 415–416
 Conservatively substituted positions, 188
 Conservative replication, 90
 Constitutive enzymes, 1261
 Constitutive heterochromatin, 182
 Constitutive mutation, 1262
 Contact inhibition, 381, 703, 1202
 Contigs, 180
 Contour lengths, 94
 Controlled rotation, 1166, 1166F
 Convergent evolution, **317**
 Cooley's anemia, 219
 Coomassie brilliant blue, 149, 411
 Cooper, J., 1048
 Cooperative binding, 326
 Cooperative process, 92
 Cooperman, B., 1122
 Coordinate expression, of enzymes, 1264
Copia, 1246
 COPII protein, 429F, 430, 437–439, 437F, 439F
 COPI protein, 429F, 430, 437
 Coproporphyrinogen I, 1056
 Coproporphyrinogen oxidase, 448, 1053
 CoQ, *see* Coenzyme Q
 Cordycepin, 1270
 Core enzyme, 1265, 1267, 1269
 Corepressor, 1297
 Core promoter element, 1282
 Core proteins, 373
 Corey, R., 221, 229
 Cori, C., 595, 651, 880
 Cori, G., 595, 651, 880
 Cori cycle, **880**, 880F
 Cori's disease, 667
 Corn:
 Bt corn, 122
 chloroplast, 902F
 and pellagra, 474
 transposons, 1236, 1244
 CORN crib mnemonic, for handedness of amino acids, 75, 75F
 Cornea, collagen fibrils in, 240T
 Cornish-Bowden, A., 623
 Cornynephage b, 1397
 Corpeus luteum, 683
 Correlated invariants, 1346
 Correlated spectroscopy (COSY), 243
 Corrin, 954
 Cortex, 679
 Corticosterone, 680
 Corticotropin-releasing factor (CRF), 673T, 682
 Cortisol, 680
 Cosmid vectors, 108
 Cosmid walking, 181
Cos site, 108
 COSY (correlated spectroscopy), 243
 Coulomb (unit), 53T
 Coulomb's law, 42, 259
 Coumadin, 543F
 Coumarin, 1167
 Coupled reactions, **60–61**, 131
 Coupling, translational, 1398
 Covalent catalysis, **510–511**
 Covalent modification, 624
 gluconeogenesis, 878–879
 glycogen metabolism, **650–651**
 Cow. *See also related topics e.g.:* Bovine pancreatic ribonuclease A
 20S proteasome, 1417
 cellulose digestion, 369
 mad cow disease, 312, 315
 as source of proteins for purification, 130
 COX, *see* Complex IV
 Cox, M. M., 1234
 COX-1, 998, 999
 COX-2, 998, 999
 COX-2 inhibitors, 999–1000, 999F
 COX-3, 1000
 Coxibs, 999
 Cozymase, 594
 CP43 subunit (PsbC), of photosystem, 916
 CP47 subunit (PsbB), of photosystem, 916
 CP-320473, 982
 CPDs (cyclobutane pyrimidine dimers), 1214–1215
 CPE (cytoplasmic polyadenylation element), 1402
 CPE-binding protein (CPEB), 1402, 1402F
 CpG islands, 1249
 cPLA₂ (cytosolic phospholipase A₂), 727
 Cpn10, 294
 Cpn60, 294
 CPS (carbamoyl phosphate synthetase), **1025–1028**
 CPS I (carbamoyl phosphate synthetase I), 1116
 CPS II (carbamoyl phosphate synthetase II), 1115F–1116
 CPSF (cleavage and polyadenylation specifying factor), 1302–1303
 Craik, C., 530
 Cramer, W., 920
c-Ras proto-oncogene, 705, 708–709
 Crassulacean acid metabolism (CAM), 937
 Cravatt, B., 404
 CRE (cAMP response element), 879
 Creatine, 1092
 Creatine kinase, 505, 570, 571F, 583
 CREB, *see* CRE binding protein
 CREB-binding protein (CBP), 283–284, 283F
 CRE binding protein (CREB), 283–284, 283F, 879
 Cre recombinase, 1243–1244, 1245F, 1259
 Cretinism, 677
 Creutzfeldt-Jakob disease (CJD), 312, 314, 315
 CRF, *see* Corticotropin-releasing factor
 Crick, F., 88, 95, 236, 1146, 1173, 1260, 1340, 1341, 1345, 1360
 Cristae, 9, 824, 825F
 Critical Assessment of Structure Prediction (CASP), 305
 Critical micelle concentration (cmc), 394
 Critical nucleus, 347F
 CrkII, 737F
 Crocodiles, blood oxygen utilization, 357
 Crofts, A., 840
 Crop plants, 122. *See also* Corn; Rice
 Cro protein, 1288, 1289F
 Crossing-over, 20, 24, 24F, 1226–1228
 Cross-linking agents, 270–271, 270F
 Cross- β spines, 310–311
 Crotoxin, 153T
 CRP (cAMP receptor protein), 1286
 CRT, *see* Calreticulin
 Cryo-electron microscopy (cryo-EM), 301, 1365–1366
 in citric acid cycle research, 798
 in secretory pathway research, 422–426, 422F, 423F, 425F
 Cryptic branch site, 1306
 Cryptic splice sites, 1318
 α Crystallin, 140F
 γ -B Crystallin, 252, 253F

- Crystallization, protein, 134–135, 241–243
 Crystallographic symmetry, 270
 Crystal proteins, 122
 CS (Cockayne syndrome), 1218
 CsA, *see* Cyclosporin A
 c-Src gene, 705, 706
 CstF (cleavage stimulating factor), 1302–1303
 CT, *see* Cholera toxin
 CTAB (cetyltrimethylammonium bromide), 399F
 CTD (C-terminal domain), 1277, 1286–1288
 CTD (C-terminal domain) kinases, 1277
 CTD (C-terminal domain) phosphatases, 1277
 C-terminal domain, *see* CTD
 C-terminus, 73
 in protein sequencing, 165–168
 in protein synthesis, 1371–1373
 CTP (cytidine triphosphate), 475F
 in ATCase feedback inhibition, 475
 synthesis of, 1118–1119
 in transcription, 96, 1265
 CTP synthetase, 1118
 C-Type cytochromes, 316, 316F, 317F
 Cud chewing, 369
 Cul1, 1412, 1413F
 Cullin family, 1412, 1421
 Curare, 781
 Curie (Ci), 591
 Curry, S., 944
 Curved arrow convention, 564
 Cusack, S., 1381
 Cushing's syndrome, 681, 743
 Cut-and-paste transposition mechanism, 1237–1239, 1238F
 Cx26, *see* Connexin 26
 Cyanate, Hb and, 340
 Cyanide, 831
 β-Cyanoalanine, 80, 80F
 Cyanobacteria, 5–6, 7F, 12, 13F
 Cyanogen bromide (CNBr), 142, 170–171
 Cyanosis, 342
 3',5'-Cyclic AMP, *see* cAMP
 Cyclic AMP response element-binding protein, *see* CREB-binding protein (CBP)
 Cyclic cascade, 647–650
 Cyclic depsipeptide, 748
 3',5'-Cyclic GMP, *see* cGMP
 2',3'-Cyclic nucleotides, 508
 Cyclic processes, 53
 Cyclic symmetry, of proteins, 268, 268F
 Cyclins, 214F, 1202
 Cyclin-dependent protein kinases, *see* CDKs
 Cyclobutane pyrimidine dimers (CPDs), 1214–1215
 Cyclobutylthymine dimer, 1214–1215, 1214F
 Cyclohexane, conformations of, 363, 363F
 Cycloheximide, 1395T, 1396F
 Cyclooxygenase (COX), 406, 997
 Cyclopentanoperhydrophenanthrene, 392
 Cyclophilin, 724F, 725
 Cyclopropane, 398
 Cycosome, 1414
 Cyclosporin A (CsA), 292–293, 724
 Cygler, M., 886
 Cys, *see* Cysteine
 CysRS, in archaeabacteria, 1359
 Cystathione, 1034
 Cysteine:
 amino groups, 208F
 biosynthesis, 1071–1072
 degradation, 1030–1034
 disulfide bonds, 71, 71F, 168–169
 and fluorescence resonance energy transfer, 286
 genetic codes for, 100T, 1343T
 α helix/β sheet propensity, 302T
 as nonessential amino acid, 1065T
 side chains, 71, 208F, 264T
 structure and general properties, 69T
 Cysteine proteases, 1417
 Cysteinyl-adenylate, 1359
 Cystic fibrosis, 122–123, 427–428
 Cystic fibrosis transmembrane regulator (CFTR) protein, 427–428
 Cystine, 71. *See also* Cysteine
 disulfide bond cleavage in protein sequencing, 168–169
 stereoisomers, 76, 76F
 D-Cystine, 76, 76F
 L-Cystine, 76, 76F
 meso-Cystine, 76, 76F
 Cys-tRNA^{Cys}, 1359
 Cyt, *see* Cytosine
 Cytidine, 83T, 1136F
 Cytidine deaminase, 1136F, 1322
 Cytidine monophosphate, *see* CTP
 Cytidine triphosphate, *see* CTP
 Cytidylc acid, *see* CMP
 Cytochalasin B, 769
 Cytochrome *a*, 838, 838F
 Cytochrome *b*, 838, 838F, 1408T
 Cytochrome *b*₅, 401–402, 402F, 971
 Cytochrome *b*₆, 920
 Cytochrome *b*/*f* complex, 915, 920–921, 921F
 Cytochrome *b*₅₅₉, 916
 Cytochrome *b*₅₆₂, 250F, 251
 Cytochrome *bc*₁, *see* Complex III
 Cytochrome *b*_h, 838
 Cytochrome *b*_h, 838
 Cytochrome *c*, 16F, 188–189, 838, 838F, 843–843
 amino acid sequence, 186T–187T, 189F, 196F
 evolution of, 189, 191F, 316, 317F
 folding of, 284, 289, 289F
 half-life, 1408T
 isoelectric point, 134T
 molecular mass, 140F
 phylogenetic tree, 189, 190F
 physical constants of, 153T
 sedimentation coefficient, 154F
 visible absorption spectrum, 838F
 X-ray structure of, 247F
 Cytochrome *c*₁, 838F, 840, 848
 Cytochrome *c*₂, 912
 Cytochrome *c* heme lyase (CCHL), 448
 Cytochrome *c* oxidase (COX), *see* Complex IV
 Cytochrome *c* reductase, 188
 Cytochrome *f*, 914–915, 915F, 920
 Cytochrome P450:
 detoxification reactions, 1225
 and drug metabolism, 543–545
 in fatty acid oxidation, 959
 in heme biosynthesis, 1055
 polymorphic, 545
 X-ray structure, 543F
 Cytochrome P450 reductase, 686
 Cytoglobin, 325
 Cytokines, 688
 γC Cytokine receptor, 123
 Cytokine receptors, 717
 Cytokinesis, 20F, 21F
 Cytomegalovirus (CMV), 1403
 Cytoplasm, 4, 1326–1327
 Cytoplasmic polyadenylation, 1401–1402
 Cytoplasmic polyadenylation element (CPE), 1402
 Cytosine, 18, 83, 83T. *See also* Watson–Crick base pairs
 and genetic code, 100T, 1343T
 IR spectra of derivatives, 1155F
 modified nucleosides occurring in tRNAs, 1347F
 and point mutation generation, 1339–1340, 1340F
 Cytoskeleton, 10–11, 10F
 Cytosol, 9–11. *See also* related topics, e.g.:
 Fatty acid biosynthesis
 citric acid cycle in, 815
 gluconeogenesis in, 876–877, 877F
 glycolysis in, 595
 heme biosynthesis in, 1054
 metabolic functions, 562T
 mRNA transport to, 101
 Cytosolic phospholipase A₂ (cPLA₂), 727
 D
 D, *see* Dihydrouridine
 D1 protein, 1314
 D1 protein (PsbA), of photosystem, 916
 D2 protein (PsbD), of photosystem, 916
 DAG, *see* Diacylglycerol
 DAHP (2-keto-3-deoxy-D-arabinoheptulosonate-7-phosphate), 1075, 1076F
 DAI (double-stranded RNA-activated inhibitor), 1400
 Dailey, H., 1055
 Dalgarno, L., 1374
 Dalton (unit), 5
 dam gene, 1247
 Dam methyltransferase (Dam MTase), 1196, 1220, 1247
Danio rerio, 177T
 Dansylamino acid, 165, 166F
 Dansylaminoethylthiogalactoside, 770
 Dansyl chloride, 165, 166F
 Dansyl polypeptide, 166F
 Dark reactions, photosynthesis, 902, 926–937
 C₄ cycle, 934–937
 Calvin cycle, 927–934
 photorespiration and C₄ cycle, 934–937
 D arm, 1346
 Darnell, J., 718
 Darst, S., 1268
 Darwin, C., 19, 28–29, 799
 Davies, D., 1076
 Davis, R., 109
 Dayhoff, M., 198
 Dbf4, 1206
 Dbf4-dependent kinase (DDK), 1206
 Dbp5, 1327
 DCCD (dicyclohexylcarbodiimide), 207–208
 dcm gene, 1247
 Dcm methyltransferase (Dcm MTase), 1247
 DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea), 915
 DDBJ (DNA Data Bank of Japan), 195T
 ddC (2',3'-dideoxyctidine), 1208
 dd CTP (2',3'-dideoxyCTP), 1178–1180
 DDE motif, 1238–1239
 ddI (2',3'-dideoxyinosine), 1208
 DDK (Dbf4-dependent kinase), 1206
 ddNTP (2',3'-dideoxynucleoside triphosphate), 117
 DEAD-box family, 1379
 Deadenylases, 1327
 DEAE-cellulose ion exchangers, 137, 141
 Deamination:
 of amino acids, *see* Amino acid deamination
 of bases, 1322–1323, 1339–1340, 1339F
 Death cap (*Amanita phalloides*), 1280

- Death-inducing signal complex, *see* DISC
 5-Deazaflavin, 1215
 6dEB (6-deoxyerythronolide B), 968
 Debranching enzyme, 370, 639, **642–643**, 667
 DEBS (deoxyerythronolide B synthase), 968
 Decamethion ion, 788
 Decapping enzyme, 1327
 Decay, of DNA, 116–117
 Decoding, 1379
 Decorin, 373T
 de Duve, C., 9
 Deep View, 257
 Deformylase, 1374
 Degeneracy, of genetic code, 99, 1338, 1341, 1343–1344, 1360–1361
 Dehydroascorbic acid, 364
 L-Dehydroascorbic acid, 364F
 7-Dehydrocholesterol, 678
 Dehydrogenases, 318, 498
Deinococcus radiodurans, 1259
 Deisenhofer, J., 294, 403, 453, 840, 910
 de Lange, T., 1213
 Deletion mutations, 1340
 DeLucia, P., 1181
 Dematin, 412
 Denaturation:
 of DNA, 90–93, 92F
 of proteins, *see* Protein denaturation
 Denaturation temperature, 57
 Denatured proteins, 57
 Denitrification, 1083
 Denitrifying bacteria, 6
De novo methods, 305
 Density gradients, 155
 3'-Deoxyadenosine, 1270
 5'-Deoxyadenosylcobalamin (AdoCbl), 954–955, 954F
 Deoxycytidine deaminase, 1124
 6-Deoxyerythronolide B (6dEB), 968
 Deoxyerythronolide B synthase (DEBS), 968
 Deoxygenated hemoglobin (deoxyHb), 329, 331
 and BPG, 341, 341F
 structure, 331–333, 334F–335F
 Deoxygenated hemoglobin S (deoxyHbS), 343–344, 343F–347F
DeoxyHb, *see* Deoxygenated hemoglobin
DeoxyHbS, *see* Deoxygenated hemoglobin S
 Deoxynucleoside triphosphates, *see* dNTPs
 Deoxynucleotides, **82–83**
 2'-Deoxy-PIP₂, 728
 Deoxyribonucleic acid, *see* DNA
 Deoxyribonucleotides, **82–83**, 82F, **1119–1130**. *See also* specific nucleotides
 Deoxy-D-ribose, 18
 β-D-2-Deoxyribose, 365
 2'-Deoxyribose, **1119–1126**
 2'-Deoxy-D-ribose, 83
 Deoxyribosomes, 1308
 Deoxy sugars, 365
 Deoxythymidine, 83T, 1136F
 Deoxythymidine monophosphate (dTMP), 83T, 1127–1128
 Dephospho-CoA kinase, 1138, 1139F
 Dephospho-CoA pyrophosphorylase, 1138, 1139F
 Depsipeptides, 1272
 Dermatan sulfate, 371F, 372
 Dermatosparaxis, 1404
 Desamido NAD⁺ (nicotinate adenine dinucleotide), 1138
 Desaturases, **969–971**
 de Saussure, T., 901
 Desensitization, 697
 Destruction box, 1414
 Detergents, 262, 265, 399F
 Detergent-resistant membranes (DRMs), 411
 Deubiquitinating enzymes, *see* DUBs
 Deuterium, *see* Hydrogen exchange
 de Vos, A., 684
 Dextran gels, 137, 139, 140T, 141
 α-Dextrinase, 370
 Dextrins, 370
 Dextrorotatory molecules, 74
 D forms (amino acids), 74–76
 d gene (fruit fly), 25F
 DGLA (dihomo-γ-linolenic acid), 994
 DH (β-hydroxyacyl-ACP dehydrase), 965
 DHA (4,7,10,13,16,19-Docosahexenoic acid), 387T, 971
 DHAP, *see* Dihydroxyacetone phosphate
 Dhe-Paganon, S., 1250
 DHF, *see* Dihydrofolate
 DHFR, *see* Dihydrofolate reductase
 DHODH, *see* Dihydroorotate dehydrogenase
 Diabetes insipidus, 743
 Diabetes mellitus, 310, **1102–1104**, 1409.
 See also Insulin
 1,2-Diacylglycerol, 941
 Diacylglycerol (DAG), 661, 994
 Diacylglycerol acyltransferase, 971
 Diacylglycerol lipase, 994
 Diacylglycerophospholipids, 1004–1005, 1005F
 Diagonal plot, 196
 Dialysis, 117, 132T, 141, 141F
 Diaminopimelic acid, 123, 377F
 Diastase, 469
 Diastereomers, 75–76
 Diatoms, 13F
 Diatomaceous earth, 144
 6-Diazo-5-oxo-L-norleucine (DON), 1142
 O-diazoacetyl-L-serine, 1142
 Diazotrophs, 1080
 Dicer (RNase), 1324
 3-(3,4-Dichlorophenyl)-1,1-dimethylurea (DCMU), 915
 Dickens, F., 892
 Dickerson, R., 89, 529, 1148
 Dicyclohexylcarbodiimide (DCCD), 207–208
 N,N'-Dicyclohexylurea, 207
 Didanosine, 1208
 2',3'-Didehydro-3'-deoxythymidine, 1208
 2',3'-DideoxyCTP (ddCTP), 1178–1180
 2',3'-Dideoxycytidine (ddC), 1208
 2',3'-Dideoxyinosine (ddI), 1208
 Dideoxy method, 176. *See also* Sanger method
 2',3'-Dideoxynucleoside triphosphate (ddNTP), 177
 Dielectric constants, 42–43, 43T
 2,4-Dienoyl-CoA reductase, 952
 3,5,2,4-Dienoyl-CoA reductase, 952
 Diethylaminoethyl cellulose ion exchangers, *see* DEAE-cellulose ion exchangers
 Diethyl ether, 43T, 144, 398
 Diethylpyrocarbonate, 312–313, 313F
 Diet-induced thermogenesis, 1099–1100
 Differential centrifugation, 130
 Diffusion coefficient, 745
 Diffusion-controlled limit, 490
 Digalactosyl diacylglycerol, 902
 Digitalis, 539, 761–762, 761F
 Digitoxin, 761–762
 Dihedral angles, 222–225
 Dihedral symmetry, of proteins, 268, 268F
 Dihomo-γ-linolenic acid (DGLA), 994
 7,8-Dihydrobiopterin, 1043–1044
 Dihydroceramide, 1009
 Dihydroceramide reductase, 1009
 Dihydrofolate (DHF), 445, 492, **1062**, 1127
 Dihydrofolate reductase (DHFR), 300, 445, 492, 1044, 1062, 1129–1130
 Dihydrolipoamide, in citric acid cycle, 794–795, 795F
 Dihydrolipoamide dehydrogenase, 796
 Dihydrolipoyl dehydrogenase, 792, 801–804, **801–814**, 802F, 804F
 Dihydrolipoyl transacetylase (E2), 792, 793F, **795–797**, 796F, 797F
 Dihydrolipoyl transsuccinylase (E2o), 810
 Dihydroorotate, 1115F, 1116, 1118
 Dihydroorotate, 1115F, 1116
 Dihydroorotate dehydrogenase (DHODH), 1115F, 1116
 Dihydropteridine reductase, 1044
 Dihydrosphingosine, 391F, 1009
 Dihydouracil, 1136F
 Dihydouracil dehydrogenase, 1136F
 Dihydouridine (D), 1346, 1347F
 Dihydroxyacetone, 361, 361F
 Dihydroxyacetone phosphate (DHAP), 596F, 600–603, 929, 944F
 Dihydroxyacetone phosphate acyltransferase, 971
 1α,25-Dihydroxycholecalciferol (1,25(OH)2D), 678, 679
 3,4-Dihydroxyphenylalanine, *see* L-DOPA
 Dihydroxythymine, 1136F
 Diimine, 1082
 Diisopropylphosphofluoridate (DIPF), 526–527
 Diketogulonic acid, 364
 L-Diketogulonic acid, 364F
 Diketopiperazine, 243F
 Dill, K., 262, 281, 287
 Dimers, 267
 Dimethoxytrityl (DMTr) protecting group, 210F, 211
 N⁶-N⁶-Dimethyladenosine, 1156–1157
 Dimethylallyl pyrophosphate, 977
 1-Dimethylamino-naphthalene-5-sulfonyl chloride, *see* Dansyl chloride
 5,6-Dimethylbenzimidazole (DMB), 954
 N,N-Dimethylformamide, *see* DMF
 N²,N²-Dimethylguanosine (m₂G), 1347F
 Dimethyloxaloacetate, 511
 Dimethylsulberimidate, 270, 270F
 Dimethyl sulfate (DMS), 1267
 Dimethyl sulfoxide, *see* DMSO
 dinB gene, 1222
 2,4-Dinitrofluorotoluene base, 1177
 2,4-Dinitrophenol (DNP), 860
 Dinoflagellates, 13F
 Dintzis, H., 1371
 Dinucleotide-binding fold, 256
 Dioldehydrase, 1018
 Dipeptides, 70, 70F
 DIPF (diisopropylphosphofluoridate), 526–527
 Diphosphatidylglycerol, 389T
 1,3-Diphosphoglycerate, 607
 2,3-Diphosphoglycerate (DPG), 329
 Diphthalamide residue, eEF2, 1398
 Diphtheria, 1397–1398
 Diphtheria toxin (DT), 123, 153T, 1395T, **1397–1398**
Diplococcus pneumoniae, 85
 Diploid cells, 24F
 Diploid genome, 182
 Diploid number, 19
 Dipolar ions, 70
 Dipole-dipole interactions, 260–261, 260F

Dipole moment, 39, 43T
 Dipyrromethane, 1053
 Directional cloning, 118, 118F
 Direct transposition, 1237–1239
 Disaccharides, **367**
 Disc electrophoresis, 148, 148F, 149F
 Discontinuous pH electrophoresis, 148, 148F
 Discriminator base, 1351F
 Disorder, entropy and, 54–57
 Dissociation constant (K_s):
 and acid strength, 45–47, 46T
 and pH, 47
 of polyprotic acids, 49–50
 Distal histidine, in hemoglobin, 332, 340
 Distance-based tree-building, 205
 Distance matrix, 204–205
 Distributive enzymes, 1177
 Disulfide bonds:
 cysteine, 71, 71F
 insulin, 165F
 interchange reaction, 280F
 keratin, 235
 in native proteins, 292
 in protein sequencing, 168–169, 172, 184
 and protein stability, 264–265
 Divergent evolution, 254, **317**
 Dixon, J., 736
 DL3, 714
 D-loops, 1213
 DM (myotonic dystrophy), 1252
 DMB (5,6-dimethylbenzimidazole), 954
 DMF (*N,N*-dimethylformamide), 134, 262
 DMS (dimethyl sulfate), 1267
 DMSO (dimethyl sulfoxide), 43T, 134, 262
 DNA (deoxyribonucleic acid), 3, 18, 1145.
 See also related topics, e.g.: DNA repair
 A-DNA, 1146F, 1147F, **1148–1149**,
 1148T, 1151
 B-DNA, 1145–1148, *see* B-DNA
 as carrier of genetic information, **85–88**
 and central dogma of molecular biology,
 95, 1260
 chemical structure, **84–85**
 cloning, *see* Molecular cloning
 complementary base pairing, 18
 decay of, 116–117
 denaturation and renaturation, 90–93,
 92F, 93F
 double helical structure, **1145–1148**
 fractionation, 156–159
 frictional ratio, 155
 junk, 1317
 mutationos of, *see* Mutations
 in nucleus, 8–9
 replication, 18F, 90, 101–104, 102F
 sedimentation coefficients, 154F
 semiconservative replication, 1173–1174
 semidiscontinuous replication,
 1175–1176, 1175F
 size, 94–95
 structure, 18F
 supercoiled, **1158–1170**
 transcription, 95–98
 X-ray diffraction, 88F
 Z-DNA, 1146F, 1147F, 1148T, **1149–1151**
 DnaA boxes, 1194, 1195
dnaA gene, 1196
 DnaA protein, 1194–1195, 1206
dnaB gene, 1182T
 DnaB helicase, 1191, 1193–1194
 DNA-binding domain, 705
 DnaB protein, 1206
 in prokaryotic replication, 1190, 1190T,
 1191, 1195, 1199
 in unwinding of DNA, 1183, **1184**, 1184T

DNA chips, *see* DNA microarrays
 DnaC protein, 1190, 1190T, 1195, 1206
 DNA Data Bank of Japan (DDBJ), 195T
 DNA-directed DNA polymerases, *see* DNA
 polymerases
 DNA-directed polymerases, 1173. *See also*
 DNA polymerases
dnaG gene, 1176
 DNA glycosylases, 1218
 DnaG protein, 1190T
 DNA gyrase, 1166–1169, **1174–1175**,
 1190, 1195
 DnaJ, 293
 DnaK, 293
 DNA ligase, 103, 103F, 1176
 in DNA replication, **1187–1188**,
 1188F, 1189F
 in mismatch repair, 1220
 DNA ligase IV, 1224
 DNA methylation, **1246–1252**
 DNA methyltransferase, 104, 105
 DNA microarrays, **211–213**, 211F, 212F,
 213F
dnaN gene, 1182T
 DNA photolyases, 1214–1215, 1215F
 DNA polymerase, 101–103, 102F,
 1173–1174, 1174F
 bypass, 1222
 error-prone, 1222
 eukaryotic, 1202–1205, 1203T
 intercalating agents inhibit, 1272
 nucleotidyl transferase mechanism, 1180F
 in PCR, 114–115
 properties of selected animal, 1203T
 replication fidelity, 1200–1201, 1201F
 RNA-directed, 95F, 116
 and semidiscontinuous replication,
 1175–1176
 and TERT, 1210
 T7 DNA polymerase, 179
 DNA polymerase α (pol α), 1202–1203
 DNA polymerase β (pol β), 1205
 DNA polymerase γ , *see* Pol γ
 DNA polymerase δ , *see* Pol δ
 DNA polymerase ϵ , *see* Pol ϵ
 DNA polymerase ζ , *see* Pol ζ
 DNA polymerase η , *see* Pol η
 DNA polymerase ι , *see* Pol ι
 DNA polymerase κ , *see* Pol κ
 DNA polymerase I (Pol I), 103, 103F,
 1176–1181
 in bacteriophage ϕ X174 replication, 1191
 in bacteriophage M13 replication, 1190F
 error correction, 103–104
 error rate, 1220
 as A family polymerase, 1202
 in gene manipulation, 109
 role in recombination repair, 1234
 in site-directed mutation, 119
 T. thermophilus, 1179F
 use in nucleic acid sequencing, 177–179
 DNA polymerase II (Pol II), **1181**, 1181T
 as B family polymerase, 1202
 role in SOS repair, 1222
 DNA polymerase III (Pol III), 103, **1182–1183**
 as C family polymerase, 1202
 error correction, 103–104
 role in mismatch repair, 1220
 DNA polymerase IV (Pol IV), 1182, 1222
 DNA polymerase V (Pol V), 1182, 1222
 DNA polymerase A family, 1202, 1205
 DNA polymerase B family, 1202
 DNA polymerase C family, 1202
 DNA polymerase D family, 1202
 DNA polymerase X family, 1202
 DNA polymerase Y family, 1202
dnaQ gene, 1182T
 DNA recombination, *see* Recombination
 DNA repair, **1213–1225**
 and carcinogen identification, 1224–1225
 direct damage reversal, 1214–1216
 double-strand break repair, **1223–1224**
 excision repair, **1216–1219**
 mismatch repair, **1220**
 and Pol I, 1181
 and replication fidelity, 1200–1201
 and SOS response, 1221–1223
 DNA replicase, 1182
 DNA replication, 18–19, 1145, **1173–1213**.
 See also related topics, e.g.: DNA ligase
 enzymes for, 1176–1189
 in eukaryotes, **1201–1213**
 in meiosis, 21F
 in mitosis, 20F
 overview, 1173–1176
 in prokaryotes, 1173, **1190–1201**, 1200F
 semiconservative nature of, 90, 91F,
 1173–1174
 semidiscontinuous, 1175–1176, 1175F
 telomeres, 1209–1213
 unwinding, 1183–1187
 and viruses, 27
 DNA–RNA hybrids, 96, 1151, 1151F
 DNase I, 1181
 DNA sequencing:
 automated, 179–180
 genome sequencing, 180–182
 Sanger method, 176–180
 sequence databases, 194–195
 DNA topoisomerases, 1176
 DnaT protein, 1190, 1190T
 DNA unwinding elements (DUEs), 1194
dnaX gene, 1182T
 DNMT1, 1250
 DNMT3a, 1250
 DNMT3b, 1250
 DNP (2,4-dinitrophenol), 860
 dNTPs (deoxyribonucleoside triphosphates):
 in chain-terminator method, 177–179
 in DNA replication, 101, 1170, 1174F
 and replication fidelity, 1200–1201
 Docking proteins, 420, 710
 4,7,10,13,16,19-Docosahexenoic acid,
 see DHA
 Dodecyltriethylammonium bromide
 (DTAB), 399F
 Dodson, E., 1216
 Dogs:
 fatty acid experiments with, 945
 number of chromosomes, 19T
 Dolichol, 883, 883F
 Dolichol-P-mannose, 884
 Dolichol-PP-oligosaccharide synthesis,
 884F, 885
 Domains:
 of life, 6
 of proteins, **247**, 281
 Domain swapping, 718
 Dominant traits, 20–22
 DON (6-Diazo-5-oxo-L-norleucine), 1142
 Donnan equilibrium, 787
 Donohue, J., 88
 Dopamine, 80, 80F, 783, 1058–1060
 Dopamine β -hydroxylase, 1060F
 Dopamine receptors, 689
 Dot matrices, sequence alignment using,
 196–197, 197F
 Dot plot, 196
 Double blind tests, 542–543
 Double-displacement reactions, 499

- Double helical structure, of DNA, **88–90**
 Double-reciprocal plot, 490, 490F, 494F–496F, 499F, 500F
doublesex (dsx) pre-mRNA (*Drosophila melanogaster*), 1318
 Double-sieve mechanism, 1356
 Double-strand breaks, *see* DSBs
 Double-stranded DNA, *see* dsDNA
 Double-stranded RNA, *see* dsRNA
 Double-stranded RNA-activated inhibitor (DAI), 1400
 Double-stranded RNA-activated protein kinase (PKR), 1400
 Doublié, S., 1303
 Doubly wound sheets, 254, 255F
 Douce, R., 1032
 Doudna, J., 1324
 Down mutations, 1266
 Down's syndrome, 312
 Doxorubicin, 1169, 1170
 DPG (2,3-diphosphoglycerate), **329**
dp gene (fruit fly), 25F
 Dpo4, 1222, 1222F
 Drew, H., 89, 1148
 Driving force, 587
 drk protein, 709
 DRMs (detergent-resistant membranes), 411
 Drosophila, 1325
Drosophila melanogaster, 23F
 alternative splicing, 1318, 1319F
cacophony pre-mRNA, 1322
 chromosome number, 19T
 complementation test, 25F
 DNA, 94, 94T, 95F
 Dscam protein, 1318
 genetic experiments, 23–26
 genetic map, 25F
 genome sequencing, 113, 177T, 182
 mitochondria, 1344T
per gene, 120F
 26S proteasome, 1411F
 replication, 1176F, 1202, 1205F
 RNA interference, 1324
 spliceosomes, 1312F
 transposons, 1244, 1246
 ubiquitin, 1409
 Drugs, 78, 396, 968
 Drug design, **539**
 cytochromes P450, 543–545
 discovery techniques, **539–541**
 HIV protease and its inhibitors, **545–551**
 pharmacology, **542–545**
 structure-based, 541
 Drug discovery, **539–541**
 Drug–drug interactions, 544
 Druker, B., 719
 DSBs (double-strand breaks), 1223–1224, 1225F
 Dscam protein, 1318
 dsDNA (double-stranded DNA), 18F
 detection by electrophoresis, 158, 158F
 Klentaq1, 1180
 replication, 102–103, 1174
 from semiconservative replication, 90
 dsRNA (double-stranded RNA), 85, 1323
 DSX-F protein, 1318
 DSX-M protein, 1318
dsx (doublesex) pre-mRNA (*Drosophila melanogaster*), 1318
 DT, *see* Diphtheria toxin
 DTAB (dodecyltriethylammonium bromide), 399F
 dTMP, *see* Deoxythymidine monophosphate
 DUBs (deubiquitinating enzymes), 1417–1418, 1420–1421
 Ducruix, A., 709
 DUEs (DNA unwinding elements), 1194
 Dunathan, H., 1033
 Dunn, M., 1078
 Duplex DNA, 90, 1159F. *See also* dsDNA (double-stranded DNA)
 dUTP, 1218
 dUTP diphosphohydrolase (dUTPase; dUTP pyrophosphate), 1126
 Dutton, L., 841
 Dutzler, R., 755
 du Vigneaud, V., 205
 Duysens, L., 909
 Dwarfism, 685
 Dyad axes, 89F
 Dynamic mutations, 1251–1252
 Dynamin, 436
 Dynamite, 743
 Dynein, 440
 Dystrophin, 1305
 E
 E1, *see* Pyruvate dehydrogenase; Ubiquitin-activating enzyme
 E1b, 799
 EI (Enzyme I), of PTS system, 765
 E1o, *see* α -Ketoglutarate dehydrogenase
 E1p, 799
 E1 protein:
 of bovine papillomavirus, 1184–1185, 1185F
 and Rho factor, 1275–1276
 E2, *see* Dihydrolipoyl transacetylase; Ubiquitin-conjugating enzyme
 EIIA, of PTS system, 765, 767F
 EIIb, of PTS system, 765
 EIIIC, of PTS system, 765
 EIII, of PTS system, 765
 E2o (dihydrolipoyl transsuccinylase), 810
 E3, *see* Ubiquitin-protein ligase
 E3 α , 1413, 1414
 E3b, 799
 E3 binding protein (E3BP), 798
 E3o, 799
 E3p, 799
 E4P, *see* Erythrose-4-phosphate
 E6-associated protein (E6AP), 1414
 E6 protein, 1414
E. coli, 4–5, 4F. *See also* Bacteriophages;
 specific genes
 χ^{176} strain, 123
 Ada protein, 1216
 ampicillin resistant, 26
 aspartokinase isoenzymes, 1075T
 BAC replication in, 109
 and bacteriophages, 27F, 28, 86F, 87, 87F, 1340
 biological structures, 14, 15F
 carbamoyl phosphate synthetase, 1027–1028
 catabolite repression, 1285–1288
 CCA-adding polymerase, 1331
 chromosome replication, 1195
 cloning of, 104
 DNA conservation experiments, 90, 91F
 DNA ligase, 1188, 1188F, 1189F
 DNA methylation, 1247
 DNA of humans vs., 8–9
 DNA repair, 1214
 DNA replication, 102F–104F, 103, 104
 DNA size, 94T
 elongation cycle, 1380F
 enzyme induction, 1260–1264
 fMet, 1374
 genome sequencing, 113, 177T
 GroEL/ES system, 294–302
 heat shock proteins, 293, 294
 homologous recombination, 1228–1230
 inclusion bodies, 117, 117F
 internal structure, 5F
 isocitrate dehydrogenase, 817
lac operon, 97, 97F
 membrane of, 398T
 mismatch repair, 1220, 1220F
 molecular composition, 5T
 multiforked chromosomes, 1195, 1195F
 nonsense suppressors, 1362, 1362T
 nucleotide excision repair, 1217
 Pol I, 1176–1181, 1181T
 Pol I Klenow fragment, 1178F
 Pol II, 1181, 1181T
 Pol III, 1181T, 1182–1184, 1182T, 1183F
 Pol IV, 1182
 Pol V, 1182
 pre-mRNA splicing, 1312
 primase, 1188–1189, 1189F
 promoters, 1266, 1267F
 pUC18 cloning vector, 107, 107F, 111
 pyrimidine synthesis regulation, 1118F
 pyruvate dehydrogenase complex, 792, 793F
 RecBCD, 1231, 1231F
 replication, 1174F, **1192–1200**, 1193F
 replication in, 1183–1187, 1184T
 replication rate, 1174–1175
 replisome, 1196F
 Rho factor, 1276F
 ribosome complements, 1363T
 ribosomes, 367F, 1363–1365, 1363F–1366F, 1370–1371
 RNAP, 1265–1266
 RNA primers, 1176
 RNAP subunits, 1277T
 RuvABC in, 1231
 semidiscontinuous replication, 1175
 shuttle vectors, 118
 σ factors, 1284
 simultaneous transcription and translation, 1283F
 as source of proteins for purification, 130
 termination pathway in ribosomes, 1391F
Ter sites and *oriC* locus, 1199F
 thymine dimer effect, 1214–1215
 topoisomerase III Y328F mutant, 1163–1165, 1164F
 TPP-sensing riboswitch, 1300, 1300F
 transcription in, 1271–1273, 1273F, 1284–1301
 translation in, 1375F, 1376F
 tRNA genes, 1330
 two-dimensional gel electrophoresis of, 152F
 type II restriction enzymes, 105T
 UV absorbance spectra of nucleic acids and DNA, 92F
 E54K mutation, 1239
 Ealick, S., 1116
 Early endosome antigen 1 (EEA 1), 734, 735F
early genes, 1284
 Eaton, W., 346, 353
 4E-BP1, 1401
 4E-BP2, 1401
 4E-BP3, 1401
 Ebright, R., 1286, 1287
 EBS (epidermolysis bullosa simplex), 235
 Echinoderms, 13F
 Eck, M., 716, 721
 Eclipsed conformation, 222, 223F
 EcoRI, 105, 105F, 105T, 106
 EcoRII, 105T

- EcoRV, 105–106, 105F, 105T, 109
 ECS (editing site complementary sequence), 1323F
 Ectodomains, 453, 684
 ED_{50} , 539
 Edema factor (EF), anthrax, 715
 Edidin, M., 408
 Editing:
 of DNA chain, 1170, 1201
 of tRNA, 1355–1358
 Editing site complementary sequence (ECS), 1323F
 Editosome, 1321–1322
 Edman, P., 165, 171
 Edman degradation, 165, 167F, 171, 172, 174
 Edman's reagent, 165, 167F
 EDTA (ethylenediaminetetraacetic acid), 157
 EEA 1, *see* Early endosome antigen 1
 eEF1A, 1388, 1389F
 eEF1B, 1388
 eEF1B α , 1388, 1389F
 eEF1B β , 1388
 eEF2, 1388, 1398, 1405
 EF, *see* Edema factor; Elongation factor
 EF1A, 1379
 EF1B, 1379
 EF2, 1385–1386
 Effectors, 348, 442, 625T, 688
 EF-G, 1375T, **1385–1388**, 1386F
 EF hands, 656
 Efstradiadis, A., 1306
 EF-Ts, 1375T, 1379, 1381–1382, 1382F
 EF-Tu, 1375T
 in chain initiation, 1382F, 1386, 1386F, 1388
 and Gln-tRNA^{Gln}, 1358
 and initiator tRNA, 1380–1381, 1381F
 and translational accuracy, 1390–1391
 Egelman, E., 1228
 EGF, *see* Epidermal growth factor
 EGF receptor, 699
 EH, *see* Enoyl-CoA hydratase
 E Helix, 242F
 EHK (epidermolytic hyperkeratosis), 235
 Ehlers-Danlos syndromes, 240, 1404
 Eicosanoid metabolism, 993–997
 for leukotrienes and lipoxins, 996F, **1000–1004**
 for prostaglandins, prostacyclins, and thromboxanes, 996F, **997–1000**
 Eicosanoid receptors, 689
 5,8,11,14,17-Eicosapentaenoic acid, *see* EPA
 8,11,14-Eicosatrienoic acid, 994
 eIF1A, 1376
 eIF2, 1376, 1399–1401
 eIF2 α , 1399–1401
 eIF2B, 1379, 1399
 eIF2 α kinases, 1399
 eIF2 phosphatase, 1399
 eIF3, 1376
 eIF4A, 1378
 eIF4B, 1365, 1379
 eIF4E, 1378, 1399, 1401
 eIF4F, 1378
 eIF4G, 1327, 1378
 eIF4H, 1379
 eIF5, 1379
 eIF5B, 1379
 eIF n , 1376, 1378
 80S subunit (ribosome), 1370T, 1371, 1371F
 Einstein (unit), 903
 Eisenberg, D., 310, 929, 1068, 1315
 EJC (exon junction complex), 1326–1327
 Eklund, H., 934, 1119, 1122
 Elastase, 525, 525T, 529T
 gene duplication, 194
 proenzymes, 538
 specificity, 169T
 X-ray structure, 528F, 530, 534–535, 535F
 Elasticity coefficient, 622
 Elastin, 240, 529
 Electrical synapses, 778
 Electric fish, 779–780
 Electroblotting, 111
 Electrochemical cells, 584, 584F
 Electrochemical potential, 745
 Electromotive force (emf), 584
 Electron acceptor, 584
 Electron density, 241
 Electron density maps, 241–242, 242F
 Electron donor, 584
 Electronic complementarity, Watson–Crick base pairs, 1155–1156
 Electron pairs, acids/bases and, 45
 Electron paramagnetic resonance (EPR), 911
 Electron paramagnetic spectroscopy, 911
 Electron spin resonance (ESR) spectroscopy, 911
 Electron-transfer flavoprotein (ETF), 948
 Electron-transfer reactions, 586
 Electron transport, 188, 386, 823, **828–845**.
 See also Oxidative phosphorylation and acyl-CoA reoxidation, 948
 Coenzyme Q:Cytochrome c oxidoreductase, 832T, 840–841, 841F
 components of, **833–845**
 cytochrome c oxidase, 832T, 841–845, 842F
 inhibitor studies, 831, 831F
 in mitochondria, **823–828**
 NADH:Coenzyme Q reductase, 832T, 834–835
 net reaction, 824F, 829–830, 829F
 P/O ratio, 831–833, 833F
 in purple photosynthetic bacteria, 909–913, 913F
 reduction potentials of components in testing mitochondria, 832T
 rigid coupling of phosphorylation and oxidation, 831–833
 sequence of, **829–833**
 thermodynamics, **828–829**
 two-center, in photosynthesis, 913–925
 Electron tunneling, 841
 Electrophiles, 563–564, 563F
 Electrophilic catalysis, 510
 Electrophoresis, 132T
 capillary, 152
 gel, 147–150
 isoelectric focusing, 150–152
 nucleic acids, **158–159**
 paper, 146–147
 proteins, **146–152**
 SDS-PAGE, 150
 Electrophoretic mobility, 146
 Electroplaques, 779, 779F
 Electrospray ionization (ESI), 172
 Electrostatic forces, protein stability and, 259–261
 Elements, periodic table, 31F
 Elemental composition, humans, 29T
 Elementary reactions, **483**
 11S regulator, 1418–1419
 Elimination reactions, **566–567**, 567F
 Elion, G., 1130
 ELISA, *see* Enzyme-linked immunosorbent assay
 Ellis, J., 293
 Elongases, **969–971**
 Elongation factor (EF), 422–423, 1375T, **1379**
 Elution, in ion exchange chromatography, 136
 Embden, G., 595
 Embden–Meyerhoff–Parnas pathway, 595.
 See also Glycolysis
 EMBL Nucleotide Sequence Database, 195T
 Embryogenesis, 1250–1251, 1401
 Embryological development, 14F
 Embryonic stem cells, 121
 Emergent properties, 738
 Emf (electromotive force), 584
 Enantiomers, 74–76, 74F
 Endergonic processes, 57
 End groups, amino acid, 165–168
 Endo, T., 446
 Endo conformation, 1153, 1153F
 Endocrine glands, 671, 672F
 Endocrine hormones, 671
 Endocrine system, 672F
 Endocytosis, 8, 433, 454F, 986, 986F
 Endoglycosidases, 366
 Endonucleases, 1160, 1407–1408. *See also* Restriction endonucleases
 Endopeptidases, 169–170, 169T, 194
 Endopeptidase Arg-C, 169T
 Endopeptidase Asp-N, 169T
 Endopeptidase Glu-C, 169T
 Endopeptidase Lys-C, 169T
 Endoplasmic reticulum (ER), 7F, 9, 11F
 metabolic functions, 562T
 rough, *see* Rough endoplasmic reticulum
 secretory pathway in, 420–424
 smooth, *see* Smooth endoplasmic reticulum
 β -Endorphin, 673T, 685
 Endosomes, 455F
 Endothelial NOS (eNOS), 686–688
 Endothelium, 686
 Endothermic processes, 53
 Endotoxins, 688
 δ -Endotoxins, 122
 End point, of titration, 51
 Enediolate anion, 567
 Energy, **53**
 conservation of, 52–54
 and evolution, 34
 metabolic processes, 17
 for prokaryotes, 5–6
 Energy barrier, to free rotation of polypeptide backbone, 222–223
 Energy coupling, 845
 Energy metabolism strategies, 1088–1090
 Energy-rich bonds, 580
 Energy-rich compounds, *see* High-energy compounds
 Energy surface, in landscape theory, 287
 Energy transduction, 587, 845
 Engelman, D., 1365
 Englander, W., 285
 Englemann, T., 901
 Enhancers, 1282–1283, 1337
 Enolase, 567, 596F, **612–613**, 613F
 Enolate, 568, 569F
 Enol forms, of nucleotide bases, 88
 Enol phosphates, 580
 5-Enolpyruvylshikimate-3-phosphate synthase (ESPS-synthase), 1075
 eNOS (endothelial NOS), 686–688
 Enoyl-ACP reductase (ER), 965
 Enoyl-CoA hydratase (EH), 947, 947F, 948, 958
 Enoyl-CoA isomerase, 950
 3,2-Enoyl-CoA isomerase, 952
 Enteropeptidase, 537

