TABLE 21-1
Classes of Lipoprotein Particles

Class			Composition (weight percent)[a]				
			Surface components			Core lipids	
	Diameter (nm)	Density (g/ml)	Protein	Phospholipid	Cholesterol	Cholesteryl esters	Triacylglycerol
Chylomicrons	75–1200	0.93	2	7	2	3	86
VLDL	30–80	0.93–1.006	8	18	7	12	55
IDL	25–35	1.006–1.019	19	19	9	29	23
LDL	18–25	1.019–1.063	22	22	8	42	6
HDL2	9–12	1.063–1.125	40	33	5	17	5
HDL3	5–9	1.125–1.21	45	35	4	13	3
Lp (a); slow pre-β	25–30	1.04–1.09					

[a] Data from Havel, R. J., and Kane, J. P. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. II (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1841–1852, McGraw-Hill, New York.

These are averages and there is considerable variation.

TABLE 21-2
Properties of Major Plasma Apolipoproteins

Designation	No. residues	Mass (kDa)	Source	Function
A-I	243	29		Major HDL protein
A-II	—	17.4	Liver and intestine	
A-IV	376	44.5		
B-100	4536	513	Liver	VLDL formation; ligand for LDL receptor
B-48	2152	241	Intestine	Chylomicron formation ligand for liver chylomicron receptor
C-I	57	6.6		
C-II	79	8.9	Liver	Cofactor for lipoprotein lipase
C-III	79	8.8		
D	—	31	Many tissues	A lipocalin
E	299	34	Liver, VLDL	Ligand for LDL receptor
(a)	Variable			Ligand for liver chylomicron receptor

store phospholipid in eggs whether from nematodes, frogs, or chickens.[39] There is some sequence similarity to that of human apolipoprotein B-100.

2. Movement of Lipid Materials Between Cells

After the synthesis and release of chylomicrons into the lymphatic circulation, various exchange processes occur by which apolipoproteins, as well as enzymes and other proteins, may be added or removed. These very complex and incompletely understood phenomena are presented in simplified form in Fig. 21-1. Chylomicrons donate apolipoproteins of the A and C families to HDL particles which, in turn, donate apoE and may also return some apoC protein to the chylomicrons.

Both chylomicrons and VLDL particles undergo similar processes in the capillary blood vessels, where their triacylglycerols are hydrolyzed to glycerol and free fatty acids by **lipoprotein lipase**.[40–42a] This enzyme requires for its activity the apolipoprotein C-II which is present in the chylomicrons and VLDL particles. Lipoprotein lipase is also known as the "clearing

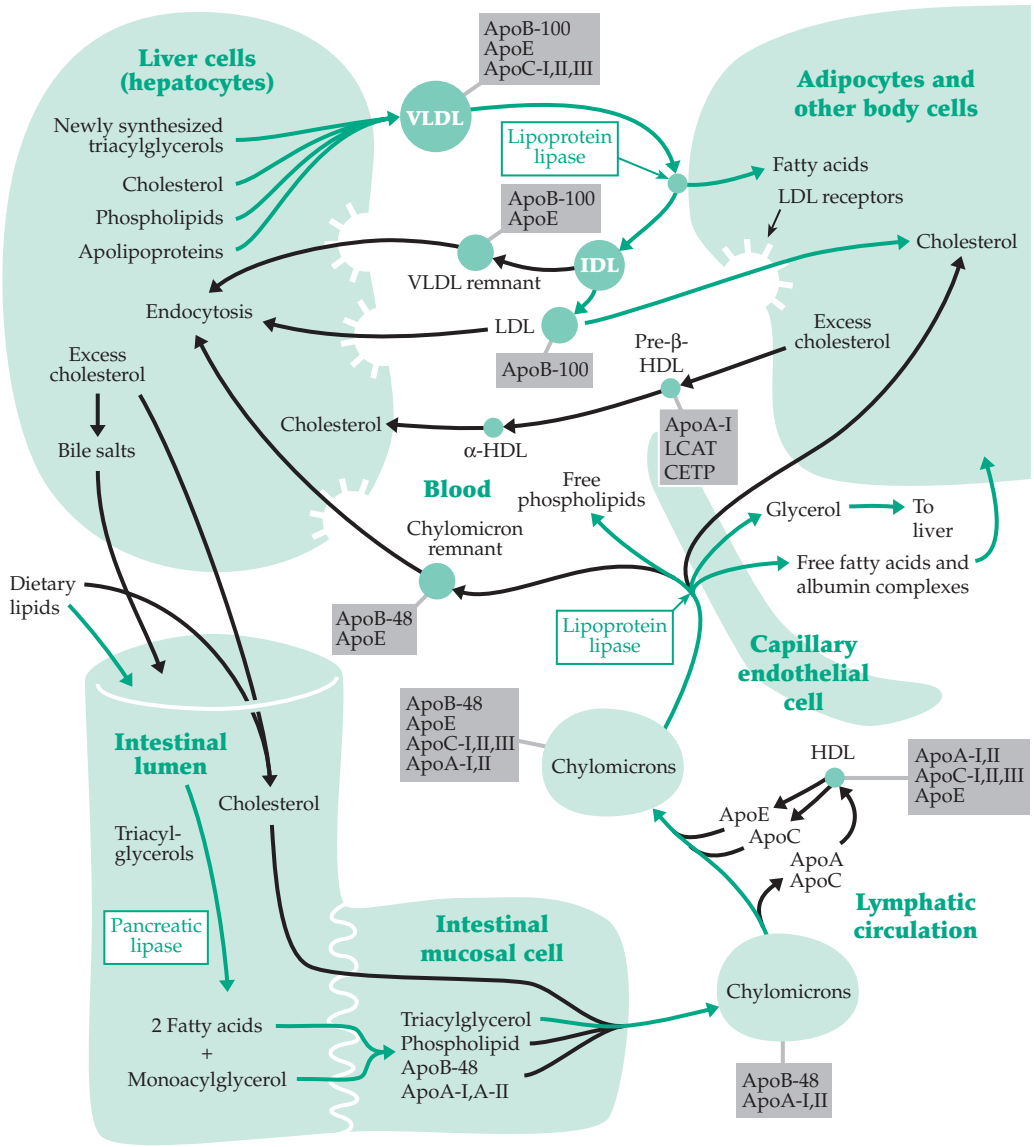


Figure 21-1 Movement of triacylglycerols from liver and intestine to body cells and lipid carriers of blood. VLDL; very low density lipoprotein which contains triacylglycerols, phospholipids, cholesterol, and apolipoproteins B, and C. IDL; intermediate density lipoproteins found in human plasma. LDL; low density lipoproteins which have lost most of their triacylglycerols. ApoB-100, etc., are apolipoproteins listed in Table 21-2. LCAT, lecithin: cholesterol acyltransferase; CETP, cholesteryl ester transfer protein (see Chapter 22).

factor" because it clears the milky chylomicron-containing lymph. It is secreted by adipocytes and other cells and becomes attached to heparan sulfate proteglycans on surfaces of capillary endothelial cells, a major site of its action.[43] Hereditary absence of functional lipoprotein lipase causes **chylomicronemia**, a massive buildup of chylomicrons in plasma.[41,44] The condition does not cause atherosclerosis but may lead to pancreatitis if not treated. Restriction of dietary fat to 20 g / day or less usually prevents problems. Naturally occurring mutations of lipoprotein lipase involving both the aspartate of the catalytic triad (p. 635)[45] and the flexible loop that covers the active site[46] have been discovered.

Both lipoprotein lipase and the less well understood **hepatic lipase** are related structurally to pancreatic lipase.[42,42b] In addition to hydrolysis of the triacylglycerols, the uptake of materials from lipoproteins probably involves shedding of intact phospholipids, perhaps as liposome-like particles.[40]

The free fatty acids and glycerol are taken up by mammalian tissue cells leaving the cholesterol and some of the phospholipids of the VLDL particles as LDL. In humans **intermediate density lipoproteins** (IDL) are formed initially, but some are converted to LDL later. Both LDL particles and the shrunken **chylomicron remnants** and **VLDL remnants** are taken up by endocytosis in coated pits and are degraded by body cells, principally of the liver.[47,48] The best known of these receptors is the 839-residue **LDL receptor,** which has a specific affinity for ApoB-100. The related **VLDL receptor** (apoE receptor) has a higher affinity for apoE[48–50] and may function in uptake of both VLDL and chylomicron remnants. The **LDL receptor-related protein** functions as a third lipoprotein receptor.[51] In addition, a series of **scavenger receptors**, found in abundance in macrophages, take up oxidized lipoproteins and other materials.[51,52] Scavenger receptor B1 (SR-B1), which is also found in liver cells, is involved in uptake of cholesterol from HDL particles by hepatocytes [53] (see also Chapter 22). Liver cells, and other cells as well, contain **lipocalins** and **fatty acid binding proteins** (Box 21-A) that help to carry these relatively insoluble acids to their destinations within the cells. Serum albumin (Box 2-A) is also a major carrier of free fatty acids.[53a] Within the adipocytes the fatty acids are reconverted to triacylglycerols. The low density (LDL) and high density (HDL) lipoproteins are involved primarily in transporting cholesterol to and from cells, a topic that is discussed in Chapter 22, Section D,2.

Fatty acids are carried to tissues for use in synthesis of triacylglycerols, phospholipids, and other membrane lipids. The mobilization of fatty acids from triacylglycerol stores and from cholesterol esters depends upon **hormone-sensitive lipase** (p. 635).[53b, 53c] This enzyme is activated by cAMP-dependent phosphorylation and moves from the cytoplasm to the surfaces of lipid droplets in response to catecholamines and other lipolytic hormones. Fatty acids are a major fuel for aerobic cells. Their conversion to acyl-CoA derivative and oxidation to CO_2 by beta oxidation (Fig. 17-1) and other pathways are discussed in Chapter 17 (pp. 939 – 950).

B. The Biosynthesis of Fatty Acids and Their Esters

The synthesis of fatty acids two carbon atoms at a time from acetyl-CoA has been considered in Chapter 17 and is outlined in Fig. 17-12. In this pathway, which resembles the β oxidation sequence in reverse, the products are saturated fatty acids with an even number of carbon atoms as shown in Fig. 21-2. In this section, we will consider some of the factors that lead to variations in the chain lengths and types of fatty acids.

1. Fatty Acid Synthases

Both bacteria and plants have separate enzymes that catalyze the individual steps in the biosynthetic sequence (Fig. 17-12). The fatty acyl group grows while attached to the small acyl carrier protein (ACP).[54–58] Control of the process is provided, in part, by the existence of isoenzyme forms. For example, in *E. coli* there are three different β-oxoacyl-ACP synthases. They carry out the transfer of any acyl primer from ACP to the enzyme, decarboxylate malonyl-ACP, and carry out the Claisen condensation (steps *b*, *e*, and *f* in Eq. 17-12)[58a – e] One of the isoenzymes is specialized for the initial elongation of acetyl-ACP and also provides feedback regulation.[58c] The other two function specifically in synthesis of unsaturated fatty acids.

In a few bacteria and protozoa and in higher animals the fatty acid synthase consists of only one or two multifunctional proteins. That from animal tissues contains six enzymes and an acyl carrier protein (ACP) domain as well. The human enzyme contains 2504 amino acid residues organized as a series of functional domains.[59–59b] Pairs of the 272-kDa chains associate to form 544-kDa dimers. The complex protein may have arisen via an evolutionary process involving fusion of formerly separate genes.[60] The enzyme contains an ACP-like site with a bound 4'-phosphopantetheine near the C terminus as well as a cysteine side chain near the N terminus in the second acylation site. Since the two –SH groups can be crosslinked by dibromopropanone,[61,62] an antiparallel linear arrangement of the two chains was proposed.[63–65] Locations of the six enzymatic activities in each chain are indicated on page 1187. According to this picture,

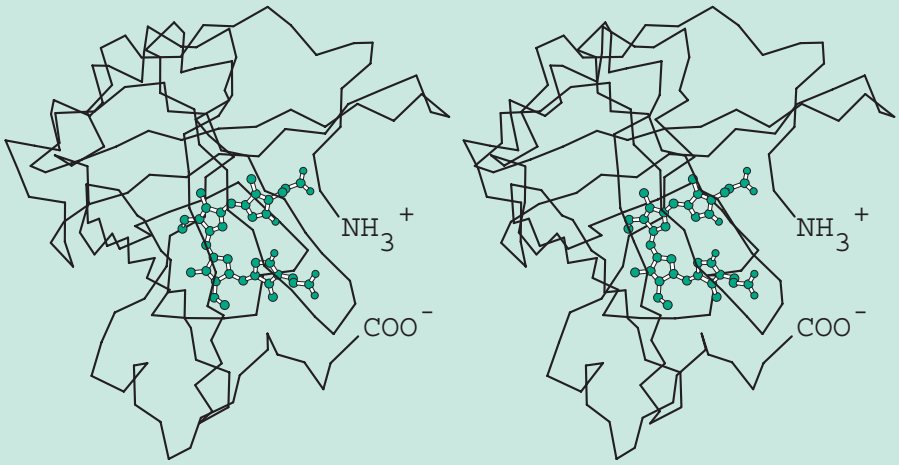
BOX 21-A LIPOCALINS, FATTY ACID-BINDING PROTEINS, AND LIPOPHORINS

Small hydrophobic molecules, which might easily bind in biologically undesirable ways, are chaperoned in animals, plants, and bacteria by binding proteins that provide hydrophobic cavities or crevices appropriate for holding these molecules in readily releasable forms. The **lipocalins**, most of which are extracellular proteins, have a conserved structural motif consisting of an 8-stranded β barrel arranged as two stacked orthogonal sheets with a C-terminal α helix that blocks one end. The other end is able to open and allow a small hydrophobic molecule to bind in the internal cavity[a–c] (see figure). Only three short amino acid sequences are conserved within a large family of lipocalins[b,d] which includes **plasma retinol-binding protein**,[e] mammalian **odorant-binding proteins**,[f] **α-lactalbumin**, **apolipoprotein D**,[a] and the blue biliverdin-binding protein **insecticyanin** of insect hemolymph.[g,h] Most lipocalins are soluble, but some such as the plasma **α1-microglobulin**,[i] which plays a role in the immune system (Chapter 31), have additional functions that require them to bind to other proteins or to cell surfaces.[j] The **gelatinase-associated lipocalin** of human neutrophils binds bacterially derived *N*-formylpeptides that act as chemotactic agents (Chapter 19) and induces release of materials from intracellular granules.[c] A few lipocalins have enzymatic activity. For example, **prostaglandin D synthase** is both an enzyme and a carrier of bile pigments and thyroid hormones.[k] Most lipocalins have been found in higher animals, but at least a few bacterial proteins belong to the family.[d] One is the 77-residue *E. coli* outer membrane lipoprotein.[a]

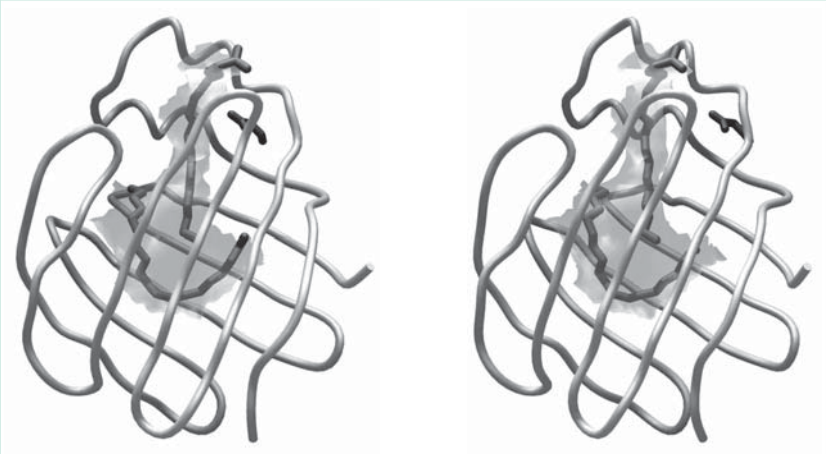
A related family of proteins are represented by **fatty acid-binding proteins**[l–o] and by the intracellular **retinol-** and **retinoic acid-binding proteins** (see also Box 22-A).[p] These are 10-stranded antiparallel β-barrels with two helices blocking an end (see Figure).

A third group of lipid-binding proteins have a four-helix bundle structure. They include the insect **lipophorins**, which transport diacylglycerols in the hemolymph (see main text), and nonspecific lipid carriers of green plants.[q] An 87-residue four-helix protein with a more open structure binds acyl-coenzyme A molecules in liver.[r]

A small 98-residue sterol-binding protein from



Stereoscopic view of an α-carbon model of an insecticyanin subunit with the bound biliverdin. The N and C termini are labeled NH_3^+ and COO^-, respectively. The positions of several amino acid residues are indicated. From Holden *et al.*[g] Courtesy of Hazel Holden.



Structure of a crystalline fatty acid-binding protein from liver with two molecules of bound oleate (dark rods). The lower molecule is more deeply embedded in the protein and more tightly bound than the second molecule, which is closer to the outer surface of the protein. Semitransparent grey marks the solvent-accessible surface of the binding cavity. An unknown molecule, perhaps butanoic acid (as modeled), binds also at top of the protein. See Thompson *et al.*[n] Courtesy of Leonard Banaszak.

BOX 21-A (continued)

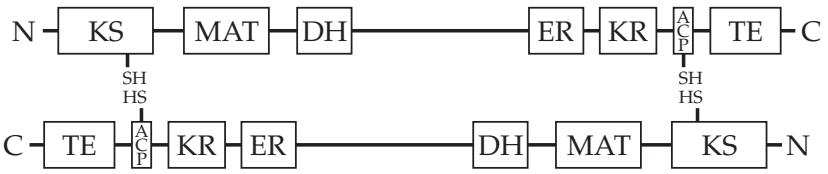
the fungus *Phytophthora*, an agriculturally important plant pathogen, has a very different folding pattern. The sterol binds into a cavity formed by six helices and two loops. The protein is not only a sterol carrier but an **elicitin**, which induces a defensive response in the invaded plant. The function of the protein for the invader may be to acquire sterols for the fungus, which is unable to synthesize them.[s]

Many larger lipid carrier proteins are known. The 476-residue plasma cholesteryl ester transfer protein is discussed briefly in Chapter 22. Plasma phospholipid transfer proteins are of similar size.[t,u] A 456-residue human phospholipid-binding protein interacts with the lipopolysaccharide of the surfaces of gram-negative bacteria (Fig. 8-30) and participates in the immune response to the bacteria. It has an elongated boomerang shape with two cavities, both of which bind a molecule of phosphatidylcholine. Other plasma lipid transfer proteins may have similar structures.[v]

[a] Bishop, R. E., Penfold, S. S., Frost, L. S., Höltje, J.-V., and Weiner, J. H. (1995) *J. Biol. Chem.* **270**, 23097–23103
[b] Flower, D. R., North, A. C. T., and Attwood, T. K. (1993) *Protein Sci.* **2**, 753–761
[c] Coles, M., Diercks, T., Muehlenweg, B., Bartsch, S., Zölzer, V., Tschesche, H., and Kessler, H. (1999) *J. Mol. Biol.* **289**, 139–157
[d] Flower, D. R., Sansom, C. E., Beck, M. E., and Attwood, T. K. (1995) *Trends Biochem. Sci.* **20**, 498–499
[e] Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990) *Proteins* **8**, 44–61
[f] Spinelli, S., Ramoni, R., Grolli, S., Bonicel, J., Cambillau, C., and Tagoni, M. (1998) *Biochemistry* **37**, 7913–7918
[g] Holden, H. M., Rypniewski, N. R., Law, J. H., and Rayment, I. (1987) *EMBO J.* **6**, 1565–1570
[h] Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. (1987) *J. Mol. Biol.* **198**, 499–513
[i] Åkerström, B., and Lögdberg, L. (1990) *Trends Biochem. Sci.* **15**, 240–243
[j] Bishop, R. E., and Weiner, J. H. (1996) *Trends Biochem. Sci.* **21**, 127
[k] Beuckmann, C. T., Aoyagi, M., Okazaki, I., Hiroike, T., Toh, H., Hayaishi, O., and Urade, Y. (1999) *Biochemistry* **38**, 8006 – 8013
[l] Sacchettini, J. C., and Gordon, J. I. (1993) *J. Biol. Chem.* **268**, 18399–18402
[m] Wiesner, S., Kurian, E., Prendergast, F. G., and Halle, B. (1999) *J. Mol. Biol.* **286**, 233–246
[n] Thompson, J., Winter, N., Terwey, D., Bratt, J., and Banaszak, L. (1997) *J. Biol. Chem.* **272**, 7140–7150
[o] Hohoff, C., Börchers, T., Rüstow, B., Spener, F., and van Tilbeurgh, H. (1999) *Biochemistry* **38**, 12229–12239
[p] Thompson, J. R., Bratt, J. M., and Banaszak, L. J. (1995) *J. Mol. Biol.* **252**, 433–446
[q] Heinemann, B., Andersen, K. V., Nielsen, P. R., Bech, L. M., and Poulsen, F. M. (1996) *Protein Sci.* **5**, 13–23
[r] Andersen, K. V., and Poulsen, F. M. (1992) *J. Mol. Biol.* **226**, 1131–1141
[s] Boissy, G., O'Donohue, M., Gaudemer, O., Perez, V., Pernollet, J.-C., and Brunie, S. (1999) *Protein Sci.* **8**, 1191–1199
[t] Wirtz, K. W. A. (1991) *Ann. Rev. Biochem.* **60**, 73–99
[u] Tall, A. (1995) *Ann. Rev. Biochem.* **64**, 235–257
[v] Beamer, L. J., Carroll, S. F., and Eisenberg, D. (1997) *Science* **276**, 1861–1864

the ACP domain of one chain would cooperate with the β-oxacyl synthase (KS) domain of the second chain. However, more recent studies indicate greater flexibility with the ACP, MAS, and KS domains of a single chain also being able to function together.[58,62] Animal fatty acid synthases produce free fatty acids, principally the C_{16} palmitate. The final step is cleavage of the acyl-CoA by a thioesterase, one of the six enzymatic activities of the synthase.

Yeast fatty acid synthase[66,67] has an $\alpha_6\beta_6$ structure where the 208-kDa α subunit contains the ACP-like site, the active site –SH, and three catalytic activities. The 220-kDa β subunit has five catalytic activities. The yeast enzyme contains the FMN thought to act as $FMNH_2$ in the second reduction step. As in bacteria, the products of the complex are molecules of acyl-CoA of chain lengths C_{14}, C_{16}, and C_{18}.[68]



Abbreviation	Enzymatic activity	Residue numbers
KS	β-Oxoacyl (ketoacyl) synthase	1 – 406
MAT	Malonyl and acetyl transferase	428 – 815
DH	Dehydratase	829 – 969
Central region	Structural core (?)	970 – 1629
ER	Enoyl reductase	1630 – 1850
KR	β-Oxoacyl (ketoacyl) reductase	1870 – 2100
ACP	Acyl carrier protein	2114 – 2190
TE	Thioesterase	2200 – 2505

Organization of eukaryotic fatty acid synthase. From Joshi *et al.*[61]

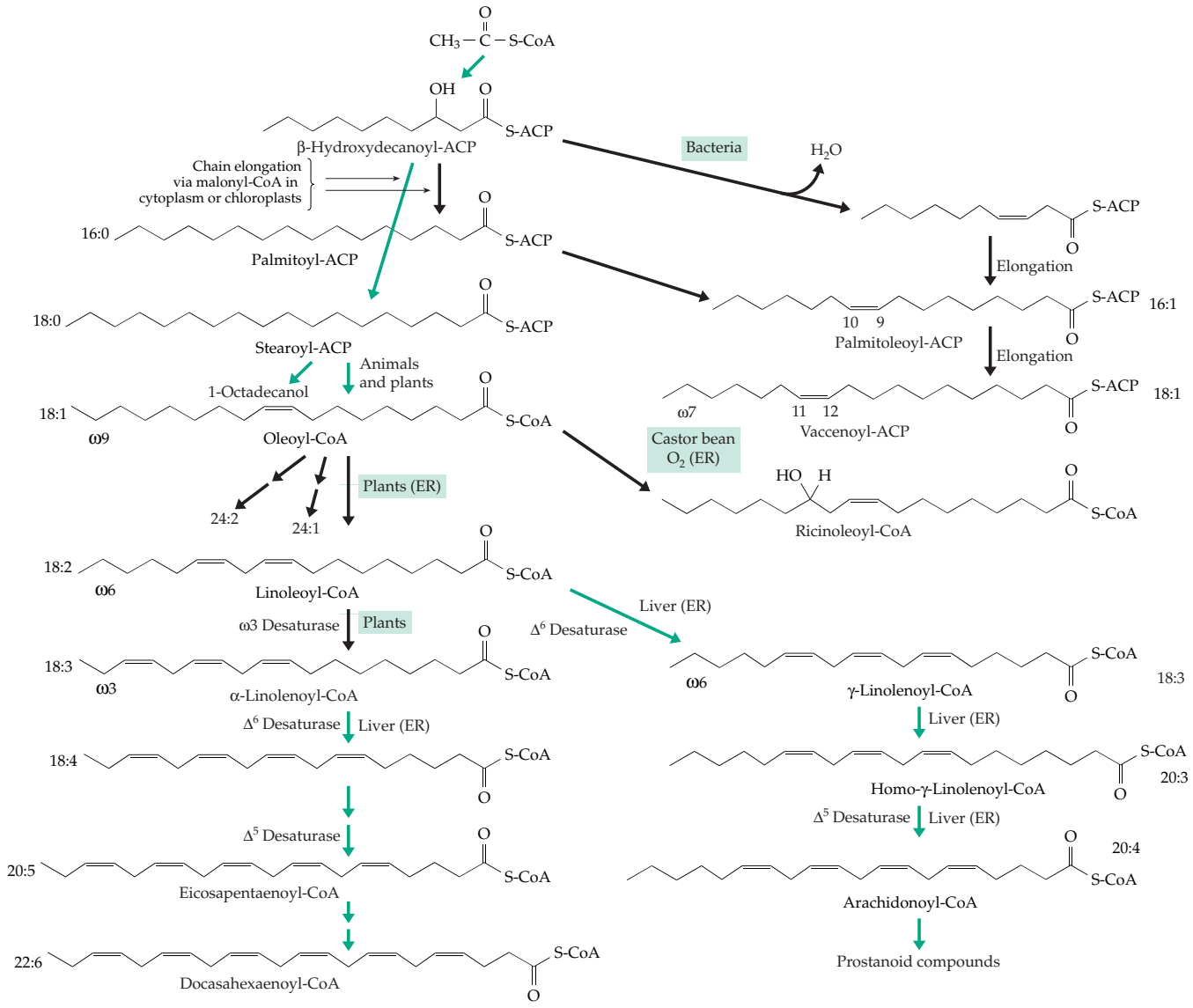


Figure 21-2 Some biosynthetic reactions of fatty acids. Green arrows indicate transformations carried out by the human body.

2. Control of Chain Length

The length of fatty acid chains is controlled largely by the enzymatic activity that releases the fatty acyl-CoA molecules or the free fatty acids from the synthase complex. In the animal enzymes the thioesterase, which is built into the synthase molecule, favors the release primarily of the 16-carbon saturated (16:0) palmitic acid. However, in mammary glands and in the uropygial glands (preen glands) of waterfowl shorter chain fatty acids predominate. These are released from the synthase by reaction with a second thioesterase, a 29-kDa protein[69,70] that catalyzes the otherwise premature release of shorter fatty acids. Cow's milk contains significant amounts of C_4-C_{14} acids as well as those with longer chains, whereas rabbit's milk contains largely C_8 to C_{10} fatty acids.[71]

In plants most biosynthesis occurs in the chloroplasts or in the protoplastids of seeds.[72–75] There are two different synthase systems in chloroplasts, one that forms primarily the 16:0 palmitoyl-ACP and the other the 18:0 stearoyl-ACP. Hydrolysis of the palmitoyl-ACP releases palmitate, one major product of chloroplasts. However, the stearoyl-ACP is desaturated to oleoyl-ACP[75a] before hydrolysis to free oleate or conversion to oleoyl-CoA. In many species oleic acid is almost the sole fatty acid exported by the chloroplasts. However, it undergoes a variety of modification reactions in the plant cytosol.

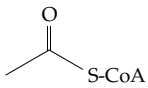
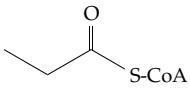
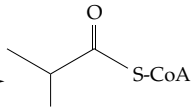
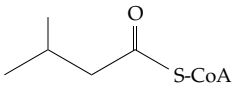
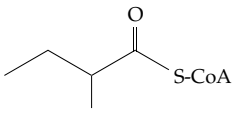
Plants, animals, and fungi all have fatty acid elon-

gation systems in the endoplasmic reticulum. Using malonyl-CoA and NADPH,[76,77] chain lengths of fatty acids may be increased to C_{20} to C_{26}. Elongation of fatty acids can also occur in mitochondria by reactions that are essentially the reverse of β oxidation. The only deviation from an exact reversal of oxidation is the use of NADPH as the reductant for enoyl-CoA reductase. Elongation of fatty acids in the *outer* membrane of mitochondria, followed by transport of the elongated chains into the mitochondria, may even constitute another shuttle for transport of reducing equivalents from NADH into mitochondria (Chapter 18, Section D).[78] Elongation reactions may also occur in peroxisomes.[78a]

3. Starter Pieces and Branches

Acetyl-CoA is most often the primer or starter piece for fatty acid synthesis, but butyryl-CoA is a better primer for rabbits. Butyryl-CoA arises from acetyl-CoA by a reversal of β oxidation, the necessary enzymes occurring in significant amounts in the cytosol.[79] If either acetyl-CoA or butyryl-CoA is the starter piece, chain elongation via malonyl-CoA (Fig. 18-12) leads to fatty acids with an even number of carbon atoms. However, degradation of the branched chain amino acids valine, isoleucine, and leucine creates a series of branched starter pieces (Table 21-3), whose utilization leads to formation of branched fatty acids of the iso and anteiso series. These are found in bacteria, in the lipids of tobacco and wool, in the "sound lens" of echo-locating porpoises,[80] and in many other materials.[81] Propionyl-CoA serves as an intermediate for introduction (via methylmalonyl-CoA) of branches at various other points in a fatty acid chain.[82] For example, 2*R*- and 4*R*-methylhexanoic acids, 2,4,6,8-tetramethyldecanoic acid, and a variety of other branched chain acids are esterified with long- chain alcohols (mainly 1-octadecanol) to form the waxes of the preen glands of ducks and geese.[83] The C_{32} **mycocerosic acid** of *Mycobacterium tuberculosis* is also formed using both malonyl-CoA and methylmalonyl-CoA for chain elongation.[84] This acid is present in mycobacterial cell walls esterified with long-chain diols (Box 21-C).[85]

TABLE 21-3
Starter Pieces for Biosynthesis of Fatty Acids

Starter piece	Fatty acid products
 Acetyl-CoA	Acid with even number of carbon atoms
 Propionyl-CoA	Acid with odd number of carbon atoms
Valine →  Isobutyryl-CoA	Iso series (even)
Leucine →  Isovaleryl-CoA	Iso series (odd)
Isoleucine → 	Anteiso series (odd)

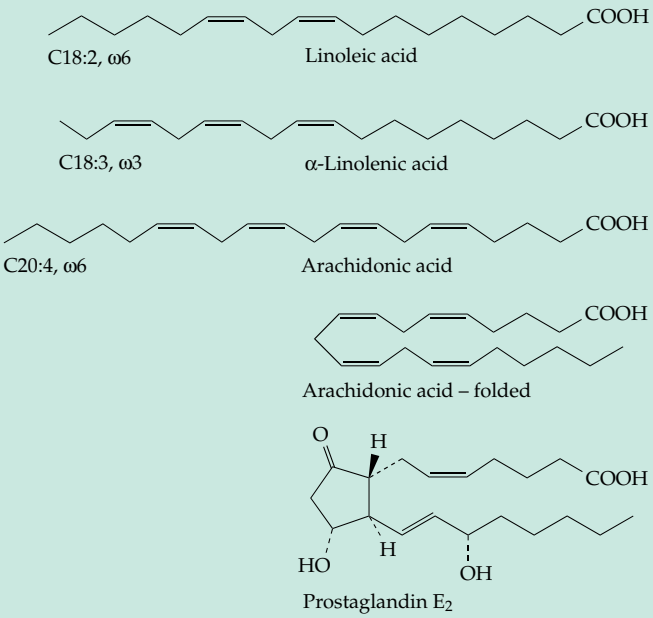
4. Synthesis by the Oxoacid Chain Elongation Process

The carbon skeleton of leucine is derived from that of valine by elongation of the corresponding oxoacid by a single carbon atom that is derived from acetyl-CoA (Fig. 17-18). At least in plants some branched-chain fatty acids of medium length are formed via the same oxoacid elongation process, which extends the chain one carbon atom at a time using 2-oxobutyrate as a starting compound. The same process can be used to form medium length straight-chain fatty acids (up to ~C_{12}) with either an odd or an even number of carbon atoms.[86]

2-Oxoglutarate can also serve as a starter piece for elongation by the oxoacid pathway. Extension by three carbon atoms yields 2-**oxosuberate** (Eq. 21-1). This dicarboxylate is converted by reactions shown in Eq. 24-39 into biotin and in archaebacteria into the coenzyme 7-mercaptoheptanoylthreonine phosphate (HTP), Eq. 21-1.[87] Lipoic acid is also synthesized from a fatty acid, the eight-carbon octanoate.[88,89] A fatty acid synthase system that utilizes a mitochondrial ACP may have as its primary function the synthesis of octanoate for lipoic acid formation.[90] The mechanism of insertion of the two sulfur atoms to form lipoate (Chapter 15) is uncertain. It requires an iron–sulfur protein[91,91a,b] and is probably similar to the corresponding process in the synthesis of biotin (Eq. 24-39)[92-93a] and in formation of HTP (Eq. 21-1). One component of the archaebacterial cofactor **methanofuran** (Chapter 15) is a tetracarboxylic acid that is formed from 2-oxoglutarate by successive condensations with two malonic acid units as in fatty acid synthesis.[94]

BOX 21-B THE ESSENTIAL FATTY ACIDS

In 1930, George and Mildred Burr reported that the C18:2 ($\Delta^{9,12}$) **linoleic acid**, a fatty acid of exclusively plant origin, cured a disease condition observed in rats raised on a highly purified fat-free diet.[a,b] These animals grew poorly, developed a scaly dermatitis, and suffered kidney damage and impaired fertility. The symptoms could be prevented if 1% of the dietary energy was provided by linoleic acid. This C18:2 fatty acid can be converted in animals into a series of other fatty acids by chain elongation and desaturation. All of this series have a double bond six carbon atoms from the $-CH_3$ terminus and form an ω6 (or n-6) family.[c]



$$\omega 6:\ C18{:}2 \rightarrow C18{:}3 \rightarrow C20{:}3 \rightarrow \underset{\text{Arachidonate}}{C20{:}4} \rightarrow C22{:}4 \rightarrow C22{:}5$$

$$\downarrow \qquad \downarrow$$

$$PGE_1,\ PGF_{1\alpha} \qquad PGE_2,\ PGF_{2\alpha}$$

The major known essential function of linoleic acid is conversion to the C20:4 ($\Delta^{5,8,11,14}$) **arachidonic acid**, the major precursor to prostaglandins and other prostanoid compounds (Section D). This conversion occurs in infants as well as adults,[d] but the rate may not always be adequate, and arachidonic acid is usually classified as an essential fatty acid. It is not clear whether linoleic acid has any essential role of its own. However, while arachidonic acid can be converted into the prostaglandins designated PGE_2 and $PGF_{2\alpha}$, linoleic acid can also give rise, via the C20:30 **dihomolinolenic acid**, to PGE_1 and $PGF_{1\alpha}$ (see Eq. 21-16). The C18:3 ($\Delta^{9,12,15}$) **α-linolenic acid**, another plant acid, can partially replace linoleic acid and can be converted into PGE_3 and $PGF_{3\alpha}$.

$$\omega 3:\ C18{:}3 \rightarrow C18{:}4 \rightarrow C20{:}4 \rightarrow C20{:}5 \rightarrow C22{:}5 \rightarrow C22{:}6$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

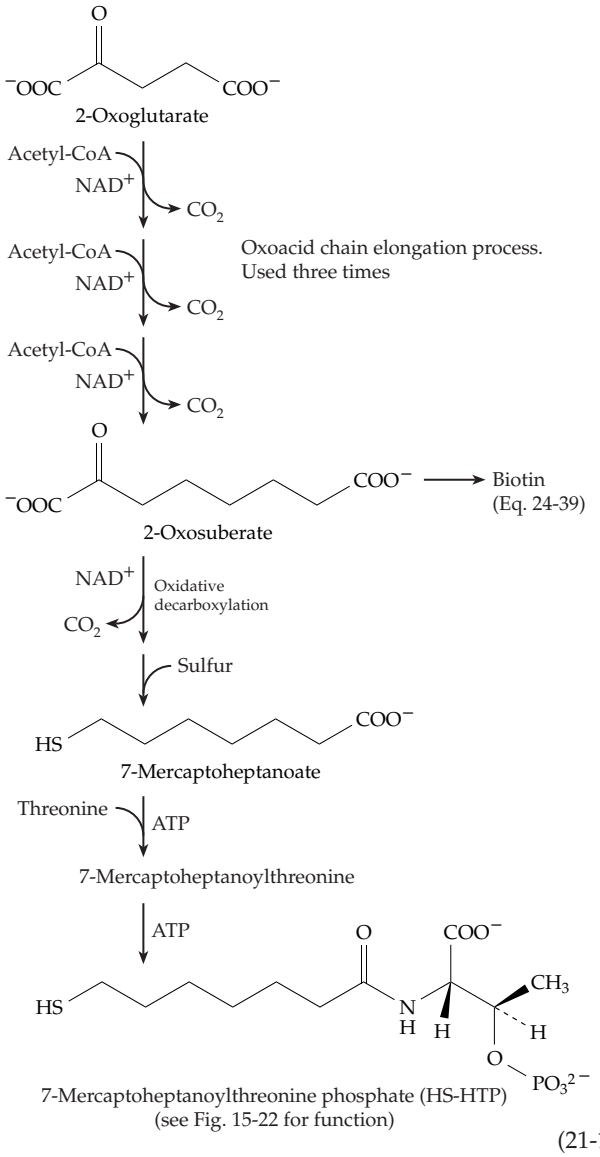
$$PGE_3,\ PGE_{3\alpha} \qquad PG_4 \qquad PG_5$$

Thus, it is not surprising that the three acids are not completely equivalent.[e–g]

Recent interest has focused on the C20:5 **eicosapentaenoic acid** (EPA) and the C22:6 **docosahexaenoic acids** (DHA). These ω3 (or n-3) polyunsaturated acids are formed from linolenic acid by marine algae and are found in fish oils.[h] The C22:5 and C22:6 acids can be converted to prostaglandins of the PG_4 and PG_5 series. DHA together with the ω6 C22:4 acid constitutes over 30% of the fatty acids in brain phospholipids. In the retina DHA accounts for over 60% of the total in the rod outer segments (Fig. 23-40). DHA may be formed in the human body from α-linolenic acid obtained from plant sources (Fig. 21-2). However, the rate of synthesis may be inadequate especially in infants and in persons of old age.[i–m] Deficiency of either EPA or DHA may lead to poor brain development during prenatal and infant life. Formation of new synapses between neurons as well as growth of new neurons in some areas of the brain is associated with thinking and with memory formation (Chapter 30). Therefore, a lack of the essential ω3 and ω6 fatty acids may contribute to mental deterioration in older adults. Eskimo populations, which consume large amounts of fish, have a very low incidence of coronary heart disease. An inverse relation between fish consumption and heart disease has also been demonstrated in other populations.[n] Inclusion of fish oils to 20–30% of total caloric intake in the diet causes a marked decrease in plasma triacylglycerols and very low density lipoproteins (VLDL).[o] This effect has been attributed to the altered composition of the prostanoid compounds known as thromboxanes and prostacyclins (PGI). For similar reasons the ω3 acids may have an antiinflammatory effect.[p,q] Administration of fish oil to patients with kidney disease has proved beneficial[r] and may decrease risk of some cancers.[s] However, a diet high in ω3 fatty acids has also been reported to increase cancer risk.[f] Long-chain ω3 fatty acids may protect against sudden death from heart disease.[t] They may promote lateral phase separation within membranes to form regions low in cholesterol (see references 95a and 119d).

BOX 21-B (continued)

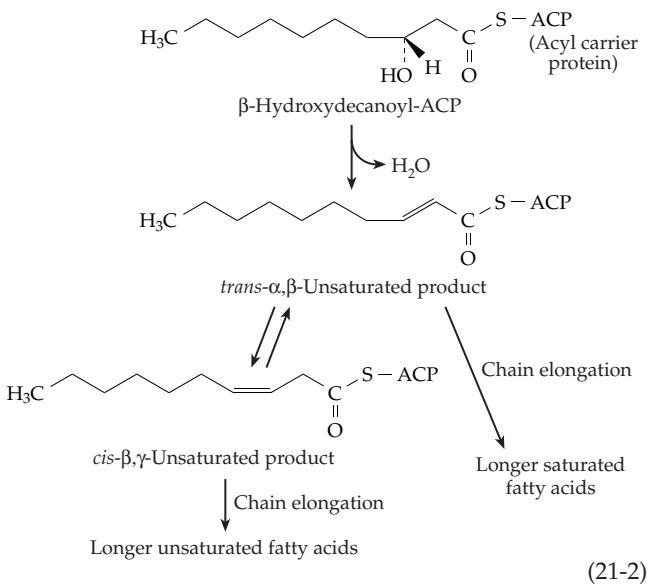
[a] Burr, G. O., and Burr, M. M. (1930) *J. Biol. Chem.* **86**, 587–621
[b] Burr, G. (1980) *Trends Biochem. Sci.* **5**, 28
[c] Horrobin, D. F., ed. (1990) *Omega-6 Essential Fatty Acids*, Wiley-Liss, New York
[d] Salem, N., JR, Wegher, B., Mena, P., and Uauy, R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 49–54
[e] Leat, W. M. F. (1981) *Trends Biochem. Sci.* **6**, IX–X
[f] Cave, W. T., Jr. (1991) *FASEB J.* **5**, 2160–2166
[g] Lands, W. E. M. (1992) *FASEB J.* **6**, 2530–2536
[h] Lees, R. S., and Karel, M., eds. (1990) *Omega-3 Fatty Acids in Health and Disease*, Dekker, New York
[i] Cho, H. P., Nakamura, M., and Clarke, S. D. (1999) *J. Biol. Chem.* **274**, 37335–37339
[j] Farkas, T., Kitajka, K., Fodor, E., Csengeri, I., Lahdes, E., Yeo, Y. K., Krasznai, Z., and Halver, J. E. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6362–6366
[k] Qiu, X., Hong, H., and MacKenzie, S. L. (2001) *J. Biol. Chem.* **276**, 31561–31566
[l] Kang, Z. B., Ge, Y., Chen, Z., Cluette-Brown, J., Laposata, M., Leaf, A., and Kang, J. X. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4050–4054
[m] Carper, J. (2000) *Your Miracle Brain*, Harper Collins Publ., New York
[n] Kromhout, D., Bosschieter, E. B., and Coulander, C. L. (1985) *N. Engl. J. Med.* **312**, 1205–1209
[o] Phillipson, B. E., Rothrock, D. W., Connor, W. E., Harris, W. S., and Illingworth, D. R. (1985) *N. Engl. J. Med.* **312**, 1210–1216
[p] Lee, T. H., Hoover, R. L., Williams, J. D., Sperling, R. I., Ravalese, J., III, Spur, B. W., Robinson, D. W., Corey, E. J., Lewis, R. A., and Austen, K. F. (1985) *N. Engl. J. Med.* **312**, 1217–1224
[q] Hwang, D. (1989) *FASEB J.* **3**, 2052–2061
[r] Donadio, J. V., Jr., Bergstralh, E. J., Offord, K. P., Spencer, D. C., and Holley, K. E. (1994) *N. Engl. J. Med.* **331**, 1194–1199
[s] Hilakivi-Clarke, L., Clarke, R., Onojafe, I., Raygada, M., Cho, E., and Lippman, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9372–9377
[t] Albert, C. M., Campos, H., Stampfer, M. J., Ridker, P. M., Manson, J. E., Willett, W. C., Ma, J. (2002) *N. Engl. J. Med.* **346**, 1113 – 1118 See also **347**, 531 – 533



(21-1)

5. Unsaturated Fatty Acids

Fatty acids containing one or more double bonds provide necessary fluidity to cell membranes[95,95a] and serve as precursors to other components of cells. Significant differences in the methods of introduction of double bonds into fatty acids are observed among various organisms. Bacteria such as *E. coli* that can live anaerobically often form **vaccenic acid** as the principal unsaturated fatty acid. It is formed by chain elongation after introduction of a cis double bond at the C_{10} stage of synthesis. The bacteria possess a **β-hydroxydecanoyl thioester dehydratase,** which catalyzes elimination of a β-hydroxyl group to yield primarily the *cis*-β,γ rather than the *trans*-α,β-unsaturated product (Eq. 21-2).[96] The mechanism may resem-



(21-2)

ble that of enoyl hydratase (Eq. 13-7), the indicated *trans*-α,β-unsaturated intermediate (enzyme bound) being isomerized to the *cis*-β,γ-unsaturated product through an allylic rearrangement. The product can then be elongated to the C_{16} **palmitoleoyl-ACP** and C_{18} **vaccenoyl-ACP** derivatives (Fig. 21-2, right side). However, dehydration of β-hydroxydecanoyl-ACP lies at a branch point in the biosynthetic sequences. The *trans*-α,β-unsaturated fatty acyl compound lies on the usual route of chain elongation to palmitoyl-CoA (left side, Fig. 21-2).

In higher plants, animals, protozoa, and fungi, saturated fatty acids are acted upon by **desaturases** to introduce double bonds, usually of the cis (Z) configuration. The substrates may be fatty acyl-ACP, fatty acyl-CoA molecules, membrane phospholipids,[97] or glycolipids.[98] The Δ^9 desaturase, isolated from liver or from yeast, converts stearoyl-CoA to oleoyl-CoA (Eq. 21-3).[99–102] This membrane-associated enzyme system utilizes NADH as a reductant, passing electrons via cytochrome b_5 reductase and cytochrome b_5 itself to the desaturase.[103,104] The pro-*R* hydrogens are removed at both C-9 and C-10.

$$O_2 + 2\,H^+ + 2\,e^- + \text{stearoyl-CoA} \rightarrow \text{oleoyl-CoA} + 2\,H_2O \qquad (21\text{-}3)$$

In plants a similar enzyme catalyzes formation of the first double bond in a fatty acyl group converting stearoyl-ACP into oleoyl-ACP in the chloroplasts.[72,75a,105–108] The soluble Δ^9 stearoyl-ACP desaturase has a diiron-oxo active site (Fig. 16-20, B, C).[109,110] Electrons are donated from light-generated reduced ferredoxin (see Chapter 23). In addition to the Δ^9 desaturase both plants and cyanobacteria usually desaturate C_{18} acids also at the Δ^{12} and Δ^{15} positions and C_{16} acids at the Δ^7, Δ^{10}, and Δ^{13} (ω3) positions.[111,112] Desaturation of oleate occurs primari-

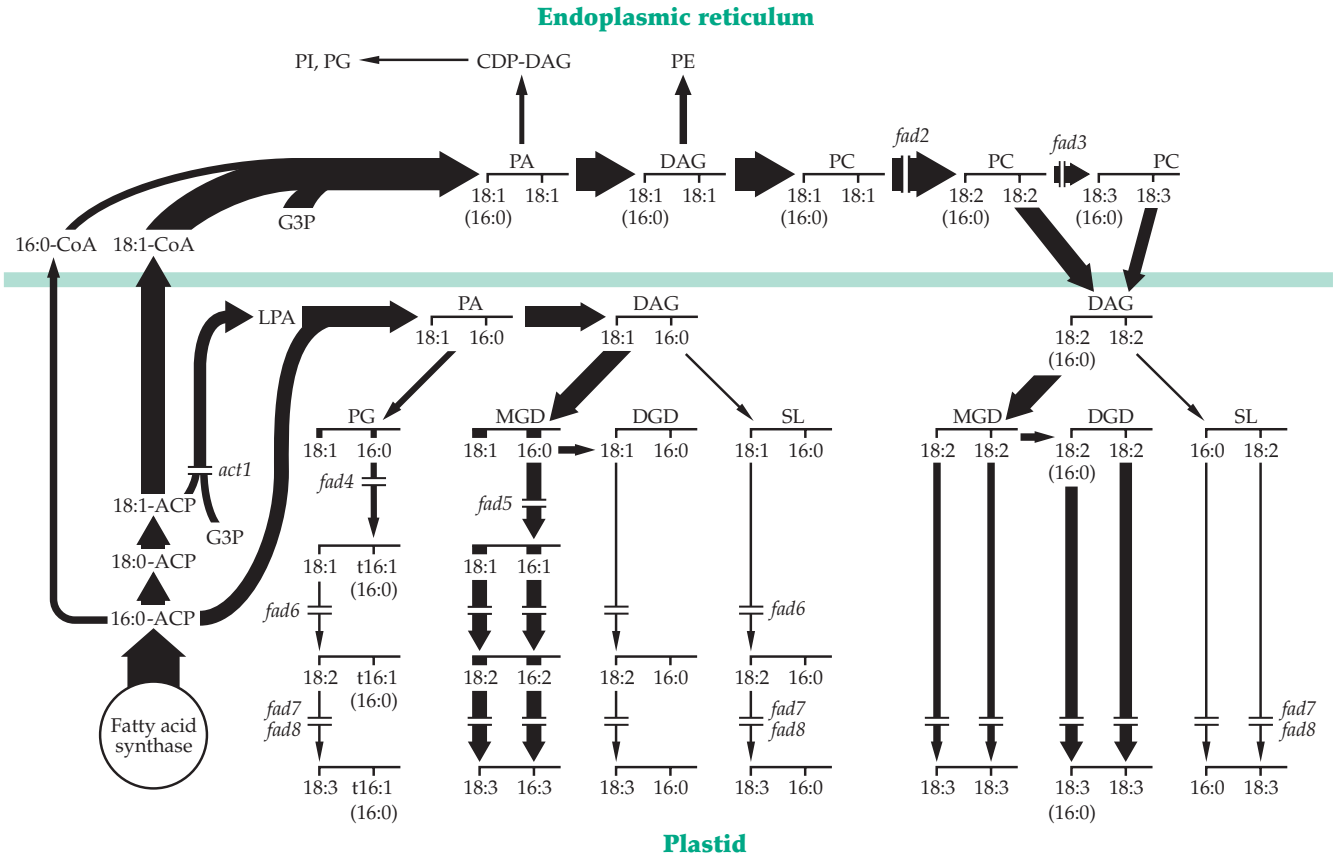


Figure 21-3 Major pathways of synthesis of fatty acids and glycerolipids in the green plant *Arabidopsis*. The major site of fatty acid synthesis is chloroplasts. Most is exported to the cytosol as oleic acid (18:1). After conversion to its coenzyme A derivative it is converted to phosphatidic acid (PA), diacylglycerol (DAG), and the phospholipids: phosphatidylcholine (PC), phosphatidylinositol (PI), phosphatidylglycerol (PG), and phosphatidylethanolamine (PE). Desaturation also occurs, and some linoleic and linolenic acids are returned to the chloroplasts. See text also. From Sommerville and Browse.[106] See also Figs. 21-4 and 21-5. Other abbreviations: monogalactosyldiacylglycerol (MGD), digalactosyldiacylglycerol (DGD), sulfolipid (SL), glycerol 3-phosphate (G3P), lysophosphatidic acid (LPA), acyl carrier protein (ACP), cytidine diphosphate-DAG (CDP-DAG).

ly in the ER after conversion of the free acid to its coenzyme A derivative or to phosphatidylcholine. Consecutive introduction of two double bonds forms linoleoyl-CoA (18:2, $\Delta^{9,12}$) and linolenoyl-CoA (18:3, $\Delta^{9,12,15}$; see Fig. 21-3).[108] All double bonds are cis. The membrane lipids of chloroplasts contain both linoleic and linolenic acids, which have apparently been returned to the chloroplasts from the cytosol[72] as indicated in Fig. 21-3. Plants grown at colder temperatures have a higher content of these trienoic acids than those grown at higher temperature.[72a]

The origin of ricinoleic acid, an abundant constituent of castor beans, is also shown in Fig. 21-2. It is formed by an **oleate hydroxylase** that has an amino acid sequence similar to those of oleate desaturases.[113] Both hydroxylation and desaturation are reactions catalyzed by diiron centers.[114] Other fatty acid hydroxylases act on the alpha[115] and the omega positions. The latter are members of the cytochrome P450 family.[116,117]

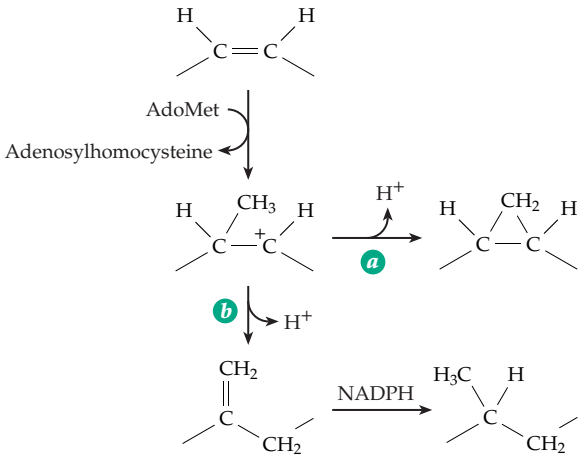
The conversion of oleoyl-CoA to linoleoyl-CoA is accomplished by some insects[118] but does not take place in most animals. As a result of this biosynthetic deficiency, polyunsaturated fatty acids such as linoleic, linolenic, and the C_{20} arachidonic acid are necessary in the diet (Box 21-B). One essential function of linoleic acid is to serve as a precursor of **prostaglandins** and related **prostanoids** (Section D). Dietary linoleate is converted to its CoA derivative and then by sequential Δ^6 desaturation,[119] elongation, and then Δ^5 desaturation, to the 20:4 ($\Delta^{5,8,11,14}$) arachidonoyl-CoA (Fig. 21-2, lower right). These acids are referred to as ω6 because of the position of the last double bond. Linolenic acid can be converted in an analogous fashion to the CoA derivative of the 20:5 ($\Delta^{5,8,11,14,17}$ ω6) eicosapentaenoic acid (**EPA**). The 22:6 docasahexaenoic acid (**DHA**; Fig. 21-2) is apparently formed by elongation of the 22:5 acyl-CoA to 24:5, desaturation, transfer to a peroxisome or mitochondrion, and β oxidation to shorten the chain.[95a]

These acids are very important in human nutrition.[119a–d] (See also Box 21-B.) In the absence of adequate essential fatty acids oleate is desaturated and elongated in a similar sequence to the unusual 20:3 ($\Delta^{5,8,11}$ ω9) acid. Vertebrate tissues also carry out desaturation at the Δ^4 and Δ^3 positions.[111] Lepidoptera, which synthesize a great diversity of pheromones, are rich in unusual desaturases such as the gland-specific acyl-CoA Δ^{11} desaturase of cabbage looper moths.[120]

Polyunsaturated fatty acids, containing 4, 5, or 6 double bonds and chain lengths of up to C_{36}, are found in phosphatidylcholine of vertebrate retinas.[121] Although the double bonds are rarely in conjugated positions in food fats and in animal bodies, some plants convert oleic or linoleic acid into fatty acids with as many as three or four conjugated double bonds.[121a] **Conjugated linoleic acid** (9*c*, 11*t* 18:2) can be formed from 11-*trans*-octadecenoate in the human body.[121b] This compound is also found in meat and dairy products. It has been reported to have anticancer properties[121c] and may be another beneficial dietary constituent. An isomerase isolated from red algae converts polyunsaturated acids into forms with conjugated double bonds. For example, arachidonic acid (5*Z*, 8*Z*, 11*A*, 14*Z*) is converted to (5*Z*, 7*E*, 9*E*, 14*Z*)-eicosatetraenoic acid.[122,123]

6. Cyclopropane Fatty Acids and Mycolic Acids

Fatty acids containing one or more cyclopropane rings are present in many bacteria (p. 381).[124,125] The extra carbon of the cyclopropane ring is added from *S*-adenosylmethionine (AdoMet) at the site of a cis double bond in a fatty acyl group of a phosphatidylethanolamine molecule in a membrane (Eq. 21-4).[126,126a] The same type of intermediate carbocation can yield either a cyclopropane fatty acid (Eq. 21-4, step *a*) or a methenyl fatty acid (Eq. 21-4, step *b*). The latter can be reduced to a branched fatty acid. This is an alternative way of introducing methyl branches that is used by some bacteria.[127]



(21-4)

Mycobacteria are rich in cyclopropane-containing fatty acids. These **mycolic acids** are major components of the cell walls and may account for 30% of the dry weight of the cells.[128] The most abundant mycolic acid of *M. tuberculosis* consists of C_{52} fatty acid containing two cyclopropane rings joined via a Claisen-type condensation with a C_{26} carboxylate fatty acid (Eq. 21-5). A similar mycolic acid formed by *M. smegmatis* has double bonds instead of cyclopropane rings as indicated below Eq. 21-5.[128,129] There are other variations. In place of a double bond or cyclopropane group there may be –OH, $-OCH_3$, C=O, epoxide, or CH_3[127]

Cyclopropane fatty acids are catabolized via β oxidation,[130] which is modified as in Eq. 21-6 when the chain degradation reaches the cyclopropane ring. The

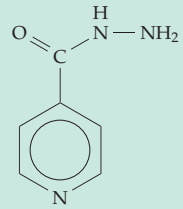
BOX 21-C TUBERCULOSIS

As many as one-third of the inhabitants of the earth are infected by *Mycobacterium tuberculosis*. For most the infection is dormant, but in some the slow-growing bacteria cause a progressive and deadly destruction of the lungs. There are still about three million deaths annually, and *M. tuberculosis* now, as in the past, kills more people then any other pathogen.[a,b] The development of drug-resistant strains of the bacterium and the threat of a worldwide resurgence of tuberculosis[c,d] has spurred new efforts to understand the unusual metabolism of mycobacteria and to develop new drugs. The complete 4.4 million base-pair sequence of the circular genome is known.[c] An unusually large fraction of the genes encode enzymes involved in synthesis and breakdown of lipids, including the synthesis of **mycolic acids** (Eq. 21-5) which are characteristic of mycobacteria.

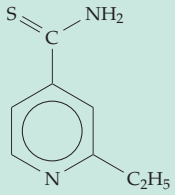
The mycobacterial cell wall, discussed in Chapter 8, contains mycolic acids bound covalently at the nonreducing ends of arabinogalactans that are attached to the inner peptidoglycan layer,[e] as well as phosphatidylinositol-anchored lipoarabinomannans. Other unusual lipids that are also present and account for some of the difficulty of treatment with antibiotics include esters of **mycocerosic acid** with long-chain diols known as phenolphthiocerols and phthiocerols.[f,fg]

Streptomycin (Boxes 20-B, 20-H) was introduced into clinical use against tuberculosis in about 1943. However, resistant mutants always survived until newer drugs were developed. Isonicotinylhydrazide (**isoniazid**) is especially effective in combinations with suitable antibiotics and other drugs.[g] The four-drug combination isoniazid, rifampicin (Box 28-A), pyrazinamide, and ethambutol is often used. Nevertheless, bacteria resistant to all of these have developed.

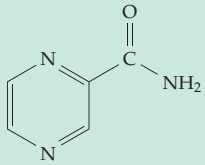
Although isoniazid has been in use for about 45 years, the enzyme that it inhibits has been recognized only recently. It is a specific NADH-dependent **enoyl reductase** involved in synthesis of mycolic acids.[h,i] The isoniazid must be activated by action of a bacterial catalase-peroxidase.[j,k] This enzyme may convert the drug to a reactive radical that combines with a NADH-derived radical to form an adduct in the active site of the enzymes. One possible reaction sequence follows.[h] However, the mechanisms are not clear.



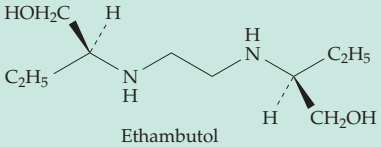
Isonicotinylhydrazide (Isoniazid)



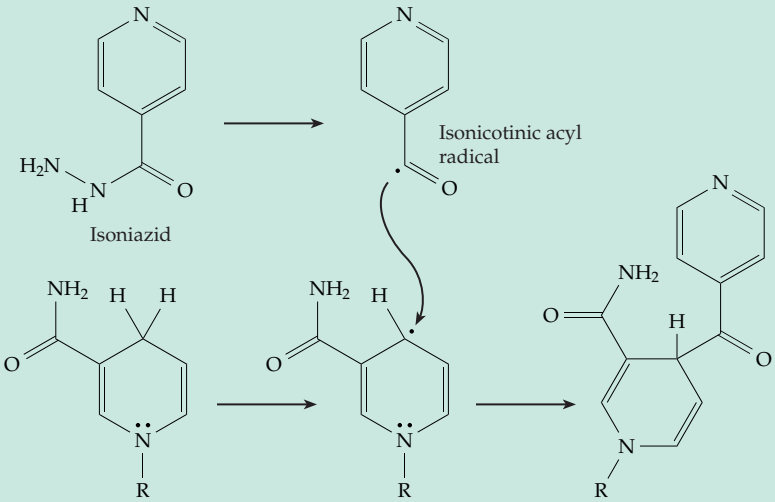
Ethionamide

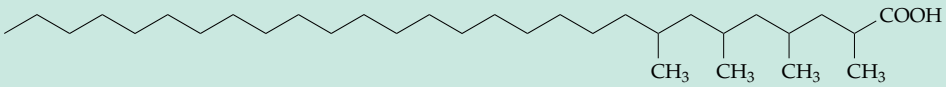


Pyrazinamide



Ethambutol

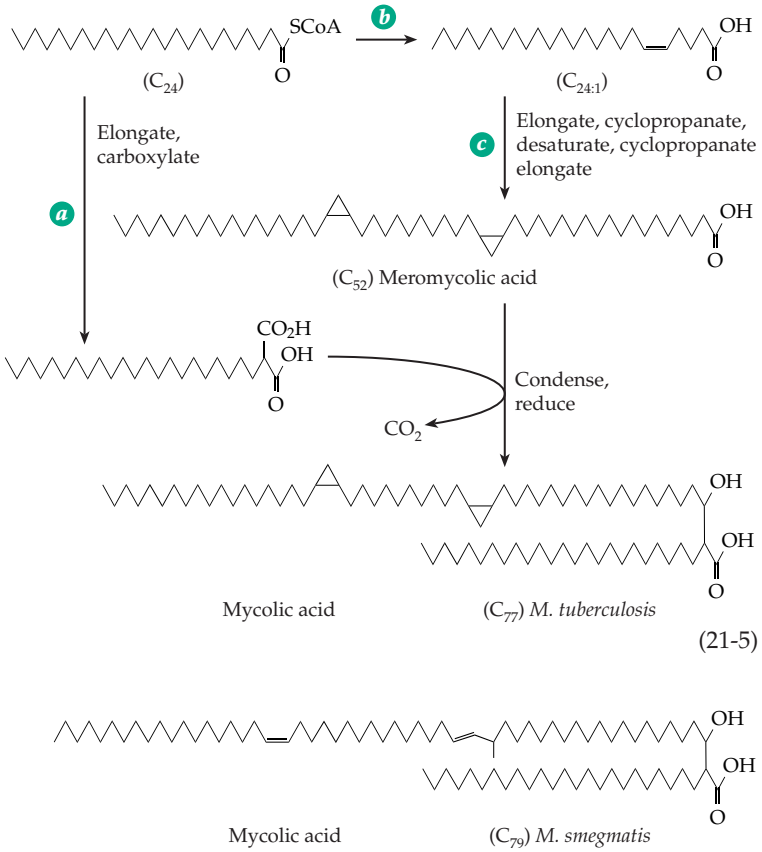




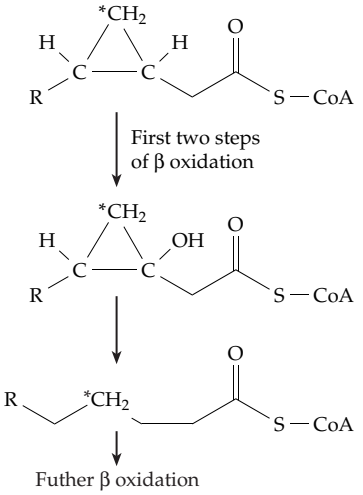
2*R*,4*R*,6*R*,8*R*-Tetramethyloctacosanoic acid (mycocerosic acid)

BOX 21-C (continued)

[a] Young, D. B. (1998) *Nature (London)* **393**, 515–516
[b] Venisse, A., Rivière, M., Vercauteren, J., and Puzo, G. (1995) *J. Biol. Chem.* **270**, 15012–15021
[c] Cole, S. T., and and 41 other authors. (1998) *Nature (London)* **393**, 537–544
[d] Iseman, M. D. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2428–2429
[e] Scherman, M. S., Kalbe-Bournonville, L., Bush, D., Xin, Y., Deng, L., and McNeil, M. (1996) *J. Biol. Chem.* **271**, 29652–29658
[f] Fitzmaurice, A. M., and Kolattukudy, P. E. (1998) *J. Biol. Chem.* **273**, 8033–8039
[fg] Patterson, J. H., McConville, M. J., Haites, R. E., Coppel, R. L., and Billman-Jacobe, H. (2000) *J. Biol. Chem.* **275**, 24900–24906
[g] Blanchard, J. S. (1996) *Ann. Rev. Biochem.* **65**, 215–239
[h] Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs, W. R. J., and Sacchettini, J. C. (1998) *Science* **279**, 98–102
[i] Baldock, C., Rafferty, J. B., Stuitje, A. R., Slabas, A. R., and Rice, D. W. (1998) *J. Mol. Biol.* **284**, 1529–1546
[j] Sherman, D. R., Mdluli, K., Hickey, M. J., Arain, T. M., Morris, S. L., Barry, C. E., III, and Stover, C. K. (1996) *Science* **272**, 1641–1643
[k] Wengenack, N. L., Lopes, H., Kennedy, M. J., Tavares, P., Pereira, A. S., Moura, I., Moura, J. J. G., and Rusnak, F. (2000) *Biochemistry* **39**, 11508–11513



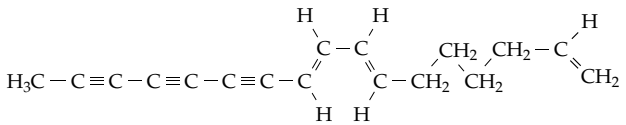
(21-5)



(21-6)

ring opening of the cyclopropanol derivatives occurs readily, even with mild nonenzymatic acid–base catalysis.