- Enthalpy, 53–54
 Entner–Doudoroff pathway, 636–637
 Entropy, 55–57
 of peptidyl transferase reaction, 1384–1385
 and protein folding, 284, 287
 and temperature, 56F
 unitary entropy change, 263T
 env, 1246
 Enzymatic interconversion, *see* Covalent modification
 Enzymes, 4, 163, 469–479. *See also* specific enzymes
 activity control, 474–479
 allosteric control, 476–479
 in assays, 131
 catalytic activity of, 243
 chemical catalysts vs., 469
 coenzymes, 473–474
 for DNA replication, 1176–1189
 DNA replication error repairing, 103–104
 feedback inhibition control for, 475
 geometric specificity, 472–473
 as globular proteins, 241
 historical perspective, 469–470
 for liposome delivery, 396
 nomenclature, 479
 one gene–one enzyme maxim, 26
 processive and distributive, 1177
 selective catalysis by, 588–589
 stereospecificity, 470–472
 substrate specificity, 470–473, 470F
 Enzyme I (EI), of PTS system, 765
 Enzyme catalysis. *See also* Drug design
 acid–base catalysis, 506–510
 covalent catalysis, 510–511
 electrostatic catalysis, 512
 lysozyme, 517–525
 metal ion catalysis, 511–512
 and preferential transition state binding, 515–516, 516F
 and proximity/orientation effects, 512–515
 serine proteases, 525–537
 zymogens, 537–538
 Enzyme Commission, 470
 Enzyme induction, 1260–1264
 Enzyme inhibition, 493–496
 Enzyme kinetics, 482, 487–492
 bisubstrate reactions, 497–501
 competitive inhibition, 493–494
 data analysis, 490–491
 and flux control coefficient, 621–622
 Haldane relationship, 491–492
 Michaelis–Menten kinetics, 488–491, 496–497, 501–503
 mixed inhibition, 495–496
 pH effects, 496–497
 reversible reaction, 491–492
 uncompetitive inhibition, 495
 Enzyme-linked immunosorbent assay (ELISA), 131–132, 132F
 Enzyme-substrate complex, 470F, 488
 Enzymology, 469
 EPA (5,8,11,14,17-eicosapentaenoic acid), 387T, 971, 994, 1002–1003
 Epidermal growth factor (EGF), 453–454, 997, 1317
 Epidermal growth factor (EGF) receptor, 699
 Epidermis, 16F
 Epidermolysis bullosa simplex (EBS), 235
 Epidermolytic hyperkeratosis (EHK), 235
 Epigenetic genome changes, 1249–1250
 Epigenetic reprogramming, 1250–1251
 Epimers, 361
 Epimerization, 567
 Epinephrine, 673T, 679–680
 in fatty acid metabolism regulation, 973
 in glycogen metabolism regulation, 660
 synthesis, 1058–1060
 Epithelial cells, 373, 410
 Epitopes, 205
 EPO, *see* Erythropoietin
 Epoxy-activated agarose, 142, 143F
 2',3'-Epoxy-ATP, 1180
 EPR (electron paramagnetic resonance), 911
 E protein, 1314
 Equatorial groups, in sugars, 363
 Equilibrium, 55, 58–61, 587
 Equilibrium constants, 45–47, 58–59, 58T
 Equilibrium density gradient ultracentrifugation, 90, 91F, 155, 156, 159
 Equilibrium thermodynamics, 587–588
 Equivalence point, 48
 ER, *see* Endoplasmic reticulum; Enoyl-ACP reductase
 Erabutoxin, 781
 ERAD process, 427
 Erbitux, 720
 eRF1, 1392
 eRF3, 1392
 Ergocalciferol, 678
 Ergosterol, 393, 678
 ERKs (extracellular-signal-regulated kinases), 713
 Ernst, O., 694
 ERp57, 886
 Error catastrophe, 587
 Error correction:
 in DNA replication, 103–104
 in RNA replication, 1281, 1281F
 Error-prone DNA polymerases, 1222
 Erythromycin, 1395T, 1396F
 Erythrocyruorins, 325
 Erythrocytes, 324–325
 heme synthesis regulation, 1055–1056
 and sickle-cell anemia, 185–188, 185F, 343–347
 Erythrocyte ghosts, 411
 Erythrocyte glucose transporter, 746–748, 746F, 747F
 Erythrocyte membranes, 411–414, 411F
 and blood groups, 414–415
 composition, 398T
 freeze-etch micrograph, 410F
 glycophorin A, 401F
 Erythromycin A, 968, 969F
 Erythropoietic protoporphyrin, 1055
 Erythropoietin (EPO), 119, 123, 717
 D-Erythrose, 360F
 Erythrose-4-phosphate (E4P), 636, 896, 929
 D-Erythritolose, 361F
 Escherich, T., 4
Escherichia coli, *see* *E. coli*
 ESEs, *see* Exonic splicing enhancers
 ESI (electrospray ionization), 172
 ESI-MS, 173
 E (exit) site, 1373, 1385, 1387
 ESPS-synthase (5-enolpyruvylshikimate-3-phosphate synthase), 1075
 ESR (electron spin resonance) spectroscopy, 911
 Essen, L.-O., 1215
 Essential amino acids, 1019, 1065T, 1072–1078
 Essential fatty acids, 971
 ESSs (exonic splicing silencers), 1306
 Esters, rates of reaction and motional freedom, 595T
 β-Estradiol, 681
 Estrogens, 458, 673T, 680, 991
 ESTs (expressed sequence tags), 181
 ETF (electron-transfer flavoprotein), 948
 ETF:ubiquinone oxidoreductase, 837, 948
 Ethanol, 783. *See also* Yeast alcohol dehydrogenase
 dielectric constant and dipole moment, 43T
 from Entner–Doudoroff pathway, 636, 637F
 from fermentation, 616
 prochiral atoms of, 77F
 protein solubility in, 134
 Ethanolamine, 389T, 390, 1004
 Ethics, of gene therapy, 123–124
 Ethidium ion, 157, 1161–1162, 1162F
 Ethyl-CoA, 874
 Ethylenediaminetetraacetic acid (EDTA), 157
 Ethylene glycol, 265
 O⁶-Ethylguanine residues, 1216
 N-Ethylmaleimide (NEM), 440–441
 Ethylnitrosurea, 1340
 Etoposide, 1169, 1170
 Eubacteria, 6, 192
 Euchromatin, 182
 Eukarya, 6, 7F
 Eukaryotes, 3, 6–14
 cellular architecture, 8–12
 chromosome number, 19T
 citric acid cycle in, 789
 DNA, 86, 88, 94T
 DNA methylation, 1248–1251
 DNA replication, 1201–1213
 evolution of, 13F, 34
 genome sequencing, 176
 glucose transporters, 751
 introns of, 1316
 metabolic pathway location, 562
 mismatch repair, 1220
 mRNA posttranscriptional modification, 1302–1327
 nuclear pores of, 1327
 nucleotide excision repair, 1217
 phylogeny and differentiation, 12–14
 polypeptide synthesis, 1376–1379, 1392
 RecA-like proteins, 1230
 ribosomes, 1369–1371
 rRNA processing, 1328–1329
 secretary proteins, 421F
 translational in, 1376–1379, 1377F
 translation in, 1398–1402
 transposons, 1244–1246
 Eukaryotic gene expression:
 in bacteria vs. eukaryotic cells, 118–119
 and colony hybridization, 114
 and mRNA lifetime, 101
 posttranscriptional modification, 97–98
 RNAP, 96, 1276–1283
 transcriptional initiation, 96
 European corn borer, 122
 Evans, M., 262
 Evans, R., 1104
 Even-chain fatty acids, 945F
 Evolution, 29, 30F. *See also* Chemical evolution; Mutations
 of citric acid cycle, 814–815
 convergent and divergent, 252, 254, 317, 531
 and divergence of kingdoms, 192
 of protein structures, 316–318
 and transposition, 1242
 Evolutionary distance, 189
 Evolutionary tree, 13F
 Excinuclease, 1217
 Excision repair, 1216–1219
 Excitatory synapses, 778
 Exciton transfer, 905

- Exclusion limit, in gel filtration chromatography, 138
- Exercise, ATP for, 1092F
- Exergonic processes, 57
- Exit site, *see* E site
- Exo conformation, 1153, 1153F
- Exocrine gland, 675
- Exocytosis, 8, 440F, 751–752
- Exoglycosidases, 366
- Exons, 98, **1304–1305**
- cassette, 1318
 - and protein modules, 1317
 - splicing, 1305–1308
- Exonic splicing enhancers (ESEs), 1306, 1314
- Exonic splicing silencers (ESSs), 1306
- Exon junction complex (EJC), 1326–1327
- Exon skipping, 1314
- 3' → 5' Exonuclease, 103–104, 104F
- Pol I, 1177, 1180–1181
 - and reverse synthesis of DNA, 1200–1201, 1201F
- 5' → 3' Exonuclease, 103, 103F
- in chain-terminator method, 177
 - Pol I, 1177, 1181
- Exonuclease I, 1220
- Exonuclease VII, 1220
- Exopeptidases, 165–166, 168T
- Exosomes, 1327, 1327F
- Exothermic processes, 53
- Expressed sequences, 97, **1304**
- Expressed sequence tags (ESTs), 181
- Expression, *see* Gene expression
- Expression platform, 1300–1301
- Expression profile, 213
- Expression vector, 117
- Extein, **1405**, 1406F, **1407**
- Extensive quantities, 61
- External nodes, phylogenetic trees, 203, 203F
- Extracellular-signal-regulated kinases (ERKs), 713
- Extrinsic membrane proteins, 400
- Ex vivo* gene therapy, 122
- Eyring, H., 484
- F**
- F1,6P, *see* Fructose-1,6-bisphosphate
- F1F0-ATPase (Complex V, proton-translocating ATP synthase), 758, **852–859**, 865
- F1P, *see* Fructose-1-phosphate
- F2,6P, *see* Fructose-2,6-bisphosphate
- f'C (5-Fluorocytosine) residue, 1248
- F6P, *see* Fructose-6-phosphate
- FAAH, *see* Fatty acid amide hydrolase
- Fab New hemoglobin fragment, 251F
- Fabry's disease, 1013T
- Facilitated diffusion, 420, **750–752**
- Factor IX, 1317
- Factor X, 1317
- Facultative anaerobes, 6
- FAD (flavin adenine dinucleotide), 82, 790F
- biosynthesis, 1138
 - in catabolism, 561F
 - in citric acid cycle, 791, 794, 795T
 - reactions of, 565, 566F
- FADH[·] (flavin adenine dinucleotide, radical form), 565F
- FADH₂ (flavin adenine dinucleotide, fully reduced form), 565F
- in catabolism, 561F
 - in citric acid cycle, 789, 790F, 791, 819
 - in electron-transport chain, 823, 824F, 828, 829
- FAD pyrophosphorylase, 1138
- FAICAR (5-formaminoimidazole-4-carboxamide ribotide), 1109F, 1111
- Familial amyloid polyneuropathy, 310
- Familial Creutzfeldt–Jakob disease, 314
- Familial hypercholesterolemia (FH), 456–457, 466, 989
- Familial type III hyperlipoproteinemia, 459
- Faraday (unit), 53T, 584
- Farber's lipogranulomatosis, 1013T
- Farnesyl group, 406, 838
- Farnesyl pyrophosphate (FPP), 978, 979F
- Farnesyl pyrophosphate synthase, 978
- FAS, *see* Fatty acid synthase
- Fasman, G., 302, 303
- Fast-twitch muscle fibers, 619
- Fat, *see* Adipose tissue
- Fats, 940. *See also* Triacylglycerols
- Fatal familial insomnia (FFI), 314
- Fat cells, *see* Adipocytes
- Fat-soluble vitamins, 474, 993
- Fatty acids, **386–388**
- activation, **945–946**, 946F
 - in citric acid cycle, 789
 - essential, 971
 - glycogen vs., 638
 - and insulin resistance, 1102, 1103F
 - names of, 387T
 - transport in blood, 973
- Fatty acid amide hydrolase (FAAH), 404–406, 405F
- Fatty acid anions, 44, 44F
- Fatty acid biosynthesis, 964F
- acetyl-CoA carboxylase, **962–964**
 - carbon–carbon bond formation, 965F
 - citric acid cycle intermediates, 818
 - elongases and desaturases, **969–971**
 - fatty acid synthase, **964–968**
 - mitochondrial acetyl-CoA transport, **968–969**
 - and other pathways, 1089–1090
 - regulation, 974F
 - triacylglycerol synthesis, **971–973**, 972F
- Fatty acid oxidation, **945**
- and AMP-dependent protein kinase, 1096
 - glycolysis inhibition, 864
 - hormonal regulation, 973–975, 974F
 - minor pathways, **958–959**
 - odd-chain fatty acids, **952–957**
 - and other pathways, 1089–1090
 - β oxidation, **947–950**, 958
 - regulation, **973–975**, 974F
 - and transport, 946–947, 946F
 - unsaturated fatty acids, **950–952**
 - ω – 3 Fatty acids, 1003
 - ω – 6 Fatty acids, 1003
- Fatty acid synthase (FAS), 568, 962, **964–968**
- Fatty acylated proteins, 407
- Fatty acyl-CoA, 947F
- Fatty acyl-CoA desaturases, 970–971, 971F
- Fatty streak lesions, 458
- Favism, 897
- FBLD (fragment-based lead discovery), 541–542
- F-box, 1412, 1413F
- F-box protein family, 1412
- FBP, *see* Fructose-1,6-bisphosphate
- FBPase, *see* Fructose-1,6-bisphosphatase
- FBPase-2 (fructose bisphosphatase-2), 663
- FC0 mutation (phage T4), 1340
- FC1 mutation (phage T4), 1340
- FC2 mutation (phage T4), 1340
- FC3 mutation (phage T4), 1340
- FC4 mutation (phage T4), 1340
- FC5 mutation (phage T4), 1340
- FCCP (carbonylcyanide-*p*-trifluoromethoxyphenylhydrazone), 860
- F⁺ cell, 1262–1263
- F[–] cell, 1262–1263
- Fd, *see* Ferredoxin
- FDP (fructose-1,6-diphosphate), 600
- FdUMP (5-fluorodeoxyuridylate), 1127–1129
- Feathers, 233
- Feedback inhibition, 474–479, 620, 623–624
- Feher, G., 910
- Feigon, J., 1212
- Female sex hormones, 681
- FeMo-cofactor, 1081–1082
- Fen (fenfluramine), 543
- FEN-1 (flap endonuclease-1), 1207
- Fenfluramine (fen), 543
- Fenn, J., 172
- Fen–phen, 543
- Fe-protein, 1080–1082
- Fermentation, 6, 469–470, 593–594
- alcoholic, 593–594, **616–619**
 - energetics, 619
 - homolactic, 593, **614–616**
- Ferns, 13F
- Ferredoxin (Fd), 835, 835F, 934F
- Ferredoxin-NADP⁺ reductase (FNR), 924, 924F
- Ferredoxin-thioredoxin reductase (FTR), 933–934, 934F
- Ferritin, 140F, **381**
- Ferrochelatase, 1055
- Fe–S cluster, *see* Iron–sulfur clusters
- Fesik, S., 707
- Fetal hemoglobin, 193–194, 331
- Fetulin, 140F
- F factor, 1262–1263
- F' factor, 1263
- Ffh polypeptide, 422
- FFI (fatal familial insomnia), 314
- FGAM, *see* Formylglycinamide ribotide
- FGAM synthetase (PurL), 1109F, 1110
- FGAR, *see* Formylglycinamide ribotide
- FGF, *see* Fibroblast growth factor
- FH, *see* Familial hypercholesterolemia
- Fibrillarin, 1329
- Fibrils, 238–240
- Fibrin, 191, 310
- Fibrinogen, 191
- in amyloid diseases, 310
 - frictional ratio, 155
 - isoelectric point (human), 134T
 - molecular mass, 140F
 - physical constants (human), 153T
 - solubility, 133F, 134
- Fibrinogen amyloidosis, 310
- [Glu¹]Fibrinopeptide B, 175F
- Fibrinopeptides, 190, 191F
- Fibroblasts, 9F, 454F, 1211–1212
- Fibroblast growth factor (FGF), 375, 700
- Fibroblast interferon, 1399
- Fibromodulin, 373T
- Fibronectin, 684
- Fibrous proteins, 226, **232–240**. *See also*
- Collagen; Keratin
 - collagen, **235–240**
 - keratin in, **233–235**, 234F
- Fick's first law of diffusion, 745
- Fields, S., 705
- Fiers, W., 176
- 50S subunit (ribosome), 1363T, 1366–1368, 1368F, 1371, 1384F
- Fight or flight response, 664
- Filamentous bacteriophage M13, 108, 108F
- Fillingame, R., 853

- Filmer, D., 352
 Fingerprinting, 147, 174
 Finishing, 181
 Fink, G., 1245
 Finzel, B., 1151
 Fire, A., 1323
 FirstGlance, 256T, 257
 First law of thermodynamics, 52–54
 First-order rate equation, 483–484
 First-order reactions, 483, 484F
 Fischer, A., 123
 Fischer, E., 70, 75, 352, 360, 470, 651, 722
 Fischer convention, 75–76
 Fischer projections, 75, 75F
 Fish, embryonic development of, 14F
 Fission yeast, *see Schizosaccharomyces pombe*
 Fitzgerald, P., 549
 5' end, of nucleic acids, 84
 FK506, 292–293
 FK506 binding protein (FKBP12), 292, 293, 724–725
 FKBP12-rapamycin-associated protein (FRAP), 737F
 Flagella, 4, 4F, 10
 Flagellates, 7F
 Flagellin, 10, 1243, 1388
 FLAP, *see* 5-lipoxygenase-activating protein
 Flap endonuclease-1 (FEN-1), 1207
 Flatworms, trans-splicing in, 1320
 Flavin, 566
 Flavin adenine dinucleotide, *see* FAD
 Flavin adenine dinucleotide, fully reduced form, *see* FADH₂
 Flavin adenine dinucleotide, radical form (FADH·), 565F
 Flavin coenzymes, 473T, 1138
 Flavin mononucleotide, *see* FMN
 Flavin mononucleotide, radical form, *see* FMNH·
 Flavin mononucleotide, reduced form, *see* FMNH₂
 Flavobacteria, 7F
 Flavodoxin, 135F
 Flavokinase, 1139F
 Flavoprotein, 801
 Flavoprotein dehydrogenase, 828, 828F
 Flavr Savr tomato, 1403
 Fleming, A., 376
 Fletterick, R., 433, 530, 640
 Flipases, 420
 Flip-flop, 396, 419–420
fljA gene, 1243
fljB gene, 1243
fljC gene, 1243
 Fluid mosaic model, of membranes, 408–411
 9-Fluorenylmethoxycarbonyl (Fmoc) group, 207
 Fluorescamine, 149
 Fluorescence, 286, 904–905
 Fluorescence recovery after photobleaching (FRAP), 396–397, 396F
 Fluorescence resonance energy transfer (FRET), 284, 286, 286F, 299, 300
 Fluoroacetyl-CoA, 809
 Fluorochlorobromomethane, enantiomers, 74F
 Fluorocitrate, 809
 (2*R*,3*R*)-2-Fluorocitrate, 809
 5-Fluorocytosine (¹⁴C) residue, 1248
 5-Fluorodeoxyuridine, 1129
 5-Fluorodeoxyuridylate (FdUMP), 1127–1129
 Fluorophore, 396
 5-Fluorouracil, 1129
 Flurbiprofen, 998
 Flux, 589, 593, 619, 745
 Flux control coefficient, 620–622
 fMet, *see* N-Formylmethionine
 fMet-tRNA_f^{Met}, 1373–1374, 1376
 FMN (flavin mononucleotide), 686, 834, 836F, 1138
 FMNH· (flavin mononucleotide, radical form), 836F
 FMNH₂ (flavin mononucleotide, reduced form), 836F
 Fmoc (9-fluorenylmethoxycarbonyl) group, 207
FMRI gene, 1251, 1252
 FMR protein (FMRP), 1251
 FNR, *see* Ferredoxin-NADP⁺ reductase
 Foam cells, 457
 Fodor, S., 212
 Folate, 1062
 Folds, 249F, 250–251
 Folding-trap mechanism, 448
 Foldons, 289, 289F
 Fold recognition, 305
 Folic acid, 474T, 1062
 Follicle-stimulating hormone (FSH), 673T, 683
 Following substrate, 498
 Fomivirsen (Vitravene), 1403
 Fontecilla-Camps, J., 941
 Food, energy content of, 941T
 Footprinting, 1267
 Forespore, 1284
 Fork-junction promoter DNA fragment, 1268
 Formaldehyde, 505
 Formamide, 43T
 5-Formaminoimidazole-4-carboxamide ribotide, *see* FAICAR
 Formate dehydrogenases, 1361
 Formic acid, 32T
N-Formiminoglutamate, 1034
*N*⁵-Formimino-tetrahydrofolate, 1034, 1062
 Formylglycynamide ribotide (FGAR), 1109F, 1110
 Formylglycaminide ribotide (FGAM), 1109F, 1110
N-Formylmethionine (fMet), 79, 79F, 1373–1374
N-Formylmethionine-tRNA_f^{Met} (Met-tRNA_f^{Met}), 1373–1374
 Formylmethionyl-tRNA, 1063
*N*¹⁰-Formyl-tetrahydrofolate, 1062
 Forskolin, 698
 Förster, T., 286
 Förster distance, 286
 48S Initiation complex, 1379
 40S subunit (ribosome), 1370T
 43S Preinitiation complex, 1376–1377
 Fos, 713, 737F
 Fosamax, 679F
 Fossil fuels, 901
 Fossil record, 29, 192
 Foster, S., 376
 4-Helix bundle, 241, 250F
 454 Life Sciences, 182
 454 sequencing system, 182–184, 183F
 4-Helix bundle, of SNAREs, 441–442
 4.4₁₆ Helix, 228
 Fowler's solution, 799
 Fowlpox virus, 94T
 F pili, 1262
 FPP, *see* Farnesyl pyrophosphate
 F protein, 1314
 Fractional saturation, 325
 Fractionation, 132, 156–159
 Fragile X syndrome, 1251–1252
 Fragment-based lead discovery (FBLD), 541–542
 Frame, 1341
 Frameshift mutations, 1222, 1341
 Frameshift suppressors, 1362
 Franceschii, F., 1376
 Frank, H., 262
 Frank, J., 422, 1365, 1371, 1388
 Franklin, B., 393–394
 Franklin, R., 88
 FRAP (FKBP12-rapamycin-associated protein), 737F
 Free energy, 57–58
 and concentration, 58, 61
 and enzyme catalysis, 537
 and equilibrium constants, 58T
 membrane transport, 744
 and protein folding, 284, 287
 standard state, 58–60
 unitary Gibbs free energy change, 263T
 Free energy of activation, 486, 487
 Free energy of formation, 59–60, 59T
 Freeman, H., 921
 Freeze-etch technique, 408–410, 409F
 Freeze-fracture technique, 408–410, 409F
 French press, 130
 FRET, *see* Fluorescence resonance energy transfer
 Frey, P., 536
 Frictional coefficient, 146, 155
 Frictional ratio, 155
 Fridovich, I., 865
 Frogs, chromosome number of, 19T
 Fromme, P., 922
 Fructofuranose, 361
 β-D-Fructofuranose, 362F
 β-Fructofuranosidase, 488
 Fructokinase, 630
 Fructose, 361, 361F, 362F, 630–631, 631F
 Fructose-1,6-bisphosphatase (FBPase), 582, 630, 877, 933F
 and glycolysis, 593, 594F, 629
 and metabolic cycles, 1088–1089
 Fructose-1,6-bisphosphate (FBP; F1,6P), 479, 596F, 600
 Fructose-1,6-diphosphate (FDP), 600
 Fructose-1-phosphate, 630
 Fructose-1-phosphate aldolase, 630
 Fructose-2,6-bisphosphate (F2,6P), 627, 663–664
 Fructose-6-phosphate (F6P), 582, 582F
 and erythrose-4-phosphate, 636
 in glycolysis, 596F, 598–600
 Fructose bisphosphatase-2 (FBPase-2), 663
 Fructose intolerance, 630
 Fruit flies, *see* *Drosophila melanogaster*
 Frydman, J., 301
 FSH, *see* Follicle-stimulating hormone
 FTR, *see* Ferredoxin-thioredoxin reductase
 FtsY, 423
 F-type ATPases, 758
 F-type H⁺-ATPase, 852, 854F
 L-Fucose, 365
 Fumarase, 481, 567, 947
 in citric acid cycle, 790F, 791, 812F
 diffusion-limited, 490
 Michaelis–Menten kinetic constants, 489T
 molecular mass, 140F
 Fumarate, 493, 947
 in citric acid cycle, 790F, 791, 811
 degradation of phenylalanine and tyrosineto, 1043–1047
 hydration, 481
 Fumarate hydratase, 812–813

- Functional groups:
 in Cahn–Ingold–Prelog system, 76
 hydrogen bonding, 43F
 p*K* values of, 72
- Fungi, 7F, 12, 13F
 divergence of, 192
 FAS-I, 967–968, 967F
- Furan, 361
- Furanoses, 361
- Furey, W., 616
- Furylfuramide, 1225
- Fusidic acid, 1395T, 1396F
- Fusion pore, 442, 443F
- Fusion proteins, 118, 431
- Futai, M., 857
- Futile cycle, 629
- Fyn, 708, 715
- FYVE domains, 734, 735F
- G**
- G_0 phase, 1163, 1206
- G_{ta} , *see* Transducin
- $G_{1,6}\text{P}$ (glucose-1,6-bisphosphate), 642
- $G_{1}\text{P}$, *see* Glucose-1-phosphate
- G_1 phase, 1202, 1205, 1206F
- G*, *see* Free energy
- G_2 phase, 1202, 1206
- $G_6\text{P}$, *see* Glucose-6-phosphate
- $G_6\text{Pase}$, *see* Glucose-6-phosphatase
- $G_6\text{PD}$, *see* Glucose-6-phosphate dehydrogenase
- $G_6\text{P}$ translocase, 661
- $G_{\text{A}2}$ ganglioside, 1011
- GABA*, *see* γ -Aminobutyric acid
- Gab-1* (Grb2-associated binder-1), 738
- gag*, 1245
- Gag-pol* polyprotein, 546
- Gag* polyprotein, 546
- Gal*, *see* Galactose
- Galactitol, 633
- Galactans, 366
- Galactocerebrosides, 391, 1009–1010
- Galactocerebroside-3-sulfate, 1010
- Galactokinase, 631
- D -Galactosamine, 365
- D -Galactose, 110, 360F, 361, 392F
- Galactose (Gal):
 and glucose in *E. coli*, 1286
 in glycolysis, 630–633, 632F
- Galactose-1-phosphate uridylyltransferase, 631
- Galactosemia, 632–633
- β -Galactosidase, 107F, 110, 366, 367
 in enzyme induction, 1261–1262
 molecular mass, 140F
 sequencing, 165
- Galactoside permease, 769. *See also*
 Lactose permease
- 1- β -Galactosylceramide, 1010
- Galactosyl diacylglycerol, 902
- Galactosyltransferase, 882
- D -Galacturonic acid, 364
- Gallo, R., 545
- Gallus gallus*, genome sequencing of, 177T
- GalNAc transferase, 890
- Gametes, 20
- γ Complex, Pol III, 1182
- Gamma radiation, 572
- $\gamma\delta$ Resolvase, 1239–1241, 1241F
- $\gamma\delta$ Transposon, 1239–1241, 1241F
- Gangliosides, 392, 392F, 1008
 biosynthesis, 1010–1011, 1011F
- GAP*, *see* Glyceraldehyde-3-phosphate; GTPase-activating protein
- GAP-containing inositol phosphatase (GIP), 735
- GAPDH, *see* Glyceraldehyde-3-phosphate dehydrogenase
- Gap junctions, membranes, 415–416, 416F
- Gap penalties, 201
- GAPs, 705–711
- GAR* (glycynamide ribotide), 1109F
- Garavito, M., 997
- Garrod, A., 25–26
- GAR* synthetase (PurD), 1109F, 1110
- GAR* transformylase (PurN), 1109F, 1110
- Gas constant, 53T
- Gas–liquid chromatography (GLC), 366
- Gassman, P., 536
- Gastric inhibitory peptide (GIP), 673T, 675
- Gastric mucosa, 764–765
- Gastrin, 673T, 675
- Gastroesophageal reflux disease (GERD), 764
- Gastrointestinal hormones, 675–676
- Gastrula, 1250
- GatA subunit of Glu-Adt, 1358
- GatB subunit of Glu-Adt, 1358
- GatC subunit of Glu-Adt, 1358
- Gated pore, 750–751
- Gating, of ion channels, 771–775
- Gaucher's disease, 1013T
- Gay-Lussac, J., 469
- GC (guanylate cyclase), 687
- GC box, 1282
- GCN2, 1400–1401
- GCN4, 1400–1401
- GCPR kinases (GRKs), 697
- G-CSF (granulocyte colony-stimulating factor), 717
- $G_{\text{D}3}$ ganglioside, 1011
- GDH*, *see* Glutamate dehydrogenase
- GDI*, *see* GDP dissociation inhibitor; Guanine nucleotide dissociation inhibitor
- GDP (guanosine diphosphate):
 in citric acid cycle, 790F, 791
 and EF-Tu, 1380–1381, 1381F
 in secretory pathway, 420, 423–424
 in vesicle formation, 434–435, 437
- GDP dissociation inhibitor (GDI), 441
- GEFs, *see* Guanine nucleotide exchange factors
- Geiger counting, 572
- Gelatin, 236
- Gel electrophoresis:
 for measuring supercoiling, 1162
 of nucleic acids, 158–159, 158F, 179
 of proteins, 132T, 147–150, 148F
 pulsed-field, 158–159, 159F
 for restriction maps, 106, 106F
 two-dimensional, 152, 152F
- Gel filtration chromatography, 138–141, 139F, 157, 169
- Geminin, 1206–1207
- GenBank, 195T
- Genes, 21–22. *See also* Chromosomes; DNA and bacterial conjugation, 1262–1264
 chemical synthesis, 118
 duplication of, 193–194, 317–318
 fruit fly experiments, 22–25
 liposome delivery, 396
 manipulation, 109–111
 one gene–one enzyme maxim, 26
 reporter, 120–121, 120F
 self-splicing, 1306–1308
 sex linked, 24
 and tetranucleotide hypothesis, 85
- Gene A protein, 1191, 1193
- Gene expression, 25–26, 95–104, 95F
 DNA replication, 101–104
- eukaryotic, *see* Eukaryotic gene expression overview, 18–19, 95–101
- prokaryotic, *see* Prokaryotic gene expression
- and RNAi, 1325–1326
- transcription, 95–98
- translation, 95, 98–101
- General acid catalysis, 506–507
- General base catalysis, 507
- General import protein (GIP), 446
- General recombination, 1225
- Gene sequence databanks, 195T
- Gene splicing, 97–98, 1305. *See also*
 Spliceosomes
 alternative splicing, 1317–1318
 cis- and trans-splicing, 1319–1320
 exons, 1305–1308
 pre-mRNAs, 1312
 role of splicing-associated factors, 1312, 1314
 self-splicing genes, 1306–1308
 significance of, 1316–1317
- Gene therapy, 122–124, 991
- Genetic anticipation, 1251
- Genetic code, 87, 88, 1338–1345
 briefly summarized, 1421
 deciphering, 1341–1343
 mitochondrial deviations from ideal, 1344T
 nature of, 1343–1344
 nonuniversality of, 185, 1344–1345
 standard, 1343T
 in translation, 99, 100T
- Genetic control, metabolic flux and, 624
- Genetic crosses, 20, 21F
- Genetic diseases, 26, 112–113, 122–123.
See also specific diseases
- Genetic engineering, 104. *See also*
 Molecular cloning; Recombinant DNA technology
- Genetic information, 18–19. *See also* Genes
- Genetic maps, 24, 25F
- Genetic mutations, *see* Mutations
- Genetic recombination, 1225
- Genetics, 19–28. *See also specific topics*
 bacterial, 26
 chromosomal theory of inheritance, 22–26
 chromosomes, 19–20
 Mendelian inheritance, 20–22
 viral, 26–28
- Genome, 8–9, 94, 1173. *See also* Human genome
- Genome Analyzer, 184
- GenomeNet, 195T
- Genome sequencing, 176, 177T, 180–185, 180F
- Genomes OnLine Database (GOLD), 195T
- Genomic imprinting, 1251
- Genomic libraries, 113–114
- Genotypes, 21–22, 22F
- Gentisic acid, 173
- Geometric complementarity, of Watson–Crick base pairs, 1154–1155
- Geometric specificity, enzymes, 472–473
- George III, King of England, 1056
- Geranylgeraniol, 903
- Geranylgeranyl groups, 406
- Geranyl pyrophosphate, 978–979
- GERD (gastroesophageal reflux disease), 764
- Gerhart, J., 475
- Gerlt, J., 536
- Germ cells, 19, 20, 21F
- Germ plasm theory, 19

- Gerstmann–Sträussler–Schneiker (GSS) syndrome, 314
- GFP (green fluorescent protein), 120–121, 120F, 121F
- GGAs, 436
- GH, *see* Growth hormone
- Ghrelin, 1099
- Giardia intestinalis*, 1324, 1324F
- Gibbs, J. W., 57
- Gibbs free energy, 57–60, 58T. *See also* Free energy
- Gierasch, L., 294
- Gigantism, 685
- Gilbert, W., 1284
- Gilman, A., 690
- GIP, *see* GAP-containing inositol phosphatase; Gastric inhibitory peptide; General import protein
- Girvin, M., 853
- GK, *see* Glucokinase
- GKRP (glucokinase regulatory protein), 662
- GLA, *see* γ -Linolenic acid
- GLC (gas-liquid chromatography), 366
- GlcNAc, *see* NAG
- Gleevec, 719–720
- Gln, *see* Glutamine
- GlnRS, 1352–1354, 1353F, 1355F
- Gln-tRNA^{Gln}, 1358–1359, 1359F
- Globins, 193–194, 193F, 324
- Globin fold, 251
- Globosides, 1008, 1010–1011, 1011F
- Globular proteins, 241
- α helix in, 226
 - β pleated sheets in, 230
 - hierarchical organization, 281–282, 281F
 - nonrepetitive structures in, 230–233
 - structural bioinformatics, 256–259
 - tertiary structure, 245–256
 - X-ray and NMR structure, 241–245
- γ-Globulins, 134T, 140F
- Glu, *see* Glutamic acid
- Glu-AdT, *see* Glu-tRNA^{Gln} amidotransferase
- Glutathione peroxidase, 897
- Glucagon, 140F, 660, 673T, 973
- Glucans, 366
- Glucocerebrosides, 391, 1010
- Glucocorticoids, 673T, 680, 991, 1090
- Glucocorticoid hormone response element (GRE), 879
- Glucocorticoid receptor, 879
- Glucogenic amino acids, 1029–1030, 1090, 1095
- Glucogenesis, 359
- Glucokinase (GK), 597, 662–664
- Glucokinase regulatory protein (GKRP), 662
- Gluconeogenesis, 562, 571, 871–880
- and AMP-dependent protein kinase, 1096
 - and citric acid cycle, 818
 - and Cori cycle, 880
 - gluconeogenesis pathway, 872–878
 - and glycolysis, 595, 878F
 - and other pathways, 1088–1089
 - regulation of, 878–879
 - and starvation, 1101
 - stimulation by FBPase-1, 664
- Gluconic acid, 364
- 1,5-Gluconolactone, 642
- D-Glucono- δ -lactone, 364F
- Glucopyranose, 361
- α-D-Glucopyranose, 362F
- β-D-Glucopyranose, 362F, 363F
- D-Glucosamine, 365
- Glucose, 565F. *See also* Blood glucose; Electron-transport chain
- complete oxidation of, 823
 - and *E. coli*, 1285–1286, 1286F
 - in Entner–Doudoroff pathway, 636, 637F
 - membrane permeability coefficients, 747T
 - mutarotation of, 507–508
 - and noninsulin-dependent diabetes, 1103F
 - transport in blood, 973
- Glucose-1,6-bisphosphate (G1,6P), 642
- Glucose-1-phosphate (G1P), 370, 631, 639, 932
- Glucose-6-phosphatase, 877
- Glucose-6-phosphatase (G6Pase), 661, 666, 1088–1089
- Glucose-6-phosphate (G6P), 565F, 638–639
- as energy-rich compound, 580
 - in Entner–Doudoroff pathway, 637F
 - in glycolysis, 596F, 597–598
 - and metabolism, 1089, 1094
- Glucose-6-phosphate dehydrogenase (G6PD), 894, 894F, 897–898
- Glucose-6-phosphate isomerase, 598
- D-Glucose, 359–362, 362F, 392F
- L-Glucose, 360
- α-D-Glucose, 363F, 507
- β-D-Glucose, 507
- Glucose-alanine cycle, 878, 1022–1023
- Glucose-dependent insulinotropic polypeptide, 675
- Glucose-fatty acid cycle, 864
- Glucose metabolism, 567
- Glucose transport:
- ATP-driven active, 758–768
 - in erythrocytes, 746–748, 746F, 747F and facilitated diffusion, 750–752, 750F, 751F
 - ion-driven active, 768–771
 - maltoporin in, 749–750
 - in muscle and fat, 751F
- α-Glucosidase, 370
- α,1,4-Glucosidase deficiency, 666
- Glucosidase II, 886
- 1-β-Glycosylceramide, 1009–1010
- D-Glucuronic acid, 364
- D-Glucurono- δ -lactone, 364F
- GluProRS, 1351
- GLUT1, 751
- GLUT2, 662, 751
- GLUT3, 751
- GLUT4, 661, 737F, 751
- GLUT4 storage vesicles, 751
- GLUT5, 751
- GLUT12, 751
- Glutamate, 71. *See also* Glutamic acid
- biosynthesis, 1065–1071
 - degradation, 1034
 - and nitrogen fixation, 1080
 - as nonessential amino acid, 1065T
- Glutamate-5-phosphate, 1070F, 1071
- Glutamate-5-semialdehyde, 1034, 1070F, 1071
- Glutamate–aspartate aminotransferase, 1022
- Glutamate–aspartate carrier, 827–828
- Glutamate dehydrogenase, 266
- Glutamate dehydrogenase (GDH), 153T, 818, 1020, 1023–1033, 1024F
- Glutamate receptor pre-mRNA, 1322
- Glutamate synthase, 934, 1065–1067
- Glutamic acid, *see also* Glutamate
- degradation, 1034
 - genetic codes for, 100T, 1343T
 - in globular proteins, 246
 - half-life, 1413T
 - α helix/β sheet propensity, 302T
- isoelectric point, 72
- in Miller–Urey experiments, 32T
- in native unfolded proteins, 283
- in PEST proteins, 1413
- side chains, 71, 264T
- structure and general properties, 69T
- Glutaminase, 1023, 1025–1027, 1034
- Glutamine:
- amino group, 208F
 - biosynthesis, 1067–1073
 - degradation, 1034
 - and *E. coli* nonsense suppressor, 1362T
 - genetic codes for, 100T, 1343T
 - in globular proteins, 246–247
 - half-life, 1413T
 - α helix/β sheet propensity, 302T
 - in native unfolded proteins, 283
 - as nonessential amino acid, 1065T
 - side chains, 71, 208F, 264T
 - structure and general properties, 69T, 71F
- Glutamine amidotransferase, 1025–1027, 1066–1067, 1078
- Glutamine PRPP aminotransferase (PurF), 1108, 1111
- Glutamine synthetase, 269F, 591, 1023, 1029, 1068–1071
- γ-Glutamyl cycle, 1061, 1061F
- γ-Glutamyl cyclotransferase, 1061
- γ-Glutamylcysteine synthetase, 1061
- Glutamyl group, 73T
- γ-Glutamyl kinase, 1070F, 1071
- γ-Glutamylphosphate, 1068
- γ-Glutamyltransferase, 1001
- γ-Glutamyl transpeptidase, 1061
- Glutaraldehyde, 270, 270F
- Glutaredoxin, 1124
- Glutathione, 142
- Glutathione (GSH), 544–545, 544F, 803, 1060–1062
- Glutathione disulfide (GSSG), 803
- Glutathione peroxidase, 866, 1003, 1060F
- Glutathione reductase (GR), 801, 803, 1060F
- Glutathione-S-transferase, *see* GST
- Glu-tRNA^{Gln}, 1358
- Glu-tRNA^{Gln} amidotransferase (Glu-AdT), 1358, 1359F
- Glu-tRNA^{Glu}, 1358
- Glx, 68T, 73. *See also* Glutamic acid; Glutamine
- Gly, *see* Glycine
- Glycans, 365–366
- Glyceraldehyde, 75, 75F
- D-Glyceraldehyde, 75, 75F, 360F, 361
- L-Glyceraldehyde, 75, 75F, 76, 77F
- Glyceraldehyde-3-phosphate, 636
- from Calvin cycle, 927–929, 931
 - in Entner–Doudoroff pathway, 637F
 - in glycolysis, 595, 596F, 600–603
 - topological diagram, 254F
- Glyceraldehyde-3-phosphate dehydrogenase (GAPDH):
- domains, 248, 249F
 - and erythrocyte membrane, 412
 - in glycolysis, 596F, 607–608, 607F, 608F
 - half-life, 1408T
 - molecular mass, 140F
- (S)-Glyceraldehyde, 76, 77F
- Glyceraldehyde kinase, 630
- Glycerate, 935
- L-Glycero-D-mannoheptose, 379F
- Glyceroglycolipids, 1004
- Glycerol, 364, 388, 389T
- Glycerol-3-phosphate, 580, 828, 828F
- sn-Glycerol-3-phosphate, 389, 389F
- Glycerol-3-phosphate acyltransferase, 971

- Glycerol-3-phosphate dehydrogenase, 828, 828F, 944
- Glycerolipids, 1004F
- Glycerol kinase, 630, 766, 944
- Glyceroneogenesis, 971, 973, 1093
- Glycerophosphate shuttle, 828, 828F
- Glycerophospholipids, 389–390, 389F, 394–395, **1004–1008**
- N*⁶- β -Glyceral lysine, 601
- Glycidol phosphate, 604
- Glycinamide ribotide (GAR), 1109F
- Glycine:
- biosynthesis, 1071–1072
 - as chemical messenger, 79–80
 - conjugates of, 993, 993F
 - degradation, 1030–1034
 - electrostatic influences, 72
 - genetic codes for, 100T, 1343T
 - half-life, 1413T
 - α helix/ β sheet propensity, 302T, 304
 - in heme biosynthesis, 1048
 - isoelectric point, 72
 - in Miller–Urey experiments, 32T
 - in native unfolded proteins, 283
 - as nonessential amino acid, 1065T
 - optical activity, 74
 - steric hinderance, 224–225, 224F, 225F
 - structure and general properties, 68T
 - titration curve, 72, 72F
- Glycine cleavage system, 1031–1033
- Glycine decarboxylase multienzyme system, 1031–1033
- Glycine synthase, 1031–1033
- Glycocalyx, 381, 414, 414F
- Glycoconjugates, 359
- Glycoforms, 380
- Glycogen, 359, **370**, 370F, 638
- fatty acids vs., 638
 - particle fabrication, 646–647
- Glycogen branching, **646–647**
- Glycogen breakdown, **638–644**
- glycogen debranching enzyme, 370, 639, **642–643**, 667
 - glycogen phosphorylase, **639–641**
 - and other pathways, 1089
 - phosphoglucomutase, 639, **642**
- Glycogen debranching enzyme, 370, 639, **642–643**, 667
- Glycogenin, 380, 645–646, 646F
- Glycogen metabolism, **638–668**
- blood glucose maintenance, **661–664**
 - control of, **647–666**
 - covalent modification in, **650–651**
 - glycogen breakdown, **638–644**
 - glycogen phosphorylase bicyclic cascade, **651–660**
 - glycogen storage diseases, **666–668**
 - glycogen synthase bicyclic cascade, **660**
 - glycogen synthesis, **644–647**
 - stress response, **664–665F**
 - thermodynamics, 643–644
- Glycogen phosphorylase, 370, **639–641**
- allosteric control, 479, **647–650**
 - bicyclic cascade, **651–660**
 - covalent modification, 1405
 - deficiency, 667
- Glycogen storage diseases, **666–668**
- Glycogen synthase, **644**, 645, **647–650**
- Glycogen synthase kinase-3 (GSK3), 660
- Glycogen synthesis:
- glycogen branching, **646–647**
 - glycogen synthase, **644**
 - and other pathways, 1089
 - UDP-glucose pyrophosphorylase, **644**
- Glycolate, 935
- Glycolate oxidase, 935
- Glycolate phosphatase, 935
- Glycolic acid, 32T
- Glycolipids, 359, 381, 398T, 400F
- Glycolipid metabolism, 1004
- glycerophospholipids, **1004–1008**
 - sphingoglycolipids, **1008–1013**
 - sphingophospholipids, **1008–1009**
- Glycolysis, 569, **593–634**. *See also* Electron-transport chain; Fermentation
- aerobic, 864
 - aldolase, 596F, **600–603**
 - and AMP-dependent protein kinase, 1096
 - anaerobic, 614–619, 864
 - and citrate, 863–864
 - effectors of nonequilibrium enzymes, 625T
 - electron-transport chain, 824F
 - enolase, 596F, **612–613**
 - and fatty acid oxidation, 864
 - fermentation, **614–619**
 - and free energy changes, 625T
 - fructose utilization, 630–631
 - galactose utilization, 630–633, 632F
 - and gluconeogenesis, 878F
 - glyceraldehyde-3-phosphate dehydrogenase, 596F, **607–608**
 - hexokinase, 596F, **597–598**
 - mannose utilization, 630, 633, **633**, 633F
 - and metabolic regulation and control, **619–630**
 - net reaction, 594F, 595
 - and other pathways, 1088
 - overview of, **595–596**, 596F
 - phosphofructokinase, 596F, **600**, 625–627, 626F, 630
 - phosphoglucose isomerase, 596F, **598–600**
 - phosphoglycerate kinase, 596F, **608–610**
 - phosphoglycerate mutase, 596F, **610–612**
 - pyruvate kinase, 596F, **613–614**
 - for rapid ATP production, 619
 - regulation in muscle, **624–630**
 - triose phosphate isomerase, 596F, **603–606**
- Glycomes, 382–383
- Glycomics, **382–383**, 578
- Glycone, 366
- Glycophorin A, 400–401, 401F, 413F
- Glycoproteins, 359, **373**
- in bacterial cell walls, 375–379
 - biosynthesis, **880–892**
 - formation, 101
 - glycomics, **382–383**
 - proteoglycans, **373–375**
 - structure and function, **379–382**
- Glycoprotein carbohydrates, **380–381**
- Glycosaminoglycans, **370–372**
- Glycosidases, 364
- Glycosides, 363
- Glycosidic bonds, 83, **363–365**, 880–881, 883F
- Glycosphingolipids (GSLs), 391, 1004
- Glycosylation, 101
- Glycosylceramides, 1009–1010
- Glycosyl-enzyme intermediate, 522
- Glycosyl group transfer, 565
- Glycosylphosphatidylinositol (GPI), 407–408, 408F, 890–892, 891F
- Glycosylphosphatidylinositol (GPI) membrane anchors, 882, 891F
- Glycosyltransferases, 881, 1010
- Glyoxylate, 935
- Glyoxylate cycle, **880**, 881F
- Glyoxylate pathway, 10
- Glyoxylic acid, 1135F
- Glyoxosome, 10, 880
- Glycosate, 1075
- G_{M1} Gangliosidosis, 1013T
- G_{M2}-activator protein, 1011, 1012F
- GM-CSF, *see* Granulocyte-macrophage colony-stimulating factor
- GM-CSF receptor, 717–718, 718F
- GMP (guanosine monophosphate), 83T, **1111–1113**, 1112F, 1130F
- GMPPNP (guanosine-5'-(β , γ -imido triphosphate)), 710–711
- GMP synthase, 1112F
- GMR α , 717–718
- GnRF, *see* Gonadotropin-releasing factor
- Goiter, 677
- GOLD (Genomes OnLine Database), 195T
- Gold, L., 213
- Goldberg, J., 434, 438, 1223
- Golden rice, 122, 122F
- Goldman equation, 775
- Goldsmith, E., 713
- Goldstein, J., 452, 453, 987
- Golgi, C., 9
- Golgi apparatus, 7F, 9, 11F. *See also* Post-translational modification
- metabolic functions, 562T
 - vesicle formation, **428–440**
 - vesicle fusion, **440–445**
- Golgi network, 428, 429F
- Golgi stack, 428, 429F
- Gonadotropin-releasing factor (GnRF), 673T, 683
- Gonadotropins, 683
- Gonads, 680, 681
- Goodman, M., 1222
- Google Scholar, 35
- Gorter, E., 395
- Gouaux, E., 417
- Gourse, R., 1266
- Gout, **1134–1135**, 1142
- GPCRs, *see* G-protein-coupled receptors
- GPI, *see* Glycosylphosphatidylinositol
- GPI-linked proteins, 407–408, 408F
- GPI membrane anchors, *see* Glycosylphosphatidylinositol membrane anchors
- G-protein-coupled receptors (GPCRs), 689, 993
- G proteins, 407, 423, 688, 692, 695F, 1314, **1376**, 1395
- G-quartets, 1212–1213, 1212F
- GR, *see* Glutathione reductase
- Gradient elution, 137
- Gram, C., 6
- Gramicidin A, 79, 748, 787
- Gram-negative bacteria, 6, 376, 376F, 403
- Gram-positive bacteria, 6, 7F, 13F, 376, 376F, 398T
- Gram stain, 6
- Grana, 902
- Granulocyte colony-stimulating factor (G-CSF), 717
- Granulocyte-macrophage colony-stimulating factor (GM-CSF), 381, 717. *See also* GM-CSF receptor
- Granulocytes, 381
- GRASP (Graphical Representation and Analysis of Surface Properties), 259
- Gray, H., 841
- Grb2, 708, 709
- Grb2-associated binder-1 (Gab-1), 738
- GRE (glucocorticoid hormone response element), 879
- GreA, 1281
- GreB, 1281, 1282
- Greek key motif, 249F, 250
- Green algae, 13F, 19T
- Greenberg, G. R., 1108

- Green fluorescent protein, *see* GFP
 Green photosynthetic bacteria, 6
 Green revolution, 122
 Green sulfur bacteria, 923
 Greider, C., 1210
 Grendel, F., 395
 GRF, *see* Growth hormone-releasing factor;
 Guanine nucleotide releasing factor
 GRIF (growth hormone release-inhibiting
 factor), 683
 Griffith, F., 86
 Griffith, J., 1213
 Grindley, N., 1240
 GRKs (GPCR kinases), 697
 gRNAs (guide RNAs), 1321, 1321F, 1324
 GroEL/ES system, 293–302, 295F–297F, 307,
 307F, 352
 Gronenborn, A., 656
 Ground substance, 371
 Group I chaperonins, 293
 Group I introns, 1306–1309, 1307T,
 1309F, 1310F
 Group II chaperonins, 293, 301–302
 Group II introns, 1307T, 1308
 Group III introns, 1307T
 Group transfer reactions, 564–565, 564F
 Group translocation, 765–766
 Growth hormone (GH), 86, 571, 683. *See also* Human growth hormone
 Growth hormone release-inhibiting factor
 (GRIF), 683
 Growth hormone-releasing factor (GRF),
 673T, 683
 Grunberg-Manago, M., 1341
 G-segment, of DNA, 1169
 GSH, *see* Glutathione
 GSH synthetase, 1061
 GSK3 (glycogen synthase kinase-3), 660
 GSLs, *see* Glycosphingolipids
 GSSG (glutathione disulfide), 803
 GSS (Gerstmann–Sträussler–Schneiker)
 syndrome, 314
 GST (glutathione-S-transferase), 142,
 1001, 1060F
 GT (UDP-glucose:glycoprotein
 glucosyltransferase), 886
 GTP (guanosine triphosphate):
 in citric acid cycle, 789, 790F, 791
 and EF-Tu, 1380–1381
 gluconeogenesis, 877–878
 hydrolysis of, 1394–1395
 in secretory pathway, 420, 423–424
 in transcription, 96, 1265
 in vesicle formation, 436, 437
 in vesicle fusion, 441
 GTPase:
 in secretory pathway, 423
 in vesicle formation, 434–437
 in vesicle fusion, 441
 GTPase-activating protein (GAP), 423, 437,
 692, 1376
 GTP-binding factor, 1395
 GTP γ S, 436F, 692
 Gua, *see* Guanine
 GU-AG introns, 1307T
 Guanidinium ion, 117, 169, 284
 Guanidino group, 580
 Guanine, 18, 83, 83T. *See also*
 Watson–Crick base pairs
 and genetic code, 100T, 1343T
 IR spectra of, 1155F
 modified nucleosides of, 1347F
 and point mutations, 1339–1340
 in purine catabolism, 1130F
 tautomeric forms, 88F
 Guanine-7-methyltransferase, 1302
 Guanine deaminase, in purine
 catabolism, 1130F
 Guanine nucleotide dissociation inhibitor
 (GDI), 692
 Guanine nucleotide exchange factors
 (GEFs), 423, 434, 692, 705–711, 1379
 Guanine nucleotide releasing factor
 (GRF), 692
 Guanosine, 83T, 1130F
 Guanosine binding site, 1310F
 Guanosine diphosphate, *see* GDP
 Guanosine monophosphate, *see* GMP
 Guanosine triphosphate, *see* GTP
 Guanylate cyclase (GC), 687
 Guanylic acid, *see* GMP
 Guide RNAs, *see* gRNAs
 Guillemin, R., 683
 D-Gulose, 360F
 GW1516, 1104
 Gymnosperms, 13F
 Gypsy moth, 122
 GyrA, 1166–1169
 GyrA intein, 1407, 1407F
 Gyrase, 1166. *See also* DNA gyrase
 GyrB, 1166–1168, 1168F
- H
H, *see* Enthalpy
 H1, 1243
Hi gene, 1243
 H2, 1243
H2 gene, 1243
H₂PO₄⁻, *see* Hydrogen phosphate ion
H₃PO₄ (phosphoric acid), 49F
 Haber, F., 1083
 Haber process, 1083
 HAD (3-L-hydroxyacyl-CoA
 dehydrogenase), 947, 947F, 948, 958
 Hadju, J., 534
 HaeII, 105T
 HaeIII, 105T, 109
 Haeckel, E., 12
Haemophilus influenzae, 176, 177T
 Hair, 233–235, 275
 Haldane, J. B. S., 32, 488
 Haldane relationship, 491–492
 Hales, S., 901
 Half cell, 584
 Half-chair conformation, 519F, 1153
 Half-life, 484
 Half reactions, 585–586
 Half-time, 484
Haloarcula marismortui, 1365, 1368
 Halobacteria, 6, 398T, 402
 Halophiles, 7F
 Hamm, H., 692
 Hammerhead ribozymes, 1310–1311, 1311F
 Hanson, R., 571, 879
 H antigen, 415
 Haploid cells, 24F
 Haploid genomes, 182
 Haploid number, 19
 Haplotype, 106
 Harden, A., 594
 Hard water, 465
 Harrison, S., 424, 431, 433, 716, 725, 1288
 Hartl, U., 300
 Hartley, B., 525
 Hasemann, C., 663
 Hatch, M., 936
 Haworth, N., 366
 Haworth projection formulas, 361, 362F
Hb, *see* Hemoglobin
HbA, *see* Hemoglobin A
- Hb Bibba, 342
 Hb Boston, 342, 343F
 Hb Bristol, 342
 Hb Cowtown, 358
 HbE, 342
 Hb Hammersmith, 342
 Hb Harlem, 344
 Hb Hyde Park, 358
 Hb Kansas, 343
 Hb KorleBu, 344–345
 HbM, 342
 Hb Memphis, 345, 358
 Hb Milwaukee, 342, 343F
 HbM Iwate, 342
 Hb Philly, 342
 Hb Rainer, 357
 Hb Riverdale-Bronx, 358
 HbS, *see* Hemoglobin S
 Hb Savannah, 342
 Hb Yakima, 343
 HCR (heme-controlled repressor), 1399
 HD gene, 1252
 HD (Huntington's disease), 1252
 HDL (high density lipoprotein), 449, 449T,
 451, 456, 458
 HDPR (6-hydroxy-1,6-dihydropurine
 ribonucleoside), 1132
 Heart, 625T, 816, 817, 1093
 Heart attack, 456, 865
 Heartburn, 764
 Heart muscle, 1092
 Heat, 53
 Heat-labile enterotoxin (LT), 696
 HEAT repeats, 722–723
 Heat shock locus UV, *see* Hs1UV
 Heat shock proteins, 293, 294. *See also*
 Molecular chaperones
 Heat shock protein 10 (Hsp10), 294
 Heat shock protein 40 (Hsp40), 293
 Heat shock protein 60 (Hsp60), 294F
 Heat shock protein 70, *see* Hsp70
 Heat shock protein 90, *see* Hsp90
 Heavy chains, clathrin, 430
 HECT domain, 1410
 Heinrich, R., 620
 Heinz bodies, 342
 Helical structures, 225–229, 225F
 collagen triple helical structure,
 236–238, 237F
 Helical symmetry, proteins, 269, 269F
 Helicases, 1163, 1176, 1183–1187
 Helicase II, 1217
 Heliozoans, 13F
 Helix:
 α , 226–227
 basic helix-loop-helix/leucine
 zipper, 987
 C, 242F
 DNA as, 88–90, 88F
 E, 242F
 4-helix bundle, 241, 250F
 H, 248F
 interwound, 1160, 1160F
 left-handed, 224F, 225F, 1149–1151
 π , 224F, 228
 and pitch of protein, 225
 right-handed, 224F–226F
 superhelix, 433, 1158–1160, 1162
 β_{10} , 224F, 226, 228
 β_{13} (α), 226, 227F
 toroidal, 1160, 1160F
 Helix-capping position, 262
 Helix-turn-helix (HTF) motif, 1240,
 1288–1290, 1289F, 1290F
 Helper T cells, 545–546

- Hemagglutinin, 380
 β-Hematin, 1058
 Hematocrit, 324
 Heme *a*, 839F
 Heme *b*, 839F. *See also* Iron-protoporphyrin IX
 Heme *b*₅₆₂ (heme *b*_H), 838
 Heme *b*₅₆₆ (heme *b*_L), 838
 Heme *c*, 839F
 Heme *c*_i (heme *x*), 920
 Heme-controlled repressor (HCR), 1399
 Heme group, 324–325, 574, 574F
 biosynthesis, 1047–1058
 degradation, 1056
 Heme oxygenase, 1056
 Heme-regulated inhibitor (HRI), 1399, 1399F
 Hemerythrin, 325
 Heme *x* (heme *c*_i), 920
 Hemiacetals, 361, 520F
 Hemifusion, 442, 443F
 Hemiketals, 361
 Hemin, 1055
 Hemmings, E., 116
 Hemocyanin, 153T, 325
 Hemoglobin:
 abnormal, 341–347
 affinity for CO, 1056–1057
 allosteric regulation, 347–354
 α and β chains, 332T–333T
 and Bohr effect, 328–329, 328F, 329, 340
 BPG and oxygen binding of, 329–331, 329F–331F, 340–341
 carbon dioxide transport, 329
 conformational fluctuations, 306, 306F
 cooperativity, 352–354
 and distal histidine residue, 340
 and erythrocyte membrane, 412
 function, 323–331
 gene duplication, 193–194
 heme group, 324–325
 Hill plot, 327F
 as “honorary enzyme,” 323
 hydrogen exchange for, 309F
 isoelectric point (human), 134T
 isolation, 129
 lamprey, 135F
 molecular pathology, 342–343
 optimal sequence alignment for, 197F
 oxygen binding, 325–328, 334–339
 Perutz mechanism, 334–340
 rate of evolution, 191, 191F
 and sickle-cell anemia, 81, 185–188, 343–347
 solubility of, 133F
 in Southern blotting, 112
 structure and mechanism, 267F, 332–341
 Hemoglobin A (HbA), 175F, 186
 Hemoglobin E, 342
 Hemoglobin F, *see* Fetal hemoglobin
 Hemoglobinopathies, 342–343
 Hemoglobin S (HbS), 175F, 186–188, 277, 343–347
 α-Hemolysin, 416–418, 418F
 Hemolytic anemia, 185, 342, 343
 Hemophilia, 119
 Hemozoin, 1058
 Henderson, R., 402
 Henderson–Hasselbalch equation, 47–48, 72
 Hendrickson, W., 702
 Hen egg white (HEW) lysozyme, 275–276, 276T, 470, 517–525, 517F, 518F, 524F
 Henri, V., 488
 Henry's law, 61
 Henseleit, K., 791, 1025
 Heparan sulfate, 373
 Heparin, 371F, 372–373, 372F, 375, 700
 Hepatitis B vaccine, 119
 Hepoxilins, 1000
 Heptahelical receptors, 689
 Heptoses, 360F
 HER2 receptor, 119–120
 Herbivores, cellulose digestion by, 369
 Herceptin, 119–120
 Hereditary elliptocytosis, 414
 Hereditary nonpolyposis colorectal cancer (HNPPCC), 1220
 Hereditary ovalcytosis, 414
 Hereditary spherocytosis, 414
 Hermit crabs, chromosome number, 19T
 Herriott, R., 86
 Hers, H.-G., 663
 Hers' disease, 667
 Hershey, A., 87
 Hershey–Chase experiment, 87F
 Hershko, A., 1410
 (15R)-HETE ([15R]-hydroxyeicosatetraenoic acid), 1004
 (15S)-HETE ([15S]-hydroxyeicosatetraenoic acid), 1003
 Heterocysts, 1083
 Heterogeneous nuclear ribonucleoprotein (hnRNP), 1314
 Heterogeneous nuclear RNA (hnRNA), 1304
 Heterogeneous nucleation, 346, 347F
 Heterokaryon, 408
 Heterolytic bond cleavage, 563, 956
 Heteropolysaccharides, 365–366
 Heterotrimeric G proteins, 688–689
 adenylate cyclases, 697–698
 G protein-coupled receptors (GPCRs), 689–690
 phosphodiesterases, 688–689
 structure and function, 690–697
 Heterotrophs, 6
 Heterotropic allosteric effect, 348, 351, 351F, 354
 Heterozygous genotypes, 21–22
 Heuristic algorithms, 201
 Heuser, J., 444
 HEW, *see* Hen egg white lysozyme
 1-O-Hexadec-1'-enyl-2-acetyl-sn-glycero-3-phosphocholine, 1008
 Hexane, 43, 43T, 144
 Hexokinase (HK), 565
 in gluconeogenesis, 872
 in glycolysis, 596F, 597–598
 mannose utilization, 633
 Hexokinase D, 662
 Hexokinase IV, 662
 Hexosaminidase A, 123, 1011, 1012F
 Hexosaminidase B, 1013
 Hexoses, 360F
 Hexose monophosphate (HMP) shunt, 892.
See also Pentose phosphate pathway
 Hfr cells (high frequency of recombination), 1263
 hGH, *see* Human growth hormone
 hGHbp (human growth hormone binding protein), 684
 HGPRT (hypoxanthine-guanine phosphoribosyltransferase), 1114
 H Helix, 248F
 HIC, *see* Hydrophobic interaction chromatography
 Hidden Markov models (HMMs), 202
 High-altitude adaptation, 330–331, 330F, 357
 High density lipoproteins, *see* HDL
 High-energy compounds, 580–581. *See also* specific compounds
 High frequency of recombination (Hfr) cells, 1263
 High-mannose oligosaccharides, 888
 High-performance liquid chromatography, *see* HPLC
 High potential iron-sulfur proteins, *see* HIPPs
 High-throughput screening, 541
 HI/HA, *see* Hyperinsulinism/hyperammonemia
 Hihydrosphingosine, 390
 Hill, A., 326
 Hill, C., 1419
 Hill, R., 903
 Hill constant, 327
 Hill equation, 326–328, 348
 Hill plot, 327–328, 327F
 Hill reaction, 903
 HindIII, 105T
 Hin DNA invertase, 1243
 hin gene, 1243
 Hinny, 1251
 HIPPs (high potential iron–sulfur proteins), 202F, 281F
 Hippuric acid, 945
 His, *see* Histidine
 his operon, 1299, 1299T
 His tag, 145, 857
 Histamine, 80, 80F, 783, 1058, 1059
 Histamine receptors, 689
 Histidine:
 biosynthesis, 1078
 degradation, 1034
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 in globular proteins, 246
 α helix/β sheet propensity, 302T
 and hemoglobin, 332, 341
 isoelectric point, 72
 side chains, 71, 264T
 structure and general properties, 69T
 Histidinodihydroxymerodesmosine, 239F
 Histones, 78, 134T
 Histone H4, 190–191
 Histrionicatoxin, 781
 Hitchings, G., 1130
 HIV (human immunodeficiency virus), 208–209, 1207
 HIV-1, 545–546, 545F
 polyproteins of, 546, 546F
 protease of, 208–209, 546, 548–551, 548F
 reverse transcriptase of, 1207–1209, 1209F
 HIV-2, 545
 HIV protease:
 of HIV-1, 208–209, 546, 548–551, 548F
 inhibitors of, 545–551
 and polyproteins, 1404
 HK, *see* Hexokinase
 (H⁺–K⁺)-ATPase, 764–765
 HMG-CoA, 960, 976–977, 1040–1041
 HMG-CoA lyase, 960
 HMG-CoA reductase, 977, 987–989
 HMG-CoA reductase kinase (RK), 989
 HMG-CoA synthase, 960, 976
 HMIT, 751
 HMMs (hidden Markov models), 202
 HMP (hexose monophosphate) shunt, 892.
See also Pentose phosphate pathway
 HNPCC (hereditary nonpolyposis colorectal cancer), 1220
 hnRNA (heterogeneous nuclear RNA), 1304
 hnRNP (heterogeneous nuclear ribonucleoprotein), 1314

- HO₂• 865
 Hoagland, M., 1345
 Hodgkin, A., 776
 Hodgkin, D. C., 241, 331
 Hofmeister, F., 70
 Hofmeister series, 266
 Hofmeyr, J.-H., 623
 Hofrichter, J., 346
 Hol, W., 696, 792, 1165
holA gene, 1182T
holB gene, 1182T
holC gene, 1182T
holD gene, 1182T
holE gene, 1182T
 Holley, R., 176, 1345
 Holliday, R., 1226
 Holliday junction, 1226, 1226F, 1233F
 in recombination repair, 1234–1236
 and RuvABC, 1230–1234
 Holoenzymes, 473, 722, 1265. *See also* Pol III
 holoenzyme; RNAP holoenzyme
 Homeostasis, **619–620**, 1096–1101
 Homing endonuclease, 1407–1408
 Hominy grits, 481
 Homocitrate, 1081–1082
 Homocysteine, 80, 80F, 1034
 Homocysteine methyltransferase, 1037
 Homogeneous nucleation, 346, 347F
 Homogenizer, 130
 Homogentisate, 1045, 1047F
 Homogentisate dioxygenase, 1045
 Homogentisic acid, 25, 569
 Homolactic fermentation, 593, **614–616**
 Homologous end-joining, 1235–1236, 1235F
 Homologous pairs, 19
 Homologous proteins, **188–192, 194–196**
 Homologous recombination, 1222,
1225–1236, 1227F
 Homology modeling, 304–305
 Homolytic bond cleavage, 563, 956
Homo neanderthalensis, 116
 Homopolyptides, 205
 Homopolysaccharides, 365–366
Homo sapiens, 14, 116. *See also* Humans
 Homoserine, 759
 Homotropic allosteric effect, 348,
 350–351, 354
 Homozygous genotypes, 21–22
 Honig, B., 259
 Hood, L., 181
 Hoogsteen geometry, 1154
 Hooke, R., 3
 Hop diffusion, 411, 411F
 Hopene, 983
 Horecker, B., 892
 Hormones, **671–674**, 673T. *See also*
 Glucagon; Insulin
 assays for, 131
 in brown fat, 860–862, 861F
 for calcium metabolism, **677–678**
 classification, 671, 672F
 epinephrine and norepinephrine,
679–680
 for fatty acid metabolism, 973–975, 974F
 gastrointestinal, **675–676**
 growth, **684–685**
 hypothalamus and pituitary, **682–683**,
 1099, 1100F
 menstrual cycle control, **683–684**
 nitric oxide, **685–688**
 opioid peptides, **685**
 pancreatic islet, **675**
 quantitative measurements, **674–675**
 steroid, **680–682**
 thyroid, **676–677**
- Hormone-sensitive triacylglycerol lipase,
 861, 944, 973
 Horn, 233, 234
 Horowitz, N., 3, 34
 β-Horseshoe, GyrA intein, 1407
 Horwitz, A., 294, 297, 298
 Host organism, 104
 Host-specific modifications, 104
 Hough, E., 1044
 Housekeeping genes, 1282
 HpaII, 105T, 1249
 (15S)-HPETE ([15S]-hydroperoxyeicosatetraenoic acid), 1003
 5-HPETE, *see* 5-
 Hydroperoxyeicosatetraenoic acid
 12-HPETE, *see* 12-
 Hydroperoxyeicosatetraenoic acid
 15-HPETE, *see* 15-
 Hydroperoxyeicosatetraenoic acid
 HPETEs (hydroperoxyeicosatetraenoic acids), 1000
 HPK1, 714
 HPLC (high-performance liquid chromatography), 145–146, 157, 171, 366
 HPr, of PTS system, 765
 H-protein, 1031
 HPV (human papillomavirus), 1414
 H-Ras, 709
 HRI, *see* Heme-regulated inhibitor
 Hsc70, 436
 HslU caps, 1419
 HslUV (heat shock locus UV), 1419, 1420F
 HslUV protease, 1419
 HslV subunit, 1419–1420, 1420F
 Hsp10 proteins, 294
 Hsp40 proteins, 293
 Hsp60 proteins, 294F
 Hsp70 proteins, 293, 421F, 445
 Hsp90 proteins, 293, 720
 HTH, *see* Helix–turn–helix motif
 HU, 1195
 Hubbard, S., 700
 Huber, R., 403, 653, 806, 910, 1414, 1415,
 1419, 1420
 Hugging dimer, 1050
 Human cells, 12F
 Human cytomegalovirus protease, 531
 Human genome, 176, 177T, 1402
 Human genome project, 181–182
 Human Genome Sequencing Consortium,
 181, 182
 Human growth hormone (hGH), 119, 250F,
 251, 259F, **684–685**
 Human growth hormone binding protein
 (hGHbp), 684
 Human growth hormone receptor, **684–685**
 Human immunodeficiency virus, *see* HIV
 Human papillomavirus (HPV), 1414
 Humans, 16F. *See also* *Homo sapiens*
 apoB, 1322
 BBP, 1314F
 biological structures, 14
 c-Cbl-UbcH7-ZAP-70 complex, 1412F
 chromosome number, 19T
 cloning, 1251
 Cul1-Rbx1-Skp1-F-box^{Skp2} complex, 1412F
 cytochrome c, 196F
 DNA repair, 1214
 DNA size, 94T
 E. coli DNA vs., 8–9
 elemental composition, 29T
 embryonic development, 14F
 evolutionary divergence of, 116
 fuel reserves for, 1101T
 genomic libraries, 113
 hormones, 673T
 Pot1, 1213
 RFLPs of, 106
 Skp1-Skp2 complex, 1412
 telomere length, 1211
 transposons, 1244, 1246
 ubiquitin, 1409
 UDG, 1219F
 Hünefeld, F., 323
 Hunger, 1098, 1099
 Huntington, 1252
 Huntington's disease (Huntington's chorea), 1252
 Hurley, T., 646
 Hurwitz, J., 1265
 Huxley, A., 776
 Hwang, P., 431
 Hyaluronan, **371–372**
 Hyaluronate, 371F, 372, 372F
 Hyaluronic acid, **371–372**, 375
 Hyaluronidase, 372
 Hybridization, 93
 Hybrid oligosaccharides, 888
 Hybridoma, 131
 Hybrid proteins, 118
 Hyde, C., 1076
 Hydrated ions, 43
 Hydrazine, 166, 168
 Hydrazinolysis, 166, 168
 Hydrochloric acid, 46
 Hydrocortisone, 680
 Hydrodynamic volume, 155
 Hydrogen, in Miller-Urey experiments, 32
 Hydrogen bonding:
 in base pairs, 84, 89, 1156
 in collagen triple helical structure, 237F
 and enzyme substrate specificity, 470F
 in functional groups, 43F
 in α helix, 227F
 in HIV protease-1, 549F
 low-barrier vs. short, strong, **535–536**
 in β pleated sheet, 229–230, 229F
 in proteins, **261–262**
 and tRNA tertiary structure, 1348–1349
 in water, 41–42, 41F
 Hydrogen exchange, **285–286**, 309F
 Hydrogen ions, 60
 Hydrogen peroxide, 10
 Hydrogen phosphate ion (H₂PO₄),
 47, 47F
 Hydrogen sulfide, 324
 Hydrolases, 479T
 Hydrolysis, base-catalyzed, 85, 85F
 Hydronium ion, 45, 45F, 46
 Hydropathies, 264, 264T, 304
 5-Hydroperoxyeicosatetraenoic acid
 (5-HPETE), 996F, 997, 1000
 12-Hydroperoxyeicosatetraenoic acid
 (12-HPETE), 996F, 997, 1000
 15-Hydroperoxyeicosatetraenoic acid
 (15-HPETE), 996F, 997, 1000
 Hydroperoxyeicosatetraenoic acids
 (HPETEs), 1000
 (15S)-Hydroperoxyeicosatetraenoic acid([15S]-HPETE), 1003
 Hydrophilic materials, 42
 Hydrophobic bonding, 264
 Hydrophobic collapse, 286–287
 Hydrophobic effect, **262–264**
 Hydrophobic forces, 44
 and enzyme substrate specificity, 470F
 in integral membrane proteins, 406
 in nucleic acid bases, 1156–1157
 and protein stability, **262–264**
 and reverse turns, 304

Hydrophobic interaction chromatography (HIC), 132T, 145
 Hydrophobic interactions, 44
 Hydrophobic materials, 42
 Hydropyrimidine hydratase, 1136F
 Hydroxide ions, 45, 45T, 46, 688
 6-Hydroxy-1,6-dihydropurine ribonucleoside (HDPR), 1132
 8-Hydroxy-7,8-dimethyl-5-deazariboflavin, 1215
 β -Hydroxyacyl-ACP dehydrase (DH), 965
 3-L-Hydroxyacyl-CoA, 947, 947F
 3-L-Hydroxyacyl-CoA dehydrogenase, *see* HAD
 3-Hydroxyanthranilate, 1042
 Hydroxyapatite, 540, 677
 Hydroxyapatite chromatography, 144, 157
 β -Hydroxybutyrate, 831
 D- β -Hydroxybutyrate, 959
 D- β -Hydroxybutyrate dehydrogenase, 960
 α -Hydroxybutyric acid, 32T
 25-Hydroxycholecalciferol, 678
 (15R)-Hydroxyeicosatetraenoic acid ([15R]-HETE), 1004
 (15S)-Hydroxyeicosatetraenoic acid ([15S]-HETE), 1003
 Hydroxyethylthiamine pyrophosphate, 616–617
 3-Hydroxynurenine, 1041–1042
 Hydroxylamine, 312–313
 point mutations caused by, 312, 313F
 prion reactivation with, 312–313, 313F
 Hydroxyl groups, 43F
 Hydroxyl radical, 325, 865
 5-Hydroxylysine (Hyl), 78, 79F, 236
 Hydroxymethylbilane, 1053
 Hydroxymethylbilane synthase, 1053
 5-Hydroxymethylcytosine residue, 1247
 Hydroxymethylglutaryl-CoA, *see* HMG-CoA
 p-Hydroxyphenylpyruvate, 1045, 1047F
 p-Hydroxyphenylpyruvate dioxygenase, 1045
 4-Hydroxyproline (Hyp), 78, 79F, 236
 3-Hydroxyproline, 236
 Hydroxypyruvate, 935
 8-Hydroxyquione, 1119
 5-Hydroxytryptophan, 1046, 1060
 Hydroxyurea, 347, 1119
 Hyl (5-hydroxylysine), 236
 Hyp (4-hydroxyproline), 236
 Hyperacute rejection, 121
 Hyperammonemia, 1023, 1025
 Hypercholesterolemia, 456–457, 466, 989–991
 Hyperchromic effect, 92
 Hyperglycemia, 1102, 1103
 Hyperhomocysteinemia, 1034–1035
 Hyperinsulinism/hyperammonemia (HI/HA), 1023, 1025
 Hyperlysinemia, 1041
 Hyperlysinuria, 1041
 Hyperphenylalaninemia, 1045
 Hypertension, 547F
 Hyperthermophiles, 266
 Hyperthyroidism, 677
 Hypervariable position, 188
 Hyperventilation, 357
 Hypoglycemia, 666
 Hypoglycin A, 948, 948F
 Hypothalamus, 672F, 682
 hunger signal transmission, 1099
 integration of signals, 1099, 1100F
 leptin expression, 1098
 Hypothyroidism, 677

Hypotonic solutions, 130
 Hypoxanthine, 127, 1108, 1130F, 1135, 1339–1340
 Hypoxanthine–guanine phosphoribosyltransferase (HGPRT), 1114
 Hypoxia, 331, 864
 Hypoxic, 324
 I
 i^6A , *see* N⁶-Isopentenyladenosine
 IAPP (islet amyloid polypeptide), 310–311, 311F
 I κ B α , 1414
 Ibandronate, 679F
 Ibuprofen, 78, 78F
 IC₅₀, 539
 ICATs, *see* Isotope-coded affinity tags
 Ice, 41–42, 41F
 I-cell disease, 439
 Icosahedral symmetry, 268, 268F
 IDL (intermediate density lipoproteins), 449, 449T, 986, 1322
 D-Idose, 360F
 IEF (isoelectric focusing), 150–152
 IF, *see* Initiation factor
 IF-1, 1375–1376, 1375T
 IF₁ protein, 865
 IF-2, 1375–1376, 1375T
 IF-3, 1375–1376, 1375T
 I-FABP (intestinal fatty acid-binding protein), 943–944, 944F
 I gene, 1262
 IgG, *see* Immunoglobulin G
 IHF (integration host factor), 1195
 IHP (inositol hexaphosphate), 330
 I κ B kinase (IKK), 1420
 IL-1, *see* Interleukin-1
 IL-3 receptors, 717
 IL-5 receptors, 717
 Ile, *see* Isoleucine
 IleRS, 1352, 1356–1357, 1357F
 Illegitimate recombination, 1236
 Illumina system, 184
 ilv operon, 1299, 1299T
 Image reconstruction, 1363F
 Imatinib, 719
 Imidazole, 513
 β , γ -Imido nucleoside triphosphate, 1275
 Iminoaceticpropionic acid, 32T
 Iminodiacetic acid, 32T
 Immobilized pH gradients, 151
 Immune electron microscopy, 1365
 Immune system, 1132
 Immunoaffinity chromatography, 143
 Immunoblotting, 149–150, 150F
 Immunochemical procedures, 131–132
 Immunofluorescence, 408
 Immunoglobulins, 163
 Immunoglobulin fold, 251, 251F
 Immunoglobulin G (IgG), 154F, 282, 380
 Immunophilins, 292, 724
 IMP (inosine monophosphate), 1108
 animal catabolism pathway, 1130F
 conversion to AMP or GMP, 1112F
 synthesis, 1108–1111, 1109F
 IMP cyclohydrolase (PurJ), 1109F, 1111
 IMP dehydrogenase, 1112F
 IMS, *see* Intermembrane space
 Inactivation, of ion channels, 774–775
 Inactivators, 494
 Inborn errors of metabolism, 26
 Inclusion bodies, 117, 117F
 Indels, *see* Insertion/deletion mutations
 Independent assortment, 21–22, 22F

Independent segregation, 23F
 Indinavir, 550F, 551
 Indirect readout, 1290
 Indole-3-glycerol phosphate, 1075
 Indolepropanol phosphate (IPP), 1078
 Induced-fit hypothesis, 352
 Induced-fit model, of allosteric interactions, 352
 Inducers, 97, 97F, 1261
 Inducible enzymes, 1261
 Inducible NOS (iNOS), 686–688
 -ine suffix, 73
 Infarction, 456
 Inflammatory response, 993, 1003–1004
 Influenza virus neuraminidase, 466
 Ingen-Housz, J., 901
 Ingold, C., 76
 Ingram, V., 186
 Inheritance, 20–26
 Inherited diseases, *see* Genetic diseases
 Inhibition, of enzymes, *see* Enzyme inhibition
 Inhibitors, 351, 492
 Inhibitory synapses, 778
 Initial velocity, 489
 Initiating codon, 100–101, 100T
 48S Initiation complex, 1379
 Initiation factor (IF), 1375–1376, 1375T
 Inorganic pyrophosphatase, 582, 946
 iNOS (inducible NOS), 686–688
 Inosine, 1130F, 1347F
 Inosine monophosphate, *see* IMP
 Inosine triphosphate, *see* ITP
 myo-Inositol, 364, 389T
 Inositol-1,4,5-triphosphate, *see* IP₃
 Inositol hexaphosphate (IHP), 330
 Inositol polyphosphate 1-phosphatase, 736
 Inositol polyphosphate 3-phosphatases, 736
 Inositol polyphosphate phosphatases, 734–736
 Insects, mutation rate of, 192
 Insertion/deletion mutations (indels), 196, 1340
 Insertion sequences (IS), 1236
In situ gene therapy, 122–123
In situ hybridization, 113–114, 114F
 InsR, *see* Insulin receptor
 Insulin, 164–165, 165F, 673T, 1099
 affinity chromatographic isolation, 142
 in fatty acid metabolism regulation, 974
 in glycogen metabolism regulation, 660
 glycogen synthase kinase-3 inhibition, 660
 isoelectric point, 134T
 and passive-mediated glucose transport, 751–752
 pyruvate dehydrogenase phosphatase activation, 805
 recombinant human, 119
 renaturation, 280–281
 sequencing of bovine, 165F
 Insulin-dependent diabetes, 1102, 1400
 Insulinlike growth factor II receptor, 890
 Insulin receptor (InsR), 142, 699–703, 737F
 Insulin receptor substrate (IRS), 710, 1103
 Insulin resistance, 1103, 1103F
 Insulin signaling system, 736–738
 Insulin-stimulated protein kinase, 659–660
 Integral membrane proteins, 399–406
 Integrase, 545, 1245
 Integration host factor (IHF), 1195
 Intein Database, 1407
 Inteins, 1405–1407, 1406F
 Intensive quantities, 61
 Intercalating agents, 1160–1162, 1272
 Intercalation, 157
 Interfacial activation, 941
 Interferons, 1399–1400, 1400F

- Intergenic suppressors, 1362
 Interleukins, 657, 671, 717
 Interleukin-1 (IL-1), 671, 1414
 Interleukin-1 β - converting enzyme, *see* ICE
 Interleukin-2, 671
 Intermediates, reaction, 483
 Intermediate density lipoproteins, *see* IDL
 Intermediate fibers, 10–11
 Intermediate filaments, 10–11
 Intermembrane space (IMS), 9, 446, 448
 Internal conversion, 904
 Internally compensated molecules, 76
 Internal nodes, phylogenetic trees, 203–204, 203F
 Internal resolution site, 1237
 Internal ribosome entry site (IRES), 1379
 Interphase, 20F, 21F
 Intervening sequences, 97, 1304
 Interwound helix, 1160, 1160F
 Intestinal fatty acid-binding protein, *see* I-FABP
 Intragenic suppressors, 1340
 Intrasteric mechanism, 658
 Intrinsic dissociation constants, 348
 Intrinsic factor, 956
 Intrinsic membrane proteins, *see* Integral membrane proteins
 Intrinsic proteins, 399–406
 Intrinsic terminators, 1273–1275, 1273F
 Intronic splicing enhancers (ISEs), 1306
 Intronic splicing silencers (ISSs), 1306
 Introns, 98, 1304–1305
 AU-AC, 1319
 Group I, 1306–1309, 1307T
 Group II, 1307T, 1308
 in prokaryotes vs. eukaryotes, 1316
 types of, 1307T
 Invariant residues, 188
 Invertase, 367, 488
 Inverted micelles, 51
 Invert sugar, 367
In vivo gene therapy, 123
 Iodoacetic acid, 168–169
 Ions, solvation of, 43F
 Ion-driven active transport, 768–771
 Ion exchange, 135
 Ion exchange chromatography, 132T, 135–138, 136F, 157
 Ion exchangers, 137–138, 138T
 Ionic charge, protein purification and, 132
 Ionic interactions, 259–260, 1158
 Ionic mobility, 44–45, 45T
 Ionic strength, 133–134, 133F
 Ionization constants, 47, 49–50, 497, 502–503
 Ionizing radiation, 1214
 Ionophores, 748–749, 748F
 Ion pair, 259–260, 260F
 IP₃ (inositol-1,4,5-triphosphate), 661, 665, 725–726
 IP₃ receptor, 726
 IPP (indolepropanol phosphate), 1078
 IPTG (isopropylthiogalactoside), 117, 621, 1261, 1295
 IRES (internal ribosome entry site), 1379
 Iron, in heme biosynthesis, 1056
 Iron-protoporphyrin IX, 316, 316F, 324, 838, 904F. *See also* Heme b
 Iron-sulfur clusters, 808F, 809, 834–835
 Iron-sulfur protein (ISP), 840
 Irreversible thermodynamics, 587
 IRS, *see* Insulin receptor substrate
 IS (insertion sequences), 1236
 IS1, 1236, 1236T
 IS2, 1236, 1236T
 IS4, 1236T
 IS5, 1236T
 Ischemia, 325, 864
 IS elements, 1236
 ISEs (intronic splicing enhancers), 1306
 Isle of Langerhans, 675
 Islet amyloid polypeptide, *see* IAPP
 Isoaccepting tRNAs, 1351
 Isoalloxazine, 566
 Isocaudamers, 128
 Isocitrate, 789, 790F, 809
 Isocitrate dehydrogenase, 568, 789, 809–810, 810F, 815–816
 Isocitrate lyase, 880
 Isoelectric focusing (IEF), 132T, 150–152
 Isoelectric point (pI), 72, 134, 134T, 150
 Isoelectric precipitation, 134
 Isoflurane, 398–399
 Isoforms, 543
 Isoionic point, 81
 Isolated systems, 52
 Isolation, protein, 129–133
 Isoleucine:
 biosynthesis, 1075
 chiral centers, 76, 77F
 degradation, 1034–1039
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 in globular proteins, 246
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 in native unfolded proteins, 283
 side chains, 70, 264T
 stereoisomers, 77F
 structure and general properties, 68T
 (2S,3S)-Isoleucine, 76, 77F
 Isomaltose, 367, 367F
 Isomerasases, 479T
 Isomerization, 567
 N⁶-Isopentenyladenosine (i⁶A), 976, 1347F
 Isopentenyl pyrophosphate, 977, 977F, 978F
 Isopentenyl pyrophosphate isomerase, 978, 978F
 Isopeptide bond, 1409
 Isoprene, 406, 975
 Isoprenoids, 976–977, 976F
 Isoprenoid groups, 406
 Isoprenylation site, 406
 α -Isopropylmalate, 1075
 Isopropylthiogalactoside, *see* IPTG
 Isoproterenol, 680
 Isopycnic ultracentrifugation, 156, 157F
 Isoschizomers, 127–128
 Isotonic saline solution, 161
 Isotopes, 573T
 Isotope-coded affinity tags (ICATs), 577–578, 577F
 Isotope effects, 572
 Isotope studies, 500–501, 572–575
 Isotope tests, 813–814
 Isozymes, 202F, 277, 543
 ISP (iron-sulfur protein), 840
 ISSs (intronic splicing silencers), 1306
 Iterative annealing model (GroEL/ES system), 299–300
 ITP (inosine triphosphate), 1273
 Iwata, S., 838, 916
 J
 Jacob, F., 1262, 1264–1265
 Jaenisch, R., 1251
 Jaffe, E., 1050
 Jahn, R., 441, 444
 JAK (Janus kinase), 718
 JAK1, 718
 JAK2, 718
 JAK3, 718
 JAK4, 718
 JAK–STAT pathway, 717–718
 Jamaican vomiting sickness, 948
 James, M., 523
 Jansonius, J., 403
 Janus, 283
 Janus kinase (JAK), 718
 Jap, B., 840
 Jaundice, 1056
 Jefferson, T., 116
 Jelly roll, 252, 253F
 Jencks, W., 515
 JIP-1, *see* JNK interacting protein-1
 Jmol, 256T, 257
 JNK, 714
 JNK interacting protein-1 (JIP-1), 714, 743
 Johnson, A., 424
 Johnson, L., 640, 658
 Joliet, P., 917
 Jones, M. E., 1117
 Jorgensen, R., 1323
 Joshua-Tor, L., 1184
 Joule (unit), 53, 53T
 Jpred3 algorithm, 304
 Jun, 705, 713, 737F
 Junk DNA, 1317
 Juvenile-onset diabetes, 310, 1102
- K
 K_M (Michaelis constant), 489
 K_a, *see* Dissociation constant
 Kaback, R., 770
 Kacser, H., 620
 Kahn, R., 1104
 Kaiser, D., 109
 Kaiser, T., 450
 Kalckar, H., 578
 Kallikrein, 525T
 Kamen, M., 813, 903
 Kaplan, N., 791
 Kaptein, R., 1295
 Karplus, M., 307, 857
 Kauzmann, W., 264
 kb (kilobase pairs), 95
 k_B, *see* Boltzman constant
 K⁺ channels, 752–755, 753F, 754F
 KDEL receptors, 440
 KDO (2-keto-3-deoxyoctanoate), 379F
 Ke, H., 725
 Keilin, D., 838
 Kelvin temperature scale, 53T
 Kendrew, J., 246, 331
 Kennedy, E., 419, 823, 946
 Kent, S., 209
 K_{eq}, *see* Equilibrium constant
 Keratan sulfate, 371F, 372, 373
 Keratin, 10–11, 10F, 233–235
 α Keratins, 233–235, 234F
 β Keratins, 233–235
 2-Keto-3-deoxy-6-phosphogluconate, 637F
 2-Keto-3-deoxy-D-arabinoheptulosonate-7-phosphate, *see* DAHP
 2-Keto-3-deoxyoctanoate (KDO), 379F
 2-Keto acid dehydrogenases, 799
 Ketoacidosis, 961, 1102
 β -Ketoacyl-ACP reductase (KR), 965
 β -Ketoacyl-ACP synthase (KS), 965
 β -Ketoacyl-CoA, 947F, 949F
 β -Ketoacyl-CoA thiolase, *see* KT
 3-Ketoacyl-CoA transferase, 961
 α -Ketoadipate, 1042
 α -Ketobutyrate, 1034, 1075
 Keto-enol tautomerization, 506–507, 507F