Another alteration of unsaturated fatty acids is the formation of acetylenic groups (–C≡C–). This apparently occurs by dehydrogenation of –CH=CH–. Examples of naturally occurring acetylenes are **crepenynic acid** (p. 381), **alloxanthin** (p. 1240), and the following remarkable hydrocarbon from the common cornflower *Centaurea cyanus*[131]:



7. The Lipids of Skin and Other Surfaces

Special fatty materials are often secreted to form external surfaces of organisms.[81,132] An example already mentioned is the secretion of the uropygial glands (preen glands) of water fowl. In the goose 90% of this material is a wax consisting of monoesters of various acids with predominantly **1-octadecanol** as the long-chain fatty alcohol.[81] The latter is formed by reduction of stearoyl-CoA as indicated in Fig. 21-2. Waxes are also important constituents of marine environments, where they are not limited to surfaces. For example, copepods, which constitute a major component of marine zooplankton, may contain up to 70% of their dry weight as wax esters. Some marine animals, such as sperm whales, accumulate the same esters in major amounts as energy stores.[133]

Among the compounds present in the lipids of human skin are a variety of branched fatty acids, both free and combined. They may play a role in maintaining the ecological balance among microorganisms of the skin, and they also impart to each individual a

distinct odor or "chemical fingerprint."[132] Some of the skin lipids are incorporated into the cornified outer skin surface (Box 8-F).[134] See also Section C, 3.

Surface lipids of plants. The thick cuticle (Fig. 1-6) that covers the outer surfaces of green plants consists largely of waxes and other lipids but also contains a complex polymeric matrix of **cutin** (stems and leaves) or **suberin** (roots and wound surfaces).[135,135a] Plant waxes commonly have $C_{10} - C_{30}$ chains in both acid and alcohol components. Methyl branches are frequently present. A major function of the waxes is to inhibit evaporation of water and to protect the outer cell layer. In addition, the methyl branched components may inhibit enzymatic breakdown by microbes. Free fatty acids, free alcohols, aldehydes, ketones, β-diketones, and alkanes are also present in plant surface waxes. Chain lengths are usually $C_{20} - C_{35}$.[136] Hydrocarbon formation can occur in other parts of a plant as well as in the cuticle. Thus, normal **heptane** constitutes up to 98% of the volatile portion of the turpentine of *Pinus jeffreyi*.[81]

Cutin is largely a polyester with a high content of ω-hydroxypalmitic acid and related fatty acids, which are also hydroxylated at a second position:

$$HO-CH_2-(CH_2)_x-\underset{\underset{OH}{|}}{CH}-(CH_2)_y-COOH$$

Cutin monomers. C_{16} acids in which y = 8, 7, 6, or 5 and x + y = 13

This allows branching of the polymer. Monomers of other chain lengths as well as aromatic components related to lignin are also present and polymerized into a high molecular mass branched structure. Suberin is a more complex ligninlike polymer with a high content of phenolic constituents[135] such as vanillin (Fig. 25-8).

Formation of hydrocarbons. Alkanes and alkenes occur in plants, in preen gland secretions, and in insects. The alkanes of plant cuticle are thought to be formed by elongation of a C_{16} acid followed by loss of the carboxyl group. The mechanisms are not obvious. However, these hydrocarbons are often two carbon atoms shorter than the starting fatty acid. The pathway between them might begin by α-oxidation to form an α-peroxy acid which would decarboxylate to form an aldehyde, a reaction similar to that of Eq. 15-36. Alternatively, a long-chain acyl-CoA may be reduced directly to an aldehyde. In fact, when suitable inhibitors are present aldehydes do accumulate in tissues that are forming hydrocarbons.[137] Conversion of an aldehyde intermediate to an alkane may occur by **decarbonylation** (loss of CO). This has been demonstrated in pea (*Pisum sativum*) leaves,[138] in uropygial glands,[139] in flies, and in a colonial green alga, *Botyrococcus braunii*.[138] In the last case 32% of the dry weight of the cells is C_{27}, C_{29}, and C_{31} hydrocarbons. They appear to be formed by action of a decarbonylase that apparently contains a cobalt porphyrin.[137] Plants require cobalt for growth, but an enzymatic function has not previously been established.

In contrast, the sex pheromone of the female housefly is (Z)-9-tricosene, a hydrocarbon apparently formed by an oxidative decarboxylative process from a precursor aldehyde by an enzyme that requires NADPH and O_2 and is apparently a cytochrome P450.[140] Oxidative deformylation by a cytochrome P450 converts aldehydes to alkenes, presumably via a peroxo intermediate.[117] Formation of an alkene by decarboxylation has also been proposed,[141] but a mechanism is not obvious.

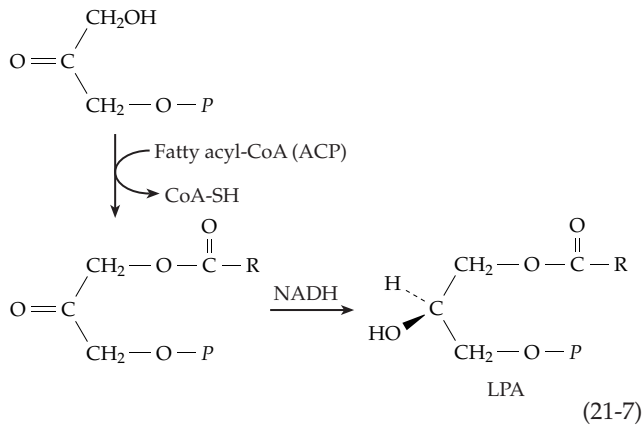
Insect waxes, hydrocarbons, and pheromones. Surface lipids of many insects contain esters of long-chain (as long as 66 carbon atoms) alcohols and long-chain acids.[142] On the other hand, waxes of the tobacco hornworm consist largely of 11- and 12-oxo derivatives of a C_{28} alcohol, which may be esterified to short-chain acids.[142] A major hydrocarbon of cockroaches is 6,9-heptacosadiene.[81]

Insects communicate through the use of a great variety of volatile pheromones. As mentioned in Chapter 8, Section A,1, some moths utilize acetate esters of various isomers of Δ^7 and Δ^{11} unsaturated C_{14} fatty acids as sex pheromones. Some other moths convert the *trans*-11-tetradecenyl acetate into the corresponding C_{14} aldehyde or alcohol, while others use similar compounds of shorter ($C_{11} - C_{12}$) chain length.[143] Some ants use ketones, such as 4-methyl-3-heptanone, as well as various isoprenoid compounds and pyrazines as volatile signaling compounds.[144] Other insects also utilize isoprenoids,[145] alkaloids,[146] and aromatic substances as pheromones.

C. Synthesis of Triacylglycerols, Glycolipids, and Phospholipids

Reduction of dihydroxyacetone phosphate yields *sn*-glycerol 3-phosphate, the starting compound for formation of the glycerol-containing lipids (Fig. 21-4 step *a*).[146a,b] Transfer of two acyl groups from ACP or CoA to the hydroxyl groups of this compound (steps *b* and *c*) yields 1,2-diacylglycerol 3-phosphate (phosphatidic acid). Two different acyltransferases are required.[147] Unsaturated fatty acids are incorporated preferentially into the 2-position. The intermediate 1-acyl-*sn*-glycero-3-phosphate, often called **lysophosphatidic acid** (LPA), is formed in excess in activated platelets and has a variety of signaling activities.[148,149] LPA for signaling is derived by turnover of existing phospholipids. An alternative route of LPA formation in liver is the transfer of one acyl group onto dihy-

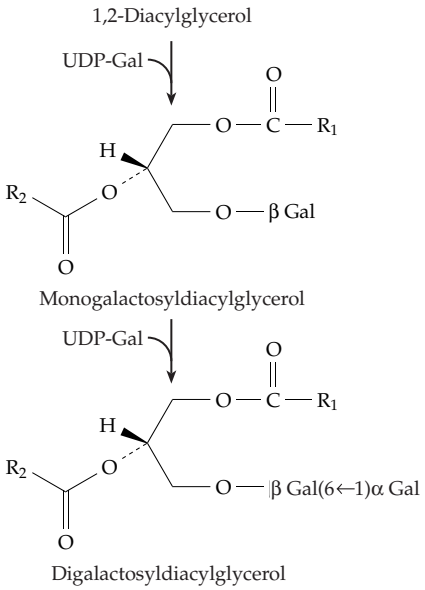
droxyacetone phosphate and reduction prior to addition of the second acyl group (Eq. 21-7).



(21-7)

Phosphatidic acid lies at a metabolic branch point. On the one hand, the phospho group can be removed by a specific phosphatase (step *d*)[150] and another acyl group (most often an unsaturated acyl group) may be transferred onto the resulting diacylglycerol (DAG, diglyceride, step *e*)[150a,b] to form a **triacylglycerol** (triglyceride). Alternatively, the phosphatidic acid may be converted to a **CDP-diacylglycerol** (step *g*), a key intermediate in phospholipid synthesis both in eukaryotes and in bacteria.[151] Not only can phosphatidic acid be hydrolyzed to 1,2-diacylglycerols, but the reverse process can occur by action of a kinase. This presumably permits recycling of the diacylglycerol formed by turnover of membrane phospholipids.[152]

Diacylglycerols can also be converted to a variety of glycolipids such as the **galactolipids** of chloroplasts (Eq. 21-8). See also Chapter 8. These are the major lipids of photosynthetic membranes.[98,153–155] Some bacteria, e.g., the mycoplasma *Acholeplasma laidlawii*, contain both monoglucosyl- and diglucosyl-DAG. Changes in the ratio of these two membrane components may regulate the phase equilibrium between bilayer and nonbilayer forms.[156] 1,2-Diacylglycerol can also react with UDP-sulfoquinovose (Eq. 20-12) to form the characteristic sulfolipid of chloroplasts.[154]



(21-8)

In animals a principal regulatory point for lipid synthesis is in the activation of acetyl-CoA carboxylase by citrate (Fig. 17-20).[156a,b] Beyond that, a complex hormonal control is exerted on both biosynthesis and the catabolism of triglycerides stored in liver and adipose tissues.[157] For example, adrenaline and glucagon, by stimulating production of cAMP, stimulate acetyl-CoA carboxylase,[158] activate lipases that cleave triacylglycerols, and mobilize depot fats.[159] Insulin, on the other hand, promotes lipid storage. It increases the activity of the enzymes of lipogenesis from the ATP-dependent citrate cleavage enzyme (Eq. 13-39) and inhibits cAMP production, thus blocking lipolysis within cells. At the transcriptional level sterols bind to activator proteins (**sterol regulatory element binding proteins,** SREBPs) and activate genes for acetyl-CoA carboxylase[158] and for stearoyl-CoA desaturases.[160] Fatty acid synthases, which play a central role in lipid formation, are controlled by both hormonal and nutritional factors at the transcriptional[161] and translational[162] levels. Environmental factors also have indirect effects. For example, the Δ^9 fatty acid desaturase activity of poikilothermic (cold-blooded) animals is increased at low temperatures. The resulting increased synthesis of unsaturated fats leads to increased fluidity of the membrane bilayer.[163] As mentioned on p. 1193; the same is true for green plants.

1. Phospholipids

Bacterial and also some eukaryotic phospholipids are formed following conversion (Fig. 21-4, step *g*) of phosphatidic acids to CDP-diacylglycerols, which are able to react with a variety of nucleophiles with displacement of CMP.[164–166] Reaction with L-serine (step *h*)[167] leads to **phosphatidylserine,** and reaction with glycerol 3-phosphate (step *i*),[168] which enters cells via a special transporter,[168a] produces **phosphatidyglycerol 3-*P*.** The enzyme catalyzing the formation of phosphatidylserine appears to occur naturally as an integral membrane protein of the ER. Some is also bound to ribosomes and to mitochondria.[169,169a] In contrast, most of the other enzymes of phospholipid formation are closely associated with or embedded in the cytoplasmic membrane. One of these, a pyruvoyl group dependent enzyme (Chapter 14, Section F), catalyzes decarboxylation of phosphatidylserine to **phosphatidylethanolamine** (PE, step *j*, Fig. 21-4).[170] This reaction had been thought unimportant in animals, but

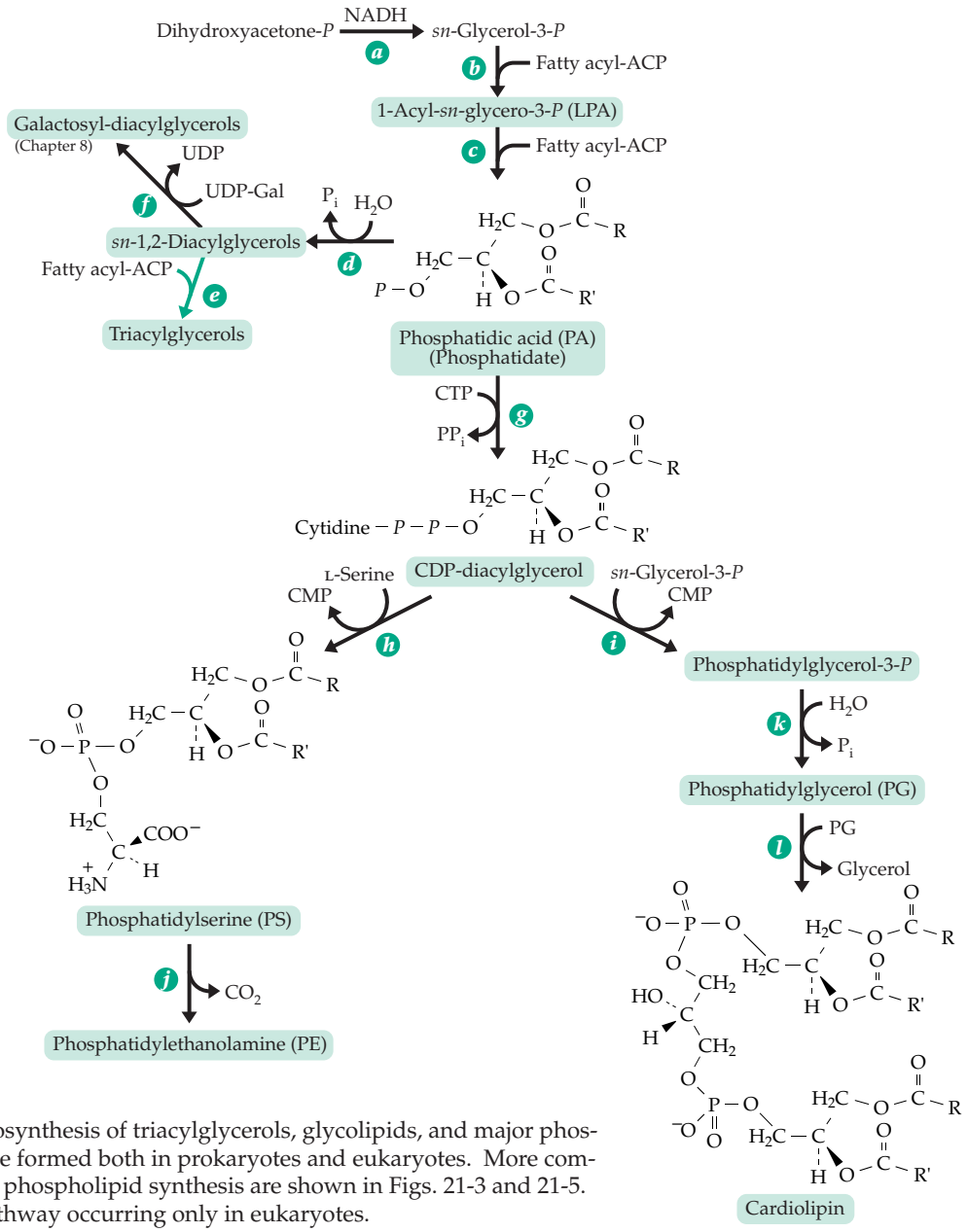


Figure 21-4 Biosynthesis of triacylglycerols, glycolipids, and major phospholipids that are formed both in prokaryotes and eukaryotes. More complete schemes of phospholipid synthesis are shown in Figs. 21-3 and 21-5. Green arrow: pathway occurring only in eukaryotes.

results with cultured cells show that decarboxylation of phosphatidylserine is often the major route of formation of phosphatidylethanolamine in mammalian cells.[171,172] This phospholipid also accounts for 75% of total phospholipid of the *E. coli* cell envelope. It is synthesized on the cytosolic side of the inner membrane, but it is also translocated to the outer membrane, where it is a major constituent of the inner bilayer leaflet.[173] PE is essential for viability of *E. coli* cells.[174] It provides dipolar ionic head groups and apparently serves as a chaperone for folding of some membrane proteins.[175]

After removal of a phosphate from phosphatidylglycerol 3-*P*, the resulting phosphatidylglycerol can be converted to **diphosphatidylglycerol** (known as **cardiolipin**). One manner in which this is accomplished in bacteria is indicated by step *l* of Fig. 21-4. One molecule of glycerol is displaced as two molecules of phosphatidylglycerol are coupled. The alternative pathway of Eq. 21-9 is followed in eukaryotic mitochondria and perhaps in some bacteria. The entire phosphatidic acid group is transferred from CDP-diacylglycerol to phosphatidylglycerol with displacement of CMP.[176–178] Gram-negative bacteria also synthesize a second set of membrane phospholipids, compounds such as **lipid A** (Figs. 8-30, 20-10) that are based on acylated glucosamine.[165]

Phosphatidylcholine, which is rarely present in bacteria, is formed in eukaryotes from phosphatidylethanolamine by three consecutive steps of methylation

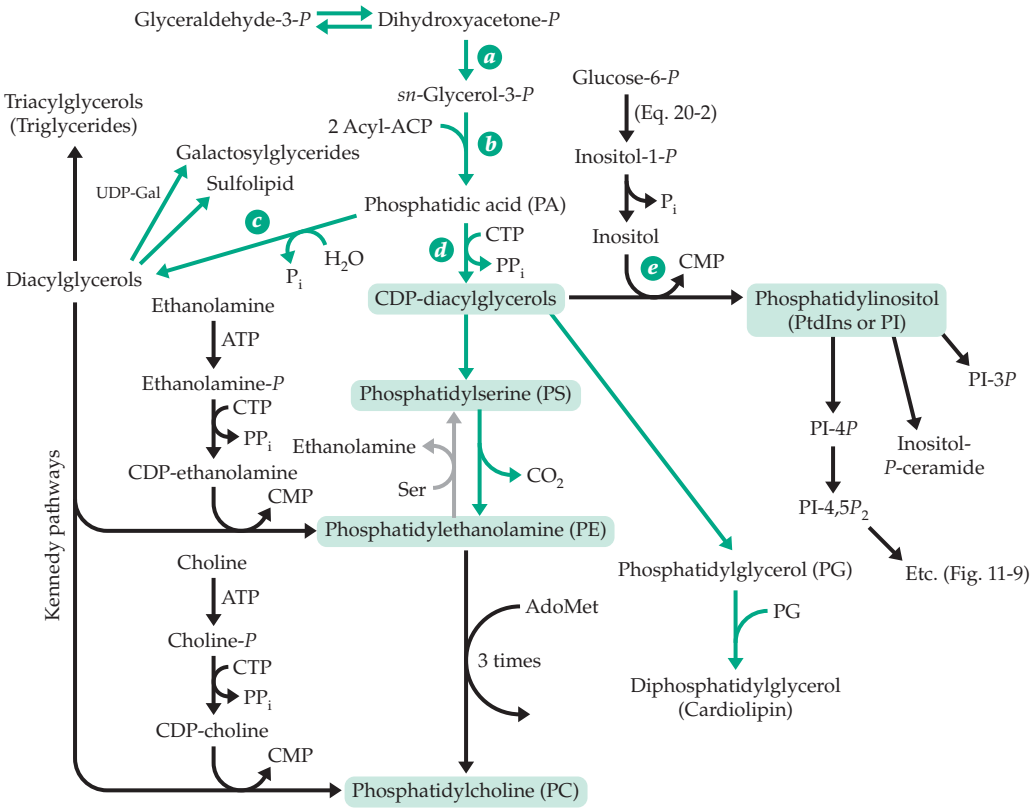
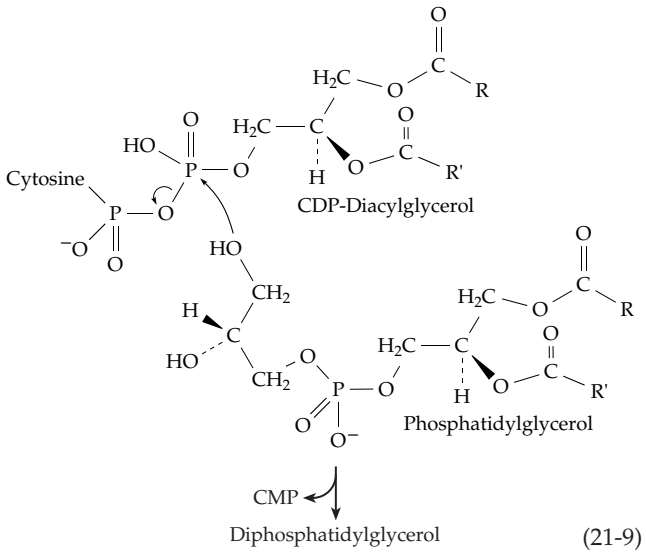


Figure 21-5 A more complete outline of the biosynthesis of triacylglycerols, glycolipids, and phospholipids including characteristic eukaryotic pathways. Green lines indicate pathways utilized by both bacteria and eukaryotes. Structures of some of the compounds are shown in Fig. 21-4. The gray arrows show the formation of phosphatidylserine by exchange with ethanolamine (Eq. 21-10).

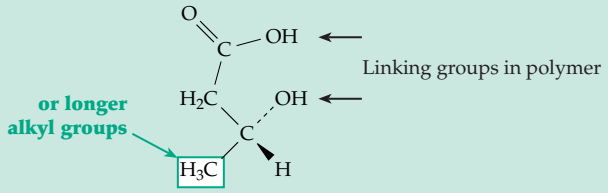


(21-9)

by *S*-adenosylmethionine (Fig. 21-5). This pathway is of major importance in eukaryotic cells.[179,180] However, alternative pathways (the Kennedy pathways),[166] which are represented by black lines on the left side of Fig. 21-5, are also used for formation of both phosphatidylcholine and phosphatidylethanolamine. In both cases, the free base, choline, or ethanolamine[180a,b] is phosphorylated with ATP. Choline phosphate formed in this manner is then converted by reaction with CTP to CDP-choline (Eq. 17-58).[181] Phosphatidylcholine is formed from this intermediate[181a,b] while CDP-ethanolamine is used to form phosphatidylethanolamine (Fig. 21-5). These synthetic reactions occur within cell nuclei as well as on surfaces of cytoplasmic membranes.[181c]

The formation of phosphatidylserine and possibly other phospholipids in animal tissues may also be accomplished by exchange reactions (Eq. 21-10, step *a*).[182,183] At the same time, decarboxylation of phosphatidylserine back to phosphatidylethanolamine (Eq. 21-10, step *b*) also takes place, the net effect being a catalytic cycle for decarboxylation of serine to ethanolamine. The latter can react with CTP to initiate synthesis of new phospholipid molecules or can be converted to phosphatidylcholine (step *c*). However, unless there is an excess of methionine and folate in the diet, choline is an essential human nutrient.[184]

BOX 21-D POLY-β-HYDROXYBUTYRATE AND BIODEGRADABLE PLASTICS

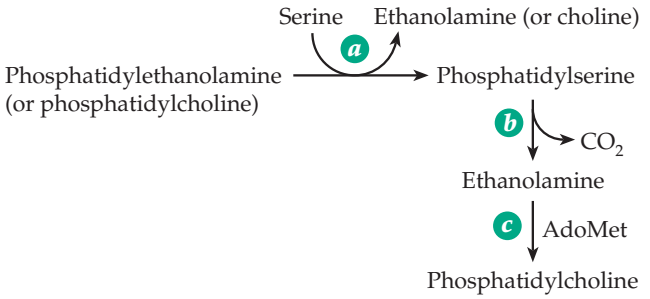


D-β-Hydroxybutyric acid

The important bacterial storage material polyhydroxybutyric acid is related metabolically and structurally to the lipids. This highly reduced polymer is made up of D-β-hydroxybutyric acid units in ester linkage, about 1500 residues being present per chain. The structure is that of a compact right-handed coil with a twofold screw axis and a pitch of 0.60 nm.[a] Within bacteria it often occurs in thin lamellae ~5.0 nm thick. Since a chain of 1500 residues stretches to 440 nm, there must be ~88 folds in a single chain. Present in both cytoplasmic granules and in membranes,[b] polyhydroxybutyrate can account for as much as 50% of the total carbon of some bacteria.[c] In *E. coli* and many other bacteria polyhydroxybutyrate is present in a lower molecular mass form bound to calcium polyphosphates, proteins, or other macromolecules.[d,e] It has also been extracted from bovine serum albumin and may be ubiquitous in both eukaryotes and prokaryotes.[d,e] The polymer may function in formation of Ca^{2+} channels in membranes.[b,d]

Biosynthesis occurs from 3-hydroxybutyryl-CoA. Some bacteria incorporate other β-hydroxyacids into the polymer.[f] Apparently various hydroxyacyl-CoAs can be diverted from the β oxidation pathway to polymer synthesis,[g] and synthases that will accept a variety of β-hydroxyacyl-CoA substrates have been isolated.[h,i] More than 80 different hydroxyacyl groups can be incorporated into the polymer.[i] A bacterially produced copolymer of β-hydroxybutyrate and β-hydroxyvalerate resembles polypropylene but is biodegradeable. It not only can be used for sutures and other medical implants[j] but also could compete with petroleum-derived plastics[h] and be derived from renewable sources. To this end the synthase genes have been cloned, engineered, and transferred into other microorganisms and plants.[k–n] Transgenic cotton plants incorporate polyhydroxybutyrate granules into the cotton fibers altering the properties of the fibers.[m] The polyhydroxybutyrate synthases apper to be related mechanistically to bacterial lipases.[o]

[a] Okamura, K., and Marchessault, R. H. (1967) in *Conformation of Biopolymers*, Vol. 2 (Ramachandran, G. N., ed), pp. 709–720, Academic Press, New York
[b] Reusch, R. N., and Sadoff, H. L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4176–4180
[c] Jacob, G. S., Garbow, J. R., and Schaefer, J. (1986) *J. Biol. Chem.* **261**, 16785–16787
[d] Reusch, R. N., Huang, R., and Bramble, L. L. (1995) *Biophys. J.* **69**, 754–766
[e] Huang, R., and Reusch, R. N. (1996) *J. Biol. Chem.* **271**, 22196–22202
[f] Peoples, O. P., and Sinskey, A. J. (1989) *J. Biol. Chem.* **264**, 15298–15303
[g] de Waard, P., van der Wal, H., Huijberts, G. N. M., and Eggink, G. (1993) *J. Biol. Chem.* **268**, 315–319
[h] Müh, U., Sinskey, A. J., Kirby, D. P., Lane, W. S., and Stubbe, J. (1999) *Biochemistry* **38**, 826–837
[i] Gerngross, T. U., and Martin, D. P. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6279–6283
[j] Pool, R. (1989) *Science* **245**, 1187–1189
[k] Poirier, Y., Dennis, D. E., Klomparens, K., and Somerville, C. (1992) *Science* **256**, 520–523
[l] Mittendorf, V., Robertson, E. J., Leech, R. M., Krüger, N., Steinbüchel, A., and Poirier, Y. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13397–13402
[m] John, M. E., and Keller, G. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12768–12773
[n] García, B., Olivera, E. R., Minambres, B., Fernández-Valverde, M., Canedo, L. M., Prieto, M. A., García, J. L., Martínez, M., and Luengo, J. M. (1999) *J. Biol. Chem.* **274**, 29228–29241.
[o] Crandall, W. V., and Lowe, M. E. (2001) *J. Biol. Chem.* **276**, 12505–12512



(21-10)

Apparently the synthesis via serine and phosphatidylserine cannot provide an adequate amount of choline, which is present in the body not only in phosphatidylcholine but in plasmalogens, sphingomyelins, and the neurotransmitter **acetylcholine**.[185]

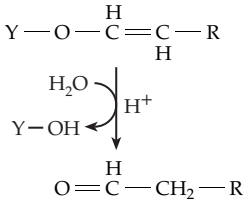
Phosphatidylinositol (PI), a major component of membrane lipids, is formed by displacement of CMP from CMD-dialylglycerol by *myo*-inositol.[186] It is also converted into a variety of less abundant phosphorylated derivatives that engage in signaling activities (see Fig. 11-9). In addition, PI is a component of the glycosylphosphatidylinositol (GPI) membrane anchors for suface proteins (Fig. 8-13). Free GPI anchors, lacking bound proteins, are also present in membranes.

They are especially abundant in many parasitic protozoa and may carry additional glycosyl groups.[186a]

Regulation of phospholipid synthesis, which has been studied in detail in yeast,[187–191] is complex but highly coordinated. The committed step in the synthesis of PE and PC is the hydrolysis of phosphatidate (PA) by a phosphatase to generate diacylglycerols (Fig. 21-5, step *c*). Reaction of PA with CTP (step *d*) also affects synthesis of the other major phospholipids. Much of the coordinate regulation arises at the transcriptional level. For example, genes for the synthesis of PC or PI are repressed by inositol alone and in combination with choline.[187,188] Regulation of CTP synthetase controls the formation of CDP-diacylglycerols.[190] In mammalian cells PC synthesis appears to occur only during the S-phase of the cell cycle (Fig. 11-15).[191] The CTP: phosphocholine cytidyltransferase that catalyzes CDP-choline formation is controlled by storage in a resevoir in the nucleus from which it is transferred to ER membranes.[181b]

2. The Ether-Linked Lipids

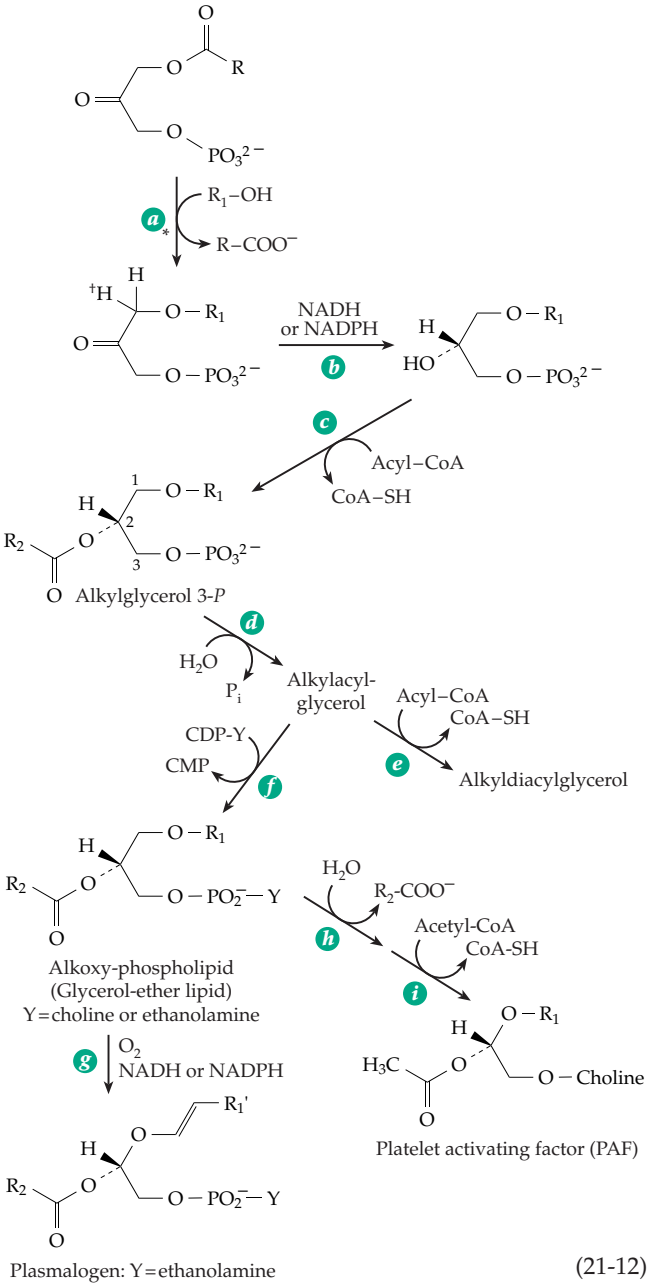
Closely related to both the triacylglycerols and phospholipids, the ether-linked lipids contain in place of one ester group an *alkoxy* (–OR) or *alkenyl* (–O–CH=CH–R) group.[192] Phospholipids containing the alk-1-enyl group, the **plasmalogens**, were first recognized in 1924 by Feulgen and Voit, who were developing histological staining procedures. They observed that treatment of tissue slices with acid resulted in the liberation of aldehydes, which were later shown to be formed by breakdown of the alkenyl lipids (Eq. 21-11). Over 10% of the lipid in the human central nervous system is plasmalogen and about 1% is alkoxy lipid. Among the latter is the **platelet activating factor** (Box 8-A).[193-194a] In many mammalian cells the ethanolamine plasmalogen **plasmenylethanolamine** represents the major storage depot of arachidonic acid.[195]



(21-11)

Ether-linked lipids constitute up to 35% of the total phospholipid in molluscs. Although they are usually regarded as animal constituents, small amounts of ether-linked lipids have been identified in plants. The major phospholipids of archaebacterial membranes are ether-linked derivatives of the polyprenyl phytanyl group and of the dimeric biphytanyl group (Chapter 8).[196–198]

Biosynthesis of ether lipids begins with formation of fatty acyl derivatives of dihydroxyacetone phosphate. The acyl group is then displaced, along with the oxygen atom to which it is attached, by an alkoxy group of a long-chain fatty alcohol (Eq. 21-12, step *a*), which is formed by reduction of the corresponding acyl-CoA.[199] The oxygen of the alcohol (designated by an asterisk) is retained in the product.[200,201] The reaction differs significantly from displacements discussed in Chapter 12. The pro-*R* hydrogen atom (marked by the dagger, †) at C-1 exchanges with the medium during the reaction suggesting that enolization of the dihydroxyacetone phosphate takes place. A possible mechanism would be to add the incoming R–O⁻, generated as in serine proteases, to the double bond of



(21-12)

the enol at C-1. This would generate a transient carbanion on C-2. It could then eliminate the carboxylate containing the original C-1 acyl group, and the enol could then ketonize.

Once an alkoxy derivative of dihydroxyacetone is formed, reduction to the 2-OH form, further acylation, and conversion to various alkyl phospholipids and neutral lipids can occur. The pathways (Eq. 21-12, steps *b*–*f*) are closely akin to those of Fig. 21-4. The conversion of alkoxy lipids to plasmalogens occurs by oxidative desaturation (Eq. 21-12, step *f*).[202] The initial steps in the synthesis of ether-linked lipids take place principally in the peroxisomes. Enzymes catalyzing both the acylation of dihydroxyacetone phosphate and the synthesis of alkyl-dihydroxyacetone-*P* (step *a*, Eq. 21-12) are found in high amounts in animal peroxisomes. In the rare autosomal recessive disorder known as the **Zellweger syndrome** peroxisomes are completely lacking.[203] Both the synthesis of ether-linked lipids[204] and the β oxidation of very-long-chain fatty acids are depressed. These acids, principally C26:0 and C26:1, accumulate in tissues[205,206] of patients with this severe disease, which is usually fatal during the first four months of life.

The platelet activating factor (PAF, Box 8-A) is formed in neutrophils and macrophages from alkylacyl-*sn*-glycero-3-phosphocholine by the action of phospholipase A_2. This enzyme removes the C2 acyl group, which is then replaced by an *acetyl group* transferred from acetyl-CoA to form PAF. Alternatively, a phosphocholine group may be transferred onto 1-alkyl-2-acyl-*sn*-glycerol from CDP-choline as in the formation of phosphatidylcholine (Fig. 21-5). PAF can undergo hydrolytic removal of its acetyl group in tissues but can also transfer it to such acceptors as lysoplasmalogens or sphingosine.[194] Hydrolytic loss of the acetyl group from PAF destroys biological activities including induction of allergic and inflammatory responses.[194a] The various signaling activities of PAF arise from binding to G-protein linked receptors in many cells and tissues.[207]

3. Sphingolipids

Sphinoglipids are phospholipids and glycolipids derived from **sphingosine** and other "long-chain bases."[208] At least 60 bases of this type have been identified.[209] They vary in chain length from C_{14} to C_{26} and include members of the iso and anteiso series. Up to two double bonds may be present. The C_{18} compound, usually called sphingosine, is derived from condensation of palmitoyl-CoA with serine.[209a] Carbon dioxide is lost from the serine during the condensation reaction (Fig. 21-6, step *a*; Chapter 14), and the resulting ketone is reduced with NADPH (step *b*) to form **sphinganine**, a common component of animal sphingolipids. It may be hydroxylated to phytosphingosine in plants and fungi (step *c*).[210,211] Sphinganine is converted to long-chain amides by acyl transfer from acyl-CoA (step *d*) and then undergoes desaturation (step *e*)[212–213a] to form **ceramides**, the precursors to more than 100 gangliosides (Fig. 20-11),[214,215] to the phospholipid **sphingomyelins** (Fig. 21-6, step *g*), and also to free sphingosine (step *h*). This last reaction is degradative and on the pathway of breakdown of gangliosides (Fig. 20-11). Further catabolism of sphingosine is thought to take place by a PLP-mediated chain cleavage to palmitaldehyde.[216]

The essential functions of sphingolipids, including the complex gangliosides, are only now being clarified.[215,217,218] The latter are abundant in brain and are thought to function in cell–cell recognition. On blood cell surfaces they carry blood group antigens (Box 4-C). They play an essential role in spermatogenesis[218] and may function in various signaling processes.[218a] In the outer cornified layers of skin, ceramides with very long chain (C_{28}–C_{36}) fatty acyl groups undergo ω hydroxylation (Fig. 21-6) and become esterified to glutamate side chains of specific skin proteins called **involucrins**. The long hydrocarbon chains are thought to pass entirely through the lipid bilayer to form rigid lamellae of a water-impermeable outer skin barrier.[134] An important hypothesis is that sphingolipids associate with cholesterol to form "lipid rafts," which float in a sea of glycerolipids and serve as bases for various signaling processes. The long hydrophobic acyl chains of the sphingolipids pack well with cholesterol to form a rigid lipid structure of high melting temperature.[218b,c]

4. Complex Lipids in Signaling

While pathways of synthesis of complex lipids have been described, we are far from understanding the dynamics of the synthesis and turnover of the membranous structures built from them. The fact that the lipid bilayer of a cell membrane is so thin means that any sudden changes in composition at a particular location will cause changes in physical properties and a wave of diffusion that will travel along the membrane. The membrane seems to be ideally structured to receive and propagate messages from outer surface or internal receptors, or messages sent along the bilayer.

One of the most studied examples of signaling with membrane lipids is provided by the **phosphoinositide cascade**, which is pictured in Fig. 11-9. Six or more phosphate esters of phosphatidylinositol (PI) are generated by the action of kinases.[219,220] More than 100 extracellular signaling molecules activate specific isozyme forms of **phospholipase C**,[221–224] releasing 20 or more different inositol phosphates from these

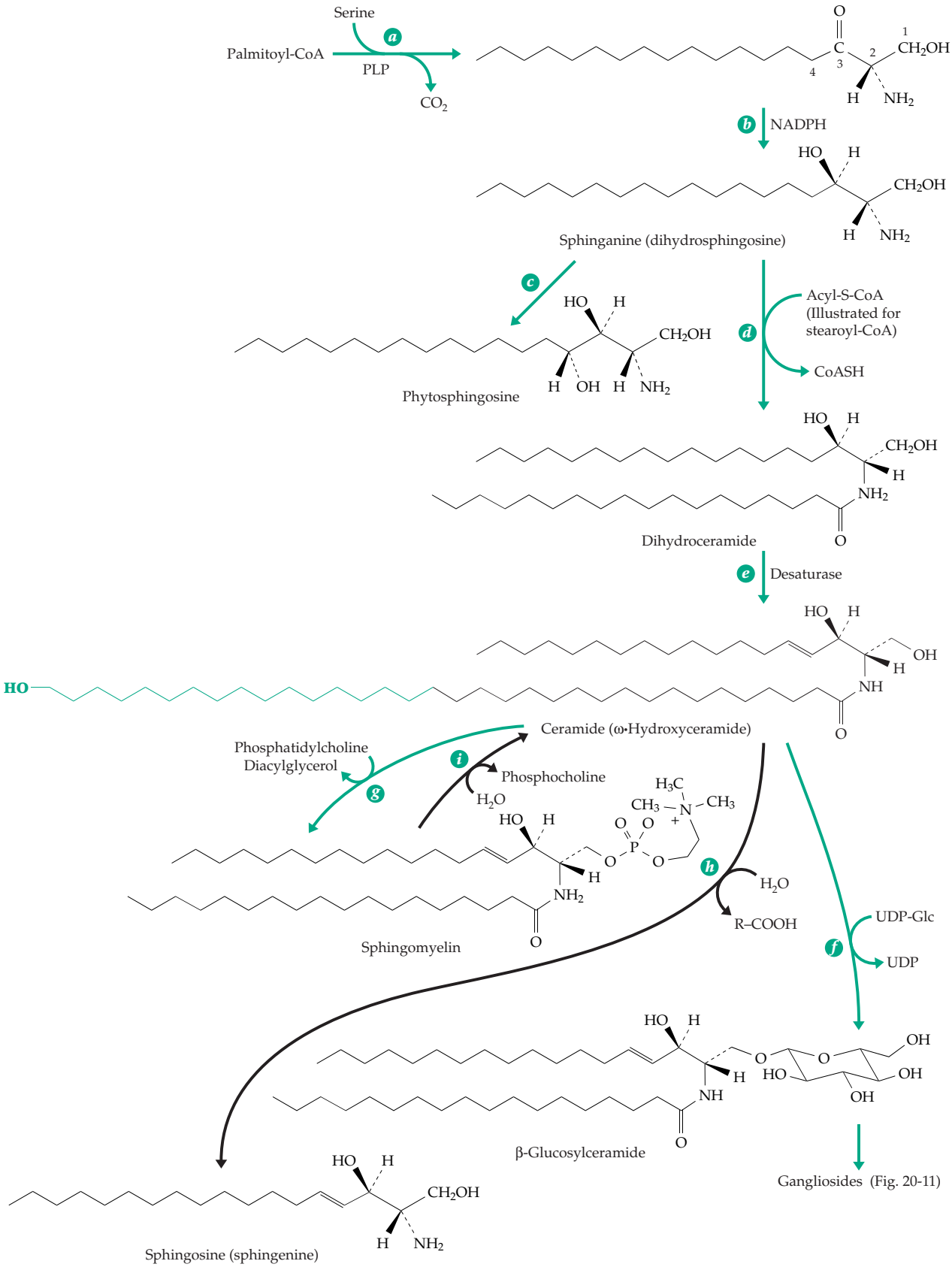


Figure 21-6 Pathways of synthesis and metabolism of sphingolipids. Gray arrows indicate catabolic pathways. See also Fig. 20-11. The green extension on the ceramide structure is that of a long-chain ω-hydroxyceramide that is covalently bound to protein in human skin.

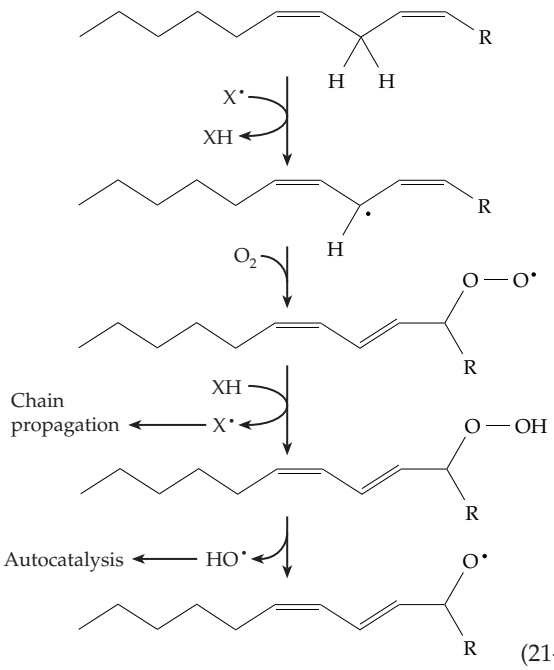
phosphoinositide esters. The released inositol phosphates, which act as water-soluble messengers, are further modified by the action of several phosphatases (Fig. 11-9).[225] At the same time, **diacylglycerols** are left in the membrane. With loss of the negative charges of the PI phosphates there will be immediate electrostatic effects in the membrane, which may alter the ionic environment, open ion channels, etc. The diacylglycerols, which diffuse within the membrane, may lose arachidonic acid from the *sn*-2 position to supply substrate for the arachidonate cascades described in Section D (Eq. 21-16). Diacylglycerols also activate the 11 isozyme forms of **protein kinase C**.[226–228] Some of these enzymes not only are activated by diacylglycerols, but also require **phosphatidylserine** for activity.

Other lipid-based signaling cascades arise from reactions that modify phosphatidylcholine molecules. Unsaturated fatty acids in the *sn*-2 position are readily oxidized by free radicals with cleavage of the hydrocarbon chains to form alcohols, aldehydes, and carboxylic acids. These mimic PAF in their biological activities.[194a,228a] Phosphatidate molecules with saturated or monounsaturated fatty acids in the *sn*-2 position arise from breakdown of phosphatidylcholine catalyzed by **phospholipase D**.[229] Phosphatidate may also be formed by a family of lipid **diacylglycerol kinases**.[230,230a] Phosphatidates containing saturated or monounsaturated fatty acids also have a variety of signaling activities.[230a] An arachidonoyl-diacylglycerol kinase is thought to function in many processes. An example is a PI-mediated cycle in invertebrate vision.[231] Sphingomyelin breakdown (black arrows in Fig. 21-6) releases diffusible ceramides that have been implicated as signaling molecules in cell proliferation, differentiation, growth arrest, and apoptosis.[232–237a] Sphingosine and sphingosine 1-*P* also have signaling functions.[211,237b-e]

Phospholipids have been shown to exchange between different membranes, e.g., of mitochondria and the ER. Exchanges of phosphatidylcholine, phosphatidylinositol, and sphingomyelin are catalyzed by specific **exchange proteins** (Box 21-A).[238,238a] These proteins may also participate in signaling, but their major function may be to transport the phospholipids from their sites of synthesis to the various membranes of the cell.

5. Peroxidation of Lipids and Rancidity

Storage of fats and oils leads to **rancidity**, a largely oxidative deterioration that causes development of unpleasant tastes, odors, and toxic compounds.[239] Similar chemical changes account for the "drying" of oil-based paints and varnishes. These reactions occur most readily with polyunsaturated fatty acids, whether free or in ester linkage within triacylglycerols. The reactions are initiated by free radicals, which may be generated by oxidative enzymes within or outside of cells, or by nonenzymatic reactions catalyzed by traces of transition metals or by environmental pollutants. Characteristic of rancidity is an autocatalytic chain reaction (Eq. 21-13).[239–241]

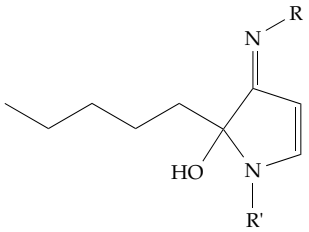


(21-13)

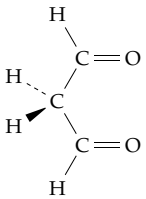
Radical X•, which initiates the reaction, is regenerated in a chain propagation sequence that, at the same time, produces an organic peroxide. The latter can be cleaved to form two additional radicals, which can also react with the unsaturated fatty acids to set up the autocatalytic process. Isomerization, chain cleavages, and radical coupling reactions also occur, especially with polyunsaturated fatty acids. For example, reactive unsaturated aldehydes can be formed (Eq. 21-14).

$$\mathrm{R{-}CH{=}CH{-}CH(\bullet O){-}R' \rightarrow R{-}CH{=}CH{-}CHO + {\bullet}R'} \qquad (21\text{-}14)$$

An intermediate in Eq. 21-13 may be converted to **4-hydroxy-2-nonenal**, a prominent product of the peroxidation of arachidonic or linoleic acids (Eq. 21-15).[242–243a] However, other biosynthetic pathways to this compound are possible.[244,244a] 4-Hydroxy-2-nonenal can react with side chains of lysine, cysteine, and histidine[245] to form fluorescent products such as the following cyclic compound generated by an oxidative reaction.[246]



Polyenoic acids also give rise to malondialdehyde, a reactive mutagenic compound, which can be reduced



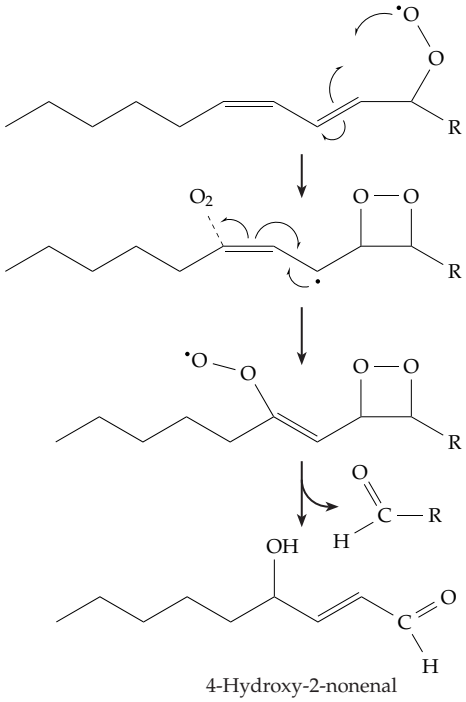
Malondialdehyde

and dehydrated to acrolein, a toxic compound which also reacts with both lysine and serine to

$$CH_2{=}CH{-}CHO$$

Acrolein

produce products excreted in the urine.[247,248] More dangerous are similar reactions of these aldehydes with proteins of the body in conditions such as diabetes or renal insufficiency.[242] The bifunctional malondialdehyde forms Schiff bases with protein amino groups and acts as a crosslinking agent.[249] **Age pigments** (also called **lipofuscin**), which tend to accumulate within neurons and other cells, are



4-Hydroxy-2-nonenal (21-15)

thought to represent precipitated lipid-protein complexes resulting from such reactions.[250] The reactions are similar to those of proteins with the products of sugar breakdown (glycation; p. 69).[250a] Organisms have developed multiple enzymatic mechanisms for detoxification of products of both glycation and oxidative degradation.[243a]

The oxidative degradation represented by the foregoing reactions is referred to as peroxidation. Peroxidation can lead to rapid development of rancidity in fats and oils. However, the presence of a small amount of tocopherol inhibits this decomposition, presumably by trapping the intermediate radicals in the form of the more stable tocopherol radicals (Eq. 15-54), which may dimerize or react with other radicals to terminate the chain.

Catalytic hydrogenation of vegetable oils is widely used to form harder fats and to decrease the content of polyunsaturated fatty acyl groups. The products have a greatly increased resistance to rancidity. However, they also contain fats with trans double bonds as well as isomers with double bonds in unusual positions.[251–253] Such compounds may interfere with normal fatty acid metabolism and also appear to affect serum lipoprotein levels adversely. Trans fatty acids are present in some foods. One hundred grams of butter contain 4–8 g, but hydrogenated fats often contain much more. It has been estimated that in the United States trans fatty acids account for 6–8% of total dietary fat.[253]

6. Some Nutritional Questions

While many of the poorer people on earth starve to death the problems of atherosclerosis and obesity affect many in wealthier societies.[253a–c] The fat content of foods is often blamed, and, as discussed in Boxes 21-B and 22-B, the quality of fatty acids in the diet is very important. However, like fatty acids, carbohydrates are also metabolized via acetyl-CoA and can readily be converted to both fatty acids and cholesterol.[253d] Obesity is largely a problem of excessive total caloric intake.

Why do some people stay slim while others become obese? What are the regulatory mechanisms that affect appetite and body composition? The human body weight tends to be stable or to increase slowly during adult life.[253e] Is there a natural set point for each individual? No, an apparent set point is just a result of action of a multitude of factors including genetic variations[253f] and psychological factors that affect exercise levels, eating habits, etc.[253g] It is worthwhile to recognize that the basal metabolic rate, which is also affected by many factors, accounts for a very high fraction of a person's energy expenditure (p. 283).

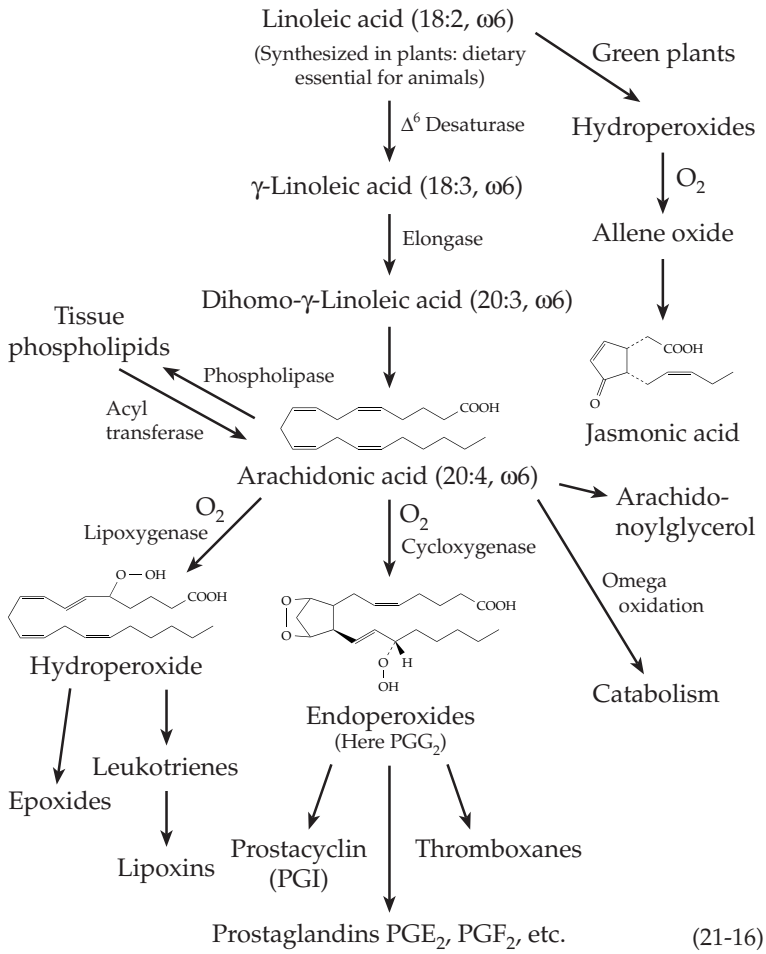
The following are among specific biochemical

factors that act on the energy balance: the activity of acetyl-CoA carboxylase and the associated level of malonyl-CoA[253h,i]; the activity of mitochondrial uncoupling proteins (Box 18-C)[253j,k]; actions of the hormone leptin[253l] (which have been hard to interpret)[253m,n]; and other hormones including cholecystokinins and neuropeptide Y (Chapter 30).

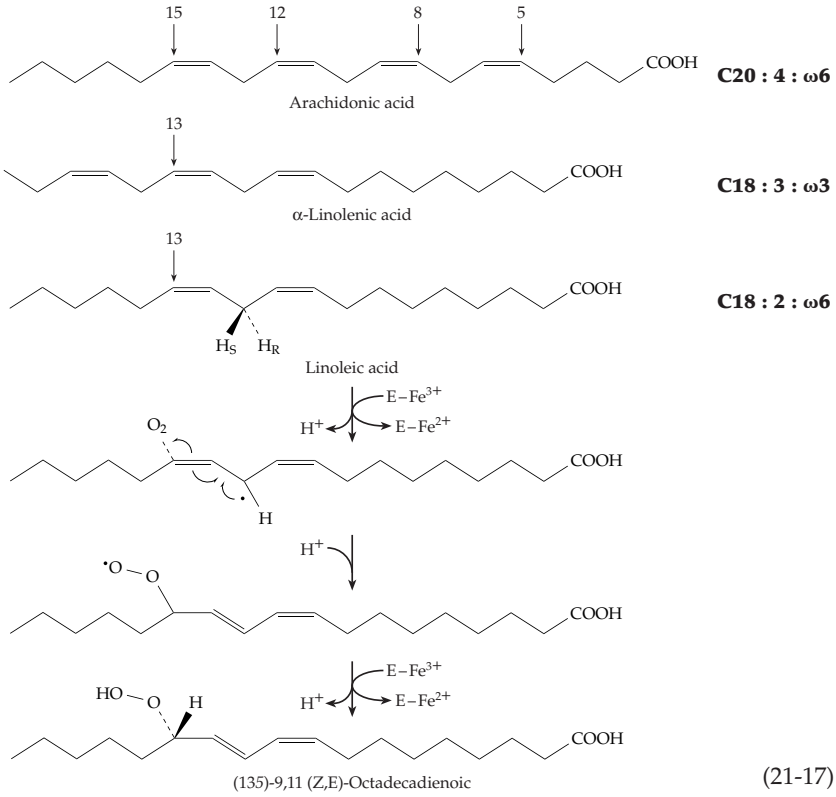
D. Prostaglandins and Related Prostanoid Compounds

Lipid peroxidation has often been regarded simply as an undesirable side reaction, but it is also a normal part of metabolism. Initiated by enzymatically generated radicals, peroxidation occurs as specific metabolic pathways, such as the **arachidonate cascade**, which leads to a variety of local hormones and other substances (Eq. 21-16).[254–256a]

As early as 1930, it was recognized that seminal fluid contains materials that promote contraction of uterine muscles. The active compounds, the **prostaglandins**, were isolated and crystallized in 1960 and were identified shortly thereafter.[257,258] As many as 14 closely related compounds are found in human seminal fluid, one of the richest known sources. Prostaglandins are present in seminal fluid at a total concentration of ~1 mM, but their action on smooth muscles has been observed at a concentration as low as 10^{-9} M. The structures and biosynthetic pathways of several of the prostaglandins are indicated in Fig. 21-7. Prostaglandins are usually abbreviated PG with an additional letter and numerical subscript added to indicate the type. The E type are β-hydroxyketones, the F type 1,3-diols, and the A type α, β-unsaturated ketones. Series 2 prostaglandins arise from arachidonic acid, while series 1 and 3 arise from fatty acids containing one fewer or one more double bond, respectively (Fig. 21-7). Additional forms are known.[257,259]



(21-16)



(21-17)

1. Metabolism of the Prostaglandins

Prostaglandins are not stored by cells but are synthesized in response to external stimuli. Arachidonic acid and other polyenoic acids are present in relatively small amounts (e.g., ~1% of total plasma fatty acids), but they are concentrated in the 2-position of phospholipids. This is in part a result of phospholipid "remodeling." Acyl groups are hydrolyzed from the *sn*-2 position by action of phospholipase A_2. An acyltransferase with a preference for arachidonoyl groups then transfers esterified arachidonic acid from

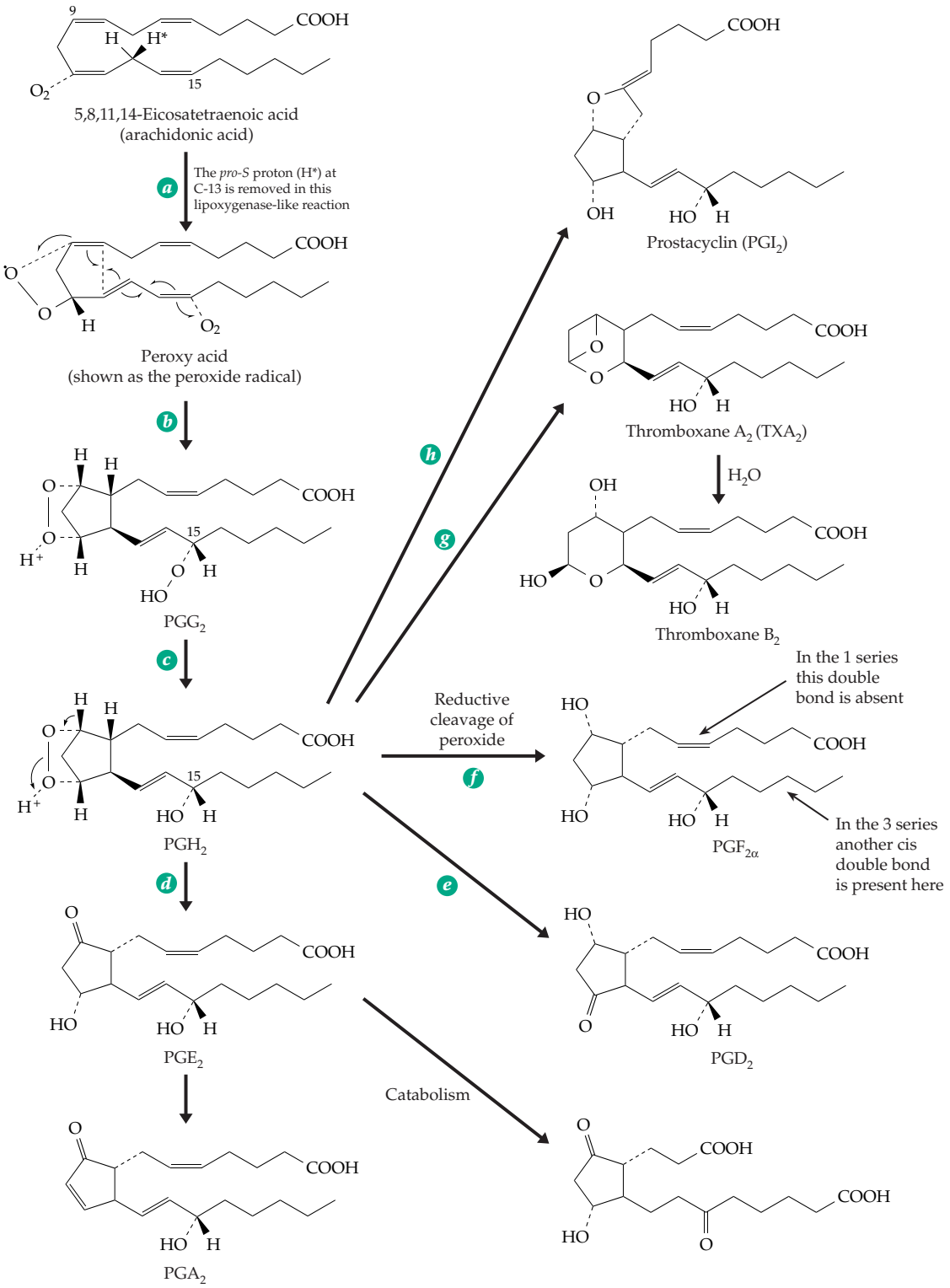


Figure 21-7 Pathway of synthesis and catabolism of the prostaglandins.

phosphatidylcholine and other phospholipids to the lyso forms of the phospholipids, which lack a 2-acyl group. The enzyme has a strong preference for the lyso-ethanolamine plasmalogens. As a consequence, in the plasmenylethanolamine of platelets arachidonoyl groups account for 66% of the acyl residues at the 2-position.[260] An arachidonate-specific acyl-CoA synthetase rapidly reconverts any free arachidonate that is not used for prostanoid synthesis back into phospholipids.[261]

The synthesis of prostaglandins, which was elucidated by Samuelsson,[258–260] begins with the release of arachidonate and other polyenoic acid precursors from phospholipids through the action of phospholipase A_2. The released arachidonate is then acted upon by **prostaglandin H synthases,** which catalyze two consecutive reactions at adjacent but distinct sites in a single protein.[262–262d] The first, **cyclooxygenase** or prostaglandin endoperoxide synthase reaction, forms PGG_2 from arachidonate and the second, a **peroxidase** reaction, generates PGH_2. There are two major mammalian isozymes of prostaglandin synthase, which are often called cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2). From studies of stereospecifically synthesized ^{3}H-containing fatty acid precursors, it was established that the first step in cyclooxygenase action involves removal of the *pro*-S proton at C-13 of the fatty acid (step *a*, Fig. 21-7). The O_2-requiring cyclooxygenase resembles lipoxygenase (Eq. 21-17).[181] The product is a peroxy acid, possibly in the form of the peroxide radical shown in Fig. 21-7. This radical (or peroxide anion) undergoes cyclization with synchronous attack by a separate O_2 molecule at C-15 (Fig. 21-7, step *b*) to give the endoperoxide PGG. Reduction of the latter to an OH group by the NADPH-dependent peroxidase (step *c*) yields PGH. The entire sequence is catalyzed by the single 70-kDa PGH synthase, which contains a single heme prosthetic group.[263] During the cyclooxygenase reaction the enzyme appears to accept electrons from cytochrome b_5. During the peroxidase step the heme group undergoes formation of the characteristic peroxidase intermediate compounds I and II[263] (Fig. 16-14). It has been suggested that a tyrosyl radical is generated in the peroxidase active site (on Y385 of COX-1) and is used to form an arachidonate radical that reacts with O_2 in the cyclooxygenase reaction.[264–266b] Alternatively, a carbocation mechanism is also possible.[267]

PGH can break down in three ways to give the E and F series of prostaglandins.[268] In one the proton at C-9 is eliminated (step *d*) as indicated by the small arrows by the PGH_2 structure of Fig. 21-7. An alternative isomerization (step *e*) gives PGD_2. The F prostaglandins are formed by reductive cleavage of the endoperoxide (step *f*). The A series and other prostaglandins arise by secondary reactions, one of which is shown in Fig. 21-7.