- Keto forms, of nucleotide bases, 88
 Ketogenic amino acids, 1029–1030, 1090, 1094
 α -Ketoglutarate, 789, 809, 1034, 1067F
 α -Ketoglutarate dehydrogenase (E1o), 789, 799, 810, 815–816
 α -Ketoglutarate dehydrogenase complex, 799
 Keto groups, hydrogen bonding in, 43F
 α -Ketoisovalerate, 1075
 α -Ketoisovalerate dehydrogenase, 1039
 Ketones, 361F
 Ketone bodies, 959–961, 973, 1095, 1101
 Ketonogenesis, 959–961
 Ketopentose, 360
 Ketoses, 360F, 361, 361F
 Ketosis, 961
 3-Ketosphinganine, 1008
 3-Ketosphinganine reductase, 1009
 3-Ketosphinganine synthase, 1008
 KF (Klenow fragment), 177, 1177–1178, 1178F, 1180–1181
 KFERQ proteins, 1409
 Khorana, H. G., 209, 211, 1342
 Kidneys, 871, 1091F, 1095
 Kidney stones, 1134
 Kieselguhr, 144
 Kilobase pairs (kb), 95
 Kim, J.-J., 948
 Kim, K. K., 1151
 Kim, P., 282
 Kim, S., 1346
 Kim, S.-H., 840
 Kinases, 513, 581, 597
 Kinematics, 257
 Kinesin, 440
 Kinetics:
 chemical, 54, 482–486
 enzyme, 487–492
 membrane transport, 745–757
 Kinetic barrier, 486
 KiNG, 256T, 257
 King, J., 289
 Kinoshita, K., Jr., 857, 1271
 Kirchhausen, T., 431
 Kjeldgaard, M., 1381, 1388
 Kleinschmidt procedure, 94F
 Klenow fragment, *see* KF
 Klentaq1, 1178–1180, 1223
 Klinefelter's syndrome, 682
 Klug, A., 1346, 1363
 KNF model, of allosteric interactions, 352
 Knockout mice, 121, 572
 Knoop, F., 572, 791, 945
 Knowles, J., 565
 Kohda, D., 446
 Kok, B., 917
 Kool, E., 1177
 Kornberg, A., 1116, 1176, 1194, 1195, 1281
 Kornberg, R., 1277–1278
 Kornfeld, S., 884, 888
 Koshland, D., 352, 513, 514, 621, 817
 Kossiakoff, A., 684
 KR (β -ketoacyl-ACP reductase), 965
 Krabbe's disease, 1013T
 K-Ras 4A, 709
 K-Ras 4B, 709
 Krauss, N., 922
 Kraut, J., 534
 Krebs, E., 722
 Krebs, H., 651, 791, 1025
 Krebs cycle, 789. *See also* Citric acid cycle
 Kreevoy, M., 536
 KS (β -ketoacyl-ACP synthase), 965
 KT (β -ketoacyl-CoA thiolase), 947, 947F, 948, 949F
 Ku, 1223–1224, 1223F
 Ku70, 1223
 Ku80, 1223
 Kuhn, H., 1001
 Kühne, F. W., 470
 Kunitz, M., 470
 Kunkel, T., 1205
 Kurisu, G., 924
 Kuriyan, J., 706, 710, 720, 1125, 1182, 1196, 1197, 1204
 Kuru, 312
 Kushmerick, M., 863
 Kv1.2 channels, 773, 773F, 774F
 Kv channels, 772–774
 K_w (ionization constant of water), 47
 Kwashiorkor, 1087
 Kynureninase, 1042
 L
lacI gene, 107F, 621, 1264
 S -lac-NAD⁺, 614, 615F
lac operators, 1284–1285, 1285F, 1294
lac operon, 97, 97F, 1261–1262, 1294
lac promoter, 117, 621, 1266, 1294
lac repressor, 97, 107F, 117, 621, 1263–1264
 allosterically important residues, 1296
 isolation of, 129
 in prokaryote transcription control, 1284–1285
 structure, 1294–1296, 1296F
 α -Lactalbumin, 882
 β -Lactamase, 1237
 Lactase, 367
 Lactate:
 converting to oxaloacetate, 872F
 in Cori cycle, 880
 from fermentation, 594, 594F, 614–619
 transport in blood, 973
 Lactate dehydrogenase, *see* LDH
 Lactate dehydrogenase H, 153T
 Lactic acid, 32T
 D-Lactic acid, 365
 β -Lactoglobulin, 133F, 134
 Lactoperoxidase, 140F
 Lactose, 367, 367F, 631
 and glucose, 1286
 metabolism in *E. coli*, 97
 synthesis, 882
 Lactose intolerance, 367
 Lactose permease, 766, 768–771, 769F, 770F, 1261
 Lactose synthase, 883F
 Lactosyl ceramide, 1010
lacZ gene, 107F, 120, 1294
lacZ' gene, 107F, 110, 111
 LADH, *see* Liver alcohol dehydrogenase
 Lagging strand synthesis, 103, 103F, 1175, 1175F, 1190–1191
 Lake, J., 1365
 Lamarck, Jean Baptiste de, 19
 Lamarckism, 19
 λ integrase, 1243–1244
 Lamprey, hemoglobin of, 194
 Lander, E., 181
 Landick, R., 1273
 Lands, W., 1005
 Landscape theory, of protein folding, 287–289, 288F
 Landsteiner, K., 414–415
 Lanosterol, 982–984, 985F
 Lanosterol synthase, 982
 Lanzilotta, W., 1055
 Large calorie (Cal), 53, 53T
 Large-T antigen, SV40, 1211
 Lariat structure, 1306, 1306F
late genes, 1284
 Lateral diffusion, 396
 Lathyrism, 238–239
 α -Latrotoxin, 780
 Lattman, E., 282
 Lauric acid, 387T
 Lavoisier, A., 823, 901
 LBHBs (low-barrier hydrogen bonds), 536
 LCa (clathrin light chain), 430, 432, 432F
 LCA2 (Leber's congenital amaurosis 2), 123
 LCAT, *see* Lecithin-cholesterol acyltransferase
 LCB (clathrin light chain), 430, 432, 432F
 Lck, 715
 LD₅₀, 539
 LDH (lactate dehydrogenase):
 half-life, 1408T
 in homolactic fermentation, 614–615, 615F
 isozymes, 277
 molecular mass, 140F
 X-ray structure of dogfish, 254, 255F, 256
 LDL (low density lipoproteins), 449, 449F
 and apoB-100, 1322
 and atherosclerosis, 458
 as cholesterol carrier, 449F
 cholesterol takeup by, 453–455
 properties, 449T
 LDL receptors (LDLR), 453–457, 453F, 454, 1317
 and apoB-100, 1322
 and atherosclerosis, 456–457
 and cholesterol homeostasis, 989–991, 990F
 and endocytosis mediation, 986, 986F
 L-DOPA, 1046, 1059–1060F
 Lead compounds (drug discovery), 539
 Leader sequence, 1297–1298
 Leading strand synthesis, 102–103, 1175, 1175F, 1190–1193
 Leading substrate, 498
 Lead poisoning, 1134
 Leaves, phylogenetic trees, 203, 203F
 Leberman, R., 1381–1382
 Leber's congenital amaurosis 2 (LCA2), 123
 Le Châtelier's principle, 59
 Lecithins, 389T, 390
 Lecithin-cholesterol acyltransferase (LCAT), 450T, 452, 453F
 Lectins, 366, 885
 Leder, P., 1342
 Lederberg, J., 1262
 Leflunomide, 1116
 Left-handed helix, 225F, 1149–1151
 Left-handed α helix, 224F
 Leghemoglobin, 219, 1083
 Lehninger, A., 823, 946
 Leloir, L., 631, 883
 Leloir pathway, 631
 Lennarz, W., 291
 Leptins, 1096–1099, 1101
 Leptin-E100, 1098F
 Lesch–Nyhan syndrome, 1114, 1135
 Leslie, A., 853, 857, 865
let-7 gene, 1326
 Lethal factor (LF), anthrax, 715
 Letsinger, R., 211
 Leucine, 26
 basic helix-loop-helix/leucine zipper, 987
 biosynthesis, 1075
 degradation, 1040–1041
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 in globular proteins, 246
 half-life, 1413T

- α helix/ β sheet propensity, 302T
 inserted by *E. coli* nonsense suppressor, 1362T
 in native unfolded proteins, 283
 side chains, 70, 264T
 specific rotation, 74
 structure and general properties, 68T
 Leucine aminopeptidase, 168T
 Leucine-rich repeats (LRRs), 1412
Leuconostoc mesenteroides, 139
 Leu-enkephalin, 673T, 685
 Leukocytes, 85
 Leukocyte elastase, 533
 Leukocyte interferon, 1399
 Leukotrienes, 993
 arachidonic acid precursor, 995–997
 biosynthesis, 1000–1002, 1002F, 1060F
 series-5, 1003
 Leukotriene A₄ (LTA₄), 1000
 Leukotriene B₄ (LTB₄), 1001
 Leukotriene C₄ (LTC₄), 1001
 Leukotriene D₄ (LTD₄), 1001
 Leukotriene E₄ (LTE₄), 1001
 Leukotriene receptors, 689
leu operon, 1299T
 Leupeptin, 556
 Levine, P., 83
 Levinthal, C., 284, 287
 Levinthal paradox, 284
 Levitt, M., 523
 Levorotatory molecules, 74
 Lewis, G. 45
 Lewis, M., 1294
 Lewis acids, 45, 511
 Lewis bases, 45
 Lewisite, 822
 LexA, 1221–1222, 1228
lexA gene, 1221
 LF (lethal factor), anthrax, 715
 L forms (amino acids), 74–76
 LH, *see* Luteinizing hormone
 LHC-II, 908–909, 909F
 LHCs, *see* Light-harvesting complexes
 Licensing, *see* Pre-replicative complex (Pre-RC)
 Lienhard, G., 515
 Life, 3–37
 and biochemical literature, 34–36
 and biochemistry, 14–19
 chiral molecules and, 78
 defined, 3
 divergence of major kingdoms, 192
 eukaryotes, 6–14
 evolutionary tree, 13F
 and genetics, 19–28
 origin of, 28–34
 prokaryotes, 3–6
 and thermodynamics, 52, 586–589
 and water's properties, 40, 41
 Life Science Directory, 194–195
 LigA, 1187
 Ligament, 235
 Ligands:
 Adair equation, 348–349
 in affinity chromatography, 142, 142F
 Hill equation for binding, 326–328
 Ligand-gated binding, 606
 Ligand-gated channels, 771–772
 Ligases, 479T
 Light absorption, 903–909
 Light chains, clathrin, 430
 Light-harvesting complexes (LHCs), 906–909, 908F
 Light reactions, photosynthesis, 902–926
 electron transport in purple bacteria, 909–913
 light absorption, 903–909
 photophosphorylation, 925–926
 two-center electron transport, 913–925
 Lignin, 369
 Lignoceric acid, 387T, 958
 Liljas, A., 1385
 Lima, C., 1327
 Limit dextrans, 385
 Limited proteolysis, 170
lin-4 gene, 1326
 LIN-14 protein, 1326
 Linear concentration gradients, 137F
 Linear gradient elution, 137
 Linear polymers, 70
 LINEs (long interspersed nuclear elements), 1246, 1249
 Lineweaver–Burk plot, 490, 490F, 494F–496F, 499F, 500F
 Linked genes, 24
 Linkers, in gene manipulation, 109–110, 110F
 Linking number, 1158, 1162
 Link proteins, 375
 Linoleic acid, 386, 387F, 387T, 950F, 971
 α -Linolenic acid, 387F, 387T, 971, 994
 γ -Linolenic acid (GLA), 387T, 994
 Lipase, 153T, 941–942, 941F
 Lipids, 15, 16F. *See also* related topics, e.g.:
 Fatty acids
 and acid-base groups, 48
 aggregates of, 393–399
 biosynthesis, 818, 973
 classification, 386–393
 fluidity, 388, 397–399
 metabolism, 561F
 in thylakoid membrane, 902
 Lipid bilayers, 44F, 395F, 397F. *See also*
 Membranes
 dynamics, 396–399, 397F
 fluidity, 397–399
 liposomes, 395–396
 micelles, 394–395, 394F
 phospholipid diffusion, 396F
 Lipid-linked membrane proteins, 406–408
 Lipid metabolism, 561F, 940–944
 cholesterol, 975–993
 eicosanoid, 993–1004
 fatty acid biosynthesis, 961–973
 fatty acid oxidation, 945–959
 ketone bodies, 959–961
 phospholipids and glycolipids, 1004–1013
 regulation of fatty acid metabolism, 973–975
 Lipidomics, 578
 Lipid-soluble vitamins, 474
 Lipid storage disease, 1011, 1013
 Lipid-water interfaces, 940–941
 Lipinski, C., 542
 Lipinski's rule of five, 542
 Lipitor, 990–991, 991F
 Lipmann, F., 578, 791, 892
 Lipoamide, in citric acid cycle, 794, 795F
 Lipogenesis, 1096
 Lipoic acid, 473T, 794, 795T
 Lipolysis, 1096
 Lipopolysaccharides, 378
 Lipoproteins:
 characteristics of major classes, 449T
 dysfunctions, 456–459
 function, 451–456
 structure, 449–451
 Lipoprotein lipase (LPL), 450T, 451, 943
 Liposomes, 122, 395–396
 Lipoxins (LXs), 993, 1003–1004
 15-*epi*-Lipoxin A₄ (15-*epi*-LXA₄), 1004
 Lipoxin A₄ (LXA₄), 1003F, 1004
 Lipoxin receptors, 689
 5-Lipoxygenase (5LO), 1000
 12-Lipoxygenase (12LO), 1000
 15-Lipoxygenase (15LO), 1000, 1001
 Lipoxigenase-1, 1001
 5-Lipoxygenase-activating protein (FLAP), 1001, 1002, 1002F
 Lipoyllysyl arm, dihydrolipoyl dehydrogenase, 797
 Lipscomb, W., 476
 Liquid scintillation counting, 572
 Literature search, 34–35
 Lithgow, T., 445
 Lithium ion, 45T, 1158
 Liver:
 AMP-dependent protein kinase in, 1096, 1097F
 and cholesterol biosynthesis, 987
 citric acid cycle in, 816, 817
 fructose utilization in, 630–631
 glucokinase in, 597
 gluconeogenesis in, 871
 glycogen storage in, 638–639, 667–668
 half-lives of enzymes in, 1408T
 heme synthesis regulation, 1055–1056
 and other organs, 1091F, 1093–1094
 rat liver cytoplasmic ribosomes, 1370T
 stress response of, 665F
 Liver alcohol dehydrogenase (LADH), 472, 505, 618–619
 Liver glycogen synthase deficiency, 668
 Liverworts, 13F
 Living systems, rise of, 33–34
 LKB1, 1096
 5LO (5-lipoxygenase), 1000
 12LO (12-lipoxygenase), 1000
 15LO, *see* 15-Lipoxygenase
 Local energy maxima, 289
 Local mediators, 671
 Lock-and-key hypothesis, 470
 Loewenstein, W., 416
 Logan, D., 1385
 Log odds, 198
 Log odds substitution matrix, 198
 Lohman, T., 1186
 Lon, 1419
 London dispersion forces, 260–261, 261F
 Longevity, and caloric restriction, 1101
 Long interspersed nuclear elements, *see* LINEs
 Long terminal repeats (LTRs), 1245–1246
 Long-term regulation, of fatty acid metabolism, 973
 Loop conformation, 230, 232
 Looped rolling circle replication, 1191–1193, 1192F
 Lopez, A., 717
 Lord, R., 1155
 Lorimer, G., 300
 Lovastatin, 990, 991F
 Low-barrier hydrogen bonds (LBHBs), 536
 Low density lipoproteins, *see* LDL
 Lowenstein, J., 1132
 Lowe syndrome, 735
 Lowry, T., 45
loxP gene, 1244
 LPL, *see* Lipoprotein lipase
 L-protein, 1032
 LRRs (leucine-rich repeats), 1412
 LT (heat-labile enterotoxin), 696
 LTA₄ (leukotriene A₄), 1000

- LTB₄ (leukotriene B₄), 1001
 LTC₄ (leukotriene C₄), 1001
 LTC₄ synthase, 1001
 LTD₄ (leukotriene D₄), 1001
 LTE₄ (leukotriene E₄), 1001
 LTRs (long terminal repeats), 1245–1246
 Lu, P., 1294
 Luciferase, 183
 Luciferin, 183
 Ludwig, M., 1036, 1037, 1125
 Lührmann, R., 1314, 1315
 Luisi, B., 800
 Lungfish, 94T
 Luteinizing hormone (LH), 154F, 673T, 683
 LXA₄, *see* Lipoxin A₄
 15-*epi*-LXA₄ (15-*epi*-lipoxin A₄), 1004
 LXs, *see* Lipoxins
 Lyases, 479T
 Lycopenes, 905F, 907
 Lydon, N., 719
 Lynen, F., 791
 Lyophilization, 141
 Lys, *see* Lysine
 Lysate, 130
 Lysidine, 1347F, 1352
 Lysine:
 amino group, 208F
 biosynthesis, 1072–1073, 1075
 degradation, 1040–1041
 and *E. coli* nonsense suppressor, 1362T
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 isoelectric point, 72
 in native unfolded proteins, 283
 side chains, 71, 208F, 264T
 structure and general properties, 69T
 trypsin-catalyzed cleavage, 170
 Lysis, 27, 130
 Lysogenic mode, 1243
 Lysolecithin, 452
 Lysophosphatidic acid, 971
 Lysophospholipase A₂, 942
 Lysophospholipids, 942
 Lysosomal protease, 525T
 Lysosomes, 7F, 9–10, **1408–1409**
 metabolic functions, 562T
 in secretory pathway, 420, 421F
 and vesicle formation, 428–440
 and vesicle fusion, 440–445
 Lysozyme, 470, 479, **517**
 in amyloid diseases, 310–311
 and bacterial cell walls, 376
 catalytic mechanism, **520–525**, 521F
 covalent intermediate of, 524F
 hen egg white, 275–276, 276T, 470,
 517–525, 517F, 518F
 isoelectric point, 134T
 for protein solubilization, 130
 structure, **517–520**
 substrate interactions with, 510F
 transition state, 523F
 Lysozyme amyloidosis, 310
 LysRS, 1351
 Lysyl group, 73T
 Lysyl hydroxylase, 240, 1405
 Lysyl oxidase, 238, 240
 Lytic mode, 1243
 α -Lytic protease, 525T
 D-Lyxose, 360F
- m⁴C(*N*⁴-Methylcytosine) residue, 1246
 m⁵C-MTase, *see* 5-Methylcytosine methyltransferase
 m⁵C (5-Methylcytosine) residue, 1246
 m⁶A (N⁶-Methyladenine), 1246
 M6P/IGF-II receptor, 890
 m⁷G, *see* N⁷-Methylguanosine
 m⁷GDP, 1378
 m²G (N²,N²-Dimethylguanosine), 1347F
 MacKinnon, R., 751, 752, 755, 774
 MacLeod, C., 86
 Macrocytes, 1064
 Macrofibrils, 234, 234F
 α_2 -Macroglobulin, 154F
 Macromolecular mimicry, 1386
 Macromolecules, 15, 16F, 17, 257
 Macrophages, 381, 457, 671
 Macrophage colony-stimulating factor (M-CSF), 717
 Macroscopic dissociation constants, 348
 Mad cow disease, 312, 315
 Magnesium ion, 45T, 1158, 1308–1309
 Maintenance methylation, 1249–1251, 1249F
 Maize, *see* Corn
 Major facilitator superfamily (MFS), 751
 Major groove, in DNA 90, 1145
 Major histocompatibility complex (MHC) proteins, 1102
 Malaria, 187–188, 343, 898, 1058
 Malate, 790F, 791, **813**, 936
 L-Malate, 481, 947
 Malate-aspartate carrier, 827, 827F
 Malate dehydrogenase, 876, 947
 in citric acid cycle, 790F, 791, **813**
 molecular mass, 140F
 Malate dehydrogenase, decarboxylating, 956
 Malate- α -ketoglutarate carrier, 827
 Malate synthase, 880
 Malathion, 527
 MALDI (matrix-assisted laser desorption/ionization), 172–173
 Male sex hormones, 681
 Malic enzyme, 956
 Malignant transformation, 1211–1212
 Malignant tumors, 381, 703. *See also* Cancer
 Malonate, 493, 791, **811**
 Malonic semialdehyde, 1136F
 Malonyl/acetyl-CoA-ACP transacylase (MAT), 964
 Malonyl-CoA, 961, 1136F
 Malonyl/palmitoyl transferase (MPT), 967
 Maltodextrins, 749
 Maltoheptaose, 640F
 Maltoporin, **749–750**, 749F
 Maltose, 367, 367F
 Maltose binding protein, 300
 Maltose phosphorylase, 501
 Maltotriose, 370
 Mammals. *See also* specific types
 cloning, 1251
 CpG methylation, 1249
 cytoplasmic adenylation of mRNAs, 1402
 DNA base composition, 84
 energy metabolism strategies, **1088–1090**
 genomic imprinting, 1251
 glycogen synthase, 645
 homolactic fermentation, **614–615**
 isoprenoid metabolism, 976F
 keratin, 232–233
 LDL-receptor mediated endocytosis, 986, 986F
 mitochondrial DNA replication, 1207, 1207F
 mitochondrial deviations from ideal, 1344T
- mRNAs, 1302
 mutation rate, 192
 myoglobin, 324
 POH1, 1418
 signal recognition particles, 422
 snoRNAs, 1329
 Mammalian target of rapamycin (mTOR), 734, 738
 Man, *see* Mannose
 Manganese ion, 1158
 Maniatis, T., 1306
 Manley, J., 1312
 Mannose, 630, 633, 633F, 747T
 D-Mannose, 360F, 361
 Mannosidase, 366
 D-Mannuronic acid, 364
 Map-based genomic sequencing, 180–181, 181F
 MAP (mitogen-activated protein) kinase, 712–714
 MAP kinase kinase (MKK), 713
 MAP kinase kinase kinase (MKKK), 713
 MAP-kinase signaling cascades, **712–714**
 Maple syrup urine disease (MSUD), 1039
 Margaric acid, 1018
 Margoliash, E., 189
 Margulis, L., 9
 Marmur, J., 93
 Mars, 39
 Martin, A., 144
 Martius, C., 791
 Mas37, 449
 Masking of mRNA, 1401–1402
 Massey, V., 801
 Mass spectrometry, *see* MS
 Mast cells, 372
 MAT (malonyl/acetyl-CoA-ACP transacylase), 964
 Matrix, 9, 445, 448, 824, 825F
 Matrix-assisted laser desorption/ionization (MALDI), 172–173
 Matrix-associated regions, *see* MARs
 Matrix processing peptidase (MPP), 447
 Matrix protein 1, *see* M1
 Matrix space, 9
 Mattevi, A., 1066
 Matthaei, H., 1341, 1343
 Matthews, B., 282
 Matthews, D., 1128
 Matthews, R., 1036, 1037
 Maturity-onset diabetes, 310, 1102–1103
 Maximal velocity (V_{max}), 489
 Maximum likelihood (ML) tree-building criteria, 205
 Maximum parsimony (MP) tree-building criteria, 205
 Mayer, R., 901
 Mb, *see* Myoglobin
 MBD (methyl-CpG binding domain), 1249
 MCAD (medium-chain acyl-CoA dehydrogenase), 948
 McArdle's disease, 644, 667
 McCarty, M., 86
 McClintock, B., 1236
 McDermott, A., 604, 606
 Mcm2, 1205
 Mcm3, 1205
 Mcm4, 1205
 Mcm5, 1205
 Mcm6, 1205
 Mcm7, 1205
 Mcm10, *see* DDK
 MCM complex, 1205, 1206
 MCPA-CoA (methylene cyclopropylacetyl-CoA), 948, 948F

- M-CSF (macrophage colony-stimulating factor), 717
- Mdm10, 449
- MDPK (myotonic dystrophy protein kinase), 1252
- MDR (multidrug resistance) transporter, 766–767
- Mechanism-based inhibitors, 1128
- Mechanosensitive channels, 772
- Medial cisternae, 428, 429F
- Mediated membrane transport, 745–748
- Medium-chain acyl-CoA dehydrogenase (MCAD), 948
- MedLine, 35, 195
- Medulla, 679
- Megaloblastic anemia, 474T
- Meiosis, 20, 21F, 24F, 1224
- Meister, A., 1061
- MEK, 713
- MEKK1, 714
- Melanin, 1046
- α-Melanocyte stimulating hormone (α-MSH), 1099
- Melezitose, 385
- Mello, C., 1323
- Melting curve:
- DNA, 93, 93F
 - proteins, 265F
- Melting temperature (T_m):
- DNA, 93, 93F
 - proteins, 265
- Memapsin 2, 547
- Membranes:
- assembly and protein targeting, 418–449
 - and blood groups, 414–415
 - composition of, 398T
 - Donnan equilibrium for, 787
 - of erythrocytes, 411–414, 411F
 - fluid mosaic model, 408–411
 - function of, 386
 - gap junctions, 415–416
 - lipid distribution in, 419–420
 - secretory pathway, 420–428
 - vesicle formation, 428–440
 - vesicle fusion, 440–445
- Membrane potential, 447, 745
- Membrane proteins, 399–406
- channel-forming, 416–418
 - integral, 399–406
 - lipid-linked, 406–408
- Membrane transport:
- ABC transporters, 766–768
 - action potentials, 775–777
 - aquaporins, 756–757
 - ATP–ADP translocator, 771
 - ATP-driven active, 758–768
 - Ca²⁺-ATPase, 762–764
 - Cl[−] channels, 755–756
 - general scheme, 747F
 - group translocation, 765–766
 - and (H⁺–K⁺)-ATPase, 764–765
 - ion-driven active, 768–771
 - in ionophores, 748–749
 - K⁺ channels, 752–755
 - kinetics of mediated transport, 746–748
 - lactose permease, 769–770
 - maltoxin, 749–750
 - mediated, 745–748
 - Na⁺–glucose symport, 768–769
 - (Na⁺–K⁺)-ATPase, 758–762
 - and neurotransmission, 771–784
 - nonmediated, 745–746
 - passive-mediated glucose transport, 750–752
- thermodynamics of, 744–745
- voltage-gated ion channels, 771–775
- Menaquinone, 910
- Mendel, G., 20–22
- Mendelian inheritance, 20–22, 107F
- Menstrual cycle control, 683–684
- Menten, M., 488
- 2-Mercaptoethanol, 168
- 6-Mercaptopurine, 1142
- Merozygote, 1263
- Merrifield, B., 206, 208
- Mertz, J., 109
- Meselson, M., 90, 1173, 1264
- Meso form, 76
- Mesophiles, 266
- Mesophyll cells, 904
- Mesosomes, 4, 4F
- Messenger RNA, *see* mRNA
- Met, *see* Methionine
- Metabolic adaptation, 1102–1104
- Metabolic control, 619–621
- Metabolic control analysis, 620–624
- Metabolic flux, 620, 624
- Metabolic homeostasis, 619–620, 1096–1101
- Metabolic inhibitors, 569
- Metabolic pathways, 14, 17, 559–562
- and carbon–carbon bonds, 567–569, 568F
 - in cells, 560F
 - chemical logic, 563–565
 - eliminations, isomerizations, and rearrangements, 566–567
 - energy metabolism, 1088–1090
 - evolution, 34
 - group transfer reactions, 564–565
 - interorgan specialization, 1090–1095
 - organelles in, 562T
 - organic reaction mechanisms, 562–569
 - oxidations and reductions, 565–566
- Metabolic regulation, 619
- Metabolic syndrome, 1104
- Metabolism, 559. *See also* specific types of metabolism
- cancer cells, 864–865
 - experimental study of, 569–575
 - inborn errors of, 26
 - isotope studies of, 572–575
 - metabolic inhibitors, growth studies, and biochemical genetics, 569–572
 - organs, cells, and subcellular organelles in, 575
 - oxidation–reduction reactions, 584–586
 - phosphate compound thermodynamics, 578–583
 - systems biology, 576–578
 - and thermodynamics of life, 586–589
- Metabolites, 559
- Metabolite carrier proteins, 448
- Metabolome, 576
- Metabolomics, 578
- Metabolons, 595, 817
- Metachromatic leukodystrophy, 1013T
- Metal-activated enzymes, 511
- Metal chelation affinity chromatography, 145
- Metal ion catalysis, 511–512
- Metalloenzymes, 511
- Metaphase, 20F, 21F
- Metastasis, 703
- Metazoa, 12, 1202, 1206–1207
- Met-enkephalin, 673T, 685
- Methane, 32, 42
- Methanococcus*, 7F
- Methanococcus jannaschii*, 424, 425F
- Methanogens, 6
- Methanol, 43T, 505
- Methemoglobin (metHb), 324, 342
- Methemoglobinemia, 342
- Methemoglobin reductase, 324–325
- N⁵,N¹⁰-Methenyltetrahydrofolate, *see* MTHF
- Methionine:
- biosynthesis, 1072–1073, 1075
 - degradation, 1034–1039
 - as essential amino acid, 1065T
 - genetic codes for, 100T, 1343T
 - in globular proteins, 246
 - half-life, 1413T
 - α helix/β sheet propensity, 302T
 - in native unfolded proteins, 283
 - side chains, 70, 264T
 - structure and general properties, 68T
- Methionine sulfone, 1066
- Methionine synthase, 955, 1037
- Methionyl-tRNA synthetase (MetRS), 100, 1352
- Methotrexate, 445, 492–493, 1129
- 9-Methyladenine, 1156F
- N⁶-Methyladenine (m⁶A) residue, 1246
- 1-Methyladenosine (m¹A), 1347F
- N-Methylalanine, 32T
- Methylamine methyltransferase, 1361
- Methyl arachidonyl fluorophosphonate, 405F
- ω-N-Methylarginine, 79F
- α-Methylaspartate, 1033F
- Methylation analysis, 366
- Methylcobalamin, 1037
- Methyl-CpG binding domain (MBD), 1249
- β-Methylcrotonyl-CoA carboxylase, 962
- Methylcytidine (m³C), 1347F
- N⁴-Methylcytosine (m⁴C) residue, 1246
- 5-Methylcytosine methyltransferase (m⁵C-MTase), 1247–1248, 1247F
- 5-Methylcytosine (m⁵C) residue, 1246
- N⁶-Methyl-dA, 85
- 5-Methyl-dC, 85
- α-β-Methylene-ADP, 1108F
- N,N'-Methylenebisacrylamide, 147, 147F
- Methylenecyclopropylacetyl-CoA, *see* MCPA-CoA
- N⁵-N¹⁰-Methylene-tetrahydrofolate, 1031, 1062
- N⁵-N¹⁰-Methylene-tetrahydrofolate reductase (MTHFR), 1035–1037
- Methyl-d-glucosides, 363F
- Methylglyoxal, 606
- O⁶-Methylguanine residue, 1216
- N⁷-Methylguanosine (m⁷G), 1302, 1347F
- 3-Methylhistidine, 79F
- Methylmalonic aciduria, 955
- Methylmalonyl semialdehyde, 737F
- Methylmalonyl-CoA, 1136F
- L-Methylmalonyl-CoA, 567
- (S)-Methylmalonyl-CoA, 952, 952F, 953
- Methylmalonyl-CoA mutase, 567, 953–956, 954F, 955F, 957F
- Methylmalonyl-CoA racemase, 953
- Methyl-p-nitrobenzene sulfonate, 536–537
- N-Methyl-N'-nitro-N-nitrosoguanidine (MNNG), 1215–1216
- Methylphosphonate, 1402
- O²-Methylribose, 1328
- Methyl silicones, 31
- N⁵-Methyl-tetrahydrofolate, 1034, 1062
- Methyl thiocyanate, 170
- 4-Methylumbelliferyl-β-D-N-acetylglucosamine, 1011
- 5-Methyluracil, *see* Thymine
- N-Methylurea, 32T
- MetI, 1290
- Metmyoglobin (metMb), 324
- Metmyoglobin reductase, 325
- met repressor, 1290–1291, 1291F

- MetRS, *see* Methionyl-tRNA synthetase
 Met-tRNA_i^{Met}, 1373–1374
 Met-tRNA_i^{Met}, 1376, 1401
 Metzler, D., 1021
 Mevacor, 990, 991F
 Mevalonate, 977
 Mevalonate-5-phototransferase, 977
 Mevalonolactone, 987
 Mevinolin, 990
 Meyerhof, O., 595
 Mfd protein, 1218
 MFS (major facilitator superfamily), 751
 Mge1, 447
 MHC (major histocompatibility complex) proteins, 1102
 M.HhaI, 1248, 1250
 M.HhaI DNA methyltransferase, 1248F
 mHsp70, 447
 Mice:
 cancer cells, 382F
 carcinogenesis tests, 1224
 genome sequencing, 176, 177T
 injection with transformed pneumococci, 86
 knockout, 121
 liver membrane composition, 398T
 monoclonal antibodies from, 119
 obese and diabetic, 1098–1099
 telomere length, 1211–1212
 transgenic, 86, 86F, 121F
 Micelles, 44, 44F, 51, **394–395**, 394F
 Michaelis, L., 488
 Michaelis complex, 488
 Michaelis constant (K_m), 489
 Michaelis–Menten equation, **488–491**, 488F
 Michaelis–Menten kinetics:
 derivations, 501–503
 for enzymes, 489T
 inhibition, 493–496
 pH dependence, 496–497
 Michel, H., 403, 840, 907, 910
 Microbodies, 10
 Microdomains, 410
 Microfibrils, 234, 234F
 Microfilaments, 10
 Microheterogeneity, 373
 MicroRNAs (miRNAs), 1325–1326
 Microscopic dissociation constants, 348
 Microscopic ionization constants, 49, 50
 Microsomes, 154F, 883, **1362**
 Microsporidia, 7F
 Microtubules, 10
 Microvilli, 38
 middle genes, 1284
 Miescher, F., 85
 Migration rate (paper chromatography), 144
 Miles, E., 1076
 Miller, C., 774
 Miller, S., 32–33
 Miller–Urey experiments, 32T
 Milstein, C., 420
 Mineral acids, 46
 Mineralcorticoids, 673T, 680, 991
 Minor groove, in DNA, 90, 1145
 Minot, G., 956
 miRBase database, 1326
 MiRNAs (microRNAs), 1325–1326
 Mismatch repair, **1220**, 1220F
 Missense suppressors, 1362
 Mitchell, P., 846
 Mitochondria, 7F, **9**, 11F, 16F, **823–828**, 825F.
 See also Electron-transport chain
 acetyl-CoA transport in, **968–969**
 anatomy, 824, 825F
 citric acid cycle in, 809, 815
 cytoplasmic shuttle systems, 827–828
 fatty acid transport in, 946–947, 946F
 gluconeogenesis in, 876–877, 877F
 heme biosynthesis in, 1054
 lipid composition of beef heart membrane, 398T
 membranes of, 853F
 metabolic functions of, 562T
 mouse liver, 833F
 photorespiration in, 935
 protein importing in, 445, 446F
 protein targeting in, **445–449**
 sedimentation coefficient, 154F
 transport systems, 824–828, 826F
 wobble pairings of tRNA, 1361
 Mitochondrial carrier family, 771
 Mitochondrial DNA, *see* mtDNA
 Mitochondrial electron-transport chain, *see* Electron-transport chain
 Mitochondrial import stimulation factor (MSF), 445
 Mitochondrial inner membrane, 398T
 Mitochondrial outer membrane, 398T
 Mitochondrial trifunctional protein, 948–949
 Mitogens, 1202, 1401
 Mitogen-activated protein (MAP) kinase, 712–714
 Mitosis, **20**, 20F, 1202
 Mitotic spindle, 10, 20, 1202
 Mixed inhibition, **495–496**, 496F, 502
 Mixed triacylglycerols, 388
 MJ33, 943
 MK-591, 1001–1002, 1002F
 MK0886, 1001
 MKK (MAP kinase kinase), 713
 MKK4, 714
 MKK7, 714
 MKKK (MAP kinase kinase kinase), 713
 MKKK kinase (MKKKK), 714
 ML (maximum likelihood tree-building criteria), 205
 MLCK (myosin light chain kinase), 656
 MLK3, 714
 MMDB, *see* Molecular Modeling Database
 MNNG (*N*-methyl-*N'*-nitro-*N*-nitrosoguanidine), 1215–1216
 Mobile genetic elements, 1225–1236
 Mobile loop, GroES system, 294, 307F
 Mobile phase, 135
 Mobility, 45–46, 45F, 45T, 745
 Modrich, P., 1220
 Modulator, 348
 MoFe-protein, 1080–1081
 Mohammadi, M., 701
 Molar extinction coefficient, 90
 Molecular archeology, 116
 Molecular biology, 95, 95F, 1260
 Molecular chaperones, 290, **293–302**.
 See also Chaperone proteins
 Molecular clock, 192
 Molecular cloning, **104–124**
 cloning vectors, **106–109**
 gene manipulation, **109–111**
 gene therapy, **122–123**
 genomic libraries, **113–114**
 and nuclear sequencing, 176
 and polymerase chain reaction, **114–117**
 protein production, **117–121**
 and protein purification, 130
 restriction endonucleases, **104–106**
 social, ethical, and legal considerations, 123, 124F
 Southern blotting, **111–113**
 transgenic organisms, **121–123**
 Molecular disease, **185–187**
 Molecular dynamics simulation, 307–308, 308F, 397
 Molecular ionization constants, 49–50
 Molecularity, reactions, 483
 Molecular mass, 5n
 by gel filtration chromatography, 138, 139, 140F
 in protein sequencing, 169
 by SDS-PAGE, 150, 151F
 Molecular Modeling Database (MMDB), 256T, 258
 Molecular sieve chromatography, 138
 Molecular size, protein purification and, 132
 Molecular weight, 5n
 Molecular weight cutoff, 141
 Mollusks, 13F
 Molten globule, 287
 Molybdopterin complex (Mo-pt), 1132
 Mondragón, A., 1164–1165, 1331
 Monera, 6
 Monoclonal antibodies, 119–120, 131–132, 143
 Monocyclic cascade, 650
 Monod, J., 349, 1262, 1264–1265
 Monogalactosyl diacylglycerol, 902
 Monooxygenases, 543
 Monoprotic acids, 49
 Monosaccharides, **359–363**. *See also* specific Monosaccharides
 Monosodium glutamate (MSG), 1064
 Monotypic proteins, 404, 426
 Montagnier, L., 545
 Moody, P., 1216
 Moore, P., 1365, 1366, 1383
 Mo-pt (molybdopterin complex), 1132
 Moras, D., 1357
 Morgan, T. H., 23, 24
 Morikawa, K., 1199, 1231
 Morphoains, 1050
 Mosses, 13F
 Most Representative NMR Structure in an Ensemble, 256T
 Mother cell, 1284
 Motifs, 249–251
 Mouse, *see* Mice
 Moving boundary electrophoresis, 146
 Mozzarelli, A., 353
 MPA (mycophenolic acid), 1112
 M phase, **1202**, 1206
 MPP (matrix processing peptidase), 447
 MPT (Malonyl/palmitoyl transferase), 967
 MP (maximum parsimony) tree-building criteria, 205
 Mre11 complex, 1224
 mRNA (messenger RNA), 1145, 1260, **1264–1265**
 affinity chromatographic isolation, 158
 base pairing with 16S rRNA, 1374–1375
 degradation, 1327
 and enzyme induction, 1260–1264
 masking and cytoplasmic polyadenylation, 1401–1402
 methylation, 1320
 posttranscriptional modification, 1302–1327
 reading of, 1342–1343, 1342F, 1372
 transport of, 1326–1327
 MS (mass spectrometry), 169, 172–174, 172F, 366
 MsbA, 767–768, 767F
 MSF (mitochondrial import stimulation factor), 445
 MSG (monosodium glutamate), 1064
 α-MSH (α-melanocyte stimulating hormone), 1099
 MS/MS, *see* Tandem mass spectrometry

- mSos, 709
 MspI, 105T
MspI, 1249
 MSUD (maple syrup urine disease), 1039
 MTase, 1247–1248
 mtDNA, 116, 1207, 1207F
 MTHF (N^5,N^{10} -methylene-tetrahydrofolate), 1129, 1215
 MTHFR (N^5,N^{10} -methylene-tetrahydrofolate reductase), 1035–1037
 mtHsp70, 447
 mTOR, *see* Mammalian target of rapamycin
 Mtr3, 1327
 Mucin, 381, 890, 890F
 Mucolipidosis II, 439
 Mucopolysaccharides, 371
 Mucus, 381
 Mule, 1251
 Müller-Hill, B., 1284, 1294
 Mullis, K., 114
 Multicellular organisms, 12, 34, 1202.
See also specific organisms
 Multidrug resistance (MDR) transporter, 766–767
 Multienzyme complexes, 792, 794
 Multiforked chromosomes, 1195, 1195F
 Multipass transmembrane proteins, 426
 Multiple sclerosis, 777
 Multiple sequence alignment, 201–203
 Munc18, 441
 Mupirocin, 1356, 1356F
 Murein, 376
 Murine eIF4e, 1378F
 Murphy, W., 956
 Musacchio, A., 708
 Muscarine synapses, 778
 Muscarinic synapses, 778
 Muscle, 10
 AMP-dependent protein kinase in, 1096, 1097F
 anaerobic glycolysis in, 615
 citric acid cycle for, 792, 817
 fibrous proteins, 232
 free energy changes of glycolysis in, 625T
 fructose utilization in, 630
 glucose uptake regulation in, 751F
 glycogen storage in, 638–639, 667
 glycolysis regulation in, 624–630
 and other organs, 1091–1093, 1091F
 slow and fast-twitch fibers, 619
 Muscle fatigue, 615, 1092–1093
 Muscle glycogen synthase deficiency disease, 668
 Muscular dystrophy, 1252
 Mutagens, 104, 570
 Mutagenesis, 119–120, 1224–1225, 1225F
 Mutarotation, 362–363, 507–508
 Mutase, 610
 Mutations:
 in DNA replication, 102
 dynamic, 1251–1252
 and evolution, 18–19, 185
 nonsense, 1362
 point, 192, 1339–1340
 random nature of, 195
 rates of, 192
 up and down, 1266
 and viruses, 27, 27F
 MutH, 1220
 MutL, 1220
 MutS, 1220
 MWC model, of allosteric interactions, 349–351
 Myasthenia gravis, 781
 Myc, 713, 737F
Mycobacterium leprae, 1234
Mycobacterium tuberculosis, 177T
Mycobacterium xenopi, 1407, 1407F
 Mycophenolic acid (MPA), 1112
Mycoplasma, 4F
Mycoplasma genitalium, 177T
Mycoplasma hominis, 94T
 Mycoplasmas, 6
 Myelin, 398T, 777
 Myelination, 777–778, 778F
 Myelin membrane, 398T
 Myeloma, 131
 Myoadenylate deaminase deficiency, 1132
 Myocardial infarction, 456, 865
 Myoglobin (Mb), 323, 331–332, 332F
 α and β chains of, 332T–333T
 conformation of, 306F
 electron density map, 242F
 ESI-MS spectrum, 173, 173F
 gene duplication, 193–194
 α helix, 227F
 H helix, 248F
 Hill plot, 327F
 isoelectric point, 134T
 mobility, 308F
 in molecular dynamics simulation, 308F
 molecular mass, 140F
 nitric oxide detoxification by, 325
 oxygen binding by, 325–326, 326F
 physical constants of, 153T
 sequence alignment, 197F
 solubility in ammonium sulfate solution, 133F
 tertiary structure, 245–246
 X-ray structure, 241F, 245F
 Myokinase, 627. *See also* Adenylate kinase (AK)
 Myosin, 10, 301
 Myosin light chain kinase (MLCK), 656
 Myotonic dystrophy (DM), 1252
 Myotonic dystrophy protein kinase, 1252
 Myristic acid, 387T, 407
 Myxobacteria, 13F
 Myxothiazol, 849
 N
 Na⁺, *see* Sodium ion
N-Acetylneurameric acid, *see* NANA
 Na⁺ channels, 772, 776–777
 NAD⁺ (nicotinamide adenine dinucleotide), 82, 471F, 791
 790F, 766F
 biosynthesis, 1137F, 1138
 in catabolism, 561F
 in citric acid cycle, 794, 795T
 and DNA ligase, 1187
 domains, 248
 in glycolysis, 585, 594F
 in glyoxylate cycle, 880
 in oxidation-reduction reactions, 565
 and yeast alcohol dehydrogenase, 471–472
 NADD, 471
 NADH (nicotinamide adenine dinucleotide, reduced form):
 766F, 766F
 in catabolism, 561F
 in citric acid cycle, 789, 791, 819
 as electron-transfer coenzyme, 586
 in electron-transport chain, 823, 824F, 828, 829
 in fatty acid oxidation, 950
 in glycolysis, 585, 594F
 in oxidation-reduction reactions, 565
 and yeast alcohol dehydrogenase, 471–472
 NADH:Coenzyme Q reductase, *see* Complex II
 NADH-cytochrome *b*₅ reductase, 971
 NADH dehydrogenase, *see* Complex I
 NADH-dependent dehydrogenases, 472
 NAD⁺ kinase, 1137F, 1138
 NADP⁺ (nicotinamide adenine dinucleotide phosphate), 471F, 473, 931, 1137F
 NADPH (nicotinamide adenine dinucleotide phosphate, reduced form):
 in Calvin cycle, 931
 in metabolic pathways, 561, 561F
 and photorespiration, 935
 in photosynthesis, 902, 926
 and yeast alcohol dehydrogenase activity, 471F
 NADPH-P450 reductase, 543–545
 NAD⁺ pyrophosphorylase, 1137F, 1138
 NAG (*N*-Acetylglucosamine), 365, 371, 376, 377F
 lysozyme catalytic mechanism, 517–520, 517F, 522–525
 in peptidoglycan, 377F
 rates of hydrolysis, 518T
 X-ray structure of, 518F
 NAG-β(1 → 4)-2-deoxy-2-fluoro-β-D-glucopyranosyl fluoride (NAG2FGlcF), 524
 Nagai, K., 1314, 1315
 Na⁺-glucose symport, 768–769, 768F, 769F
 Nails, 233, 234, 275
 Nakamura, Y., 1394
 (Na⁺–K⁺)-ATPase ((Na⁺–K⁺) pump), 758–762, 761F
 NAM (*N*-acetylmuramic acid), 365, 365F, 376, 377F
 lysozyme catalytic mechanism, 517, 517F, 519–520, 522–525
 rates of hydrolysis, 518T
 NANA (*N*-acetylneurameric acid), 365, 365F, 1010
 NAS, *see* Normalized alignment score
 Nathans, D., 105
 National Center for Biotechnology Information (NCBI), 35, 201
 Native chemical ligation, 209, 209F
 Native form (DNA), 88
 Native proteins, 221, 264
 crystalline, 243
 folding into, 278
 free energy, 57
 unfolded forms, 283–284, 284F
 Natural selection, *see* Evolution
 Navia, M., 549, 724
 NCBI, *see* National Center for Biotechnology Information
 NDB, *see* Nucleic Acid Database
 NDP, *see* Nucleoside diphosphate
 Neandertals, 116
 Nearest-neighbor analysis, 127
 NEDD8, 1421
 Needleman, S., 199
 Needleman–Wunsch sequence alignment algorithm, 199–201, 200F
 Negative allosteric effect, 348
 Negative design, of proteins, 305
 Negatively cooperative binding, 327
 Negative regulator, 1286
 Neighbor-joining (N-J) method, for phylogenetic tree construction, 204
 Nelfinavir, 550F
 Nelson, N., 923
 NEM (*N*-ethylmaleimide), 440–441
 Nematodes, 13F, 1320. *See also* *Caenorhabditis elegans*

- Némethy, G., 352
 NEMO subunit, 1420
 NEM-sensitive fusion protein,
see NSF protein
 N-end rule, 1413
 NER, *see* Nucleotide excision repair
 Nernst, W., 584
 Nernst equation, 584
 Nernst–Planck equation, 745
 Nerve growth factor (NGF), 717
 Nerve transmission, 440F
 Nervonic acid, 387T
 Nested primers, 116
 Neuberg, C., 595
 Neupert, W., 445
 Neural tube defects, 1035
 Neurocan, 373T
 Neurofibrillary tangles, 311, 312
 Neuroglobin, 325
 Neurohypophysis, 683
 Neurons, 12F
 Neuronal membranes, anesthetics and,
 398–399
 Neuronal NOS (nNOS), 686–687, 687F
 Neuropeptides, 783–784
 Neuropeptide Y (NPY), 1099
Neurospora, 13F, 26
Neurospora crassa, 1319, 1320F, 1344T
 Neurotoxins, 442, 442F, 776–777
 Neurotransmission:
 action potentials, 775–777
 voltage-gated ion channels, 771–775
 Neurotransmitters, 527, 778–784
 amino acids as, 80
 synthesis of, 1058–1060
 in vesicles, 440
 Neurotransmitter receptors, 778–784
 Neutral drift, 188–192
 Neutral fats, 388
 Neutral solutions, 47
 Neutrophils, 688
 Newsholme, E., 629, 973
 New variant Creutzfeldt–Jakob disease
 (nvCJD), 315
 N-extein, 1406F, 1407
 NFAT_p, 724
 NF-κB (nuclear factor-κB), 1414
 NGF (nerve growth factor), 717
 NH₄⁺, *see* Ammonium ion
 NHEJ (nonhomologous end-joining),
 1223–1224, 1224F
 N^ω-hydroxyl-L-arginine (NOHA), 686, 686F
 Niacin, 474
 Niacinamide, 474
 Nicholls, A., 259
 Nicholson, G., 408
 Nick translation, 1181, 1187
 Nicotinamide, 474T, 1137F
 Nicotinamide adenine dinucleotide,
see NAD⁺
 Nicotinamide adenine dinucleotide, reduced
 form, *see* NADH
 Nicotinamide adenine dinucleotide
 phosphate, *see* NADP⁺
 Nicotinamide adenine dinucleotide
 phosphate, reduced form, *see* NADPH
 Nicotinamide coenzymes, 471F, 473T,
1136–1138
 Nicotinamide mononucleotide, *see* NMN
 Nicotinamide phosphoribosyltransferase,
 1137F, 1138
 Nicotinate, 1137F
 Nicotinate adenine dinucleotide (desamido
 NAD⁺), 1138
 Nicotinate mononucleotide, 1137F, 1138
 Nicotinate phosphoribosyltransferase, 1136,
 1137F, 1138
 Nicotine, 778
 Nicotinic acid, 474
 Nicotinic synapses, 778
 NIDDM, *see* Noninsulin-dependent diabetes
 Niemann–Pick disease, 1013T
 Nierhaus, K., 1373, 1388
 Nigericin, 869
 NIH shift, 1045
 19S caps, 1414, 1417
 19S regulator, 1414
 Nirenberg, M., 1341–1343
 Nitrate ion, 324
 Nitrate reductase, 1083
 Nitric acid, 46
 Nitric oxide (NO), 324
 binding to heme group, 324
 detoxification by myoglobin, 324
 hormonal roles, **685–688**
 Nitric oxide synthase, 685–687
 Nitritration, 1083
 Nitroacetic acid, 857
 Nitrite, 1340
 Nitrite ion, 324
 Nitrite reductase, 934, 1083
 3-Nitro-2-(S)-hydroxypropionate, 812
 Nitrocellulose, 111, 113
 Nitrogen, 31
 Nitrogenase, 1080–1082
 Nitrogen cycle, 1083, 1083F
 Nitrogen excretion, 1134–1135
 Nitrogen fixation, 5–6, **1078–1083**
 Nitrogen-fixing bacteria, 1080–1082
 Nitrogen mustard, 1340
 Nitrogenous bases, 83. *See also* Nucleic
 acid bases
 Nitroglycerin, 686
 p-Nitrophenolate ion, 513, 525–526
 p-Nitrophenylacetate, 513, 525–526, 526F
 Nitrosamines, 1340
 Nitrosoheme, 687
 Nitrous acid, 1339, 1340F
 N-J method, for phylogenetic tree
 construction, 204
 N-Linked glycoproteins, 882–889
 N-Linked oligosaccharides, 374F, 379–380,
 379F, 881, 888F
 synthesis of glycoproteins from, 882–889
 NLS (nuclear localization signal), 1414
 NMN (nicotinamide mononucleotide),
 1137F, 1138, 1187
 NMR spectroscopy:
 interpretation of structures, 243–244
 for metabolism studies, 569, 572
 with pulsed H/D exchange, 285–286
 nNOS, *see* Neuronal NOS
 NO, *see* Nitric oxide
 Nodes of Ranvier, 777
 NOESY, *see* Nuclear Overhauser effect
 spectroscopy
 NOHA, *see* N^ω-hydroxyl-L-arginine
 Noller, H., 1364, 1366, 1387, 1388, 1392, 1394
 Nomenclature:
 amino acid, 72–73
 enzymes, **479**
 Nomura, M., 1365
 Noncompetitive (mixed) inhibition,
495–496, 496F, 502
 Nonconjugate flow, 587
 Noncooperative binding, 327
 Nonequilibrium thermodynamics, 587
 Nonessential amino acids, 1019,
1064–1072, 1065T
 Nonhomologous end-joining, *see* NHEJ
 Noninsulin-dependent diabetes (NIDDM),
 310, 1102–1104, 1103F
 Nonketonic hyperglycinemia, 1032–1033
 Nonmediated membrane transport,
745–746
 Nonpolar amino acids, 70–71
 Nonreceptor tyrosine kinases (NRTKs), 715
 Nonsense codons, 1343
 Nonsense mutations, 1362
 Nonsense suppression, 1362
 Nonshivering thermogenesis, 629,
 860–861
 Nonstandard amino acids, **78–80**, 79F, 80F
 Nonsteroidal anti-inflammatory drugs
 (NSAIDs), 994, 995, 998–999, 999F
 Nonviral retrotransposons, 1246, 1246F
 Noonan syndrome, 721
 Norepinephrine, 673T, **679–680**, 1046
 controls fatty acids in brown fat, 861
 role in fatty acid metabolism regulation, 973
 role in glycogen metabolism regulation, 660
 synthesis from tyrosine, 1058–1060
 Normalized alignment score (NAS),
 196, 198F
 Northern transfer (Northern blot), 113
 Northrop, J., 470
 NO synthase (NOS), 685–687
 Novobiocin, 1167, 1169, 1170, 1175
 NPCs (nuclear pore complexes), 1327
 NPY (neuropeptide Y), 1099
 NPY/AgRP cell type, 1099
 N-Ras, 709
 NRTKs (nonreceptor tyrosine kinases), 715
 NSAIDs, *see* Nonsteroidal anti-
 inflammatory drugs
 nSec1, 433F, 442–444
 NSF (NEM-sensitive fusion) protein, 441,
 444–445, 444F, 445F, 1159
 nt (nucleotides), 1175
 N-terminal nucleophile (Ntn) family,
 1066, 1417
 N-terminus (amino acids), 73, 165,
 1371–1373
 Ntn family, *see* N-terminal nucleophile family
 NTPase, 1275–1276
 NTPs, *see* Nucleoside triphosphates
 Nuclear envelope, 9
 Nuclear factor κB (NF-κB), 1414
 Nuclear localization signal (NLS), 1414
 Nuclear membrane, 7F, 9
 Nuclear Overhauser effect spectroscopy
 (NOESY), 243, 244F
 Nuclear pore complexes (NPCs), 1327
 Nucleases, 101
 Nucleation, 346, 347F
 Nuclei, in landscape theory of protein
 folding, 287–289, 288F
 Nucleic acids, 15, 82–95. *See also*
 DNA; RNA
 and amino acid sequence, 70
 base pairing, 1154–1156
 base stacking, 1156–1157
 as buffer components, 48
 chemical structure, 84–85, 84F
 dietary, 1130
 and DNA function, 85–88
 and DNA structure, 88–95
 double-helical structures, **1145–1148**
 evolution, 33
 functional groups, 43
 hydrophobic forces, 1157–1158
 and nucleotides, 82–84
 oligonucleotide synthesis, 209–213
 organization of, 163–164
 polymeric organization, 17F

- sedimentation coefficients, 154F
stability of, 1151–1158
sugar-phosphate chain conformation, 1152–1154
supercoiled DNA, **1158–1170**
Nucleic acid bases, **82–83**. *See also*
 Nucleotide bases
 Chargaff's rules for, 84, 85
 modification of, 85
 names, structures, and abbreviations, 83T
 tautomeric forms of, 88, 88F
 UV absorbance spectra, 92F
Nucleic Acid Database (NDB), 256T, **257**
Nucleic acid fractionation, **156–159**
Nucleic acid sequencing, **176**, 180–182
 amino acid sequencing contrasted, 184–185
 genome sequencing, 180–184
 Sanger method, 176–180
Nuclei scaffolding, 287
Nucleoid, 4, 4F
Nucleolus, 7F, 9, 11F, 1328
Nucleophiles, 563, 563F, 564
Nucleophilic catalysis, 510–513
Nucleoplasmin, 293
Nucleoporins, 1327
Nucleosidases, 1130
Nucleosides, **82–83**, 83T
Nucleoside-5'-phosphate, 83
Nucleoside diphosphate (NDP), 582, 1125
Nucleoside diphosphate kinases, 582, 644, 1113
Nucleoside monophosphate kinases, 1113
Nucleoside phosphorylases, 1130
Nucleoside triphosphates, 82
Nucleoside triphosphates (NTPs), 95–96, 582
Nucleosomes, 291–293
Nucleotidase, 1130F, 1136F
3'-Nucleotide, 83
5'-Nucleotide, 83
Nucleotides, 15, 18, **82–83**
 defined, 82
 degradation, **1130–1136**
 names, structures, and abbreviations, 83T
 as units, 1175
Nucleotide bases. *See also* Nucleic acid bases
 in aqueous solution, 1156–1157
 ionic interactions, 1158
 in prebiotic conditions, 33
Nucleotide coenzyme biosynthesis, **1136–1139**
Nucleotide excision repair (NER), 1216–1218, 1217F
Nucleotide metabolism, **1107–1140**
 biosynthesis of nucleotide coenzymes, 1136–1139
 degradation of nucleotides, 1130–1136
 deoxyribonucleotide formation, 1119–1130
 purine ribonucleotide synthesis, 1107–1114
 pyrimidine ribonucleotide synthesis, 1114–1119
Nucleotide sugars, 882F
Nucleus, 3, **8–9**. *See also* DNA replication; animal cell, 7F
 metabolic functions, 562T
 plant cell, 11F
 transport of mRNA from, **1326–1327**
NutraSweet (aspartame), 1087
nvCJD (new variant Creutzfeldt-Jakob disease), 315
Nyborg, J., 1380, 1381, 1388
Nylon, 111
- O
O-antigens, 85, **378–379**
Obesity, 629, 862, 1093, 1096–1099, 1104
OB folds, 1213
Obligate aerobes, 6
Obligate anaerobes, 6
O blood type, 22
OB-R protein, 1098
Ochoa, S., 791, 1341, 1342
ochre codon, 1343
ochre mutants, 1408
ochre suppressor, 1362
O^c mutation, 1264
OCRL, 735
Octahedral symmetry, proteins, 268, 268F
Octanoyl-CoA, 948
Oculocerebrorenal dystrophy, 735
ODCase, *see* OMP decarboxylase
Odd-chain fatty acids, 945F, **952–957**
O'Donnell, M., 1182, 1196, 1197, 1222
OEC (oxygen-evolving complex), 916
OE (outside end) sequences, 1237, 1238F
OGDH (2-Oxoglutarate dehydrogenase), 799
O⁺ gene, 1264
Ogston, A., 814
1, 25(OH)2D (1 α , 25-Dihydroxycholecalciferol), 678
Oils, 387, 393–394, 394F
Okazaki, R., 1175
Okazaki fragments, **1174–1176**, 1187, 1219
 in *E. coli* replication, 1193F, 1194, 1198
 eukaryotes, 1207
Oleate, 44F
Oleic acid, 386, 387F, 387T, 950F
(2',5')-Oligoadenylate synthetase (2,5A synthetase), 1400
Oligomers, 267, 270–271, 271F
Oligomycin B, 870
Oligomycin sensitivity conferring protein, *see* OSCP
Oligonucleotides, 209–214, 1402–1403
Oligonucleotide adaptors, 110F
Oligonucleotide probes, 112–113, 113F
Oligopeptides, 70
Oligosaccharides, **359**
 biosynthesis, **880–892**
 N- and *O*-linked, 374F, 379–380, 379F
 in plasma membranes, 400F
 synthesizing, 382–383
Oligosaccharyltransferase, *see* OST
O-Linked glycoproteins, 890–892
O-Linked oligosaccharides, 374F, **379–381**, 380F, 881
 synthesis of glycoproteins from, 890–892
 ω oxidation, fatty acids, 959
 ω protein, 1163. *See also* Topoisomerase I
Omeprazole, 765
2'-*O*'-methyltransferase, 1302
OMP, *see* Orotidine-5'-monophosphate
Omp85, 449
OMP decarboxylase (ODCase), 1115F, 1116–1117
OmpF porin, 403–404, 405F
Oncogenes, 705–711
One gene-one enzyme maxim, 26
One gene-one polypeptide maxim, 26
One-intermediate model, of reversible enzyme reactions, 491
Ontogeny, phylogeny and, 12
Oocytes, mRNA masking and, 1401
opal codon, 1343
opal suppressor, 1362
Oparin, A., 32
Open complex, 1267–1268
- Open conformation:
 of Klentaq1, 1179–1180, 1179F
 of Rep helicase, 1186
Open reading frames (ORFs), 182, **304**
Open β sheets, 251, 254–256
Open systems, 52, 587, 588F
Open-X conformation, 1227F
Operators, 97
Operator constitutive mutation, 1264
Operon, 1262
Opiate receptors, 685
Opsin, 689, 694
Optical activity, **73–78**
Optical density, in Beer–Lambert law, 90
Optical isomers, 75
Optical traps, 1273–1274
Optical tweezers, 1274, 1274F
Oral rehydration therapy, 768
ORC (origin recognition complex), 1205–1207
Orc1, 1205
Orc2, 1205
Orc3, 1205
Orc4, 1205
Orc5, 1205
Orc6, 1205
Order, probability of, 55F
Order-disorder transition, 397, 397F
Ordered Bi Bi reactions, 498–500
Ordered mechanisms, 498
Order of reaction, 483
ORFs, *see* Open reading frames
Organs, 16F, **1090–1095**. *See also*
 specific organs
Organelles, 16F, 562T
Organic arsenicals, 799
Organic compounds, synthesizing, 32F
Organic reaction mechanisms, **562–569**
Organic solvents, 134
Organisms, 16F, 192. *See also* specific organisms
Organ perfusion, 575
oriC locus, 1194–1195, 1194F, 1199F
Orientation effects, enzyme catalysis by, **512–515**
Origin-independent replication restart, 1235
Origin recognition complex, *see* ORC
Ornithine, 80, 80F, 571F, 1028, 1071
Ornithine- δ -aminotransferase, 1070F, 1071
Ornithine decarboxylase, 1408T
Ornithine transcarbamoylase, **1028**
Ornithorhynchus anatinus, 177T
Orotate, 1115F, 1116
Orotate phosphoribosyl transferase, 1115F, 1116
Orotic acid, 1114
Orotic aciduria, 1118–1119
Orotidine-5'-monophosphate (OMP), 1115F, 1116–1117
Orthologous proteins, 194
Orthophosphate (P_i), 578
Orthophosphate cleavage, 582
OSCP (oligomycin sensitivity conferring protein), 855, 870
Osmotic coefficient, 1156, 1157F
Osmotic lysis, 130
OST (oligosaccharyltransferase), 885–886, 885F
Osteoarthritis, 240
Osteoblasts, 540, 677–678
Osteoclastoma, 540
Osteoclasts, 540, 677–678
Osteocytes, 12F
Osteogenesis imperfecta (brittle bone disease), 240

- Osteoglycin, 373T
 Osteomalacia, 677–678
 Osteoporosis, 540, 678–679, **1400**
 Ouabain, 761–762, 761F
 Outer membrane, 6
 Outgroups, 204
 Outside end sequences, *see* OE sequences
 Ovalbumin, 134T, 140F, 1304, 1304F, 1305F
 Ovarian cancer, 1236
 Ovaries, 672F
 Overproducers, 117
 Ovomucoid, 140F
 Oxa1, 448
 Oxalic acid, 49
 Oxaloacetate, 947
 in citric acid cycle, 791, 814
 synthesis of, 872F, 1034, 1067F
 Oxalosuccinate, 789, 809
 Oxidants, 584
 Oxidation, 583
 Oxidation-reduction reactions, 364–365, **565–566, 584–586**
 Oxidative deamination, **1023–1025**, 1340F
 Oxidative phosphorylation, 386, 561F, 845–862
 and ATP synthesis, **852–859**
 control of, **862–863**
 and electron-transport chain, 823, 829, 831–833
 energy coupling hypothesis for, **845–846**
 and glycolysis, 594F
 and other pathways, 1090
 proton gradient generation, **846–852**
 uncoupling, **859–862**, 860F
 Oxidizing agent, 584
 Oxidizing atmosphere, 32
 Oxidoreductases, 479T
 Oxidoreduction reactions, 583
 2,3-Oxidosqualene, 982, 983f
 Oxidosqualene cyclase, 983F, 984F
 2,3-Oxidosqualene cyclase, 982, 983
 2-Oxoacid dehydrogenase family, 799
 2-Oxoglutarate dehydrogenase (OGDH), 799
 Oxonium ion, 520
 5-Oxoprolinase, 1061
 5-Oxoproline, 1062
 Oxyanion hole, 534
 Oxygen:
 adaptation of life to, 34
 partial reduction of, 865
 production of, 914–920, 934–937
 transport of, 56, 323–234, 329
 Oxygenated hemoglobin (oxyHb), 325, 329, 332, 333F–335F
 Oxygenated myoglobin (oxyMb), 325
 Oxygen binding, by hemoglobin, **325–328**
 and BPG, **329–331**, 329F–331F
 cooperativity, 334–340, 351–354
 free energy and saturation curves, 339F
 Oxygen debt, 880
 Oxygen electrode, 830, 830F
 Oxygen-evolving complex (OEC), 916
 Oxygen tension, 325
 oxyHb, *see* Oxygenated hemoglobin
 Oxyluciferin, 183
 oxyMb (oxygenated myoglobin), 325
 Oxytocin, 205, 673T, 683
Oxytricha nova, 1212, 1212F
- P
 P1 plasmids, 1259
 p44 subunit, 1218
 p51, 1208–1209, 1209F
 p53, 1211, 1414
 p66, 1208–1209, 1209F
 p85, PI3K subunit, 732
 p101, PI3K subunit, 733
 p150, PI3K subunit, 733
 P680, 916–917
 P700, 922–923
 P870, 909
 P960, 912
 PA (protective antigen), anthrax, 715
 PA26, 1418–1419, 1418F
 PA28, 1418
 PA700, 1414
 Pääbo, S., 116
 PAB II (poly(A)-binding protein II), 1303
 PABP (poly(A)-binding protein), 1303, 1304F
 Pace, N., 1331
 Pacific Biosciences, 184
 Packing density, 247
 PAF (platelet-activating factor), 1008
 PAGE (polyacrylamide gel electrophoresis), 147–148. *See also* SDS-PAGE
 PAH (phenylalanine hydroxylase), 1043
 Pai, E., 1133
 Pain, prostaglandins and, 993–994
 Pairwise alignment, 196, 201
 PajMo experiment, 1236F, 1262–1263
 Palade, G., 9, 1362
 PALA (*N*-(phosphonacetyl)-L-aspartate), 476–477
 Palindromes, 105, 106
 Palmitate, 44F
 Palmitic acid, 386, 387T, 407
 Palmitoleic acid, 387T
 1-Palmitoleyl-2,3-dioleoyl-glycerol, 940
 1-Palmitoleyl-2-linoleoyl-3-stearoyl-glycerol, 388
 Palmitoyl-ACP, 965
 Palmitoyl thioesterase (TE), 965
 PAM (presequence translocase-associated motor), 447
 PAM-1 matrix, 198
 Pam16, 447
 Pam17, 447
 Pam18, 447
 PAM-250 log odds substitution matrix, **198–199**, 199F, 201, 203
 PAM units (Percentage of Accepted point Mutations), 189, 197–199
 Pancreas, 672F
 Pancreatic δ cells, **675**
 Pancreatic acinar cells, 12F, 525
 Pancreatic α cells, **675**, 973
 Pancreatic β cells, 661, **675**, 973, 1102–1103
 Pancreatic cancer, 1236
 Pancreatic DNase, 157
 Pancreatic DNase I, 1160
 Pancreatic islet hormones, **675**
 Pancreatic phospholipase A₂, 942–943, 942F, 943F
 Pandit, Jayvardhan, 982
 Pangenesis, 19
 Pangenesis, 19
 Pantothenate, 474T, 1138, 1139F
 Pantothenate kinase, 1138, 1139F
 Pantothenic acid, 792
Pan troglodytes, 177T
 PAP (poly(A) polymerase), 1302, 1303F
 Paper chromatography, 132T, 144–145, 144F
 Paper electrophoresis, 146–147, 147F
 PapG-protein, 382
 PAPS (3'-phosphoadenosine-5'-phosphosulfate), 1010, 1072
 Parabiosis, 1096–1099
 Paracrine glands, 671
- Parallel β pleated sheets, 224F, 229–230
 Paralogous proteins, 194
 Parathion, 527
 Parathyroid, 672F
 Parathyroid hormone (PTH), 673T, 677
 Pardee, A., 475, 1262
 Parent DNA, 1174F
 Parker, M., 718
 Parkinson's disease, 309, 720, 1059, 1060
 Parnas, J., 595
 Parodi, A., 883
 Paromomycin, 1395T, 1396F, **1397**, 1397F
 Partial molar free energy, 58
 Partial specific volume, 153
 Partition coefficient, 144, 540, 745–746
 PAS (periodic acid-Schiff's reagent), 411–412
 pas (primosome assembly site), 1190
 Passenger RNAs, 1324
 Passive-mediated transport, 745, **750–752**
 Pastan, I., 1286
 Pasteur, L., 469, 593, 619, 864
 Pasteur effect, 619, 864, 1088
 Patel, D., 1300, 1324
 Path dependence, 53
 Pauling, L., 185–186, 221, 226, 229, 515, 1355
 Pavletich, N., 736, 1412, 1413
 PBG (porphobilinogen), 1048
 PBGS, *see* Porphobilinogen synthase
 P bodies, 1325
 PbRCs (purple photosynthetic bacteria RCs), 909
 PC, *see* Plastocyanin
 PCC, *see* Protein-conducting channel
 P-cluster, 1081
 PCNA (proliferating cell nuclear antigen), 1204, 1204F, 1206
 PCR, *see* Polymerase chain reaction
 PDB, *see* Protein Data Bank
 PDBid, 256–257
 PDC, *see* Pyruvate decarboxylase; Pyruvate dehydrogenase multienzyme complex
 PDGFR (platelet-derived growth factor receptor), 699
 PDI (protein disulfide isomerase), 280, **290–292**, 290F, 291F
 PDK-1 (phosphoinositide-dependent protein kinase-1), 731
 PDZ domain, 708
 PE, *see* Phosphatidylethanolamine
 Pearl, L., 1234
 Peas, 19T, 20–22
 PEG (polyethylene glycol), 1132
 PEG-ADA, 1132
 Pellagra, 474, 1138
 Penetrance, of diseases, 1251
 Penicillin, **377–378**, 377F, 378F, 1395
 Penicillinase, 378
 Penicillinoic acid, 378F
 Pentoses, 82–83, 360F
 Pentose phosphate pathway, **892–898**, 893F, 1090
 PEP, *see* Phosphoenolpyruvate
 PEP carboxykinase (PEPCK), 818, 1089
 in gluconeogenesis, 872–873, 876, 876F, 879
 PEP carboxylase, 1408T
 PEPCK, *see* PEP carboxykinase
 Pepsin, 134T, 169T, 470, 547, 547F
 Pepsinogen, 675
 Peptidases, 79
 Peptides, 70. *See also* Polypeptides
 mass spectrometry of, 172–175, 172F
 in protein sequencing, 169–171
 release of, 1392–1394

- Peptide-*N*⁴-(*N*-acetyl- β -D-glucosaminyl) asparagine 253F, 252F
- Peptide-*N*⁴-(*N*-acetyl- β -D-glucosaminyl) asparagine amidase F, 252
- Peptide bonds, 70, 99F, 222, 1384–1385
- Peptide group, 221–225
- cis*-Peptide group, 222F
- trans*-Peptide group, 222F
- Peptide mapping, 174–175, 175F
- Peptide recognition particle (prp73), 1409
- Peptide units, 70
- Peptidoglycan, 376
- Peptidoglycan *N*-acetylmuramoyhydrolase, 479
- Peptidoleukotrienes, 1000, 1001
- Peptidomimetics, 550–551
- Peptidyl-L-amino acid hydrolase, 470
- Peptidyl homoserine lactone, 170
- Peptidyl prolyl cis-trans isomerases (PPIs), 290, 292
- Peptidyl site, *see* P site
- Peptidyl transferase, 1382, 1383, 1383F, 1385F
- Peptidyl-tRNA, 98F, 1373, 1391
- Percentage of accepted point mutations, *see* PAM units
- Perchloric acid, 46
- per gene, 120F
- Perham, R., 796, 800
- Periodic acid-Schiff's reagent (PAS), 411–412
- Periodic table of elements, 31F
- Peripheral proteins, 400
- Peripheral subunit-binding domain (PSBD), 796, 800, 800F
- Periplasmic space, 117, 378
- PERK (PKR-like endoplasmic reticulum kinase), 1400
- Perlecan, 373T
- Permeability coefficient, 745, 747T, 775
- Permeases, 745
- Permethylolation analysis, 366
- Permissive conditions, 27
- Pernicious anemia, 474T, 956–957, 1064
- Peroxisomal β oxidation, fatty acids, 958
- Peroxisomes, 10, 562T, 935, 958F. *See also* Amino acid metabolism; Glyoxylate pathway
- Peroxisome proliferator-activated receptor γ (PPAR γ), 1104
- Peroxynitrite, 688
- Pertussis toxin (PT), 697, 1398
- Perutz, M., 331, 341, 1252
- Perutz mechanism, 334–340
- Petsko, G., 308
- 6PF-2-K (phosphofructokinase-2), 663–664
- Pfam, *see* Protein families
- Pfanner, N., 445
- PFGE, *see* Pulsed-field gel electrophoresis
- PFK, *see* Phosphofructokinase
- PFK-2 (phosphofructokinase-2), 663–664
- PFK-2/FBPase-2, 663
- Pfu* DNA polymerase, 114
- 2PG, *see* 2-Phosphoglycerate
- 3PG, *see* 3-Phosphoglycerate
- PG (polygalacturonase), 1403
- PGAs, 994
- PGE₂, 997
- PGEs, 994
- PGF _{α} , 994
- PGF _{β} , 994
- PGFs, 994
- PGF_{2 α} , 997
- PGG₂, 997
- PGH₂, 997
- PGH synthase (PGHS), 997–999, 997F, 999F
- PGI, *see* Phosphoglucone isomerase
- PGI₂, *see* Prostacyclin I₂
- PGK, *see* Phosphoglycerate kinase
- P-glycoprotein, 766–767
- PGM, *see* Phosphoglycerate mutase
- PGs, *see* Prostaglandins
- pH, 47–48
- and biochemical standard state, 60
- and buffers, 48
- in disc electrophoresis, 148
- immobilized pH gradients, 150
- in ion exchange chromatography, 136
- and isoelectric focusing, 150, 151
- and Michaelis-Menten kinetics, 496–497
- and protein denaturation, 265
- and protein solubility, 133F, 134
- and protein stabilization, 130–131
- Phages, *see* Bacteriophages
- Pharmacodynamics, 539
- Pharmacogenomics, 545
- Pharmacokinetics, 542
- Pharmacology, 539, 542–545
- Phase variation, 1243, 1243F
- PHAS-I, 1401
- PHAS-II, 1401
- PH (pleckstrin homology) domain, 435, 708, 729
- Phe, *see* Phenylalanine
- pheA operon, 1299T
- Phen (phenetermine), 543
- Phenolphthalein, 51
- Phenotypes, 21–22, 22F
- Phenoaxone ring system, 1272
- Phentermine (phen), 543
- Phentolamine, 680
- Phenylacetic acid, 945
- Phenylaceturic acid, 945
- Phenylalanine:
- aromatic ring flipping and protein core mobility, 308–309
 - biosynthesis, 1075–1078
 - degradation, 570, 570F, 1043–1047
 - as essential amino acid, 1065T
 - genetic codes for, 100T, 1341–1342, 1343T
 - in globular proteins, 246
 - half-life, 1413T
 - α helix/ β sheet propensity, 302T
 - in native unfolded proteins, 283
 - Ramachandran diagram, 225
 - side chains, 70–71, 264T
 - structure and general properties, 68T, 71F
- Phenylalanine hydroxylase (PAH), 1043
- Phenylethanolamine *N*-methyltransferase, 1060F
- 2-Phenylethyl boronic acid, 556
- Phenylisothiocyanate, *see* PITC
- Phenylketonuria (PKU), 569–570, 948, 1045–1047
- β -Phenylpropionate, 527
- Phenylpyruvate, 570, 1045
- Phenylthiocarbamyl (PTC), 165
- Phenylthiohydantoin (PTH), 165
- Pheophytin *a* (Pheo *a*), 916
- Pheromones, 671–672, 976
- PheRS, 1351
- Philadelphia chromosome, 718
- Phillips, D., 252, 517, 519, 521–524
- Phillips, S., 1290, 1291
- PhK, *see* Phosphorylase kinase
- Phlorizin, 768
- PHLPP protein phosphatase, 732
- pH meter, 47
- PhoE porin, 403–404
- Phosphagens, 583
- Phosphatase, 140F
- 5-Phosphatase II, 735
- Phosphate, 578
- Phosphate carrier, 448, 826
- Phosphate compound thermodynamics, 578–583
- Phosphatidic acid phosphatase, 971
- Phosphatidic acids, 389, 389T, 398T
- Phosphatidylcholine, 389T, 398T
- Phosphatidylethanolamide: serine transferase, 1004
- Phosphatidylethanolamine (PE), 389T, 398T, 419
- Phosphatidylglycerol, 389T, 398T, 1004, 1006F
- Phosphatidylglycerol phosphate, 1005
- Phosphatidylinositol (PI), 891F
- Phosphatidylinositol (PtdIns), 389T, 398T, 732, 1004–1006, 1006F
- Phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5-P₂), 435, 665, 725
- Phosphatidylserine, 389T, 398T, 1004, 1005F
- Phosphoinothrinicin, 1068
- 3'-Phosphoadenosine-5'-phosphosulfate, *see* PAPS
- 3'-Phosphoadenosine-5'-phosphosulfate (APS) reductase, 1072
- Phosphoanhydride bond, 578
- Phosphoarginine, 580
- Phosphocholine, 1004
- Phosphocreatine, 570–571, 580, 583, 1092
- (2',5')-Phosphodiesterase, 1400
- Phosphodiesterases, 176, 653, 688–689
- Phosphodiester groups, 84
- Phosphodiester method, of oligonucleotide synthesis, 209, 211
- Phosphoenolpyruvate (PEP), 581F, 596F, 612, 872
- Phosphoenolpyruvate carboxykinase, *see* PEP carboxykinase
- Phosphoenolpyruvate carboxykinase (PEPCK), 571
- Phosphoenolpyruvate-dependent phosphotransferase (PTS) system, 765, 765F
- Phosphoester bond, 578
- Phosphoethanolamine, 1004
- Phosphofructokinase (PFK), 667
- allosteric control, 479
 - binding to erythrocyte membrane, 412
 - in gluconeogenesis, 872
 - in glycolysis, 596F, 600, 625–627, 626F, 627F, 630
 - and integration of pathways, 1088
- Phosphofructokinase-2 (PFK-2), 663–664
- Phosphogluokinase, 642
- Phosphoglucomutase, 632, 639, 642
- 6-Phosphogluconate, 637F, 894, 894F
- 6-Phosphogluconolactonase, 894
- 6-Phosphoglucono- δ -lactone, 894
- Phosphoglucose isomerase (PGI), 596F, 598–600, 599F
- Phosphoglycerate kinase (PGK), 596F, 608–610, 609F
- Phosphoglycerate mutase (PGM), 596F, 610–612, 610F
- 2-Phosphoglycerate (2PG), 594–595, 610
- 3-Phosphoglycerate (3PG), 594–595, 608
- Phosphoglycerides, 389–390
- Phosphoglycohydroxamate, 603
- 2-Phosphoglycolate, 603, 934–935
- Phosphoguanidines, 580, 581F
- 3-Phosphohydroxypyruvate, 1070F, 1071
- Phosphoimager, 111
- Phosphoinositide 3-kinases (PI3Ks), 732–734
- Phosphoinositide 4-kinases (PIP4Ks), 732
- Phosphoinositide 5-kinases (PIP5Ks), 732

- Phosphoinositide cascade, **725–738**
 inositol polyphosphate phosphatases, **734–736**
 phosphoinositide 3-kinases, **732–734**
 phospholipases C, **727–730**
 protein kinases C, **730–732**
 second messengers, **725–727**
 Phosphoinositide-dependent protein kinase-1 (PDK-1), **731**
 Phosphoinositol, **994**
 Phospholipases, **408, 410**
 Phospholipase A₂, **942–943, 942F, 943F, 994, 1017**
 Phospholipase C, **664, 706, 725, 727–730, 728F, 994**
 Phospholipase C- β (PLC- β), **690, 692**
 Phospholipase C- γ , **706**
 Phospholipids, **400F, 410F, 419–420**
 Phospholipid exchange proteins, **420**
 Phospholipid metabolism, **1004**
 glycerophospholipids, **1004–1008**
 sphingoglycolipids, **1008–1013**
 sphingophospholipids, **1008–1009**
 Phospholipid translocases, **420**
 Phosphomannose isomerase, **633**
 Phosphomevalonate kinase, **977**
 N-(Phosphonacetyl)-L-aspartate (PALA), **476–477**
 Phosphopantetheine transferase, **961**
 Phosphopantetheine transferase (PPT), **967**
 Phosphopantothenoylcysteine decarboxylase, **1138, 1139F**
 Phosphopantothenoylcysteine synthetase, **1138, 1139F**
 Phosphopentose epimerase, **929**
 Phosphoprotein phosphatase-1 (PP1), **651, 658–660**
 Phosphoprotein phosphatase-2A, **963**
 Phosphoramidite method, of oligonucleotide synthesis, **210F, 211**
 β -5-Phosphoribosylamine (PRA), **1108, 1109F, 1111**
 5-Phosphoribosyl- α -pyrophosphate (PRPP), **1078, 1108, 1109F, 1114**
 5-Phosphoribosylpyrophosphate synthetase, **1108**
 Phosphoribulokinase, **927**
 Phosphoric acid (H_3PO_4), **49F**
 Phosphorothioate oligonucleotides, **1402**
 Phosphorus, **31**
 Phosphorylase, *see* Glycogen phosphorylase
 Phosphorylase kinase (PhK), **651, 655, 658, 667, 670**
 Phosphoryl groups, **578**
 Phosphoryl group-transfer potentials, **578–579**
 Phosphoryl group-transfer reactions, **565, 565F, 578–579**
 O-Phosphoserine, **79F**
 Phosphotyrosine-binding (PTB) domain, **707**
 Photoautotrophs, **5–6**
 Photooxidation, **905**
 Photophosphorylation, **582, 912–913, 925–926**
 Photoprotective agents, **976**
 Photoreactivating enzymes, **1214–1215**
 Photorespiration, **934–937, 935F**
 Photosynthesis, **5–6, 559, 901**. *See also* Calvin cycle
 carbohydrates product of, **359**
 chloroplasts as site of, **11–12, 901–902**
 dark reactions, **902, 926–937**
 electron transport in purple bacteria, **909–913**
 evolution, **34**
 light absorption, **903–909**
 light reactions, **902–926**
 membrane process, **386**
 photophosphorylation, **925–926**
 photorespiration and C₄ cycle, **934–937**
 two-center electron transport, **913–925**
 Photosynthetic bacteria, **6**
 Photosynthetic reaction center (PRC), **905**
 integral membrane proteins, **403, 404F**
 purple bacteria, **909–911, 909F, 911F**
 Photosystem I (PS I), **913–925, 922F, 923F, 925F**
 Photosystem II (PS II), **913–925, 918F, 925F**
 Phototrophs, **559**
 Phycobiliproteins, **909**
 Phycocyanin, **905F, 909**
 Phycocyanobilin, **909**
 Phycoerythrin, **905F, 909**
 R-Phycoerythrin, **140F**
 Phycoerythrobilin, **909**
 Phylloquinone, **922**
 Phylogenesis, **6**
 Phylogenetic trees, **6, 7F, 13F**
 and bioinformatics, **203–205, 203F**
 and protein sequencing, **189, 190F**
 Phylogeny, eukaryotes, **12–14**
 Phytonic acid, **958**
 Phytonic acid storage syndrome, **959**
 Phytyanoyl-CoA hydroxylase, **958**
 Phytol, **903**
 PI, *see* Phosphatidylinositol
 pI, *see* isoelectric point
 P_i (orthophosphate), **578**
 PI3Ks (phosphoinositide 3-kinases), **732–734**
 PIC (preinitiation complex), **1376–1377**
 Pickart, C., **1417**
 Picket-fence Fe-porphyrin complexes, **328, 328F**
 Picot, D., **920**
 Pigs, **121, 130**
 π Helix, **224F, 228**
 Pili, **4, 4F**
 Pilks, S., **663**
 Ping Pong reactions, **498–499, 499F, 501**
 PIP₂, *see* Phosphatidylinositol-4,5-bisphosphate
 PIP4Ks (phosphoinositide 4-kinases), **732**
 PIP5Ks (phosphoinositide 5-kinases), **732**
 PIR (Protein Information Resource), **195T**
 PISA (Protein Interfaces, Surfaces and Assemblies), **256T**
Pisum sativum, **20**
 PITE (phenylisothiocyanate), **165, 167F**
 Pitch, of helix, **89, 225**
 Pituitary, **672F, 682**
 PK, *see* Pyruvate kinase
 pK, **47**
 of amino acid ionizable groups, **68T–69T, 70**
 and pH, **48**
 of polyprotic acids, **49–50**
 of weak acids, **46T**
 PKA, *see* Protein kinase A
 PKB, *see* Protein kinase B
 PKC, *see* Protein kinase C
 PKR (double-stranded RNA-activated protein kinase), **1400**
 PKR-like endoplasmic reticulum kinase (PERK), **1400**
 PKU, *see* Phenylketonuria
 Placebo, **542**
 Planck's constant, **903**
 Planck's law, **903**
 Plants, **7F, 12, 13F**. *See also* Photosynthesis
 C₃ and C₄, **936**
 cell structure of, **11, 11F**
 divergence from other kingdoms, **192**
 glycogen synthase, **645**
 glyoxylate cycle, **880–881**
 mitochondrial deviations from ideal, **1344T**
 mutation rate, **192**
 transgenic, **122**
 Plantae, **12**
 Plant cells, **11, 11F**
 Plant medicines, **539**
 Plaques, **27**
 Plasma glucose, **1103F**
 Plasmalogens, **390, 574, 1005, 1007–1008, 1007F**
 Plasma membranes, **4**. *See also* Cell membrane
 composition, **398T**
 of eukaryotes, **8, 11F**
 of prokaryotes, **4F**
 schematic diagram, **400F**
 secretory pathway, **420–428**
 sedimentation coefficient, **154F**
 vesicle formation, **428–440**
 vesicle fusion, **440–445**
 Plasmids, **106–108, 1228F, 1262, 1262F**
 Plasmid-based cloning vectors, **106–108**
 Plasmin, **525T**
 Plasminogen, **458**
 Plasmodesma, **11F**
Plasmodium falciparum, **177T, 188, 898, 1058**
 Plastids, **901**
 Plastocyanin (PC), **916, 921–922, 921F**
 Plastoquinol (QH₂), **915**
 Plastoquinone (Q), **915**
 Platelet-activating factor (PAF), **1008**
 Platelet-derived growth factor receptor (PDGFR), **699**
 PLC- β , *see* Phospholipase C- β
 Pleckstrin, **735**
 Pleckstrin homology domain, *see* PH domain
Pleurodeles waltli, **1271F**
 P-loop, **710**
 PLP, *see* Pyridoxal-5'-phosphate
 Pmf (proton-motive force), **846**
 PMP (pyridoxamine-5'-phosphate), **1020**
 Pneumococci, S and R form, **86, 86F**
 PNP, *see* Purine nucleoside phosphorylase
 POH1, **1418**
 Point mutations, **192, 1339–1340**
 Point symmetry, **267**
 pol, **1246**
 Pol α (DNA polymerase α), **1202–1203**
 Pol α /primase, **1203, 1206, 1207**
 Pol β (DNA polymerase β), **1205**
 Pol γ (DNA polymerase γ), **1202, 1205**
 Pol δ (DNA polymerase δ), **1202–1207, 1203F**
 Pol ϵ (DNA polymerase ϵ), **1202, 1205, 1206**
 Pol ζ (DNA polymerase ζ), **1205, 1223**
 Pol η (DNA polymerase η), **1205, 1223**
 Pol ι (DNA polymerase ι), **1205, 1223**
 Pol κ (DNA polymerase κ), **1205, 1223**
 Pol I, *see* DNA polymerase I; RNAP I
 Pol II, *see* DNA polymerase II; RNAP II
 Pol III, *see* DNA polymerase III; RNAP III
 Pol III core, **1182, 1198, 1222**
 Pol III holoenzyme, **1182, 1183F, 1190, 1191**
 in *E. coli* replication, **195, 1193F, 1194**
 in SOS repair, **1222**
 Pol IV, *see* DNA polymerase IV
 Pol V, *see* DNA polymerase V
 Pol V mutasome, **1222**
 POLA, *see* Pol α
 Polarimeter, **74, 74F**
 Polarity, **132, 1398**
 Polar side chains, amino acid, **71**
 POLB, *see* Pol β
 polC gene, **1182T**