A biochemical characteristic of the prostaglandins is rapid catabolism. The product shown in Fig. 21-7 (lower right) arises by oxidation of the 15-OH to a carbonyl group, permitting reduction of the adjacent trans double bond. Two steps of β oxidation as well as ω oxidation are also required[269] to produce the dicarboxylic acid product shown. However, a series of products appears, and the distribution varies among species. Catabolism of prostaglandins is especially active in the lungs, and any prostaglandins entering the bloodstream are removed by a single pass through the lungs. This observation has led to the conclusion that prostaglandins are not hormones in the classical sense but act on a more local basis.

2. Thromboxanes and Prostacyclins

In blood platelets and in some other tissues PGG is also transformed to another series of compounds, the **thromboxanes**,[270] which were identified in 1975. Labile hemiacetals, the thromboxanes A (TXA, Fig. 21-7), are derived by rearrangement of PGH (step *g*). Thromboxane synthase,[271–273] which catalyzes the reaction, has characteristics of a cytochrome P450. Cytochromes P450 are known to react with peroxides as well as with O_2, and the endoperoxide of PGH may be opened by the synthase prior to rearrangement to TXA.[273] Thromboxane A_2 is so unstable that its half-life at 37°C in water is ~36 s. It is spontaneously converted to TXB_2 (Fig. 21-7), which contains an –OH group at C-15. The thromboxanes B are much more stable than TXA but are not very active physiologically.

By 1976, Vane and associates had identified another prostanoid compound, **prostacyclin** (or PGI_2).[274–275a] This compound also arises from PGH_2 by action of a cytochrome P450-like prostacyclin synthase (Fig. 21-7).[273,275,276] It is thought to be an important vasoprotective molecule. As with the thromboxanes, prostacyclin undergoes rapid inactivation[277] by hydrolysis to the physiologically inactive 6-oxo-$PGF_1\alpha$.

3. Lipoxygenases

Lipoxygenases, of which the enzyme from soy beans has been studied the most, also catalyze oxidation of polyunsaturated fatty acids in lipids as indicated in Eq. 21-17. Formation of the hydroperoxide product is accompanied by a shift of the double bond and conversion from cis to trans configuration. Soybean lipoxygenase is a member of a family of related lipoxygenases that are found in all eukaryotes. All

appear to have similar iron- or manganese-containing active sites and to act by similar mechanisms.[278–280e] The major substrate in animals is arachidonic acid (probably as the arachidonate ion). As marked on the structures above Eq. 21-17, there are 5-, 8-, 12-, and 15-lipoxygenases, which catalyze reaction with dioxygen at the indicated places.[281,282] Linoleic and linolenic acids are the primary substrates in plants. Soybean lipoxygenase acts on the 13-position of linoleic acid as shown in Eq. 21-17. However, this enzyme is often referred to as a *15-lipoxygenase* because it acts on arachidonate at C15. The 100-kDa enzyme from

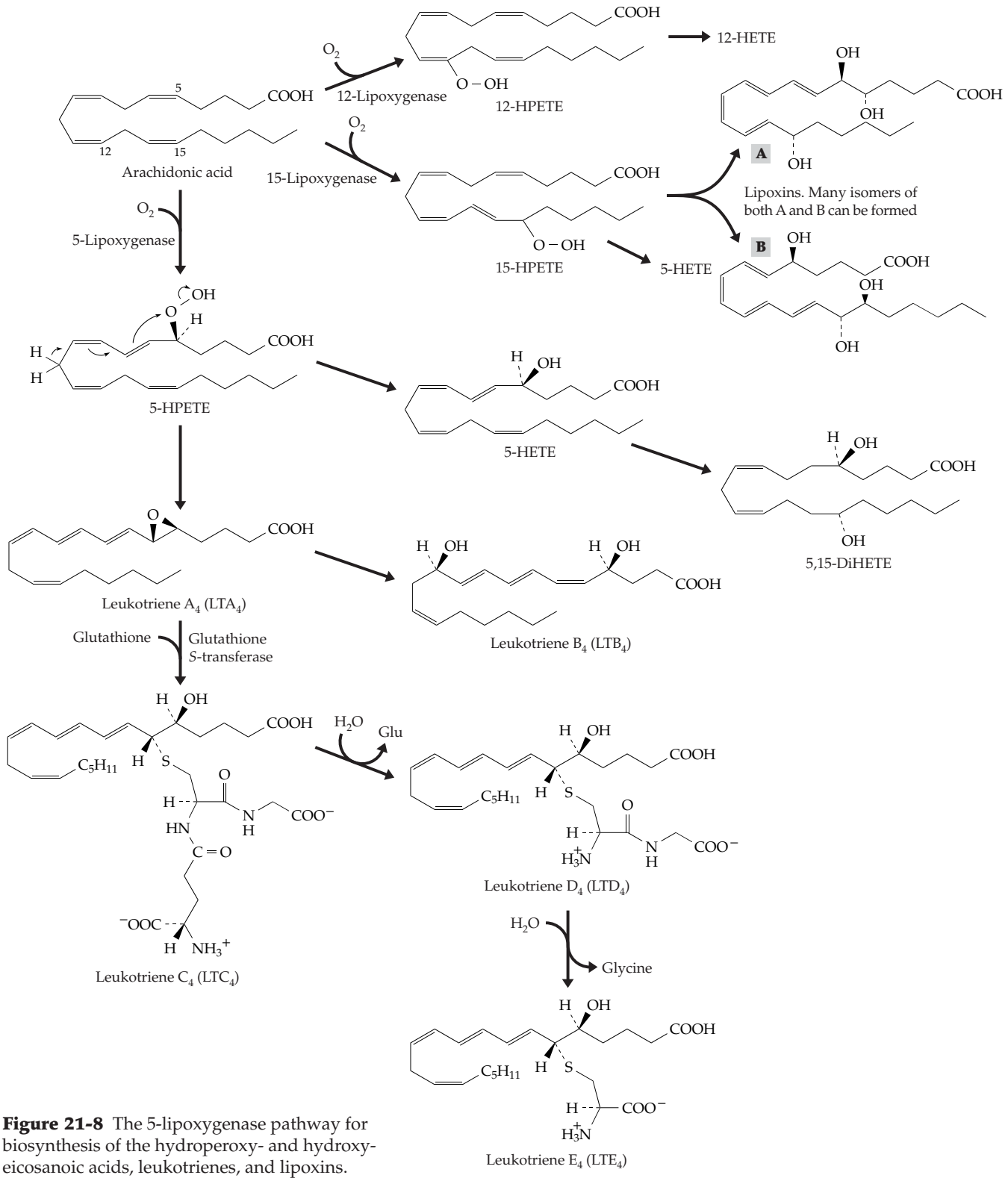


Figure 21-8 The 5-lipoxygenase pathway for biosynthesis of the hydroperoxy- and hydroxy-eicosanoic acids, leukotrienes, and lipoxins.

soybeans contains one atom of Fe(II), which is bound by a cluster of His and Tyr side chains.[283] It must be oxidized to the Fe(III) state before becoming active.[279,284] The initial reaction with O_2 may occur via an intermediate radical.

4. Leukotrienes, Lipoxins, and Related Compounds

Yet another series of products results from the action on arachidonate of tissue lipoxygenases, which compete with the prostaglandin-forming cyclooxygenases. The 5-lipoxygenase (Fig. 21-8) produces the unstable peroxide **5-hydroperoxy 6,8,11,14-eicosatetraenoic acid** (usually abbreviated **5-HPETE**). This enzyme requires ATP and Ca^{2+} and appears to be regulated by a series of other metabolites.[285] The 12- and 15-lipoxygenases, whose distribution varies among different mammalian organs and tissues, form the corresponding 12- and 15-HPETES as well as 5, 15-, 8,15-, 14,15-diHPETEs.[255,286] Some of these peroxides have physiological effects of their own, but they are largely transformed by peroxidases to more stable compounds such as the corresponding alcohols (hydroxy-icosatetraenoic acids or HETEs; Fig. 21-8).

The **leukotrienes** are formed from 5-HPETE (Fig. 21-8).[287,288] Dehydration of HPETE produces the unstable epoxide **leukotriene A_4** (LTA_4), which can be hydrolyzed enzymatically by leukocytes to the diol **leukotriene B_4** (LTB_4).[289,290] Alternatively leukotriene synthase, present in many cells, catalyzes the addition of glutathione (Box 11-B) to the LTA. This is a ring-opening reaction of the epoxide that can be visualized as a nucleophilic displacement by the thiolate anion of glutathione at C-6 (Fig. 21-8). The product is **leukotriene C_4** (LTC_4),[291] which can undergo consecutive removal of glutamate and glycine to form **leukotrienes D** and **E** (LTD_4, LTE_4), respectively. Removal of the glutamate occurs by the action of γ-glutamyl transpeptidase (Box 11-B), whereas removal of the glycine is hydrolytic. LTC_4 and the more potent LTD_4 have been identified as the **slow-reacting substance of anaphylaxis** (SRS-A), a long-sought mediator of bronchial asthma.[292,293] Leukotrienes can be formed from polyunsaturated acids other than arachidonic acid. Thus, eicosapentaenoic acid yields LTC_5 and LTD_5.[294] A lipoxygenase-derived product from the C18:2 linoleic acid is 13-hydroxylinoleic acid, which is made principally by endothelial cells that line blood vessels. It may contribute to resistance to blood clotting.[294]

Products of the 15-lipoxygenase pathway include a group of trihydroxytetraenes formed by leukocytes.[295] Several routes of biosynthesis, which may involve epoxide intermediates, are known.[296,297] The structures of two of these compounds, **lipoxin A** and **lipoxin B,** are shown in Fig. 21-8. Several stereoisomers and cis–trans isomers can be formed. These compounds can all arise from 15-HPETE, either by enzyme action or nonenzymatically. In fact, the entire series of prostanoid compounds arise by reactions related to but more specific than those that occur during nonenzymatic autoxidation of arachidonate.[255,298] Cytochrome P450-catalyzed reaction with O_2 can convert arachidonic acid into four different **epoxytrienoic acids** (**EETs**), which may also exist as stereoisomers. They are vasodilators which affect a variety of signaling pathways.[298a,b]

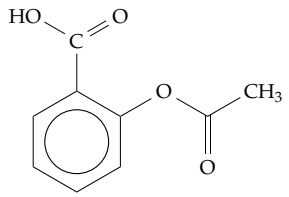
5. Physiological Effects of the Prostanoids

The release of arachidonate and initiation of the arachidonate cascade is induced by hormones, various inflammatory and immunological stimuli, and even mechanical agitation. Tissues do not all behave the same in response to the arachidonate cascade.[299] Blood platelets form largely thromboxane A_2, whereas tissues of the aorta form prostacyclin. Prostaglandin D_2 is a major prostanoid in the central nervous system.[300] Biological functions of prostanoids are also varied.[256a,301] The **primary prostaglandins** PGE and PGF were first recognized as mediators of inflammation. However, PGE_2 and PGF_2 sometimes have opposite effects. The unstable precursors PGG_2 and PGH_2, which have half-lives of only a few minutes, are much more powerful than the more stable PGEs and PGFs. Prostaglandin D_2, which is released in lungs during attacks of asthma, is thought to be a major bronchoconstrictor, but it may also serve as a neurotransmitter.[300] Thromboxanes released from platelets cause smooth muscle contraction and aggregation of the platelets, the first step in blood clot formation. Thromboxanes have half life-times of only seconds but are extremely potent not only in inducing platelet aggregation but also in causing contraction of blood vessels. Prostacyclin has the opposite effect, being a potent vasodilator that causes relaxation of smooth muscle. Upon release from blood vessel walls it acts to prevent clot formation.

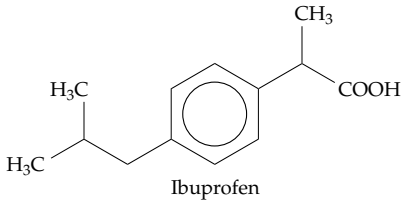
The lipoxygenase pathway (Fig. 21-8) leading to the leukotrienes, lipoxins, and other products is especially active in leukocytes and in mast cells.[302] The leukotrienes promote inflammation, but lipoxins A_4 and B_4 are antiinflammatory.[303] The release of leukotrienes LTC_4, LTD_4, and LTE_4 in lung tissue is correlated with the long-lasting contractions of smooth muscle of the bronchi that are characteristic of asthma.[304] Leukotriene LTC_4 is ~1000 times more powerful than histamine in inducing such contraction. Leukotrienes have also been found in the central nervous system.[305]

Some effects of prostaglandins are mediated through cell surface G-protein coupled receptors (see Chapter 11).[306] Some other prostanoids bind to and activate nuclear peroxisome proliferator-activated receptors.[306] PGJ_2 may inhibit fatty acid synthesis and fat deposition in adipose tissue through these receptors. Some of the prostanoid derivatives enter membranes and may become incorporated into phospholipids and exert their effects there.

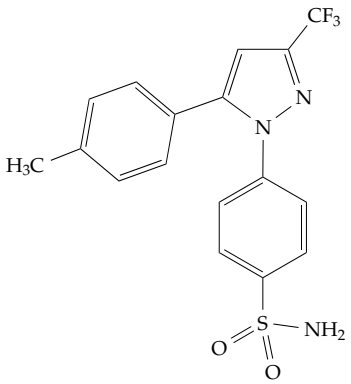
A number of medical uses of prostaglandins have been discovered and more will probably be developed. While prostaglandins may be required for conception, small amounts of PGE_2 or $PGF_{2\alpha}$ induce abortion. $PGF_{2\alpha}$ is also used to induce labor. Prostaglandins are widely employed to control breeding of farm animals, to synchronize their estrus cycles, and to improve the efficiency of artificial insemination.[307]



Aspirin (acetylsalicylic acid)



Ibuprofen



Celecoxib, a new COX-2 inhibitor

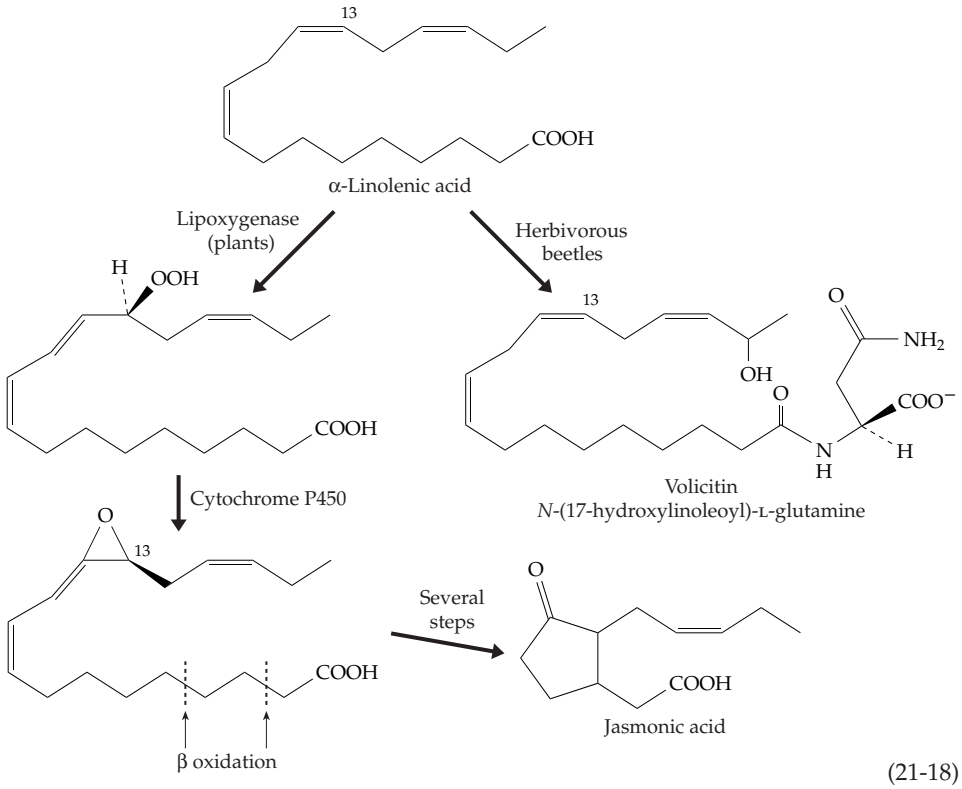
6. Inflammation

Special interest in the prostaglandins has focused on pain of inflammation and allergic responses. The medical significance is easy to see. Five million Americans have **rheumatoid arthritis**, an inflammatory disease. Bronchial asthma and other allergic diseases are equally important. Our most common medicine is **aspirin**, an anti-inflammatory drug. Both the inflammatory response and the immune response are normal parts of the defense mechanisms of the body, but both are potentially harmful, and it is their regulation that is probably faulty in rheumatoid arthritis and asthma. Overproduction of prostaglandins may be a cause of menstrual cramps.[308]

Prostaglandins have been implicated both in the induction of inflammation and in its relief. In inflammation small blood vessels become dilated, and fluid and proteins leak into the interstitial spaces to produce the characteristic swelling (edema). Many polymorphonuclear leukocytes attracted by chemotactic factors that include LTB_4[309] (Chapter 19) migrate into the inflamed area, engulfing dead tissue and bacteria. In this process lysosomes of the leukocytes release phospholipase A, which hydrolyzes phospholipids and initiates the arachidonate cascade. The leukotrienes that are formed promote the inflammatory response. However, cAMP can suppress inflammation, and PGE_2 has a similar effect. Indeed, E prostaglandins, when inhaled in small amounts, relieve asthma.

The synthesis of prostaglandins is inhibited by aspirin[310] and many other analgesic drugs. Aspirin is an acetylating reagent, and the inhibition has been traced to acetylation of the side chain –OH group of a single serine residue, Ser 530 of COX-1 or Ser 516 of COX-2 in the arachidonate binding channel.[311–313] Other nonsteroidal antiinflammatory drugs (NSAIDs), e.g., ibuprofen, are competitive inhibitors of the cyclooxygenases.[314] COX-3, a variant form of COX-1, may be the target for acetaminohen (Box 18-E).[314a] However, the same drugs also inhibit activation of neutrophils and may thus exercise their anti-inflammatory action in more than one way.[315] Since PGE_1 is a potent **pyrogen** (fever-inducing agent), a relationship to the ability of aspirin to reduce fever is also suggested. Unfortunately, all of these drugs inhibit COX-1 of platelets. Small regular doses of aspirin may be useful in preventing blood clots in persons with arterial disease, but they can be disastrous. Thousands of people die annually of hemorrhage caused by aspirin.[316,317]

Recently it was recognized that COX-1 provides eicosinoids for homeostatic purposes, while it is COX-2 that is inducible and generates prostoglandins for production of leukotrienes and induction of the inflammatory response. Now there is a major effort, with the first drugs already in use, to develop specific inhibitors for COX-2, which do not inhibit COX-1. It is hoped that these will be safer than aspirin.[256a,266a,311,317,317a-c] However, these drugs can also cause dangerous side effects.[317d] COX-2 of macrophages is also inhibited by γ-tocopherol, a major form of vitamin E.[317e]



(21-18)

7. Plant Lipoxygenases and Jasmonic Acid

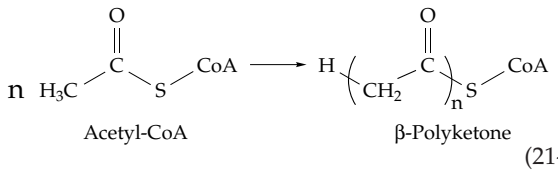
An octadecenoid signaling pathway (Eq. 21-18), which resembles the arachidonate cascade in some respects, plays an important role in green plants.[318–320] Alpha-linolenic acid is acted upon by a lipoxgenase in plastids to form a 13-hydroperoxy derivative. This is dehydrated and cyclized by allene oxide synthase. Although this doesn't appear to be an oxidation–reduction process, the enzyme seems to be a cytochrome P450 and to initiate the cyclization to the unstable epoxide **allene oxide** by homolytic cleavage of the peroxy group.[321,321a] Allene oxide synthase and the cyclase that acts in the next step may be cytosolic, while the β oxidation that shortens the chain occurs in plants exclusively in peroxisomes or glyoxysomes.[319]

Jasmonic acid is a plant growth regulator that affects many aspects of plant development as well as responses to environmental signals. A very important function is mobilization of plant defenses in response to damage by herbivores, by bacterial or fungal pathogens, or by ultraviolet light.[322,323] The synthesis of protease inhibitors as well as phytoalexins is induced. A curious variant of the jasmonate pathway is the acquisition of α-linoleic acid from plants by chewing caterpillars. The linoleic acid is hydroxylated by the insect, and conjugated with glutamine to form **volicitin** (*N*-(17-hydroxylinoleoyl)-L-glutamine; Eq. 21-18). Some of this compound reenters the plant from material regurgitated into wounds by the caterpillars. Volicitin induces the plant to release volatile terpenes and other compounds. The value to the caterpillars is not clear, but not only does volicitin induce defensive reactions in plants but also the released volatile compounds attract wasps that parasitize the caterpillars.[324,325] Plants also form a group of **isoprostanes** E_1 from α-linolenic acid.[326]

Allene oxides are unusual biological products. However, they are formed from arachidonic acid by some corals and are evidently precursors to prostaglandin esters, which may be present in high concentrations.[327] Allene oxides are also present in starfish oocytes.[321,328]

E. The Polyketides

In 1907, Collie proposed that polymers of ketene ($CH_2{=}C{=}O$) might be precursors of such compounds as **orsellinic** acid, a common constituent of lichens. The hypothesis was modernized in 1953 by Birch and Donovan, who proposed that several molecules of acetyl-CoA are condensed (Eq. 21-19) but *without the two reduction steps required in biosynthesis of fatty acids* (Fig. 17-12).[329] As we now know they were correct in assuming that the condensation occurs via malonyl-CoA and an acyl carrier group of an enzyme. The resulting **β-polyketone** can react in various ways to give the large group of compounds known as polyketides.



(21-19)

β-Polyketones can be stabilized by ring formation through ester or aldol condensations. Remaining carbonyl groups can be reduced (prior to or after cyclization) to hydroxyl groups, and the latter can be eliminated as water to form benzene or other aromatic rings. Figure 21-9 illustrates two ways in which cyclization can occur. One involves a Claisen ester condensation during which the enzyme and its SH group are eliminated. Enolization of the product gives a trihydroxy-acetophenone. The second cyclization reaction is the aldol condensation. Following the condensation water is eliminated, and the product is hydrolyzed and enolized to form orsellinic acid. Another product of fungal metabolism is **6-methylsalicylic acid,** which lacks one OH group of orsellinic acid. This synthesis can be explained by assuming that the carbonyl group at C-5 of the original β-polyketone was reduced to an OH group at some point during the biosynthesis. Elimination of two molecules of water together with enolization of the remaining ring carbonyl gives the product (Fig. 21-9).[330]

By allowing a few variations in the basic polyketone structure, the biosynthesis of a large number of unusual compounds can be explained. Extra oxygen atoms can be added by hydroxylation, and methyl groups may be transferred from *S*-adenosylmethionine to form methoxyl groups.[331] Occasionally a methyl group may be transferred directly to the carbon chain. Glycosyl groups may also be attached.[332,333] Many starter pieces other than acetyl-CoA may initiate polyketide synthesis. These include the branched-chain acids of Table 21-3, nicotinic and benzoic acids, 4-coumaroyl-CoA, and a 14:1 Δ^9-ACP. The last of these starter pieces is formed by desaturation of the corresponding 14:0-ACP and is converted via polyketide synthesis to one of a family of **anacardic acids,** which provide pest resistance to a variety of dicotyledenous plants (Fig. 21-10, bottom).[334] The CoA derivative of malonic acid amide is the starter piece for synthesis of the antibiotic **tetracycline** as indicated in Fig. 21-10).[335] Polyketide origins of some other antibiotics are also indicated in this figure.

The cloning and sequencing of genes for enzymes involved in synthesis of polyketides of fungi and actinomycetes has shown that these enzymes are closely related to the fatty acid synthases and, like the latter, have a multidomain structure (Fig. 21-11). The possibilities of engineering these genes, together with the urgent need for new antibiotics, has led to an

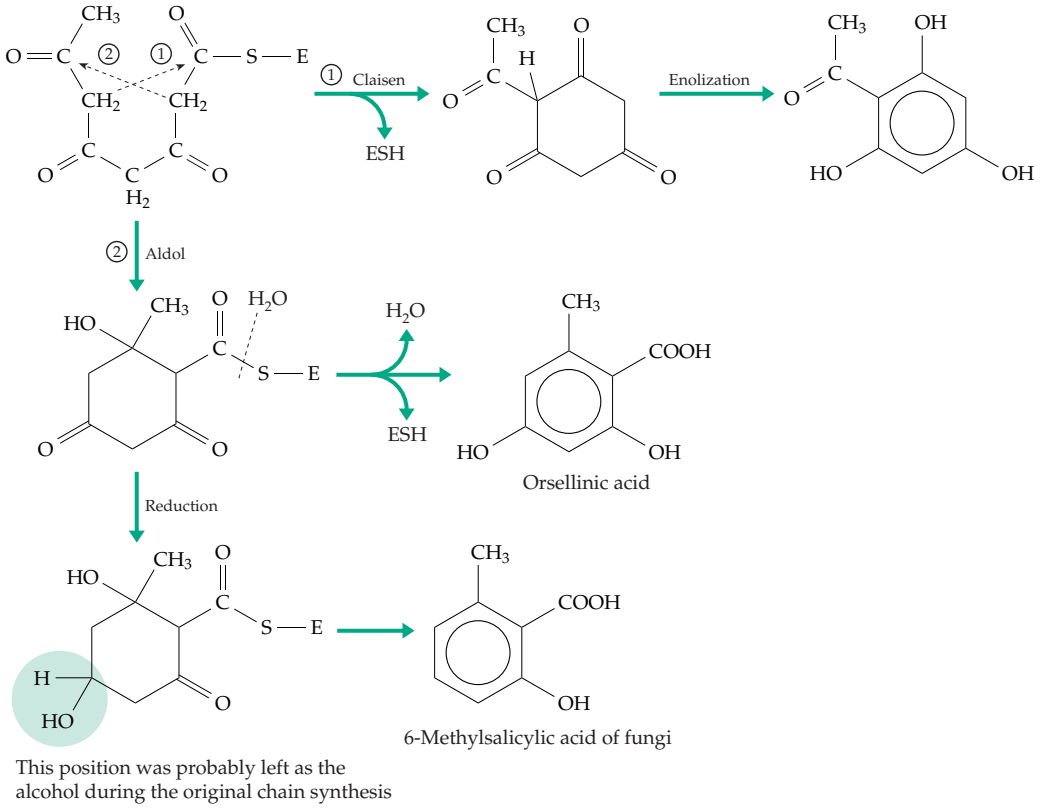
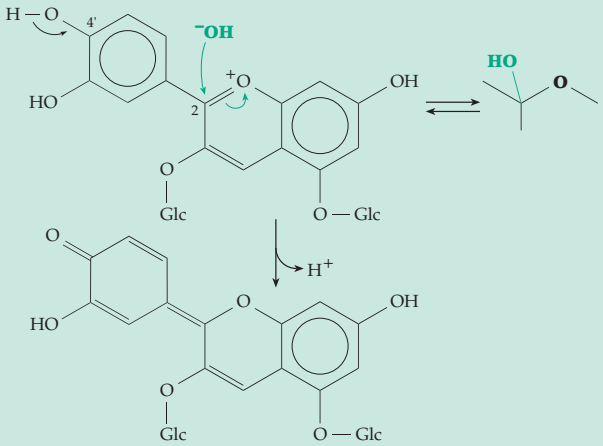


Figure 21-9 Postulated origin of orsellinic acid and other polyketides.

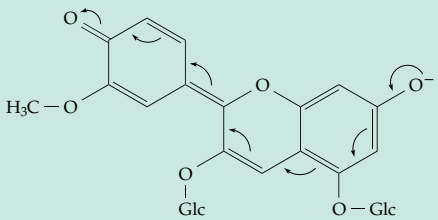
BOX 21-E HOW THE FLOWERS MAKE THEIR COLORS

Most of the pigments of flowers arise from a single polyketide precursor. Phenylalanine is converted to ***trans*-cinnamic acid** (Eq. 14-45) and then to cinnamoyl-CoA. The latter acts as the starter piece for chain elongation via malonyl-CoA (step *a* in the accompanying scheme). The resulting β-polyketone derivative can cyclize in two ways. The aldol condensation (step *b*) leads to **stilbenecarboxylic acid** and to such compounds as **pinosylvin** of pine trees. The Claisen condensation (step *c*) produces **chalcones**, **flavonones**, and **flavones**. These, in turn, can be converted to the yellow **flavonol pigments** and to the red, purple, and blue **anthocyanidins**.[a–c]

At the bottom of the synthetic scheme on the next page the structures and names of three common anthocyanidins are shown. The names are derived from those of flowers from which they have been isolated. The colors depend upon the number of hydroxyl groups and on the presence or absence of methylation and glycosylation. In addition to the three pigments indicated in the diagram, three other common anthocyanidins are formed by methylation. **Peonidin** is 3'-methylcyanidine. Methylation of delphinidin at position 3' yields **petunidin**, while methylation at both the 3' and 5' positions gives **malvidin**. There are many other anthocyanidins of more limited distribution. Anthocyanidins are nearly insoluble, but they exist in plants principally as glycosides known as **anthocyanins**. The number of different glycosides among the many species of flowering plants is large. Both the 3 and 5 –OH groups may be glycosylated with Glc, Gal, Rha, Ara, and by a large variety of oligosaccharides. The colors of the anthocyanins vary from red to violet and blue and are pH dependent. For example, **cyanin** (diglucosyl cyanidin) is red in acid solution and becomes violet upon dissociation of the 4'-hydroxyl group:

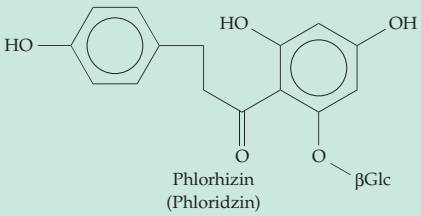


Dissociation of the 7–OH generates an anion with an extended conjugated π electron system, which will favor absorption of long-wavelength light (Chapter 23) and a blue color. Notice that a large number of resonance structures can be drawn for both the anthocyanin and the dissociated forms. Formation of complexes of Mg^{2+} or other metal ions with the 4' $-O^-$ and adjacent OH groups may also stabilize blue colored forms.[d]

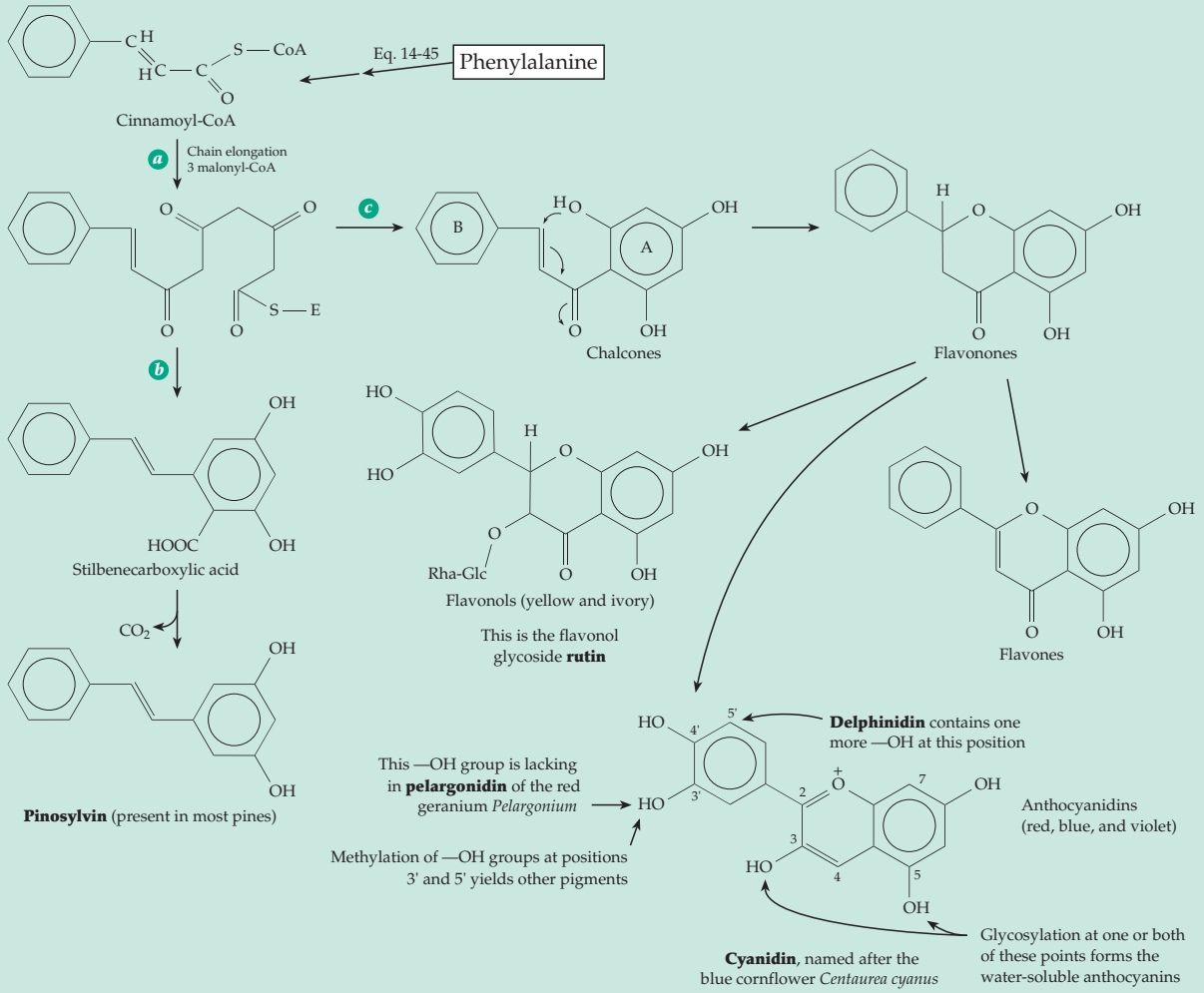


Most blue flower pigments are based on delphinidin,[b] but the "heavenly blue" of the morning glory is a peonidin with a complex caffeoylglucose-containing glycosyl group on the 3-position. Its blue color has been attributed to the relatively high pH of ~7.7 in vacuoles.[e] The aromatic rings within the glycosyl group of this and other complex anthocyanins may fold over the primary chromophore and stabilize the colored forms. A competing reaction, which is indicated in green on the first structure in this box, is the addition of a hydroxyl ion at C-2 to give a nearly colorless adduct.[f]

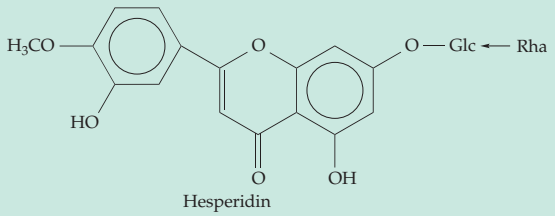
The yellow pigments of flowers are usually flavonols. The most common of all is **rutin**, the 3α-rhamnosyl-D-glucosyl derivative of **quercitin** (see diagram). An extraordinary number of other flavonols, flavones, and related compounds are found throughout the plant kingdom.[g] One of these is **phlorhizin**, a dihydrochalcone found in the root bark of pears, apples, and other plants of the rose family. Phlorhizin specifically blocks resorption of glucose by kidney tubules. As a result, the drug induces a strong glucosuria. The biochemical basis is uncertain, but the action on kidney tubules may be related to inhibition of mutarotase.[h]

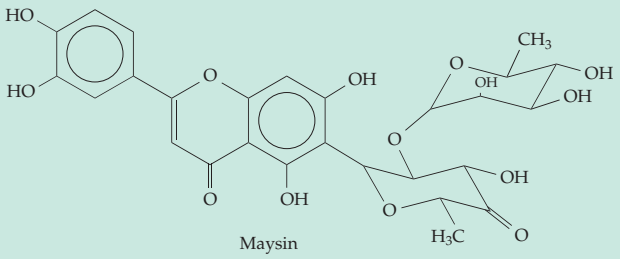


BOX 21-E (continued)



The flavone glycoside **hesperidin** makes up 80% of the dry weight of orange peels. It has been claimed (but not proved) that this compound, also known as **vitamin P** and **citrus bioflavonoid**, is essential to good health. Another flavone, **maysin**, is a resistance factor for the corn earworm and is present in silks of resistant strains of *Zea mays*.[i]





[a] Clevenger, S. (1964) *Sci. Am.* **210**(Jun), 85–92
[b] Harborne, J. B. (1988) in *Plant Pigments* (Goodwin, T. W., ed), pp. 299–343, Academic Press, London
[c] Lloyd, A. M., Walbot, V., and Davis, R. W. (1992) *Science* **258**, 1773–1775
[d] Kondo, T., Yoshida, K., Nakagawa, A., Kawai, T., Tamura, H., and Goto, T. (1992) *Nature (London)* **358**, 515–518
[e] Yoshida, K., Kondo, T., Okazaki, Y., and Katou, K. (1995) *Nature (London)* **373**, 291
[f] Figueiredo, P., Elhabiri, M., Saito, N., and Brouillard, R. (1996) *J. Am. Chem. Soc.* **118**, 4788–4793
[g] Nakayama, T., Yonekura-Sakakibara, K., Sato, T., Kikuchi, S., Fukui, Y., Fukuchi-Mizutani, M., Ueda, T., Nakao, M., Tanaka, Y., Kusumi, T., and Nishino, T. (2000) *Science* **290**, 1163–1166.
[h] White, A., Handler, P., and Smith, E. L. (1973) *Principles of Biochemistry*, 5th ed., McGraw-Hill, New York (pp. 415–416)
[i] Byrne, P. F., McMullen, M. D., Snook, M. E., Musket, T. A., Theuri, J. M., Widstrom, N. W., Wiseman, B. R., and Coe, E. H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8820–8825

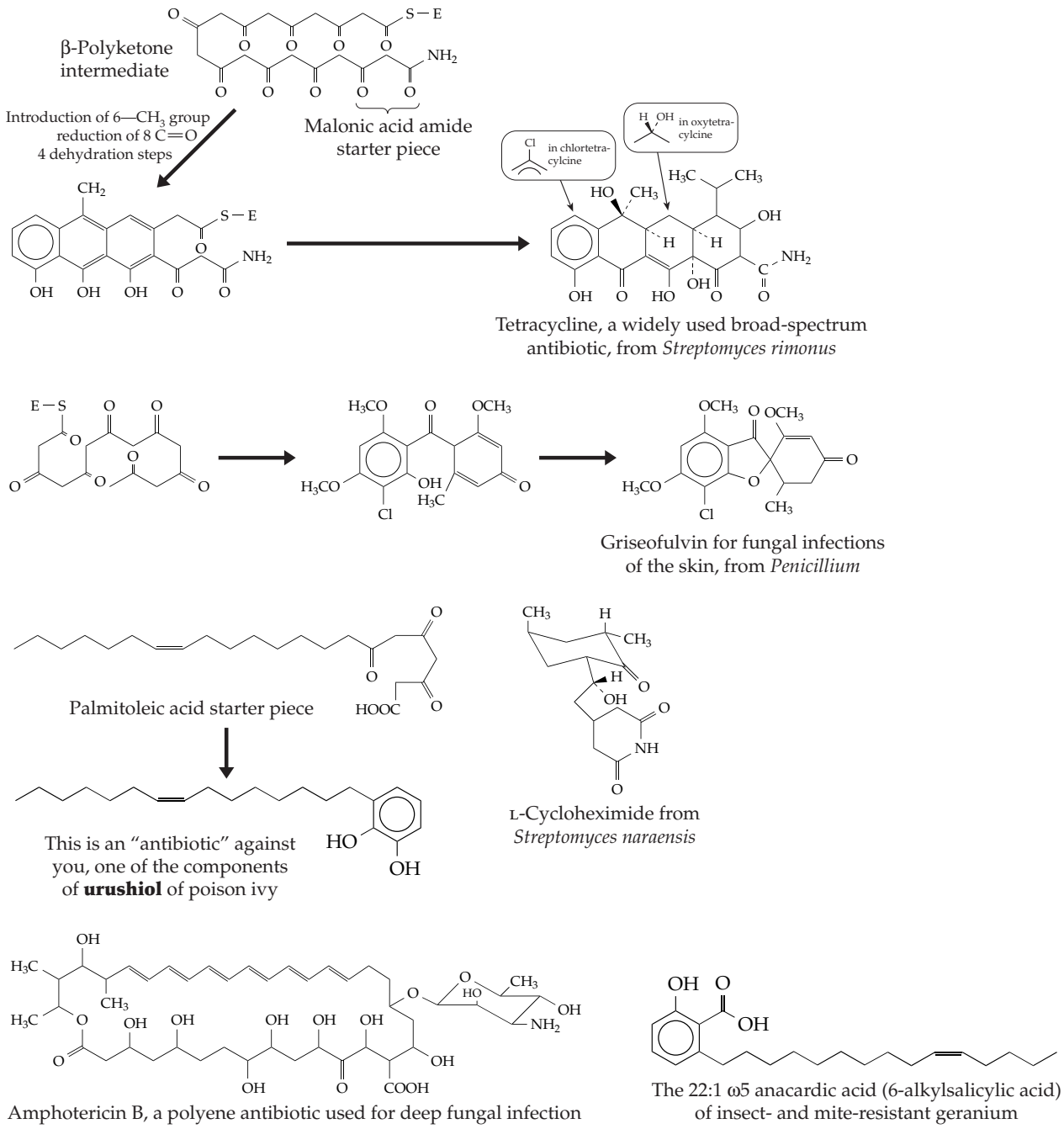


Figure 21-10 Some important polyketide antibiotics and plant defensive compounds.

explosion of information about polyketide synthases.[336,336a–c]

A 26-kb gene cluster encoding enzymes for synthesis of the blue antibiotic **actinorhodin** by *Streptomyces coelicolor* has been cloned and sequenced.[332,337] The three large ~10-kb genes required for formation of the broad-spectrum antibiotic **erythromycin** by *Saccharopolyspora erythraea* have also been cloned and sequenced.[337–339] In both cases, the genes encode large proteins with structures resembling those of the eukaryotic fatty acid synthases (Section B,1). However, a new feature is evident. As shown in Fig. 21-11, each of the three polypeptides of the deoxyerythronolide synthase, which synthesizes the aglycone of erythromycin, consists of two multidomain modules, each able to catalyze one round of reaction with a new molecule of malonyl-CoA. When reduction of an oxo group or dehydration and

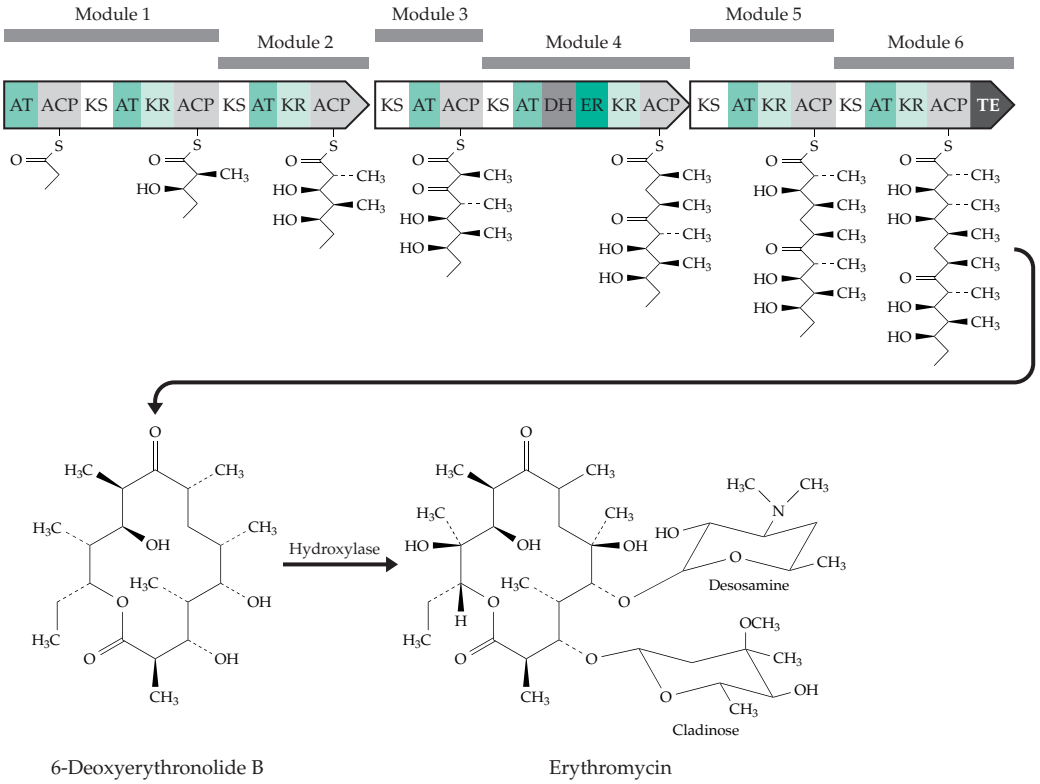


Figure 21-11 Catalytic domains within three polypeptide chains of the modular polyketide synthase that forms 6-deoxyerythronolide B, the aglycone of the widely used antibiotic erythromycin. The domains are labeled as for fatty acid synthases; AT, acyltransferase; ACP, acyl carrier protein; KS, β-ketoacyl-ACP synthase; KR, ketoreductase; DH, dehydrase; ER, enoylreductase; TE, thioesterase. After Pieper *et al.*[338] Courtesy of Chaitan Khosla.

reduction of an enoyl-CoA are not needed in a round, the KR, DH, and ER domains are absent (as in module 3 of Fig. 21-11). A final domain contains a thioesterase that releases and cyclizes the product. The "assembly line" sequence of synthetic steps, beginning with a propionyl group from propionyl-CoA, is pictured in Fig. 21-11. Two hydroxylation steps[340] and transfer of two unusual glycosyl groups complete the synthesis of the antibiotic.

Other medically important polyketides include the antibiotics **doxorubicin** (14-hydroxydaunomycin; Fig. 5-23),[341] rifamycin (Box 28-A),[342] and the antifungal **pimaricin**,[343] **griseofulvin**, and **amphotericin** (Fig. 21-10), the HMG-CoA reductase inhibitor **lovastatin**,[344] the 2-butanyl-4-methylthreonine of cyclosporin A (Box 9-F),[345] and other immunosuppressants such as **rapamycin**.[346] Many characteristic plant products, including **stilbenes**[347] and **chalcones**[348,348a] (Box 21-E), are polyketides. A variety of different polyketides serve as phytoalexins.[349] Some such as **aflatoxin**[350] are dangerous toxins. Ants and ladybird beetles make toxic polyamine alkaloids using a polyketide pathway.[350a]

Avermectin (Fig. 30–25), a widely used antibiotic against canine heartworms, is formed by a polyketide synthase with an unusually broad specificity for starter units. More than 40 alternative carboxylic acids are accepted. By grafting the first multidomain module of the erythromycin-forming synthase (of Fig. 21-11) onto the wide-specificity loading module of the avermectin-forming synthase, a whole new series of antibiotics have been created.[351] This is only one of many steps being taken to create new aliphatic and aromatic, linear and macrocyclic polyketides by genetic engineering.[336,352,353] Combinatorial biosynthesis (see Chapter 3) is also being developed[336,354] and has even been discovered in nature.[355]

References

1. Ayvazian, L., Crenon, I., Hermoso, J., Pignol, D., Chapus, C., and Kerfelec, B. (1998) *J. Biol. Chem.* **273**, 33604–33609
2. Bezzine, S., Ferrato, F., Ivanova, M. G., Lopez, V., Verger, R., and Carrière, F. (1999) *Biochemistry* **38**, 5499–5510

2a. Crandall, W. V., and Lowe, M. E. (2001) *J. Biol. Chem.* **276**, 12505–12512

3. Hermoso, J., Pignol, D., Kerfelec, B., Crenon, I., Chapus, C., and Fontecilla-Camps, J. C. (1996) *J. Biol. Chem.* **271**, 18007–18016
4. Kobayashi, J., Applebaum-Bowden, D., Dugi, K. A., Brown, D. R., Kashyap, V. S., Parrott, C., Duarte, C., Maeda, N., and Santamarina-Fojo, S. (1996) *J. Biol. Chem.* **271**, 26296–26301
5. Winkler, F. K., D'Arcy, A., and Hunziker, W. (1990) *Nature (London)* **343**, 771–774
6. Levy, E., Mehran, M., and Seidman, E. (1995) *FASEB J.* **9**, 626–635
7. Havel, R. J., and Kane, J. P. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1841–1852, McGraw-Hill, New York

7a. Liu, H., Talmud, P. J., Lins, L., Brasseur, R., Olivecrona, G., Peelman, F., Vandekerckhove, J., Rosseneu, M., and Labeur, C. (2000) *Biochemistry* **39**, 9201–9212

7b. Rozek, A., Sparrow, J. T., Weisgraber, K. H., and Cushley, R. J. (1999) *Biochemistry* **38**, 14475–14484

8. Orlova, E. V., Sherman, M. B., Chiu, W., Mowri, H., Smith, L. C., and Gotto, A. M., Jr. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8420–8425
9. Brown, M. S., and Goldstein, J. L. (1974) *Sci. Am.* **251**(Nov), 58–66
10. Phillips, J. C., Wriggers, W., Li, Z., Jonas, A., and Schulten, K. (1997) *Biophys. J.* **73**, 2337–2346

10a. Segrest, J. P., Jones, M. K., Klon, A. E., Sheldahl, C. J., Hellinger, M., De Loof, H., and Harvey, S. C. (1999) *J. Biol. Chem.* **274**, 31755–31758

10b. Tricerri, M. A., Agree, A. K. B., Sanchez, S. A., Bronski, J., and Jonas, A. (2001) *Biochemistry* **40**, 5065–5074

11. Rosseneu, M., and Labeur, C. (1995) *FASEB J.* **9**, 768–776
12. Chan, L. (1992) *J. Biol. Chem.* **267**, 25621–25624
13. Ferguson, S. J. (1987) *Trends Biochem. Sci.* **12**, 353–357
14. MacGinnitie, A. J., Anant, S., and Davidson, N. O. (1995) *J. Biol. Chem.* **270**, 14768–14775
15. Richardson, N., Navaratnam, N., and Scott, J. (1998) *J. Biol. Chem.* **273**, 31707–31717
16. Lawn, R. M. (1992) *Sci. Am.* **266**(Jun), 54–60
17. Utermann, G. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1887–1912, McGraw-Hill, New York
18. Trieu, V. N., and McConathy, W. J. (1995) *J. Biol. Chem.* **270**, 15471–15474
19. Mochalkin, I., Cheng, B., Klezovitch, O., Scanu, A. M., and Tulinsky, A. (1999) *Biochemistry* **38**, 1990–1998

19a. Wang, J., and White, A. L. (1999) *J. Biol. Chem.* **274**, 12883–12889

20. Fless, G. M., Santiago, J. Y., Furbee, J., Jr., and Meredith, S. C. (1997) *Biochemistry* **36**, 11304–11313
21. Williams, K. J., Fless, G. M., Petrie, K. A., Snyder, M. L., Brocia, R. W., and Swenson, T. L. (1992) *J. Biol. Chem.* **267**, 13284–13292

21a. Goldstein, J. L., and Brown, M. S. (2001) *Science* **292**, 1310–1312

22. Baggio, G., Donazzan, S., Monti, D., Mari, D., Martini, S., Gabelli, C., Dalla Vestra, M., Previato, L., Guido, M., Pigozzo, S., Cortella, I., Crepaldi, G., and Franceschi, C. (1998) *FASEB J.* **12**, 433–437
23. Bergeron, J., Frank, P. G., Scales, D., Meng, Q.-H., Castro, G., and Marcel, Y. L. (1995) *J. Biol. Chem.* **270**, 27429–27438
24. Deeb, S. S., Cheung, M. C., Peng, R., Wolf, A. C., Stern, R., Albers, J. J., and Knopp, R. H. (1991) *J. Biol. Chem.* **266**, 13654–13660
25. Laccotripe, M., Makrides, S. C., Jonas, A., and Zannis, V. I. (1997) *J. Biol. Chem.* **272**, 17511–17522

25a. Gorshkova, I. N., Liu, T., Zannis, V. I., and Atkinson, D. (2002) *Biochemistry* **41**, 10529–10539

25b. Klon, A. E., Segrest, J. P., and Harvey, S. C. (2002) *Biochemistry* **41**, 10895–10905

26. Borhani, D. W., Rogers, D. P., Engler, J. A., and Brouillette, C. G. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12291–12296
27. Wang, G., Sparrow, J. T., and Cushley, R. J. (1997) *Biochemistry* **36**, 13657–13666
28. Rogers, D. P., Roberts, L. M., Lebowitz, J., Engler, J. A., and Brouillette, C. G. (1998) *Biochemistry* **37**, 945–955
29. Lopez, J., Roghani, A., Bertrand, J., Zanni, E., Kalopissis, A., Zannis, V. I., and Chambaz, J. (1994) *Biochemistry* **33**, 4056–4064
30. Boisfer, E., Lambert, G., Atger, V., Tran, N. Q., Pastier, D., Benetollo, C., Trottier, J.-F., Beaucamps, I., Antonucci, M., Laplaud, M., Griglio, S., Chambaz, J., and Kalopissis, A.-D. (1999) *J. Biol. Chem.* **274**, 11564–11572

30a. Pennacchio, L. A., Olivier, M., Hubacek, J. A., Cohen, J. C., Cox, D. R., Fruchart, J.-C., Krauss, R. M., and Rubin, E. M. (2001) *Science* **294**, 169–173

31. Dong, L.-M., Wilson, C., Wardell, M. R., Simmons, T., Mahley, R. W., Weisgraber, K. H., and Agard, D. A. (1994) *J. Biol. Chem.* **269**, 22358–22365
32. Wilson, C., Wardell, M. R., Weisgraber, K. H., Mahley, R. W., and Agard, D. A. (1991) *Science* **252**, 1817–1822
33. Mahley, R. W., and Rall, S. C. J. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1953–1980, McGraw-Hill, New York

33a. Kypreos, K. E., Morani, P., van Dijk, K. W., Havekes, L. M., and Zannis, V. I. (2001) *Biochemistry* **40**, 6027–6035

33b. Morrow, J. A., Arnold, K. S., Dong, J., Balestra, M. E., Innerarity, T. L., and Weisgraber, K. H. (2000) *J. Biol. Chem.* **275**, 2576–2580

34. Law, J. H., and Wells, M. A. (1989) *J. Biol. Chem.* **264**, 16335–16338
35. Breiter, D. R., Kanost, M. R., Benning, M. M., Wesenberg, G., Law, J. H., Wells, M. A., Rayment, I., and Holden, H. M. (1991) *Biochemistry* **30**, 603–608
36. Zhang, Y., Lewis, R. N. A. H., McElhaney, R. N., and Ryan, R. O. (1993) *Biochemistry* **32**, 3942–3952
37. Soulages, J. L., Rivera, M., Walker, F. A., and Wells, M. A. (1994) *Biochemistry* **33**, 3245–3251

37a. Dettloff, M., Weers, P. M. M., Niere, M., Kay, C. M., Ryan, R. O., and Wiesner, A. (2001) *Biochemistry* **40**, 3150–3157

38. Heinemann, B., Andersen, K. V., Nielsen, P. R., Bech, L. M., and Poulsen, F. M. (1996) *Protein Sci.* **5**, 13–23
39. Banaszak, L., Sharrock, W., and Timmins, P. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 221–246
40. Deckelbaum, R. J., Ramakrishnan, R., Eisenberg, S., Olivecrona, T., and Bengtsson-Olivecrona, G. (1992) *Biochemistry* **31**, 8544–8551
41. Brunzell, J. D. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1913–1932, McGraw-Hill, New York
42. van Tilbeurgh, H., Egloff, M.-P., Martinez, C., Rugani, N., Verger, R., and Cambillau, C. (1993) *Nature (London)* **362**, 814–820

42a. Borén, J., Lookene, A., Makoveichuk, E., Xiang, S., Gustafsson, M., Liu, H., Talmud, P., and Olivecrona, G. (2001) *J. Biol. Chem.* **276**, 26916–26922

42b. Hime, N. J., Barter, P. J., and Rye, K.-A. (1998) *J. Biol. Chem.* **273**, 27191–27198

43. Lookene, A., Savonen, R., and Olivecrona, G. (1997) *Biochemistry* **36**, 5267–5275
44. Dichek, H. L., Fojo, S. S., Beg, O. U., Skarlatos, S. I., Brunzell, J. D., Cutler, G. B., Jr., and Brewer, H. B., Jr. (1991) *J. Biol. Chem.* **266**, 473–477
45. Ma, Y., Bruin, T., Tuzgol, S., Wilson, B. I., Roederer, G., Liu, M.-S., Davignon, J., Kastelein, J. J. P., Brunzell, J. D., and Hayden, M. R. (1992) *J. Biol. Chem.* **267**, 1918–1923
46. Dugi, K. A., Dichek, H. L., Talley, G. D., Brewer, J., HB, and Santamarina-Fojo, S. (1992) *J. Biol. Chem.* **267**, 25086–25091
47. Mamo, J. C. L., Elsegood, C. L., Gennat, H. C., and Yu, K. (1996) *Biochemistry* **35**, 10210–10214
48. Strickland, D. K., Kounnas, M. Z., and Argraves, W. S. (1995) *FASEB J.* **9**, 890–898
49. Sakai, J., Hoshino, A., Takahashi, S., Miura, Y., Ishii, H., Suzuki, H., Kawarabayasi, Y., and Yamamoto, T. (1994) *J. Biol. Chem.* **269**, 2173–2182
50. Willnow, T. E., Sheng, Z., Ishibashi, S., and Herz, J. (1994) *Science* **264**, 1471–1474
51. Hajjar, D. P., and Haberland, M. E. (1997) *J. Biol. Chem.* **272**, 22975–22978
52. Suzuki, K., Yamada, T., and Tanaka, T. (1999) *Biochemistry* **38**, 1751–1756
53. Krieger, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4077–4080

53a. Bhattacharya, A. A., Grüne, T., and Curry, S. (2000) *J. Mol. Biol.* **303**, 721–732

53b. Saltiel, A. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 535–537

53c. De Simone, G., Galdiero, S., Manco, G., Lang, D., Rossi, M., and Pedone, C. (2000) *J. Mol. Biol.* **303**, 761–771

54. Holak, T. A., Kearsley, S. K., Kim, Y., and Prestegard, J. H. (1988) *Biochemistry* **27**, 6135–6142
55. Heath, R. J., and Rock, C. O. (1995) *J. Biol. Chem.* **270**, 15531–15538
56. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. 1998) *EMBO J.* **17**, 1183–1191
57. Heath, R. J., and Rock, C. O. (1995) *J. Biol. Chem.* **270**, 26538–26542
58. Joshi, A. K., Witkowski, A., and Smith, S. (1998) *Biochemistry* **37**, 2515–2523