- polC* mutants, 1182
POLD1, *see* Pol δ
POLE, *see* Pol ϵ
POLG, *see* Pol γ
POLH, *see* Pol η
POLI, *see* Pol ι
 Poliomyelitis, 1404
POLK, *see* Pol κ
 Polubiquitin, 1404
 Poly(A), 1157F, 1302–1303
 Poly(A)-binding protein, *see* PABP
 Poly(A)-binding protein II (PAB II), 1303
 Polyacrylamide gel, 139, 140T, 147, 147F
 Polyacrylamide gel electrophoresis, *see* PAGE
 Polyadenylation, 1401–1402
 Polyadenylic acid [poly(A)] tail, 97
 Poly(A) polymerase, *see* PAP
 Poly(A) (polyadenylic acid) tail, 97
 Poly(AUAG), 1343
 Polybrene, 171
 Polycistronic mRNA, 1264
 Polycythemia, 343
 Polydispersity, 375
 Polyelectrolytes, 136
 Polyethylene glycol (PEG), 1132
 Polygalacturonase (PG), 1403
 Polyglycine, 205, 224F, 228–229, 229F
 Poly(GUAA), 1343
 Polyketide biosynthesis, 968, 969F
 Polylinkers, 107
 Polylysine, 205
 Polymerase chain reaction (PCR), **114–117**, 115F, 212–213
 Polymerase switching, 1204
 Polymorphic enzymes, 545
 Polynucleotides, 84, 209–213. *See also* Nucleic acids
 Polyomavirus, 94T, 1172
 Polyoxyethylene-*p*-isoctylphenyl ether, 399F
 Polyoxyethylene lauryl ether, 399F
 Polypeptides, 70. *See also* specific types, e.g.: Proteins
 backbone conformations, **222–225**
 chemical synthesis of, **205–209**
 one gene-one polypeptide maxim, 26
 in protein sequencing, 169
 titration curves, 72, 73F
 Polypeptide hormones, 673T
 Polypeptide synthesis. *See also* Protein synthesis
 chain elongation, 1372–1373, 1373F, 1379–1388
 chain initiation, 1373–1379
 chain termination, 1391–1395
 chemical, **205–209**
 direction of, 1371–1373, 1372F
 summary of, 1365
 translational accuracy of, 1388–1391
 Polyproline, 224F
 Polyproline II helix, 228–229, 229F
 Polyproteins, HIV-1, 546
 Polyproteins, of HIV-1, 546F
 Polyprotic acids, **48–50**, 49F
 Polyribonucleotides, 1342
 Polyribosomes (polysomes), 154F, 1372, 1373F
 Polysaccharides, 15, 359, **365–366**. *See also* specific polysaccharides
 carbohydrate analysis, **366–367**
 disaccharides, **367**
 glycosaminoglycans, **370–373**
 glycosidic bonds in, **363–365**
 polymeric organization, 17F
 storage, **369–370**
 structural, **367–369**
 Polyserine, 205
 Polysomes, 154F, 1372, 1373F
 Polystyrene gels, 137
 Polytopic proteins, 402–403
 Poly(UAUC), 1343
 Polyubiquitin (PolyUb), 1411
 Polyunsaturated fatty acids, 387
 Polyvinylidene difluoride, *see* PVDF
POLZ, *see* Pol ζ
 POM/CART cell type, 1099
 Pompe's disease, 666
 Popot, J.-L., 920
 P/O ratio, **831–833**, 833F
 Porcine somatotropin (pST), 119
 Porins, 403–404, 405F, 749–750, 824
 Porphobilinogen (PBG), 1048
 Porphobilinogen deaminase, 1048, 1053
 Porphobilinogen synthase (PBGS), 1048–1051, 1051F
 Porphyrias, 1056
 Porphyrin, 324, 818
 Porters, 745
 Porter, D., 812
 Porter, K., 9
 Position-Specific Iterated BLAST, *see* PSI-BLAST
 Position-specific score matrices, 202
 Positive allosteric effect, 348
 Positive-inside rule, 427
 Positively cooperative binding, 327
 Positive regulator, 1286
 Posttranslational transport, of proteins, 428, 428F
 Posttranscriptional modification, 97–98, 97F, **1301–1331**
 covalent modification, 1405
 mRNA processing, 1302–1327
 protein splicing, 1405–1408
 proteolytic cleavage, 1403–1404
 rRNA processing, 1328–1329
 tRNA processing, 1329–1331
 Posttranslational modification, 101, **1403–1408**
 and renaturation, 280–281
 role in secretory pathway, 422
 Posttranslocational state, 1387–1388, 1387F
 POT1 (protection of telomeres 1), 1213
 Potassium channels, **752–755**, 753F, 754F
 Potassium ion, 45T, 1158
 Potassium ion channels, **752–755**, 753F, 754F
 Potatoes, 19T
 Poulos, T., 686
 Poultier, D., 979
 PP1, *see* Phosphoprotein phosphatase-1
 PPAR γ (peroxisome proliferator-activated receptor γ), 1104
 ppGpp, 1301
 PP_i, *see* Pyrophosphate
 P pili, 382
 PPI, *see* Peptidyl prolyl cis-trans isomerase
 PPM family, 722
 pppA(2'p⁵'A)_n, 1400
 PPP family, 722–725
 pppGpp, 1301
 (p)ppGpp, 1301
 P-protein, 1031
 PPRP, *see* 5-Phosphoribosyl- α -pyrophosphate
 PPT (phosphopantetheine transferase), 967
 PRA, *see* β -5-Phosphoribosylamine
 Prader–Willi syndrome (PWS), 1251
 Pravachol, 990, 991F
 Pravastatin, 990, 991F
 PRC, *see* Photosynthetic reaction center
 Prealbumin, 268, 268F, 269F, 677
 Prebiotic era, 29
 Precipitation, isoelectric, 134
 Precursor-product relationships, 574
 Preinitiation complex (PIC), 1376–1377
 Prelog, V., 76
 Pre-miRNAs, 1325, 1326F
 Pre-mRNAs, 1304–1305, 1312
 Prenylated proteins, 406–407, 407F
 Prenylation site, 406
 Prenyltransferase, 679, 978
 Preparative ultracentrifugation, **155–156**, 155F
 Prephenate, 1075
 Preprimosome, 1190
 Preprocollagen, 1404
 Preproinsulin, 1404
 Preproproteins, 1404
 Preproteins, 422, 1404
 Pre-replicative complex (Pre-RC), 1205–1207, 1206F
 Presequences, 422
 Presequence translocase-associated motor (PAM), 447
 Presqualene pyrophosphate, 979F, 980–981, 981F
 Presynaptic membrane, 440, 440F
 Pretranslocational state, 1387F, 1388
 Pre-tRNA introns, 1307T
 Pre-tRNAs, 1331
 PRF (Protein Research Foundation), 195T
pr gene (fruit fly), 25, 25F
 PriA helicase, 1191
 PriA protein, 1184T, **1185–1186**, 1190, 1190T, 1191
 Pribnow, D., 1266
 Pribnow box, 1266
 PriB protein, 1190, 1190T
 PriC protein, 1190, 1190T
 Priestley, J., 901
 Primaquine, 897–898
 Primary structure, 163, 289–290, 1345–1346. *See also* Nucleic acid sequencing; Protein sequencing
 Primary transcripts, 1301
 Primase, 101, 113F, 1176, **1188–1189**, 1190T
 Primers, 101–102, 103F, 116, 1176
 Primer strand, 1209
 Pri-miRNAs, 1325, 1325F
 Primordial atmosphere, 32
 Primosome, 1190–1193, 1190T
 Primosome assembly site (*pas*), 1190
 Prions, 313
 Prion diseases, **312–316**
 Prion hypothesis, 314–315
 PrionProtein (PrP), 313–315
 Prn-p, 313, 314
 Pro, *see* Proline
 Procarboxypeptidase A, 538
 Procarboxypeptidase B, 538
 Processive enzymes, 1177
 Processivity, of transcription, 1271–1272
 Prochiral atoms, 77
 Prochiral differentiation, 471F
 Prochymosin, 117F
 Procollagen, 1403–1404, 1403F–1405F
 ProCysRS, 1359
 Product inhibition studies, 500
 Proelastase, 538
 Proenzymes (zymogens), **537–538**
 Profiles, sequence alignment, 202
 Proflavin, 1340
 Progeria, 1211
 Progesterone, 681
 Progestins, 673T, 681, 991
 Proinsulin, 280–281, 280F
 Prokaryotae, 6

Prokaryotes, 3–6, 4F. *See also specific organisms*
 chain elongation, 1379–1388
 chain initiation, 1373–1376
 chain termination, 1391–1392, 1392F
 citric acid cycle, 789, 814–815
 DNA replication, 1173, 1190–1201
 DNA size, 94T
 eukaryotic protein production in, 118
 evolution, 13F, 34
 fMet, 1373–1374
 genome sequencing, 176
 introns of, 1316
 mismatch repair, 1220
 mRNA posttranscriptional modification, 1302
 nucleotide excision repair, 1217
 ribosomes of, 1369–1371
 transcription control, 1283–1301
 transformation by DNA, 85–86
 translational initiation, 1375–1376
 transposons of, 1236–1244
 Prokaryotic gene expression:
 mRNA lifetime, 101
 RNAP, 1265
 transcription, 96–97, 101, 1269
 translation, 101
 Proliferating cell nuclear antigen, *see* PCNA
 Proline:
 amino group, 67
 biosynthesis, 1071
 degradation, 1034
 genetic codes for, 100T, 1343T
 α helix/ β sheet propensity, 302T
 low α helix propensity, 304
 in native unfolded proteins, 283
 as nonessential amino acid, 1065T
 in PEST proteins, 1413
 Ramachandran diagram, 225
 side chains, 70, 264T
 specific rotation, 74
 structure and general properties, 68T
 Proline racemase, 516
 Prolyl 3-hydroxylase, 1405
 Prolyl 4-hydroxylase, 1405
 Prolyl hydroxylase, 236
 Promoters, 97, 571, 1266–1267, 1284
 Pronucleus, 121, 121F
 Proofreading, 1201
 DNA chain, 1177
 of synthesized proteins, 1390–1391
 of tRNA, 1355–1358
 1,2-Propanediol, 1018
 β Propeller, 433
 Propeptides, 1403–1404
 Prophase, 20F, 21F
 Prophospholipase A₂, 538
 Propionaldehyde, 1018
 Propionic acid, 32T
 Propionyl-CoA carboxylase, 952–953
 Propionyl-CoA carboxylase reaction, 953F
 Proportional counting, 572
 Propranolol, 680
 Proproteins, 1403–1404
pro-R, 77, 77F
pro-S, 77, 77F
 Prostacyclins, 993, 995–1000
 Prostacyclin I₂ (PGI₂), 997, 998
 Prostacyclin receptors, 689
 Prostacyclin synthase, 998
 Prostaglandins (PGs), 971, 993–995, 994F, 995F, 997–1000
 Prostaglandin endoperoxide synthase, 997
 Prostaglandin H₂ synthase, 406, 997
 Prostaglandin receptors, 689

Prostanoic acid, 993–994
 Prostate cancer, 1236
 Prosthetic group, 473
 Proteases, 101, 131, 1419–1420
 Protease A, 525T
 Protease B, 525T
 20S Proteasome, 1411F, 1415, 1415F
 26S Proteasome, 1411–1412, 1414
 Proteasomes, in ubiquitination, 1411–1412
 Protection of telomeres 1 (POT1), 1213
 Protective antigen (PA), anthrax, 715
 Proteins, 15. *See also related topics, e.g.:*
 Amino acids
 amino acid derivatives in, 78–79, 80F
 amino acids of, 67–73, 68T–69T
 beta structures, 229–232
 in biological processes, 67
 and buffers, 48
 chemical evolution of, 189–192, 191F
 and chemical synthesis of polypeptides, 206–209
 in chromatin, 8
 conformational diseases, 309–316
 crystal, 122
 cyclic symmetry, 268, 268F
 denatured, 57
 design of, 305–306
 dihedral symmetry, 268, 268F
 essential amino acids in, 1064
 evolution of, 192–196, 196F, 316–318
 fibrous, *see* Fibrous proteins
 functional groups of, 43
 fusion, 118
 globular, *see* Globular proteins
 helical structures, 225–229
 helical symmetry, 269, 269F
 homologous, 189–192
 icosahedral symmetry, 268, 268F
 from inclusion bodies, 117
 and molecular cloning, 117–121
 motifs, 249–251
 number of, 70
 octahedral (cubic) symmetry, 268, 268F
 one gene–one polypeptide maxim, 26
 organization of, 17F, 163–164, 281–282, 281F
 physical constants of, 153T
 polyproteins, 1404
 posttranslational modification and degradation of, 101
 primary structure, *see* Protein sequencing
 pseudosymmetric, 268
 for purification, 130
 quaternary structure, 266–271, 267F
 roles of, 163
 secondary structure, 221–233, 302–304
 solubilities of, 133–135
 structures of, 230–233
 subunit interactions, 267
 supersecondary structures, 249–251
 symmetry, 267–270
 tertiary structure, 245–256, 304–305
 tetrahedral symmetry, 268, 268F
 titration curves, 72, 72F
 in translation, 98–101
 ubiquitin attachment to, 1410F
 α_1 -Proteinase inhibitor, 533
 Proteinase K, 1383
 Protein assays, 131–132
 Protein C, 1317
 Protein-conducting channel (PCC), 424, 426
 Protein crystals, 134F
 Protein Data Bank (PDB), 256–257, 256T, 259
 Protein degradation, 1408–1411. *See also* Polyubiquitin; Ubiquitin
 Protein denaturation, 57, 221, 262, 265–266
 Protein disulfide isomerase, *see* PDI
 Protein dynamics, 306–309
 Protein families (Pfam), 256T, 258
 Protein folding, 537
 accessory proteins, 290–302
 determinants, 281–284
 endoplasmic reticulum, 427–428
 GroEL/ES system, 294–302
 hierarchical nature of, 289
 landscape theory of, 287–289, 288F
 pathways, 284–290
 peptidyl prolyl cis-trans isomerases, 292
 protein disulfide isomerase, 280, 290–292, 290F, 291F
 and pulsed H/D exchange, 285–286
 and renaturation, 278–281
 Protein folding funnels, 287–289, 288F
 Protein folding problem, 278
 Protein G, 282
 Protein GB1, 282, 282F
 Protein growth factors, 671
 Protein Information Resource (PIR), 195T
 Protein Interfaces, Surfaces and Assemblies (PISA), 256T
 Protein isolation, 129–133
 Protein kinases, 670, 973, 975
 Protein kinase A (PKA), 651–654
 Protein kinase B (PKB), 734, 737F
 Protein kinase C (PKC), 660, 698, 713, 727, 730–732
 Protein metabolism, 561F
 Protein phosphatases, 721–725
 Protein phosphatase-2A, 699, 723F
 Protein purification, 129–156
 chromatographic separations, 135–146
 electrophoresis, 146–152
 general strategy, 132–133
 and protein isolation, 129–133
 and protein solubilities, 133–135
 ultracentrifugation, 152–156
 Protein renaturation, 278–281
 Protein Research Foundation (PRF), 195T
 Protein sequence databanks, 195T
 Protein sequencing, 164–165
 C-terminus identification, 165–168
 disulfide bond cleavage, 168–169
 disulfide bond position assignment, 172
 Edman degradation, 165, 167F, 171, 172, 174
 end group analysis, 165–168
 N-terminus identification, 165
 nucleic acid sequencing contrasted, 184–185
 peptide characterization and sequencing by mass spectrometry, 172–174, 172F
 peptide cleavage reactions, 169–171
 peptide fragments, ordering, 171
 peptide fragments, separation and purification of, 171
 peptide mapping, 174–175, 175F
 separation, purification, and characterization of polypeptide chains, 169
 sequence alignment, 194–203
 sequence databases, 194–195
 sequence determination, 171, 171F
 Protein Ser/Thr phosphatases, 722
 Protein solubilization, 130
 Protein splicing, 1405–1408
 Protein stability, 130–131
 and denaturation, 265–266
 and disulfide bonds, 264–265
 electrostatic forces, 259–261
 hydrogen bonding, 261–262
 hydrophobic forces, 262–264
 thermostable proteins, 266

- Protein synthesis, 1260, 1362–1398. *See also related topics, e.g.: Translation and antibiotics, 1395–1398 chain elongation, 1372–1373, 1379–1388 chain initiation, 1373–1379 chain termination, 1391–1395 direction of, 1371–1373 secretory pathway, 420–428 translational accuracy, 1388–1391 vesicle formation, 428–440 vesicle fusion, 440–445*
- Protein tyrosine kinases (PTKs), 690, 699–703
- Protein tyrosine phosphatases (PTPs), 721
- Proteoglycans, 373–375, 373T
- Proteoglycan subunit, 373, 375
- Proteolysis, limited, 170
- Proteome, 152, 185, 258
- Proteomics, 152, 576–578, 577F
- Proteopedia, 256T, 257
- Protista, 12, 13F
- Protofilaments, keratin, 234F
- Protomers, 267
- Protons, 45–46, 45F, 45T
- Proton-motive force (pmf), 846
- Proton pump mechanism, 846–847, 850, 850F
- Proton-translocating ATP synthase, 852–859
- Proton-translocating ATP synthase (Complex V), *see* F_1F_0 -ATPase
- Proton wire, 800
- Proto-oncogenes, 705
- Protoplasts, 378
- Protoporphyrin IX, 316, 316F, 324 biosynthesis, 1053–1055 and electron-transport chain, 838
- Protoporphyrinogen IX, 1054
- Protoporphyrinogen oxidase, 1053–1055
- Protosterol, 982
- Protozoa, 13F, 185, 192, 1344T. *See also specific organisms*
- Pro-tRNA^{Pro}, 1359
- Proximal histidine, in hemoglobin, 332
- Proximity effects, enzyme catalysis by, 512–515
- PrP (PrionProtein), 313–315
- PrP 27–30, 315
- Prp73 (peptide recognition particle), 1409
- PrP^C, 314–315
- PRPP, *see* 5-Phosphoribosyl- α -pyrophosphate
- PrP^{Sc}, 314–315
- Prusiner, S., 312
- PS I, *see* Photosystem I
- PS II, *see* Photosystem II
- PsbB, of photosystem, 916
- PsbC, of photosystem, 916
- PSBD, *see* Peripheral subunit-binding domain
- Pseudoglobulin, 133F
- Pseudorotation, 591
- Pseudosymmetric proteins, 268
- Pseudouridine, 1219, 1346, 1347F
- [PSI], 315
- PSI-BLAST (Position-Specific Iterated BLAST), 202–203, 304
- D-Psicose, 361F
- P site (peptidyl site), 98F, 1373, 1385, 1387
- Psoriasis, 993
- PST (porcine somatotropin), 119
- PstI, 105T
- PT, *see* Pertussis toxin
- PTB (phosphotyrosine-binding) domain, 707
- PTC (phenylthiocarbamyl), 165
- PtdIns, *see* Phosphatidylinositol
- PtdIns-4,5-P₂, *see* Phosphatidylinositol-4,5-bisphosphate
- PTEN, 736
- Pteridine, 1043
- Pterins, 1043–1045
- Pterin-4a-carbinolamine, 1044
- Pterin-4a-carbinolamine dehydratase, 1044
- PTH, *see* Parathyroid hormone; Phenylthiohydantoin
- PTKs, *see* Protein tyrosine kinases
- PTPs (protein tyrosine phosphatases), 721
- PTS, *see* Phosphoenolpyruvate-dependent phosphotransferase system
- pUC18 cloning vector, 107, 107F
- Puckers, sugar-phosphate chain, 1152–1153, 1153F
- Pulmonary emphysema, 533
- Pulse-chase experiment, 1175
- Pulsed-field gel electrophoresis (PFGE), 158–159, 159F
- Pulsed H/D exchange, 284–286, 309
- Pulse labeling, 1175F
- Pumped protons, 843
- PurB, *see* Adenylosuccinate lyase
- PurC, *see* SAICAR synthetase
- PurE, 1111
- PurF, *see* Glutamine PRPP aminotransferase
- PurH, *see* AICAR transformylase
- Purification tables, 161
- Purine, 83
- Purine nucleoside phosphorylase (PNP), 1130F, 1131
- Purine nucleotide cycle, 1132, 1133F
- Purine repressor (PurR), 1296
- Purine ribonucleotides, 1107–1108. *See also Adenine; Guanine AMP synthesis, 1111–1113, 1113F catabolism, 1130–1134 GMP synthesis, 1111–1113, 1113F IMP synthesis, 1108–1111, 1113F orientations of, 1152F and purine ring atoms, 1107F salvage, 1114 synthesis regulation, 1113–1114*
- Purity, 132
- PurJ, *see* IMP cyclohydrolase
- PurK, 1111
- PurL, *see* FGAM synthetase
- PurM, *see* AIR synthetase
- PurN, *see* GAR transformylase
- Puromycin, 424, 1376, 1382, 1382F, 1395, 1395T
- Purple membrane, 402
- Purple photosynthetic bacteria, 6, 7F, 13F, 403 electron transport in, 909–913, 913F light-harvesting complexes, 906–909
- PVDF (polyvinylidene difluoride), 111, 171
- PvuII, 105T
- PWS (Prader–Willi syndrome), 1251
- Pycnodynatosi, 540
- Pyl (pyrrolysine), 1361–1362
- PylRS, 1362
- Pyran, 361
- Pyranoses, 361
- Pyridine nucleotides, 471F
- α -Pyridone, 508
- Pyridoxal-5'-phosphate (PLP), 511, 640, 1020
- Pyridoxal phosphate, 473T
- Pyridoxamine-5'-phosphate (PMP), 1020
- Pyridoxine, 474T, 1020
- Pyrimidine, 83, 475–479
- Pyrimidine dimers, 1214–1215, 1214F
- Pyrimidine ribonucleotides, 736–738, 1136. *See also Cytosine; Thymine; Uracil CTP synthesis, 1118–1119*
- sterically allowed orientations, 1152F
- synthesis regulation, 1118
- UMP synthesis, 1114–1118
- UTP synthesis, 1118–1119
- Pyrobaculum aerophilum*, 1315
- Pyroccocus furiosus*, 114
- Pyrolobus fumarii*, 266
- Pyrrolidine, 1361–1362
- Pyrrophosphate (PP_i), 102F, 578, 1265
- Pyrrophosphate cleavage, 582, 582F
- Pyrrophosphate ion, 96, 96F
- 5-Pyrophosphatevalonate, 977
- Pyrophosphatevalonate decarboxylase, 977–978, 978
- Pyrosequencing, 183, 184F
- Pyrrole-2-carboxylate, 516
- Pyrrole rings, 324
- Δ -1-Pyrroline-2-carboxylate, 516
- Δ -1-Pyrroline-5-carboxylate, 1070F, 1071
- Pyrroline-5-carboxylate reductase, 1070F, 1071
- Pyrrolysine (Pyl), 1361–1362
- Pyruvamide, 617
- Pyruvate: alanine biosynthesis from, 1067F biosynthesis of, 561F in citric acid cycle, 766F, 790–792, 790F, 794–795, 816 degradation of amino acids to, 1030–1034 in Entner–Doudoroff pathway, 637F fermentation of, 594F, 614–619 in glycolysis, 593, 594F, 595, 596F and integration of pathways, 1090 and oxaloacetate, 872–876, 872F
- Pyruvate carboxylase, 818–819, 872–876, 874F, 875F
- Pyruvate decarboxylase (PDC), 616, 617F
- Pyruvate dehydrogenase (E1), 794–795, 798F, 799–800, 800F
- Pyruvate dehydrogenase kinase, 799, 804–805, 805F
- Pyruvate dehydrogenase multienzyme complex (PDC), 792–800, 1111 coenzymes and prosthetic groups, 795T reactions of, 794F regulation, 804–805, 805F
- Pyruvate dehydrogenase phosphatase, 799, 805
- Pyruvate:ferredoxin oxidoreductase, 814–815
- Pyruvate kinase (PK), 581–582, 1089 in gluconeogenesis, 872 in glycolysis, 596F, 613–614, 614F
- Pyruvate-phosphate dikinase, 936
- Pythagoras, 897
- PYY_{3–36}, 1099
- Q
- Q (plastoquinone), 915
- Q cycle, 848–849, 848F
- QFR (quinol–fumarate reductase), 840
- QH₂ (plastoquinol), 915
- QSAR, *see* Quantitative structure-activity relationships
- Q-SNAREs (SNAP receptors), 441, 442, 779, 780
- Quantitative structure-activity relationships (QSAR), 540, 540F
- Quantum yield, 912
- Quasi-equivalent subunits, 269
- Quaternary structure, proteins, 164, 266–271, 267F
- Questran, 990
- Quiescence, 1202
- Quinine, 539, 1058
- Quinolate, 1137F, 1138
- Quinolate phosphoribosyltransferase, 1137F, 1138

- Quinol-fumarate reductase (QFR), 840
 Quinolone antibiotics, 1167, 1169–1170
 Quiocho, F., 1131
- R**
 R5P, *see* Ribose-5-phosphate
 Rab5, 734
 Rab-GAP, 441
 Rab-GEF, 441
 Rab proteins, 406, 441–444
 Rac1, 727–728, 728F
 Racemic mixtures, 78
 Racker, E., 852, 860, 892
 Rack mechanism, 515
 RACKs, 730
 Rad50, 1224
 RAD51, 1230, 1236
 Radioactive counting, 149
 Radioactive decay, 572
 Radioimmunoassays, 130, 674
 Raf, 712–713
 Raffinose, 385
 Rafts, 411
 Ramachandran, G. N., 223
 Ramachandran diagram, 223–226, 224F, 229
 Ramakrishnan, V., 1365, 1389, 1392, 1394
 Ran, 709
 Randle, P., 864
 Randle cycle, 864
 Random Bi Bi reactions, 498–500
 Random coil conformation, 90, 155, 230, 232
 Random mechanism, 498
 Rao, Z., 838
 RAP1, 1213
 Rap1A, 713
 Rapamycin, 734, 737F
 Rapid equilibrium random Bi Bi reactions, 499, 500
 Rapoport, T., 424, 620
 RasGAP, 711
 Ras proteins, 407, 434, 688, 708–709
 Rats:
 carcinogenesis tests, 1224
 chromosome number, 19T
 genome sequencing, 177T
 liver cytoplasmic ribosomes, 1370T
 liver enzymes, 1408T
 liver membrane composition, 398T
 proteins for purification from, 130
 Rat1 exonuclease, 1304
 Rate constant, 483
 Rate-determining step, 487
 Rate equations, 483
 Ratner, S., 1025
 Rayment, I., 613, 874, 1238
 Rb, 1211
 RB69 pol, 1223
 RBD, *see* RNA-binding domain
 RBP1, 1318
 Rbx1, 1412–1413, 1413F
 RdRP (RNA-dependent RNA polymerase), 1324
 Reaction coordinate diagram, 485–486, 485F
 Reaction kinetics, *see* Enzyme kinetics; Kinetics
 Reaction rates, non-enzymatic reactions, 483–484
 Reactions, coupled, 60–61
 Reactive oxygen species (ROS), 325, 840, 865
 Reading frames, 101, 101F, 1342F
 Reads, 181
 Rearrangement reactions, 567
 RecA, 1221, 1222, 1239
 and recombination in *E. coli*, 1228–1230, 1229F, 1230F
 and recombination repair, 1234
 recA gene, 1221
 RecBCD, 1230–1231, 1230F, 1231F, 1235
 recB gene, 1230
 recC gene, 1230
 recD gene, 1230
 Receptors, 660
 hormone binding, 674–675
 target of drug design, 539
 Receptor-mediated endocytosis, 454
 Receptor tyrosine kinases (RTKs), 293, 699–703
 Recessive traits, 20–22
 RecG, 1234
 RecJ, 1220
 Recognition helix, 1288
 Recognition sites, Type II restriction enzymes, 105, 105T
 Recombinant DNA, 104, 109F, 110F, 621. *See also* Molecular cloning
 Recombinant proteins, 119
 Recombinases, 1239
 Recombination, 24, 1225–1246
 chromosomal rearrangement by, 1242F
 homologous, 1225–1236, 1227F
 site-specific, 1239–1240, 1242–1246
 transposition, 1236–1246
 viruses, 27–28, 28F
 Recombination repair, 1222, 1234–1236, 1234F, 1235F
 Recommended names, enzymes, 479
 Red algae, 13F
 Redox loop mechanism, 847–849, 847F
 Redox reactions, 583. *See also* Oxidation-reduction reactions
 Red tide, 777, 939
 Reducing agent, 584
 Reducing atmosphere, 32
 Reducing sugars, 364
 Reductant, 584
 Reduction, 583. *See also* Oxidation-reduction reactions
 Reduction potentials, 585, 585T
 Reductive pentose phosphate cycle, 927. *See also* Calvin cycle
 Reed, G., 613
 Reed, L., 792, 798
 Rees, D., 910, 1080
 re face, 77, 77F
 Refsum's disease, 958
 REG, 1418–1419
 REG α , 1418–1419
 REG β , 1418
 Reichard, P., 1119, 1122
 Reichert, E., 323
 RelA (stringent factor), 1301
 relA gene, 1301
 Relatedness odds, 198
 Relatedness odds matrix, 198
 Related-to-ubiquitin-1 (RUB1), 1421
 Relative elution volume, in gel filtration chromatography, 138
 Relaxed circles, 1160
 Relaxed control, 107, 1301
 Release factors, 682, 1375T. *See also* RF-1; RF-2; RF-3
 Release-inhibiting factors, 682
 Remington, J., 766, 806, 807
 Remnant receptor, 454
 Renaturation:
 of DNA, 93, 93F
 of proteins, *see* Protein renaturation
 Renin, 547, 547F
 Reperfusion injury, 325
 Repetitive sequences, 181
 Repetitive sequences, ultracentrifugation of nucleic acids, 159
 Rep helicase, 1183–1186, 1191–1193
 Replica DNAs, 1174F
 Replica plating, 26, 26F
 Replication, *see* DNA replication
 Replication bubbles, 1173–1174, 1173F
 Replication eyes, 1173–1174, 1173F, 1176F
 Replication factor C, *see* RFC
 Replication forks, 102–103, 1173–1174
 in *E. coli* replication, 1199
 eukaryotes, 1206
 in recombination repair, 1234–1235, 1234F, 1235F
 Replication fork collapse, 1222
 Replication origin, 106, 1174, 1205–1206
 Replication protein A (RPA), 1206
 Replicative forms, *see* RF; RFI; RFII
 Replicative transposition, 1237, 1239, 1239F
 Replicative transposons, 1242
 Replicons, 1205, 1226
 Replisome, 1193F, 1194, 1196, 1196F
 Reporter genes, 120–121, 120F, 571
 Rep protein, 1184T
 Repressors, 96
 RER, *see* Rough endoplasmic reticulum
 Research articles, reading of, 35–36
 Resequencing, 184
 Resins, 137–138, 207–208
 Resistin, 1104
 Resolvase, 1239–1241, 1241F
 Resonance energy transfer, 905
 Respiration, 9, 34, 323
 Respiratory chain, 823
 Respiratory distress syndrome, 1005
 Restart primosome, 1235
 Restriction endonucleases, 104–106, 105F
 digests from, 106F
 for gene manipulation, 109
 use in nucleic acid sequencing, 176
 Restriction enzymes, 104, 105T
 Restriction-fragment length polymorphisms (RFLPs), 106, 106F, 107F
 Restriction maps, 107F
 Restriction-modification systems, 104
 Restrictive conditions, 27
 Retaining β -glycosidases, 524
 Reticulocytes, 111–112, 1056
 Retinal, 123, 402, 402F, 689, 850–851, 976
 Retinol, 252
 Retinol binding protein, 252, 252F, 310
 Retrograde transport, 428
 Retrotranslocation, 428
 Retrotransposons, 1244–1246
 Retroviruses, 116, 545, 1207, 1245–1246, 1245F
 Reverse genetics, 185
 Reverse gyrase, 1163
 Reverse-phase chromatography (RPC), 132T, 145, 171, 172
 Reverse transcriptase, 95F, 116, 545, 1207–1209, 1208F
 Reverse transcriptase inhibitors, 546
 Reverse translation, 313
 Reverse turns, 232, 233F, 246, 304
 Reversible processes, 56–57
 Reversible reactions, 491–492, 501–502
 Reznikoff, W. S., 1237, 1238
 RF (replicative form), 1190, 1266
 RF-1 (release factor), 1375T, 1391–1394, 1393F
 RF-2 (release factor), 1375T, 1391–1394
 RF-3 (release factor), 1375T, 1392

- RFC (replication factor C), 1204–1205, 1204F
 RF I (replicative form), 1190, 1191, 1191F
 RF II (replicative form), 1190
 RFLPs, *see* Restriction-fragment length polymorphisms
 R form (*RS* system), 76–77
 R Groups, amino acids, 70
 RGS proteins, 692
rh1 gene, 1243
 L-Rhamnose, 365
 Rh (Rhesus) blood group system, 414–415
 Rheumatoid arthritis, 1409
Rhizobium, 1080F
 Rho13N, 1276F
 Rhodopsin, 163, 689
 Rho factor, 1275–1276, 1276F
 Rho proteins, 688, 709, 1365
 Ribitol, 364
 Riboflavin, 474T, 565F, 1138
 Ribonuclease A, *see* RNase A
 Ribonuclease B, *see* RNase B
 Ribonuclease T1, 176
 Ribonucleic acid, *see* RNA
 Ribonucleoproteins, 1145
 Ribonucleotides, 82, 82F
 Ribonucleotide reductase (RNR), 844, 1119–1126
 D-Ribose, 19, 82, 360F, 361
 Ribose phosphate isomerase, 929
 Ribose phosphate pyrophosphokinase, 1108, 1109F, 1111
 Ribose-5-phosphate (R5P):
 in AMP/GMP synthesis, 1112F
 in Calvin cycle, 929
 in IMP synthesis, 1108
 in pentose phosphate cycle, 892–894
 Ribosomal recycling factor, *see* RRF
 Ribosomal RNA, *see* rRNA
 5S Ribosomal RNA, 176
 70S Ribosome, 1363T, 1369F, 1375
 Ribosomes, 1145, 1338. *See also* Protein synthesis
 architecture, 1366–1368
 binding sites, 1368
 of *E. coli*, 367F, 1363F–1366F
 of eukaryotes, 7F, 9, 1369–1371
 of prokaryotes, 4, 4F
 reading direction in, 1342–1343, 1372
 as ribozyme, 1383–1384
 RNA synthesis regulation, 1301
 sedimentation coefficient, 154F
 self-assembly of, 1365
 structure, 1363–1371
 synthesis in secretory pathway, 421F
 translation by, 98–101, 1363–1398
 vesicle formation, 428–440
 vesicle fusion, 440–445
 Riboswitches, 1299–1301, 1300F, 1318–1319, 1320F
 Ribothymidine (T), 1347F
 Ribozymes, 82, 1078
 and Group I introns, 1308–1309
 hammerhead, 1310–1311, 1311F
 ribosomes as, 1383–1384
 D-Ribulose, 359, 361F
 Ribulose-5-phosphate epimerase, 894, 895F
 Ribulose-5-phosphate isomerase, 894, 895F
 Ribulose-5-phosphate (Ru5P), 892–895, 895F, 927
 Ribulose bisphosphate carboxylase activase, 931, 933
 Ribulose-1,5-bisphosphate carboxylase-oxigenase, *see* RuBisCO
 Ribulose bisphosphate carboxylase (RuBP carboxylase), 929–931, 930F
 Ribulose-1,5-bisphosphate (RuBP), 927
 Rice, 122, 122F, 177T
 Rich, A., 236, 1149–1151, 1155, 1323, 1346, 1372
 Richter, J., 1402
 Ricin, 1395T
 Rickets, 677
Rickettsia, 4F
Rickettsia prowazekii, 177T
 Rieske, J., 835
 Rieske iron–sulfur proteins, 835
 Rifampcin, 1176, 1269
 Rifamycin B, 1269, 1395
 Right-handed helix, 224F, 225F, 227F
rIIA complementation group, 28
rIIB cistron, 1340
rIIB complementation group, 28
 Rilling, H., 979
 Ringe, D., 308, 434
 RING finger, 1410
 RISC (RNA-induced silencing complex), 1324
 Risedronate, 679F
 Ritonavir, 550F, 551
 Rittenberg, D., 574, 961, 1048
 RK (HMG-CoA reductase kinase), 989
 RMM (RNA-recognition motif), 1365
 RNA-11, 1151
 16S RNA, 159
 23S RNA, 1329F
 45S RNA, 1328
 RNA (ribonucleic acid), 4, 1145. *See also related topics, e.g.:* Translation and central dogma of molecular biology, 95, 1260
 chemical structure, 85
 on DNA chips, 213
 in DNA replication, 101–102
 and early evolution, 33–34
 fractionation, 156–159
 hybridization, 93
 messenger, *see* mRNA [messenger RNA]
 metabolite-sensing, 1299–1301
 modes of growth for, 1270F
 Northern blot for, 113
 PCR amplification, 116
 ribosomal, *see* rRNA [ribosomal RNA]
 synthesis of, 95–98
 transcript editing, 185
 in transcription, 1260–1265
 transfer, *see* tRNA [transfer RNA]
 translation of, 98–101
 zonal ultracentrifugation, 159
 RNAase D, 1328
 RNAase M5, 1328
 RNAase M16, 1328
 RNAase M23, 1328
 RNA-binding domain (RBD), 1303. *See also RRM* (RNA recognition motif)
 RNA-dependent RNA ATPases, 1313F
 RNA-dependent RNA helicases, 1313F
 RNA-dependent RNA polymerase (RdRP), 1324
 RNA-directed DNA polymerase, 95F, 116, 1207–1208. *See also* Reverse transcriptase
 RNA-directed RNA polymerase, 95F
 RNA–DNA hybrid helix, 96, 1151, 1151F, 1207
 RNA editing, 1320, 1322–1323
 RNAi, *see* RNA interference
 RNA-induced silencing complex (RISC), 1324
 RNA interference (RNAi), 1145, 1323–1326, 1323F
 RNA ligase, 1322
 RNAP (RNA polymerase), 96F, 1265–1283
 and AraC, 1292–1294
 of bacteriophage M13, 1190
 in catabolite repression, 1286
 chain elongation, 1270–1272, 1270F
 chain initiation, 1267–1269
 chain termination, 1271F, 1273–1275
 eukaryotic, 1276–1283, 1277T
 and intercalating agents, 1272
 and *lac* repressor, 1285
 in replication, 1176
 RNA-directed, 95F
 and stringent response, 1301
 structure, 1268, 1268F–1269F
 template binding, 1266–1267, 1266F
 of *Thermus aquaticus*, 1268, 1269F
 of *Thermus thermophilus*, 1268, 1268F–1269F
 in transcription, 95–96, 1265–1283
 transcription cycle and translocation mechanism of, 1280F
 RNAP A, *see* RNAP I
 RNAP B, *see* RNAP II
 RNAP C, *see* RNAP III
 RNAP core enzyme, 1265, 1267, 1269
 RNAP holoenzyme, 1265–1266
 chain elongation, 1270–1272, 1270F, 1271F
 chain initiation, 1267–1269
 chain termination, 1271F, 1273–1275
 and *lac* repressor, 1285
 and template binding, 1266–1267, 1266F
 RNAP I (RNA polymerase I) (Pol I), 1276, 1280–1282, 1408T
 RNAP II (RNA polymerase II) (Pol II), 1276–1283
 and amatoxins, 1280
 error correction by, 1281, 1281F
 and mRNA capping, 1302
 and splicing factors, 1314
 structure of yeast, 1277–1280, 1278F
 RNAP II elongation complex, 1279F
 RNAP III (RNA polymerase III) (Pol III), 1280, 1281, 1283
 RNA polymerase, *see* RNAP
 RNA polymerase I, *see* RNAP I
 RNA polymerase II, *see* RNAP II
 RNA polymerase III, *see* RNAP III
 RNA primers, 101–102, 103F, 1176
 of bacteriophage M13, 1190
 in *E. coli* replication, 1193F
 eukaryotic, 1207
 Pol I excises, 1181, 1187
 RNA-recognition motif, *see* RRM
 RNase, for nucleic acid fractionation, 157
 RNase III, 1324, 1328
 RNase A (ribonuclease A), 63, 176, 470
 acid–base catalytic mechanism, 508–510, 509F
 denaturation of, 265F, 266
 folding rate, 284, 381
 isoelectric point, 134T
 physical constants, 153T
 renaturation, 278–281, 279F
 state of, 308
 titration curve, 73F
 RNase B (ribonuclease B), 380–381, 380F
 RNase E, 1328
 RNase F, 1328
 RNase H, 1151, 1207–1209
 RNase H1, 1207
 RNase L, 1400
 RNase P, 1328
 RNA sequencing, 176, 179
 RNA triphosphatase, 1302
 RNA world, 33–34, 1126
 RNR, *see* Ribonucleotide reductase

- Roach, P., 646
 Roberts, R., 1248, 1304
 Robison, R., 595
 Rodbell, M., 690
 Rodnina, M., 1384
 Rofecoxib, 999–1000
 Rolling circle replication, 1191–1193, 1192F
 Rooted trees, 203F, 204
 Rop protein, 282–283, 283F
 ROS, *see* Reactive oxygen species
 Rosbash, M., 1306
 Rose, G., 262, 281, 282, 304
 Rose, I., 1119
 Roseman, S., 765, 766
 Rosetta algorithm, 305, 306
 Rossman, M., 256, 615
 Rossman fold, 256
 Rotamases, 292
 Rotamers, 306
 Rotation angles, 222–225
 Rotation symmetry, protein, 268
 Rotenone, 831
 Rothman, J., 419, 442, 888
 Rough endoplasmic reticulum (RER), 7F, 9
 metabolic functions, 562T
 secretory pathway in, 420–424
 vesicle formation, 428–440
 vesicle fusion, 440–445
 Round-up, 1075
 Roundworms, trans-splicing in, 1320
 Rous, P., 704
 Rous sarcoma virus (RSV), 704
 RPA (replication protein A), 1206
 RPC, *see* Reverse-phase chromatography
 RPE65 gene, 123
 Rpn11, 1418
 Rpr4, 1327
 Rpr6, 1327
 Rpr40, 1327
 Rpr41, 1327
 Rpr42, 1327
 Rpr43, 1327
 Rpr44, 1327
 Rpr45, 1327
 Rpr46, 1327
 Rpt5, 1417
 RRF (ribosomal recycling factor), 1375T, 1392, 1394, 1394F
 RRM (RNA recognition motif), 1303–1304, 1304F, 1365
 5S rRNA, 1328, **1370**
 5S rRNA, 1301, 1328, 1364, **1370**
 16S rRNA, 1328, 1364–1366, 1370F
 base pairing with rRNA, 1374–1375
 and stringent response, 1301
 18S rRNA, 1328, **1370**
 23S rRNA, 1301, 1328, 1364, 1366
 28S rRNA, 1328, **1370**
 30S rRNA, **1370**
 40S rRNA, 1370T
 50S rRNA, **1370**
 60S rRNA, 1370T
 80S rRNA, 1370T, 1371, 1371F
 rRNA (ribosomal RNA), 1260. *See also*
 Ribosomes
 methylation, 1328
 posttranscriptional modification, 1328–1329, 1328F
 in protein synthesis, 98
 structure of, 213
 zonal ultracentrifugation, 159
 RS75091, 1001
 R-SNAREs (SNAP receptors), 441, 442, 779, 780
 RS system, 76–77
- R state (oxyHb quaternary conformation), 333–334, 336–339, 336F, 337F
 and hemoglobin cooperativity, 352–353
 and sequential model of allosterism, 351
 and symmetry model of allosterism, 349–351
 R state, ATCase, 476–479, 477F
 RSV (Rous sarcoma virus), 704
 RTKs, *see* Receptor tyrosine kinases
 Rtt103 protein, 1304
 R-type pneumococci, 86, 86F
 Ru5P, *see* Ribulose-5-phosphate
 RUB1 (related-to-ubiquitin-1), 1421
 Ruben, S., 813, 903
 RuBisCo (ribulose-1,5-bisphosphate carboxylase-oxygenase), 299–300, **934–935**, 934F
 RuBP (ribulose-1,5-bisphosphate), 927
 RuBP carboxylase, *see* Ribulose bisphosphate carboxylase
 Rubredoxin, 135F
 Running gel, in disc electrophoresis, 148
 Rupley, J., 520
 Rutter, W., 530
 RuvA, 1231–1233, 1232F, 1234
 RuvAB, 1231, 1234, 1235
 RuvABC, 1230–1234
 RuvABC resolvosome, 1234
 RuvB, 1231–1234
 RuvC, 1233, 1234
 Rye, H., 299
- S
 S, *see* Entropy
 S1P, *see* Site-1 protease
 S2P (site-2 protease), 988
 s⁴U (4-thiouridine), 1347F
 S⁶, 1417
 S7P, *see* Sedoheptulose-7-phosphate
 S20/L26, 1364
 Sabatini, D., 420
 Sac1, 736
 Saccharides, **359**. *See also* Glycoproteins;
 Monosaccharides; Polysaccharides
 Saccharomyces cerevisiae (Baker's yeast):
 Cdc6, 1206–1207
 genome sequencing, 177T, 182
 mitochondrial deviations from ideal, 1344T
 replication origins, 1205
 replication studies, 1202
 RNAP II structure, 1277–1280, 1278F
 RNAP subunits, 1277T
 as source of proteins for purification, 130
 topoisomerase II, 1167–1168, 1168F
 Saccharopine, 1040
 Sacchettini, J., 944, 1407
 Saenger, W., 916, 922
 Saibil, H., 297
 SAICAR, *see* 5-Aminoimidazole- 4-(N-succinylcarboxamide) ribotide
 SAICAR synthetase (PurC), 1109F, 1111
 Salamanders, 14F
 Sall, 105T
 Sali, A., 1371
 Salmine, 134T
Salmonella typhimurium, 1224, 1243
 Salts:
 dissolution in water, 42–43
 and protein denaturation, 265–266
 protein solubilities and concentrations of, 133–134, 133F
 Saltatory conduction, 777
 Salt bridges, 259–260, 262, 339
 Salting in, 132T, 133–134
- Salting out, 132T, 134
 Salvage pathways, **1114**
 SAM (*S*-adenosylmethionine), 80, 80F, 1004, 1034
 and Dam/Dcm MTases, 1247, 1247F
 as methylating agent, 1062, 1214F, 1218
 met repressor complexed with, 1290–1291, 1291F, 1297
 and mRNA capping, 1302
 rRNA methylation, 1328
 Sam35, 449
 Sam37, 449
 Sam50, 449
 SAM complex, 448–449
 Samuelson, B., 994
 Sancar, A., 1217
 Sandhoff's disease, 1013T
 Sanger, F., 164, 165, 176, 1344
 Sanger method, of nucleic acid sequencing, 176–180, 184
 Santi, D., 1127, 1247
 SAP-A, 1011
 SAP-B, 1011
 SAP-C, 1011
 SAP-D, 1011
 Saposins, 1011
 SAPs (sphingolipid activator proteins), 1011
 Saquinavir, 550F, 551
 Sar1, 430
 Sar1p, 437
 α-Sarcin, 1395T
 Sarcomas, 704
 Sarcoplasmic reticulum (SR), 761, 763
 Sarcosine, 32T
 Sareste, M., 708
 Sarin, 782
 SARs (structure–activity relationships), 540
 SASPs (small acid-soluble spore proteins), 1149
 Satellite bands, 159
 Sattler, M., 1314
 Saturated enzymes, 489
 Saturated fatty acids, 387–388, 387T
 Saturation function, 747
 Saturation kinetics, 747
 Sax, M., 616
 Saxitoxin, 777
 Sazanov, L., 835
 SBPase, *see* Sedoheptulose bisphosphatase
 SCA (spinocerebellar ataxia), 1252
 Scaffold-attachment regions, *see* SARs
 Scaffold proteins, 713–714
 Scalar protons, 843
 Scanning microcalorimeter, 59
 SCAP, *see* SREBP cleavage-activating protein
 Scatchard, G., 674
 Scatchard plot, 674
 Scavenger Receptor Class B type I (SR-BI), 456
 Scavenger receptors, 457
 SCF^{Skp2}, 1412, 1413F
 SCF complexes, 1412
 Schachman, H., 475
 Schaftingen, E., 663
 Schally, A., 683
 Schatz, G., 445
 Schiff base, 510, 568, 569F, 1020–1021
 Schiffer, M., 910
 Schimmel, P., 1351, 1356
 Schimmel, P.R., 1329
 Schindelin, H., 291
 Schirmer, H., 803
 Schirmer, T., 749
Schistosoma mansoni, 1310–1311, 1311F

- Schizosaccharomyces pombe*, 1202, 1206–1207
 Schleiden, M., 3
 Schleif, R., 1292, 1293
 Schlessinger, J., 729
 Schoenheimer, R., 1408
 Schofield, C., 534
 Schulman, L., 1351
 Schultz, S., 1212
 Schulz, G., 403, 628, 803, 983
 Schwann, T., 3
 Schwann cell, 778F
 SCID, *see* Severe combined immunodeficiency disease
 SCID-X1, 123
 SciFinder Scholar, 35
 Scintillation counter, 149
 SCIPs, 735
 Scissile peptide bond, 169
 SCOP, *see* Structural Classification of Proteins
 Scorpion venom, 777
 Scott, W., 1310
 Scrapie, 312–315, 313F
 Screens, for drugs, 539
 Scrunching, 1268
 Scurvy, 236, 364
 SDS (sodium dodecyl sulfate), 150, 169, 171, 399F
 SDS-PAGE (polyacrylamide gel electrophoresis), 150, 151F, 169, 411F. *See also* PAGE (polyacrylamide gel electrophoresis)
 Sec (selenocysteine), 1361
 Sec1, 441
 Sec7, 435
 Sec12p, 437
 Sec13/31, 430, 438–439, 438F
 Sec23/24, 430
 Sec61, 424
 Sec62/Sec63 complex, 428
 SecA, 428
 SecB, 428
 Secondary lysosome, 455F
 Secondary structure, 163, 221–233
 helical, 225–229, 225F
 nonrepetitive, 230, 232
 peptide group in, 221–225
 predicting, for proteins, 302–304
 proteins, 221–233
 tRNA, 1345–1346, 1346F
 Second law of thermodynamics, 54–57
 Second messengers, 661, 725–727
 Second-order rate equation, 484
 Second-order reactions, 483, 484F
 Secretary proteins, 421F
 β -Secretase, 312, 547
 γ -Secretase, 312
 Secretin, 673T, 675
 Secretory pathway, 420–428
 Secretory vesicles, 429F, 1405
 SecY, 424–425, 425F
 SeCys (selenocysteine), 1361
 Sedimentation, 152–153
 Sedimentation coefficient, 153, 154F
 Sedoheptulose-7-phosphate (S7P), 636, 894
 Sedoheptulose bisphosphatase (SBPase), 929, 933F
 Seeberger, P., 383
 Seed sequence, 1325
 Séguin, A., 823
 SELB, 1361
 Selectable markers, 111
 Selenocysteine (Sec) (SeCys), 1361
 SELEX (Systematic Evolution of Ligands by Exponential enrichment), 213–214
 Self-assembly, 17, 278, 302
 Self-compartmentalized proteases, 1419–1420
 Self organization, 29, 30F
 Self-splicing genes, 1306–1308
 Sem-5, 709
 Semiconservative replication, 90
 Semiconservative replication, of DNA, 90, 1173–1174
 Semidiscontinuous replication, of DNA, 1175–1176, 1175F
 Semi-invariant position, 1346
 Sendai virus, 408, 409F
 Senebier, J., 901
 Senescence, 703, 1210
 Sense RNA, 1323
 Sense strand, 203, 1266, 1267F
 Sephadex gels, 139
 Sepharose gels, 139
 Septic shock, 688
seqA gene, 1196
 SeqA protein, 1196
 Sequence databases, 194–195
 Sequence-tagged connectors (STCs), 181
 Sequence-tagged sites (STSs), 181
 Sequencing, *see* Nucleic acid sequencing; Protein sequencing
 Sequencing gel, 177–178, 178F, 179F
 Sequential model, of allosteric interactions, 351–352
 Sequential reactions, 498–501, 500F
 Sequestration, 1195
 SER, *see* Smooth endoplasmic reticulum
 Ser, *see* Serine
 SERCA complex, 762–764, 762F, 763F
 Series-5 leukotrienes, 1003
 Serine:
 amino group, 208F
 biosynthesis, 1071–1072
 degradation, 1030–1034
 genetic codes for, 100T, 1343T
 in globular proteins, 246–247
 in glycerophospholipids, 389T
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 inserted by *E. coli* nonsense suppressor, 1362T
 in native unfolded proteins, 283
 as nonessential amino acid, 1065T
 in PEST proteins, 1413
 side chains, 71, 208F, 264T
 structure and general properties, 69T
 Serine carboxypeptidase II, 531, 531F
 Serine dehydratase, 1031, 1408T
 Serine hydroxymethyltransferase, 1031, 1033–1034, 1129
 Serine palmitoyltransferase, 1008
 Serine proteases, 525T
 catalytic mechanism, 531–537, 532F
 evolutionary relationships among, 530–531
 kinetics and catalytic groups, 525–527
 specificity pockets, 529F
 transition state stabilization, 534F
 X-ray structures, 527–531
 zymogens, 537–538
 Serine recombinase, 1239
 Serotonin, 783, 1046, 1059
 Serotonin receptors, 689
 Serpentine receptors, 689
 SerRS, 1351
 Serum, disc electrophoresis of, 149F
 Serum albumin, 134T, 140F, 944–945, 945F
 Serum albumin C, 133F
 Serum albumin dimer, 140F
 Seryl-adenylate, 1358
 Severe combined immunodeficiency disease (SCID), 122, 123, 1132
 Sex chromosomes, 22–23
sex-lethal (sxl) gene (*Drosophila melanogaster*), 1318
 Sex linked genes, 24
 SF1 (splicing factor 1), 1312
 S form (*RS* system), 76–77
 SH1 domains, 706
 SH2-containing inositol-5-phosphatases (SHIPs), 735
 SH2 domains, 706–707, 1317
 SH3 domains, 708, 1317
 Sharon, N., 523
 Sharp, K., 259
 Sharp, P., 1304, 1306
 Shc, 707, 709–710
 Shear degradation, DNA, 95
 Sheep, scrapie in, 312–315
 Shelterin, 1213
 Shemin, D., 574, 1048
 Shi, Y., 722
 Shimomura, O., 120
 Shine, J., 1374
 Shine-Dalgarno sequences, 1300, 1301, 1374–1375, 1398
 SHIP2, 735
 Shipley, G., 696
 SHIPs (SH2-containing inositol-5-phosphatases), 735
 Shirakawa, M., 1250
 Shoelson, S., 721
 Shoolingin-Jordan, P., 1053
 Short, strong hydrogen bonds, 536
 Short tandem repeats (STRs), 115–116
 Short-term regulation, of fatty acid metabolism, 973
 Shotgun cloning, 113
 Shotgun sequencing, 181
 Shotton, D., 529
 SHP-2, 721–722
 Shulman, G., 1103
 Shuman, S., 1188
 Shuttle vectors, 118
 Sialic acid, 1010
 Sialic acids, 365, 392F. *See also* NANA
 Sialidase, 381
 Sickle-cell anemia, 185–188, 343–347
 gene therapy, 123
 Southern blot for diagnosis or prenatal detection, 112–113
 Sickle-cell hemoglobin, 81. *See also* Hemoglobin S
 Sickle-cell trait, 187–188
 Side chains, of amino acids, 70–71, 208F, 264T, 511
 SIDS (sudden infant death syndrome), 948
 si face, 77, 77F
 Sigler, P., 294, 692, 729, 942, 1288
 σ^{52} , 1284
 σ^{54} , 1284
 σ^{gp28} , 1284
 $\sigma^{gp33/34}$, 1284
 σ^{70} , 1265, 1284
 σ factors, 1265, 1266, 1284
 σ -Replication, 1191–1193, 1192F
 Sigmoidal shape, oxygen-binding curve, 326, 326F, 339
 Signal-anchor sequence, 426
 Signal-gated channels, 772
 Signal peptidase, 422, 1404
 Signal peptides, 420, 1404

- Signal recognition particle, *see* SRP
 Signal sequence, 117
 Signal transducers and activators of transcription (STATs), 718
 Signal transduction, 671, 1099. *See also related topics, e.g.: Hormones*
 Sildenafil, 699, 743
 Silica, 138, 144, 145
 Silica gel, 144
 Silicon, 31
 Silicones, 31
 Silk, 233F
 Silver stain, 149
 Simian sarcoma virus, 705
 Simian virus 40, *see* SV40
 Simmons, D., 1000
 Simon, M., 1243
 Simple sugars, 359
 Simple transposition, 1237
 Simple triacylglycerols, 388
 Simvastatin, 990, 991F
 Singer, J., 408
 Single blind tests, 542
 Single-displacement reactions, 498, 501
 Single nucleotide polymorphisms (SNPs), 184, 213
 Single-pass transmembrane proteins, 426
 Single-strand binding protein, *see* SSB
 Single-stranded DNA, *see* ssDNA
 Single-strand nicks, 1230, 1235F
 Single-tailed lipids, 394, 394F
 Sinning, I., 424
 siRNAs (small interfering RNAs), 1323–1324
 Site-1 protease (S1P), 988, 989
 Site-2 protease (S2P), 988
 Site-directed mutagenesis, 119–120
 Site-specific recombination, 1237, 1239–1240, 1242–1246
 β-Sitosterol, 393
 60S subunit (ribosome), 1370T, 1371
 Size exclusion chromatography, 138
 Sjöstrand, F., 9
 Skeletal muscle, 1091–1092
 Skin, 16F, 235, 238F, 240T
 Skp1, 1412–1413, 1412F, 1413F
 Skp2, 1412–1413, 1412F
 Slab gel electrophoresis, 148F. *See also* Gel electrophoresis
 Slack, R., 936
 Slater, E., 845
 Slicer, 1324
 Sliding clamp, Pol III, 1182–1183, 1183F, 1196–1198
 Slime molds, 7F, 13F
 Slow reacting substances of anaphylaxis (SRS-A), 1000
 Slow-twitch muscle fibers, 619
 SL (spliced leader) RNA, 1320
 Small acid-soluble spore proteins (SASP), 1149
 Small interfering RNAs (siRNAs), 1323–1324
 Small nuclear ribonucleoproteins, *see* snRNPs
 Small nuclear RNAs, *see* snRNAs
 Small nucleolar RNAs, *see* snoRNAs
 Small ubiquitin-related modifier (SUMO), 1421
 Sm core domain, 1315–1316
 Smith, C., 158
 Smith, E., 189
 Smith, H., 105, 181
 Smith, J., 920
 Smith, M., 119
 Smith, T., 200, 1023
 Smith–Waterman alignment algorithm, 201
 Smoking, 458
 Smooth endoplasmic reticulum (SER), 7F, 9, 562T
 SM proteins, 441
 Sm proteins, 1314, 1315F
 Sm RNA motif, 1314
 S_N2 reaction, 514, 514F
 Snake venom, 781, 1017
 Snake venom phosphodiesterase, 176
 SNAP-25, 441, 441F, 443, 444
 SNAP receptors (SNAREs), 441–445, 443F
 SNAPs, *see* Soluble NSF attachment proteins
 SNAREs, *see* SNAP receptors
 Snell, E., 1021
 Snorkeling, 406
 snoRNAs, 1329
 snoRNPs, 1329
 SNPs, *see* Single nucleotide polymorphisms
 snRNAs, 1311–1312
 snRNPs (small nuclear ribonucleoproteins), 1312, 1314
 Snu114, 1313F
 SOD, *see* Superoxide dismutase
 Sodium azide, 131
 Sodium channels, 772, 776–777
 Sodium cholate, 399F
 Sodium deoxycholate, 399F
 Sodium dodecyl sulfate, *see* SDS
 Sodium ion, 45T, 1158
 Solid phase peptide synthesis (SPPS), 206, 206F
 SOLiD system, 184
 Söll, D., 1358
 Solubility, protein, 132–135
 Solubilization, protein, 130
 Soluble NSF attachment proteins (SNAPs), 441, 444
 Solution methods, for nucleic acid fractionation, 156–157
 Solvated ions, 43
 Solvents:
 dielectric constants and dipole moments, 43T
 water and nonpolar, 262–264, 263T
 water as, 42–44
 Somatic cells, 19, 20
 Somatomedins, 673T, 683
 Somatostatin, 673T, 675, 683
 Somatotropin, 119, 673T, 683
 Sondek, J., 728
 Sonenberg, N., 1378
 Sonication, 95, 130, 395
 D-Sorbose, 361F
 Sorensen, S., 47
 Soret bands, 838
 Sorting sequence, of LDL receptors, 457
 Sos, 709–711
 SOS box, 1221
 SOS repair, 1222–1223
 SOS response, 1181, 1221–1223, 1221F, 1234
 Southern, E., 111
 Southern blotting (Southern transfer technique), 111–113, 112F, 149, 1207–1208
 Sowadski, J., 653
 Spacer gel, 148
 Spacer groups, 142
 Special pair, in photosynthesis, 910–912, 913F
 Specific rotation, 74
 Spectrin, 412, 413F
 Sperm cells, 12F, 1211
 Sperm whale myoglobin, *see* Myoglobin
 SPGA (*Streptomyces griseus* protease A), 530–531
- S phase, 1202, 1205, 1207
 Spheroplasts, 378
 Sphinganine, 1009
 Sphingoglycolipids, 391, 1004, 1008–1013, 1009F
 Sphingolipids, 390–392, 395, 1004F
 Sphingolipid activator proteins (SAPs), 1011
 Sphingolipid storage diseases, 392, 1011, 1013
 Sphingomyelins, 390–391, 391F, 398T, 1008, 1008F
 Sphingophospholipids, 390–391, 1004, 1008–1009
 Sphingosine, 390, 391F, 392F
 Spiegelman, S., 1264–1265
 Spina bifida, 1035
 Spinocerebellar ataxia (SCA), 1252
 Spirilla, 4
Spirillum, 4F
 Spirochetes, 4F
 Spleen phosphodiesterase, 176
 Spliced leader (SL) RNA, 1320
 Spliceosomes, 1308, 1312–1313, 1312F, 1313F
 AU-AC, 1319
 structures, 1314–1316, 1316F
 Splicing, *see* Gene splicing
 Splicing factor 1 (SF1), 1312
 Sponges, 13F
 Spontaneity, 52, 57–58, 57T
 Spontaneous processes, 52, 57–58, 57T
 Spores, 4, 117, 1149, 1284
 Sporulation, 1149
 SPPS, *see* Solid phase peptide synthesis
 SQS, *see* Squalene synthase
 Squalene, 975, 975F, 983F
 cyclization, 982–983
 formation, 978–982, 979F
 Squalene epoxidase, 982
 Squalene-hopene cyclase, 406, 983
 Squalene synthase (SQS), 978, 982F
 SR, *see* Sarcoplasmic reticulum; SRP receptor
 SRα, 423
 SRβ, 423
 SR1 mutant, 299–300
 SRA domain, 1250
 SR-BI (Scavenger Receptor Class B type I), 456
 Src, 244F, 706–707, 715
 Src family, 715
 Src homology domains, *see* SH 1 domains; SH2 domains; SH3 domains
 SRE (sterol regulatory element), 987
 SREBP, *see* Sterol regulatory element binding protein
 SREBP cleavage-activating protein (SCAP), 987–989
 SRP (signal recognition particle), 420–424, 421F, 423F, 1404
 SRP9, 422
 SRP14, 422
 SRP19, 422
 SRP54, 422
 SRP68, 422
 SRP72, 422
 SRP receptor (SR), 420–421, 421F, 423
 SR proteins, 1314
 SRS-A (slow reacting substances of anaphylaxis), 1000
 SSB (single-strand binding protein), 1183–1184, 1184F, 1184T, 1187, 1187F, 1195
 (SSB)₃₅, 1187
 (SSB)₆₅, 1187
 ssDNA (single-stranded DNA), 1180, 1185, 1187

- Stabilization, protein, 130–131
 Stacked-X conformation, 1227F
 Stacking associations, in tRNAs, 1348–1349
 Stacking gel, 148
 Stadtman, E., 1068
 Stadtman, T., 1361
 Staggered conformation, 222, 223F
 Stahl, F., 90, 1173
 Staining, 149, 158
 Standard free energy changes, 59–60
 Standard redox potential, 584, 585T
 Standard state, 58–60
 Stanley, C., 1023
 Staphylococcal nuclease, 143F
Staphylococcus, 4F
 Starch, 359, 369–370, 932F
 Starch-branched enzyme, 931
 Starch synthase, 645, 931
 Stark, H., 1315–1316
 Starvation, 871, 973, 1101
 State, of a system, 53, 56
 State functions, 53, 54
 Statins, 990–991
 Stationary phase, 135
 Statistical factors, 348
 STATS (signal transducers and activators of transcription), 718
 Stavudine, 1208
 STCs (sequence-tagged connectors), 181
 Ste5p, 713–714
 Steady state, 488–490, 588
 Stearic acid, 386, 387F, 387T, 392F
 1-Stearoyl-2-oleoyl-3-phosphatidylcholine, 390F
 Steatorrhea, 943
 Steitz, J., 1311, 1329, 1356
 Steitz, T., 597, 968, 1177, 1178, 1208, 1229, 1240, 1286, 1353, 1366, 1375, 1383
 Stem cells, 121
 Step-wise elution, 136–137, 136F
 Stercobilin, 1056, 1057
 Stereoelectronic assistance, 514–515
 Stereoelectronic control, 514
 Stereoisomers, 75
 Stereo pair, 271
 Stereo pictures, 271, 271F
 Stereospecificity, enzymes, 470–472
 Stereospecific numbering, 389F
 Steric interference, polypeptides, 222–225, 223F
 Steroid hormone receptors, 291–293
 Steroid hormones, 392, 393, 673T, 680–682, 976
 biosynthesis, 991–993, 992F
 Sterol, 392
 Sterol regulatory element (SRE), 987
 Sterol regulatory element binding protein (SREBP), 988, 988F, 989
 Sterol-sensing domain, 987
 Steven, A., 436
 Stevens, R., 404
 Sticky ends, 105, 117–118
 Stigmastanol, 393
 Stigmatellin, 849
 Stöffler, G., 1365
 Stokes equation, 155
 Stokes radius, 155
 Stomata, 904
 Stoops, J., 798
 Stop codons, 99, 100T, 1343
 Stopped-flow device, 284, 284F
 Stop-transfer anchor sequence, 426
 Storage polysaccharides, 369–370
 Stout, D., 808
 Strain, lysozyme catalytic mechanism and, 522–523
 Strand passage mechanism, 1163–1165, 1164F
Streptavidin, 298, 857–858, 873
Streptomyces alboniger, 1382
Streptomyces antibioticus, 1272
Streptomyces griseus protease A (SPGA), 530–531
Streptomyces mediterranei, 1269
Streptomycin, 1263F, 1388, 1395, 1395T, 1396F, 1397
Streptomycin-resistant mutants, 1395, 1397
 Stringent control, 107, 1301
 Stringent factor, *see* RelA
 Stringent response, 1301, 1397
 Stroke, 865
 Stroma, 11–12, 902
 Stromal lamellae, 902
 Strong acids, 46–47
 Strong bases, 47–48, 48F
 Stroud, R., 529, 817
 STRs (short tandem repeats), 115–116
 Structural Classification of Proteins (SCOP), 256T, 258
 Structural genes, 117, 203, 1262
 Structural genomics project, 258–259
 Structural polysaccharides, 367–369
 Structure-activity relationships (SARs), 540
 Structure-based drug design, 541
 STSs (sequence-tagged sites), 181
 Stubbe, J., 1119
 Stueher, D., 686
 Sturtevant, A., 24
 S-type pneumococci, 86, 86F
su1 nonsense suppressor, 1362T
su2 nonsense suppressor, 1362T
su3 nonsense suppressor, 1362T
su4 nonsense suppressor, 1362T
su5 nonsense suppressor, 1362T
su6 nonsense suppressor, 1362T
su7 nonsense suppressor, 1362T
 Sublactam, 378
 Submitochondrial particles, 852
Substania nigra, 1060
 Substituents, of prochiral centers, 77
 Substrates, 58, 348, 469, 1128
 Substrate cycles, 624, 628–629
 Substrate-level phosphorylation, 582
 Substrate specificity, 470–473, 470F
 Subtilisin, 119, 525T, 531, 531F, 534
 Subunits, protein:
 composition, 270–271
 interactions, 266, 267
 and nucleic acid subunits, 164
 Subunit IV, of cytochrome *b*, 920
 Succinate, 493, 791, 811, 947
 Succinate dehydrogenase, *see* Complex II
 Succinate thiokinase, 810
 Succinic acid, 32T, 49
 Succinylcholine, 782
 Succinyl-CoA, 567, 956
 from amino acids, 1034–1039
 in citric acid cycle, 766F, 789, 810–811, 811F
 in heme biosynthesis, 1048
 Succinyl-CoA synthetase, 791, 810–811, 811F
 Succinyl-phosphate, 810
 Suck, D., 1233
 Sucrase, 370
 Sucrose (table sugar), 144, 265, 367, 367F, 488, 932F
 Sucrose-6-phosphate, 932
 Sucrose-phosphate phosphatase, 932
 Sucrose-phosphate synthase, 932
 Sucrose phosphorylase, 501
 Sudden infant death syndrome (SIDS), 948
 Sugar-phosphate chain conformation, 1152–1154, 1152F, 1153F
 Sugars, 15, 359. *See also* Monosaccharides; Polysaccharides
 synthesis in prebiotic conditions, 33
 Suicide substrates, 1128
 Sulfanilamide, 1064
 Sulfate ion, 45T
 Sulfate-reducing bacteria, 6
 Sulfatides, 391, 1008, 1010, 1010F
 Sulfite reductase, 934, 1072
Sulfolobus solfataricus, 1222, 1222F
 Sulfonamides, 1064, 1111
 Sulfoquinovosyl diacylglycerol, 902
 Sulfuric acid, 46
 Sulston, J., 181
 Sumner, J., 132, 470
 SUMO (small ubiquitin-related modifier), 1421
 Sundaralingam, M., 1153
 Sup35 protein, 315–316
 Superacids, 511
 Supercoiled DNA, 1158–1170, 1160F, 1270–1271
 measurements of supercoiling, 1160–1162
 superhelix topology, 1158–1160
 topoisomerases, 1162–1170
 Superhelix, 433, 1158–1160
 Superhelix density, 1162
 Superoxide dismutase (SOD), 489T, 491, 491F, 866
 Superoxide ion, 325, 341
 Superoxide radical, 491
 Superrepressed mutants, 1336
 Supersecondary structures, 249–251
 Supertwisting, 1158
 Suppressors, 1340
 Suppressor tRNAs, 1362
 Supramolecular assemblies, 15, 16F
 Surface-active molecules, 51
 Surface hydrophobicity, 145
 Surface labeling, 400
 Surfactants, 51
 Surroundings, 52
 Sutherland, E., 661, 722
 Sutton, W., 22
 SV40 (Simian virus 40):
 enhancers, 1282–1283
 GC boxes, 1282
 large-T antigen, 1211
 replication studies, 1202
 supercoiling, 1162, 1162F
 Svedberg, T., 152
 Svedbergs (unit), 153
 Swiss-Pdb Viewer, 256T, 257
 SWISS-PROT, 195, 203
 Swiss-Prot Protein Knowledgebase, 195T
 Swiss roll barrel, 252, 252F
sxl (*sex-lethal*) gene (*Drosophila melanogaster*), 1318
 SXL protein, 1318
 SYBR Safe, 158
 Symmetry, proteins, 267–270
 Symmetry model, of allosteric interactions, 349–351
 Symport, 758, 758F
 Synapses, 440, 440F, 527, 778
 Synaptic cleft, 440, 440F, 778
 Synaptic complex, 1237
 Synaptic vesicles, 440, 441F, 779, 780
 Synaptobrevin, 441, 441F, 444
 Synaptjanin1, 736
 Synaptosomes, 440, 779, 1241F
 Synaptotagmin I, 779

- Synchrotrons, 241
 Syn conformation, 1152F
Synechocystis sp., 177T
 Syngle, R., 144
 Synonyms, in genetic code, 99, 1343
Syntaxis, 433F, 441, 441F, 443, 444
 2',5'A synthetase ([2',5']-oligoadenylate synthetase), 1400
 Syphilis, arsenicals for, 799
 SYPRO dyes, 149
 Systematic Evolution of Ligands by Exponential enrichment (SELEX), 213–214
 Systematic names, enzymes, 479
 Systemic lupus erythematosus, 1204, 1312
 Systems, 52–53
 Systems biology, 576–578, 576F
 Szent-Györgyi, A., 791
 Szostak, J., 1210
- T
 T (ribothymidine), 1347F
 T1 G6P translocase, 665F
 T2 transporter, 665F
 T3 transporter, 665F
 T4 Endonuclease VII, 1233–1234, 1233F
 T4 Lysozyme, 282, 304
 T7 gene 4 helicase/primase, 1184, 1184F, 1188
 T7 gp4, 1184, 1184F
 Tabun, 782
 Tacrolimus, 724
 Tagamet, 764
 D-Tagatose, 361F
 Tail spike protein, 289–290
 Tainer, J., 686, 1126, 1219
 D-Talose, 360F
 Tandem mass spectrometry (MS/MS), 173–175, 174F, 175F
 Tangier disease, 458
Taq DNA polymerase, 114, 179
 TaqI, 105T
Taq RNAP, 1268, 1269, 1269F
 T arm, 1346
 Tarui's disease, 667
 TATA box, 1282F
 Tatum, E., 26, 570, 1262
 Tau protein, 312
 Taurine conjugates, 993, 993F
 Taxonomy, 6, 12, 189
 Taylor, S., 653, 654
 Tay–Sachs disease, 123, 392, 1011, 1013, 1013F
 TBSV (tomato bushy stunt virus), 277
 T_qC arm, 1346
 TC10, 738
 TCA (tricarboxylic acid) cycle, 789. *See also* Citric acid cycle
 T cell receptor, 724
 T cells, 123, 292–293, 671
 TCR (transcription-coupled repair), 1217
 TD₅₀, 539
 TDF (testes-determining factor), 682
 TE (palmitoyl thioesterase), 965
 TEBP (telomere end-binding protein), 1212–1213, 1212F
 Teeth, 235
 Teichoic acids, 378, 378F
 Telomerase, 1209–1213
 Telomere end-binding protein, *see* TEBP
 Telomere repeat-binding factor 1 (TRF1), 1213
 Telomere repeat-binding factor 2 (TRF2), 1213
 Telomeres, 109, 1209–1213
 Telomeric T-loops, 1213, 1213F
 Telophase, 20F, 21F
 Temin, H., 545, 1207
- Temperature:
 denaturation, 57
 and entropy, 56F
 and lipid bilayer fluidity, 397–399
 Temperature-jump, 284
 Template binding, RNAP, 1266–1267, 1266F
 Template strand, 96, 102F
 Tendons, 232, 233, 235, 240T
 Tensin, 736
 TER, 1210
TerA, 1199
TerB, 1199
TerC, 1199
TerD, 1199
TerE, 1199
TerF, 1199
TerG, 1199
TerH, 1199
TerI, 1199
TerJ, 1199
 Terminal deoxynucleotidyl transferase, 109, 109F
 Terminal desaturases, 970–971
 Terminal transferase, 109, 109F
 Terminal uridylyltransferase (TUTase), 1322
 Termites, 369
 Termolecular reactions, 483
 Terpenoids, 976–977
Ter sites, 1199, 1206
 TERT (telomerase subunit), 1210, 1211, 1211F
 Tertiary structure, 163
 α/β barrels, 252, 254
 domains, 247–249, 251–253, 253F
 globular proteins, 245–256
 α helices and β sheets, 246
 open β sheets, 254–256
 predicting, for proteins, 304–305
 and protein folding, 287
 side chain polarity, 246–247
 supersecondary structures, 249–251
 tRNA, 1348–1349
 Testes, 672F
 Testes-determining factor (TDF), 682
 Testicular feminization, 681
 Testosterone, 681
 Tetanus, 442
 Tethering factors, 441
 Tetracycline, 128, 1395T, 1396F, 1397
 Tetracycline-resistant bacterial strains, 1397
 Tetrahedral intermediate, 531–533
 Tetrahedral symmetry, proteins, 268, 268F
 5,6,7,8-Tetrahydrobiopterin, *see* BH₄
 Tetrahydrofolate, 473T, 492
 Tetrahydrofolate (THF), 1062–1064
 Tetrahydrofuran-2-carboxylate, 516
Tetrahymena thermophila:
 Group I introns, 1307–1310, 1307F, 1309F, 1310F
 telomerase, 1209–1210
 Tetraloop, 1308
 2,3,4,6-O-Tetramethyl-α-D-glucose, 508
 Tetramethyl-p-phenylenediamine (TMPD), 831
 Tetranucleotide hypothesis, 85
 Tetrapeptides, 73F
 Tetraubiquitin, 1411F
 Tetrazole, 211
 Tetrodotoxin, 777
 TeTx, 442
 TFIIB, 1218
 TFIIS protein, 1281–1282
 TG, *see* Thapsigargin
 TGN, *see* Trans Golgi network
- Thalassemias, 219, 342
 Thalidomide, 78, 78F
 Thapsigargin (TG), 763, 763F
 ThDP (thiamin diphosphate), 616
 Theoretical plate, in chromatography, 135
 Therapeutic index, 539
 Thermal stability, proteins, 266
Thermatoga maritima, 1394F
 Thermoacidophiles, 6
 Thermocycler, 114
 Thermodynamics, 52–61. *See also* Equilibrium
 chemical equilibria, 58–61
 defined, 52
 electron-transport chain, 828–829
 first law, 52–54
 free energy, 57–58, 61
 glycogen metabolism, 643–644
 of life, 586–589
 of membrane transport, 744–745
 of phosphate compounds, 578–583
 second law, 54–57
 transition state, 486
 units and constants, 53T
 Thermogenesis, 629, 860–861, 1099–1001
 Thermogenin, 861, 1101
 Thermolysin, 169T
Thermoplasma acidophilum, 1415, 1415F
Thermoproteus, 7F
 Thermosomes, 301
 Thermostable proteins, 266
Thermus aquaticus, 179, 466
 DNA polymerase I (Klenetq1), 1178–1180
 EF-Tu, 1381F
 in PCR, 114
 RNAP, 1268, 1269, 1269F
Thermus thermophilus:
 30S subunit, 1389–1390
 Argonaute, 1324, 1325F
 EF-G complex with GDP, 1385–1386, 1386F
 proteinase K, 1383
 ribosomal complex, 1384, 1392–1394, 1393F, 1394F
 ribosomal subunits, 1366–1368, 1371, 1386F
 RNAP, 1268, 1268F–1269F
 RuvB, 1231–1233, 1232F
 θ replication, 1174, 1175F
 θ structures, 1174
 THF, *see* Tetrahydrofolate
 Thiamin diphosphate (ThDP), 616
 Thiamine, 474T, 618
 Thiamine pyrophosphate, *see* TPP
 Thiazolidinediones (TZDs), 1103–1104
 Thiazolinone, 165
 Thiazolium ring, 616
thi box, 1300
 Thin layer chromatography (TLC), 145
 Thiogalactoside transacetylase, 1261
 Thiohemiacetal, 608
 Thiokinases, 945
 Thiolase, 947, 958, 976
 Thiolase reaction, 949–950
 Thioredoxin (Trx), 291, 933–934, 934F, 1124
 Thioredoxin reductase (TrxR), 1125
 4-Thiouridine (s⁴U), 1347F
 30S subunit (ribosome), 363, 1363T, 1368, 1368F, 1369F, 1370F
 path of mRNA through, 1375, 1376F
 Thirumalai, D., 300
 Thr, *see* Threonine
 Thr-AMS, 1358
 Threading (fold recognition), 305

- 3₁₀ Helix, 224F, 226, 228
 3' end, of nucleic acids, 84
 3.6₁₃ Helix, *see* α helix
 Threonine:
 biosynthesis, 1072–1073, 1075
 degradation, 1030–1034
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 in globular proteins, 246–247
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 PEST proteins rich in, 1413
 side chains, 71, 264T
 stereoisomers, 75–76, 76F, 77F
 structure and general properties, 69T
 D-Threonine, 75, 76, 76F
 L-Threonine, 75, 76, 76F
 (2S,3R)-Threonine, 76, 77F
 Threonine dehydrogenase, 1034
 Threonyl adenylate, 1357
 D-Threose, 360F
 Thrombin, 525T, 735
 Thromboxanes, 993, **995–1000**
 Thromboxane A₂ (TxA₂), 998
 Thromboxane receptors, 689
 Thromboxane synthase, 998
thr operon, 1299T
 ThrRS, 1358
 Thr-tRNA^{Thr}, 1358
 Thy, *see* Thymine
 Thylakoids, 12, 902, 916F, 926, 926F
 Thymidylate synthase (TS), 1062, 1126–1128
 Thymidylate synthesis, 1126–1130
 Thymine, 18, **83**, 83T, 1177. *See also* Watson-Crick base pairs
 origin, **1126–1130**
 and point mutation generation, 1339–1340
 in pyrimidine catabolism, 1136F
 tautomeric forms, 88F
 Thymine dimer, 1214–1215, 1214F
 Thyroid, 672F
 Thyroid hormones, **676–677**
 Thyroid hormone deiodinases, 1361
 Thyroid hormone receptor, 677
 Thyroid hormone response element (TRE), 879
 Thyroid-stimulating hormone (TSH), 683
 Thyroperoxidase (TPO), 676
 Thyrotropin, 673T, 683
 Thyrotropin-releasing factor (TRF), 673T, 683
 Thyroxine, 80, 80F, 310, 673T, 676–677
 TIM (triose phosphate isomerase) (TPI):
 in Calvin cycle, 929
 of chicken muscle, 231F, 252, 254F
 in glycolysis, 596F, **603–606**, 605F
 Tim9, 448
 Tim10, 448
 Tim12, 448
 Tim14, 447
 Tim16, 447
 Tim17, 447
 Tim18, 448
 Tim21, 447
 Tim22, 448
 Tim23, 446–448
 Tim44, 447
 Tim50, 447
 Tim54, 448
 TIM barrel, 252, 254F, 300
 TIM proteins, 445–449
 TIN2, 1213
 Tisarius, A., 146
 Tissues, 16F
 Tissue culture, 575
 Tissue regeneration, 1163
 Tissue slices, 575
 Tissue-type plasminogen activator (t-PA), recombinant, 119
 Titin, 1305
 Titration curves, **47–48**, 47F, 48F
 of proteins and polypeptides, 72, 72F, 73F
 TLC (thin layer chromatography), 145
 T-loops, telomeric, 1213, 1213F
 TLS (translesion synthesis), 1222
 T_m, *see* Melting temperature
 TMPD (tetramethyl-*p*-phenylenediamine), 831
 TM proteins, *see* Transmembrane proteins
 Tn3, 1237, 1239
 Tn5, 1237–1239
 Tn5 transposase, 1237–1239, 1238F
 TNBS (trinitrobenzenesulfonic acid), 419, 419F
 TNF- α , *see* Tumor necrosis factor- α
 TNFs, *see* Tumor necrosis factors
 TnpA, 1237
tnpA gene, 1237
 TnpR, 1237, 1239
tnpR gene, 1237
 Toads, 1409
 Tob35, 449
 Tob55, 449
 TOB complex, 448, 449
 Todd, A., 84
 Tollen's reagent, 364
 Toluene, 130
 Tom5, 446
 Tom6, 446
 Tom7, 446
 Tom20, 446, 447F
 Tom22, 446
 Tom37, 449
 Tom38, 449
 Tom40, 446
 Tom55, 449
 Tom70, 446
 Tomato bushy stunt virus (TBSV), 277
 TOM complex, 446, 447
 TOM core complex, 446, 447, 447F
 TOM proteins, 445–447
 Tong, L., 604
 Tonks, N., 721
 Top 7 protein, 306, 306F
 Topogenesis, 427
 Topoisomerases, **1162–1170**
 Topoisomerase I, 1165–1166, 1165F
 Topoisomerase II, 1163, 1166–1169, 1169F
 Topoisomerase III, 1163, 1164F
 Topoisomerase IV, 1166
 Topoisomerase V, 1166
 Topoisomerase VI, 1166
 Topoisomerase inhibitors, 1169–1170
 Topologically bound polynucleotide strands, 1159F
 Topological switch point, 255
 Topology:
 of integral proteins, 406
 polypeptide strands, 230
 Toprim fold, 1189
 TORC2, 1096
 Toroidal helix, 1160, 1160F
 Torpedo model, 1304
 Torsion angles, 222–225
 Tosyl-L-phenylalanine chloromethyl ketone, *see* TPCK
Toxoplasma gondii, 1116
 Toxoplasmosis, 1116
 Toyoshima, C., 760, 763
 t-PA, recombinant, 119
 TPCK (tosyl-L-phenylalanine chloromethyl ketone), 527, 527F
 TPI (triose phosphate isomerase), *see* TIM
 TPO (thyroperoxidase), 676
 TPP (thiamine pyrophosphate), 473T, 511, 795, 795T
 in alcoholic fermentation, 616, 616F
 synthesis of, 1300, 1319, 1320F
 TPP1, 1213
 T-protein, 1032
 TR (telomerase subunit), 1211
tra-2 (*transformer-2*) gene (*Drosophila melanogaster*), 1318
 TRA2 protein, 1318
 TRAM (translocating chain-associated membrane protein), 426
 Trans-acting factors, 1264
 Transaldolase, 894, 896–897, 896F
 Transaminases, 498, 1020–1023
 Transamination, 935, **1020–1023**
 Trans cisternae, 428, 429F
 Transcobalamins, 957
Trans configuration gene mutations, 28
 Transcortin, 681
 Transcription, 9, 19, **95–98**, 1145, **1260–1333**. *See also* Posttranscriptional modification; RNAP
 DNA replication vs., 1173
 monitoring, 120–121
 posttranscriptional modification, 1301–1331
 prokaryotic control of, 1283–1301
 proportion of genome transcribed, 182
 RNA in, 1260–1265
 RNAP in, 1265–1283
 speed and accuracy, 1271–1272
 Transcription bubble, 96, 96F, 1267, 1270–1271
 Transcription-coupled repair (TCR), 1217
 Transcription factors, 96, 572
 Transcription repair coupling factor (TRCF), 1218
 Transcriptome, 213
 Transcriptomics, **576**
 Transducin (G_{ta}), 691, 693F
 Transduction, 1225
 Transfection, 87
 Transferase reactions, 497–498
 Transferases, 479T
 Transferrin, 140F, 1056
 Transfer RNA, *see* tRNA
 Transformation, 86, 110–111, 704
 Transformation competent cells, 108
transformer-2 (*tra-2*) gene (*Drosophila melanogaster*), 1318
transformer (*tra*) pre-mRNA (*Drosophila melanogaster*), 1318
 Transforming principle, DNA as, 85–86
 Transgenes, 121
 Transgenic organisms, 86, 86F, **121–122**, 570–572
 Trans Golgi network (TGN), 428, 429F
 Transient phase, 488
 Transimination, 1022
 Transitions, 1339
 Transition state, **485–486**, **515–516**
 Transition state analogs, 516
 Transition state diagram, 485–486, 485F, 487F
 Transition state theory, **484–487**
 Transition temperature, 397
 Transketolase, 894–896, 895F, 929
 Translation, 19, 95, **98–101**, 98F, 1145. *See also* Genetic code; Ribosomes
 accuracy of, 1388–1391
 and antibiotics, 1395