58a. Zhang, Y.-M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) *J. Biol. Chem.* **276**, 8231–8238

58b. Scarsdale, J. N., Kazanina, G., He, X., Reynolds, K. A., and Wright, H. T. (2001) *J. Biol. Chem.* **276**, 20516–20522

58c. Qiu, X., Janson, C. A., Smith, W. W., Head, M., Lonsdale, J., and Konstantinidis, A. K. (2001) *J. Mol. Biol.* **307**, 341–356

References

58d. McGuire, K. A., Siggaard-Andersen, M., Bangera, M. G., Olsen, J. G., and von Wettstein-Knowles, P. (2001) *Biochemistry* **40**, 9836–9845
58e. Moche, M., Dehesh, K., Edwards, P., and Lindqvist, Y. (2001) *J. Mol. Biol.* **305**, 491–503
59. Jayakumar, A., Tai, M.-H., Huang, W.-Y., Al-Feel, W., Hsu, M., Abu-Elheiga, L., Chirala, S. S., and Wakil, S. J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8695–8699
59a. Chirala, S. S., Jayakumar, A., Gu, Z.-W., and Wakil, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3104–3108
59b. Witkowski, A., Joshi, A. K., and Smith, S. (2002) *Biochemistry* **41**, 10877–10887
60. McCarthy, A. D., and Hardie, D. G. (1984) *Trends Biochem. Sci.* **9**, 60–63
61. Joshi, A. K., Witkowski, A., and Smith, S. (1997) *Biochemistry* **36**, 2316–2322
62. Witkowski, A., Joshi, A. K., Rangan, V. S., Falick, A. M., Witkowska, H. E., and Smith, S. (1999) *J. Biol. Chem.* **274**, 11557–11563
63. Wakil, S. J. (1989) *Biochemistry* **28**, 4523–4530
64. Jayakumar, A., Chirala, S. S., and Wakil, S. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12326–12330
65. Smith, S. (1994) *FASEB J.* **8**, 1248–1259
66. Wakil, S. J., Stoops, J. K., and Joshi, V. C. (1983) *Ann. Rev. Biochem.* **52**, 537–579
67. Kolodziej, S. J., Penczek, P. A., Schroeter, J. P., and Stoops, J. K. (1996) *J. Biol. Chem.* **271**, 28422–28429
68. Orme, T. W., McIntyre, J., Lynen, F., Kühn, L., and Schweizer, E. (1972) *Eur. J. Biochem.* **24**, 407–415
69. Poulose, A. J., Rogers, L., Cheesbrough, T. M., and Kolattukudy, P. E. (1985) *J. Biol. Chem.* **260**, 15953–15958
70. Witkowska, H. E., Green, B. N., and Smith, S. (1990) *J. Biol. Chem.* **265**, 5662–5665
71. Smith, S., and Abraham, S. (1975) *Adv. Lipid Res.* **13**, 195–239
72. Stumpf, P. K. (1981) *Trends Biochem. Sci.* **6**, 173–176
72a. Murakami, Y., Tsuyama, M., Kobayashi, Y., Kodama, H., and Iba, K. (2000) *Science* **287**, 476–479
73. Ohlrogge, J. B., Jaworski, J. G., and Post-Beittenmiller, D. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 3–32, CRC Press, Boca Raton, Florida
74. Heinz, E. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 33–90, CRC Press, Boca Raton, Florida
75. Stumpf, P. K. (1984) in *Fatty Acid Metabolism and its Regulation: New Comprehensive Biochemistry*, Vol. 7 (Numa, S., ed.), pp. 155–179, Elsevier, Amsterdam
75a. Lyle, K. S., Möenne-Loccoz, P., Ai, J., Sanders-Loehr, J., Loehr, T. M., and Fox, B. G. (2000) *Biochemistry* **39**, 10507–10513
76. Choi, J.-Y., and Martin, C. E. (1999) *J. Biol. Chem.* **274**, 4671–4683
77. Watkins, P. A., Lu, J.-F., Steinberg, S. J., Gould, S. J., Smith, K. D., and Braiterman, L. T. (1998) *J. Biol. Chem.* **273**, 18210–18219
78. Whereat, A. F., Orishimo, M. W., Nelson, J., and Phillips, S. J. (1969) *J. Biol. Chem.* **244**, 6498–6506
78a. Das, A. K., Uhler, M. D., and Hajra, A. K. (2000) *J. Biol. Chem.* **275**, 24333–24340
79. Lin, C. Y., and Kumar, S. (1971) *J. Biol. Chem.* **246**, 3284–3290
80. Varanasi, U., Feldman, H. R., and Malins, D. C. (1975) *Nature (London)* **255**, 340–343
81. Kolattukudy, P. E. (1968) *Science* **159**, 498–505
82. Dillwith, J. W., Nelson, J. H., Pomonis, J. G., Nelson, D. R., and Blomquist, G. J. (1982) *J. Biol. Chem.* **257**, 11305–11314
83. Odham, G., and Stenhagen, E. (1971) *Acc. Chem. Res.* **4**, 121–128
84. Cole, S. T., and and 41 other authors. (1998) *Nature (London)* **393**, 537–544
85. Azad, A. K., Sirakova, T. D., Fernandes, N. D., and Kolattukudy, P. E. (1997) *J. Biol. Chem.* **272**, 16741–16745
86. Kroumova, A. B., Xie, Z., and Wagner, G. J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11437–11441
87. Howell, D. M., Harich, K., Xu, H., and White, R. H. (1998) *Biochemistry* **37**, 10108–10117
88. Hayden, M. A., Huang, I. Y., Iliopoulos, G., Orozco, M., and Ashley, G. W. (1993) *Biochemistry* **32**, 3778–3782
89. Sulo, P., and Martin, N. C. (1993) *J. Biol. Chem.* **268**, 17634–17639
90. Wada, H., Shintani, D., and Ohlrogge, J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1591–1596
91. Busby, R. W., Schelvis, J. P. M., Yu, D. S., Babcock, G. T., and Marletta, M. A. (1999) *J. Am. Chem. Soc.* **121**, 4706–4707
91a. Miller, J. R., Busby, R. W., Jordan, S. W., Cheek, J., Henshaw, T. F., Ashley, G. W., Broderick, J. B., Cronan, J. E., Jr., and Marletta, M. A. (2000) *Biochemistry* **39**, 15166–15178
91b. Gueguent, V., Macherel, D., Jaquinod, M., Douce, R., and Bourguignon, J. (2000) *J. Biol. Chem.* **275**, 5016–5025
92. Duin, E. C., Lafferty, M. E., Crouse, B. R., Allen, R. M., Sanyal, I., Flint, D. H., and Johnson, M. K. (1997) *Biochemistry* **36**, 11811–11820
93. Hayden, M. A., Huang, I., Bussiere, D. E., and Ashley, G. W. (1992) *J. Biol. Chem.* **267**, 9512–9515
93a. Ugulava, N. B., Sacanell, C. J., and Jarrett, J. T. (2001) *Biochemistry* **40**, 8352–8358
94. White, R. H. (1987) *Biochemistry* **26**, 3163–3167
95. deMendoza, D., and Cronan, J. E., Jr. (1983) *Trends Biochem. Sci.* **8**, 49–52
95a. Wallis, J. G., Watts, J. L., and Browse, J. (2002) *Trends Biochem. Sci.* **27**, 467–473
96. Henderson, B. S., Larsen, B. S., and Schwab, J. M. (1994) *J. Am. Chem. Soc.* **116**, 5025–5034
97. Roughan, G., and Slack, R. (1984) *Trends Biochem. Sci.* **9**, 383–386
98. Sperling, P., Linscheid, M., Stöcker, S., Mühlback, H.-P., and Heinz, E. (1993) *J. Biol. Chem.* **268**, 26935–26940
99. Jeffcoat, R., and James, A. T. (1984) in *Fatty Acid Metabolism and its Regulation: New Comprehensive Biochemistry*, Vol. 7 (Numa, S., ed), pp. 85–112, Elsevier, Amsterdam
100. Thiede, M. A., and Strittmatter, P. (1985) *J. Biol. Chem.* **260**, 14459–14463
101. Shanklin, J., Whittle, E., and Fox, B. G. (1994) *Biochemistry* **33**, 12787–12794
102. Buist, P. H., and Behrouzian, B. (1998) *J. Am. Chem. Soc.* **120**, 871–876
103. Vergères, G., Ramsden, J., and Waskell, L. (1995) *J. Biol. Chem.* **270**, 3414–3422
104. Mitchell, A. G., and Martin, C. E. (1995) *J. Biol. Chem.* **270**, 29766–29772
105. McKeon, T. A., and Stumpf, P. K. (1982) *J. Biol. Chem.* **257**, 12141–12147
106. Somerville, C., and Browse, J. (1991) *Science* **252**, 80–87
107. Töpfer, R., Martini, N., and Schell, J. (1995) *Science* **268**, 681–686
108. Browse, J., McConn, M., James, D., Jr., and Miquel, M. (1993) *J. Biol. Chem.* **268**, 16345–16351
109. Lindqvist, Y., Huang, W., Schneider, G., and Shanklin, J. (1996) *EMBO J.* **15**, 4081–4092
110. Fox, B. G., Shanklin, J., Ai, J., Loehr, T. M., and Sanders-Loehr, J. (1994) *Biochemistry* **33**, 12776–12786
111. Spychalla, J. P., Kinney, A. J., and Browse, J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1142–1147
112. Mustardy, L., Los, D. A., Gombos, Z., and Murata, N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10524–10527
113. Broun, P., Shanklin, J., Whittle, E., and Somerville, C. (1998) *Science* **282**, 1315–1317
114. van de Loo, F. J., Broun, P., Turner, S., and Somerville, C. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6743–6747
115. Kaya, K., Ramesha, C. S., and Thompson, G. A., Jr. (1984) *J. Biol. Chem.* **259**, 3548–3553
116. Ortiz de Montellano, P. R., Chan, W. K., Tuck, S. F., Kaikaus, R. M., Bass, N. M., and Peterson, J. A. (1992) *FASEB J.* **6**, 695–699
117. Vaz, A. D. N., McGinnity, D. F., and Coon, M. J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3555–3560
118. de Renobales, M., Cripps, C., Stanley-Samuelson, D. W., Jurenka, R. A., and Blomquist, G. J. (1987) *Trends Biochem. Sci.* **12**, 364–366
119. Cho, H. P., Nakamura, M. T., and Clarke, S. D. (1999) *J. Biol. Chem.* **274**, 471–477
119a. Bernoud-Hubac, N., Davies, S. S., Boutaud, O., Montine, T. J., and Roberts, L. J., II. (2001) *J. Biol. Chem.* **276**, 30964–30970
119b. Mirnikjoo, B., Brown, S. E., Kim, H. F. S., Marangell, L. B., Sweatt, J. D., and Weeber, E. J. (2001) *J. Biol. Chem.* **276**, 10888–10896
119c. Brzustowicz, M. R., Cherezov, V., Zerouga, M., Caffrey, M., Stillwell, W., and Wassall, S. R. (2002) *Biochemistry* **41**, 12509–12519
119d. Albert, C. M., Campos, H., Stampfer, M. J., Ridker, P. M., Manson, J. E., Willett, W. C., and Ma, J. (2002) *M. Engl. J. Med.* **346**, 1113–1118 (and discussion in Vol. 347, pp. 531–533)
120. Knipple, D. C., Rosenfield, C.-L., Miller, S. J., Liu, W., Tang, J., Ma, P. W. K., and Roelofs, W. L. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15287–15292
121. Aveldaño, M. I., and Sprecher, H. (1997) *J. Biol. Chem.* **262**, 1180–1186
121a. Cahoon, E. B., Ripp, K. G., Hall, S. E., and Kinney, A. J. (2001) *J. Biol. Chem.* **276**, 2637–2643
121b. Adlof, R. O., Duval, S., and Emken, E. A. (2000) *Lipids* **35**, 131–135
121c. Majumder, B., Wahle, K. W. J., Moir, S., Schofield, A., Choe, S.-N., Farquharson, A., Grant, I., and Heys, S. D. (2002) *FASEB J.* **16**, 1447-1449
122. Wise, M. L., Hamberg, M., and Gerwick, W. H. (1994) *Biochemistry* **33**, 15223–15232
123. Wise, M. L., Rossi, J., and Gerwick, W. H. (1997) *Biochemistry* **36**, 2985–2992
124. Cox, G. S., Thomas, E., Kaback, H. R., and Weissbach, H. (1973) *Arch. Biochem. Biophys.* **158**, 667–676
125. Wang, A.-Y., Grogan, D. W., and Cronan, J. E., Jr. (1992) *Biochemistry* **31**, 11020–11028
126. Packter, N. M. (1973) *Biosynthesis of Acetate-Derived Compounds*, Wiley, New York
126a. Glickman, M. S., Cahill, S. M., and Jacobs, W. R., Jr. (2001) *J. Biol. Chem.* **276**, 2228–2233
127. Yuan, Y., Crane, D. C., Musser, J. M., Sreevatsan, S., and Barry, C. E., III. (1997) *J. Biol. Chem.* **272**, 10041–10049
128. Yuan, Y., Lee, R. E., Besra, G. S., Belisle, J. T., and Barry, C. E., III. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6630–6634
129. Liu, J., and Nikaido, H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4011–4016
130. Tipton, C. L., and Al-Shathir, N. M. (1974) *J. Biol. Chem.* **249**, 886–889
131. Bohlmann, F., Burkhardt, T., and Zdero, C. (1973) *Naturally Occuring Acetylenes*, Academic Press, New York

References

132. Nicolaides, N. (1974) *Science* **186**, 19–26
133. Bauermeister, A., and Sargent, J. R. (1979) *Trends Biochem. Sci.* **4**, 209–211
134. Nemes, Z., Marekov, L. N., Fésüs, L., and Steinert, P. M. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8402–8407
135. Kolattukudy, P. E. (1980) *Science* **208**, 990–1000
135a. Wellesen, K., Durst, F., Pinot, F., Benveniste, I., Nettesheim, K., Wisman, E., Steiner-Lange, S., Saedler, H., and Yephremov, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9694–9699
136. von Wettstein-Knowles, P. M. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 127–166, CRC Press, Boca Raton, Florida
137. Dennis, M., and Kolattukudy, P. E. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5306–5310
138. Cheesbrough, T. M., and Kolattukudy, P. E. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6613–6617
139. Cheesbrough, T. M., and Kolattukudy, P. E. (1988) *J. Biol. Chem.* **263**, 2738–2743
140. Reed, J. R., Vanderwel, D., Choi, S., Pomonis, J. G., Reitz, R. C., and Blomquist, G. J. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10000–10004
141. Gorgen, G., and Boland, W. (1989) *Eur. J. Biochem.* **185**, 237–242
142. Buckner, J. S., Nelson, D. R., Fatland, C. L., Hakk, H., and Pomonis, J. G. (1984) *J. Biol. Chem.* **259**, 8461–8470
143. Morse, D., and Meighen, E. (1984) *Science* **226**, 1434–1436
144. Hölldobler, B. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 19–22
145. Prestwich, G. D. (1987) *Science* **237**, 999 – 1006
146. Dussourd, D. E., Harvis, C. A., Meinwald, J., and Eisner, T. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9224–9227
146a. Athenstaedt, K., and Daum, G. (2000) *J. Biol. Chem.* **275**, 235–240
146b. Kalhan, S. C., Mahajan, S., Burkett, E., Reshef, L., and Hanson, R. W. (2001) *J. Biol. Chem.* **276**, 12928–12931
147. Rock, C. O., Goelz, S. E., and Cronan, J. E., Jr. (1981) *J. Biol. Chem.* **256**, 736–742
148. Eberhardt, C., Gray, P. W., and Tjoelker, L. W. (1997) *J. Biol. Chem.* **272**, 20299–20305
149. Guo, Z., Liliom, K., Fischer, D. J., Bathurst, I. C., Tomei, L. D., Kiefer, M. C., and Tigyi, A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14367–14372
150. Kai, M., Wada, I., Imai, S.-i, Sakane, F., and Kanoh, H. (1997) *J. Biol. Chem.* **272**, 24572–24578
150a. Buhman, K. K., Chen, H. C., and Farese, R. V., Jr. (2001) *J. Biol. Chem.* **276**, 40369–40372
150b. Abo-Hashema, K. A. H., Cake, M. H., Power, G. W., and Clarke, D. (1999) *J. Biol. Chem.* **274**, 35577–35582
151. Shen, H., Heacock, P. N., Clancey, C. J., and Dowhan, W. (1996) *J. Biol. Chem.* **271**, 789–795
152. Badola, P., and Sanders, C. R., II. (1997) *J. Biol. Chem.* **272**, 24176–24182
153. Maréchal, E., Block, M. A., Joyard, J., and Douce, R. (1994) *J. Biol. Chem.* **269**, 5788–5798
154. Joyard, J., Block, M. A., Malherbe, A., Maréchal, E., and Douce, R. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 231–258, CRC Press, Boca Raton, Florida
155. Dörmann, P., Balbo, I., and Benning, C. (1999) *Science* **284**, 2181–2184
156. Karlsson, O. P., Rytömaa, M., Dahlqvist, A., Kinnunen, P. K. J., and Wieslander, A. (1996) *Biochemistry* **35**, 10094–10102
156a. Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A. H., and Wakil, S. J. (2001) *Science* **291**, 2613–2616
156b. Jelenska, J., Crawford, M. J., Harb, O. S., Zuther, E., Haselkorn, R., Roos, D. S., and Gornicki, P. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2723–2728
157. Thompson, G. A., Jr. (1992) *The Regulation of Membrane Lipid Metabolism*, 2nd ed., CRC Press, Boca Raton, Florida
158. Lopez, J. M., Bennett, M. K., Sanchez, H. B., Rosenfeld, J. M., and Osborne, T. F. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1049–1053
159. Honnor, R. C., Dhillon, G. S., and Londos, C. (1985) *J. Biol. Chem.* **260**, 15130–15138
160. Tabor, D. E., Kim, J. B., Spiegelman, B. M., and Edwards, P. A. (1999) *J. Biol. Chem.* **274**, 20603–20610
161. Soncini, M., Yet, S.-F., Moon, Y., Chun, J.-Y., and Sul, H. S. (1995) *J. Biol. Chem.* **270**, 30339–30343
162. Dudek, S. M., and Semenkovich, C. F. (1995) *J. Biol. Chem.* **270**, 29323–29329
163. Tiku, P. E., Gracey, A. Y., Macartney, A. I., Beynon, R. J., and Cossins, A. R. (1996) *Science* **271**, 815–818
164. Jackson, B. J., Gennity, J. M., and Kennedy, E. P. (1986) *J. Biol. Chem.* **261**, 13464–13468
165. Raetz, C. R. H., and Dowhan, W. (1990) *J. Biol. Chem.* **265**, 1235–1238
166. Kennedy, E. P. (1992) *Ann. Rev. Biochem.* **61**, 1–28
167. Bae-Lee, M. S., and Carman, G. M. (1984) *J. Biol. Chem.* **259**, 10857–10862
168. Kawasaki, K., Kuge, O., Chang, S.-C., Heacock, P. N., Rho, M., Suzuki, K., Nishijima, M., and Dowhan, W. (1999) *J. Biol. Chem.* **274**, 1828–1834
168a. Auer, M., Kim, M. J., Lemieux, M. J., Villa, A., Song, J., Li, X.-D., and Wang, D.-N. (2001) *Biochemistry* **40**, 6628–6635
169. Louie, K., and Dowhan, W. (1980) *J. Biol. Chem.* **255**, 1124–1127
169a. Stone, S. J., and Vance, J. E. (2000) *J. Biol. Chem.* **275**, 34534–34540
170. Trotter, P. J., and Voelker, D. R. (1995) *J. Biol. Chem.* **270**, 6062–6070
171. Shiao, Y.-J., Lupo, G., and Vance, J. E. (1995) *J. Biol. Chem.* **270**, 11190–11198
172. Heikinheimo, L., and Somerharju, P. (1998) *J. Biol. Chem.* **273**, 3327–3335
173. Huijbregts, R. P. H., de Kroon, A. I. P. M., and de Kruijff, B. (1998) *J. Biol. Chem.* **273**, 18936–18942
174. Rietveld, A. G., Chupin, V. V., Koorengevel, M. C., Wienk, H. L. J., Dowhan, W., and de Kruijff, B. (1994) *J. Biol. Chem.* **269**, 28670–28675
175. Bogdanov, M., Umeda, M., and Dowhan, W. (1999) *J. Biol. Chem.* **274**, 12339–12345
176. Ohtsuka, T., Nishijima, M., and Akamatsu, Y. (1993) *J. Biol. Chem.* **268**, 22908–22911
177. Zhao, M., Schlame, M., Rua, D., and Greenberg, M. L. (1998) *J. Biol. Chem.* **273**, 2402–2408
178. Chang, S.-C., Heacock, P. N., Mileykovskaya, E., Voelker, D. R., and Dowhan, W. (1998) *J. Biol. Chem.* **273**, 14933–14941
179. McMaster, C. R., and Bell, R. M. (1994) *J. Biol. Chem.* **269**, 14776–14783
180. Kent, C., and Carman, G. M. (1999) *Trends Biochem. Sci.* **24**, 146–150
180a. Dowd, S. R., Bier, M. E., and Patton-Vogt, J. L. (2001) *J. Biol. Chem.* **276**, 3756–3763
180b. Lykidis, A., Wang, J., Karim, M. A., and Jackowski, S. (2001) *J. Biol. Chem.* **276**, 2174–2179
181. Vance, D. E., and Pelech, S. L. (1984) *Trends Biochem. Sci.* **9**, 17–20
181a. Henneberry, A. L., Wistow, G., and McMaster, C. R. (2000) *J. Biol. Chem.* **275**, 29808–29815
181b. Cornell, R. B., and Northwood, I. C. (2000) *Trends Biochem. Sci.* **25**, 441–447
181c. Hunt, A. N., Clark, G. T., Attard, G. S., and Postle, A. D. (2001) *J. Biol. Chem.* **276**, 8492–8499
182. Kuge, O., Saito, K., and Nishijima, M. (1997) *J. Biol. Chem.* **272**, 19133–19139
183. Vance, J. E. (1998) *Trends Biochem. Sci.* **23**, 423–428
184. Zeisel, S. H., da Costa, K.-A., Franklin, P. D., Alexander, E. A., LaMont, J. T., Sheard, N. F., and Beiser, A. (1991) *FASEB J.* **5**, 2093–2098
185. Lee, H.-C., Fellenz-Maloney, M.-P., Liscovitch, M., and Krzysztof Blusztajn, J. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10086–10090
186. Nikawa, J.-i, Kodaki, T., and Yamashita, S. (1987) *J. Biol. Chem.* **2 262**, 4876–4881
186a. Ilgoutz, S. C., Zawadzki, J. L., Ralton, J. E., and McConville, M. J. (1999) *EMBO J.* **18**, 2746–2755
187. Wu, W.-I., and Carman, G. M. (1996) *Biochemistry* **35**, 3790–3796
188. Griac, P., Swede, M. J., and Henry, S. A. (1996) *J. Biol. Chem.* **271**, 25692–25698
189. Kim, K.-H., Voelker, D. R., Flocco, M. T., and Carman, G. M. (1998) *J. Biol. Chem.* **273**, 6844–6852
190. Ostrander, D. B., O'Brien, D. J., Gorman, J. A., and Carman, G. M. (1998) *J. Biol. Chem.* **273**, 18992–19001
191. Wieprecht, M., Wieder, T., Paul, C., Geilen, C. C., and Orfanos, C. E. (1996) *J. Biol. Chem.* **271**, 9955–9961
192. Snyder, F., ed. (1972) *Ether Lipids: Chemistry and Biology*, Academic Press, New York
193. Lee, T.-c, Malone, B., and Snyder, F. (1988) *J. Biol. Chem.* **263**, 1755–1760
194. Karasawa, K., Qiu, X., and Lee, T.-c. (1999) *J. Biol. Chem.* **274**, 8655–8661
194a. Min, J.-H., Wilder, C., Aoki, J., Arai, H., Inoue, K., Paul, L., and Gelb, M. H. (2001) *Biochemistry* **40**, 4539–4549
195. Ford, D. A., Rosenbloom, K. B., and Gross, R. W. (1992) *JBC* **267**, 11222–11228
196. Langworthy, T. A. (1985) in *The Bacteria* (Woese, C. R., and Wolfe, R. S., eds), pp. 459–497, Academic Press, New York
197. Jones, W. J., Nagle, D. P., Jr., and Whitman, W. B. (1989) *Microbiol. Rev.* **51**, 135–177
198. DeRosa, M. A., Gambacorta, A., and Gliozzi, A. (1986) *Microbiol. Rev.* **50**, 70–80
199. Burdett, K., Larkins, L. K., Das, A. K., and Hajra, A. K. (1991) *J. Biol. Chem.* **266**, 12201–12206
200. Friedberg, S. J., Weintraub, S. T., Singer, M. R., and Greene, R. C. (1982) *J. Biol. Chem.* **258**, 136–142
201. Brown, A. J., and Snyder, F. (1983) *J. Biol. Chem.* **258**, 4184–4189
202. Paltauf, F., and Holasek, A. (1973) *J. Biol. Chem.* **248**, 1609–1615
203. Borst, P. (1983) *Trends Biochem. Sci.* **8**, 269–272
204. de Vet, E. C. J. M., Ijlst, L., Oostheim, W., Wanders, R. J. A., and van den Bosch, H. (1998) *J. Biol. Chem.* **273**, 10296–10301
205. Moser, A. E., Singh, I., Brown, F. R., Solish, G. I., Kelley, R. I., Benke, P. J., and Moser, H. W. (1984) *N. Engl. J. Med.* **310**, 1141–1146
206. Lazarow, P. B., and Moser, H. W. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2287–2324, McGraw-Hill, New York
207. Chao, W., Liu, H., Hanahan, D. J., and Olson, M. S. (1992) *J. Biol. Chem.* **267**, 6725–6735
208. Lynch, D. V. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 285–308, CRC Press, Boca Raton, Florida

References

209. Barenholz, Y., and Gatt, S. (1982) in *New Comprehensive Biochemistry*, Vol. 4 (Hawthorne, J. N., and Ansell, G. B., eds), pp. 129–177, Elsevier, Amsterdam
209a. Ikushiro, H., Hayashi, H., and Kagamiyama, H. (2001) *J. Biol. Chem.* **276**, 18249–18256
210. Stoffel, W., and Melzner, I. (1980) *Z. Physiol. Chem.* **361**, 755–771
211. Grilley, M. M., Stock, S. D., Dickson, R. C., Lester, R. L., and Takemoto, J. Y. (1998) *J. Biol. Chem.* **273**, 11062–11068
212. Sperling, P., Zähringer, U., and Heinz, E. (1998) *J. Biol. Chem.* **273**, 28590–28596
213. Toledo, M. S., Levery, S. B., Straus, A. H., Suzuki, E., Momany, M., Glushka, J., Moulton, J. M., and Takahashi, H. K. (1999) *Biochemistry* **38**, 7294–7306
213a. Savile, C. K., Fabriàs, G., and Buist, P. H. (2001) *J. Am. Chem. Soc.* **123**, 4382–4385
214. Yamakawa, T. (1988) *Trends Biochem. Sci.* **13**, 452–454
215. van Echten, G., and Sandhoff, K. (1993) *J. Biol. Chem.* **268**, 5341–5344
216. Wiegandt, H. (1971) *Adv. Lipid Res.* **9**, 249–289
217. Hakomori, S.-i. (1990) *J. Biol. Chem.* **265**, 18713–18716
218. Takamiya, K., Yamamoto, A., Furukawa, K., Zhao, J., Fukumoto, S., Yamashiro, S., Okada, M., Haraguchi, M., Shin, M., Kishikawa, M., Shiku, H., Aizawa, S., and Furukawa, K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12147–12152
218a. Mylvaganam, M., and Lingwood, C. A. (1999) *J. Biol. Chem.* **274**, 20725–20732
218b. Uragami, M., Tokutake, N., Yan, X., and Regen, S. L. (2001) *J. Am. Chem. Soc.* **123**, 5124–5125
218c. Brown, D. A., and London, E. (2000) *J. Biol. Chem.* **275**, 17221–17224
219. Rameh, L. E., and Cantley, L. C. (1999) *J. Biol. Chem.* **274**, 8347–8350
220. Anderson, R. A., Boronenkov, I. V., Doughman, S. D., Kunz, J., and Loijens, J. C. (1999) *J. Biol. Chem.* **274**, 9907–9910
221. Rhee, S. G., and Bae, Y. S. (1997) *J. Biol. Chem.* **272**, 15045–15048
222. Heinz, D. W., Ryan, M., Bullock, T. L., and Griffith, O. H. (1995) *EMBO J.* **14**, 3855–3863
223. Essen, L.-O., Perisic, O., Katan, M., Wu, Y., Roberts, M. F., and Williams, R. L. (1997) *Biochemistry* **36**, 1704–1718
224. Hondal, R. J., Zhao, Z., Kravchuk, A. V., Liao, H., Riddle, S. R., Yue, X., Bruzik, K. S., and Tsai, M.-D. (1998) *Biochemistry* **37**, 4568–4580
225. Majerus, P. W., Kisseleva, M. V., and Norris, F. A. (1999) *J. Biol. Chem.* **274**, 10669–10672
226. Newton, A. C. (1995) *J. Biol. Chem.* **270**, 28495–28498
227. Newton, A. C., and Johnson, J. E. (1998) *Biochim. Biophys. Acta.* **1376**, 155–172
228. Epand, R. M., Stevenson, C., Bruins, R., Schram, V., and Glaser, M. (1998) *Biochemistry* **37**, 12068–12073
228a. McIntyre, T. M., Zimmerman, G. A., and Prescott, S. M. (1999) *J. Biol. Chem.* **274**, 25189–25192
229. Hodgkin, M. N., Pettitt, T. R., Martin, A., Michell, R. H., Pemberton, A. J., and Wakelam, M. J. O. (1998) *Trends Biochem. Sci.* **23**, 200–204
230. Topham, M. K., and Prescott, S. M. (1999) *J. Biol. Chem.* **274**, 11447–11450
230a. Jones, D. R., Pettitt, T. R., Sanjuán, M. A., Mérida, I., and Wakelam, M. J. O. (1999) *J. Biol. Chem.* **274**, 16846–16852
231. Walsh, J. P., Suen, R., and Glomset, J. A. (1995) *J. Biol. Chem.* **270**, 28647–28653
232. Schissel, S. L., Keesler, G. A., Schuchman, E. H., Williams, K. J., and Tabas, I. (1998) *J. Biol. Chem.* **273**, 18250–18259
233. Testi, R. (1996) *Trends Biochem. Sci.* **21**, 468–471
234. Spiegel, S., and Merrill, A. H., Jr. (1996) *FASEB J.* **10**, 1388–1397
235. Hannun, Y. A., and Obeid, L. M. (1995) *Trends Biochem. Sci.* **20**, 73–77
236. Usta, J., El Bawab, S., Roddy, P., Szulc, Z. M., Hannun, Y. A., and Bielawska, A. (2001) *Biochemistry* **40**, 9657–9668
236a. Hannun, Y. A., Luberto, C., and Argraves, K. M. (2001) *Biochemistry* **40**, 4893–4903
237. Mandala, S. M., Thornton, R., Tu, Z., Kurtz, M. B., Nickels, J., Broach, J., Menzeleev, R., and Spiegel, S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 150–155
237a. Liu, Y.-Y., Han, T.-Y., Giuliano, A. E., and Cabot, M. C. (2001) *FASEB J.* **15**, 719–730
237b. Windh, R. T., Lee, M.-J., Hla, T., An, S., Barr, A. J., and Manning, D. R. (1999) *J. Biol. Chem.* **274**, 27351–27358
237c. Ammit, A. J., Hastie, A. T., Edsall, L. C., Hoffman, R. K., Amrani, Y., Krymskaya, V. P., Kane, S. A., Peters, S. P., Penn, R. B., Spiegel, S., and Panettieri, R. A., Jr. (2001) *FASEB J.* **15**, 1212–1214
237d. Dickson, R. C. (1998) *Ann. Rev. Biochem.* **67**, 27–48
237e. Hla, T., Lee, M.-J., Ancellin, N., Paik, J. H., and Kluk, M. J. (2001) *Science* **294**, 1875–1878
238. Westerman, J., de Vries, K.-J., Somerharju, P., Timmermans-Hereijgers, J. L. P. M., Snoek, G. T., and Wirtz, K. W. A. (1995) *J. Biol. Chem.* **270**, 14263–14266
238a. Yoder, M. D., Thomas, L. M., Tremblay, J. M., Oliver, R. L., Yarbrough, L. R., and Helmkamp, G. M., Jr. (2001) *J. Biol. Chem.* **276**, 9246–9252
239. Gutteridge, J. M. C., and Halliwell, B. (1990) *Trends Biochem. Sci.* **15**, 129–135
240. Porter, N. A., and Wujek, D. G. (1984) *J. Am. Chem. Soc.* **106**, 2626–2629
241. Wagner, B. A., Buettner, G. R., and Burns, C. P. (1994) *Biochemistry* **33**, 4449–4453
242. Al-Abed, Y., Liebich, H., Voelter, W., and Bucala, R. (1996) *J. Biol. Chem.* **271**, 2892–2896
243. Uchida, K., Shiraishi, M., Naito, Y., Torii, Y., Nakamura, Y., and Osawa, T. (1999) *J. Biol. Chem.* **274**, 2234–2242
243a. Burczynski, M. E., Sridhar, G. R., Palackal, N. T., and Penning, T. M. (2001) *J. Biol. Chem.* **276**, 2890–2897
244. Gardner, H. W., and Hamberg, M. (1993) *J. Biol. Chem.* **268**, 6971–6977
244a. Schneider, C., Tallman, K. A., Porter, N. A., and Brash, A. R. (2001) *J. Biol. Chem.* **276**, 20831–20838
245. Kato, Y., Mori, Y., Makino, Y., Morimitsu, Y., Hiroi, S., Ishikawa, T., and Osawa, T. (1999) *J. Biol. Chem.* **274**, 20406–20414
246. Tsai, L., Szweda, P. A., Vinogradova, O., and Szweda, L. I. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7975–7980
247. McGirr, L. G., Hadley, M., and Draper, H. H. (1985) *J. Biol. Chem.* **260**, 15427–15431
248. Hadley, M., and Draper, H. H. (1988) *FASEB J.* **2**, 138–140
249. Slatter, D. A., Paul, R. G., Murray, M., and Bailey, A. J. (1999) *J. Biol. Chem.* **274**, 19661–19669
250. Armstrong, D., Dimmitt, S., Boehme, D. H., Leonberg, S. C., Jr., and Vogel, W. (1974) *Science* **186**, 155–156
250a. Onorato, J. M., Jenkins, A. J., Thorpe, S. R., and Baynes, J. W. (2000) *J. Biol. Chem.* **275**, 21177–21184
251. Hill, E. G., Johnson, S. B., Lawson, L. D., Mahfouz, M. M., and Holman, R. T. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 953–957
252. Holman, R. T., Pusch, F., Svingen, B., and Dutton, H. J. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4830–4834
253. Mensink, R. P., and Katan, M. B. (1990) *N. Engl. J. Med.* **323**, 439–445
253a. Kopelman, P. G. (2000) *Nature (London)* **404**, 635–643
253b. Unger, R. H., and Orci, L. (2001) *FASEB J.* **15**, 312–321
253c. Taubes, G. (2001) *Science* **291**, 2536–2545
253d. Koo, S.-H., Dutcher, A. K., and Towle, H. C. (2001) *J. Biol. Chem.* **276**, 9437–9445
253e. Weigle, D. S. (1994) *FASEB J.* **8**, 302–310
253f. Barsh, G. S., Farooqi, I. S., and O'Rahilly, S. (2000) *Nature (London)* **404**, 644–651
253g. Harris, R. B. S. (1990) *FASEB J.* **4**, 3310–3318
253h. Ruderman, N., and Flier, J. S. (2001) *Science* **291**, 2558–2559
253i. Lee, J.-J., Moon, Y.-A., Ha, J.-H., Yoon, D.-J., Ahn, Y.-H., and Kim, K.-S. (2001) *J. Biol. Chem.* **276**, 2576–2585
253j. Gura, T. (1998) *Science* **280**, 1369–1370
253k. Jucker, B. M., Ren, J., Dufour, S., Cao, X., Previs, S. F., Cadman, K. S., and Shulman, G. I. (2000) *J. Biol. Chem.* **275**, 39279–39286
253l. Clément, K., Vaisse, C., Lahlou, N., Cabrol, S., Pelloux, V., Cassuto, D., Gourmelen, M., Dina, C., Chambaz, J., Lacorte, J.-M., Basdevant, A., Bougnères, P., Lebouc, Y., Froguel, P., and Guy-Grand, B. (1998) *Nature (London)* **392**, 398–401
253m. Gura, T. (2000) *Science* **287**, 1738–1741
253n. Lee, Y., Wang, M.-Y., Kakuma, T., Wang, Z.-W., Babcock, E., McCorkle, K., Higa, M., Zhou, Y.-T., and Unger, R. H. (2001) *J. Biol. Chem.* **276**, 5629–5635
254. Makita, K., Falck, J. R., and Capdevila, J. H. (1996) *FASEB J.* **10**, 1456–1463
255. Schewe, T., and Kühn, H. (1991) *Trends Biochem. Sci.* **16**, 369–73
256. Capdevila, J. H., Falck, J. R., and Estabrook, R. W. (1992) *FASEB J.* **6**, 731–736
256a. Funk, C. D. (2001) *Science* **294**, 1871–1875
257. Bergström, S. (1967) *Science* **157**, 382–391
258. Samuelsson, B., Granström, E., Green, K., Hamberg, M., and Hammarström, S. (1975) *Ann. Rev. Biochem.* **44**, 669–695
259. Newton, R. F., and Roberts, S. M., eds. (1982) *Prostaglandins and Thromboxanes*, Butterworth Scientific, London
260. Neufeld, E. J., Bross, T. E., and Majerus, P. W. (1984) *J. Biol. Chem.* **259**, 1986–1992
261. Kramer, R. M., and Deykin, D. (1983) *J. Biol. Chem.* **258**, 13806–13811
262. Picot, D., Loll, P. J., and Garavito, R. M. (1994) *Nature (London)* **367**, 243–249
262a. Seibold, S. A., Cerda, J. F., Mulichak, A. M., Song, I., Garavito, R. M., Arakawa, T., Smith, W. L., and Babcock, G. T. (2000) *Biochemistry* **39**, 6616–6624
262b. Klefer, J. R., Pawlitz, J. L., Moreland, K. T., Stegeman, R. A., Hood, W. F., Glerse, J. K., Stevens, A. M., Goodwin, D. C., Rowlinson, S. W., Marnett, L. J., Stallings, W. C., and Kurumball, R. G. (2000) *Nature (London)* **405**, 97–101
262c. Thuresson, E. D., Malkowski, M. G., Lakkides, K. M., Rieke, C. J., Mulichak, A. M., Ginell, S. L., Garavito, R. M., and Smith, W. L. (2001) *J. Biol. Chem.* **276**, 10358–10365
262d. Peng. S., Okeley, N. M., Tsai, A.-L., Wu, G., Kulmacz, R. J., and van der Donk. W. A. (2002) *J. Am. Chem. Soc.* **124**, 10785–10796
263. Lambeir, A.-M., Markey, C. M., Dunford, H. B., and Marnett, L. J. (1985) *J. Biol. Chem.* **260**, 14894–14896
264. Tsai, A.-l, Kulmacz, R. J., and Palmer, G. (1995) *J. Biol. Chem.* **270**, 10503–10508

References

265. Smith, W. L., Garavito, R. M., and DeWitt, D. L. (1996) *J. Biol. Chem.* **271**, 33157–33160
266. Tsai, A.-l, Palmer, G., Xiao, G., Swinney, D. C., and Kulmacz, R. J. (1998) *J. Biol. Chem.* **273**, 3888–3894
266a. Marnett, L. J., Rowlinson, S. W., Goodwin, D. C., Kalgutkar, A. S., and Lanzo, C. A. (1999) *J. Biol. Chem.* **274**, 22903–22906
266b. Malkowski, M. G., Ginell, S. L., Smith, W. L., and Garavito, R. M. (2000) *Science* **15**, 1933–1937
267. Dean, A. M., and Dean, F. M. (1999) *Protein Sci.* **8**, 1087–1098
268. Jakobsson, P.-J., Thorén, S., Morgenstern, R., and Samuelsson, B. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7220–7225
269. Williams, D. W., Hale, S. E., Okita, R. T., and Masters, B. S. S. (1984) *J. Biol. Chem.* **259**, 14600–14608
270. Hamberg, M., Svensson, J., and Samuelsson, B. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2994–2998
271. Hecker, M., and Ullrich, V. (1989) *J. Biol. Chem.* **264**, 141–150
272. Ruan, K.-H., Li, D., Ji, J., Lin, Y.-Z., and Gao, X. (1998) *Biochemistry* **37**, 822–830
273. Haurand, M., and Ullrich, V. (1985) *J. Biol. Chem.* **260**, 15059–15067
274. Moncada, S., Gryglewski, R., Bunting, S., and Vane, J. R. (1976) *Nature (London)* **263**, 663–665
275. Vane, J. R., and Bergström, S., eds. (1979) *Prostacyclin*, Raven Press, New York
275a. Cheng, Y., Austin, S. C., Rocca, B., Koller, B. H., Coffman, T. M., Grosser, T., Lawson, J. A., and FitzGerald, G. A. (2002) *Science* **296**, 539–541
276. Shyue, S.-K., Ruan, K.-H., Wang, L.-H., and Wu, K. K. (1997) *J. Biol. Chem.* **272**, 3657–3662
277. Wong, P. Y.-K., Malik, K. U., Taylor, B. M., Schneider, W. P., McGiff, J. C., and Sun, F. F. (1985) *J. Biol. Chem.* **260**, 9150–9153
278. Kramer, J. A., Johnson, K. R., Dunham, W. R., Sands, R. H., and Funk, M. O., Jr. (1994) *Biochemistry* **33**, 15017–15022
279. Glickman, M. H., and Klinman, J. P. (1996) *Biochemistry* **35**, 12882–12892
280. Nelson, M. J., Chase, D. B., and Seitz, S. P. (1995) *Biochemistry* **34**, 6159–6163
280a. Brash, A. R. (1999) *J. Biol. Chem.* **274**, 23679–23682
280b. Skrzypczak-Jankun, E., Bross, R. A., Carroll, R. T., Dunham, W. R., and Funk, M. O., Jr. (2001) *J. Am. Chem. Soc.* **123**, 10814–10820
280c. Knapp, M. J., Seebeck, F. P., and Klinman, J. P. (2001) *J. Am. Chem. Soc.* **123**, 2931–2932
280d. Mogul, R., and Holman, T. R. (2001) *Biochemistry* **40**, 4391–4397
280e. Tomchick, D. R., Phan, P., Cymborowski, M., Minor, W., and Holman, T. R. (2001) *Biochemistry* **40**, 7509–7517
281. Funk, C. D., Keeney, D. S., Oliw, E. H., Boeglin, W. E., and Brash, A. R. (1996) *J. Biol. Chem.* **271**, 23338–23344
282. Hamberg, M., Su, C., and Oliw, E. (1998) *J. Biol. Chem.* **273**, 13080–13088
283. Boyington, J. C., Gaffney, B. J., and Amzel, L. M. (1993) *Science* **260**, 1482–1486
284. Moiseyev, N., Rucker, J., and Glickman, M. H. (1997) *J. Am. Chem. Soc.* **119**, 3853–3860
285. Ford-Hutchinson, A. W., Gresser, M., and Young, R. N. (1994) *Ann. Rev. Biochem.* **63**, 383–417
286. Glasgow, W. C., Harris, T. M., and Brash, A. R. (1986) *J. Biol. Chem.* **261**, 200–204
287. Hammarström, S. (1983) *Ann. Rev. Biochem.* **52**, 355–377
288. Samuelsson, B., Dahlén, S.-E., Lindgren, J. Å., Rouzer, C. A., and Serhan, C. N. (1987) *Science* **237**, 1171–1176
289. Samuelsson, B., and Funk, C. D. (1989) *J. Biol. Chem.* **264**, 19469–19472
290. Blomster Andberg, M., Hamberg, M., and Haeggström, J. Z. (1997) *J. Biol. Chem.* **272**, 23057–23063
291. Carter, B. Z., Wiseman, A. L., Orkiszewski, R., Ballard, K. D., Ou, C.-N., and Lieberman, M. W. (1997) *J. Biol. Chem.* **272**, 12305–12310
292. Lewis, R. A., Austen, K. F., and Soberman, R. J. (1990) *N. Engl. J. Med.* **323**, 645–655
293. Lynch, K. R., O'Neill, G. P., Liu, Q., Im, D.-S., Sawyer, N., Metters, K. M., Coulombe, N., Abramovitz, M., Figueroa, D. J., Zeng, Z., Connolly, B. M., Bai, C., Austin, C. P., Chateauneuf, A., Stocco, R., Greig, G. M., Kargman, S., Hooks, S. B., Hosfield, E., Williams, D. L., Jr., Ford-Hutchinson, A. W., Caskey, C. T., and Evans, J. F. (1999) *Nature (London)* **399**, 789–793
294. Hammarström, S. (1983) *J. Biol. Chem.* **258**, 1427–1430
295. Serhan, C. N., Hamberg, M., and Samuelsson, B. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5335–5339
296. Rowley, A. F., Lloyd-Evans, P., Barrow, S. E., and Serhan, C. N. (1994) *Biochemistry* **33**, 856–863
297. Chang, M. S., Boeglin, W. E., Guengerich, F. P., and Brash, A. R. (1996) *Biochemistry* **35**, 464–471
298. Brash, A. R., Porter, A. T., and Maas, R. L. (1985) *J. Biol. Chem.* **260**, 4210–4216
298a. Zeldin, D. C. (2001) *J. Biol. Chem.* **276**, 36059–36062
298b. Node, K., Huo, Y., Ruan, X., Yang, B., Spiecker, M., Ley, K., Zeldin, D. C., and Liao, J. K. (1999) *Science* **285**, 1276–1279
299. Goetzl, E. J., An, S., and Smith, W. L. (1995) *FASEB J.* **9**, 1051–1058
300. Urade, Y., Fujimoto, N., and Hayaishi, O. (1985) *J. Biol. Chem.* **260**, 12410–12415
301. Brock, T. G., McNish, R. W., and Peters-Golden, M. (1999) *J. Biol. Chem.* **274**, 11660–11666
302. MacMillan, D. K., Hill, E., Sala.A, Sigal, E., Shuman, T., Henson, P. M., and Murphy, R. C. (1994) *J. Biol. Chem.* **269**, 26663–26668
303. Maddox, J. F., Colgan, S. P., Clish, C. B., Petasis, N. A., Fokin, V. V., and Serhan, C. N. (1998) *FASEB J.* **12**, 487–494
304. Dahlén, S.-E., Hansson, G., Hedqvist, P., Björck, T., Granström, E., and Dahlén, B. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1712–1716
305. Lindgren, J. Å., Hökfelt, T., Dahlén, S.-E., Patrono, C., and Samuelsson, B. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6212–6216
306. Reginato, M. J., Krakow, S. L., Bailey, S. T., and Lazar, M. A. (1998) *J. Biol. Chem.* **273**, 1855–1858
307. Nelson, N. A., Kelly, R. C., and Johnson, R. A. (1982) *Chem. Eng. News* **Aug 16**, 30–44
308. Marx, J. L. (1979) *Science* **205**, 175–176
309. Wheelan, P., and Murphy, R. C. (1995) *J. Biol. Chem.* **270**, 19845–19852
310. Weissmann, G. (1991) *Sci. Am.* **264**(Jan), 84–90
311. Kalgutkar, A. S., Crews, B. C., Rowlinson, S. W., Garner, C., Seibert, K., and Marnett, L. J. (1998) *Science* **280**, 1268–1270
312. Vane, J. (1994) *Nature (London)* **367**, 215–217
313. Lecomte, M., Laneuville, O., Ji, C., DeWitt, D. L., and Smith, W. L. (1994) *J. Biol. Chem.* **269**, 13207–13215
314. Loll, P. J., Picot, D., Ekabo, O., and Garavito, R. M. (1996) *Biochemistry* **35**, 7330–7340
314a. Warner, T. D., and Mitchell, J. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13371-13373
315. Abramson, S., Korchak, H., Ludewig, R., Edelson, H., Haines, K., Levin, R. I., Herman, R., Rider, L., Kimmel, S., and Weissman, G. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7227–7231
316. Patrono, C. (1994) *N. Engl. J. Med.* **330**, 1287–1294
317. DuBois, R. N., Abramson, S. B., Crofford, L., Gupta, R. A., Simon, L. S., van de Putte, L. B. A., and Lipsky, P. E. (1998) *FASEB J.* **12**, 1063–1073
317a. Selinsky, B. S., Gupta, K., Sharkey, C. T., and Loll, P. J. (2001) *Biochemistry* **40**, 5172–5180
317b. FitzGerald, G. A., and Patrono, C. (2001) *N. Engl. J. Med.* **345**, 433–442
317c. Killen, J. P., Nzerue, C. M., Rich, S. A., FitzGerald, G. A., and Patrono, C. (2001) *N. Engl. J. Med.* **345**, 1708–1709
317d. Vane, J. R. (2002) *Science* **296**, 474–475
317e. Jiang, Q., Elson-Schwab, I., Courtemanche, C., and Ames, B. N. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11494–11499
318. Vick, B. A. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 167–194, CRC Press, Boca Raton, Florida
319. Schaller, F., and Weiler, E. W. (1997) *J. Biol. Chem.* **272**, 28066–28072
320. Vijayan, P., Shockey, J., Lévesque, C. A., Cook, R. J., and Browse, J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7209–7214
321. Song, W.-C., Baertschi, S. W., Boeglin, W. E., Harris, T. M., and Brash, A. R. (1993) *J. Biol. Chem.* **268**, 6293–6298
321a. Ziegler, J., Stenzel, I., Hause, B., Maucher, H., Hamberg, M., Grimm, R., Ganal, M., and Wasternack, C. (2000) *J. Biol. Chem.* **275**, 19132–19138
322. Conconi, A., Smerdon, M. J., Howe, G. A., and Ryan, C. A. (1996) *Nature (London)* **383**, 826–829
323. Xie, D.-X., Feys, B. F., James, S., Nieto-Rostro, M., and Turner, J. G. (1998) *Science* **280**, 1091–1094
324. Alborn, H. T., Turlings, T. C. J., Jones, T. H., Stenhagen, G., Loughrin, J. H., and Tumlinson, J. H. (1997) *Science* **276**, 945–949
325. Paré, P. W., Alborn, H. T., and Tumlinson, J. H. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13971–13975
326. Parchmann, S., and Mueller, M. J. (1998) *J. Biol. Chem.* **273**, 32650–32655
327. Koljak, R., Boutaud, O., Shieh, B.-H., Samel, N., and Brash, A. R. (1997) *Science* **277**, 1994–1996
328. Brash, A. R., Hughes, M. A., Hawkins, D. J., Boeglin, W. E., Song, W.-C., and Meijer, L. (1991) *J. Biol. Chem.* **266**, 22926–22931
329. Birch, A. J. (1967) *Science* **156**, 202–206
330. Spencer, J. B., and Jordan, P. M. (1992) *Biochemistry* **31**, 9107–9116
331. Fu, H., Alvarez, M. A., Khosla, C., and Bailey, J. E. (1996) *Biochemistry* **35**, 6527–6532
332. Kendrew, S. G., Harding, S. E., Hopwood, D. A., and Marsh, E. N. G. (1995) *J. Biol. Chem.* **270**, 17339–17343
333. Quirós, L. M., and Salas, J. A. (1995) *J. Biol. Chem.* **270**, 18234–18239
334. Schultz, D. J., Cahoon, E. B., Shanklin, J., Craig, R., Cox-Foster, D. L., Mumma, R. O., and Medford, J. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8771–8775
335. Michal, G., ed. (1999) *Biochemical Pathways*, Wiley - Spektrum Academischer Verlag, New York - Heidelberg, Germany
336. Hutchinson, C. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3336–3338
336a. Bao, W., Sheldon, P. J., and Hutchinson, C. R. (1999) *Biochemistry* **38**, 9752–9757

References

336b. Jez, J. M., Ferrer, J.-L., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000) *Biochemistry* **39**, 890–902
336c. Khosla, C., Gokhale, R. S., Jacobsen, J. R., and Cane, D. E. (1999) *Ann. Rev. Biochem.* **68**, 219–253
337. Cane, D. E. (1994) *Science* **263**, 338–340
338. Pieper, R., Gokhale, R. S., Luo, G., Cane, D. E., and Khosla, C. (1997) *Biochemistry* **36**, 1846–1851
339. Kao, C. M., Pieper, R., Cane, D. E., and Khosla, C. (1996) *Biochemistry* **35**, 12363–12368
340. Lambalot, R. H., Cane, D. E., Aparicio, J. J., and Katz, L. (1995) *Biochemistry* **34**, 1858–1866
341. Kendrew, S. G., Katayama, K., Deutsch, E., Madduri, K., and Hutchinson, C. R. (1999) *Biochemistry* **38**, 4794–4799
342. Kim, C.-G., Yu, T.-W., Fryhle, C. B., Handa, S., and Floss, H. G. (1998) *J. Biol. Chem.* **273**, 6030–6040
343. Aparicio, J. F., Colina, A. J., Ceballos, E., and Martín, J. F. (1999) *J. Biol. Chem.* **274**, 10133–10139
344. Kennedy, J., Auclair, K., Kendrew, S. G., Park, C., Vederas, J. C., and Hutchinson, C. R. (1999) *Science* **284**, 1368–1372
345. Offenzeller, M., Santer, G., Totschnig, K., Su, Z., Moser, H., Traber, R., and Schneider-Scherzer, E. (1996) *Biochemistry* **35**, 8401–8412
346. Schwecke, T., Aparicio, J. F., Molnár, I., König, A., Ee Khaw, L., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J. B., Böhm, G. A., Staunton, J., and Leadlay, P. F. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7839–7843
347. Preisig-Müller, R., Gehlert, R., Melchior, F., Stietz, U., and Kindl, H. (1997) *Biochemistry* **36**, 8349–8358
348. Schröder, J., Raiber, S., Berger, T., Schmidt, A., Schmidt, J., Soares-Sello, A. M., Bardshiri, E., Strack, D., Simpson, T. J., Veit, M., and Schröder, G. (1998) *Biochemistry* **37**, 8417–8425
348a. Jez, J. M., Bowman, M. E., and Noel, J. P. (2001) *Biochemistry* **40**, 14829–14838
349. Guo, L., Dixon, R. A., and Paiva, N. L. (1994) *J. Biol. Chem.* **269**, 22372–22378
350. Brown, D. W., Adams, T. H., and Keller, N. P. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14873–14877
350a. Schroeder, F. C., Farmer, J. J., Smedley, S. R., Attygalle, A. B., Eisner, T., and Meinwald, J. (2000) *J. Am. Chem. Soc.* **122**, 3628–3634
351. Marsden, A. F. A., Wilkinson, B., Cortés, J., Dunster, N. J., Staunton, J., and Leadlay, P. F. (1998) *Science* **279**, 199–202
352. Shen, Y., Yoon, P., Yu, T.-W., Floss, H. G., Hopwood, D., and Moore, B. S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3622–3627
353. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999) *Science* **284**, 482–485
354. Reynolds, K. A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12744–12746
355. Schröder, F. C., Farmer, J. J., Attygalle, A. B., Smedley, S. R., Eisner, T., and Meinwald, J. (1998) *Science* **281**, 428–431
356. Metz, J. G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., Valentine, R., Lardizabal, K., Domergue, F., Yamada, A., Yazawa, K., Knauf, V., and Browse, J. (2001) *Science* **293**, 290–293
357. Nugent, C., Prins, J. B., Whitehead, J. P., Wentworth, J. M., Chatterjee, V. K. K., and O'Rahilly, S. (2001) *J. Biol. Chem.* **276**, 9149–9157
358. Gibbons, H. S., Lin, S., Cotter, R. J., and Raetz, C. R. H. (2000) *J. Biol. Chem.* **275**, 32940–32949
359. Sohlenkamp, C., de Rudder, K. E. E., Röhrs, V., López-Lara, I. M., and Geiger, O. (2000) *J. Biol. Chem.* **275**, 18919–18925
360. Moody, J. S., Kozak, K. R., Ji, C., and Marnett, L. J. (2001) *Biochemistry* **40**, 861–866
361. Nakajima, J.-i, Tanaka, Y., Yamazaki, M., and Saito, K. (2001) *J. Biol. Chem.* **276**, 25797–25803

Study Questions

1. Outline possible pathways of metabolism of dietary fats. Consider digestion, transport of fatty acids, storage, conversion to prostaglandins, steroids, etc. Will any of the fat be converted into glucose?

2. What are the functions in the human body of the following?

   Pancreatic lipase
   Lipocalins
   Lipoprotein lipase
   Very low density lipoprotein (VLDL)
   Hormone-sensitive lipase
   Chylomicrons
   Apolipoproteins

3. Describe two different types of fatty acid synthase. Compare the basic chemical reactions that are involved. Also, compare these with the reactions of fatty acid oxidation.

4. Discuss the different types of fatty acids found in the human body and the synthetic pathways by which they are formed.

5. What mechanisms are utilized for incorporation of double bonds into fatty acids? Propose a mechanism that makes use of polyketide synthase domains (Fig. 21-11) in the synthesis of polyunsaturated fatty acids. See Metz *et al.*[356]

6. In what locations would you expect to find the following?

   Tripalmitin
   Mycolic acids
   Arachidonic acid
   Propionic acid
   Docosahexaenoic acid

7. Substitution of a small percentage of ω6 fatty acids in the diet of insulin-resistant rodents with ω3 unsaturated fatty acids normalized insulin action.[357] Can you suggest possible mechanisms? Is this result significant to human nutrition?

8. Formation of the 3-hydroxymyristoyl groups of lipid A (Fig. 8-30) requires O_2. Comparisons of amino acid sequences suggest that an Fe^{2+}/2-oxoglutarate-dependent oxygenase is involved.[358] Write a balanced equation for this reaction.

9. Phosphatidylcholine can be formed by two pathways as described on pp. 1198–1199. A third pathway, used by some bacteria, involves a direct one-step reaction of choline with CDP-diacylglycerol.[359] Write a reasonable chemical mechanism.

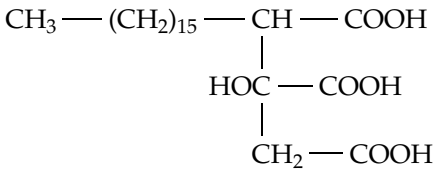
Study Questions

10. ^{14}C-Carboxyl labeled palmitic acid is fed to a fasted rat. There is no increase in liver glycogen, but the glucose units of the glycogen contain ^{14}C.
    a) Outline, using appropriate equations, the reaction sequence by which the carbon atoms of glucose become labeled.
    b) Explain why there is no net synthesis of glycogen from the fatty acid.

11. a) Write the reactions that most *dietary* tripalmitin will undergo in the body of an adult human in order to be deposited in adipose tissue as tripalmitin.
    b) What is the minimum amount of ATP (high energy bonds) normally required to deposit the one mole of dietary tripalmitin in adipose tissue? Count only ATP involved in tripalmitin metabolism and consider the source of glycerol in the adipose tissue.

12. Describe the biochemical effects on lipid metabolism of injecting into a normal animal
    a) insulin
    b) glucagon
    c) epinephrine

13. Suggest a biosynthetic pathway for formation of the fungal metabolite **agaricic acid**:



14. The ketone **palmiton** $CH_3(CH_2)14 - {}^*CO - (CH_2)_{14}CH_3$ is formed by mycobacteria. The carbon marked by an asterisk was found to be labeled after feeding of [1-^{14}C]palmitic acid. Suggest a biosynthetic pathway.

15. The following reaction occurs in the biosynthesis of fatty acids.

    Crotonyl-ACP + NADPH + H^+ $\rightarrow$ butyryl-ACP + $NADP^+$

    The reduction half-reactions for crotonyl-ACP and NADPH are

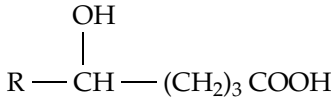
    Crotonyl-ACP + 2 H^+ + 2 e^- $\rightarrow$ butyryl-ACP
    $E^{o\prime} = -0.015$ V
    NADP + 2 H^+ + 2 e^- $\rightarrow$ NADPH + H
    $E^{o\prime} = -0.320$ V

    What is $\Delta G^{o\prime}$ for this reaction? What is the equilibrium constant for the reaction?

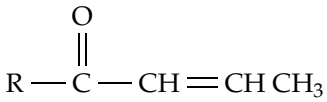
16. How does the inhibition of citrate synthase affect fatty acid synthesis?

17. Malonyl CoA is an allosteric effector of carnitine acyl transferase. What kind of effector is it, i.e., activator or inhibitor, and what is the logic behind the interaction?

18. Compound II is formed in a series of enzymatic reactions from compound I. Propose a mechanistically realistic sequence, showing by name any cofactors required.



I

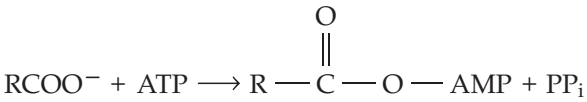


II

19. Fatty acid biosynthesis requires NADPH. Where does the NADPH come from?

20. An individual has been found who is missing malic enzyme in his cytoplasm. He has instead an enzyme that converts the oxalacetate made from the citrate lyase reaction directly to pyruvate and CO_2. Discuss this patient in terms of the likely effect of these changes on his ability to synthesize fatty acids.

21. The $\Delta G^{o\prime}$ values for the hydrolysis of any P – O – P bond of ATP, inorganic pyrophosphate, or any acyl CoA thiolester are all about –34 kJ/mole, while the corresponding figure for the hydrolysis of a mixed carboxylic phosphate anhydride is about –55 kJ/mole. Calculate the value of $\Delta G^{o\prime}$ for the following reaction describing the activation of fatty acids to the fatty acyl adenylate.



Study Questions

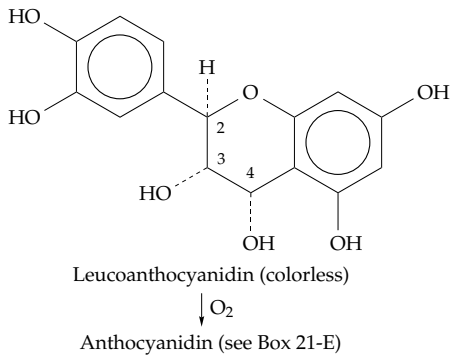
22. Fatty acid biosynthesis is made irreversible by the specific input of energy. Name the reactions or give equations for those steps in the pathway that require ATP. It is important that you consider both the mitochondrial and cytosolic components of the pathway.

23. The fatty acid biosynthesis pathway communicates with at least three other metabolic pathways either by sharing common intermediates or by regulatory mechanisms. Fill in the table below. List four molecules that have this function. You should name the additional pathway where each of these is found and briefly describe what it does in this second pathway. Do **not** use a redox cofactor as one of your choices.

Molecule	Other Pathway	Role in Second Pathway

24. The endogenous cannabinoid 2-arachidonoylglycerol is thought to play important roles both in the brain (Chapter 30) and in the immune system (Chapter 31). Leukocyte 12-oxygenase acts on this compound.[360] What products would be expected?

25. In the synthesis of anthocyanidins (Box 21-E) another Fe^{2+}/2-oxoglutarate-dependent oxygenase acts on the colorless leukoanthrocyanidin, which is then converted to the colored anthocyanidin:

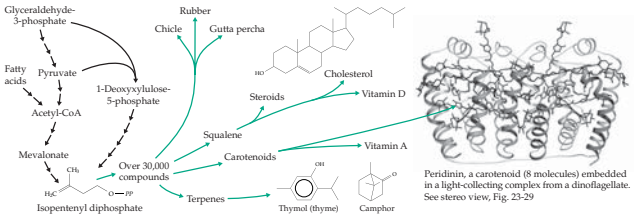


Propose a reasonable sequence for this reaction. See Nakajima *et al*.[361]

26. Why does a dietary deficiency of choline in humans induce a fatty liver, i.e., a liver in which the hepatocytes contain excess triglycerides?

27. Name two important functions of citrate in the conversion of acetate to palmitate by adipose tissues.

28. Outline the steps by which linoleic acid can be converted to prostaglandins in the human body.



Starting with the simple compounds acetyl-CoA, glyceraldehyde-3-phosphate, and pyruvate, which arise via the central pathways of metabolism, the key intermediate **isopentenyl diphosphate** is formed by two independent routes. It is then converted by bacteria, fungi, plants, and animals into thousands of different naturally occurring products. These include high polymers, such as rubber, as well as vitamins, sterols, carotenoids, and over 30,000 different terpenes and related compounds. Many of the latter are found only in specific plants where they may function as defensive compounds or pheromones.

Contents

1227 ..... **A. Isopentenyl Diphosphate and Polyprenyl Synthases**
1229 ..... 1. An Alternative Pathway for Isoprenoid Synthesis
1230 ..... 2. Isomerization and Isoprene Formation
1230 ..... 3. Polyprenyl Compounds
1231 ..... 4. Prenylation of Proteins and Other Compounds
1232 ..... **B. Terpenes**
1232 ..... 1. Biosynthesis of Monoterpenes
1234 ..... 2. Sesquiterpenes and Diterpenes
1236 ..... 3. Formation of the Symmetric Terpenes, Squalene and Phytoene
1237 ..... **C. Carotenes and Their Derivatives**
1240 ..... 1. Xanthophylls and Other Oxidized Carotenes
1243 ..... 2. Properties and Functions of Carotenes
1243 ..... **D. Steroid Compounds**
1244 ..... 1. Biosynthesis of Sterols
1245 ..... *Formation of cholesterol*
1247 ..... 2. Metabolism of Cholesterol in the Human Body
1248 ..... *Serum cholesterol*
1250 ..... *The LDL and related receptors*
1251 ..... *Abnormalities of cholesterol metabolism*
1251 ..... *Bile acids*
1252 ..... **E. The Steroid Hormones**
1253 ..... 1. Progestins
1253 ..... 2. Adrenal Cortical Steroids
1254 ..... 3. Androgens
1260 ..... 4. Estrogens
1263 ..... 5. The Steroid Receptor Family
1265 ..... **F. Other Steroids**

1266 ..... **References**
1271 ..... **Study Questions**

Boxes

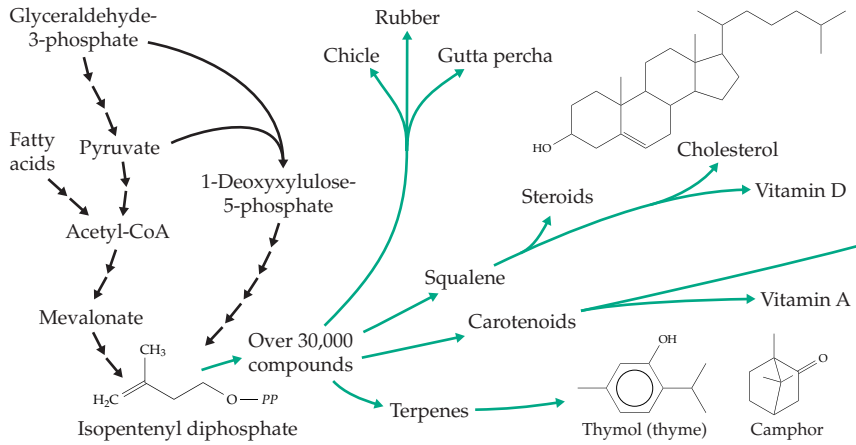
1241 ..... Box 22-A Vitamin A
1249 ..... Box 22-B Atherosclerosis
1257 ..... Box 22-C Vitamin D
1261 ..... Box 22-D The Renin–Angiotensin–Aldosterone System and the Regulation of Blood Pressure

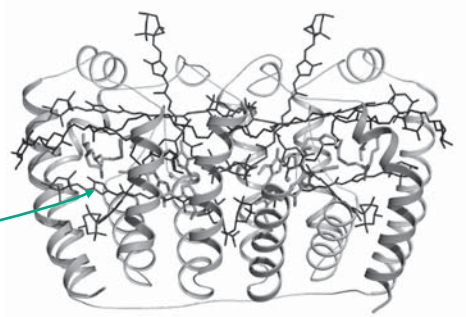
Tables

1264 ..... Table 22-1 Known Members of the Steroid Receptor Family

Polyprenyl (Isoprenoid) Compounds

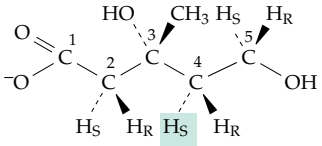
22



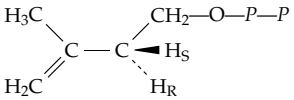


Peridinin, a carotenoid (8 molecules) embedded in a light-collecting complex from a dinoflagellate. See stereo view, Fig. 23-29

The **terpenes**, **carotenoids**, **steroids**, and many other compounds arise in a direct way from the prenyl group of **isopentenyl diphosphate** (Fig. 22-1).[1–6a] Biosynthesis of this five-carbon branched unit from **mevalonate** has been discussed previously (Chapter 17, Fig. 17-19) and is briefly recapitulated in Fig. 22-1. Distinct isoenzymes of 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) in the liver produce HMG-CoA destined for formation of ketone bodies (Eq. 17-5) or mevalonate.[7,8] A similar cytosolic enzyme is active in plants which, collectively, make more than 30,000 different isoprenoid compounds.[9,10] However, many of these are formed by an alternative pathway that does not utilize mevalonate but starts with a thiamin diphosphate-dependent condensation of glyceraldehyde 3-phosphate with pyruvate (Figs. 22-1, 22-2).



3(*R*) Mevalonate



Isopentenyl diphosphate

The two-step reduction of HMG-CoA to mevalonate (Fig. 22-1, step *a*)[11–15] is highly controlled, a major factor in regulating cholesterol synthesis in the human liver.[12,16,17] The N-terminal portion of the 97-kDa 888-residue mammalian HMG-CoA reductase is thought to be embedded in membranes of the ER, while the C-terminal portion is exposed in the cytoplasm.[16] The enzyme is sensitive to feedback inhibition by cholesterol (see Section D, 2). The regulatory mechanisms include a phosphorylation–dephosphorylation cycle and control of both the rates of synthesis and of proteolytic degradation of this key enzyme.[14,15,18–20]

A. Isopentenyl Diphosphate and Polyprenyl Synthases

In animals all isoprenoid compounds are apparently synthesized from mevalonate, which is converted by the consecutive action of two kinases[21–23] into mevalonate 5-diphosphate (Fig. 22-1, step *b*). Mevalonate kinase is found predominantly in peroxisomes, which are also active in other aspects of steroid synthesis in humans.[21,24] A deficiency of this enzyme is associated with mevalonic aciduria, a serious hereditary disease in which both blood and urine contain very high concentrations of mevalonate.[23] Mevalonate diphosphate kinase, which is also a decarboxylase, catalyzes phosphorylation of the 3-OH group of mevalonate (step *c*, Fig. 22-1) and decarboxylative elimination of phosphate (step *d*)[25] to form isopentenyl diphosphate.

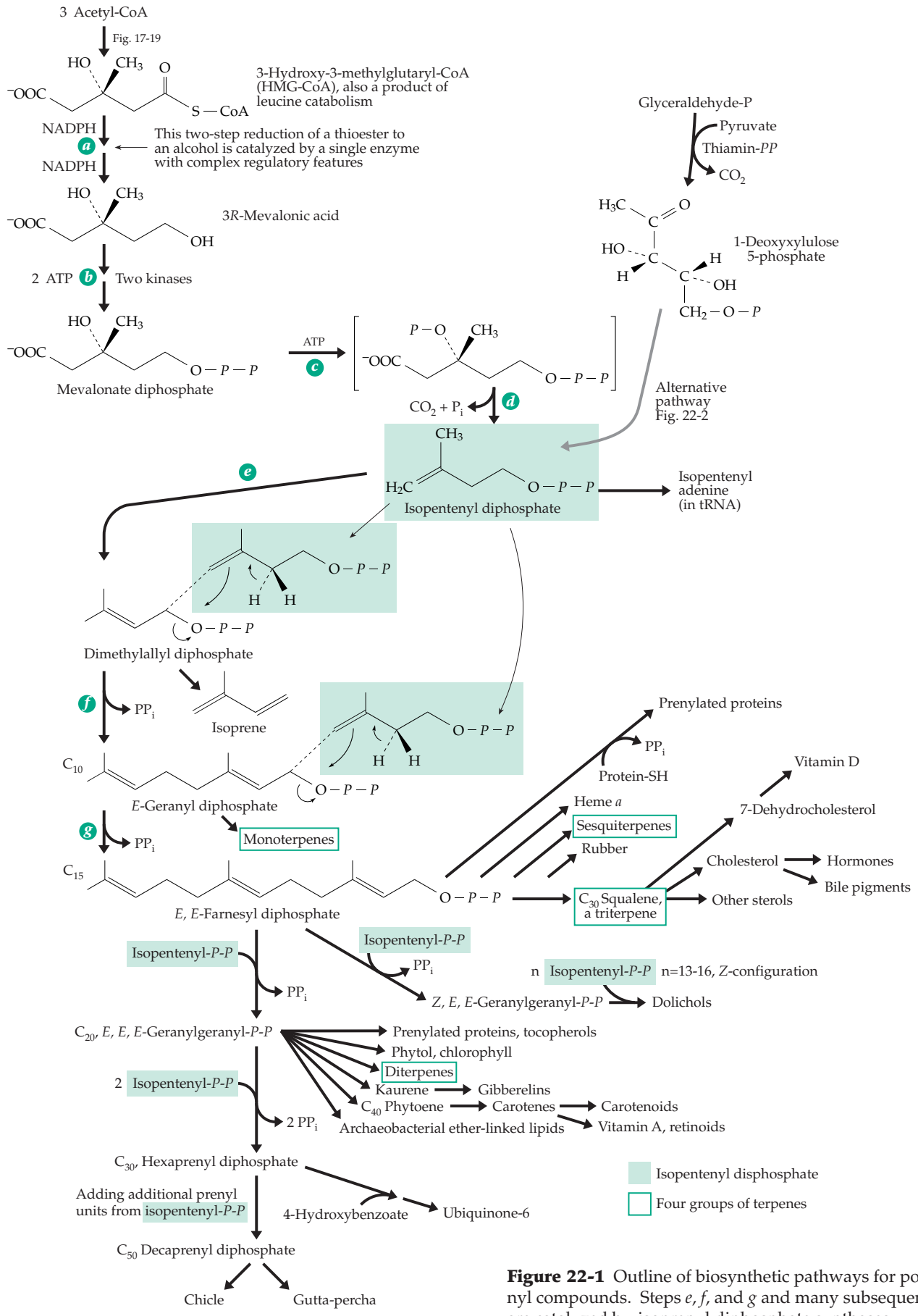


Figure 22-1 Outline of biosynthetic pathways for polyprenyl compounds. Steps *e*, *f*, and *g* and many subsequent steps are catalyzed by isoprenyl diphosphate synthases.

1. An Alternative Pathway for Isoprenoid Synthesis

It is generally agreed that mevalonate is the precursor to sterols in higher plants as well as in animals and is also the precursor to plant carotenoids. However, it is poorly incorporated into monoterpenes and into some diterpenes such as those of the taxane group.[26,27] The alternative **glyceraldehyde 3-phosphate:pyruvate pathway** explains this result. The pathway also operates in some bacteria and apparently is the sole source of isoprenoid compounds for the unicellular alga *Scenedesmus*.[28] The pathway is outlined in Fig. 22-2. Pyruvate is decarboxylated by a thiamin diphosphate-dependent enzyme,[29] and the resulting enamine is condensed with D-glyceraldehyde 3-phosphate to form 1-deoxyxylulose 5-phosphate.[28,30–31a] The latter undergoes an isomeroreductase rearrange-

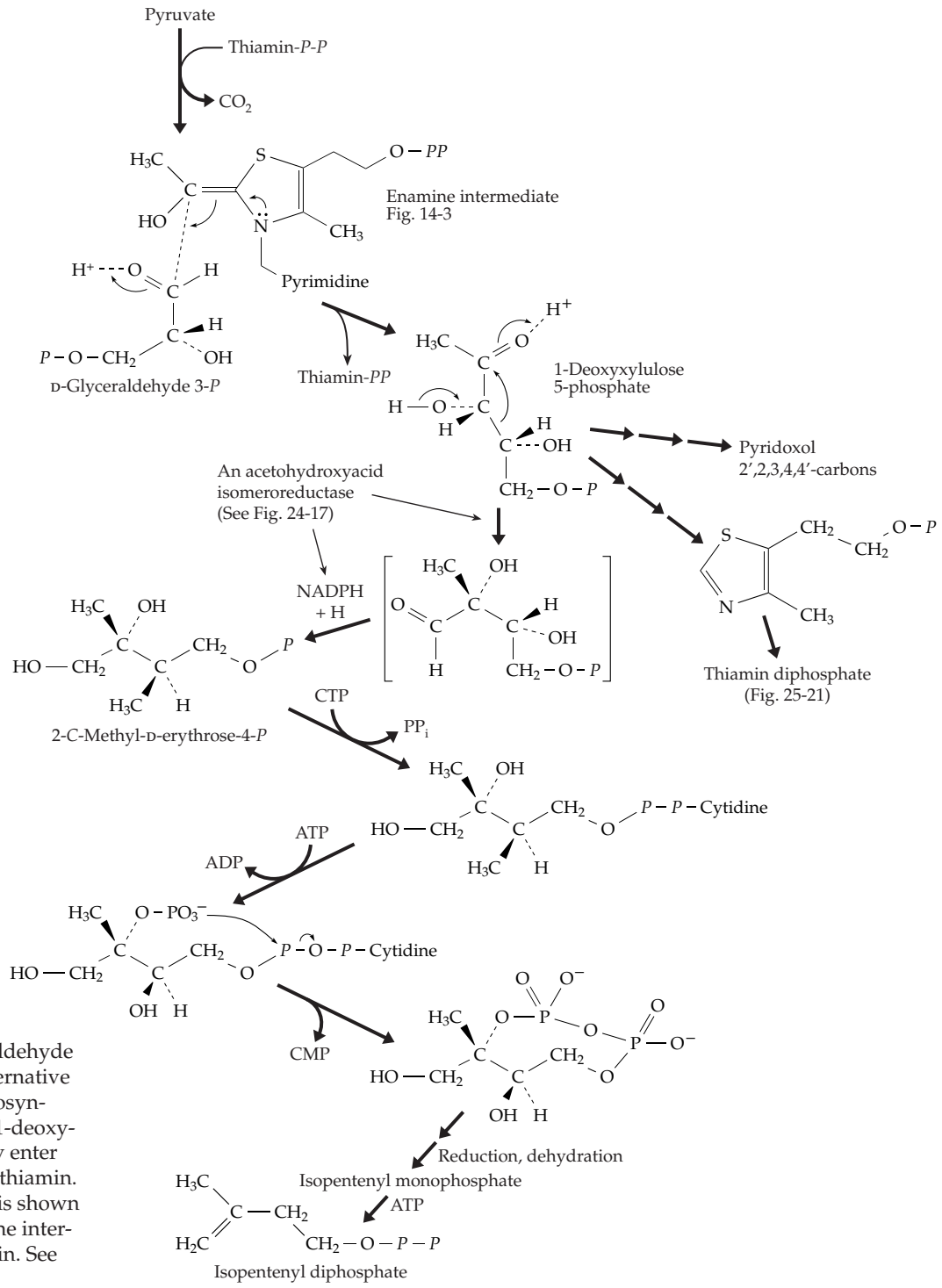
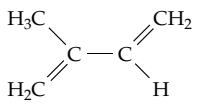


Figure 22-2 The glyceraldehyde 3-phosphate:pyruvate alternative pathway of isoprenoid biosynthesis. The intermediate 1-deoxyxylulose 5-phosphate may enter terpenes, vitamin B_6, and thiamin. Isopentenyl diphosphate is shown as the final product, but the intermediate steps are uncertain. See Lange *et al.*[32g]

ment of the type that occurs in the biosynthesis of valine and isoleucine (Fig. 24-17),[32] but the additional steps on to isopentenyl diphosphate are not obvious. However, some intermediate compounds have been identified as is indicated in Fig. 22-2.[32a–g] 1-Deoxyxylulose 5-phosphate has also been identified as an intermediate in the biosynthesis of vitamin B_6 in *E. coli*.[32a,h,33] It gives rise to the 2', 2, 3, 4, and 4' carbon atoms of pyridoxine and also to the pyrimidine ring of thiamin. See Chapter 25.

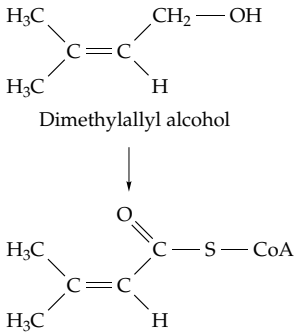
2. Isomerization and Isoprene Formation

Before polymer formation begins, one molecule of isopentenyl diphosphate must be isomerized to **dimethylallyl diphosphate** (Fig. 22-1, step *e*, Eq. 13-56).[10,33a,b] In this process the hydrogen that was in the 4-*pro-S* position of mevalonic acid (the *pro-R* position of isopentenyl pyrophosphate) is lost. Dimethylallyl diphosphate is unstable and can undergo acid-catalyzed elimination of PP_i to form isoprene, apparently by a carbocation mechanism.



Isoprene

This evidently accounts for the presence of isoprene in the breath.[34] Isoprene is also formed by many plants and is released into the atmosphere in large amounts, which contribute to photochemical formation of haze. A Mg^{2+}-dependent enzyme catalyzes the elimination of pyrophosphate.[35] Isoprene emissions rise with increasing temperature, and it has been suggested that the isoprene may dissolve in chloroplast membranes and in some way confer increased heat resistance.[36,37] Hydrolytic dephosphorylation can lead to dimethylallyl alcohol, which is oxidized in the liver to dimethylacrylyl-CoA (Eq. 22-1).



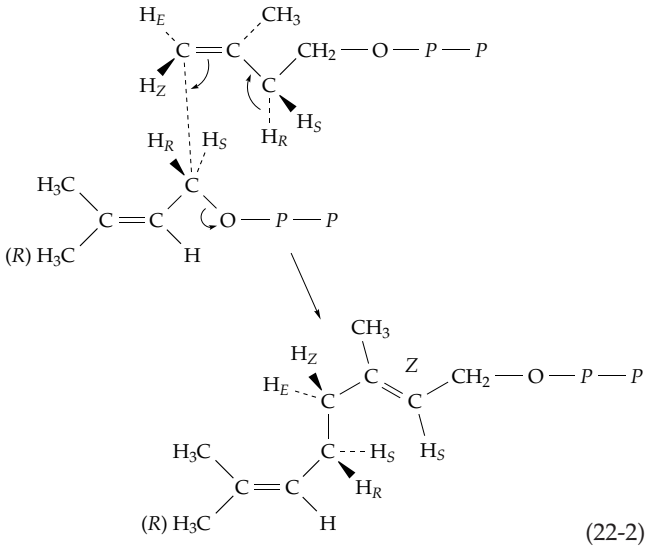
Dimethylallyl alcohol

Dimethylacrylyl-CoA (22-1)

The latter is also a catabolite of leucine and can be converted back to HMG-CoA via a biotin-dependent carboxylation (see Fig. 24-18). This provides a means of recycling the dimethylallyl alcohol back to the polyprenyl pathway.[38]

3. Polyprenyl Compounds

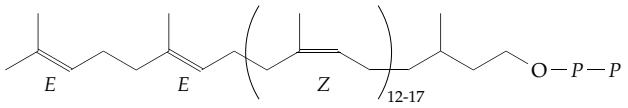
Dimethylallyl diphosphate serves as the starter piece for most polyprenyl compounds. Additional prenyl units are added, with elimination of pyrophosphate, by the action of **polyprenyl diphosphate synthases** as indicated in Fig. 22-1. Many of the products have all-trans (*E*) double bonds. A substantial number of these synthases are known and are distinguished by their chain length specificity and stereochemical properties.[39–46b] The most studied is farnesyl diphosphate synthase. The three-dimensional structure of an avian form is known.[47,47a] It catalyzes steps *f* and *g* of Fig. 22-1, joining three prenyl groups with the *E,E* (trans, trans) configuration. This protein, which consists almost entirely of packed α helices, has a large central cavity with conserved lysines and two aspartate-rich sequences (DDXXD) along its walls. These polar groups, together with magnesium ions, probably bind the pyrophosphate groups of the substrates.[48] Aspartates 224 and 225 of a bacterial form of the enzyme appear to be essential for catalytic activity.[49] The reaction is thought to be initiated by elimination of PP_i to form a carbocation to which the second prenyl unit adds as in Eq. 22-2.[50,51] For each prenyl unit a hydrogen atom that was originally the 4-*pro-S* hydrogen of mevalonic acid is lost as a proton.[52,53] Addition of another prenyl unit gives *E,E,E* geranylgeranyl diphosphate.[46a]



(22-2)

The chain length of the polyprenyl compounds appears to be determined by the protein structure of

the synthase.[45,54,54a,b] Polymerization of prenyl units can continue with the formation of high molecular weight polyprenyl alcohols such as the **dolichols** and bacterial decaprenols (Chapter 20) or of the high polymers **rubber** (all *Z* configuration), chicle, and **gutta-percha**.[6a] Dolichols, which function in the biosynthesis of glycoproteins, consist of 16–21 prenyl units and are synthesized in the endoplasmic reticulum as the diphosphates.[55–57] Farnesyl diphosphate is elongated to *Z*,*E*,*E* geranylgeranyl diphosphate, and polymerization continues with addition of 13–18 more units, all with the *Z* (cis) configuration.[46b,57a,b] However, after dephosphorylation[58] the double bond of the last unit added becomes saturated.[59] Partial absence of the required reductase causes a serious human deficiency disease involving faulty glycoprotein synthesis.[60]



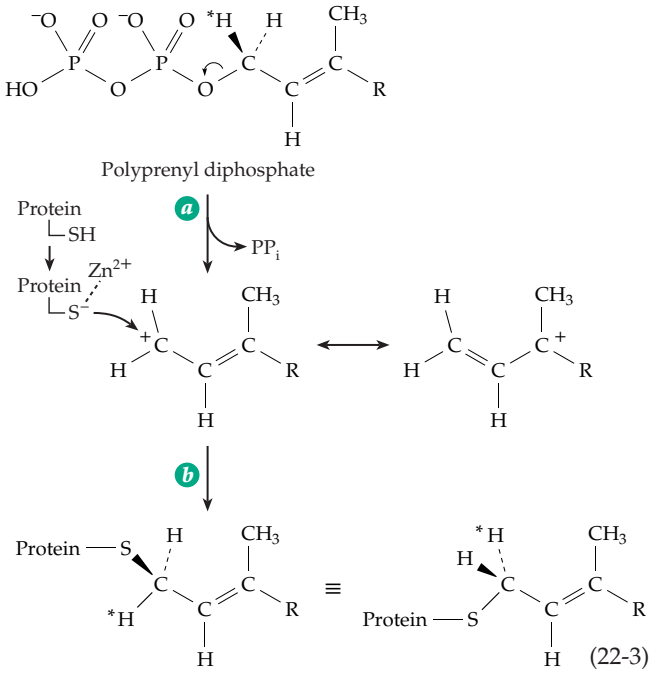
A fully extended 19-unit dolichol (dolichol-19) would have a length of about 10 nm, twice that of the bilayer in which it is dissolved. It has been suggested that the central part of the molecule has a helical structure, while the ends are more flexible. Dolichols also appear to increase the fluidity of membrane bilayers.[61] Bacterial undecaprenyl diphosphate, which has a similar function, contains only one *E* and ten *Z* double bonds[62–63a] (see p. 1152).

Rubber also contains almost entirely *Z* double bonds. Consistent with this fact is the finding that the prenyltransferases catalyzing formation of rubber promote loss of the *pro-R* proton rather than the *pro-S* proton of mevalonic acid (see Eq. 22-2). There appear to be two types of prenyltransferase in animal mitochondria giving rise to *E* and *Z* double bonds, respectively.[64] In contrast, the rubber tree contains a 137-residue protein, the **rubber elongation factor**. This small protein binds to *E* prenyltransferases causing them to form *Z* double bonds.[65] The bacterium *Micrococcus luteus* synthesizes all *E* polyprenyl alcohol diphosphates up to the C_{45} nonaprenyl compound **solanesyl diphosphate**.[66]

Chain elongation during polymerization of prenyl units can be terminated in one of a number of ways. The pyrophosphate group may be hydrolyzed to a monophosphate or to a free alcohol. Alternatively, two polyprenyl compounds may join "head to head" to form a symmetric dimer. The C_{30} terpene **squalene**, the precursor to cholesterol, arises in this way from two molecules of farnesyl diphosphate as does **phytoene**, precursor of the C_{40} carotenoids, from *E*,*E*,*E* geranylgeranyl diphosphate. The phytanyl groups of archaebacterial lipids (p. 385) arise rather directly from geranylgeranyl diphosphate by transfer of the polyprenyl group to the $-CH_2OH$ group of *sn*-glycerol 3-phosphate.[67,67a] This is followed by hydrogenation of the double bonds. Formation of diphytanyl group (p. 388) must involve additional crosslinking reactions.

4. Prenylation of Proteins and Other Compounds

Polyprenyl groups are often transferred onto thiolate ions of cysteine side chains of certain proteins that bind to membranes (p. 559).[68,69] We have previously considered the Ras family (Chapter 11). Recoverin, an important protein in the visual cycle (Fig. 23-43), is another example of a prenylated protein. Both **farnesyltransferases**[70–76f] and **geranylgeranyltransferases**[72,77–78b] have been characterized, and the three- dimensional structure of the former has been established.[73,75–76] The two-domain protein contains a seven-helix crescent-shaped hairpin domain and an α,α-barrel similar to that in Fig. 2-29. A bound zinc ion in the active site may bind the $-S^-$ group of the substrate protein after the farnesyl diphosphate has been bound into the active site.[76,79] These enzymes are thought to function by a carbocation mechanism as shown in Eq. 22-3 and with the indicated inversion of configuration.[71]



Inhibition of these prenyltransferases blocks growth of tumor cells. Many prenyltransferase inhibitors are apparently nontoxic to normal cells and are undergoing human clinical trials as anticancer drugs.[76a,79a,b] Among other important polyprenyl

compounds are the side chains of vitamin K, the ubiquinones, plastoquinones, tocopherols, and the phytyl group of chlorophyll. In all cases, a diphosphate of a polyprenyl alcohol serves as an alkyl group donor. Introduction of the polyprenyl chain into aromatic groups, such as those of the quinones (Fig. 15-24), occurs at a position ortho to a hydroxyl group in the reduced quinone (hydroquinone). The reader should be able to propose a reasonable prenyltransferase mechanism involving participation of the hydroxy group. The monoprenyl compound dimethylallyl diphosphate prenylates the N^6 position of adenine in a specific site in many tRNA molecules (Fig. 5-33)[80] as well as the C-4 position of L-tryptophan in the synthesis of ergot alkaloids.[81]

B. Terpenes

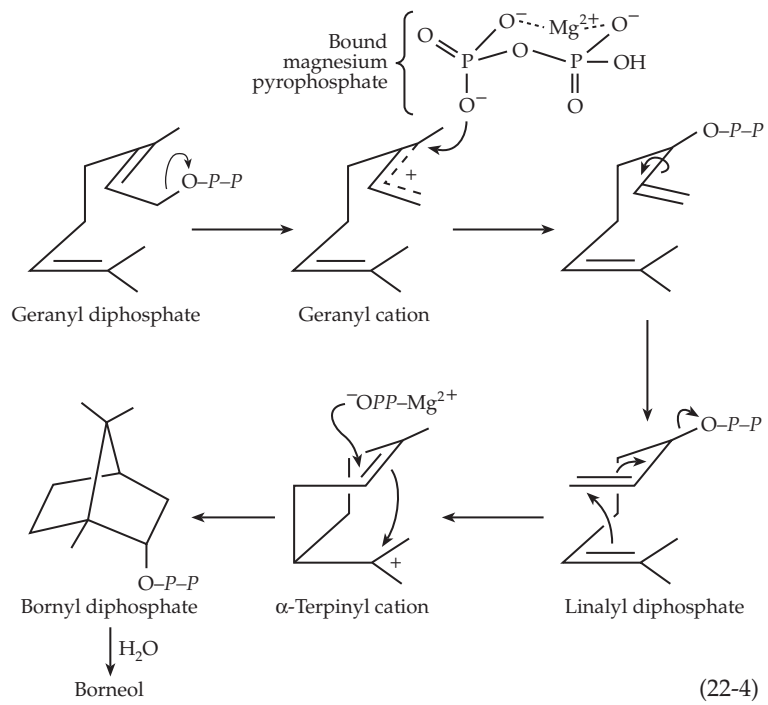
The number of small compounds that arise from isopentenyl diphosphate and are found in plants, animals, and bacteria is staggering. Just a few of these "terpenes" are shown in Figs. 22-3 and 22-4. The biosynthetic pathways have been worked out by "feeding" radioactively labeled acetate to plants and studying the characteristic labeling patterns in the terpenes. Many of the enzymes involved have been identified and studied. A given plant usually contains a large number of different terpenes, which are often concentrated in specialized "oil glands" or resinous duct tissues. Lesser amounts, often as glycosides of terpene alcohols, may be present within cells. Some terpenes occur in truly enormous amounts. For example, turpentine may contain 64% of **α-pinene** and juniper oil 65% **α-terpineol**.[82] The large quantities of α-pinene released into the air from pine trees are a major cause of photochemical smog.[83]

Terpenes have a variety of functions. Plant terpenes may deter herbivores and attract pollinators. They may participate in competition among plants and may act as antibiotics, called **phytoalexins**, to protect plants from bacteria and fungi.[84] In invertebrate animals terpenes serve as hormones, pheromones, and defensive repellants (Figs. 22-3, 22-4). The terpene squalene is the precursor to sterols. Some terpenes are toxic. For example, thujone (Fig. 22-3), which is present in the liqueur absinthe, causes serious chronic poisoning.[84]

1. Biosynthesis of Monoterpenes

The compounds of Figure 22-3 each contain ten carbon atoms and are called monoterpenes. They occur largely in plants, but some function in arthropods as pheromones. As with chain elongation, the cyclization of geranyl diphosphate to the various monoterpenes appears to occur through loss of pyrophosphate (as PP_i) with formation of an intermediate carbocation such as that depicted in Equation 22-3.[85–88a] Similar mechanisms initiate cyclization of sesquiterpenes and diterpenes. Numerous terpene cyclases have been isolated, and several have been studied carefully. A stereochemical view of the formation of borneol is illustrated in Eq. 22-4. Both linalyl-*PP* and bornyl-*PP* are intermediates. Croteau and associates suggested that a tight ion pair between carbocation and magnesium pyrophosphate is maintained at each stage.[86,89]

As is indicated in Fig. 22-3, the same intermediate cation can yield a variety of end products. For example, pure geranyl diphosphate: pinene cyclase catalyzes formation of several other terpenes in addition to α-pinene.[89] Another aspect of terpene synthesis is that insects may convert a plant terpene into new compounds for their own use. For example, **myrcene**, which is present in pine trees, is converted by bark beetles to **ipsenol** (Fig. 22-3), a compound that acts as an aggregation pheromone.[90]



(22-4)

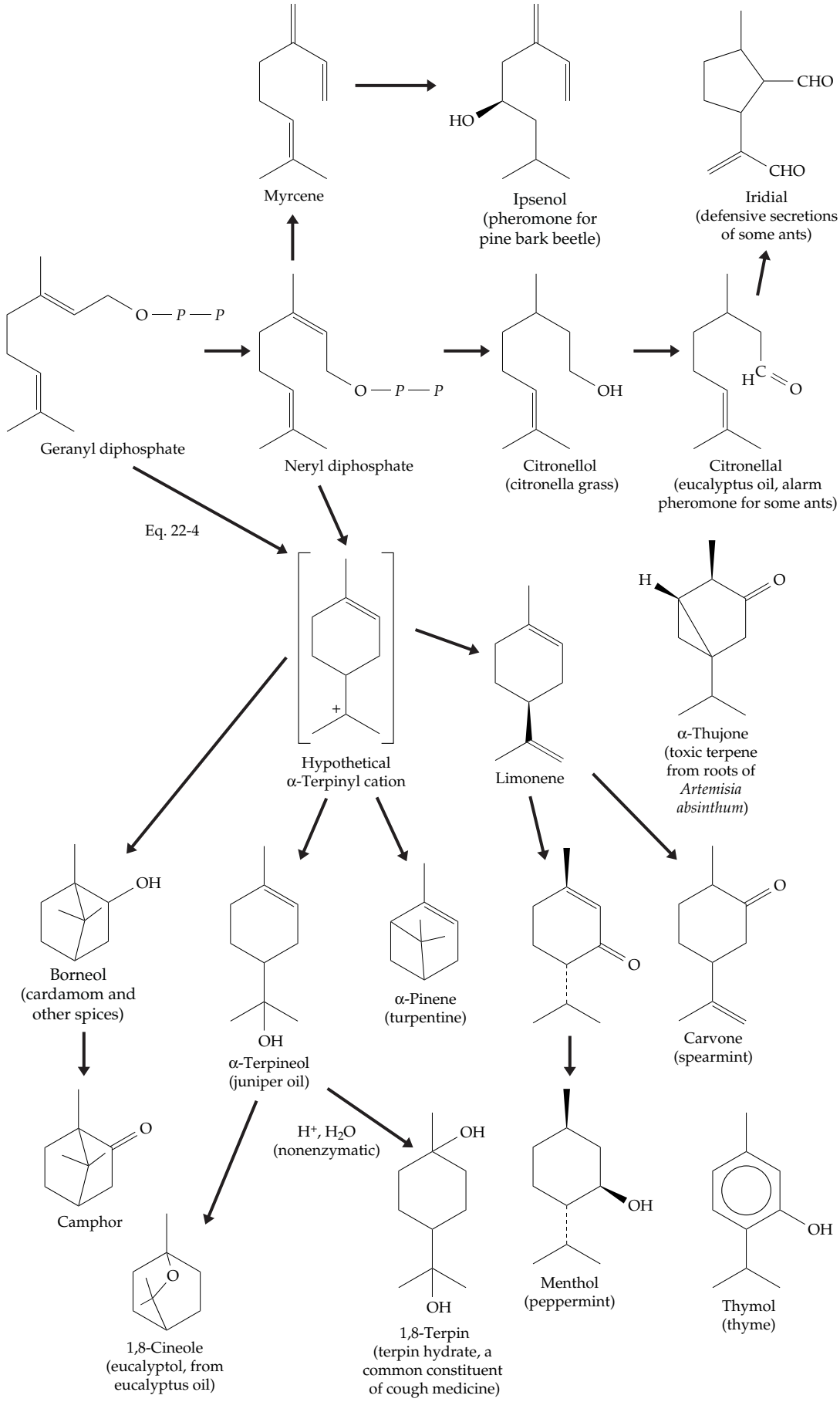
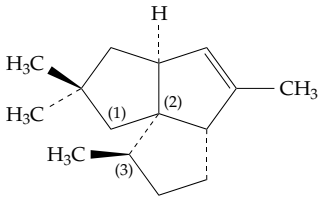


Figure 22-3 Probable biosynthetic pathways of some monoterpenes and related substances. Some of the natural sources are indicated.

2. Sesquiterpenes and Diterpenes

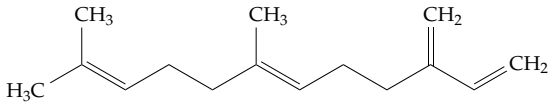
Most of the compounds shown in Figure 22-4 are derived from the C_{15} farnesyl diphosphate. There are more than 300 known cyclic structures among these **sesquiterpenes,** and many sesquiterpene synthases have been characterized.[91,91a] **Aristolochene** is formed by the action of a 38-kDa cyclase that has been isolated from species of *Penicillium* and *Aspergillus*.[92–94] Notice that the synthesis must involve two cyclization steps and migration of a methyl group. Three-dimensional structures are known for at least two terpene synthases,[95,96] and comparison of gene sequences suggests that many others have similar structures. The **5-epi-aristolochene synthase** of tobacco makes the 5-epimer of aristolochene (Fig. 22-4). It binds the diphosphate group within a central cavity using two Mg^{2+} ions held by carboxylates, some of which are in the DDXXD sequence found also in polyprenyl diphosphate synthases. The enzyme active sites of both the epi-aristolochene synthase and a **pentalenene synthase** from *Streptomyces*[96] are rich in polar groups that form hydrogen-bonded networks and which participate in proton abstraction and donation during the rearrangement reactions that must occur.



Pentalenene

Aromatic groups that are also present may assist in stabilizing intermediate carbocationic species. Deprotonation by an aspartate side chain in the epi-aristolochene synthase has been proposed to assist the cyclization; subsequent reprotonation by an Asp•Tyr•Asp triad would generate a new carbocation and promote the necessary methyl group migration. A detailed step-by-step mechanism has been proposed.[95] The fungal pentalenene synthase has an active site histidine, which is proposed to serve as proton acceptor and donor for the several steps of the reaction of farnesyl diphosphate.[96] The carbon atoms originating from C1 to C3 of the precursor have been marked on the pentalenene structure as well as on the aristolochene structure in Fig. 22-4. We now see how synthases can guide the terpene cyclase reactions to give specific products.

Another sesquiterpene synthase forms **trichodiene**, the parent compound for a family of mycotoxins and antibiotics.[97,97a,b] A different sesquiterpene synthase, present in peppermint and also in a wide range of other plants and animals, forms the acyclic **(E)-β-farnesene**.[98] It serves as alarm pheromone for aphids, has a variety of signaling functions in other insects, and is a urinary pheromone in mice.[98]



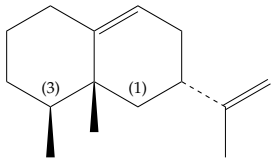
(*E*)-β-Farnesene

Abscisic acid, one of five known types of plant hormone of general distribution throughout higher plants, is not regarded as a true terpene because it arises by degradation of a carotenoid.[99–99b] However, its structure (Fig. 22-4) is that of a sesquiterpene.

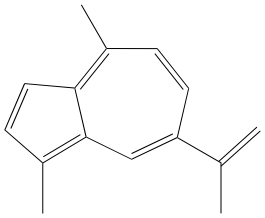
The C_{20} **diterpenes** are derived from geranylgeranyl-*PP*. Among the best known members are another group of plant hormones, the **gibberellins**.[100–102a] The first gibberellin was isolated as a product of plants infected with a *Fusarium* fungus. The rice plants grew in an abnormally tall, weak form. Subsequently, this multimembered class of over 50 highly modified diterpenes have been shown to have a variety of regulatory functions in all higher plants. For example, gibberellins are essential for stem elongation.

Equation 22-5 gives an abbreviated biosynthetic sequence for gibberellin A_1. The ring closure of step *a*, Eq. 22-5, may be initiated by protonation of the double bond at the left of the first structural formula. The resultant carbocation could initiate the consecutive closure of the two rings and the loss of a proton from a methyl group (step *b*) to yield copalyl-*PP*. Steps *c* and *d*, Eq. 22-5, each represent several reactions. In step *c*, pyrophosphate is eliminated, and the methyl group that becomes a methylene in *ent*-kaurene (enantiotopic kaurene) undergoes migration.[102b] Step *d* involves several hydroxylation and oxidation steps as well as a ring contraction through which one of the original methylene groups ends up as a carboxyl group in the final product.[102,103,103a] Deactivation of the hormone is initiated by a 2-oxoglutarate-dependent dioxygenase.[103b]

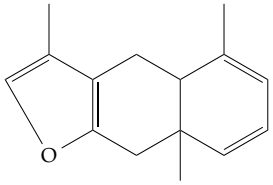
The **juvenile hormone** of insects (Fig. 22-4) is also of polyprenyl origin.[104,105] However, two of the methyl groups have been converted to (or replaced by) ethyl groups. The isolation and identification of the structure of the juvenile hormone was a difficult task. After its completion it was a surprise to researchers to discover that a large variety of synthetic compounds, sometimes with only a small amount of apparent structural similarity, also serve as juvenile hormones, keeping insects in the larval stage or preventing insect eggs from hatching. Furthermore, a number of plant products such as **juvebione** (Fig. 22-4),[106] which was originally isolated from paper, have the same effect. Thus, in nature products of plant metabolism have a profound effect upon the development of insects that



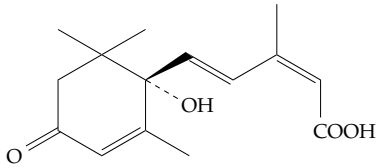
Aristolochene, parent hydrocarbon of many fungal toxins



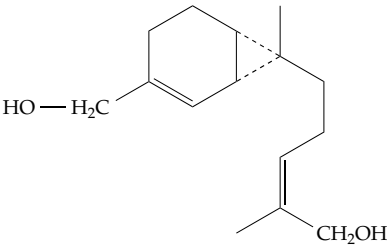
Lactarazulene, dye from the mushroom *Lactarius deliciosus*



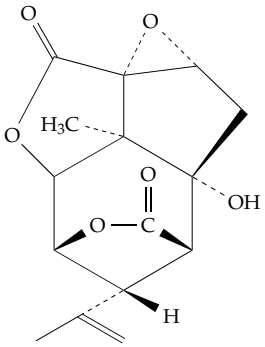
Furanoeudesma-1,3-diene, a sesquiterpene from myrrh with analgesic properties



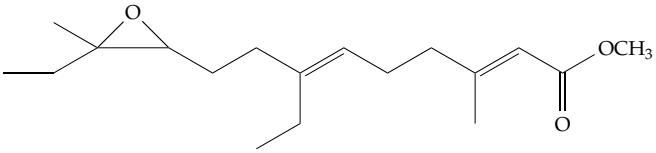
Abscisic acid, an important hormone of green plants



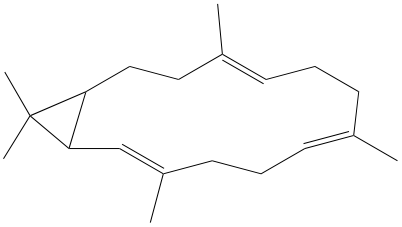
Sirenin, a sperm attractant produced by female gametes of the phycomycete *Allomyces*



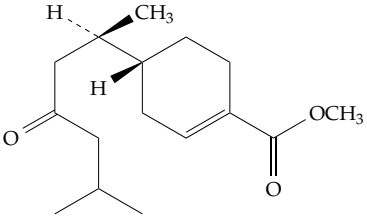
Picrotoxin, a plant toxin sometimes used to kill fish



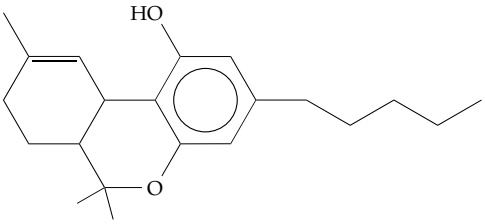
Juvenile hormone, a substance acting to maintain the larval stage of insects



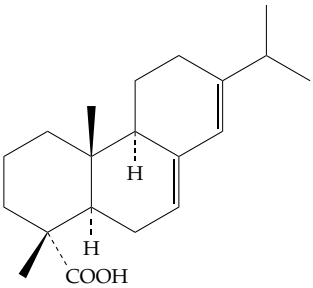
Casbene, a macrocyclic phytoalexin



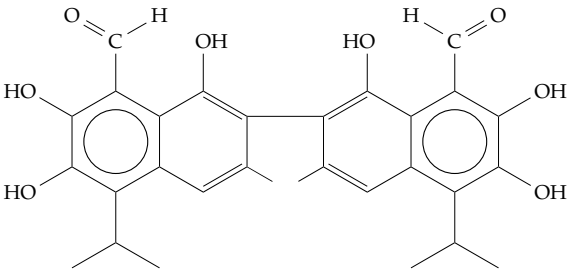
Juvebione, a product of the balsam fir with juvenile hormone activity



Δ'-3,4-*trans*-Tetrahydrocannabinol (Δ'-THC) a psychotropic component of marijuana

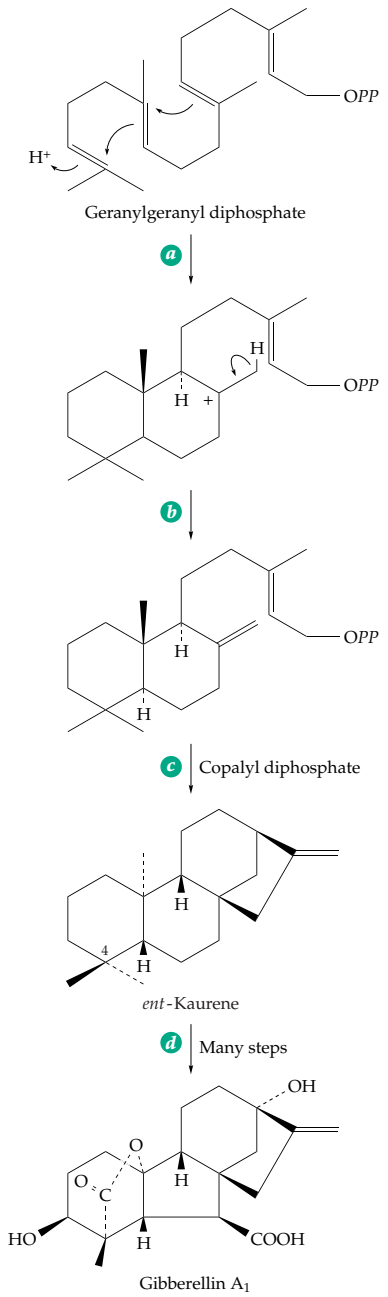


Abietic acid, a component of rosin, from conifers



Gossypol, toxic, dimeric sesquiterpene from cottonseed

Figure 22-4 More terpenes and related substances. The numbers in parentheses on the aristolochene structure are those of atoms in the precursor farnesyl diphosphate.



(22-5)

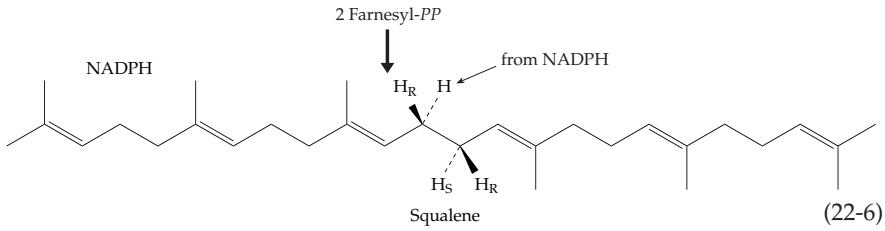
eat the plants. There is interest in the possible use of juvenile hormone, or of synthetic compounds mimicking its action, as insecticides.

Many conifers secrete **oleoresin** (pitch) in response to attack by bark beetles. The oleoresin contains approximately equal amounts of turpentine (a mixture of monoterpenes and sesquiterpenes) and diterpenoid rosin including **abietic acid** (Fig. 22-4).[91,107,107a] The oleoresin is toxic to beetles and, after evaporation of the turpentine, forms a hard rosin seal over the wounds. **Casbene** (Fig. 22-4) is a diterpene produced by castor beans as a phytoalexin (Chapter 31), an antifungal antibiotic.[108] The synthesis of the anticancer compound taxol (Box 7-D) from geranylgeranyl diphosphate involves extensive oxidative and other modification.[109]

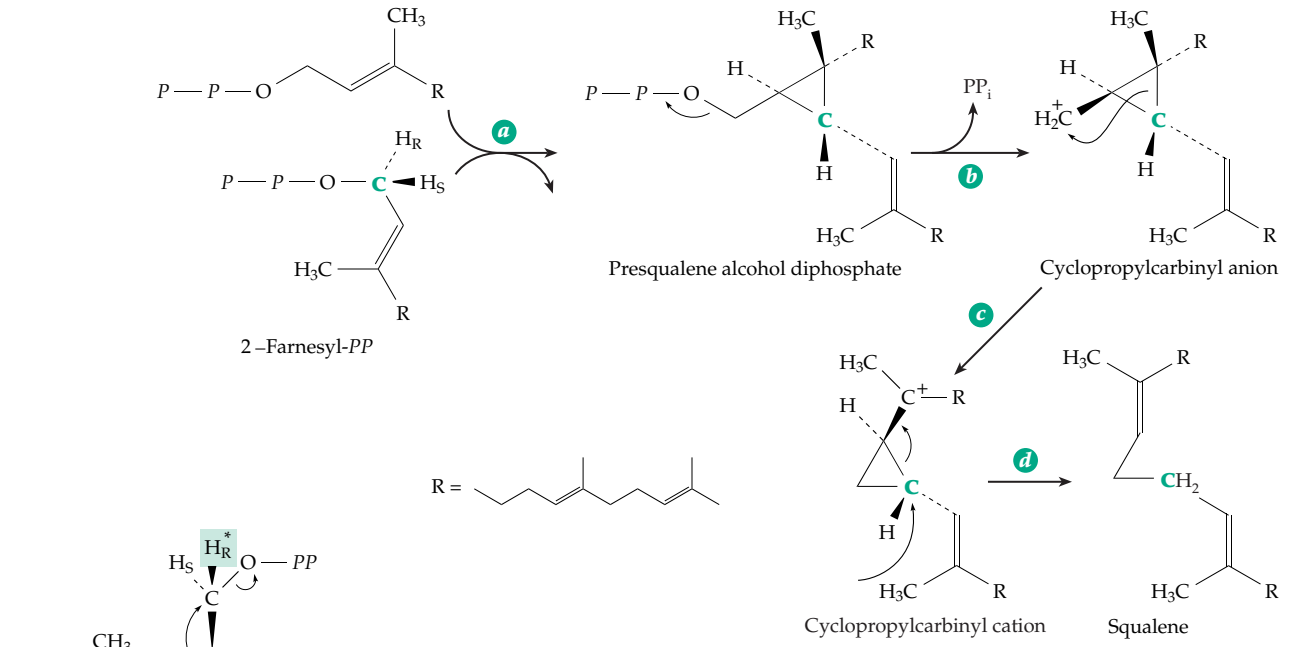
3. Formation of the Symmetric Terpenes, Squalene and Phytoene

Two molecules of the C_{15} farnesyl diphosphate can be joined "head to head" to form the C_{30} squalene (Eq. 22-6). Similarly, two C_{20} **geranylgeranyl-*PP*** molecules can be joined to form the C_{40} phytoene (Fig. 22-5), a precursor of carotenoid pigments of plants.[110] In the synthesis of squalene both pyrophosphate groups are eliminated from the precursor molecules, and one proton from C-1 of one of the molecules of farnesyl-*PP* is lost. The other three C-1 hydrogens are retained. At the same time, one proton is introduced from the *pro-S* position of NADPH. Squalene synthase has been difficult to obtain,[111,112] and its mechanism has been uncertain. However, there is strong evidence in favor of carbocationic intermediates.[113–114b] The first step (Eq. 22-7, step *a*) involves reaction of the initial carbocation with the double bond of the second farnesyl-*PP* to form the cyclopropane derivative **presqualene alcohol-*PP***, which was first isolated from yeast as the free alcohol by Rilling and associates.[115] The loss of the second pyrophosphate (Eq. 22-7, step *b*) generates a cyclopropylcarbinyl cation, which can rearrange (step *c*) to a more stable tertiary cation.[112,113] The latter is reduced by NADPH with opening of the cyclopropane ring (Eq. 22-7, step *d*). Once formed squalene diffuses within and between membranes with the help of cytosolic protein carriers.[116]

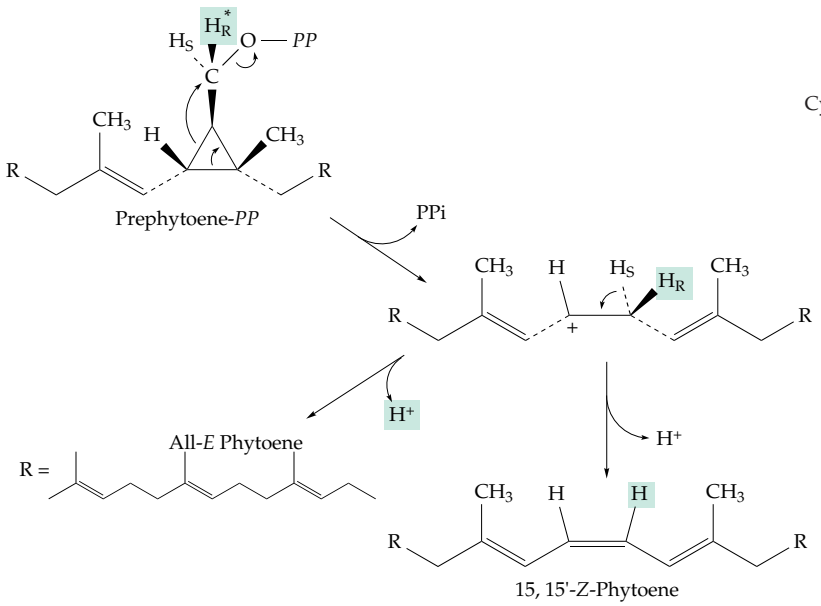
Phytoene (Fig. 22-5) is apparently formed from geranylgeranyl-*PP* via **prephytoene-*PP*,** whose structure is entirely analogous to that of presqualene-*PP*.[44,117] However, no reduction by NADH is required (Eq. 22-8). It is known that the 5-*pro-R* hydrogen atoms of mevalonate are retained in the phytoene as indicated by a shaded box in Eq. 22-8. Elimination of the other (*pro-S*) hydrogen yields 15,15'–*Z* phytoene (*cis*-phytoene), while elimination of the *pro-R* hydrogen yields all-*E* (trans) phytoene. Higher plants and fungi form mostly *cis*-phytoene, but some bacteria produce the all-*E* isomer.[118]



(22-6)

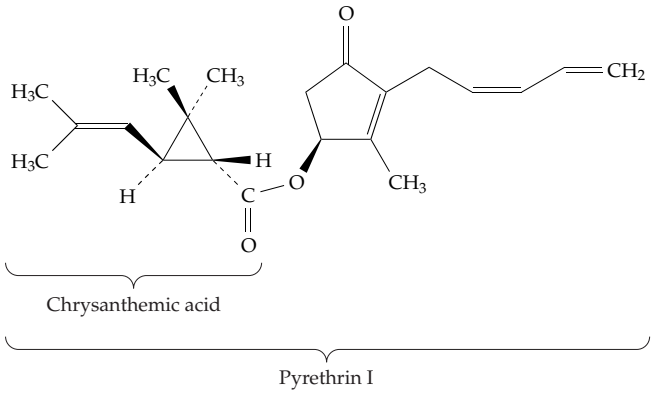


(22-7)



(22-8)

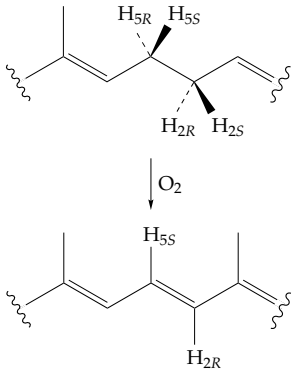
Another polyprenyl compound formed by a head to head condensation is **chrysanthemic acid**. This monoterpene component of the pyrethrum insecticides is formed by chrysanthemums from two molecules of dimethyallyl-*PP* via an intermediate analogous to presqualene alcohol-*PP*.[118a]



A quite different "tail to tail" condensation, whose chemistry is still obscure,[118b] must occur in archaebacteria whose lipids contain the C_{40} diphytanyl alcohol. An example is the diglyceryltetraether,[119,120] whose structure is shown on p. 388.

C. Carotenes and Their Derivatives

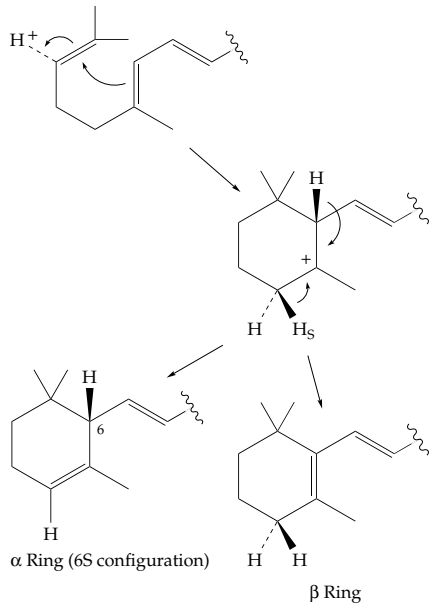
Phytoene can be converted to the carotenes by pathways indicated in part in Fig. 22-5 and Eq. 22-10. One of the first products is **lycopene**, the red pigment of tomatoes and watermelons, which is an all-trans compound. If 15-*Z* phytoene is formed, it must, at some point, be isomerized to an all-*E* isomer, and four additional double bonds must be introduced. The isomerization *may* be nonenzymatic. The double bonds are created by an oxygen-dependent **desaturation,** which occurs through the trans loss of hydrogen atoms.



(22-9)

Desaturation takes place in a stepwise fashion, and many intermediate compounds with fewer double bonds are known (Eq. 22-10).[118,121–123] The enzymes required have not been characterized well until recently. Plant enzymes are present in small amounts, and isolation has been difficult. However, the genes for carotenoid biosynthesis in such bacteria as the purple photosynthetic *Rhodobacter*,[118,124] *Rhodospirillum*,[125] and *Rubrivarax*,[126] the cyanobacterium *Synechococcus*,[127] and the nonphotosynthetic *Erwinia*[44,118] have been cloned, sequenced, and used to produce enzymes in quantities that can be studied. Matching genes from higher plants have also been cloned and expressed in bacteria.[123]

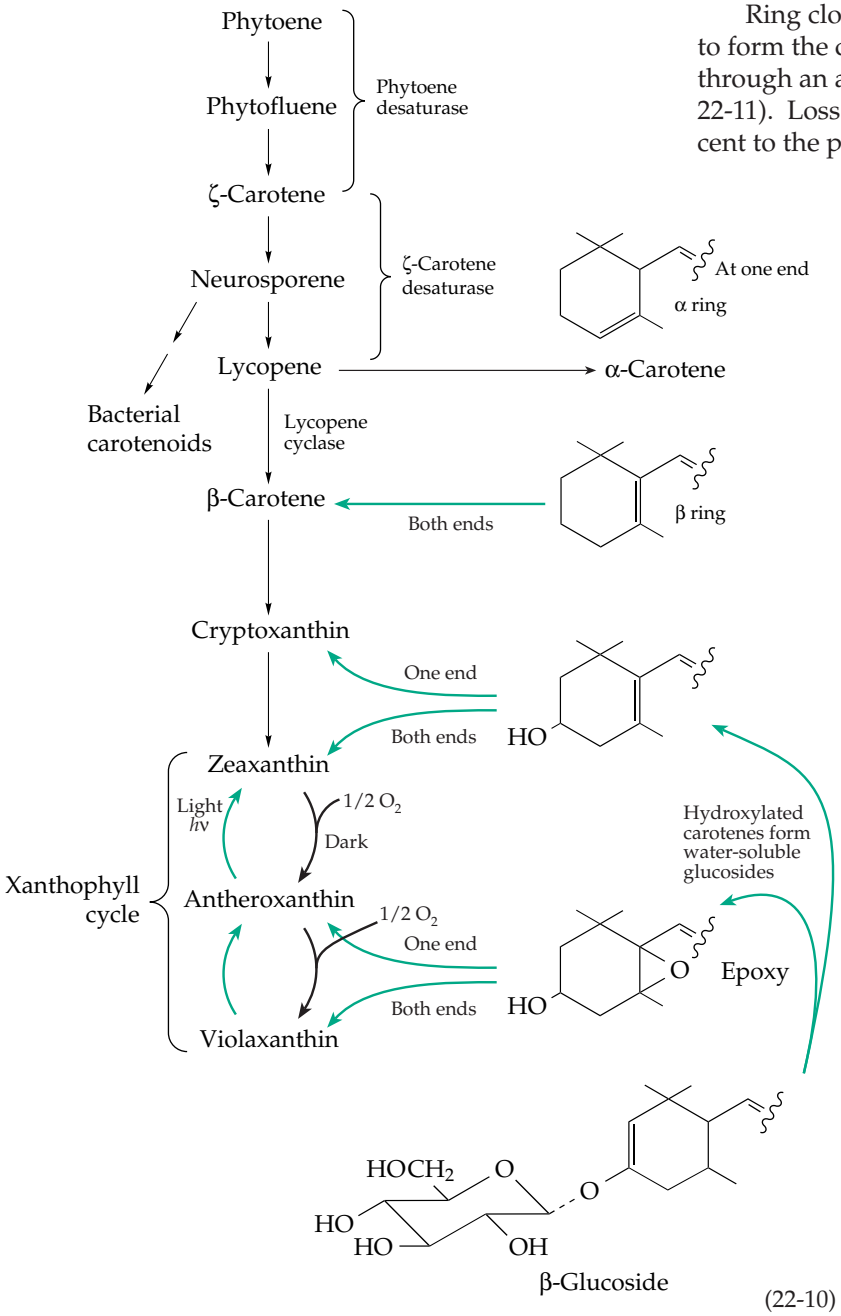
Ring closure at the ends of the lycopene molecule to form the carotenes can be formulated most readily through an acid-catalyzed carbocation mechanism (Eq. 22-11). Loss of one or the other of two protons adjacent to the positive charge leads to the β ring of β-carotene or to the α ring of α-carotene.[110,123a] Compounds with only one ring may also be formed.[123b] In many bacteria



(22-11)

these rings are not formed at all, but the open-chain (acyclic) cartenoids may be modified in ways similar to those of higher plants.[118,124,125]

A genetic engineering success is the transfer of genes for synthesis of β-carotene into rice. The resulting "golden rice" contains enough carotene in its endosperm to make a significant contribution to the vitamin A needs of people for whom



(22-10)

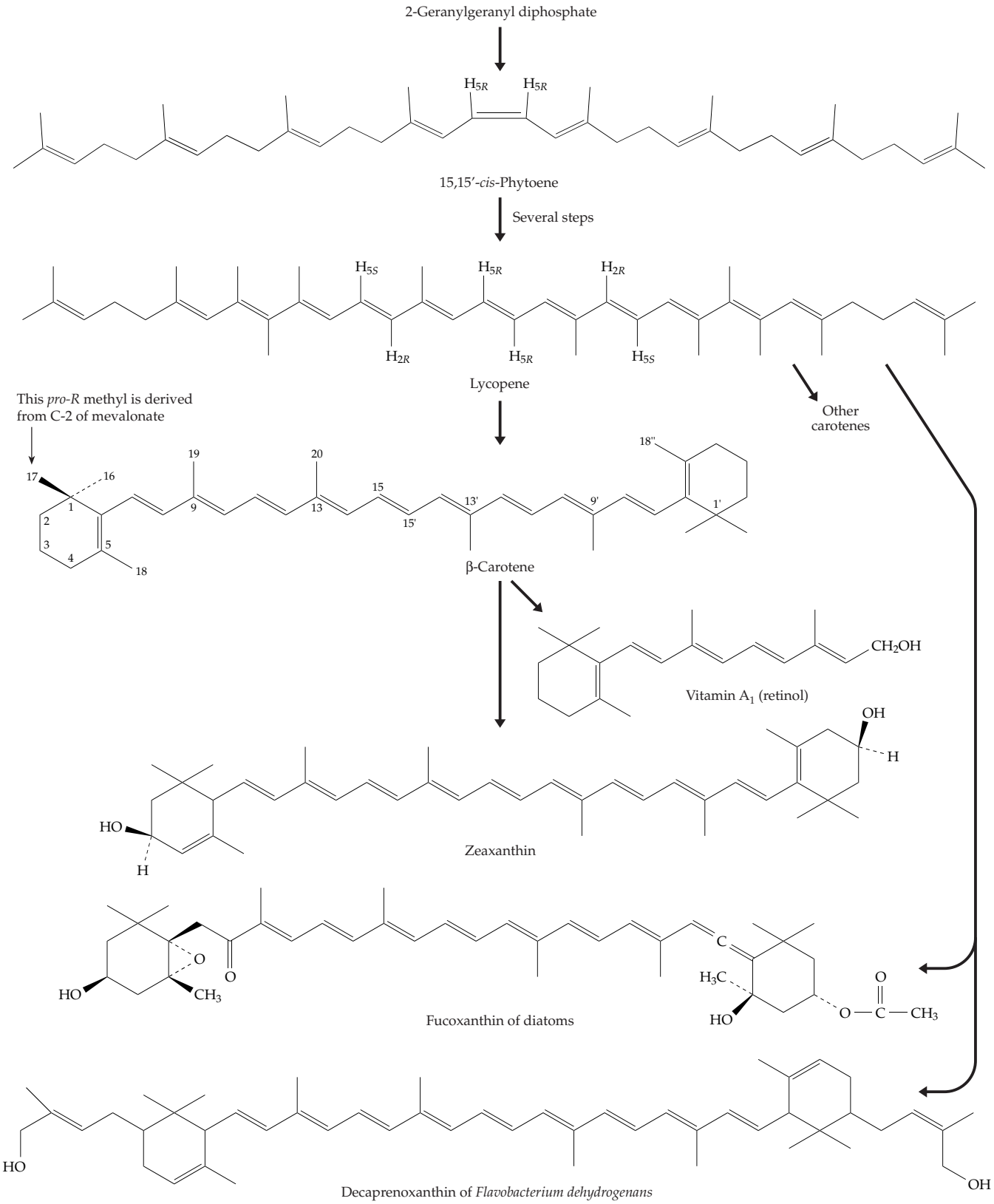
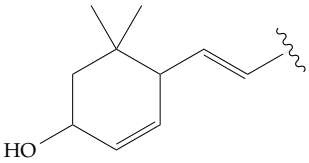


Figure 22-5 Structures and partial biosynthetic pathways for a few of the more than 600 known carotenoid compounds. The origin of some hydrogen atoms from mevalonate is shown, using the numbering for mevalonate. The numbering system for C_{40} carotenoids is also indicated.