- Translation, (*Contd.*)
 control of eukaryotic, 1398–1402
 defined, 1338
E. coli initiation sequences, 1375F
 genetic code, 1338–1345
 initiation in *E. coli*, 1376F
 posttranslational modification, 1403–1408
 protein degradation, 1408–1421
 ribosomes and polypeptide synthesis, 1362–1398
 tRNA aminoacylation, 1345–1362
 Translational coupling, 1398
 Translesion synthesis (TLS), 1222
 Translocating chain-associated membrane protein (TRAM), 426
 Translocation, 1379, 1387–1388, 1387F
 Translocators, 745
 Translocon, 420, 426–427, 1404
 Translocon-associated membrane protein (TRAP), 425F, 426
 Transmembrane helices, 426–427, 426F
 Transmembrane (TM) proteins, 251, 400–404, 426–427, 427F
 Transmissible spongiform encephalopathy (TSE), 309, 312
 Transmission coefficient, 486
 Transpeptidation, 1379, 1382–1384
 Transplantation, transgenic organisms for, 121–122
 Transporters, 745
 Transport proteins, 744, 745. *See also* Membrane transport
 Transposable elements, 1236
 Transposase, 1236
 Transposition, 1236–1246
 Transposons, 1236–1246
 Trans-SNARE complex, 442
 Trans-splicing, 1319–1320, 1321F
 Transthyretin, 268, 268F, 269F, 310
 Transverse diffusion, 396
 Transversion, 1339
 TRAP, *see* Translocon-associated membrane protein
tra (transformer) pre-mRNA (*Drosophila melanogaster*), 1318
 TRA protein, 1318
 Traps, optical, 1273–1274
 Trastuzumab, 119–120
 TRCF (transcription repair coupling factor), 1218
 β -TrCP, 1414
 TRE, *see* Thyroid hormone response element
 TRF, *see* Thyrotropin-releasing factor
 TRF1 (telomere repeat-binding factor 1), 1213
 TRF2 (telomere repeat-binding factor 2), 1213
 Triacylglycerol lipase, 941, 1089–1090
 Triacylglycerols, 388–389, 940. *See also* Fatty acids; Lipids
 and cholesterol biosynthesis, 986
 and dietary EPA, 1003
 and liver, 1094
 synthesis, 971–973, 972F
 transport in blood, 452F, 973
 Triad family, 1066
 Tricarboxylate transport system, 968
 Tricarboxylic acid (TCA) cycle, *See* Citric acid cycle
 TRiC chaperonins, 301–302, 301F
 Trichloracetic acid, 211
 Tricorn protease, 1420, 1420F
 Triggered conformational changes, 307
 Trigger factor, 293
 Triglycerides, 388–389, 940
 Triglycerols, 940
 Triiodothyronine, 673T, 676
 Trimethoprim, 1129
N,N,N-Trimethylalanine, 79F
 ϵ -*N,N,N*-Trimethyllysine, 79F
 Trinitrobenzenesulfonic acid, *see* TNBS
 Trinucleotide repeat expansions, 1251–1252
 Triolein, 388
 Trioleoylglycerol, 388
 Trioses, 360F
 Triose phosphate isomerase, 231F, *see* TIM
 Tripeptides, 70
 Triple binding assays, 1342
 Triplet codons, 1340–1341
 Tris(2,3-dibromopropyl) phosphate, 1225
 Triskelions, 430F, 431
 Tristearin, 388
 Tristearoylglycerol, 388
 Triton X-20, 399F
 Triton X-100, 399F, 411
 tRNA (transfer RNA), 1145, 1260, 1338
 aminoacylation, 1345–1362
 bases of, 85
 cloverleaf form, 98F
 isoaccepting, 1351
 modified bases, 1346, 1347F
 posttranscriptional modification, 1329–1331
 primary structure, 1345–1346
 role in gene expression, 95F
 role in protein synthesis, 98–101
 secondary structure, 1345–1346, 1346F
 sedimentation coefficients, 154F
 sequencing, 176
 structure and function of, 213
 tertiary structure, 1348–1349
 tRNA^{Ala}, 176, 1345, 1351F, 1352, 1352F
 tRNA^{Arg}, 1352
 tRNA^{Asp}, 1349, 1351F, 1354–1355, 1354F, 1355F
 tRNA^{Cys}, 1352
 tRNA^{Gln}, 1351F, 1352–1354, 1353F, 1355F
 tRNA^{Ile}, 1352, 1355–1358, 1356F, 1357F
 tRNA^{Met}, 1374, 1374F
 tRNA^{Met}_m, 1374
 tRNA^{Phe}, 1346, 1346F, 1348, 1352
 tRNA^{Pyl}, 1362
 tRNA^{Sec}, 1361
 tRNA^{Ser}, 1351F
 tRNA^{Tyr}, 1331
 tRNA^{Val}, 1356
 Trophic hormones, 682
 Tropocollagen, 240
 Tropomodulin, 412F–413F
 Tropomyosin, 134T, 412F–413F
 α -Tropomyosin, 1317, 1317F
 Trout, 1409
 Trp, *see* Tryptophan
 trpL sequence, 1297–1299, 1297F–1299F
 trp operon, 1296–1299, 1297F–1299F, 1299T
 trp repressor, 1288, 1290, 1290F
 Trx, *see* Thioredoxin
 TrxR (thioredoxin reductase), 1125
Trypanosoma brucei, 1418
 Trypanosomiasis, arsenicals for, 799
 Trypanosomes, 1320
 Trypsin, 470, 525, 525T, 529T
 catalytic mechanism, 533, 533F
 cytochrome b_5 cleavage, 401
 gene duplication, 194
 proenzymes, 537, 537F
 specificity, 169–170, 169T
 X-ray structure, 528F, 529
 Trypsinogen, 537, 537F
 Tryptophan:
 biosynthesis, 1075–1078, 1297
 degradation, 1041–1042
 as essential amino acid, 1065T
 and fluorescence resonance energy transfer, 286
 genetic codes for, 100T, 1343T
 in globular proteins, 246–247
 α helix/β sheet propensity, 302T
 inserted by *E. coli* nonsense suppressor, 1362T
 in native unfolded proteins, 283
 nicotinamide coenzyme from, 1137F, 1138
 side chains, 70–71, 264T
 structure and general properties, 68T
 Tryptophan hydroxylase, 1045, 1059
 Tryptophan synthase, 1075–1078
 Tryptophanyl-tRNA^{Trp}, 1298, 1299F
 TS, *see* Thymidylate synthase
 TSE, *see* Transmissible spongiform encephalopathy
 T-segment, of DNA, 1169
 TSH (thyroid-stimulating hormone), 683
 Tsien, R., 121
 t-SNAREs (SNAP receptors), 441
 T state (deoxyHb quaternary conformation), 333–334, 336–339, 336F, 337F
 and hemoglobin cooperativity, 352–353
 and sequential model of allosterism, 351
 and symmetry model of allosterism, 349–351
 T state, ATCase, 476–479, 477F
 Tsukihara, T., 416, 1417
 Tswett, M., 135
d-Tubocurarine, 781
 Tubulin, 10, 10F, 269
 Tubulin α , 301
 Tubulin β , 301
 Tumor necrosis factor- α (TNF- α), 715, 1096, 1414
 Tumor necrosis factors (TNFs), 715, 717
 Tumor promoters, 730–731
 Tumor suppressors, 736, 1414
 Tumor viruses, 704–705, 1163
 Tunicamycin, 888, 889F
 Tunicates, 368
 Turkeys, 19T
 Turner's syndrome, 682
 Turnover number, 490
tus gene, 1199
 Tus protein, 1199, 1199F, 1206
 TUTase (terminal uridylyltransferase), 1322
 Tweezers, optical, 1274, 1274F
 Twilight zone, 196
 Twin-bulb apparatus, 54, 55F, 56
 Twintrons, 1307T
 Twist, of superhelix, 1158–1159, 1159F
 Twitchin kinase, 658
 2,2 Ribbon, 224F, 226, 228
 Two-dimensional (2D) gel electrophoresis, 152, 152F, 175
 Two-dimensional NMR spectroscopy, 243–244, 285–286
 Two-dimensional paper chromatography, 144–145, 145F
 Two-hybrid system, 705
 TxA₂ (thromboxane A₂), 998
 Ty1, 1245
 TYA, 1246
 TYB, 1246
 Tyk2, 718
 Type 0 glycogen storage disease, 668

- Type IA topoisomerases, **1162–1165**, 1163F, 1164F
 Type IB topoisomerases, 1162, **1165–1166**, 1165F
 Type IC topoisomerases, 1162
 Type I diabetes, 310, 1102, 1400
 Type I β bends, 232, 233F
 Type I glycogen storage disease, 661, 666
 Type I restriction endonucleases, 104–105
 Type I TM proteins, 426
 Type I topoisomerases, **1163–1166**
 Type IIA topoisomerases, 1163
 Type IIB topoisomerases, 1163
 Type II diabetes, 310, 1102–1103
 Type II β bends, 232, 233F
 Type II glycogen storage disease, 666
 Type II restriction endonucleases, 105, 105T
 Type II TM proteins, 426, 427F
 Type II topoisomerase inhibitors, 1169–1170
 Type II topoisomerases, 1163, **1166–1170**, 1175
 Type III glycogen storage disease, 667
 Type III restriction endonucleases, 104–105
 Type III TM proteins, 426, 427F
 Type IV glycogen storage disease, 667
 Type IV restriction endonucleases, 105
 Type IV TM proteins, 426–427
 Type V glycogen storage disease, 667
 Type VI glycogen storage disease, 667
 Type VII glycogen storage disease, 667
 Type VIII glycogen storage disease, 667
 Type IX glycogen storage disease, 667
 Type A blood, 22, **415**
 Type AB blood, 22, **415**
 Type B blood, 22, **415**
 Type O blood, 22, **415**
 Type α interferon, 1399
 Type β interferon, 1399
 Type γ interferon, 1399
 Tyr, *see* Tyrosine
 Tyrosine:
 aromatic ring flipping and protein core mobility, 308–309
 biosynthesis, 1075–1078
 degradation, 1043–1047
 and *E. coli* nonsense suppressor, 1362T
 and fluorescence resonance energy transfer, 286
 genetic codes for, 100T, 1343T
 in globular proteins, 246–247
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 in native unfolded proteins, 283
 in neurotransmitter synthesis, 1059–1060
 as nonessential amino acid, 1065T
 side chains, 71, 264T
 structure and general properties, 69T
 Tyrosine aminotransferase, 1408T
 Tyrosine hydroxylase, 1045, 1059
 Tyrosine kinase-based signaling, 699
 binding modules, adaptors, GEFs, and GAPs, **705–711**
 and cancer, **703–705**
 MAP-kinase signaling cascades, **712–714**
 protein phosphatases, **721–725**
 tyrosine kinase-associated receptors, **699–703**, **715–720**
 Tyrosine recombinase, 1239
 Tyvelose, 379F
 TZDs (thiazolidinediones), 1103–1104
 U
 U1-70K, 1315
 U1-A, 1315, 1316
 U1-C, 1315
 U1-snRNA, 1312, 1315–1316
 U1-snRNP, 1312, 1315–1316, 1316F
 U2-snRNA, 1312
 U2-snRNP, 1312
 U4atac–U6atac, 1319
 U4-snRNA, 1312
 U4–U6-snRNP, 1312
 U5, 1319
 U5-snRNP, 1312
 U6-snRNA, 1312
 U6-snRNP, 1312
 U11, 1319
 U12, 1319
 UAA (codon), 1343, 1362, 1362T
 UAG (codon), 1343, 1362, 1362T
 UbcH7, 1412–1413, 1412F
 UBDs (ubiquitin-binding domains), 1420
 Ubiquinol, 836F
 Ubiquinone, *see* Coenzyme Q
 Ubiquitin, 989, 1404, **1410–1421**, 1410F. *See also* Polyubiquitin
 Ubiquitin-activating enzyme (E1), 1410
 Ubiquitination, 1414–1421
 Ubiquitination signals, 1410
 Ubiquitin-binding domains (UBDs), 1420
 Ubiquitin-conjugating enzyme (E2), 1410–1421
 Ubiquitin-like modifiers (UBLs), 1421
 Ubiquitin-protein ligase (E3), 1410–1421, 1410F
 Ubisemiquinone, 836F
 UBLs (ubiquitinlike modifiers), 1421
 Ubr1, 1413
 UCP, *see* Uncoupling protein
 UCP1, 1100
 UCP2, 862, 1101
 UCP3, 862, 1101
 UDG (uracil–DNA glycosylase), 1218–1219, 1219F
 UDPG, *see* UDP-glucose
 UDP-galactose, 631
 UDP-galactose-4-epimerase, 632
 UDP-glucose (UDPG), 631, 644
 UDP-glucose:glycoprotein glucosyltransferase (GT), 886
 UDP-glucose pyrophosphorylase, **644**
 3'-U-exo (3'-U-exonuclease), 1322
 UGA (codon), 1343, 1362, 1362T
 UGA-1 nonsense suppressor, 1362T
 UGA-2 nonsense suppressor, 1362T
 Uhlenbeck, O., 1351
 UHRF1 protein, 1250, 1250F
 Ultracentrifugation, 132T, **152–156**, 159
 Ultrafiltration, 117, 132T, 141
 Umami, 1064
 UMP (uridine monophosphate), 83T, **1114–1118**, 1136F
umuC gene, 1222
umuD, 1222
umuD'2C, 1222
umuD gene, 1222
 Uncompetitive inhibition, **495**, 495F, 502
 Uncoupling protein (UCP), 861–862, 1100–1101
 Unexpressed DNA, 1304–1305, 1316–1317
 UNG (uracil N-glycosylase), 1218
ung *E. coli* mutants, 1218–1219
 Unidirectional replication, 1174, 1175F
 Uniform resource locator (URL), 194
 Unimolecular reactions, 483
 Uniport, 758, 758F
 Unit cells, 270
 Unit evolutionary period, 190
 Universal donors (blood), 466
 Universal recipients (blood), 466
 Unrooted trees, 203F, 204
 Unsaturated fatty acids, 386–388, 387T, **950–952**, 950F
 Unwin, N., 402, 781
 uORFs (upstream open reading frame), 1400–1401
 Up-and-down β Barrels, 251–252
 UP elements, *see* Upstream promoter elements
 Up mutations, 1266
 Upstream open reading frame (uORFs), 1400–1401
 Upstream promoter elements (UP elements), 1266, 1282
 Ura, *see* Uracil
 Uracil, 19, **83**, 83T. *See also* Watson–Crick base pairs
 and genetic code, 100T, 1343T
 and mutations, 1218–1219, 1339–1340
 in pyrimidine catabolism, 1136F
 in tRNAs, 1347F
 Uracil–DNA glycosylase, *see* UDG
 Uracil N-glycosylase (UNG), 1218
 Urate oxidase, 1134
 URE2 gene (yeast), 315
 Ure2p, 315
 Urea, 1025
 in Miller–Urey experiments, 32T
 in protein sequencing, 169, 171
 for renaturation of proteins, 117
 in stopped-flow apparatus, 284
 from uric acid breakdown, 1135F
 Urea cycle, 791
 arginase, **1025–1029**
 argininosuccinase, **1028**
 argininosuccinate synthetase, **1028**
 carbamoyl phosphate synthetase, **1025–1028**
 ornithine transcarbamoylase, **1028**
 regulation of, 1028–1029
 Urease, 132, 140F, 323, 470, 489T
 β -Ureidoisobutyrate, 1136F
 β -Ureidopropionate, 1136F
 β -Ureidopropionate, 1136F
 Ureotelic organism, 1025
 Urey, H., 32–33
 Uric acid, 1025, 1107, **1134–1135**
 Uricotelic organism, 1025
 Uridine, 83T, 1136F
 Uridine diphosphate glucose (UDP-glucose), 644
 Uridine monophosphate, *see* UMP
 Uridine phosphorylase, 1136F
 Uridine triphosphate, *see* UTP
 Uridylic acid, *see* UMP
 Uridyltransferase, 1071
 Urine, 1095
 Urkingdoms, 6
 URL (uniform resource locator), 194
 Urobilin, 1056, 1057
 Urobilinogen, 1056, 1057
 Urochordates, 13F
 Uronic acids, 364
 Uroporphyrinogen decarboxylase, 1053, 1054
 Uroporphyrinogen III, 1053, 1054
 Uroporphyrinogen III synthase, 1052–1054
 Uroporphyrinogen synthase, 1053
 UTP (uridine triphosphate), 96, **1118–1119**, 1265
 UV absorbance spectra, 90, 92, 92F, 285F
 UV absorption, gel electrophoresis and, 149
 UvrABC endonuclease, 1217
 UV radiation, 32, 104, 1214
uvrA gene, 1217

- UvrA protein, 1217
uvrB gene, 1217
 UvrB protein, 1217
uvrC gene, 1217
 UvrC protein, 1217
 UvrD protein, 1217, 1220
 Uyeda, K., 663
- V
 V_{\max} (maximal velocity), 489
 Vaccines, 205
 Vacuoles, 7F, 11, 11F
 Vagelos, R., 961
 Val, *see* Valine
 Val-AMS (5'-O'-[N-(L-valyl)sulfamoyl]adenosine), 1356–1357
- Valine:
 and α -amino- β -chlorobutyrate, 1408
 biosynthesis, 1075
 degradation, 1034–1039
 as essential amino acid, 1065T
 genetic codes for, 100T, 1343T
 in globular proteins, 246
 half-life, 1413T
 α helix/ β sheet propensity, 302T
 in native unfolded proteins, 283
 side chains, 70, 264T
 structure and general properties, 68T
 Valinomycin, 79, 748, 749F
 ValRS, 1356–1357
 Valyl-adenylate, 1356–1357
 5'-O'-[N-(L-valyl)sulfamoyl]adenosine (Val-AMS), 1356–1357
- VAMP (Vesicle Associated Membrane Protein), 441
 Vanadate, 758
 van der Waals distance, 41, 224F
 van der Waals forces, 260
 Van Duyne, G., 1244
 Vane, J., 994
 van Helmont, J.-B., 901
 van Leeuwenhoek, A., 4
 van Niel, C., 903
 Van Roey, P., 252
 van't Hoff plot, 59
 Variable arm, 1346
 Variant surface glycoprotein (VSG), 892
 Varmus, H., 705
 Varshavsky, A., 1413
 Vasopressin, 673T, 683, 784
 Vassilyev, D., 1268
 VAST (Vector Alignment Search Tool), 256T, 258
 vCJD (new variant Creutzfeldt-Jakob disease), 315
 V(D)J recombination, 1223
 Vector Alignment Search Tool, *see* VAST
 Vectorial nature, of active transport, 760
 Vectorial protons, 843
 Vectors:
 cloning, 104, **106–109**
 shuttle, 118
 Vegetative cell, 1284
 Velocity, of reactions, 483, 488, 489F
 Vennesland, B., 470, 471
 Venom, *see* Snake venom
 Venter, C., 176, 181, 182
 v-erbB oncogene, 705
 Verdine, G., 1216
 Versican, 373T
 Vertebrates, 13F. *See also* specific types
 collagen of, 235
 introns of, 1316
 open reading frames for, 182
 rRNA methylation sites of, 1328–1329
- Very low density lipoproteins, *see* VLDL
 Vesicle Associated Membrane Protein (VAMP), 441
 Vesicle formation, **428–440**
 Vesicle fusion, 430, **440–445**
 Vesicular stomatitis virus (VSV), 885
v-fos proto-oncogene, 705
vg gene (fruit fly), 25F
 Viagra, 699, 743
Vibrio cholerae, 576
 Villafranca, E., 724
 Villafranca, J., 1128
 Vimentin, 10F
 Vinograd, J., 1160
 Viox, 543, 999–1000
 Viral strand, bacteriophage M13, 1190
 Virions, 545
 Virus-based cloning vectors, 108
 Viruses, 3. *See also* specific viruses
 budding of, 466
 and cancer, 704–705
 DNA size, 94T
 in gene therapy, 122
 genetics of, 26–28, 27F, 28F
 life cycle, 27F
 RNA and DNA, 85
 and RNAi, 1324–1325
 and RNA polymerases, 96
 sedimentation coefficients, 154F
 vaccines for, 205
 Viscosity, 155
 Vitalism, 587
 Vitamins, 474, 474T, 993
 Vitamin A, 122, 252, 474, 943. *See also* Retinol
 Vitamin B₁, 474T, 618
 Vitamin B₂, 474T, 565F
 Vitamin B₆, 474T, 640, 1020
 Vitamin B₁₂, 474T, 956–957
 Vitamin C (ascorbic acid), 236, **364–365**, 364F
 Vitamin D, 474, 673T, **677–679**, 679F, 943
 Vitamin D₂, 678
 Vitamin D₃, 677–678
 Vitamin D-binding protein, 678
 Vitamin D intoxication, 678
 Vitamin E, 866, 943
 Vitamin K, 943
 Vitamin K₁, 922
 Vitamin K₂, 910
 Vitravene (Fomivirsen), 1403
v-Jun proto-oncogene, 705
 VLDL (very low density lipoproteins), 449, 456, 944
 and apoB-100, 1322
 and atherosclerosis, 458
 degradation, 452
 formation, 986
 properties, 449T
 transport in blood, 973
 Void volume, in gel filtration chromatography, 138
 Voltage-gated ion channels, **771–775**, 772F, 779
 von Euler, U., 993
 von Gierke's disease, 666, 1135
 von Liebig, J., 470
v-Ras proto-oncogene, 705
 VSG (variant surface glycoprotein), 892
v-sis oncogene, 705
 v-SNAREs (SNAP receptors), 441
 v-Src protein, 704–705
 VSV (vesicular stomatitis virus), 885
 VSV G-protein, 885, 887F
 V-type ATPases, 758
- W
 Wagner, G., 1216, 1378
 Wakil, S., 961
 Waksman, G., 1178, 1186
 Waksman, S., 1395
 Walker, J., 853, 857, 865
 Walker A motif, 855
 Walker B motif, 855
 Walsh, C., 562
 Walz, T., 431
 Wang, A., 1149
 Wang, J., 1162
 Warburg, O., 575, 595, 823, 864, 892
 Warfarin, 543F
 Warshel, A., 523
 Water, **40–45**
 activity of, 60
 dielectric constant and dipole moment, 43T
 in glycerophospholipids, 389T
 ionization constant, 47
 and nitrogen excretion, 1134
 and nonpolar solvents, 262–264, 263T
 proton mobility, 44–45
 reactant in Miller–Urey experiments, 32, 33
 solvent properties, **42–44**
 structure and interactions, **40–42**, 40F–42F
- Waterman, M., 200
 Water-oxidizing center (WOC), 916
 Water-soluble vitamins, 474
 Watson, H., 529
 Watson, J., 88, 182, 1146, 1173, 1363
 Watson–Crick base pairs, 89–90, 89F, 104, 1145–1146, **1154–1156**
 and DNA polymerase I, 1178–1180
 electronic complementarity, **1155–1156**
 geometric complementarity, **1154–1155**
 and Pol I, 1178–1180
 Watson–Crick structure:
 of B-DNA, **88–90**
 real DNA vs., 1146–1148
 Waxes, 387
 WD40 service motif, 433
 WD repeat, 987
 Weak acids, 46–48, 48F
 Weber, P., 686
 Web of Science, 35
 Weinberg, R., 1211
 Weis, W., 443, 444
 Weismann, A., 19
 Weiss, S., 1265
 Westbrook, E., 696
 Westerhoff, H., 863
 Western blot, 149
 Westheimer, F., 470, 471
 WGS, *see* Whole genome shotgun sequencing
 Whelan, W., 645
wh gene (fruit fly), 24, 25, 25F
 Whole genome shotgun sequencing (WGS), 180F, 181
 Wigley, D., 1231
 Wikipedia, 35
 Wikström, M., 844
 Wild type, 23–24
 Williams, C., 803, 1125
 Williams, J., 606
 Williams, L., 1148
 Williams, R., 728, 733
 Wilson, K., 1112
 Winged helix motif, 1291
 Withers, S., 524
 Wittinghofer, A., 711
 Wittmann, H.-G., 1365
 Wittmann-Liebold, B., 1365

- Wlodawer, A., 549
 Wobble pairing, 1360–1361, 1360F, 1360T
 WOC (water-oxidizing center), 916
 Woese, C., 6, 8
 Wolberger, C., 1293
 Wolcott–Rallison syndrome, 1400
 Wolfenden, R., 515, 1384
 Wolman's disease, 466
 Wolynes, P., 287
 Woodward, R. B., 975
 Wool, 233, 235
 Work, 53–54, 57–58
 World Wide Web, 195T, 256T
 Writting number, of superhelix, 1159–1160, 1159F
 Wunsch, C., 199
 Wüthrich, K., 243, 314
 WW domain, 708
 Wyman, J., 349
- X**
 Xanthine, 1112, 1130F
 Xanthine dehydrogenase, 1133
 Xanthine oxidase (XO), 1130F, 1132–1133
 Xanthomas, 457
 Xanthosine, 1130F
 Xanthosine monophosphate, *see* XMP
 X chromosome, 22–23
 Xenobiotics, 542
Xenopus borealis, 1283
Xenopus laevis, 1163
 Xenotransplantation, 121–122
 XerC, 1243
 XerD, 1243
Xeroderma pigmentosum (XP), 1217–1218, 1223, 1224
 X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside), 110
 Xhol, 105T
 X-linked phosphorylase kinase deficiency, 667
 XMP (xanthosine monophosphate), 1112, 1130F
XO, see Xanthine oxidase
XP, see Xeroderma pigmentosum
 XPA, 1217
 XPB, 1217, 1218
 XPC, 1217
 XPD, 1217, 1218
 XPE, 1217
- XPF, 1217
 XPG, 1217, 1218
 XPV, 1217
 X-ray crystallography:
 and double helix of DNA, 88, 88F
 of enzyme–substrate complexes, 470
 interpretation of, 241–245
 and NMR, 244
 protein dynamics studies, 306
 Xrcc4, 1223–1224
 Xrn1 exonuclease, 1327
 Xrn2 exonuclease, 1304
 Xu, W., 722
 Xu5P, *see* Xylulose-5-phosphate
 Xylitol, 365
 XylNAc residue (*N*-acetylxylosamine), 523
 Xylose, 597–598
 D-Xylose, 360F, 361
 D-Xylulose, 361
 Xylulose-5-phosphate (Xu5P), 892–894, 929
- Y**
 YACs (yeast artificial chromosomes), 108–109, 113, 114, 180–181
YADH, see Yeast alcohol dehydrogenase
 Yallow, R., 674
 Yang, W., 1222
 Yanoffsky, C., 1296, 1298, 1337
 Yarus, M., 1383
 Y chromosome, 23
 Yeast, 13F. *See also* *Saccharomyces cerevisiae* (Baker's yeast)
 20S proteasome, 1416F–1417F
 chromosome number, 19T
 cloning of, in host organism, 104. *See also* YACs [yeast artificial chromosomes]
 DNA size, 94T
 eEF1A-eEF1Bb, 1389F
 exon splicing, 1305–1306
 fermentation, 593–594, 616–619
 GCN2, 1400–1401
 genome sequencing, 176, 177T, 182
 introns in, 1316
 prions in, 315–316
 protein disulfide isomerase, 291–292, 291F
 RAD51, 1230
 Rpn11, 1418
 80S rRNA, 1371F
 rRNA methylation sites, 1328–1329
- shuttle vectors, 118
 snoRNAs, 1329
 transposons, 1244
 tRNA^{Ala}, 1345F, 1346
 tRNA^{Asp}, 1349
 tRNA^{Phe}, 1346, 1346F, 1348
 vesicle fusion, 440
Yeast alanine tRNA, 176
Yeast alcohol dehydrogenase (YADH), 618, 619
 geometric specificity, 472–473
 molecular mass, 140F
 stereospecificity, 470–472
Yeast artificial chromosomes, *see* YACs
Yellow mosaic virus protein, 153T
Yersinia pestis, 1420–1421
Y⁺ gene, 1261–1262, 1264
 Ylid, 616
 -yl suffix, 73
 Yokoyama, S., 1355–1357
 Yonath, A., 1365, 1366, 1376
 Yonetani, T., 354
 YopJ, 1420
 Yoshida, M., 857
 Yoshikawa, S., 841
 Young, W., 594
- Z**
 Z α , 1150–1151, 1150F
 Zalcitabine, 1208
 Zamecnik, P., 1345, 1362
 Zamore, Phillip, 1323
 ZAP-70, 1412, 1412F
 Z-DNA, 1146F, 1147F, 1148T, **1149–1151**
 Zero order reactions, 488
Z⁺ gene, 1261–1262, 1264
Zidovudine, see 3'-Azido-3'-deoxythymidine
 α Zigzag, 433
 Zimm, B., 94
 Zinc, 473, 511–513
 Zinc fingers, 1402
 Zipping, 443F
 Zocor, 990, 991F
 Zonal ultracentrifugation, 155, 156F, 159
 Zone electrophoresis, 146
 Z-scheme, 914, 915F, 917F
 Zwitterions, 67F, 70
 Zygote, 20
 Zymase, 594
 Zymogens, **537–538**

This page intentionally left blank

One- and Three-Letter Symbols for the Amino Acids^a

A	Ala	Alanine
B	Asx	Asparagine or aspartic acid
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glutamine or glutamic acid

^aThe one-letter symbol for an undetermined or nonstandard amino acid is X.

Thermodynamic Constants and Conversion Factors

Joule (J)

$$1 \text{ J} = 1 \text{ kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \quad 1 \text{ J} = 1 \text{ C} \cdot \text{V} \text{ (coulomb volt)}$$

$$1 \text{ J} = 1 \text{ N} \cdot \text{m} \text{ (newton \cdot meter)}$$

Calorie (cal)

$$1 \text{ cal heats 1 g of H}_2\text{O from 14.5 to 15.5}^{\circ}\text{C}$$

$$1 \text{ cal} = 4.184 \text{ J}$$

Large calorie (Cal)

$$1 \text{ Cal} = 1 \text{ kcal} \quad 1 \text{ Cal} = 4184 \text{ J}$$

Avogadro's number (N)

$$N = 6.0221 \times 10^{23} \text{ molecules} \cdot \text{mol}^{-1}$$

Coulomb (C)

$$1 \text{ C} = 6.241 \times 10^{18} \text{ electron charges}$$

Faraday (F)

$$1 \text{ F} = N \text{ electron charges}$$

$$1 \text{ F} = 96,485 \text{ C} \cdot \text{mol}^{-1} = 96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1}$$

Kelvin temperature scale (K)

$$0 \text{ K} = \text{absolute zero} \quad 273.15 \text{ K} = 0^{\circ}\text{C}$$

Boltzmann constant (k_B)

$$k_B = 1.3807 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$$

Gas constant (R)

$$R = Nk_B \quad R = 8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} \quad R = 1.9872 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

$$R = 0.08206 \text{ L} \cdot \text{atm} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

The Standard Genetic Code

First Position (5' end)	Second Position				Third Position (3' end)
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met ^a	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

^aAUG forms part of the initiation signal as well as coding for internal Met residues.

Some Common Biochemical Abbreviations^a

A	adenine	E4P	erythrose-4-phosphate
aa	amino acid	EPR	electron paramagnetic resonance
aaRS	aminoacyl-tRNA synthetase	ER	endoplasmic reticulum
ACAT	acyl-CoA:cholesterol acyltransferase	ESI	electrospray ionization
ACh	acetylcholine	EST	expressed sequence tag
AChE	acetylcholinesterase	ETF	electron-transfer flavoprotein
ACP	acyl-carrier protein	FAD	flavin adenine dinucleotide, oxidized form
ADA	adenosine deaminase	FADH [.]	flavin adenine dinucleotide, radical form
ADH	alcohol dehydrogenase	FADH ₂	flavin adenine dinucleotide, reduced form
AdoCbl	5'-deoxyadenosylcobalamin	FAS	fatty acid synthase
AdoMet	adenosylmethionine	FBP	fructose-1,6-bisphosphate
ADP	adenosine diphosphate	FBPase	fructose-1,6-bisphosphatase
ADPNP	adenosine-5'-(β , γ -imido)triphosphate	Fd	ferredoxin
AIDS	acquired immunodeficiency syndrome	FGF	fibroblast growth factor
AKAP	A-kinase anchoring protein	FH	familial hypercholesterolemia
ALA	δ -aminolevulinic acid	fMet	N-formylmethionine
AMP	adenosine monophosphate	FMN	flavin mononucleotide
AMPK	AMP-dependent protein kinase	FNR	ferredoxin-NADP ⁺ reductase
AMPPNP	adenosine-5'-(β , γ -imido)triphosphate	F1P	fructose-1-phosphate
ARS	autonomously replicating sequence	F2,6P	fructose-2,6-bisphosphate
ATCase	aspartate transcarbamoylase	F6P	fructose-6-phosphate
ATP	adenosine triphosphate	G	guanine
BAC	bacterial artificial chromosome	GABA	γ -aminobutyric acid
BChl	bacteriochlorophyll	Gal	galactose
BCKDH	branched chain α -keto acid dehydrogenase	GalNAc	N-acetylgalactosamine
BH ₄	5,6,7,8-tetrahydrobiopterin	GAP	glyceraldehyde-3-phosphate
bHLH	basic helix-loop-helix	GAP	GTPase activating protein
bp	base pair	GAPDH	glyceraldehyde-3-phosphate dehydrogenase
BPG	D-2,3-bisphosphoglycerate	GDH	glutamate dehydrogenase
BPheo	bacteriopheophytin	GDP	guanosine diphosphate
BPTI	bovine pancreatic trypsin inhibitor	GEF	guanine nucleotide exchange factor
C	cytosine	GK	glucokinase
CaM	calmodulin	Glc	glucose
CAM	crassulacean acid metabolism	GlcNac	N-acetylglucosamine
cAMP	3',5'-cyclic AMP	GLUT	glucose transporter
CAP	catabolite gene activator protein	GMP	guanosine monophosphate
CCV	clathrin-coated vesicle	gp	gene product
CD	circular dichroism	G1P	glucose-1-phosphate
CDK	cyclin-dependent protein kinase	G6P	glucose-6-phosphate
cdDNA	complementary DNA	G6PD	glucose-6-phosphate dehydrogenase
CDP	cytidine diphosphate	GPI	glycosylphosphatidylinositol
CDR	complementarity-determining region	gRNA	guide RNA
CE	capillary electrophoresis	GSH	glutathione
cGMP	3',5'-cyclic GMP	GSSG	glutathione disulfide
CGN	cis Golgi network	GTF	general transcription factor
Chl	chlorophyll	GTP	guanosine triphosphate
CM	carboxymethyl	HA	hemagglutinin
CMP	cytidine monophosphate	Hb	hemoglobin
CoA or CoASH	coenzyme A	HDL	high density lipoprotein
COP	coat protein	HPGRT	hypoxanthine-guanine phosphoribosyl transferase
CoQ	coenzyme Q (ubiquinone)	HIV	human immunodeficiency virus
COX	cyclooxygenase or cytochrome c oxidase	HMG-CoA	β -hydroxy- β -methylglutaryl-CoA
CPS	carbamoyl phosphate synthetase	hnRNA	heterogeneous nuclear RNA
CTD	C-terminal domain	HPETE	hydroperoxyeicosatetraenoic acid
CTP	cytidine triphosphate	HPLC	high-performance liquid chromatography
D	dalton	Hsp	heat shock protein
d	deoxy	HTH	helix-turn-helix
DAG	1,2-diacylglycerol	Hyl	5-hydroxylysine
dd	dideoxy	Hyp	4-hydroxyproline
DEAE	diethylaminoethyl	IDL	intermediate density lipoprotein
DG	sn-1,2-diacylglycerol	IF	initiation factor
DHAP	dihydroxyacetone phosphate	IgG	immunoglobulin G
DHF	dihydrofolate	IHP	inositol hexaphosphate
DHFR	dihydrofolate reductase	IMP	inosine monophosphate
DMF	N,N-dimethylformamide	IP ₃	inositol-1,4,5-trisphosphate
DMS	dimethyl sulfate	IPTG	isopropylthiogalactoside
DNA	deoxyribonucleic acid	IR	infrared
DNP	2,4-dinitrophenol	IS	insertion sequence
dNTP	2'-deoxynucleotide triphosphate	ISP	iron-sulfur protein
Dol	dolichol	ITP	inosine triphosphate
L-DOPA	L-3,4-dihydroxyphenylalanine	JAK	Janus kinase
dsDNA	double-stranded DNA	K _M	Michaelis constant
DUB	deubiquitinating enzyme	kb	kilobase pair
EF	elongation factor	kD	kilodalton
EGF	epidermal growth factor	KF	Klenow fragment
ELISA	enzyme-linked immunosorbent assay	LCAT	lecithin:cholesterol acyl transferase
EM	electron microscopy	LDH	lactate dehydrogenase
emf	electromotive force	LDL	low density lipoprotein

^aThe three-letter and one-letter abbreviations for the “standard” amino acid residues are given in Table 4-1.

LHC	light-harvesting complex	PtdIns	phosphatidylinositol
LT	leukotriene	PTK	protein tyrosine kinase
LX	lipoxin	PTP	protein tyrosine phosphatase
MALDI	matrix-assisted laser desorption/ionization	Q	ubiquinone (CoQ) or plastoquinone
Man	mannose	QH ₂	ubiquinol
MAPK	mitogen-activated protein kinase	QSAR	quantitative structure–activity relationship
Mb	myoglobin	r	ribo
MHC	major histocompatibility complex	RC	photosynthetic reaction center
miRNA	microRNA	RER	rough endoplasmic reticulum
MKK	MAP kinase kinase	RF	release factor <i>or</i> replicative form
mRNA	messenger RNA	RFLP	restriction-fragment length polymorphism
MS	mass spectrometry	RK	HMG-CoA reductase kinase
MurNAc	<i>N</i> -acetylmuramic acid	RNA	ribonucleic acid
NA	neuraminidase	RNAi	RNA interference
NAD ⁺	nicotinamide adenine dinucleotide, oxidized form	RNAP	RNA polymerase
NADH	nicotinamide adenine dinucleotide, reduced form	RNR	ribonucleotide reductase
NADP ⁺	nicotinamide adenine dinucleotide phosphate, oxidized form	R5P	ribose-5-phosphate
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form	RPC	reverse phase chromatography
NAG	<i>N</i> -acetylglucosamine	RRM	RNA-recognition motif
NAM	<i>N</i> -acetylmuramic acid	rRNA	ribosomal RNA
NANA	<i>N</i> -acetylneurameric (sialic) acid	RS	tRNA synthetase
NDP	nucleoside diphosphate	RSV	Rous sarcoma virus
NEM	<i>N</i> -ethylmaleimide	RT	reverse transcriptase
NER	nucleotide excision repair	RTK	receptor tyrosine kinase
NeuNAc	<i>N</i> -acetylneurameric acid	RuBisCO	ribulose-1,5-bisphosphate carboxylase–oxygenase
NMN	nicotinamide mononucleotide	RuBP	ribulose-1,5-bisphosphate
NMR	nuclear magnetic resonance	Ru5P	ribulose-5-phosphate
NOESY	nuclear Overhauser effect spectroscopy	S	Svedberg unit
NOS	nitric oxide synthase	SAM	<i>S</i> -adenosylmethionine
NRK	nonreceptor tyrosine kinase	SAR	structure–activity relationship
NSAID	nonsteroidal anti-inflammatory drug	SCAP	SREBP cleavage-activating protein
NSF	NEM-sensitive fusion protein	SCID	severe combined immunodeficiency disease
NTP	nucleotide triphosphate	SDS	sodium dodecyl sulfate
OEC	oxygen-evolving complex	SH2	Src homology domain 2
OMP	orotidine monophosphate	SH3	Src homology domain 3
ORF	open reading frame	siRNA	small interfering RNA
P or p	phosphate	SNAP	soluble NSF attachment protein
P _i	orthophosphate ion	SNARE	SNAP receptor
PAGE	polyacrylamide gel electrophoresis	snoRNA	small nucleolar RNA
PAP	poly(A) polymerase	snRNA	small nuclear RNA
PBG	porphobilinogen	snRNP	small nuclear ribonucleoprotein
PC	plastocyanin	SOD	superoxide dismutase
PCNA	proliferating cell nuclear antigen	S7P	sedoheptulose-7-phosphate
PCR	polymerase chain reaction	SR	SRP receptor
PDB	Protein Data Bank	SRE	sterol regulatory element
PDC	pyruvate dehydrogenase multienzyme complex	SREBP	SRE binding protein
PDE	phosphodiesterase	SRP	signal recognition particle
PDGF	platelet-derived growth factor	SSB	single-strand binding protein
PDI	protein disulfide isomerase	ssDNA	single-stranded DNA
PE	phosphatidylethanolamine	STAT	signal transducer and activator of transcription
PEP	phosphoenolpyruvate	STC	sequence-tagged connector
PEPCK	PEP carboxykinase	STS	sequence-tagged site
PFGE	pulsed-field gel electrophoresis	SV40	simian virus 40
PKF	phosphofructokinase	T	thymine
PG	prostaglandin	TAF	TBP-associated factor
2PG	2-phosphoglycerate	TBP	TATA box-binding protein
3PG	3-phosphoglycerate	TBSV	tomato bushy stunt virus
PGI	phosphoglucone isomerase	TCA	tricarboxylic acid
PGK	phosphoglycerate kinase	TGN	trans Golgi network
PGM	phosphoglycerate mutase	THF	tetrahydrofolate
PH	phenylalanine hydroxylase <i>or</i> pleckstrin homology	TIM	triose phosphate isomerase
Pheo	pheophytin	TLC	thin layer chromatography
PhK	phosphorylase kinase	TM	transmembrane
PIC	preinitiation complex	TMV	tobacco mosaic virus
PI3K	phosphoinositide 3-kinase	topo	topoisomerase
PIP ₂	phosphatidylinositol-4,5-bisphosphate	TPP	thiamine pyrophosphate
PK	pyruvate kinase	tRNA	transfer RNA
PKA	protein kinase A	TS	thymidylate synthase
PKB	protein kinase B	TTP	thymidine triphosphate
PKC	protein kinase C	U	uracil
PKU	phenylketonuria	UCP	uncoupling protein
PLC	phospholipase C	UDP	uridine diphosphate
PLP	pyridoxal-5'-phosphate	UDPG	uridine diphosphate glucose
pmf	proton motive force	UMP	uridine monophosphate
PMP	pyridoxamine-5'-phosphate	UTP	uridine triphosphate
PNP	purine nucleotide phosphorylase	UV	ultraviolet
Pol	DNA polymerase	V _{max}	maximal velocity
PP1	phosphoprotein phosphatase-1	VLDL	very low density lipoprotein
PP _i	pyrophosphate ion	XMP	xanthosine monophosphate
PPI	peptidyl prolyl cis–trans isomerase	XP	xeroderma pigmentosum
PrP	prion protein	Xu5P	xylulose-5-phosphate
PRPP	5-phosphoribosyl- α -pyrophosphate	YAC	yeast artificial chromosome
PS	photosystem	YADH	yeast alcohol dehydrogenase