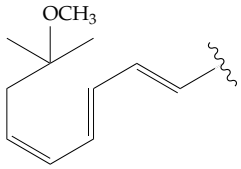
rice is a major food. There are estimated to be over 100 million vitamin A-deficient children in the world. As many as one-quarter of a million of these go blind each year.[127a] It is hoped that the golden rice will help to alleviate the problem. At the same time, an ongoing program is supplying vitamin A, which is stored in the liver, at regular intervals of time to many children.[127b]

1. Xanthophylls and Other Oxidized Carotenes

Carotenes can be hydroxylated and otherwise modified in a number of ways.[110,128–131] The structure of zeaxanthin, one of the resulting **xanthophylls**, is indicated in Fig. 22-5. Some other xanthophylls are shown in Eq. 22-10. Lutein resembles zeaxanthin, but the ring at one end of the chain has been isomerized by a shift in double bond position to the accompanying structure. The photosynthetic bacterium *Rhodospirillum rubrum* has its own special carotenoid spirilloxanthin, which has the accompanying structure at both ends of the chain.



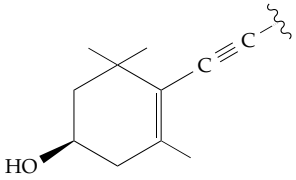
Lutein (structure at one end of molecule)



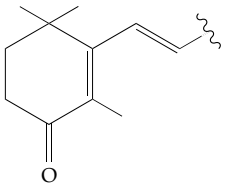
Spirilloxanthin

Fucoxanthin (Fig. 22-5) is the characteristic brown pigment of diatoms. One end of the molecule has an epoxide, also formed by the action of O_2, while the other end contains an **allene** structure rare in nature. Even so, fucoxanthin may be the most abundant carotenoid of all. The structure of the allene-containing end of the fucoxanthin molecule (turned over from that shown in Fig. 22-5) is also given in Eq. 22-12. Figure 22-5 does not indicate the stereochemistry of the allene group correctly; the carotenoid chain protrudes behind the ring as drawn in the equation. **Violaxanthin** contains epoxide groups in the rings at both ends of the molecule (Eq. 22-12). An isomerase in algae converts violaxanthin into **neoxanthin** (Eq. 22-12), which contains the allene structure at one end. Subsequent acetylation yields fucoxanthin.

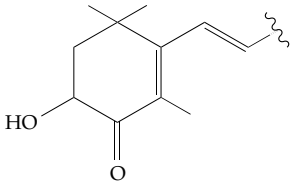
Other algal carotenoids contain acetylenic triple bonds. For example, **alloxanthin** has the following structure at both ends of the symmetric molecule. The symmetric carotenoids **canthaxanthin** and **astaxanthin** have oxo groups at both ends:



Alloxanthin



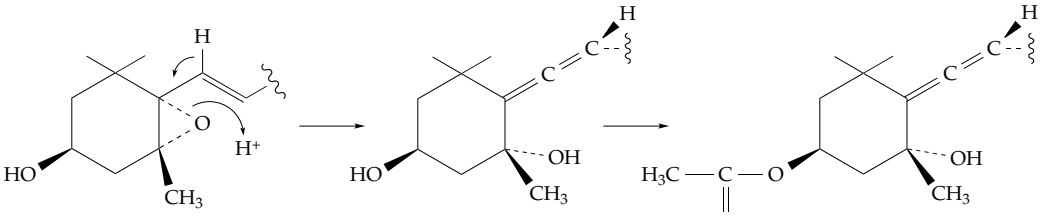
Canthaxanthin



Astaxanthin

These carotenoids have a limited distribution and occur as complexes, perhaps in Schiff base linkage, with proteins. Astaxanthin–protein complexes with absorption maxima ranging from 410 nm to 625 nm or more provide the color to the lobster's exoskeleton.[130,132] Whereas most naturally occurring carotenoids have all-*E* double bonds, mono-*Z* isomers of canthaxanthin are found in the colored carotenoproteins of the brine shrimp *Artemesia*.[133]

Some bacteria synthesize C_{50} carotenoids such as decaprenoxanthin (Fig. 22-5), the extra carbon atoms at each end being donated from additional prenyl groups, apparently at the stage of cyclization of lycopene.[134] Thus, a carbocation derived by elimination of pyrophosphate from dimethylallyl-*PP* could replace the H^+ shown in the first step of Eq. 22-11. The foregoing descriptions deal with only a few of the many known structural modifications of carotenoids.[2,135,136]



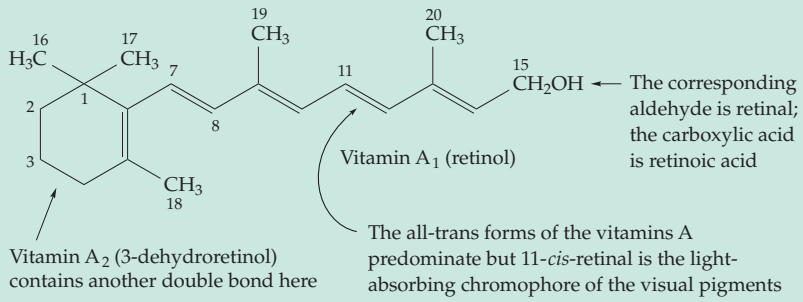
Violaxanthin (ring structure present at both ends)

Neoxanthin (this structure occurs at one end only)

Fucoxanthin

(22-12)

BOX 22-A VITAMIN A



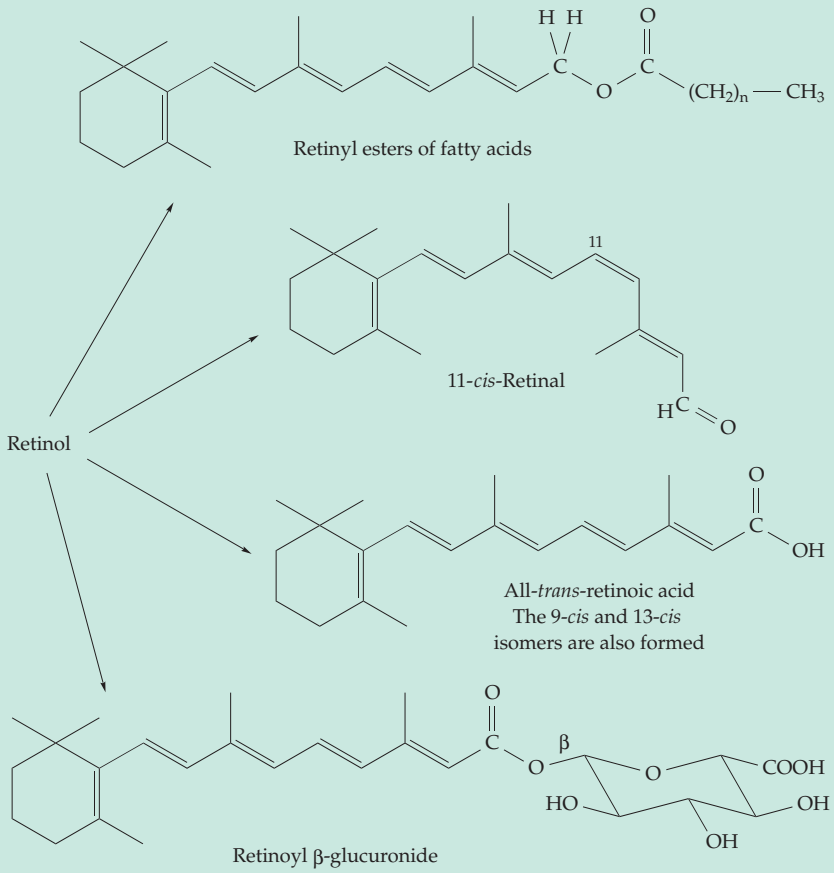
The recognition in 1913 of vitamin A (Box 14-A) was soon followed by its isolation from fish liver oils.[a–c] Both vitamin A_1 (**retinol**) and vitamin A_2 are 20-carbon polyprenyl alcohols. They are formed by cleavage of the 40-carbon β-carotene (Fig. 22-5) or other carotenoids containing a β-ionone ring. While the carotenes are plant products, vitamin A is produced only in animals, primarily within cells of the intestinal mucosa.[d–f] The carotene chains are cleaved in the center, and to some extent in other positions,[g] by oxygenases; β-carotene yields as many as two molecules of the vitamin A aldehyde **retinal**.[h–i] The retinal is reduced by NADH to retinol which is immediately esterified, usually with saturated fatty acids, by transfer of an acyl group from a fatty acyl-CoA or from phosphatidylcholine. The resulting retinyl esters are transported in chylomicrons. They remain in the chylomicron remnants (see Fig. 21-1), which are taken up by the liver where both hydrolysis and reesterification occur.[j] Vitamin A is one of the few vitamins that can be stored in animals in relatively large quantities. It accumulates in the liver, mainly as retinyl palmitate, in special storage cells termed stellate cells.[f] The human body usually contains enough vitamin A to last for several months.

Free retinol is released from the liver as a 1:1 complex of retinol with the 21-kDa **retinol-binding protein**.[k,l] This protein is normally almost saturated with retinol and is bound to another serum protein, the 127-residue **transthyretin** (prealbumin).[m,n] Some of the retinol is oxidized to **retinoic acid**. Both all-*trans* and 13-*cis*-retinoic acids as well as 5,6-epoxyretinoic acid are found in tissue.[f,o,p,pq]

Another metabolite, which may be very important, is **retinoyl β-glucuronide**.[q,r,rs]

Cell surfaces of body tissues appear to contain receptors for the retinol-binding protein. Many cells also contain cytoplasmic retinol-binding proteins[s–u] as well as proteins that bind retinoic acid.[u–y] These proteins are members of the large superfamily of hydrophobic transporter molecules described in Box 21-A. This includes the milk protein **β-lactoglobulin,** which also forms a complex with retinol.[z,aa]

A strikingly early symptom of vitamin A deficiency is **night blindness**. A variety of other symptoms include dry skin and hair, conjunctivitis of the eyes, retardation of growth, and low resistance to infection. The skin symptoms are particularly noticeable in the internal respiratory passages and alimentary canal lining. About 0.7 mg/day of vitamin A is required by an adult. The content of vitamin A in foods is often expressed in terms of international units: 1.0 mg of retinol equals 3333 I.U.



BOX 22-A VITAMIN A (continued)

Vitamin A, as retinal, has a clearly established role in vision (Chapter 23) and apparently has a specialized function in reproduction. In vitamin A deficiency no sperm cells are formed in males, and fetal resorption occurs in females. Rats deprived of vitamin A but fed retinoic acid become blind and sterile but otherwise appear healthy.[e,bb] Evidently either the alcohol or the aldehyde has an essential function in reproduction, whereas bone growth and maintenance of mucous secretions requires only retinoic acid. Indeed, retinoic acid is 100 to 1000 times more active than other forms of vitamin A in these differentiation functions.[r]

In vitamin A deficiency the internal epithelial surfaces of lungs and other tissues, which are usually rich in mucous secreting cells and in ciliated cells, develop thick layers of keratinizing squamous cells similar to those on the external surface of the body. The synthesis of some mannose- and glucosamine-containing glycoproteins consequently decreases.[cc] The major effects of retinoic acid is evidently through regulation of transcription (Chapter 28). In developing lungs retinoic acid promotes the transformation of undifferentiated epithelial cells into mucus-secreting cells.[dd]

Do we know all of the special chemistry of vitamin A that is involved in its functions? Retinal could form Schiff bases with protein groups as it does in the visual pigments. Redox reactions could occur. Conjugative elimination of water from retinol to form **anhydroretinol** is catalyzed nonenzymatically by HCl. Anhydroretinol occurs in nature and

Anhydroretinol

may serve as an inhibitor of the action of 14-hydroxy-*retro*-retinol in lymphocyte differentiation.[ee,ff]

Much recent interest has been aroused by the fact that retinoid compounds, including both retinol and retinoic acid, reduce the incidence of experimentally induced cancer. In addition, 13-*cis*-retinoic acid taken orally is remarkably effective in treatment of severe cystic acne.[gg] However, both vitamin A and retinoic acid in large doses are **teratogenic**, i.e., they cause fetal abnormalities. The use of 13-*cis*-retinoic acid during early phases of pregnancy led to a high incidence of major malformations in infants born.[hh]

[a] Moore, T. (1981) *Trends Biochem. Sci.* **6**, 115–116
[b] Moore, T. (1957) *Vitamin A*, Elsevier, Amsterdam
[c] Olson, J. A. (1968) *Vitam. Horm.(N. Y.)* **26**, 1–63
[d] Blomhoff, R., ed. (1994) *Vitamin A in Health and Disease*, Dekker, New York
[e] Goodman, D. S. (1984) *N. Engl. J. Med.* **310**, 1023–1031
[f] Blomhoff, R., Green, M. H., Berg, T., and Norum, K. R. (1990) *Science* **250**, 399–404
[g] Tang, G., Wang, X.-D., Russell, R. M., and Krinsky, N. I. (1991) *Biochemistry* **30**, 9829–9834
[h] Redmond, T. M., Gentleman, S., Duncan, T., Yu, S., Wiggert, B., Gantt, E., and Cunningham, F. X., Jr. (2001) *J. Biol. Chem.* **276**, 6560–6565
[hi] Kiefer, C., Hessel, S., Lampert, J. M., Vogt, K., Lederer, M. O., Breithaupt, D. E., and von Lintig, J. (2001) *J. Biol. Chem.* **276**, 14110–14116
[i] Symons, M. (1996) *Trends Biochem. Sci.* **21**, 178–181
[j] Blomhoff, R., Rasmussen, M., Nilsson, A., Norum, K. R., Berg, T., Blaner, W. S., Kato, M., Mertz, J. R., Goodman, D. S., Eriksson, U., and Peterson, P. A. (1985) *J. Biol. Chem.* **260**, 13560–13565
[k] Zanotti, G., Ottonello, S., Berni, R., and Monaco, H. L. (1993) *J. Mol. Biol.* **230**, 613–624
[l] Zanotti, G., Marcello, M., Malpeli, G., Folli, C., Sartori, G., and Berni, R. (1994) *J. Biol. Chem.* **26**, 29613–29620
[m] Monaco, H. L., Rizzi, M., and Coda, A. (1995) *Science* **268**, 1039–1041
[n] Blake, C. C. F., Geisow, M. J., Swan, I. D. A., Rerat, C., and Rerat, B. (1974) *J. Mol. Biol.* **88**, 1–12
[o] Napoli, J. L., Khalil, H., and McCormick, A. M. (1982) *Biochemistry* **21**, 1942–1949
[p] Napoli, J. (1996) *FASEB J.* **10**, 993–1001
[pq] Tryggvason, K., Romert, A., and Eriksson, U. (2001) *J. Biol. Chem.* **276**, 19253–19258
[q] Barua, A. B., and Olson, J. A. (1986) *Am. J. Clin. Nutr.* **43**, 481–485
[r] Formelli, F., Barua, A. B., and Olson, J. A. (1996) *FASEB J.* **10**, 1014–1024
[rs] Cullum, M. E., and Zile, M. H. (1985) *J. Biol. Chem.* **260**, 10590–10596
[s] Rong, D., Lovey, A. J., Rosenberger, M., d'Avignon, A., Ponder, J., and Li, E. (1993) *J. Biol. Chem.* **268**, 7929–7934
[t] Quadro, L., Blaner, W. S., Salchow, D. J., Vogel, S., Piantedosi, R., Gouras, P., Freeman, S., Cosma, M. P., Colantuoni, V., and Gottesman, M. E. (1999) *EMBO J.* **18**, 4633–4644
[u] Newcomer, M. E. (1995) *FASEB J.* **9**, 229–239
[v] Ross, A. C. (1993) *FASEB J.* **7**, 317–327
[w] Thompson, J. R., Bratt, J. M., and Banaszak, L. J. (1995) *J. Mol. Biol.* **252**, 433–446
[x] Chen, X., Tordova, M., Gilliland, G. L., Wang, L., Li, Y., Yan, H., and Ji, X. (1998) *J. Mol. Biol.* **278**, 641–653
[y] Mansfield, S. G., Cammer, S., Alexander, S. C., Muehleisen, D. P., Gray, R. S., Tropsha, A., and Bollenbacher, W. E. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6825–6830
[z] Godovac-Zimmermann, J. (1988) *Trends Biochem. Sci.* **13**, 64–66
[aa] Cho, Y., Batt, C. A., and Sawyer, L. (1994) *J. Biol. Chem.* **269**, 11102–11107
[bb] Clamon, G. H., Sporn, M. B., Smith, J. M., and Saffiotti, V. (1974) *Nature (London)* **250**, 64–66
[cc] Rossa, G. C., Bendrick, C. J., and Wolf, G. (1981) *J. Biol. Chem.* **256**, 8341–8347
[dd] Chytil, F. (1996) *FASEB J.* **10**, 986–992
[ee] Buck, J., Derguini, F., Levi, E., Nakanishi, K., and Hammerling, U. (1991) *Science* **254**, 1654–1656
[ff] Grün, F., Noy, N., Hämmerling, U., and Buck, J. (1996) *J. Biol. Chem.* **271**, 16135–16138
[gg] Greenwald, D., DeWys, W., Black, G., Malone, W., Sporn, M., and Underwood, B. A. (1984) *Science* **224**, 338
[hh] Lammer, E. J., Chen, D. T., Hoar, R. M., Agnish, N. D., Benke, P. J., Braun, J. T., Curry, C. J., Fernhoff, P. M., Grix, A. W., Jr., Lott, I. T., Richard, J. M., and Sun, S. C. (1985) *N. Engl. J. Med.* **313**, 837–841

2. Properties and Functions of Carotenes

The most characteristic property of carotenoids is the striking color, most often yellow to red, which is used by birds as a sexual attractant and by plants to attract pollinators.[137,138] The associated light absorption fits these compounds for a role in photosynthetic light-harvesting,[139] in photoprotection, and in photoreception,[140] matters that are dealt with in Chapter 23. One aspect of photoprotection, which involves the **xanthophyll cycle,** is also indicated in Eq. 22-10. The cycle allows green plants to adjust to varying light intensity by altering the amount of zeaxanthin available for quenching excessive amounts of photoexcited chlorophyll (Chapter 23). Zeaxanthin undergoes epoxidation by O_2 to form antheroxanthin and violaxanthin as shown in Eq. 22-10. The process requires NADPH and reduced ferredoxin.[131] When light intensity is high the process is reversed by an ascorbate-dependent violaxanthin **de-epoxidase**.[99,128,141]

Violaxanthin also functions as a precursor to the plant hormone abscisic acid. Compare the structure of the latter (Fig. 22-4) with those of carotenoids. Oxidative cleavage of violaxanthin or related epoxy-carotenoids initiates the pathway of synthesis of this hormone.[142,143]

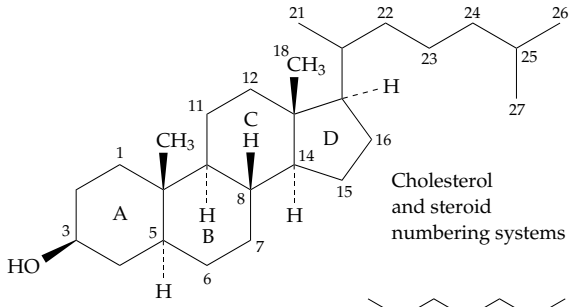
The system of conjugated double bonds responsible for carotenoid colors also helps to impart specific shapes to these largely hydrophobic molecules and ensures that they occupy the appropriate niches in the macromolecular complexes with which they associate. Information on stereochemistry is provided in a short review by Britton.[138]

β-Carotene, which can serve as an antioxidant at low oxygen pressures and can quench singlet oxygen,[144,145] has been associated with a reduced incidence of lung cancer.[137,146] While most animals do not synthesize carotenoids, they use them to make vitamin A and related retinoids and also as colorants. Yellow and red pigments of bird feathers[147] and the colors of tissues of salmon and of lobsters and other invertebrates are derived from dietary carotenoids, which are often modified further by the new host. The lobster accumulates astaxanthin, as a blue protein complex,[148] and the flamingo uses the astaxanthin of shrimp to color its feathers.[130]

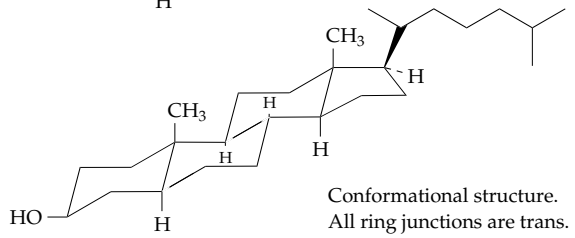
Dietary carotenes and carotenoids are absorbed and transported in the plasma of humans and animals by lipoproteins.[149] The conversion of carotenes to vitamin A (Box 22-A) provides the aldehyde **retinal** for synthesis of visual pigments (Chapter 23) and **retinoic acid**, an important regulator of gene transcription and development (Chapter 32).[150–152c] See also Section E,5.

D. Steroid Compounds

The large class of **steroids** contains a characteristic four-ring nucleus consisting of three fused six-membered rings and one five-membered ring.[153] **Cholestanol** (dihydrocholesterol) may be taken as a representative steroid alcohol or **sterol**. Most sterols, including cholestanol, contain an 8- to 10-carbon side chain at position 17. The polyprenyl origin of the side chain is suggested by the structure. Steroid compounds usually contain an oxygen atom at C-3. This atom is present in an –OH group in the sterols and



Cholesterol and steroid numbering systems



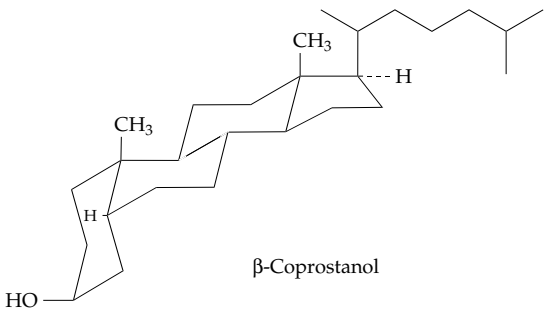
Conformational structure. All ring junctions are trans.

frequently in a carbonyl group in other steroids. Most steroids contain two axially oriented methyl groups, the "angular methyl groups,"which are attached to the ring system and numbered C-18 and C-19. In the customary projection formulas they are to be thought of as extending forward toward the viewer. In the same manner, the equatorially oriented 3-OH group of cholestanol and the side chain at C-17 also project forward toward the viewer in the projection formula.

The angular methyl groups, the 3-OH groups, and the side chain of cholestanol are all on the same side of the steroid ring in the projection formula and are all said to have a **β orientation**. Substituents projecting from the opposite side of the ring system are **α oriented**. While the methyl groups (C-18 and C-19) almost always have the β orientation, the 3-OH group has the α orientation in some sterols. Dashed lines are customarily used to connect α-oriented substituents, and solid lines are used for β-oriented substituents in structural formulas. Cholesterol is chiral and its enantiomer does not support life for *C. elegans* and presumably for other organisms.[153a]

In cholestanol the ring fusions between rings A and B, B and C, and C and D are all trans; that is, the hydrogen atoms or methyl groups attached to the

bridgehead carbon atoms project on opposite sides of the ring system. This permits all three of the six-membered rings to assume relatively unstrained chair conformations. However, the introduction of a double bond alters the shape of the molecule significantly. Thus, in cholesterol the double bond between C-5 and C-6 (Δ^5) distorts both the A and B rings from the chair conformation. In some steroid compounds the junction between rings A and B is cis. This greatly alters the overall shape of the steroid from the relatively flat one of cholestanol to one that is distinctly bent. An example is **β-coprostanol**, a product of bacterial action on cholesterol and a compound occurring in large amounts in the feces. In some sterols, notably the estrogenic hormones, ring A is completely aromatic and the methyl group at C-19 is absent.



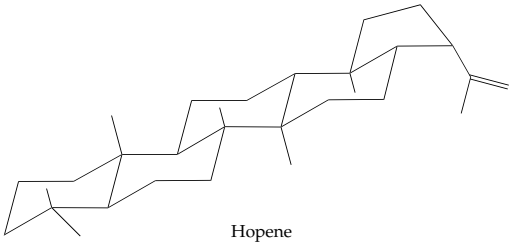
β-Coprostanol

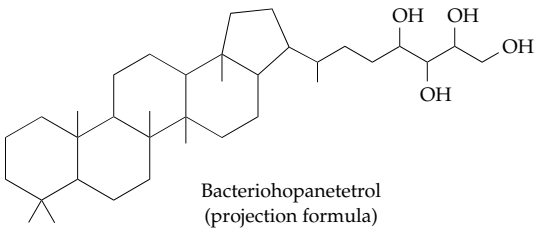
1. Biosynthesis of Sterols

Most animal steroids arise from cholesterol, which in turn is derived from squalene. This C_{30} triterpene, whose biosynthesis is described in Section B, is named after the dogfish *Squalus* in whose liver it accumulates as a result of blockage in oxidation to cholesterol. Squalene is also a prominent constituent of human skin lipids. Its conversion to **cholesterol**, which takes place in most animal tissues,[117,154–156] is initiated by a microsomal enzyme system that utilized O_2 and NADPH to form **squalene 2,3-oxide** (Fig. 22-6, step *a*). The subsequent cyclization reaction, which probably takes place through a carbocation created by attack of a proton on the oxygen atom of the epoxide ring (Fig. 22-6, step *b*), is catalyzed by the large 70- to 80-kDa **oxidosqualene cyclase**.[157–159] The enzyme from rat liver consists of 733 residues and contains a highly conserved sequence with two consecutive aspartates that are thought to be at the active site. The sequence is somewhat similar to that of prenyltransferases and sesquiterpene cyclases.[157] The cyclization step appears to require that the enzyme hold the substrate in a rigid conformation as indicated in Fig. 22-6. The flow of electrons effects the closure of all four rings. The carbocation created at C-2 of squalene (C-4 of the sterol ring that is formed) by opening of the epoxide ring reacts with electrons from the 6,7 double bond to close ring A leaving a carbocation at C-6. This in turn reacts with the 10,11 double bond leaving a carbocation at C-10, etc. At the end of this cascade a carbocation is left on C-19 of squalene, which is numbered C-20 in the incipient sterol. The closures of rings A and B both follow the Markovnikov rule by generating relatively stable tertiary carbocations. Thus, the natural chemical reactivities of the substrates are followed in these enzymatic reaction steps. However, this is not the case in the closure of ring C to form a 6-membered ring instead of a 5-membered ring. This presumably happens because the enzyme imposes the correct geometry for a 6-membered ring on the squalene and the correct stereochemistry on the ring closure.[160]

The rearrangement of this initially created C-20 carbocation to **lanosterol** (Fig. 22-6, step *c*) is also a remarkable reaction that requires the shift of a hydride ion and of two methyl groups, as indicated by the arrows in the figure. In addition, a hydrogen at C-9 (sterol numbering) is lost as a proton. Lanosterol is named for its occurrence in lanolin, the waxy fat in wool. Although the principal component of lanolin is cholesterol, lanosterol is its precursor both in sheep and in all other animals. Cholesterol is in turn the precursor to other animal sterols. The cholesterol biosynthetic pathway also provides cells with a variety of important signaling molecules.[160a]

In green plants, which contain little or no cholesterol, **cycloartenol** is the key intermediate in sterol biosynthesis.[161–162a] As indicated in Fig. 22-6, step *c′*, cycloartenol can be formed if the proton at C-9 is shifted (as a hydride ion) to displace the methyl group from C-8. A proton is lost from the adjacent methyl group to close the cyclopropane ring. There are still other ways in which squalene is cyclized,[162,163,163a] including some that incorporate nitrogen atoms and form alkaloids.[163b] One pathway leads to the **hopanoids**. These triterpene derivatives function in bacterial membranes, probably much as cholesterol does in our membranes. The three-dimensional structure of a bacterial hopene synthase is known.[164,164a] Like glucoamylase (Fig. 2-29) and farnesyl transferase, the enzyme has an $(\alpha,\alpha)_6$-barrel structure in one domain and a somewhat similar barrel in a second domain. The active site lies in a large interior cavity. The properties of the hopene synthase are similar to those of oxidosqualene synthase, and it appears to function by a similar mechanism, which resembles that of Fig. 22-6 but does not depend upon O_2. Hopene lacks polar groups, but these are provided in the hopanoids by a polyol side chain. One of these compounds, **bacteriohopanetetrol**, may be one of the most abundant compounds on earth.[160,165,166] Hopanoids appear to originate from mevalonate synthesized via the 1-deoxyxylulose pathway (Fig. 22-2). The polyol side chain is probably formed from ribose.[166]





Formation of cholesterol. The conversion of lanosterol to cholesterol requires at least 19 steps,[167,168] which are catalyzed by enzymes bound to membranes of the ER. The removal of the three methyl groups of lanosterol, the migration of the double bond within the B ring, and the saturation of the double bond in the side chain may occur in more than one sequence, two of which are indicated in Fig. 22-7. The predominant pathway in many organisms including humans is the oxidative demethylation at the C / D ring junction (C-14) by a cytochrome P450 called **lanosterol 14α-demethylase**. This single enzyme catalyzes three consecutive O_2 and NADPH-dependent reactions that convert the methyl group to hydroxymethyl, formyl, and then free formate (right side of Fig. 22-8).[169–172b] Steps *a* and *b* are typical cytochrome P450 oxygenation reactions. In step *b* a geminal diol is formed and is dehydrated to the formyl derivative. The third step is atypical. Shyadehi *et al.* proposed the sequence depicted in steps *c-f* in which an Fe(III) peroxo intermediate reacts as shown.[172] This mechanism is supported by the fact that both ^{18}O present in the formyl group and ^{18}O from $^{18}O_2$ appear in the liberated formate.

The corresponding reactions of the methyl groups at C-4 on the A ring[167,168,173] are depicted on the left side of Fig. 22-8. The 4α methyl group is first hydroxylated by a microsomal (ER) system similar to cytochrome P450 but able to accept electrons from NADH and cytochrome b_5 rather than NADPH.[173] The two-step oxidation of the resulting alcohol to a carboxylic acid is catalyzed by the same enzyme. A second enzyme catalyzes the dehydrogenation of the 3-OH group to a ketone allowing for efficient β-decarboxylation (Fig. 22-8, steps *j* and *k*).[173a] Inversion of configuration at C-4, assisted by the 3-carbonyl group (step *l*), places the second 4-methyl group in the α orientation. After reduction of the 3-carbonyl by a third enzyme the sequence is repeated on this second methyl group.

In addition to the enzymes that are embedded in the membranes of the ER, conversion of lanosterol to cholesterol depends upon soluble cytoplasmic carrier proteins.[174] See also Box 21-A. Other sterols formed in

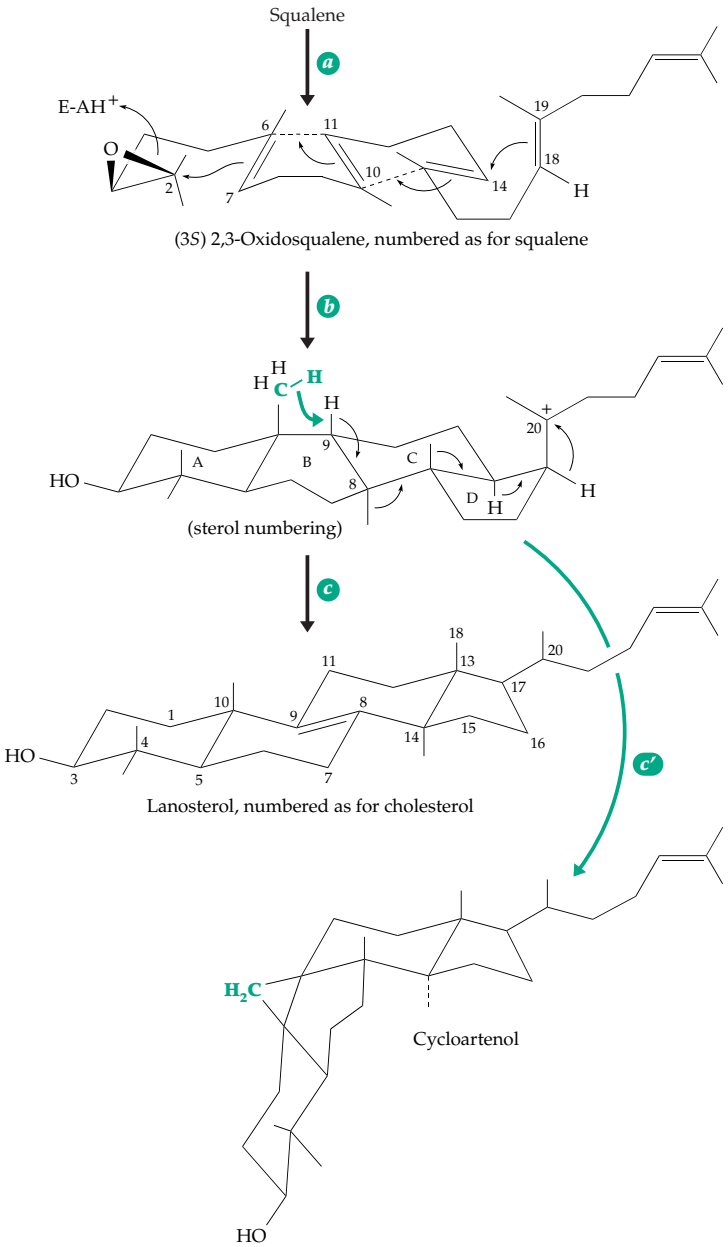
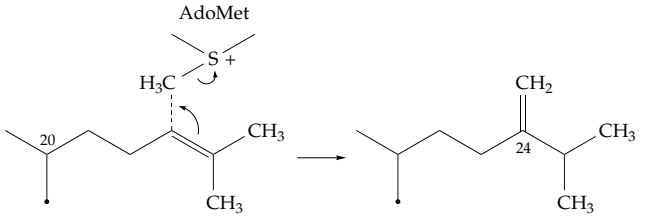


Figure 22-6 The cyclization of all-*trans* squalene to lanosterol and cycloartenol.

the animal body are **7-dehydrocholesterol**, prominent in skin and a precursor of vitamin D. Both β-cholestanol and its isomer β-coprostanol are formed by bacteria in the intestinal tract, and small amounts of cholesterol are converted to cholestanol within tissues. **Ergosterol**, the most common sterol in fungi, contains the $\Delta^{5,7}$ ring system of 7-dehydrocholesterol as well as an extra double bond in the side chain. It arises from zymosterol (Fig. 22-7).[173,174a]

As indicated previously, plant sterols are thought to be formed in most cases through cycloartenol which is often converted to **24-methylenecycloartenol**, a substance present in grapefruit peel and in many other plants. The methylene carbon is donated by *S*-adenosylmethionine (AdoMet) as shown in Eq. 22-13, which implies a transient intermediate carbocation. Saturation of the side chain and oxidative demethylation similar to that shown in Fig. 22-8[175] and introduction of a double bond[176] leads to **campesterol** (Fig. 22-9). It has the Δ^5-unsaturated ring of cholesterol but, like many other plant sterols, the side chain has one additional methyl group, which is also donated from *S*-adenosylmethionine.[161,177,178] Several more steps are



(22-13)

required to convert campesterol into the plant steroid hormone **brassinolide**.[179–180] Among higher plants, **sitosterol** and **stigmasterol** are the most common sterols. Each contains an extra ethyl group in the side chain. Sitosterol is formed by the methylation (by AdoMet) of ergosterol. For the guinea pig stigmasterol is a vitamin, the"antistiffness factor" necessary to prevent stiffening of the joints. Some other plant steroids arise without addition of the extra carbons at C-23 or C-24 but usually via a different cyclization of squalene. Of these, the cucurbitacins (Fig. 22-9) are among the bitterest substances known.[181]

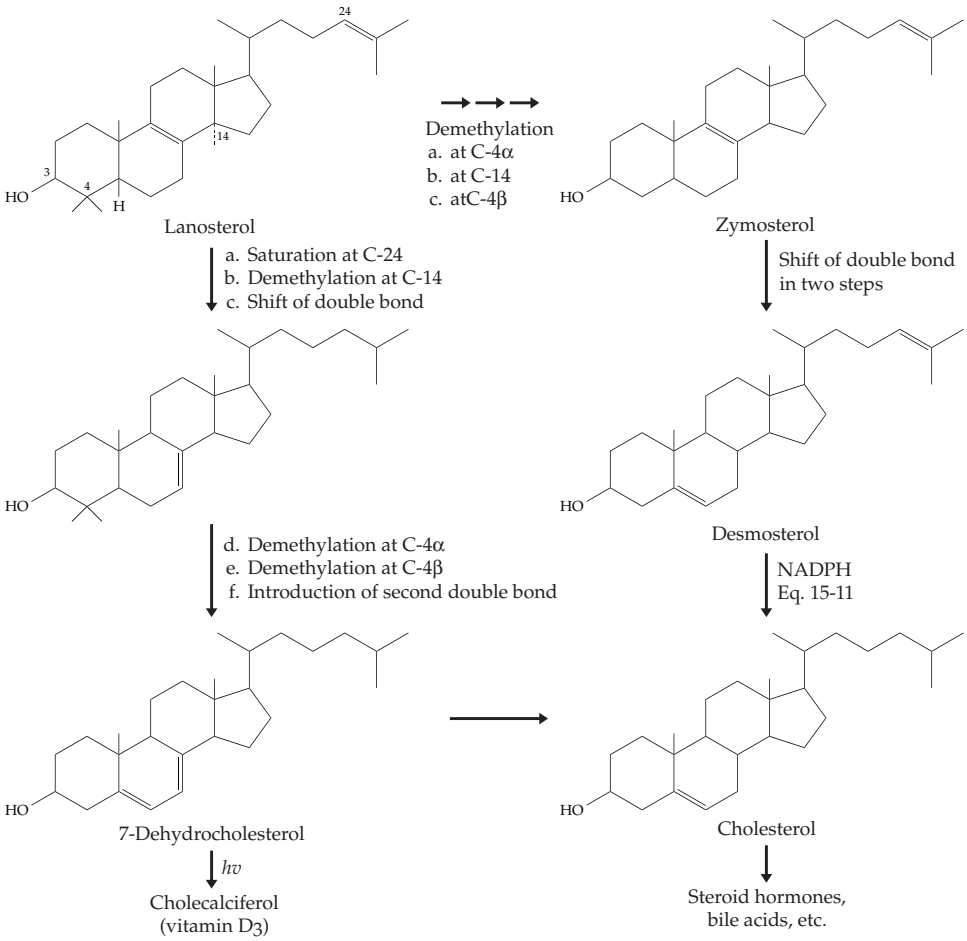


Figure 22-7 Conversion of lanosterol to cholesterol. Two of many possible sequences are shown.

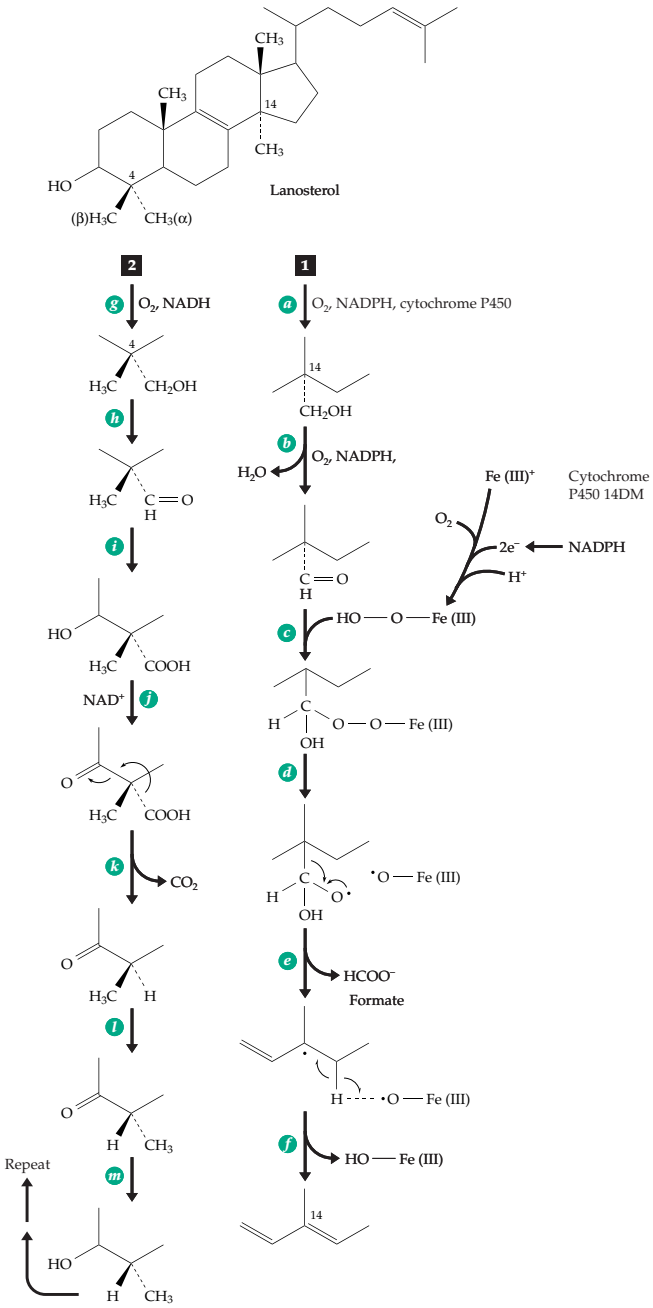


Figure 22-8 Steps in the demethylation of lanosterol. The most frequent sequence, labeled [1], begins with demethylation at C-14 by the action of a cytochrome P450 and is followed [2] by the successive demethylation of the α-CH_3 and β-CH_3 at C-4 by an NADH-dependent oxygenase.

2. Metabolism of Cholesterol in the Human Body

Cholesterol is both absorbed from the intestinal tract and synthesized from acetate via squalene, principally in the liver. The quantities produced are substantial. Daily biosynthesis is ~600 mg, and dietary uptake may supply another 300 mg.[182] Not only is there a large amount of cholesterol in the brain and other nervous tissues but also about 1.7 g of cholesterol per liter is present in blood plasma, about two-thirds of it being esterified principally to unsaturated fatty acids. The cholesterol content of plasma varies greatly with diet, age, and sex. By age 55 it averages 2.5 g / liter and may be considerably higher. Women up to the age of menopause have distinctly lower blood cholesterol than do men. Cholesterol regulates its own abundance by a variety of feedback mechanisms.[183,184] These include inhibition of the synthesis by means of reduced activities (step *a* of Fig. 22-1) of HMG-CoA reductase, farnesyl diphosphate synthase (step *g* of Fig. 22-1), and squalene synthase. All of these reactions are essential steps in cholesterol synthesis.[16,185] On the other hand, cholesterol induces an increase in acyl-CoA:cholesterol acyltransferase.

Dietary cholesterol, together with triacylglycerols, is absorbed from the intestinal tract and enters the large lipoprotein chylomicrons (see Fig. 21-1). Absorption of cholesterol is incomplete, usually amounting to less than 40% of that in the diet. Absorption requires bile salts and is influenced by other factors.[186] As it is needed cholesterol is taken from the plasma lipoproteins into cells by endocytosis. Much of the newly absorbed cholesterol is taken up by the liver. The liver also secretes cholesterol, in the form of esters with fatty acids, into the bloodstream.

Cholesterol is synthesized in the ER and other internal membranes by most cells of the body.[187,188] Newly formed cholesterol is sorted from the ER into the various membranes of the cell, the greatest abundance being in plasma membranes where cholesterol plays an essential role in decreasing fluidity. Cholesterol also aggregates with sphingolipids to form rigid lipid "rafts" floating in the plasma membrane. These rafts are thought to have important functions in signaling, in distribution of lipid materials,[188a] and in influencing protein translocation.[188b] Caveolae in cell surfaces may also arise from cholesterol-rich rafts.[188c] However, cholesterol must also be able to move out of the internal membranes back into the interior ER of the cell to provide for homeostasis and to allow formation of cholesteryl esters for transport, bile acids for excretion by liver, and the steroid hormones.[183,184,189] Movement between organelles occurs with the aid of sterol carrier proteins.[190–190c]

Liver and some intestinal cells export cholesterol into the bloodstream, together with triacylglycerols and phospholipids in the form of VLDL particles, for uptake by other tissues (see Fig. 21-1). Cholesteryl esters are formed in the ER by **lecithin:cholesterol acyltransferase** (LCAT), an enzyme that transfers the central acyl group from phosphatidylcholine to the hydroxyl group of cholesterol.[191,191a] This enzyme is also secreted by the liver and acts on free cholesterol in lipoproteins.[192] Tissue acyltransferases also form cholesteryl esters from fatty acyl-CoAs.[192a]

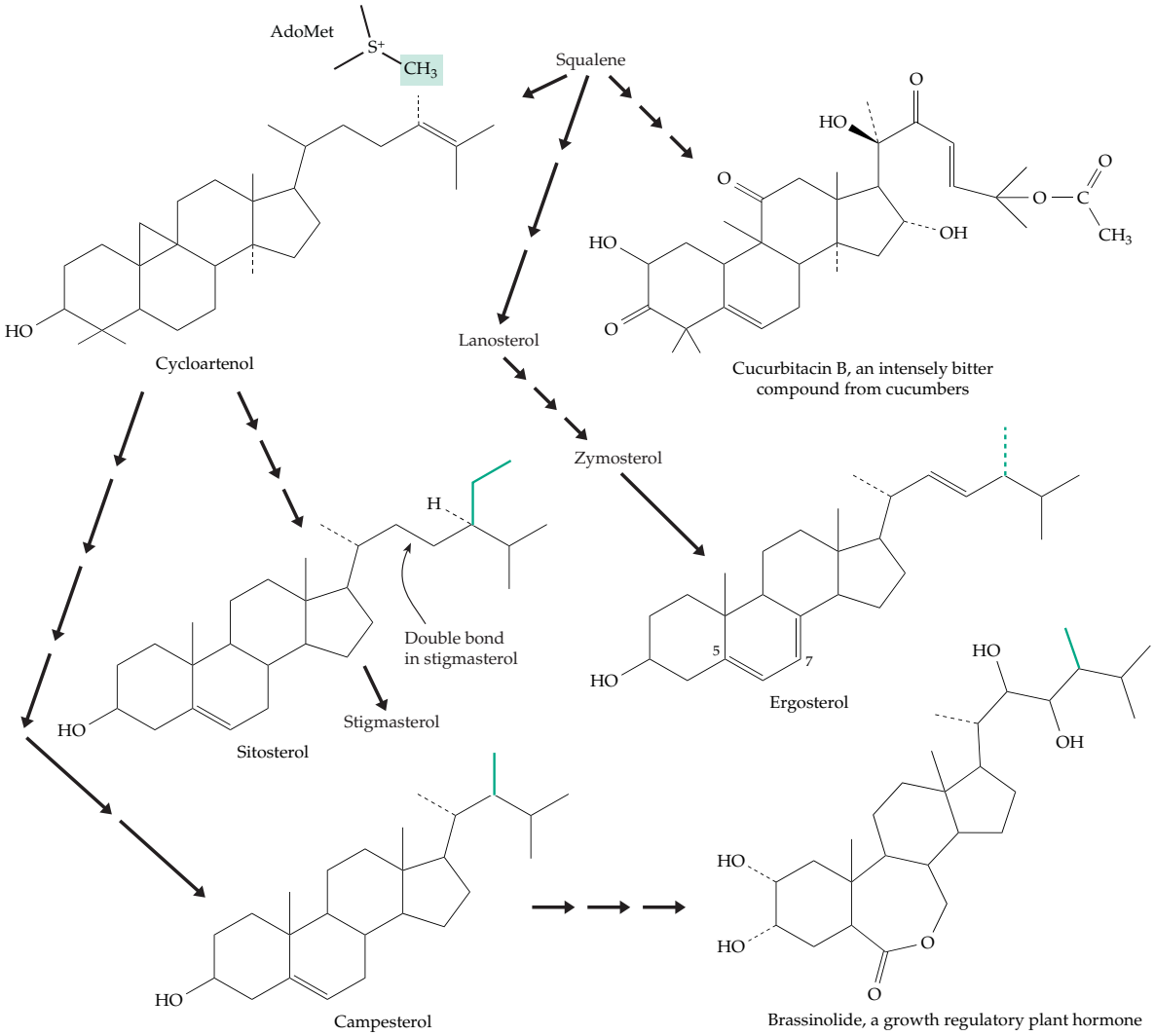


Figure 22-9 Structures and routes of biosynthesis for a few plant steroids.

In the brain a complex of cholesterol with apolipoprotein E (Table 21-2) promotes the formation of new synapses (Chapter 30). Synthesis of cholesterol for this purpose appears to occur within glial cells.[192b,c]

Serum cholesterol. Most cholesterol is carried in the blood by low density lipoprotein (LDL, Tables 21-1, 21-2), which delivers the cholesteryl esters directly to cells that need cholesterol. Both a 74-kDa **cholesteryl ester transfer protein**[193–195a] and a **phospholipid transfer protein**[196,196a] are also involved in this process. **Cholesterol esterases**, which release free cholesterol, may act both on lipoproteins and on pancreatic secretions.[197–199]

The LDL–cholesterol complex binds to LDL receptors on the cell surfaces.[167,168,200–202] These receptors are specific for apolipoprotein B-100 present in the LDL. The occupied LDL–receptor complexes are taken up by endocytosis through coated pits; the apolipoproteins are degraded in lysosomes, while the cholesteryl esters are released and cleaved by a specific **lysosomal acid lipase**[203,204] to form free cholesterol.

While the primary role of LDL appears to be the transport of esterified cholesterol to tissues, the high density lipoproteins (HDL) carry excess cholesterol away from most tissues to the liver.[205–207] The apoA-I present in the HDL particle not only binds lipid but activates LCAT, which catalyzes formation of cholesteryl esters which migrate into the interior of the HDL and are carried to the liver.

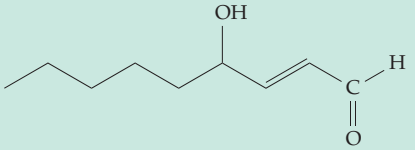
Unlike other lipoproteins, HDL particles are assembled outside of cells from lipids and proteins, some of which may be donated from chylomicrons (see Fig. 21-1) or other lipoprotein particles. HDL has a higher protein content than other lipoproteins and is more heterogeneous. The major HDL protein is apolipoprotein A-I, but many HDL particles also contain A-II,[205,208–210] and apolipoproteins A-IV, D, and E may also be present. A low plasma level of HDL cholesterol is associated with a high risk of atherosclerosis.[205,207]

BOX 22-B ATHEROSCLEROSIS

Our most common lethal disease is atherosclerosis, which causes constriction and blockage of arteries of the heart, brain, and other organs. In the United States, Europe, and Japan half of all deaths can be attributed to this ailment.[a,b] There seems to be a variety of causes. However, there is agreement that the disease begins with injury to the endothelial cells that form the inner lining of the arteries.[a,c,d] This is followed by the aggregation of blood platelets at the sites of injury and infiltration of smooth muscle cells, which may be attracted by 12-hydroxyeicosotetraenoic acid and other chemoattractants formed by activated platelets.[c] **"Foam cells"** laden with cholesterol and other lipids appear, and the lesions enlarge to become the characteristic plaques (**atheromas**).

The best understood cause of atherosclerosis is the genetic defect **familial hypercholesterolemia**, an autosomal dominant trait carried by one person in 500 all over the world.[e] Males with the defective gene tend to develop atherosclerosis when they are 35–50 years of age. The approximately one in a million persons *homozygous* for the trait develop coronary heart disease in their teens or earlier. Cultured fibroblasts from these patients have 40- to 60-fold higher levels of HMG-CoA reductase (Eq. 15-9) than are present normally, and the rate of cholesterol synthesis is increased greatly. The LDL level is very high and, as shown by Brown and Goldstein,[f–i] the LDL receptor gene is defective. Genetic defects associated with a low HDL level are also associated with atherosclerosis[b,j–l] as is a genetic variant of the metalloproteinase **stromelysin**.[m]

Other factors favoring development of atherosclerosis include hypertension and smoking. Chickens infected with a herpes virus (Marek disease virus) develop the disease after infection, and it is possible that artery damage in humans can also be caused by viruses[n] or bacterial infections.[o] In recent years it has been established that oxidative modification of the phospholipids in LDL induces the uptake of LDL by scavenger receptors of macrophages. This appears to trigger the development of foam cells and atherosclerotic plaques.[c,p,q] The initial damage is thought to be caused by lipid peroxides in the diet or generated by lipoxygenases in platelets and other cells.[c,p,r] Unsaturated fatty acids in lipoproteins can undergo oxidation (Chapter 21), especially in the presence of Cu^{2+} ions,[s,t] to yield malondialdehyde, 4-hydroxynonenal (Eq. 21-15), and other reactive compounds, which may damage the lipoproteins and cause them to have too high an affinity for their receptors in the smooth muscle cells of artery walls.[p,u] The 17β-hydroperoxy derivative of cholesterol has also been found in atherosclerotic lesions and may account for some of the toxicity of oxidized LDL.[r] Ascorbic acid may help to prevent formation of these oxidation products.[p,q,v] Chlorinated sterols may also be produced by the myeloperoxidase of the phagocytic macrophages that are abundant in atherosclerotic plaque.[w] Trans fatty acids, which are abundant in some margarines, and other hydrogenated fats raise both cholesterol and LDL levels.[x] Another cause of artery disease may be the presence of excessive homocysteine,[y,z] which can accumulate as a result of marginal deficiencies of folate, vitamin B_6, or vitamin B_{12}.



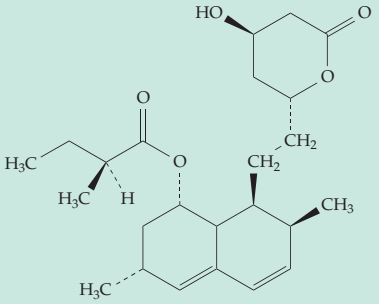
4-Hydroxy-2-nonenal

What can be done to prevent atherosclerosis? For persons with a high LDL level there is little doubt that a decreased dietary intake of cholesterol and a decrease in caloric intake are helpful. While such dietary restriction may be beneficial to the entire population, controlled studies of the effect of dietary modification on atherosclerosis have been disappointing and confusing.[aa] A diet that is unhealthy for some may be healthy for others. For example, an 88-year old man who ate 25 eggs a day for many years had a normal plasma cholesterol level of 150–200 mg / deciliter (3.9–5.2 mM)![bb] Comparisons of diets rich in unsaturated fatty acids, palmitic acid, or stearic acid have also been confusing.[cc,cd,dd] Can it be true that palmitic acid from tropical oils and other plant sources promotes atherogenesis, but that both unsaturated fatty acids and stearic acid from animal fats are less dangerous?

One of the best therapeutic approaches may be to prevent absorption of cholesterol from the intestines by inclusion of a higher fiber content in the diet.[ee] Supplementation with a cholesterol-binding resin may provide additional protection. Plant sterols also interfere with cholesterol absorption. Incorporation of esters of **sitostanol** into margarine provides an easy method of administration.[ff] Supplemental vitamin E may also be of value.[q] Another effective approach is to decrease the rate of cholesterol synthesis by administration of drugs that inhibit the synthesis of cholesterol. Inhibitors of HMG-CoA reductase,[gg,hh] (e.g., vaLostatin) isopentenyl-*PP* isomerase, squalene synthase (e.g.,

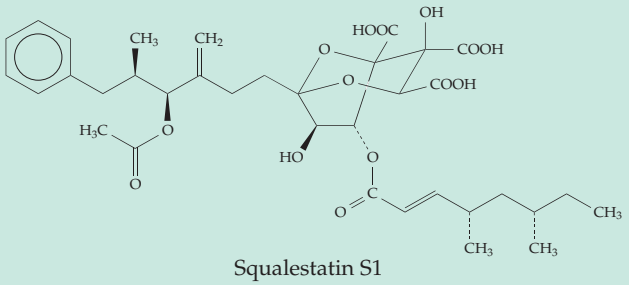
BOX 22-B ATHEROSCLEROSIS (continued)

squalestatin S1),[ii] and other enzymes in the biosynthetic pathway are targets for drug treatment.



Lovastatin
(Mevinolin)

Questions of possible long-term toxicity remain. Since 1976 there has been a greater than 25% decrease in the incidence of ischemic heart disease in the United States.[jj] Increased exercise, a decreased severity of influenza epidemics, and fluoridation of water[kk] have been suggested as explanations.



Squalestatin S1

[a] Ross, R. (1993) *Nature (London)* **362**, 801–809
[b] Krieger, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4077–4080
[c] Yagi, K. (1986) *Trends Biochem. Sci.* **11**, 18–19
[d] Breslow, J. L. (1996) *Science* **272**, 685–688
[e] Goldstein, J. L., Hobbs, H. H., and Brown, M. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1981–2030, McGraw-Hill, New York
[f] Brown, M. S., and Goldstein, J. L. (1974) *Sci. Am.* **251**(Nov), 58–66
[g] Brown, M. S., Kovanen, P. T., and Goldstein, J. L. (1981) *Science* **212**, 628–635
[h] Brown, M. S., and Goldstein, J. L. (1986) *Science* **232**, 34–47
[i] Motulsky, A. G. (1986) *Science* **231**, 126–128
[j] Breslow, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8314–8318
[k] Bergeron, J., Frank, P. G., Scales, D., Meng, Q.-H., Castro, G., and Marcel, Y. L. (1995) *J. Biol. Chem.* **270**, 27429–27438
[l] Breslow, J. L. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2031–2052, McGraw-Hill, New York
[m] Ye, S., Eriksson, P., Hamsten, A., Kurkinen, M., Humphries, S. E., and Henney, A. M. (1996) *J. Biol. Chem.* **271**, 13055–13060
[n] Benditt, E. P., Barrett, T., and McDougall, J. K. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6386–6389
[o] Gura, T. (1998) *Science* **281**, 35–37
[p] Holvoet, P., and Collen, D. (1994) *FASEB J.* **8**, 1279–1284
[q] Steinberg, D. (1997) *J. Biol. Chem.* **272**, 20963–20966
[r] Chisolm, G. M., Ma, G., Irwin, K. C., Martin, L. L., Gunderson, K. G., Linberg, L. F., Morel, D. W., and DiCorleto, P. E. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11452–11456
[s] Lynch, S. M., and Frei, B. (1995) *J. Biol. Chem.* **270**, 5158–5163
[t] Leeuwenburgh, C., Rasmussen, J. E., Hsu, F. F., Mueller, D. M., Pennathur, S., and Heinecke, J. W. (1997) *J. Biol. Chem.* **272**, 3520–3526
[u] Itabe, H., Yamamoto, H., Suzuki, M., Kawai, Y., Nakagawa, Y., Suzuki, A., Imanaka, T., and Takano, T. (1996) *J. Biol. Chem.* **271**, 33208–33217
[v] Retsky, K. L., Freeman, M. W., and Frei, B. (1993) *J. Biol. Chem.* **268**, 1304–1309
[w] Hazen, S. L., Hsu, F. F., Duffin, K., and Heinecke, J. W. (1996) *J. Biol. Chem.* **271**, 23080–23088
[x] Mensink, R. P., and Katan, M. B. (1990) *N. Engl. J. Med.* **323**, 439–445
[y] Chen, Z., Crippen, K., Gulati, S., and Banerjee, R. (1994) *J. Biol. Chem.* **269**, 27193–27197
[z] Kokame, K., Kato, H., and Miyata, T. (1996) *J. Biol. Chem.* **271**, 29659–29665
[aa] Kolata, G. (1985) *Science* **227**, 40–41
[bb] Kern, F., Jr. (1991) *N. Engl. J. Med.* **324**, 896–899
[cc] Bonanome, A., and Grundy, S. M. (1988) *N. Engl. J. Med.* **318**, 1244–1248
[cd] Merkel, M., Velez-Carrasco, W., Hudgins, L. C., and Breslow, J. L. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13294 – 13299
[dd] Hayes, K. C., and Khosla, P. (1992) *FASEB J.* **6**, 2600–2607
[ee] Jenkins, D. J. A., Wolever, T. M. S., Rao, A. V., Hegele, R. A., Mitchell, S. J., Ransom, T. P. P., Boctor, D. L., Spadafora, P. J., Jenkins, A. L., Mehling, C., Relle, L. K., Connelly, P. W., Story, J. A., Furumoto, E. J., Corey, P., and Würsch, P. (1993) *N. Engl. J. Med.* **329**, 21–26
[ff] Pedersen, T. R. (1995) *N. Engl. J. Med.* **333**, 1350–1351
[gg] Grundy, S. M. (1988) *N. Engl. J. Med.* **319**, 24–33
[hh] Brown, M. S., and Goldstein, J. L. (1996) *Science* **272**, 629
[ii] Kelly, M. J., and Roberts, S. M. (1995) *Nature (London)* **373**, 192–193
[jj] Stallones, R. A. (1980) *Sci. Am.* **243**(Nov), 53–59
[kk] Taves, D. R. (1978) *Nature (London)* **272**, 361–362

The LDL and related receptors. The LDL receptor gene extends over 50 kb of DNA and appears to be a mosaic of exons shared by several other genes that seem to have nothing to do with cholesterol metabolism.[201,211] The 839-residue receptor protein consists of five structural domains. The N-terminal domain that binds the LDL consists of seven repeated ~40-residue cysteine-rich modules.[212,213] This is followed by a large domain that is homologous to a precursor of the epidermal growth factor, a 48-residue domain containing many *O*-glycosylated serine and threonine residues, a 27-residue hydrophobic region that spans the membrane, and a 50-residue C-terminal cytoplasmic domain.[200,214–216] Synthesis of LDL receptors is regu-

lated by a feedback mechanism, the cholesterol released within cells inhibiting the synthesis of new receptors.

As mentioned in Chapter 21, there are several related receptors with similar structures. Two of them have a specificity for apolipoprotein E and can accept remnants of VLDL particles and chylomicrons.[216–220] The LDL receptor-related protein is a longer-chain receptor.[216,221] LDL particles, especially when present in excess or when they contain oxidized lipoproteins, may be taken up by endocytosis into macrophages with the aid of the quite different **scavenger receptors**.[221–225] The uptake of oxidized lipoproteins by these receptors may be a major factor in promoting development of atherosclerosis (Box 22-B). On the other hand, **scavenger receptor SR-B1**, which is also present in liver cells, was recently identified as the receptor for HDL and essential to the "reverse cholesterol transport" that removes excess cholesterol for excretion in the bile.[213,213a]

Abnormalities of cholesterol metabolism. A variety of genetic problems have been identified, many of them being associated with atherosclerosis (Box 22-B).[218,226–230] In the commonest form of **familial hypercholesterolemia** a mutation in the LDL receptor protein prevents normal synthesis, binding, clustering into coated pits, or uptake of LDL and its cholesteryl esters. Over 600 mutations have been identified.[229,229a] Some of these are present in a Ca^{2+}-binding region of the 5th cysteine-rich module.[230] Other disorders that raise the plasma LDL level include a defective apoB-100 protein (see p. 1182)[230a] and deficiency of a protein that seems to be involved in incorportation of LDL receptors into clathrin-coated pits during endocytosis or in receptor recycling.[229a] In a **cholesteryl ester storage disease** the lysosomal lipase is lacking. Absence of lecithin:cholesterol acyltransferase from plasma causes corneal opacity and often kidney failure.[231]

In the very rare and fatal **Niemann–Pick C1** disease lysosomes in cells of the central nervous system and the viscera accumulate LDL-derived cholesterol. Study of the DNA of patients led to discovery of a 1278-residue integral membrane protein, which may be required for the Golgi-mediated transport of unesterified cholesterol from lysosomes to the ER.[189,232–234c]

Some people with elevated lipoprotein levels have VLDL that migrates on electrophoresis in the β band rather than the pre-β band (see Box 2-A). The presence of the β-VLDL is associated with a high incidence of artery disease,[218] which is most likely to develop in persons homozygous for a genetic variant of apolipoprotein E. The problem may arise because apo-E is required for receptor-mediated uptake of VLDL, which interacts both with tissue LDL receptors and with hepatic apo-E receptors. Genes for many of the apolipoproteins are polymorphic, and numerous alleles are present in a normal population.[218,235] In the rare **Tangier disease** apolipoprotein A-I is catabolized too rapidly, and the HDL level is depressed, resulting in accumulation of cholesterol esters in macrophages, Schwann cells, and smooth muscles. Orange-yellow enlarged tonsils are characteristic of the disease.[236] An ABC type transporter that allows cholesterol to leave cells is defective.[236a,b,c] The faulty component is known as the **cholesterol-efflux regulatory protein**. Another ABC transporter, apparently encoded by a pair of genes, which are expressed predominantly in liver and intestinal cells, prevents excessive accumulation of plant sterols such as sitosterol (Fig. 22-9).[236d,e]

Bile acids. Among the metabolites of cholesterol the bile acids (Fig. 22-10)[182,237,238] are quantitatively the most important (100–400 mg / day). These powerful emulsifying agents are formed in the liver and flow into the bile duct and the small intestine. A large fraction is later reabsorbed in the duodenum and is returned to the liver for reuse.[238a] Formation of the bile acids involves the removal of the double bond of cholesterol, inversion at C-3 to give a 3α-hydroxyl group, followed by hydroxylation and oxidation of the side chain.[238b–f] The principal human bile acids are **cholic acid** and **chenodeoxycholic acid** (Fig. 22-10). The free bile acids are then converted to CoA derivatives and conjugated with glycine and taurine to form **bile salts**, such as **glycocholic** and **taurocholic acids**.[238d] Several rare lipid-storage diseases are associated with defective bile acid formation.[239–241] In one of these, **cerebrotendinous xanthomatosis**, cholestanol is deposited both in tendons throughout the body and in the brain. Oxidation of the cholesterol side chain is incomplete with excretion, as glucuronides, of large amounts of bile alcohols (precursors to the bile acids). The synthesis of bile acids is regulated by feedback inhibition by the bile acids, but in this disease the inhibition is absent and the rates of both cholesterol biosynthesis and oxidation are increased. The problem is not one of storage of cholesterol but of the cholestanol that arises as a minor product of the pathway. A proper ratio of bile salt, phosphatidylcholine, and cholesterol in the bile is important to prevention of **cholesterol gallstones**.[207]

A variety of other oxidative modifications of cholesterol take place in tissues to give small amounts of diols.[242] Hydroperoxides of cholesterol may also be formed.[243] Some of the products are probably toxic, but others may be essential. One of these is 26-hydroxycholesterol, a minor component of plasma but a major neonatal excretion product.[244]

The body contains sulfate esters of cholesterol and other sterols,[245] sometimes in quite high concentrations relative to those of unesterified steroids. These esters are presumably soluble transport forms. They

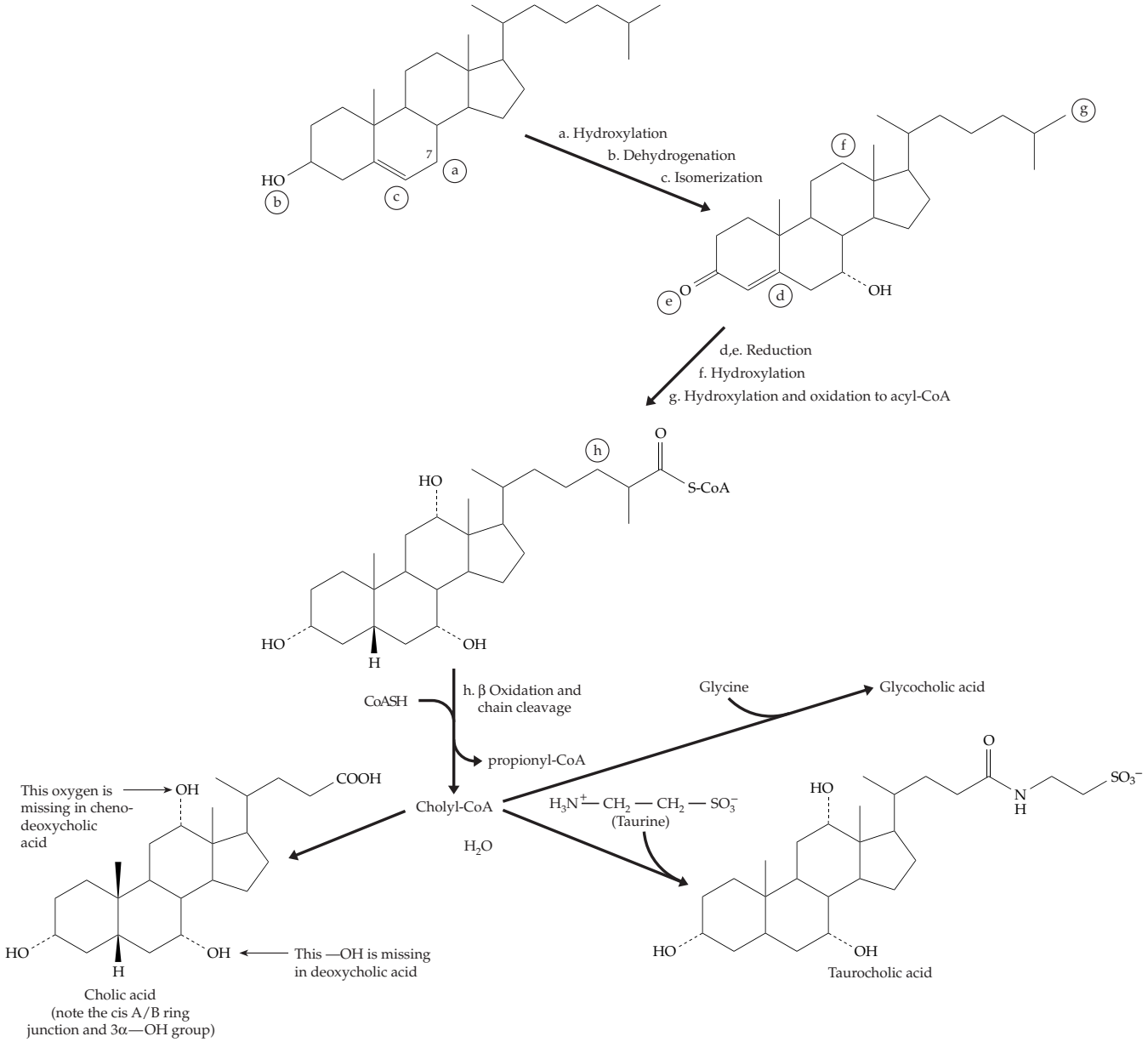


Figure 22-10 Formation of the bile acids.

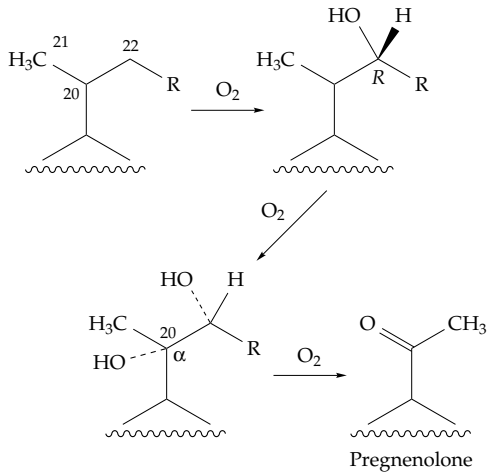
are hydrolyzed by a steroid sulfatase present within cells and whose absence causes **X-linked ichthyosis**, one of the commoner forms of scaly skin. The defect may also lead to corneal opacity but appears not to cause any other problems.[246] Glucuronides of sterols are another group of water-soluble metabolites.[247] We tend to think of cholesterol as an inert structural component of cell membranes. However, it has been found in ester linkage to a signaling protein of the "hedgehog" family, important in embryo development.[248]

E. The Steroid Hormones

In the animal body three important groups of hormones are formed by the metabolism of cholesterol: the **progestins**, the **sex hormones**, and the **adrenal cortical hormones**.[249] Their synthesis occurs principally in mitochondria of the adrenal cortex and the gonads. Steroid hormone synthesis is regulated by hormones, such as **corticotropin** (ACTH), from the anterior pituitary[250] (see Chapter 30) and is also dependent upon the recently discovered **steroidogenic acute regulatory protein,** which in some way promotes the movement of stored cholesterol into mitochondrial membranes.[251,252] Some major pathways of

biosynthesis are outlined in Fig. 22-11. The side chain is shortened to two carbon atoms through hydroxylation and oxidative cleavage to give the key intermediate **pregnenolone**. The reaction is initiated by the mitochondrial **cytochrome P450$_{ssc}$**, which receives electrons from NADPH and adrenodoxin (Chapter 16).[253] Hydroxylation occurs sequentially on C-22 and C-20 (Eq. 22-14). The chain cleavage is catalyzed by the same enzyme, an overall 6-electron oxidation occurring in three O_2-dependent steps. Dehydrogenation of the 3-OH group of pregnenolone to C=O is followed by a shift in the double bond, the oxosteroid isomerase reaction (Eq. 13-30, step *b*). In bacteria these two steps are catalyzed by different proteins, but a single human 3β-hydroxysteroid / Δ^5-Δ^4 isomerase catalyzes both reactions.[254,255] The product is the α,β-unsaturated ketone **progesterone**.

Most steroid hormones exist in part as sulfate esters and may also become esterified with fatty acids.[256] The fatty acid esters may have relatively long lives within tissues.[256] A special sex hormone-binding globulin transports sex hormones in the blood and regulates their access to target cells.[256a,b]

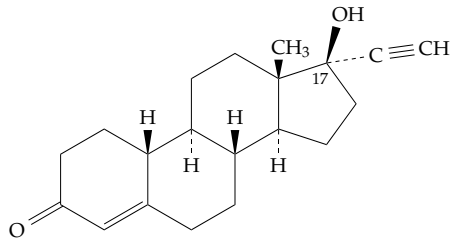


(22-14)

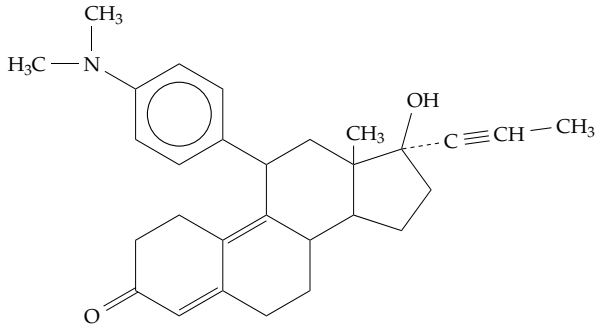
1. Progestins

Progesterone is the principal hormone of the **corpus luteum**, the endocrine gland that develops in the ovarian follicle after release of an ovum. Progesterone is also formed in the adrenals, testes, and placenta. It is metabolized rapidly, largely by reduction to alcohols, which may then be conjugated and excreted as glucuronides (see Eq. 20-16).[247] Reduction of the double bond within the A ring of progesterone leads to complete loss of activity, an indication that the α,β-unsaturated ketone group may play an essential role in the action of the hormone. Progesterone has a special role in the maintenance of pregnancy, and together with the estrogenic hormones it regulates the menstrual cycle. It is also essential for reproduction in lower animals such as birds and amphibians.[257,258]

The synthetic progesterone agonist **norethindrone** is widely used as one component of contraceptive pills. Having an opposite effect is **mifepristone** (also known as RU 486), a powerful antagonist of both progesterone and glucocorticoids. It is capable of inducing abortion and has other medical uses.[258–260] It is an effective emergency contraceptive agent that prevents implantation of an embryo.[261,262]



Norethindrone
(19-Nor-17α-ethynyltestosterone)



Mifepristone (RU 486)

2. Adrenal Cortical Steroids

Within the adrenal cortex (the outer portion of the adrenal glands) progesterone is converted into two groups of hormones of which **cortisol** and **aldosterone** are representative.[263] Two different cytochrome P450 hydroxylases, found in the ER and specific for C-21 and C-17α, respectively, together with a mitochondrial cytochrome P450 specific for C-11β (Eq. 18-55) participate in formation of cortisol.[264] Two of the same enzymes together with additional hydroxylases are required to form aldosterone.

Absence of the C-21 hydroxylase is one of the commonest of hereditary metabolic defects and is one of several enzymatic deficiencies that lead to **congenital adrenal hyperplasia**.[265–269] Cortisol, the synthesis of which is controlled by ACTH, is secreted by the adrenals in amounts of 15–30 mg daily in an adult. The hormone, which is essential to life, circulates in the blood, largely bound to the plasma protein **transcortin**. Cortisol, in turn, exerts feedback inhibi-

tion on ACTH production, and it is this feedback loop that fails when the C-21 hydroxylase is missing. Normally the circulating cortisol binds to receptors in both the pituitary and the hypothalamus of the brain to inhibit release of both ACTH and its hypothalamic releasing hormone (corticotropin-releasing hormone, CRF; see also Chapter 30). Girls are especially seriously affected by adrenal hyperplasia because, as ACTH production increases, the adrenal glands swell and produce an excessive amount of androgens. This occurs during the prenatal period of androgen release that initiates sexual differentiation. Girls with this deficiency are born with a masculine appearance of their external genitalia and continue to develop a masculine appearance. For reasons that are not clear the gene for the 21-hydroxylase is located within the HLA region (Chapter 31) of human chromosome 6.

Cortisol is a glucocorticoid which promotes gluconeogenesis and the accumulation of glycogen in the liver (Chapter 17). While it induces increased protein synthesis in the liver, it inhibits protein synthesis in muscle and many other tissues and leads to breakdown of fats to free fatty acids in adipose tissue.

Cortisol and its close relative **cortisone**, which was discovered by Kendall and Reichstein in the late 1940s, are probably best known for their anti-inflammatory effect in the body.[268,270] The effect depends upon several factors including inhibition of protein synthesis by fibroblasts, neutrophils, and antibody-forming cells. Migration of neutrophils into the inflamed area is also suppressed. Because of this action cortisone and synthetic analogs such as prednisolone and dexamethasone are among the modern "wonder drugs." They are used in controlling acute attacks of arthritis and of serious inflammations of the eyes and other organs. However, prolonged therapy can have serious side effects including decreased resistance to infections, wasting of muscle, and resorption of bone. The last results from a specific inhibition of calcium absorption from the gastrointestinal tract, glucocorticoids being antagonistic to the action of vitamin D (Box 22-C).

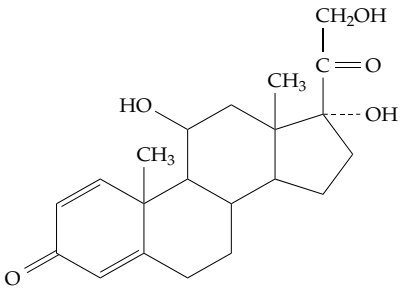
Aldosterone, which is classified as a **mineralocorticoid**, is produced under the control of the **renin–angiotensin** hormone system (Box 22-D), which is stimulated when sodium ion receptors in the kidneys detect an imbalance. It is synthesized in vascular cells of the body as well as in the adrenal cortex.[271] Aldosterone promotes the resorption of sodium ions in the kidney tubules and thus regulates water and electrolyte metabolism.[267,268] Glucocorticoids also have weak mineralocorticoid activity, and most patients with adrenocortical insufficiency (**Addison's disease**) can be maintained with glucocorticoids alone if their salt intake is adequate. Addison's disease develops when the adrenals are destroyed, most often by autoimmune disease or by tuberculosis.

3. Androgens

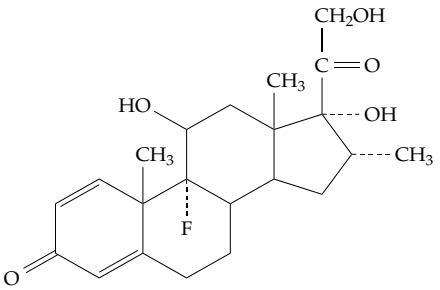
The principal **androgenic** or male sex hormone is **testosterone** formed from pregnenolone through removal of the side chain at C-17. The first step in the conversion is 17α-hydroxylation by a cytochrome P450 which may also mediate the further oxidative cleavage to **Δ^4-androstenedione** (Fig. 22-11). Reduction of the 17-carbonyl group forms testosterone. About 6–10 mg are produced daily in men, and smaller amounts (~0.4 mg) are synthesized in women. Testosterone is carried in the blood as a complex with a β-globulin and affects a variety of target tissues including the reproductive organs. Another striking effect is stimulation of the growth of the beard. Testosterone also causes premature death of follicles of head hair in genetically susceptible individuals. However, a bald man can usually grow a full beard, and follicles of the beard type, when transplanted to the head, remain immune to the action of androgen. No one knows what regulatory differences explain this fact. Baldness might be cured by use of suitable antagonists of the androgenic hormones, but the beard might fall out and sexual interest could be lost.

Androgen synthesis in the human male fetus begins at the age of about 70 days when the testes enlarge and go through an important period of activity that begins the conversion of the infant body to a male type. Other bursts of testosterone synthesis occur during infancy, but there is little further synthesis until the onset of puberty.[268,272]

Within many target tissues testosterone is converted by an NADPH-dependent 5α-reductase into **5α-dihydrotestosterone**. That this transformation is important is shown by the fact that absence of one of the two 5α-reductase isomers causes a form of **pseudohermaphroditism** in which male children are often mistaken as female and raised as girls. However, at puberty they become unmistakably male.[268,273–275] Many other metabolites of testosterone are known. These include the

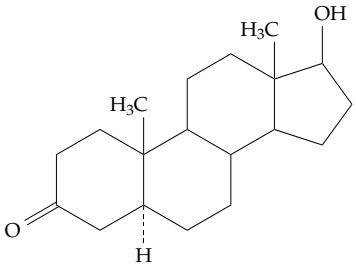


Prednisolone

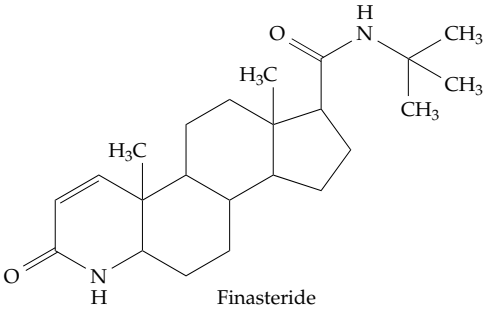


Dexamethasone

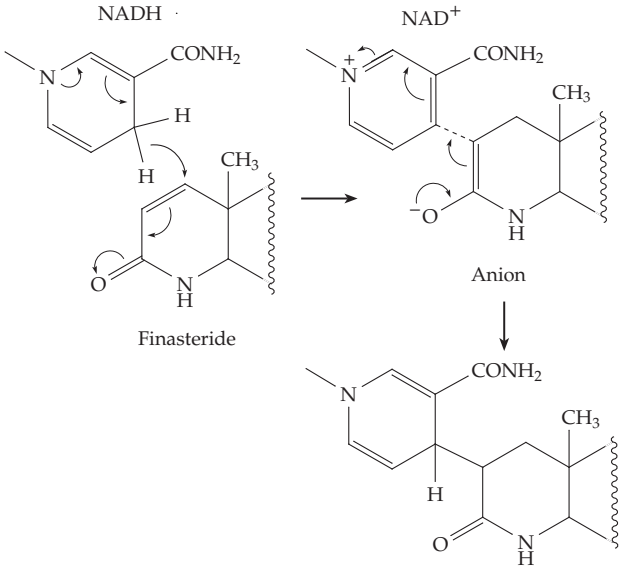
isomeric 5β-dihydrotestosterone and 5α-androstane-3β, 17β-diol which arises by reduction of the carbonyl group of 5α-dihydrotestosterone. Testosterone and dihydroxytestosterone have distinct roles in the body.



5α-Dihydrotestosterone



Finasteride



(22-15)

For example, testostosterone is required for sperm cell formation, voice deepening, and growth of pubic hair while dihydroxytestosterone stimulates development of the prostate gland and male pattern hair growth.[276]

Males with deficiency of the 5α-reductase isoenzyme do not develop acne, male pattern baldness, or enlarged prostates.[274] The last fact was some of the impetus for development of the steroid 5α-reductase inhibitor **finasteride,** which is widely used to treat benign prostate enlargement.[274,277,278] It is an enzyme-activated inhibitor in which the NADH reduces the C=C bond in the A ring, which is not in the same position as in the substrate. The resulting anion cannot become protonated but instead adds to the NAD^+ as shown in Eq. 22-15.

A number of other androgens are present in the body. The adrenal glands make **dehydroepiandrosterone** (DHEA; Fig. 22–11), which circulates in human blood as its sulfate ester in higher concentration than that of any other androgen.[279–283] However, this steroid is absent from most species. DHEA can be taken up by tissues and converted to testosterone, estrogens, or other steroids (Fig. 22–11).[280] Recent attention has been focused on this hormone because it reaches a peak plasma concentration at age 20–25 years and by age 70 has fallen to 1/5 this value or less.[283] Should older men supplement their circulating DHEA by oral ingestion of 25–50 mg per day of DHEA sulfate? The hormone depresses blood cholesterol and lowers blood glucose in diabetic individuals.[280,282] It seems to promote increased energy metabolism.[281] It may fight obesity and atherosclerosis,[282] increase levels of estrogen and other steroid hormones in the brain,[284] and enhance memory and immune function.[285–285c] However, the hormone may be metabolized differently in different tissues, and its pathway of biosynthesis in the brain is uncertain.[279,286] Will men synthesize more testosterone from DHEA or make more estrogens in their adipose tissues? Young women should not take DHEA. It may increase the testosterone and dihydrotestosterone levels in the blood manyfold, and the women may become hirsute and masculinized.[280] However, most DHEA is converted to estrogen which may be of value to older women.[285a]

In addition to their role in sexual development androgens have a generalized "anabolic" effect causing increased protein synthesis, especially in muscles.[287–289] They promote bone growth, and the adolescent growth spurt in both males and females is believed to result from androgens. The greater height attained by men results in part from the higher concentration of androgen than is present in women. Many synthetic steroids have been made in an attempt to find "anabolic hormones" with little or no androgenic activity. The effort has been at least partially successful, and the use of anabolic hormones by athletes has become both widespread and controversial.

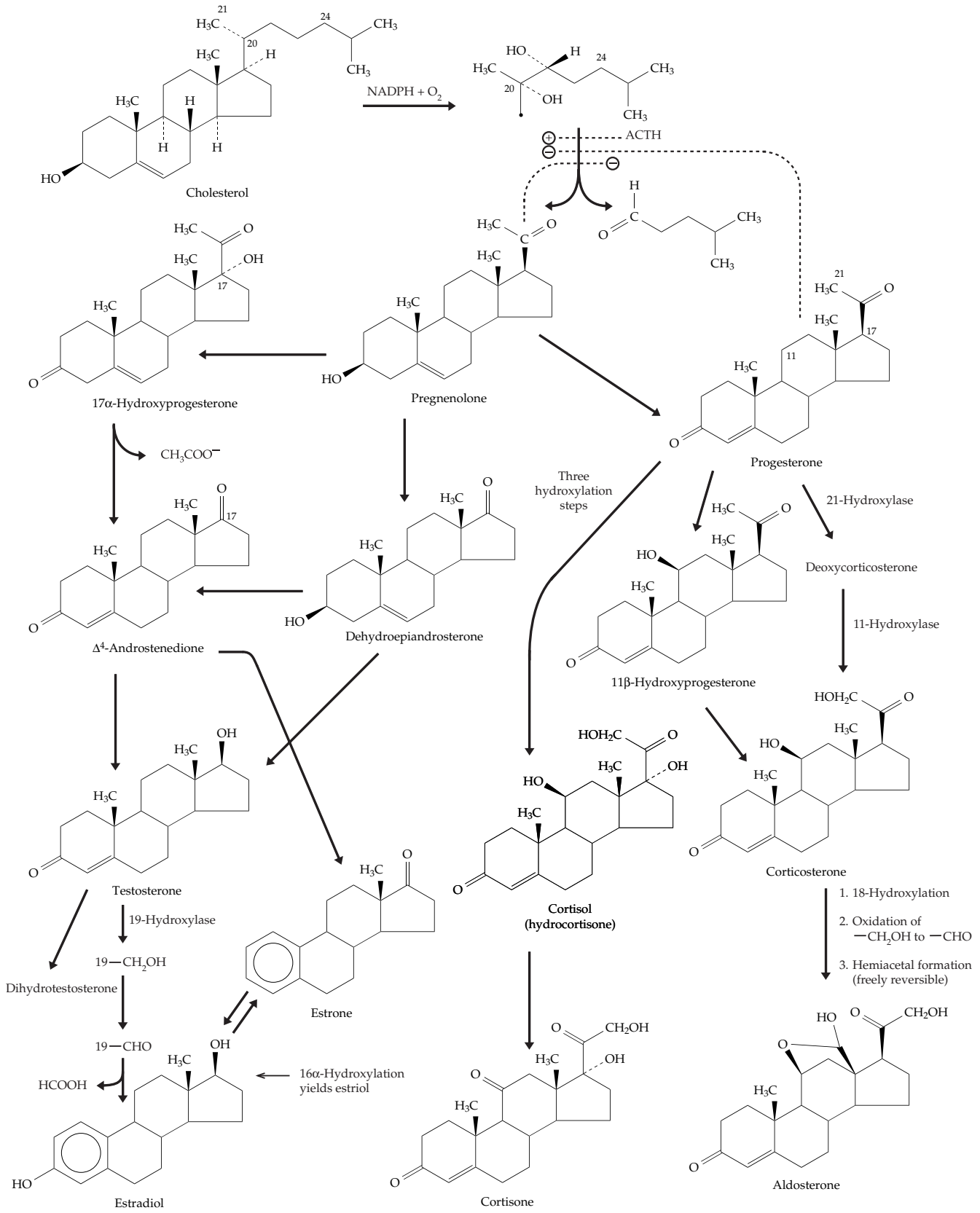
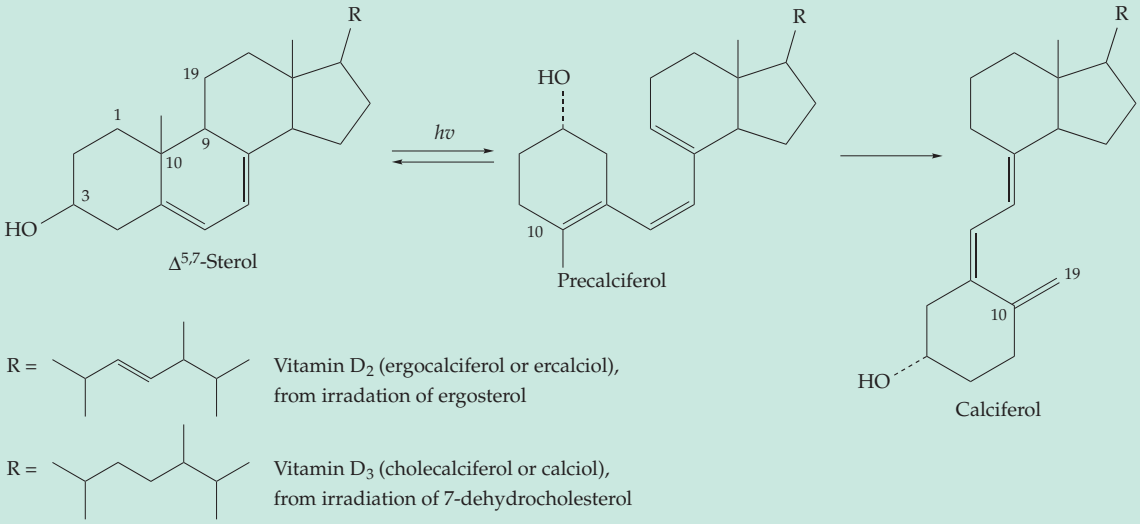


Figure 22-11 Biosynthesis of some steroid hormones.

BOX 22-C VITAMIN D

A lack of vitamin D causes **rickets**, a disease of humans and other animals in which the bones are soft, deformed, and poorly calcified. Rickets was recognized by some persons to result from a dietary deficiency well over a hundred years ago, and the use of cod liver oil to prevent the disease was introduced in about 1870. By 1890 an association of rickets with a lack of sunlight had been made.

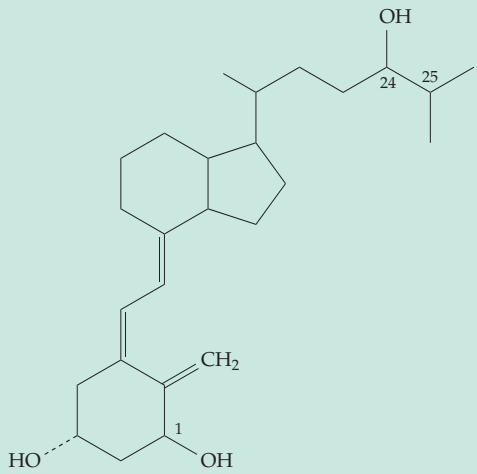


However, it was not until 1924, when Steenbock and Hess showed that irradiation of certain foods generated protective activity against the disease, that vitamin D (**calciferol**) was recognized as a second lipid-soluble vitamin. Vitamin D is a family of compounds formed by the irradiation of $\Delta^{5,7}$-unsaturated sterols such as ergosterol and 7-dehydrocholesterol.[a–e] The former yields **ergocalciferol** (vitamin D_2) and the latter **cholecalciferol** (vitamin D_3).

At low temperature the intermediate **precalciferol** can be isolated. Irradiation sets up a photochemical steady-state equilibrium between the $\Delta^{5,7}$-sterol and the precalciferol. At higher temperatures the latter is converted to calciferol.[f] Other products, including toxic ones, are produced in slower photochemical side reactions. Therefore, the irradiation of ergosterol for food fortification must be done with care.

With normal exposure to sunlight enough 7-dehydrocholesterol is converted to cholecalciferol in the skin that it was concluded that no dietary vitamin D is required by most adults except during pregnancy. However, recently it has been recognized that old and sick adults probably need 400–600 I.U. per day to maintain calcium absorption and to prevent osteoporesis and fractures.[g,gh] It is usually recommended that children receive ~20 μg (400 I.U.) of ergocalciferol per day in their diet. Larger amounts are undesirable, and at a tenfold higher level vitamin D is seriously toxic.[h]

The principal function of vitamin D is in the control of calcium metabolism. This is accomplished through the mediation of polar, hydroxylated metabolites, the most important of which is **1α,25-dihydroxyvitamin D3** (1,25-dihydroxycholecalciferol or calcitriol).[a,i,j] This compound may be properly described as a steroid hormone and vitamin D itself as a hormone precursor formed in the skin. The major hydroxylation reactions of

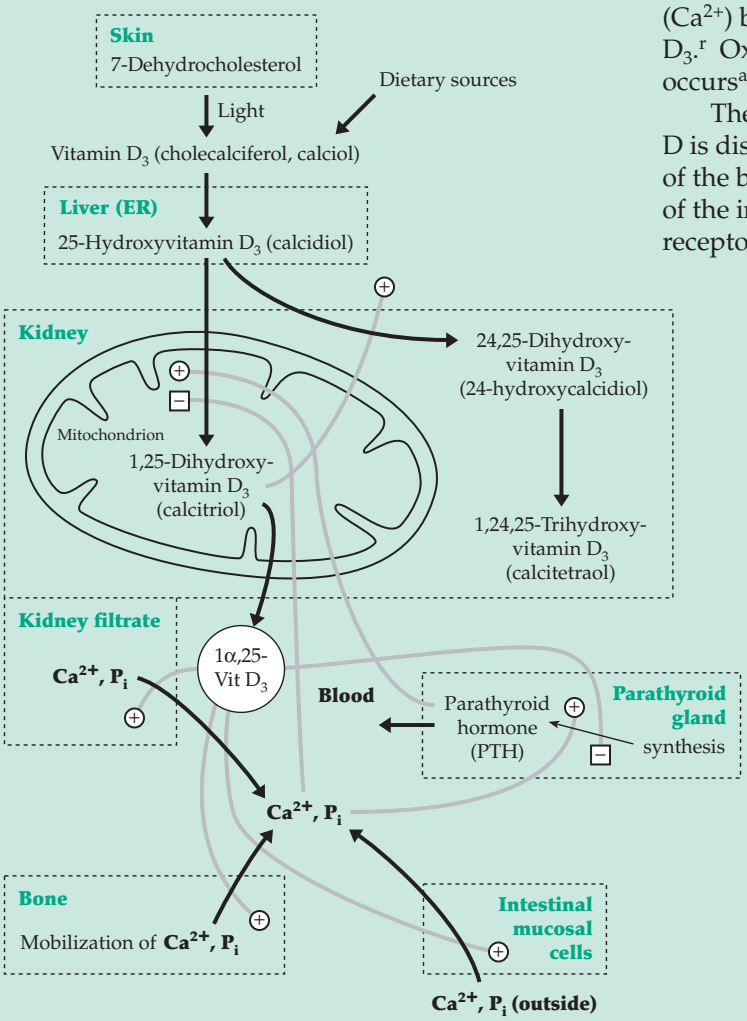


1 α,25-Dihydroxyvitamin D_3

vitamin D are summarized in the accompanying scheme. The first hydroxylation to 25-hydroxyvitamin D_3 occurs largely in the liver,[jk] but the subse-

BOX 22-C VITAMIN D (continued)

quent cytochrome P450 catalyzed 1α-hydroxylation takes place almost entirely in the kidneys.[k,l] Since it is 1,25-dihydroxy derivative that is essential for control of calcium ion metabolism, human patients



with damaged kidneys often suffer severe demineralization of their bones (renal osteodystrophy). Administration of synthetic 1α,25-dihydroxyvitamin D_3 provides an effective treatment for these persons and also for children with an inherited defect in production of this hormone.[a,m] However, large doses, tested as an antileukemia drug, caused a severe hypercalcinemia limiting its use.[n]

A second major vitamin D metabolite is 24*R*,25-dihydroxyvitamin D_3, a compound that circulates in the blood at a concentration 10 times higher than that of the 1,25-isomer.[a,b] However, no biological function has been discovered, and like a series of other polar metabolites (>30) it is probably on a pathway of inactivation and degradation of vitamin D. 1α,25-Dihydroxyvitamin D is also hydroxylated at C-24.[o] Additional hydroxylations occur at the 23- and 26-positions.[p,q] The 24-OH is often converted to an oxo group. Oxidation at C-26 together with cyclization yields 26,23 lactol and lactone species. The 25-hydroxy-26,23-lactone suppresses serum (Ca^{2+}) by competing with 1,25-dihydroxyvitamin D_3.[r] Oxidative cleavage of the side chains also occurs[a,s] as in the metabolism of cholesterol.

The hormonally active 1α,25-dihydroxyvitamin D is distributed through the bloodstream to all parts of the body. It is taken up rapidly by nuclei of cells of the intestinal lining where it binds to a 55-kDa receptor protein. In response, the cells synthesize **calbindins** (Chapter 6), Ca^{2+}-binding proteins which facilitate the uptake of calcium ions by the body.[t–v] (see also Fig. 6-7).

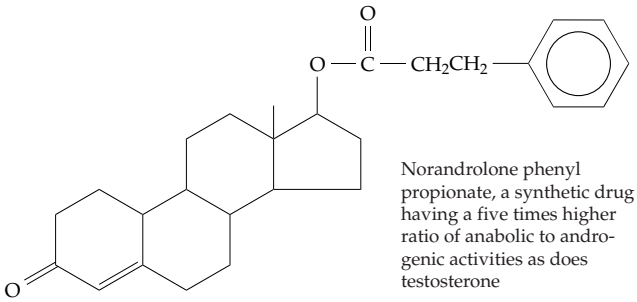
Other target organs for the action of 1,25-dihydroxyvitamin D include the kidneys, bone, muscle,[vw] and skin. The hormone promotes reabsorption of both Ca^{2+} and inorganic phosphate by kidney tubules. In bone it binds to a specific receptor where it promotes the mobilization of calcium ions. This effect may result in part from stimulation of calcium-activated ATPase of the outer membrane of bone cells. Dissolution of bone also requires the presence of **parathyroid hormone** (**PTH**), the 83-residue hormone secreted by the parathyroid gland. In women past the age of menopause and in elderly men the production of 1,25-dihydroxyvitamin D decreases.[w] This may be a cause of the serious bone loss (**osteoporosis**) frequently observed. Treatment with 1,25-dihydroxyvitamin D_3 or a synthetic analog seems to be helpful to such individuals.[x,xy] See also Chapter 30, Section A,5.

There is another important member of the Ca^{2+} homeostatic system. While vitamin D and PTH act together to increase the calcium level of the blood, **calcitonin**, a hormone of the thyroid gland, lowers the level of Ca^{2+} by promoting deposition of calcium in bone by the osteoblasts. The overall effect is to hold the concentration of Ca^{2+} in the blood at 2.2–2.6 mM in most animals with bones serving as a mobile reserve. Another role for vitamin D is suggested by the observation that 1,25-dihydroxyvitamin D_3 inhibits the growth promoting effect of interleukin-2 on mitogen-activated lymphocytes (Chapter 31).

BOX 22-C (continued)

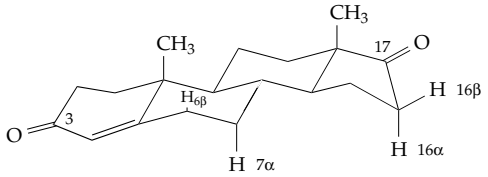
Like other steroid hormones 1,25-dihydroxyvitamin D_3 acts to regulate gene transcription. It binds to a specific receptor, a member of the v-*erb*-A superfamily of ligand-activated transcription factors and a relative of the steroid receptor family.[a,j,y] Like other hormone receptors this vitamin D_3 receptor (VDR) has some rapid "nongenomic" actions such as causing an increase in intracellular [Ca^{2+}] as well as slower effects on transcription.[y–bb] Like the steroid receptors it is found in both cytoplasm and nucleus.[y] When occupied by 1,25-dihydroxyvitamin D, the receptor binds tightly to specific dihydroxyvitamin D_3 response elements in the DNA. These are found in promoter sequences for genes such as that of bone protein **osteocalcin**.[cc] In some promoters the response element binds heterodimers of VDR with another receptor, e.g., the retinoid X receptor (Table 22-1).[cc,dd] One effect of this response element is to activate the gene for the 24-hydroxylase involved in degradation of 1,25-dihydroxyvitamin D.[dd] Heterodimers of VDR with the thyroid hormone receptor activate the transcription of genes for calbindins D_{28K} and D_{9K}.[u] VDR also binds to repressor sequences, e.g., in the parathyroid gland.[ee] See also Chapter 28 and Table 22-1.

[a] DeLuca, H. F. (1988) *FASEB J.* **2**, 224–236
[b] Reichel, H., Koeffler, H. P., and Norman, A. W. (1989) *N. Engl. J. Med.* **320**, 980–991
[c] Lawson, D. E. M. (1978) *Vitamin D*, Academic Press, New York
[d] Norman, A. W. (1979) *Vitamin D, the Calcium Homeostatic Steroid Hormone*, Academic Press, New York
[e] DeLuca, H. F. (1979) *Vitamin D – Metabolism and Function*, Springer, New York
[f] Holick, M. F., Tian, X. Q., and Allen, M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3124–3126
[g] Utiger, R. D. (1998) *N. Engl. J. Med.* **338**, 828–829
[gh] Thomas, M. K., Lloyd-Jones, D. M., Thadhani, R. I., Shaw, A. C., Deraska, D. J., Kitch, B. T., Vamvakas, E. C., Dick, I. M., Prince, R. L., and Finkelstein, J. S. (1998) *N. Engl. J. Med.* **338**, 777–783
[h] Jacobus, C. H., Holick, M. F., Shao, Q., Chen, T. C., Holm, I. A., Kolodny, J. M., Fuleihan, G. E.-H., and Seely, E. W. (1992) *N. Engl. J. Med.* **326**, 1173–1177
[i] Holick, M. F., Schnoes, T. H. K., DeLuca, H. F., Suda, T., and Cousins, R. J. (1971) *Biochemistry* **10**, 2799–2804
[j] Minghetti, P. P., and Norman, A. W. (1988) *FASEB J.* **2**, 3043–3053
[jk] Hosseinpour, F., and Wikvall, K. (2000) *J. Biol. Chem.* **275**, 34650–34655
[k] Takeyama, K.-i, Kitanaka, S., Sato, T., Kobori, M., Yanagisawa, J., and Kato, S. (1997) *Science* **277**, 1827–1830
[l] Menaa, C., Vrtovsnik, F., Friedlander, G., Corvol, M., and Garabédian, M. (1995) *J. Biol. Chem.* **270**, 25461–25467
[m] Hughes, M. R., Malloy, P. J., Kieback, D. G., Kesterson, R. A., Pike, J. W., Feldman, D., and O'Malley, B. W. (1988) *Science* **242**, 1702–1705
[n] Lee, N. E., Reddy, G. S., Brown, A. J., and Williard, P. G. (1997) *Biochemistry* **36**, 9429–9437
[o] Miyamoto, Y., Shinki, T., Yamamoto, K., Ohyama, Y., Iwasaki, H., Hosotani, R., Kasama, T., Takayama, H., Yamada, S., and Suda, T. (1997) *J. Biol. Chem.* **272**, 14115–14119
[p] Mayer, E., Bishop, J. E., Chandraratna, R. A. S., Okamura, W. H., Kruse, J. R., Popjak, G., Ohnuma, N., and Norman, A. W. (1983) *J. Biol. Chem.* **258**, 13458–13465
[q] Koszewski, N. J., Reinhardt, T. A., Napoli, J. L., Beitz, D. C., and Horst, R. L. (1988) *Biochemistry* **27**, 5785–5790
[r] Yamada, S., Nakayama, K., Takayama, H., Shinki, T., Takasaki, Y., and Suda, T. (1984) *J. Biol. Chem.* **259**, 884–889
[s] Jones, G., Kano, K., Yamada, S., Furusawa, T., Takayama, H., and Suda, T. (1984) *Biochemistry* **23**, 3749–3754
[t] Szebenyi, D. M. E., and Moffat, K. (1986) *J. Biol. Chem.* **261**, 8761–8777
[u] Schräder, M., Müller, K. M., Nayeri, S., Kahlen, J.-P., and Carlberg, C. (1994) *Nature (London)* **370**, 382–386
[v] Heizmann, C. W., and Hunziker, W. (1991) *Trends Biochem. Sci.* **16**, 98–103
[vw] Morelli, S., Buitrago, C., Vazquez, G., De Boland, A. R., and Boland, R. (2000) *J. Biol. Chem.* **275**, 36021–36028
[w] Slovik, D. M., Adams, J. S., Neer, R. M., Holick, M. F., and Potts, J. T., Jr. (1981) *N. Engl. J. Med.* **305**, 372–374
[x] Tilyard, M. W., Spears, G. F. S., Thomson, J., and Dovey, S. (1992) *N. Engl. J. Med.* **326**, 357–362
[xy] Shevde, N. K., Plum, L. A., Clagett-Dame, M., Yamamoto, H., Pike, J. W., and Deluca, H. F. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13487–13491
[y] Barsony, J., Renyi, I., and McKoy, W. (1997) *J. Biol. Chem.* **272**, 5774–5782
[z] Yukihiro, S., Posner, G. H., and Guggino, S. E. (1994) *J. Biol. Chem.* **269**, 23889–23893
[aa] Lissoos, T. W., Beno, D. W. A., and Davis, B. H. (1993) *J. Biol. Chem.* **268**, 25132–25138
[bb] de Boland, A. R., Morelli, S., and Boland, R. (1994) *J. Biol. Chem.* **269**, 8675–8679
[cc] Schräder, M., Müller, K. M., and Carlberg, C. (1994) *J. Biol. Chem.* **269**, 5501–5504
[dd] Allegretto, E. A., Shevde, N., Zou, A., Howell, S. R., Boehm, M. F., Hollis, B. W., and Pike, J. W. (1995) *J. Biol. Chem.* **270**, 23906–23909
[ee] Kremer, R., Sebag, M., Champigny, C., Meerovitch, K., Hendy, G. N., White, J., and Goltzman, D. (1996) *J. Biol. Chem.* **271**, 16310–16316



Norandrolone phenyl propionate, a synthetic drug having a five times higher ratio of anabolic to androgenic activities as does testosterone

Several sex-dependent differences have been observed in the action of cytochrome P450 isoenzymes on steroid hormones.[290,290a] Thus, androstenedione is hydroxylated by rat liver enzymes specific for the 6β, 7α, 16α, and 16β positions.[291] Of these the 16 hydroxylase is synthesized only in males, and synthesis of the 6 hydroxylase is also largely suppressed in females.



Δ^4-Androstenedione

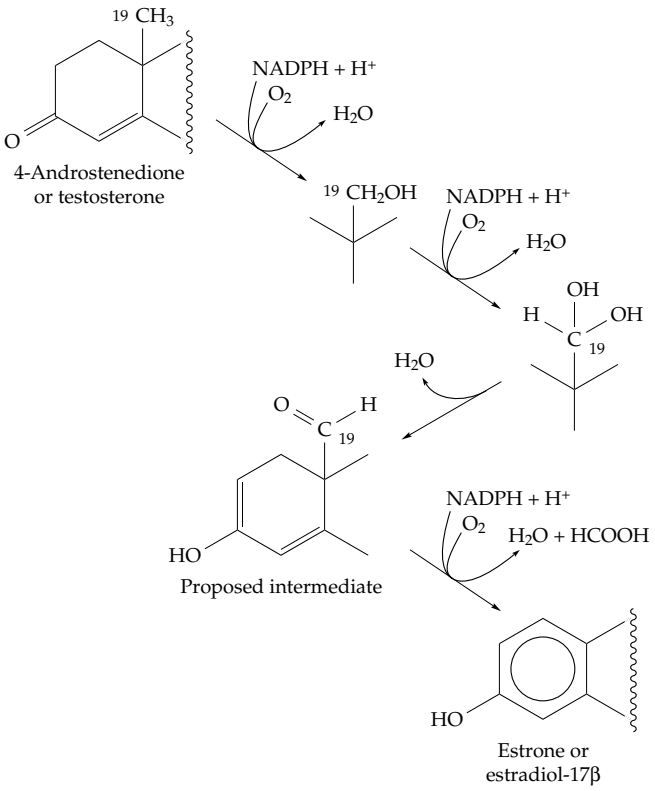
A female-specific 15β hydroxylase acts on steroid sulfates such as corticosterone sulfate and forms the major urinary excretion product of that hormone in female rats.[292] These sex-specific differences in enzymes are thought to be related to secretions of growth hormone that are in turn controlled by the "programming" of the hypothalamus by androgen during the neonatal period in rats[272] or during human fetal development.

4. Estrogens

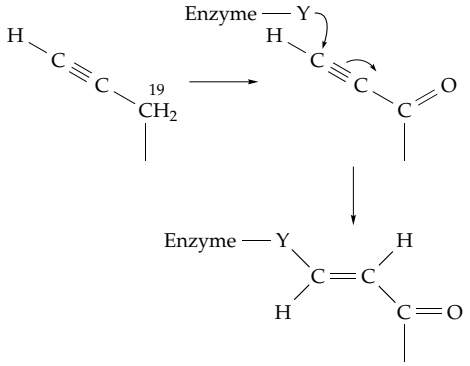
The principal **estrogenic** or female hormone is estradiol-17β. It is formed by oxidative removal of C-19 of testosterone followed by aromatization of the A ring.[268,293–295] All of the estrogenic hormones have this aromatic ring. Its formation involves three steps of hydroxylation followed by elimination of formate (Eq. 22-16). This **aromatase** appears to be a unique cytochrome P450, which catalyzes all of the steps of Eq. 22-16. It accepts electrons from NADPH via the flavoprotein NADH-cytochrome P450 reductase, which serves as the intermediate electron carrier.[296] It probably acts by a mechanism related to that illustrated for lanosterol 14α-demethylase on the right-hand side of Fig. 22-8.[294] This enzyme is the target for synthetic enzyme-activated inhibitors.[293] One of these is an androstenedione derivative with an acetylenic group attached to C-19. Passage through the first two steps of Eq. 22-16 generates a conjugated ketone to which a nucleophilic group of the enzyme can add irreversibly to inactivate the enzyme (Eq. 22-17). The C-17 acetylenic progesterone antagonist (norethyndrone) is also an enzyme activated inhibitor of the aromatase.

Estrogens are formed largely in the ovary and during pregnancy in the placenta. Estrogens are also synthesized in the testes, and the estrogen content of the horse testis is the highest of any endocrine organ. Target tissues for estrogens include the mammary glands, the uterus, and many other tissues throughout the body. Estrogens act on the growing ends of the long bones to stop growth and are therefore responsible, in part, for the shorter stature of females as compared to males. They are responsible for the overall higher fat content of the female body and for the smoother skin of the female. Recent attention has been focused on the effects of estrogens on brain neurons. Estrogens stimulate sprouting of axons and dendrites in neurons in cell cultures, and there is preliminary evidence that the hormone improves memory in healthy women and in those with Alzheimer disease.[285]

The cooperative action of progesterone and estradiol regulate the menstrual cycle. At the beginning of the cycle the levels of both estrogen and progesterone are low. Estrogen synthesis increases as a result of release of **follicle-stimulating hormone** (FSH) from the anterior pituitary. This hormone stimulates growth of the graafian follicles of the ovary which in turn produce estrogen. At about the midpoint of the cycle, as a result of the action of the pituitary **luteinizing hormone** (LH), an ovum is released and progesterone secretion begins. The latter is essential to maintenance of pregnancy. If a blastocyst is not implanted, hormone production decreases and menstruation occurs.



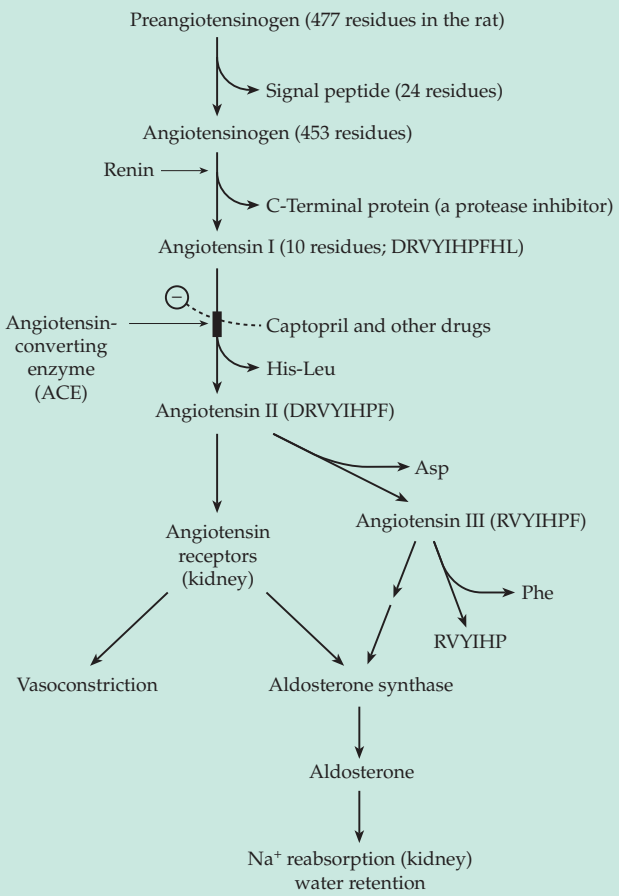
(22-16)



(22-17)

BOX 22-D THE RENIN–ANGIOTENSIN–ALDOSTERONE SYSTEM AND THE REGULATION OF BLOOD PRESSURE

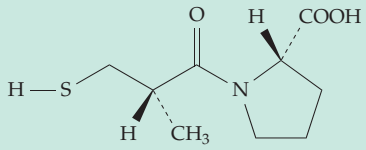
The mineralocorticoid **aldosterone** was isolated and identified in 1953. Although the function of adrenal cortical hormones in regulation of electrolytes had been known for many years, the special role of aldosterone had been overlooked.[a] Aldosterone works in concert with the aspartate protease **renin** and the octapeptide **angiotensin II** to regulate blood pressure. Angiotensin II, which is the most potent pressor substance known, is formed in the liver from the 477-residue (in the rat) **preangiotensinogen** as shown in the following cascade:



This **renin–angiotensin system** is a peptide hormone-generating system that operates in blood plasma rather than within tissues.[b,c]

Angiotensinogen, which is secreted by the liver and circulates in the blood, is converted to the physiologically inactive decapeptide angiotensin I by cleavage of a Leu-Leu peptide bond by the 328-residue renin.[d–f] Its precursor **preprorenin** is produced in the kidneys by the juxtaglomerular cells as well as in some other tissues and undergoes several steps of processing before the active enzyme is formed.[g–j] Active renin is released from the kidney cells into the bloodstream in response to various stimuli which include low arterial pressure resulting from constriction of the renal arteries or loss of blood.[k] Parathyroid hormone, glucagon, other adrenergic agonists, cAMP, some prostaglandins, low levels of aldosterone or Na^+, or high K^+ all induce secretion of renin. High blood pressure, α-adrenergic agonists, some prostaglandins, angiotensin, vasopressin, high Na^+, or low K^+ concentration decrease secretion of renin. It has been suggested that these diverse effects may be modified by a rise or fall in the Ca^{2+} concentration, high Ca^{2+} inhibiting secretion of renin, and low Ca^{2+} favoring secretion.

The only known physiological substrate for renin is angiotensinogen,[e] but it may also act on related proteins in the brain and other organs.[g] The inactive angiotensin I is converted to angiotensin II by the metal-containing carboxydipeptidase known as **angiotensin-converting enzyme**.[l,m] This enzyme is a target for drugs such as captopril, which is used to control high blood pressure (hypertension).[i] A zinc-dependent aminopeptidase may cut off the



Captopril

N-terminal aspartate to form angiotension III,[n] and degradation of angiotensions II and III can be initiated by removal of the C-terminal phenylalanine by a prolylcarboxypeptidase.[o]

Angiotensin II has a variety of effects. By constricting blood vessels it raises blood pressure, and by stimulating thirst centers in the brain it increases blood volume. Both angiotensins II and III also act on the adrenal gland to promote the synthesis and release of aldosterone. Most of the effects of angiotension II are mediated by 359-residue seven-helix G-protein linked receptors which activate phospholipase C.[p,q,qr] Like other steroid hormones aldosterone acts,via mineralocorticoid receptors, to control transcription of a certain set of proteins. The end effect is to increase the transport of Na^+ across the renal tubules and back into the blood. Thus, aldosterone acts to decrease the loss of Na^+ from the body. It promotes retention of water and raises

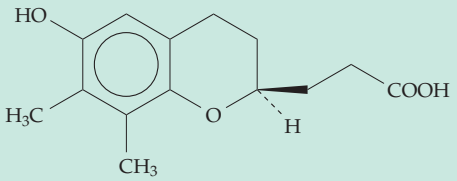
BOX 22-D THE RENIN–ANGIOTENSIN–ALDOSTERONE SYSTEM AND THE REGULATION OF BLOOD PRESSURE (continued)

blood pressure.[c,r,s] Its primary function is to provide adequate Na^+ to cells.[s] Dietary sodium appears to have little or no effect on blood pressure.[t]

The control of blood pressure is considerably more complex than it is described here. Another hormone system involving the peptide **bradykinin** and prostaglandins acts to lower blood pressure. Bradykin is also cleaved by the angiotensin-converting enzyme but is *inactivated* by the cleavage.[u,v] At least ten human genes have been shown to affect blood pressure.[c] One of these is the structural gene for angiotensin-converting enzyme, which has been linked to hypertension in both rats and humans.[w,x]

While several antagonists of angiotensin-converting enzyme are widely used to treat hypertension, they are not free of harmful side effects.[u,y,z] One alternative approach is to inhibit renin.[v,y]

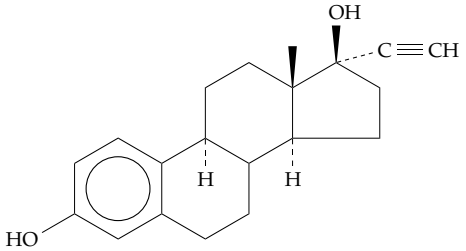
While the angiotensins promote release of aldosterone, the **atrial natriuretic hormone**[r,aa–cc] inhibits release. This group of 21- to 33-residue polypeptides, secreted by cells of the atria (auricles) of the heart, also inhibits release of renin and promotes secretion of both Na^+ and water. Thus, they antagonize the action of aldosterone, which promotes Na^+ retention. However, there is uncertainty as to the significance of these peptides. The following metabolite of γ-tocopherol (Fig. 15-24) has been isolated from urine and is proposed as a new endogenous natriuretic factor.[dd]



2, 7, 8-trimethyl-2-(β-carboxyethyl)-6-hydroxychroman

[a] Tait, J. F., and Tait, S. A. S. (1978) *Trends Biochem. Sci.* **3**, N273–N275
[b] Inagami, T. (1994) in *Essays in Biochemistry*, Vol. 28 (Tipton, K. F., ed), pp. 147– , Portland Press, Chapel Hill, North Carolina
[c] Lifton, R. P. (1996) *Science* **272**, 676–680
[d] Kem, D. C., and Brown, R. D. (1990) *N. Engl. J. Med.* **323**, 1136–1137
[e] Wang, W., and Liang, T. C. (1994) *Biochemistry* **33**, 14636–14641
[f] Tong, L., Pav, S., Lamarre, D., Pilote, L., LaPlante, S., Anderson, P. C., and Jung, G. (1995) *J. Mol. Biol.* **250**, 211–222
[g] Yanai, K., Saito, T., Kakinuma, Y., Kon, Y., Hirota, K., Taniguchi-Yanai, K., Nishijo, N., Shigematsu, Y., Horiguchi, H., Kasuya, Y., Sugiyama, F., Yagami, K.-i, Murakami, K., and Fukamizu, A. (2000) *J. Biol. Chem.* **275**, 5–8
[h] Smith, E. L., Hill, R. L., Lehman, I. R., Lefkowitz, R. J., Handler, P., and White, A. (1983) *Principles of Biochemistry, Mammalian Biochemistry*, 7th ed., McGraw-Hill, New York (pp. 157–163)
[i] Bull, H. G., Thornberry, N. A., Cordes, M. H. J., Patchett, A. A., and Cordes, E. H. (1985) *J. Biol. Chem.* **260**, 2952–2962
[j] Derkx, F. H. M., Schalekamp, M. P. A., and Schalekamp, M. A. D. H. (1987) *J. Biol. Chem.* **262**, 2472–2477
[k] Fray, J. C. S., Lush, D. J., and Valentine, A. N. D. (1983) *Fed. Proc.* **42**, 3150–3154
[l] Tipnis, S. R., Hooper, N. M., Hyde, R., Karran, E., Christie, G., and Turner, A. J. (2000) *J. Biol. Chem.* **275**, 33238–33243
[m] Ehlers, M. R. W., Schwager, S. L. U., Scholle, R. R., Manji, G. A., Brandt, W. F., and Riordan, J. F. (1996) *Biochemistry* **35**, 9549–9559
[n] Vazeux, G., Wang, J., Corvol, P., and Llorens-Cortès, C. (1996) *J. Biol. Chem.* **271**, 9069–9074
[o] Tan, F., Morris, P. W., Skidgel, R. A., and Erdös, E. G. (1993) *J. Biol. Chem.* **268**, 16631–16638
[p] Noda, K., Feng, Y.-H., Liu, X.-p, Saad, Y., Husain, A., and Karnik, S. S. (1996) *Biochemistry* **35**, 16435–16442
[q] Boucard, A. A., Wilkes, B. C., Laporte, S. A., Escher, E., Guillemette, G., and Leduc, R. (2000) *Biochemistry* **39**, 9662–9670
[qr] Heerding, J. N., Hines, J., Fluharty, S. J., and Yee, D. K. (2001) *Biochemistry* **40**, 8369–8377
[r] Flier, J. S., and Underhill, L. H. (1985) *N. Engl. J. Med.* **313**, 1330–1340
[s] Berger, S., Bleich, M., Schmid, W., Cole, T. J., Peters, J., Watanabe, H., Kriz, W., Warth, R., Greger, R., and Schütz, G. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9424–9429
[t] Taubes, G. (1998) *Science* **281**, 898–907
[u] Williams, G. H. (1988) *N. Engl. J. Med.* **319**, 1517– 1525
[v] Dealwis, C. G., Frazao, C., Badasso, M., Cooper, J. B., Tickle, I. J., Driessen, H., Blundell, T. L., Murakami, K., Miyazaki, H., Sueiras-Diaz, J., Jones, D. M., and Szelke, M. (1994) *J. Mol. Biol.* **236**, 342–360
[w] Hilbert, P., Lindpaintner, K., Beckmann, J. S., Serikawa, T., Soubrier, F., Dubay, C., Cartwright, P., De Gouyon, B., Julier, C., Takahasi, S., Vincent, M., Ganten, D., Georges, M., and Lathrop, G. M. (1991) *Nature (London)* **353**, 521–529
[x] Kreutz, R., Hübner, N., James, M. R., Bihoreau, M.-T., Gauguier, D., Lathrop, G. M., Ganten, D., and Lindpaintner, K. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8778–8782
[y] Tong, L., Pav, S., Lamarre, D., Simoneau, B., Lavallée, P., and Jung, G. (1995) *J. Biol. Chem.* **270**, 29520–29524
[z] Warren, J. B., and Loi, R. K. (1995) *FASEB J.* **9**, 411–418
[aa] Sagnella, G. A., and MacGregor, G. A. (1986) *Trend Biochem. Sci.* II, 299–302
[bb] Cantin, M., and Genest, J. (1986) *Sci. Am.* **254**(February), 76–81
[cc] Lopez, M. J., Garbers, D. L., and Kuhn, M. (1997) *J. Biol. Chem.* **272**, 23064–23068
[dd] Wechter, W. J., Kantoci, D., Murray, E. D., Jr., D'Amico, D. C., Jung, M. E., and Wang, W.-H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6002–6007

Administration of estrogens and progestins inhibits FSH and LH secretion from the pituitary (feedback inhibition) and hence ovulation. This effect is the action of contraceptive pills. A small amount of the synthetic estrogen 17-ethynylestradiol may be taken daily for 10–15 days followed by a combination of estrogen plus a progestin such as norethindrone for 0–15 days. Alternatively, a progestin alone may be

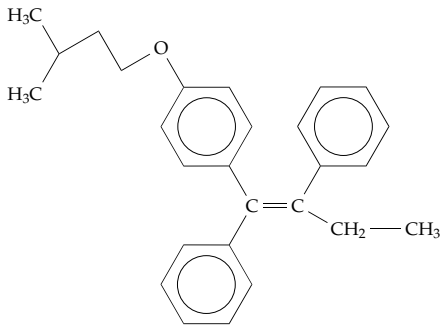


17-Ethynylestradiol

ingested over the entire period. Another synthetic compound with estrogenic activity is diethylstilbestrol. Its once widespread use in promoting growth of cattle and other animals has been discontinued because of carcinogenic action in rats fed large amounts of the compound.

Human cancers of the breast and endometrium are stimulated by estrogen. However, the mechanism is unknown.[297,298] It has usually been assumed that the proliferation of cells induced by estrogens leads to mutations and cancer. However, estrogens can form adducts with DNA after oxidation to 2- and 4-hydroxy derivatives and further oxidation to quinones.[297,298] In a similar manner prostate cancer is promoted by androgens. It has also been observed that in the United States the incidence of cancers of the ovary and endometrium has declined substantially during the past 35 years, perhaps as a result of the anti-estrogenic activity of the progestin in the widely used oral contraceptives.[299] This observation led to the idea of **hormonal chemoprevention**, the deliberate use of hormone antagonists to slow cancer formation.[299a] This may be especially attractive to persons carrying known cancer-susceptibility genes. The synthetic antiestrogen **tamoxifen** is being evaluated as a chemopreventive agent for breast cancer.[299,299b] However, a planned large-scale trial was postponed because of uncertainities about safety.[300]

In addition to steroids there are plant flavonoids (Box 21-E) that have estrogenic activity. These labile compounds are among the **"environmental estrogens."** In addition, there are many much more stable compounds, including the insecticides atrazine and DDT, PCBs, and phenolic softeners for plastics, that have weak estrogenic activity. Alarm has been sounded



trans-Tamoxifen

by some who maintain that these **xenoestrogens** are contributing to breast cancer, to reproductive difficulties in animals, and to low sperm counts in men.[301] It seems surprising that such small amounts of weakly estrogenic compounds could have such large effects. Some experiments suggest that two weak xenoestrogens may cooperate to give larger effects,[302,303] but this concept needs further evaluation.

Although estrogens are usually regarded as female hormones, they are present in small amounts in male blood and in high concentration in semen. Male reproductive tissues contain estrogen receptors, and mice deficient in these receptors are sterile and their testes degenerate.[304] Furthermore, as in females estrogen stops growth of long bones in late puberty. A few men lacking estrogen or estrogen receptors have grown very tall (>2.1 m and still growing).[305]

5. The Steroid Receptor Family

The principal mode of action of steroid hormones is to stimulate transcription of specific sets of genes. The plasma concentrations of these hormones are low, typically ~10^{-9} M, but they have a high affinity for their protein receptors, some of which are located initially in the cytosol but are found largely in the nucleus. The earliest identification of steroid receptors was accomplished with radioactive ^{3}H-labeled progesterone, estrogens, and glucocorticoids.[306,307] Autoradiographs of thin sections of tissue made after the uptake of the hormones revealed that the radioactivity was concentrated in the nuclei. The protein receptors were then isolated and were found to bind both to the hormone and to specific sequences in DNA, the hormone **response elements**.[308] The progesterone, estrogen, and glucocorticoid receptors are multidomain proteins with two of the domains having highly conserved sequences and structures. One of these domains binds to DNA and the other to the steroid hormone. Their amino acid sequences are also related to those of the v-*erb-A* oncogene (Table 11-3).[309] With this knowledge available and through use of methods of "reverse genetics," receptors for the other steroid hormones and also for vitamin D, retinoic acid, and thyroid hormone were identified as members of the family as were several "orphan" receptors of as yet unknown function (Table 22-1). Invertebrates have similar receptors. At least eight are present in *Drosophila*,[307] and the family is present in the nematode *Caenorhabditis*.[310]

Because of their hydrophobic character the steroid hormones or other nonpolar ligands diffuse through membranes into cells. There they may encounter a variety of binding proteins that affect their access to a receptor.[307,311] Some receptors, including glucocorticoid receptors, are found in the cytosol. After a hor-

TABLE 22-1
Known Members of the Steroid Receptor Family[a]

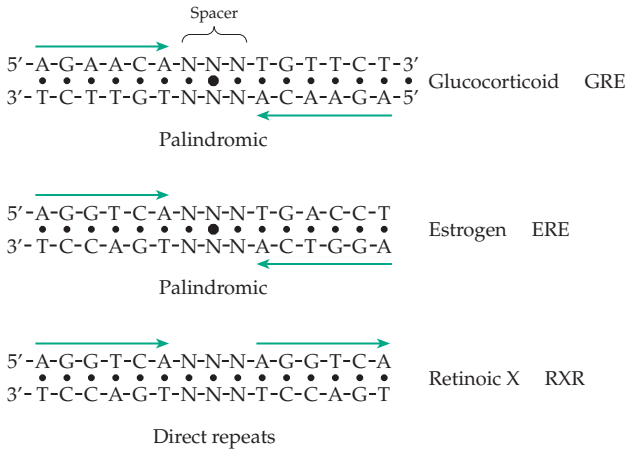
Glucocorticoid[b–d]	Thyroid α, β_1, β_2[r]
Mineralocorticoid[e,f]	Retinoic acid α, β, γ [s]
Progesterone[g,h]	Retinoid-X α,β[t–w]
Androgen[i,j,k]	Peroxisome proliferators[x,y]
Estrogen[l–o]	Farnesoid X: bile acids[z,aa]
Vitamin D_3[p,q]	Orphan receptors, 8 or more

[a] See Fuller (1991) *FASEB J.* **5**, 3092–3099
[b] Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991) *Nature (London)* **352**, 497–505
[c] Eriksson, M. A. L., Härd, T., and Nilsson, L. (1995) *Biophys. J.* **68**, 402–426
[d] La Baer, J., and Yamamoto, K. R. (1994) *J. Mol. Biol.* **239**, 664–688
[e] Funder, J. W. (1993) *Science* **259**, 1132–1133
[f] Geller, D. S., Farhi, A., Pinkerton, N., Fradley, M., Moritz, M., Spitzer, A., Meinke, G., Tsai, F. T. F., Sigler, P. B., and Lifton, R. P. (2000) *Science* **289**, 119–123
[g] Williams, S. P., and Sigler, P. B. (1998) *Nature (London)* **393**, 392–396
[h] Knotts, T. A., Orkiszewski, R. S., Cook, R. G., Edwards, D. P., and Weigel, N. L. (2001) *J. Biol. Chem.* **276**, 8475–8483
[i] Grossmann, M. E., Lindzey, J., Blok, L., Perry, J. E., Kumar, M. V., and Tindall, D. J. (1994) *Biochemistry* **33**, 14594–14600
[j] Matias, P. M., and 13 other authors. (2000) *J. Biol. Chem.* **275**, 26164–26171
[k] Sack, J. S., Kish, K. F., Wang, C., Attar, R. M., Kiefer, S. E., An, Y., Wu, G. Y., Scheffler, J. E., Salvati, M. E., Krystek, S. R., Jr., Weinmann, R., and Einspahr, H. M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4904–4909
[l] Ekena, K., Weis, K. E., Katzenellenbogen, J. A., and Katzenellenbogen, B. S. (1996) *J. Biol. Chem.* **271**, 20053–20059
[m] Tanenbaum, D. M., Wang, Y., Williams, S. P., and Sigler, P. B. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5998–6003
[n] Brzozowski, A. M., Pike, A. C. W., Dauter, Z., Hubbard, R. E., Bonn, T., Engström, O., Öhman, L., Greene, G. L., Gustafsson, J.-Å., and Carlquist, M. (1997) *Nature (London)* **389**, 753–758
[o] Kahlert, S., Nuedling, S., van Eickels, M., Vetter, H., Meyer, R., and Grohé, C. (2000) *J. Biol. Chem.* **275**, 18447–18453
[p] Craig, T. A., Veenstra, T. D., Naylor, S., Tomlinson, A. J., Johnson, K. L., Macura, S., Juranic, N., and Kumar, R. (1997) *Biochemistry* **36**, 10482–10491
[q] Hendick, M., and Carlberg, C. (2000) *J. Mol. Biol.* **304**, 793–801
[r] Kostrouch, Z., Kostrouchova, M., and Rall, J. E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 156–159
[s] Klaholz, B. P., Mitschler, A., and Moras, D. (2000) *J. Mol. Biol.* **302**, 155–170
[t] Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H., and Moras, D. (1995) *Nature (London)* **375**, 377–382
[u] Holmbeck, S. M. A., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, J., and Wright, P. E. (1998) *J. Mol. Biol.* **281**, 271–284
[v] Kersten, S., Dong, D., Lee, W.-y, Reczek, P. R., and Noy, N. (1998) *J. Mol. Biol.* **284**, 21–32
[w] Chawla, A., Repa, J. J., Evans, R. M., and Mangelsdorf, D. J. (2001) *Science* **294**, 1866–1870
[x] Nolte, R. T., Wisely, G. B., Westin, S., Cobb, J. E., Lambert, M. H., Kurokawa, R., Rosenfeld, M. G., Willson, T. M., Glass, C. K., and Milburn, M. V. (1998) *Nature (London)* **395**, 137–143
[y] Rosen, E. D., and Spiegelman, B. M. (2001) *J. Biol. Chem.* **276**, 37731–37734
[z] Makishima, M., Okamoto, A. Y., Repa, J. J., Tu, H., Learned, R. M., Luk, A., Hull, M. V., Lustig, K. D., Mangelsdorf, D. J., and Shan, B. (1999) *Science* **284**, 1362–1365
[aa] Parks, D. J., Blanchard, S. G., Bledsoe, R. K., Chandra, G., Consler, T. G., Kliewer, S. A., Stimmel, J. B., Willson, T. M., Zavacki, A. M., Moore, D. D., and Lehmann, J. M. (1999) *Science* **284**, 1365–1368

mone binds to a cytosolic receptor, the complex apparently undergoes time and temperature dependent alterations that activate the receptor before it diffuses into the nucleus and binds to its proper response element in the DNA.[312] This process has been observed directly for glucocorticoid receptors labeled by fusion with green fluorescent protein (Box 23-A).[313]

The conserved 68-residue DNA binding domain of the glucocorticoid receptor contains two Zn^{2+} ions, each coordinated by four cysteine $-CH_2-S^-$ groups with tetrahedral geometry. These two consecutive motifs form structures somewhat similar to those of the "zinc fingers" shown in Fig. 5-38.[314–316] However, the overall folding pattern is different from those considered in Chapter 5. The two zinc-binding sites lie at the N termini of a pair of helices that cross at right angles near their centers. One of these is a DNA-recognition helix that fits into the major groove of DNA thereby allowing interaction of its amino acid side chains with the bases of the DNA response elements.

The response elements for glucocorticoids and estrogen receptors contain short palindromic sequences with various three-nucleotide "spacer" sequences in the center as follows.[308,314,316–318] Two receptor proteins bind to the palindromic DNA forming a homodimeric receptor pair. For the 9-*cis* retinoic acid receptor RXR-α the response element contains a pair of direct repeats of a 6-base consensus sequence with a two-base pair spacer:



The RXR-α receptor binds differently and tends to form *heterodimeric* pairs with other receptors.[319,320] All of these receptors undergo conformational changes when agonists or antagonists bind.[320,321] Estrogens can also bind to androgen receptors, perhaps in playing their essential role in male reproductive physiology.[322] There are more than one type of receptor for each group of steroid hormones, and these may interact differently with the various response elements in DNA[323] making the effects of hormones complex and hard to analyze. Interactions with additional proteins

also affect the response of a cell to hormones.[324] Furthermore, steroid hormones have **transcription-independent effects**. For example, progesterone binds to oxytocin receptors[325] as well as to other steroid receptors, which affect a broad range of biochemical processes.[326]

F. Other Steroids

The **saponins** are a series of steroid glycosides with detergent properties that are widespread among higher plants.[327] Some are toxic, and among these toxic materials are compounds of extraordinary medical importance. Best known are the steroid glycosides of *Digitalis*, among them **digitonin** (Fig. 22-12). The particular arrangement of sugar units in this molecule imparts a specificity toward heart muscle. The compound is extremely toxic; in small amounts it acts to increase the tone of heart muscle and is widely used in treatment of congestive heart failure.[328] The maintenance dose is only 0.1 mg / day. Another toxic glycoside and heart stimulant is **ouabain** (Fig. 22-12). Ouabain is a specific inhibitor of the membrane-bound ($Na^+ + K^+$)-ATPase believed to be the ion pump that keeps intracellular K^+ concentrations high and Na^+ concentrations low (Chapter 8). Similar glycosides account for the extreme toxicity of the leaves of the oleander and the roots of the lily of the valley. A steroid glycoside from red squill is used as a rat poison. A number of alkaloids (nitrogenous bases) are derived from steroids. An example is **solanidine** (Fig. 22-12), which is present in the skins and sprouts of potatoes, making both quite toxic.

Some animals also contain toxic steroids. **Batrachotoxin** of the Columbian poison arrow frog (Fig. 22-12) is present in amounts of only 50 μg per frog.[329,330]

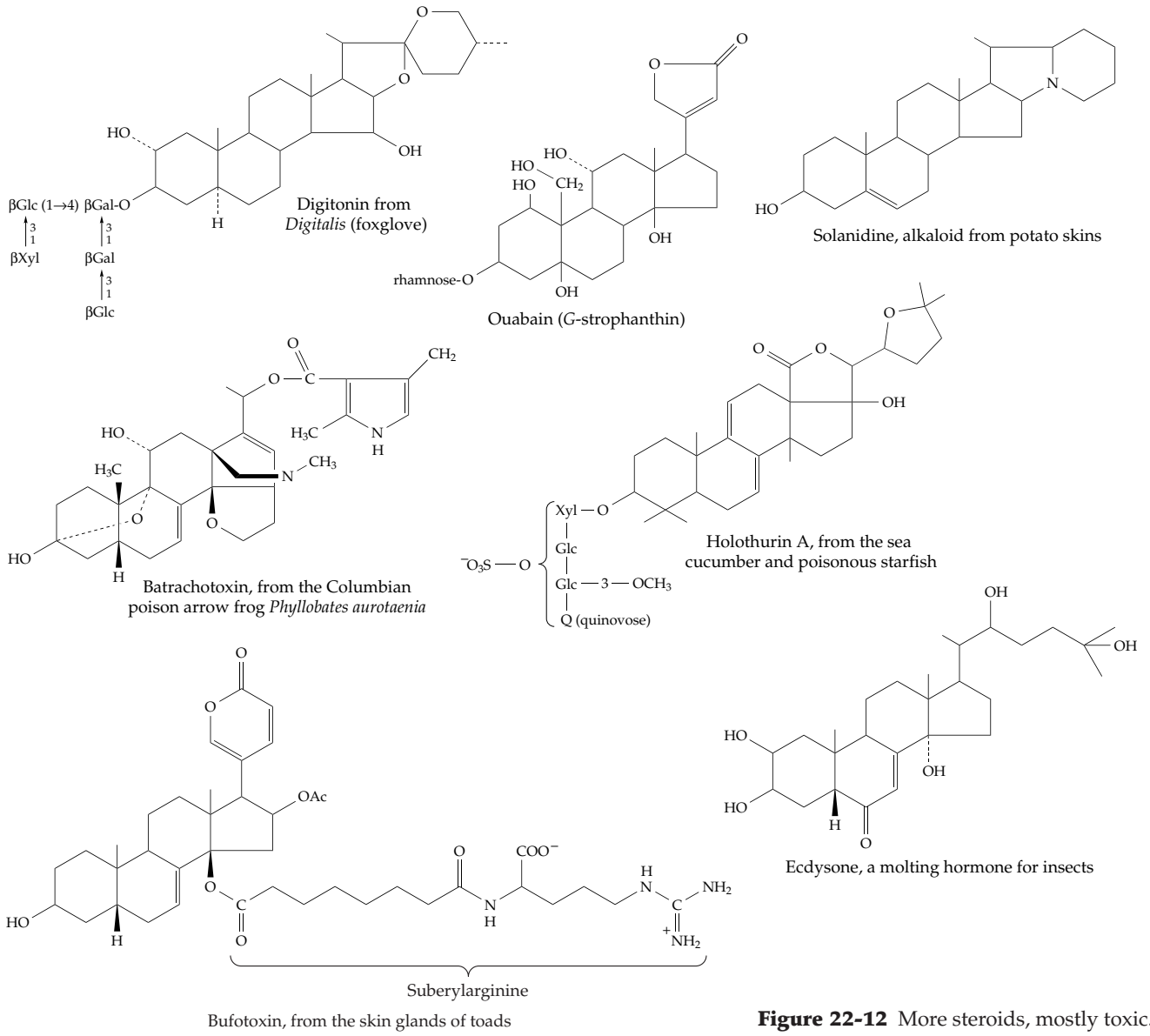
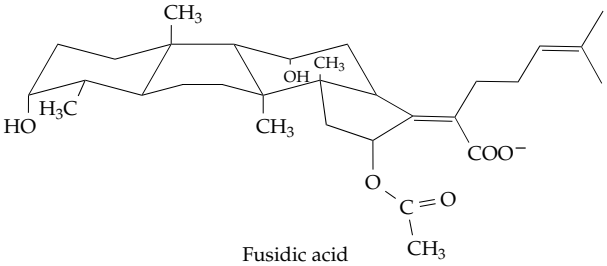


Figure 22-12 More steroids, mostly toxic.

The toxin acts on nerves to block transmissions to the muscle by increasing the permeability of membranes to sodium ions. It is specifically antagonized by **tetrodotoxin** (Fig. 30-16). Batrachotoxin alkaloids are present also in certain birds.[330a] Some echinoderms make powerful steroid toxins such as **holothurin A** (Fig. 22-12), a surface active agent that causes irreversible destruction of the excitability of neuromuscular tissues. The common toad produces in its skin steroid toxins such as **bufotoxin** (Fig. 22-12), which are sufficiently powerful to teach a dog to leave toads alone.

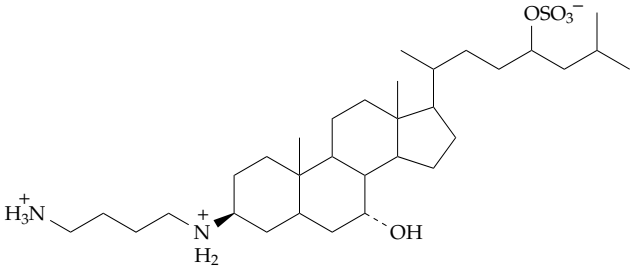
Ecdysone, a highly hydroxylated steroid (Fig. 22-12), is a molting hormone for insects.[331,332] Several molecules with ecdysone activity are known, and some of these are produced by certain plants. Although ecdysones are needed by insects for larval molting, they are toxic in excess. Perhaps plants protect themselves from insects by synthesizing these substances.

Among the antibiotics **fusidic acid** is a steroid. An inhibitor of chloramphenicol acetyltransferase[333] it is highly inhibitory to staphyococci but almost noninhibitory to *E. coli*. Note the boat conformation of the B ring.



Fusidic acid

Another steroid antibiotic, **squalamine**,[334] was isolated from the stomachs of sharks. It is effective against gram-positive and gram-negative bacteria, and some fungi as well.[335]



Squalamine

References

1. Luckner, M. (1984) *Secondary Metabolism in Plants and Animals*, 2nd ed., Academic Press, New York
2. Britton, G. (1976) in *Chemistry and Biochemistry of Plant Pigments*, 2nd ed., Vol. 1 (Goodwin, T. W., ed), pp. 262–327, Academic Press, New York
3. Haslam, E. (1985) *Metabolites and Metabolism*, Oxford Univ. Press, London
4. Herbert, R. B. (1989) *The Biosynthesis of Secondary Metabolites*, 2nd ed., Chapman and Hall, New York
5. Towers, G. H. N., and Stafford, H. A., eds. (1990) *Biochemistry of the Mevalonic Acid Pathway to Terpenoids*, Vol. 24, Plenum Press, New York and London
6. Nes, W. D., ed. (1994) *Isopentenoids and Other Natural Products*, American Chemical Society, Washington D.C.

6a. Wang, K., and Ohnuma, S.-i. (1999) *Trends Biochem. Sci.* **24**, 445–451

7. Misra, I., Narasimhan, C., and Miziorko, H. M. (1993) *J. Biol. Chem.* **268**, 12129–12135
8. Quant, P. A. (1994) in *Essays in Biochemistry*, Vol. 28 (Tipton, K. F., ed), Portland Press, Chapel Hill, North Carolina
9. Denbow, C. J., Lång, S., and Cramer, C. L. (1996) *J. Biol. Chem.* **271**, 9710–9715
10. Street, I. P., Coffman, H. R., Baker, J. A., and Poulter, C. D. (1994) *Biochemistry* **33**, 4212–4217
11. Frimpong, K., and Rodwell, V. W. (1994) *J. Biol. Chem.* **269**, 11478–11483
12. Bischoff, K. M., and Rodwell, V. W. (1997) *Protein Sci.* **6**, 156–161
13. Misra, I., and Miziorko, H. M. (1996) *Biochemistry* **35**, 9610–9616
14. Omkumar, R. V., and Rodwell, V. W. (1994) *J. Biol. Chem.* **269**, 16862–16866
15. Lawrence, C. M., Rodwell, V. W., and Stauffacher, C. V. (1995) *Science* **268**, 1758–1762
16. Luskey, K. L., and Stevens, B. (1985) *J. Biol. Chem.* **260**, 10271–10277
17. Goldstein, J. L., and Brown, M. S. (1990) *Nature (London)* **343**, 425–430
18. Meigs, T. E., Roseman, D. S., and Simoni, R. D. (1996) *J. Biol. Chem.* **271**, 7916–7922
19. Berkhout, T. A., Simon, H. M., Patel, D. D., Bentzen, C., Niesor, E., Jackson, B., and Suckling, K. E. (1996) *J. Biol. Chem.* **271**, 14376–14382
20. Correll, C. C., Ng, L., and Edwards, P. A. (1994) *J. Biol. Chem.* **269**, 17390–17393
21. Biardi, L., Sreedhar, A., Zokaei, A., Vartak, N. B., Bozeat, R. L., Shackelford, J. E., Keller, G.-A., and Krisans, S. K. (1994) *J. Biol. Chem.* **269**, 1197–1205
22. Potter, D., Wojnar, J. M., Narasimhan, C., and Miziorko, H. M. (1997) *J. Biol. Chem.* **272**, 5741–5746
23. Schafer, B. L., Bishop, R. W., Kratunis, V. J., Kalinowski, S. S., Mosley, S. T., Gibson, K. M., and Tanaka, R. D. (1992) *J. Biol. Chem.* **267**, 13229–13238
24. Paton, V. G., Shackelford, J. E., and Krisans, S. K. (1997) *J. Biol. Chem.* **272**, 18945–18950
25. Dhe-Paganon, S., Magrath, J., and Abeles, R. H. (1994) *Biochemistry* **33**, 13355–13362
26. Eisenreich, W., Menhard, B., Hylands, P. J., Zenk, M. H., and Bacher, A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6431–6436
27. Kuo, S.-M., and Aronson, P. S. (1996) *J. Biol. Chem.* **271**, 15491–15497
28. Rohmer, M., Seemann, M., Horbach, S., Bringer-Meyer, S., and Sahm, H. (1996) *J. Am. Chem. Soc.* **118**, 2564–2566
29. Sprenger, G. A., Schörken, U., Wiegert, T., Grolle, S., de Graaf, A. A., Taylor, S. V., Begley, T. P., Bringer-Meyer, S., and Sahm, H. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12857–12862
30. Himmeldirk, K., Sayer, B. G., and Spenser, I. D. (1998) *J. Am. Chem. Soc.* **120**, 3581–3589
31. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H.-J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. (1998) *J. Biol. Chem.* **273**, 16555–16560

31a. Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A., and Eisenreich, W. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1158–1163

32. Arigoni, D., Sagner, S., Latzel, C., Eisenreich, W., Bacher, A., and Zenk, M. H. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10600–10605

32a. Cane, D. E. (2000) *Science* **287**, 818–819

32b. Arigoni, D., Eisenreich, W., Latzel, C., Sagner, S., Radykewicz, T., Zenk, M. H., and Bacher, A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1309–1314

32c. Rohdich, F., Wungsintaweekul, J., Fellermeier, M., Sagner, S., Herz, S., Kis, K., Eisenreich, W., Bacher, A., and Zenk, M. H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11758–11763

32d. Herz, S., and 10 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2486–2490

32e. Rohdich, F., Wungsintaweekul, J., Eisenreich, W., Richter, G., Schuhr, C. A., Hecht, S., Zenk, M. H., and Bacher, A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6451–6456

32f. Lange, B. M., and Croteau, R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13714–13719

32g. Lange, B. M., Rujan, T., Martin, W., and Croteau, R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13172–13177

32h. Zeidler, J., Ullah, N., Gupta, R. N., Pauloski, R. M., Sayer, B. G., and Spenser, I. D. (2002) *J. Am. Chem. Soc.* **124**, 4542–4543

33. Hill, R. E., Himmeldirk, K., Kennedy, I. A., Pauloski, R. M., Sayer, B. G., Wolf, E., and Spenser, I. D. (1996) *J. Biol. Chem.* **271**, 30426–30435

33a. Paton, V. G., Shackelford, J. E., and Krisans, S. K. (1997) *J. Biol. Chem.* **272**, 18945–18950

References

33b. Kaneda, K., Kuzuyama, T., Takagi, M., Hayakawa, Y., and Seto, H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 932–937
34. Deneris, E. S., Stein, R. A., and Mead, J. F. (1985) *J. Biol. Chem.* **260**, 1382–1385
35. Silver, G. M., and Fall, R. (1995) *J. Biol. Chem.* **270**, 13010–13016
36. Sharkey, T. D., and Singsaas, E. L. (1995) *Nature (London)* **374**, 769
37. Mlot, C. (1995) *Science* **268**, 641–642
38. Brady, P. S., Scofield, R. F., Schumann, W. C., Ohgaku, S., Kumaran, K., Margolis, J. M., and Landau, B. R. (1982) *J. Biol. Chem.* **257**, 10742–10746
39. Chen, A., Kroon, P. A., and Poulter, C. D. (1994) *Protein Sci.* **3**, 600–607
40. Cunillera, N., Arró, M., Delourme, D., Karst, F., Boronat, A., and Ferrer, A. (1996) *J. Biol. Chem.* **271**, 7774–7780
41. Tarshis, L. C., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., and Poulter, C. D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15018–15023
42. Ohnuma, S.-i, Narita, K., Nakazawa, T., Ishida, C., Takeuchi, Y., Ohto, C., and Nishino, T. (1996) *J. Biol. Chem.* **271**, 30748–30754
43. Ohnuma, S.-i, Nakazawa, T., Hemmi, H., Hallberg, A.-M., Koyama, T., Ogura, K., and Nishino, T. (1996) *J. Biol. Chem.* **271**, 10087–10095
44. Math, S. K., Hearst, J. E., and Poulter, C. D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6761–6764
45. Ohnuma, S.-i, Hirooka, K., Hemmi, H., Ishida, C., Ohto, C., and Nishino, T. (1996) *J. Biol. Chem.* **271**, 18831–18837
46. Carattoli, A., Romano, N., Ballario, P., Morelli, G., and Macino, G. (1991) *J. Biol. Chem.* **266**, 5854–5859
46a. Kuzuguchi, T., Morita, Y., Sagami, I., Sagami, H., and Ogura, K. (1999) *J. Biol. Chem.* **274**, 5888–5894
46b. Kharel, Y., Zhang, Y.-W., Fujihashi, M., Miki, K., and Koyama, T. (2001) *J. Biol. Chem.* **276**, 28459–28464
47. Tarshis, L. C., Yan, M., Poulter, C. D., and Sacchettini, J. C. (1994) *Biochemistry* **33**, 10871–10877
47a. Fernandez, S. M. S., Kellogg, B. A., and Poulter, C. D. (2000) *Biochemistry* **39**, 15316–15321
48. Sacchettini, J. C., and Poulter, C. D. (1997) *Science* **277**, 1788–1789
49. Koyama, T., Tajima, M., Sano, H., Doi, T., Koike-Takeshita, A., Obata, S., Nishino, T., and Ogura, K. (1996) *Biochemistry* **35**, 9533–9538
50. Poulter, C. D., and Rilling, H. C. (1978) *Acc. Chem. Res.* **11**, 307–313
51. Poulter, C. D., Wiggins, D. L., and Le, A. T. (1981) *J. Am. Chem. Soc.* **103**, 3926–3927
52. Suga, T., Hirata, T., Aoki, T., and Shishibori, T. (1983) *J. Am. Chem. Soc.* **105**, 6178–6179
53. Davisson, V. J., Neal, T. R., and Poulter, C. D. (1993) *J. Am. Chem. Soc.* **115**, 1235–1245
54. Berdis, A. J., and Benkovic, S. J. (1996) *Biochemistry* **35**, 9253–9265
54a. Zhang, Y.-W., Li, X.-Y., and Koyama, T. (2000) *Biochemistry* **39**, 12717–12722
54b. Oh, S. K., Han, K. H., Ryu, S. B., and Kang, H. (2000) *J. Biol. Chem.* **275**, 18482–18488
55. Matsuoka, S., Sagami, H., Kurisaki, A., and Ogura, K. (1991) *J. Biol. Chem.* **266**, 3464 – 3468
56. Keller, R. K. (1987) *Trends Biochem. Sci.* **12**, 443–445
57. Ericsson, J., Scallen, T. J., Chojnacki, T., and Dallner, G. (1991) *J. Biol. Chem.* **266**, 10602–10607
57a. Pan, J.-J., Chiou, S.-T., and Liang, P.-H. (2000) *Biochemistry* **39**, 10936–10942
57b. Fujihashi, M., Zhang, Y.-W., Higuchi, Y., Li, X.-Y., Koyama, T., and Miki, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4337–4342
58. Frank, D. W., and Waechter, C. J. (1998) *J. Biol. Chem.* **273**, 11791–11798
59. Sagami, H., Igarashi, Y., Tateyama, S., Ogura, K., Roos, J., and Lennarz, W. J. (1996) *J. Biol. Chem.* **271**, 9560–9566
60. Ohkura, T., Fukushima, K., Kurisaki, A., Sagami, H., Ogura, K., Ohno, K., Hara-Kuge, S., and Yamashita, K. (1997) *J. Biol. Chem.* **272**, 6868–6875
61. Murgolo, N. J., Patel, A., Stivala, S. S., and Wong, T. K. (1989) *Biochemistry* **28**, 253–260
62. Takahashi, I., Ogura, K., and Seto, S. (1980) *J. Biol. Chem.* **255**, 4539–4543
63. Wolucka, B. A., and de Hoffmann, E. (1995) *J. Biol. Chem.* **270**, 20151–20155
63a. Schulbach, M. C., Mahapatra, S., Macchia, M., Barontini, S., Papi, C., Minutolo, F., Bertini, S., Brennan, P. J., and Crick, D. C. (2001) *J. Biol. Chem.* **276**, 11624–11630
64. Runquist, M., Ericsson, J., Thelin, A., Chojnacki, T., and Dallner, G. (1994) *J. Biol. Chem.* **269**, 5804–5809
65. Dennis, M. S., Henzel, W. J., Bell, J., Kohr, W., and Light, D. R. (1989) *J. Biol. Chem.* **264**, 18618–18626
66. Ohnuma, S.-i, Koyama, T., and Ogura, K. (1991) *J. Biol. Chem.* **266**, 23706–23713
67. Chen, A., Zhang, D., and Poulter, C. D. (1993) *J. Biol. Chem.* **268**, 21701–21705
67a. Morii, H., Nishihara, M., and Koga, Y. (2000) *J. Biol. Chem.* **275**, 36568–36574
68. Zhang, F. L., and Casey, P. J. (1996) *Ann. Rev. Biochem.* **65**, 241–269
69. Gelb, M. H. (1997) *Science* **275**, 1750–1751
70. Vogt, A., Sun, J., Qian, Y., Tan-Chiu, E., Hamilton, A. D., and Sebti, S. M. (1995) *Biochemistry* **34**, 12398–12403
71. Mu, Y. Q., Omer, C. A., and Gibbs, R. A. (1996) *J. Am. Chem. Soc.* **118**, 1819–1823
72. Del Villar, K., Mitsuzawa, H., Yang, W., Sattler, I., and Tamanoi, F. (1997) *J. Biol. Chem.* **272**, 680–687
73. Park, H.-W., Boduluri, S. R., Moomaw, J. F., Casey, P. J., and Beese, L. S. (1997) *Science* **275**, 1800–1804
74. Mathis, J. R., and Poulter, C. D. (1997) *Biochemistry* **36**, 6367–6376
75. Dunten, P., Kammlott, U., Crowther, R., Weber, D., Palermo, R., and Birktoft, J. (1998) *Biochemistry* **37**, 7907–7912
76. Long, S. B., Casey, P. J., and Beese, L. S. (1998) *Biochemistry* **37**, 9612–9618
76a. Crespo, N. C., Ohkanda, J., Yen, T. J., Hamilton, A. D., and Sebti, S. M. (2001) *J. Biol. Chem.* **276**, 16161–16167
76b. Long, S. B., Hancock, P. J., Kral, A. M., Hellinga, H. W., and Beese, L. S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12948–12953
76c. Micali, E., Chehade, K. A. H., Isaacs, R. J., Andres, D. A., and Spielmann, H. P. (2001) *Biochemistry* **40**, 12254–12265
76d. Huang, C.-c, Hightower, K. E., and Fierke, C. A. (2000) *Biochemistry* **39**, 2593–2602
76e. Pang, Y.-P., Xu, K., El Yazal, J., and Prendergast, F. G. (2000) *Protein Sci.* **9**, 1857–1865
76f. Long, S. B., Casey, P. J., and Beese, L. S. (2002) *Nature (London)* **419**, 645–650
77. Yokoyama, K., McGeady, P., and Gelb, M. H. (1995) *Biochemistry* **34**, 1344–1354
78. Witter, D. J., and Poulter, C. D. (1996) *Biochemistry* **35**, 10454–10463
78a. Clausen, V. A., Edelstein, R. L., and Distefano, M. D. (2001) *Biochemistry* **40**, 3920–3930
78b. Dursina, B., Thomä, N. H., Sidorovitch, V., Niculae, A., Iakovenko, A., Rak, A., Albert, S., Ceacareanu, A.-C., Kölling, R., Herrmann, C., Goody, R. S., and Alexandrov, K. (2002) *Biochemistry* **41**, 6805–6816
79. Fu, H.-W., Beese, L. S., and Casey, P. J. (1998) *Biochemistry* **37**, 4465–4472
79a. Ashar, H. R., James, L., Gray, K., Carr, D., Black, S., Armstrong, L., Bishop, W. R., and Kirschmeier, P. (2000) *J. Biol. Chem.* **275**, 30451–30457
79b. Huber, H. E., Robinson, R. G., Watkins, A., Nahas, D. D., Abrams, M. T., Buser, C. A., Lobell, R. B., Patrick, D., Anthony, N. J., Dinsmore, C. J., Graham, S. L., Hartman, G. D., Lumma, W. C., Williams, T. M., and Heimbrook, D. C. (2001) *J. Biol. Chem.* **276**, 24457–24465
80. Leung, H.-C. E., Chen, Y., and Winkler, M. E. (1997) *J. Biol. Chem.* **272**, 13073–13083
81. Gebler, J. C., Woodside, A. B., and Poulter, C. D. (1992) *J. Am. Chem. Soc.* **114**, 7354–7360
82. Clause, E. P., Tyler, V. E., and Brady, L. R. (1970) *Pharmacognosy*, 6th ed., Lea & Febiger, Philadelphia, Pennsylvania
83. Loreto, F., Ciccioli, P., Brancaleoni, E., Cecinato, A., Frattoni, M., and Sharkey, T. D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9966–9969
84. Harborne, J. B., and Tomas-Barberan, F. A., eds. (1991) *Ecological Chemistry and Biochemistry of Plant Terpenoids*, Clarendon Press, Oxford
85. Gershenzon, J., and Croteau, R. B. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 339–388, CRC Press, Boca Raton, Florida
86. Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4126–4133
87. Bohlmann, J., Steele, C. L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784–21792
88. Wise, M. L., Savage, T. J., Katahira, E., and Croteau, R. (1998) *J. Biol. Chem.* **273**, 14891–14899
88a. Williams, D. C., McGarvey, D. J., Katahira, E. J., and Croteau, R. (1998) *Biochemistry* **37**, 12213–12220
89. Croteau, R., and Satterwhite, D. M. (1989) *J. Biol. Chem.* **264**, 15309–15315
90. Seybold, S. J., Quilici, D. R., Tillman, J. A., Vanderwel, D., Wood, D. L., and Blomquist, G. J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8393–8397
91. Steele, C. L., Crock, J., Bohlmann, J., and Croteau, R. (1998) *J. Biol. Chem.* **273**, 2078–2089
91a. Rising, K. A., Starks, C. M., Noel, J. P., and Chappell, J. (2000) *J. Am. Chem. Soc.* **122**, 1861–1866
92. Procter, R. H., and Hohn, T. M. (1993) *J. Biol. Chem.* **268**, 4543–4548
93. Cane, D. E., and Tsantrizos, Y. S. (1996) *J. Am. Chem. Soc.* **118**, 10037–10040
93a. Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2000) *J. Biol. Chem.* **275**, 25533–25539
94. Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270**, 7375–7381
95. Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997) *Science* **277**, 1815–1820
96. Lesburg, C. A., Zhai, G., Cane, D. E., and Christianson, D. W. (1997) *Science* **277**, 1820–1824
97. Cane, D. E., and Xue, Q. (1996) *J. Am. Chem. Soc.* **118**, 1563–1564
97a. Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2002) *Biochemistry* **41**, 1732–1741

References

97b. Peters, R. J., and Croteau, R. B. (2002) *Biochemistry* **41**, 1836–1842
98. Crock, J., Wildung, M., and Croteau, R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12833–12838
99. Marin, E., Nussaume, L., Quesada, A., Gonneau, M., Sotta, B., Hugueney, P., Frey, A., and Marion-Poll, A. (1996) *EMBO J.* **15**, 2331–2342
99a. Qin, X., and Zeevaart, J. A. D. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15354–15361
99b. Seo, M., Peeters, A. J. M., Koiwai, H., Oritani, T., Marion-Poll, A., Zeevaart, J. A. D., Koornneef, M., Kamiya, Y., and Koshiba, T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12908–12913
100. He, X., Saint-Jeannet, J.-P., Woodgett, J. R., Varmus, H. E., and Dawid, I. B. (1995) *Nature (London)* **374**, 617–622
101. Spray, C. R., Kobayashi, M., Suzuki, Y., Phinney, B. O., Gaskin, P., and MacMillan, J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10515–10518
102. Xu, Y.-L., Li, L., Wu, K., Peeters, A. J. M., Gage, D. A., and Zeevaart, J. A. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6640–6644
102a. Rojas, M. C., Hedden, P., Gaskin, P., and Tudzynski, B. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5838–5843
102b. Kawaide, H., Sassa, T., and Kamiya, Y. (2000) *J. Biol. Chem.* **275**, 2276–2280
103. Lange, T., Hedden, P., and Graebe, J. E. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8552–8556
103a. Helliwell, C. A., Chandler, P. M., Poole, A., Dennis, E. S., and Peacock, W. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2065–2070
103b. Thomas, S. G., Phillips, A. L., and Hedden, P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4698–4703
104. Feyereisen, R., and Farnsworth, D. E. (1987) *J. Biol. Chem.* **262**, 2676–2681
105. Toong, Y. C., Schooley, D. A., and Baker, F. C. (1988) *Nature (London)* **333**, 170–171
106. Bohlmann, J., Crock, J., Jetter, R., and Croteau, R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6756–6761
107. Vogel, B. S., Wildung, M. R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**, 23262–23268
107a. Peters, R. J., and Croteau, R. B. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 580–584
108. Mau, C. J. D., and West, C. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8497–8501
109. Lin, X., Hezari, M., Koepp, A. E., Floss, H. G., and Croteau, R. (1996) *Biochemistry* **35**, 2968–2977
110. Goodwin, T. W., ed. (1988) *Plant Pigments*, Academic Press, London
111. Mookhtiar, K. A., Kalinowski, S. S., Zhang, D., and Poulter, C. D. (1994) *J. Biol. Chem.* **269**, 11201–11207
112. Gu, P., Ishii, Y., Spencer, T. A., and Shechter, I. (1998) *J. Biol. Chem.* **273**, 12515–12525
113. Jarstfer, M. B., Blagg, B. S. J., Rogers, D. H., and Poulter, C. D. (1996) *J. Am. Chem. Soc.* **118**, 13089–13090
114. Zurer, P. (1997) *Chem. Eng. News* January 6, 6–7
114a. Radisky, E. S., and Poulter, C. D. (2000) *Biochemistry* **39**, 1748–1760
114b. Jarstfer, M. B., Zhang, D.-L., and Poulter, C. D. (2002) *J. Am. Chem. Soc.* **124**, 8834–8845
115. Musico, F., Carlson, J. P., Kuehl, L., and Rilling, H. C. (1974) *J. Biol. Chem.* **249**, 3746–3749
116. Kojima, Y., Friedlander, E. J., and Bloch, K. (1981) *J. Biol. Chem.* **256**, 7235–7239
117. Torssell, K. B. G. (1983) *Natural Product Chemistry*, Wiley, New York (pp. 181–215)
118. Armstrong, G. A., and Hearst, J. E. (1996) *FASEB J.* **10**, 228–237
118a. Rivera, S. B., Swedlund, B. D., King, G. J., Bell, R. N., Hussey, C. E., Jr., Shattuck-Eidens, D. M., Wrobel, W. M., Peiser, G. D., and Poulter, C. D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4373–4378
118b. Soderberg, T., Chen, A., and Poulter, C. D. (2001) *Biochemistry* **40**, 14847–14854
119. Heathcock, D. H., Finkelstein, B. L., Aoki, T., and Poulter, C. D. (1985) *Science* **229**, 862–863
120. Poulter, C. D., Aoki, T., and Daniels, L. (1988) *J. Am. Chem. Soc.* **110**, 2620–2624
121. Britton, G. (1988) in *Plant Pigments* (Goodwin, T. W., ed), pp. 133–182, Academic Press, London
122. Chamovitz, D., Sandmann, G., and Hirschberg, J. (1993) *J. Biol. Chem.* **268**, 17348–17353
123. Bartley, G. E., Viitanen, P. V., Pecker, I., Chamovitz, D., Hirschberg, J., and Scolnik, P. A. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6532–6536
123a. Arrach, N., Fernández-Martín, R., Cerdá-Olmedo, E., and Avalos, J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1687–1692
123b. Cunningham, F. X., Jr., and Gantt, E. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2905–2910
124. Armstrong, G. A., Schmidt, A., Sandmann, G., and Hearst, J. E. (1990) *J. Biol. Chem.* **265**, 8329–8338
125. Komori, M., Ghosh, R., Takaichi, S., Hu, Y., Mizoguchi, T., Koyama, Y., and Kuki, M. (1998) *Biochemistry* **37**, 8987–8994
126. Ouchane, S., Picaud, M., Vernotte, C., Reiss-Husson, F., and Astier, C. (1997) *J. Biol. Chem.* **272**, 1670–1676
127. Langer, M., Pauling, A., and Rétey, J. (1995) *Angew. Chem. Int. Ed. Engl.* **34**, 1464–1465
127a. Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., and Potrykus, I. (2000) *Science* **287**, 303–305
127b. A contribution of $40 to UNICEF (www.unicefusa.com) will provide vitamin A for a year to 1000 young children, protecting them from blindness.
128. Demmig-Adams, B., Gilmore, A. M., and Adams, W. W., III. (1996) *FASEB J.* **10**, 403–412
129. Sun, Z., Gantt, E., and Cunningham, F. X., Jr. (1996) *J. Biol. Chem.* **271**, 24349–24352
130. Fraser, P. D., Miura, Y., and Misawa, N. (1997) *J. Biol. Chem.* **272**, 6128–6135
131. Bouvier, F., d'Harlingue, A., Hugueney, P., Marin, E., Marion-Poll, A., and Camara, B. (1996) *J. Biol. Chem.* **271**, 28861–28867
132. Schroeder, W. A., and Johnson, E. A. (1995) *J. Biol. Chem.* **270**, 18374–18379
133. Nelis, H. J. C. F., Lavens, P., Moens, L., Sorgeloos, P., Jonckheere, J. A., Criel, G. R., and DeLeenheer, A. P. (1984) *J. Biol. Chem.* **259**, 6063–6066
134. Swift, I. E., and Milborrow, B. V. (1981) *J. Biol. Chem.* **256**, 11607–11611
135. Searcy, D. G. (1982) *Trends Biochem. Sci.* **7**, 183–185
136. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4576–4579
137. Olson, J. A., and Krinsky, N. I. (1995) *FASEB J.* **9**, 1547–1550
138. Britton, G. (1995) *FASEB J.* **9**, 1551–1558
139. Zuber, H. (1986) *Trends Biochem. Sci.* **11**, 414–419
140. Quiñones, M. A., Lu, Z., and Zeiger, E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2224–2228
141. Bugos, R. C., and Yamamoto, H. Y. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6320–6325
142. Rock, C. D., and Zeevaart, J. A. D. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7496–7499
143. Schwartz, S. H., Tan, B. C., Gage, D. A., Zeevaart, J. A. D., and McCarty, D. R. (1997) *Science* **276**, 1872–1874
144. Kennedy, T. A., and Liebler, D. C. (1992) *J. Biol. Chem.* **267**, 4658–4663
145. Everett, S. A., Dennis, M. F., Patel, K. B., Maddix, S., Kundu, S. C., and Willson, R. L. (1996) *J. Biol. Chem.* **271**, 3988–3994
146. Burton, G. W., and Ingold, K. U. (1984) *Science* **224**, 569–573
147. Brush, A. H. (1990) *FASEB J.* **4**, 2969–2977
148. Weesie, R. J., Jansen, F. J. H. M., Merlin, J. C., Lugtenburg, J., Britton, G., and de Groot, H. J. M. (1997) *Biochemistry* **36**, 7288–7296
149. Parker, R. S. (1996) *FASEB J.* **10**, 542–551
150. Hoffman, M. (1990) *Science* **250**, 372–373
151. Gudas, L. J. (1994) *J. Biol. Chem.* **269**, 15399–15402
152. Means, A. L., and Gudas, L. J. (1995) *Ann. Rev. Biochem.* **64**, 201–233
152a. Mertz, J. R., Shang, E., Piantedosi, R., Wei, S., Wolgemuth, D. J., and Blaner, W. S. (1997) *J. Biol. Chem.* **272**, 11744–11749
152b. White, J. A., Guo, Y.-D., Baetz, K., Beckett-Jones, B., Bonasoro, J., Hsu, K. E., Dilworth, F. J., Jones, G., and Petkovich, M. (1996) *J. Biol. Chem.* **271**, 29922–29927
152c. Lamb, A. L., and Newcomer, M. E. (1999) *Biochemistry* **38**, 6003–6011
153. Parish, E. J., and Nes, W. D., eds. (1997) *Biochemistry and Function of Sterols*, CRC Press, Boca Raton, Florida
153a. Crowder, C. M., Westover, E. J., Kumar, A. S., Ostlund, R. E., Jr., and Covey, D. F. (2001) *J. Biol. Chem.* **276**, 44369–44372
154. van Tamelen, E. E. (1968) *Acc. Chem. Res.* **1**, 111–120
155. Heftmann, E. (1969) *Steroid Biochemistry*, Academic Press, New York
156. Templeton, W. (1969) *An Introduction to the Chemistry of Terpenoids and Steroids*, Butterworths, London
157. Abe, I., and Prestwich, G. D. (1994) *J. Biol. Chem.* **269**, 802–804
158. Abe, I., and Prestwich, G. D. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9274–9278
159. Shi, Z., Buntel, C. J., and Griffin, J. H. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7370–7374
160. Nes, W. D., and Venkatramesh, M. (1994) in *Isopentenoids and Other Natural Products* (Nes, W. D., ed), pp. 55–89, American Chemical Society, Washington D.C.
160a. Edwards, P. A., and Ericsson, J. (1999) *Ann. Rev. Biochem.* **68**, 157–185
161. Goad, L. J. (1991) in *Ecological Chemistry and Biochemistry of Plant Terpenoids* (Harborne, J. B., and Tomas-Barberan, F. A., eds), pp. 209–229, Clarendon Press, Oxford
162. Nes, W. D., Parker, S. R., Crumley, F. G., and Ross, S. A. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 389–426, CRC Press, Boca Raton, Florida
162a. Herrera, J. B. R., Wilson, W. K., and Matsuda, S. P. T. (2000) *J. Am. Chem. Soc.* **122**, 6765–6766
163. Nes, W. D. (1990) in *Biochemistry of the Mevalonic Acid Pathway to Terpenoids* (Towers, G. H. N., and Stafford, H. A., eds), pp. 283–327, Plenum, New York
163a. Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000) *J. Am. Chem. Soc.* **122**, 6816–6824
163b. Heathcock, C. H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14323–14327
164. Wendt, K. U., Poralla, K., and Schulz, G. E. (1997) *Science* **277**, 1811–1815
164a. Wendt, K. U., Lenhart, A., and Schulz, G. E. (1999) *J. Mol. Biol.* **286**, 175–187
165. Prince, R. C. (1987) *Trends Biochem. Sci.* **12**, 455–456

References

166. Rohmer, M., and Bisseret, P. (1994) in *Isopentenoids and Other Natural Products* (Nes, W. D., ed), pp. 31–43, American Chemical Society, Washington D.C.
167. Fukushima, H., Grinstead, G. F., and Gaylor, J. L. (1981) *J. Biol. Chem.* **256**, 4822–4826
168. Krieger, M., Kingsley, D., Sege, R., Hobbie, L., and Kozarsky, K. (1985) *Trends Biochem. Sci.* **10**, 447–452
169. Williams, J. B., and Napoli, J. L. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4658–4662
170. Lamb, D. C., Kelly, D. E., Schunck, W.-H., Shyadehi, A. Z., Akhtar, M., Lowe, D. J., Baldwin, B. C., and Kelly, S. L. (1997) *J. Biol. Chem.* **272**, 5682–5688
171. Trzaskos, J. M., Ko, S. S., Magolda, R. L., Favata, M. F., Fischer, R. T., Stam, S. H., Johnson, P. R., and Gaylor, J. L. (1995) *Biochemistry* **34**, 9670–9676
172. Shyadehi, A. Z., Lamb, D. C., Kelly, S. L., Kelly, D. E., Schunck, W.-H., Wright, J. N., Corina, D., and Akhtar, M. (1996) *J. Biol. Chem.* **271**, 12445–12450
172a. Podust, L. M., Poulos, T. L., and Waterman, M. R. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3068–3073
172b. Bellamine, A., Mangla, A. T., Nes, W. D., and Waterman, M. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8937–8942
173. Bard, M., Bruner, D. A., Pierson, C. A., Lees, N. D., Biermann, B., Frye, L., and Koegel, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 186–190
173a. Taton, M., Husselstein, T., Benveniste, P., and Rahier, A. (2000) *Biochemistry* **39**, 701–711
174. Seedorf, U., Brysch, P., Engel, T., Schrage, K., and Assmann, G. (1994) *J. Biol. Chem.* **269**, 21277–21283
174a. Gachotte, D., Barbuch, R., Gaylor, J., Nickel, E., and Bard, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13794–13799
175. Pascal, S., Taton, M., and Rahier, A. (1993) *J. Biol. Chem.* **268**, 11639–11654
176. Rahier, A., and Taton, M. (1996) *Biochemistry* **35**, 7069–7076
177. Nes, W. D., Janssen, G. G., and Bergenstrahle, A. (1991) *J. Biol. Chem.* **266**, 15202–15212
178. Shi, J., Gonzalas, R. A., and Bhattacharyya, M. K. (1996) *J. Biol. Chem.* **271**, 9384–9389
179. Russell, D. W. (1996) *Science* **272**, 370–371
179a. Rouleau, M., Marsolais, F., Richard, M., Nicolle, L., Voigt, B., Adam, G., and Varin, L. (1999) *J. Biol. Chem.* **274**, 20925–20930
180. Li, J., Biswas, M. G., Chao, A., Russell, D. W., and Chory, J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3554–3559
181. Metcalf, R. L., Metcalf, R. A., and Rhodes, A. M. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3769–3772
182. Russell, D. W., and Setchell, K. D. R. (1992) *Biochemistry* **31**, 4737–4749
183. Lange, Y., and Steck, T. L. (1994) *J. Biol. Chem.* **269**, 29371–29374
184. Liscum, L., and Underwood, K. W. (1995) *J. Biol. Chem.* **270**, 15443–15446
185. Guan, G., Dai, P.-H., Osborne, T. F., Kim, J. B., and Shechter, I. (1997) *J. Biol. Chem.* **272**, 10295–10302
186. Mackay, K., Starr, J. R., Lawn, R. M., and Ellsworth, J. L. (1997) *J. Biol. Chem.* **272**, 13380–13389
187. Lange, Y., and Muraski, M. F. (1988) *J. Biol. Chem.* **263**, 9366–9373
188. Bretscher, M. S., and Munro, S. (1993) *Science* **261**, 1280–1281
188a. Simons, K., and Ikonen, E. (2000) *Science* **290**, 1721–1726
188b. Nilsson, I., Ohvo-Rekilä, H., Slotte, J. P., Johnson, A. E., and von Heijne, G. (2001) *J. Biol. Chem.* **276**, 41748–41754
188c. Parpal, S., Karlsson, M., Thorn, H., and Strålfors, P. (2001) *J. Biol. Chem.* **276**, 9670–9678
189. Lange, Y., Ye, J., and Steck, T. L. (1998) *J. Biol. Chem.* **273**, 18915–18922
190. Puglielli, L., Rigotti, A., Greco, A. V., Santos, M. J., and Nervi, F. (1995) *J. Biol. Chem.* **270**, 18723–18726
190a. López García, F., Szyperski, T., Dyer, J. H., Choinowski, T., Seedorf, U., Hauser, H., and Wüthrich, K. (2000) *J. Mol. Biol.* **295**, 595–603
190b. Schroeder, F., Frolov, A., Starodub, O., Atshaves, B. B., Russell, W., Petrescu, A., Huang, H., Gallegos, A. M., McIntosh, A., Tahotna, D., Russell, D. H., Billheimer, J. T., Baum, C. L., and Kier, A. B. (2000) *J. Biol. Chem.* **275**, 25547–25555
190c. Choinowski, T., Hauser, H., and Piontek, K. (2000) *Biochemistry* **39**, 1897–1902
191. Yang, H., Bard, M., Bruner, D. A., Gleeson, A., Deckelbaum, R. J., Aljinovic, G., Pohl, T. M., Rothstein, R., and Sturley, S. L. (1996) *Science* **272**, 1353–1356
191a. Spady, D. K., Willard, M. N., and Meidell, R. S. (2000) *J. Biol. Chem.* **275**, 27005–27012
192. Jauhiainen, M., and Dolphin, P. J. (1986) *J. Biol. Chem.* **261**, 7032–7043
192a. Seo, T., Oelkers, P. M., Giattina, M. R., Worgall, T. S., Sturley, S. L., and Deckelbaum, R. J. (2001) *Biochemistry* **40**, 4756–4762
192b. Mauch, D. H., Nägler, K., Schumacher, S., Göritz, C., Müller, E.-C., Otto, A., and Pfrieger, F. W. (2001) *Science* **294**, 1354–1357
192c. Barres, B. A., and Smith, S. J. (2001) *Science* **294**, 1296–1297
193. Jiang, X.-c, and Bruce, C. (1995) *J. Biol. Chem.* **270**, 17133–17138
194. Agellon, L. B., Quinet, E. M., Gillette, T. G., Drayna, D. T., Brown, M. L., and Tall, A. R. (1990) *Biochemistry* **29**, 1372–1376
195. Epps, D. E., Greenlee, K. A., Harris, J. S., Thomas, E. W., Castle, C. K., Fisher, J. F., Hozak, R. R., Marschke, C. K., Melchior, G. W., and Kézdy, F. J. (1995) *Biochemistry* **34**, 12560–12569
195a. Föger, B., Chase, M., Amar, M. J., Vaisman, B. L., Shamburek, R. D., Paigen, B., Fruchart-Najib, J., Paiz, J. A., Koch, C. A., Hoyt, R. F., Brewer, H. B., Jr., and Santamarina-Fojo, S. (1999) *J. Biol. Chem.* **274**, 36912–36920
196. Tall, A. (1995) *Ann. Rev. Biochem.* **64**, 235–257
196a. Desrumaux, C., Labeur, C., Verhee, A., Tavernier, J., Vandekerckhove, J., Rosseneu, M., and Peelman, F. (2001) *J. Biol. Chem.* **276**, 5908–5915
197. Feaster, S. R., Quinn, D. M., and Barnett, B. L. (1997) *Protein Sci.* **6**, 73–79
198. Lopez-Candales, A., Bosner, M. S., Spilburg, C. A., and Lange, L. G. (1993) *Biochemistry* **32**, 12085–12089
199. Sutton, L. D., Froelich, S., Hendrickson, H. S., and Quinn, D. M. (1991) *Biochemistry* **30**, 5888–5893
200. Brown, M. S., and Goldstein, J. L. (1986) *Science* **232**, 34–47
201. Motulsky, A. G. (1986) *Science* **231**, 126–128
202. Brown, M. S., and Goldstein, J. L. (1983) *Ann. Rev. Biochem.* **52**, 223–261
203. Griffiths, G., and Simons, K. (1986) *Science* **234**, 438–443
204. Golgi, C. (1898) *Arch. Ital. Biol.* **30**, 60 and 278
205. Breslow, J. L. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2031–2052, McGraw-Hill, New York
206. Dietschy, J. M. (1997) *Am. J. Clin. Nutr.* **65**, 1581S–1589S
207. Krieger, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4077–4080
208. Lopez, J., Roghani, A., Bertrand, J., Zanni, E., Kalopissis, A., Zannis, V. I., and Chambaz, J. (1994) *Biochemistry* **33**, 4056–4064
209. Bergeron, J., Frank, P. G., Scales, D., Meng, Q.-H., Castro, G., and Marcel, Y. L. (1995) *J. Biol. Chem.* **270**, 27429–27438
210. Gong, E. L., Stoltzfus, L. J., Brion, C. M., Murugesh, D., and Rubin, E. M. (1996) *J. Biol. Chem.* **271**, 5984–5987
211. Südhof, T. C., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1985) *Science* **228**, 815–822
212. Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroon, P. A., and Smith, R. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6334–6338
213. Kozarsky, K. F., Donahee, M. H., Rigotti, A., Iqbal, S. N., Edelman, E. R., and Krieger, M. (1997) *Nature (London)* **387**, 414–417
213a. Krieger, M. (1999) *Ann. Rev. Biochem.* **68**, 523–558
214. Brown, M. S., and Goldstein, J. L. (1974) *Sci. Am.* **251**(Nov), 58–66
215. Lehrman, M. A., Russell, D. W., Goldstein, J. L., and Brown, M. S. (1987) *J. Biol. Chem.* **262**, 3354–3361
216. Strickland, D. K., Kounnas, M. Z., and Argraves, W. S. (1995) *FASEB J.* **9**, 890–898
217. Chappell, D. A., Inoue, I., Fry, G. L., Pladet, M. W., Bowen, S. L., Iverius, P.-H., Lalouel, J.-M., and Strickland, D. K. (1994) *J. Biol. Chem.* **269**, 18001–18006
218. Rosseneu, M., and Labeur, C. (1995) *FASEB J.* **9**, 768–776
219. Sakai, J., Hoshino, A., Takahashi, S., Miura, Y., Ishii, H., Suzuki, H., Kawarabayasi, Y., and Yamamoto, T. (1994) *J. Biol. Chem.* **269**, 2173–2182
220. Takahashi, S., Suzuki, J., Kohno, M., Oida, K., Tamai, T., Miyabo, S., Yamamoto, T., and Nakai, T. (1995) *J. Biol. Chem.* **270**, 15747–15754
221. Krieger, M., and Herz, J. (1994) *Ann. Rev. Biochem.* **63**, 601–637
222. Doi, T., Kurasawa, M., Higashino, K.-i, Imanishi, T., Mori, T., Naito, M., Takahashi, K., Kawabe, Y., Wada, Y., Matsumoto, A., and Kodama, T. (1994) *J. Biol. Chem.* **269**, 25598–25604
223. Krieger, M., Acton, S., Ashkenas, J., Pearson, A., Penman, M., and Resnick, D. (1993) *J. Biol. Chem.* **268**, 4569–4572
224. Krieger, M. (1992) *Trends Biochem. Sci.* **17**, 141–146
225. Hajjar, D. P., and Haberland, M. E. (1997) *J. Biol. Chem.* **272**, 22975–22978
226. Schaefer, E. J., and Levy, R. I. (1985) *N. Engl. J. Med.* **312**, 1300–1310
227. Breslow, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8314–8318
228. Havel, R. J., and Kane, J. P. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1841–1852, McGraw-Hill, New York
229. Goldstein, J. L., Hobbs, H. H., and Brown, M. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1981–2030, McGraw-Hill, New York
229a. Goldstein, J. L., and Brown, M. S. (2001) *Science* **292**, 1310–1312
229b. Garcia, C. K., and 12 other authors. (2001) *Science* **292**, 1394–1398
230. Fass, D., Blacklow, S., Kim, P. S., and Berger, J. M. (1997) *Nature (London)* **388**, 691–693
230a. Borén, J., Ekström, U., Ågren, B., Nilsson-Ehle, P., and Innerarity, T. L. (2001) *J. Biol. Chem.* **276**, 9214–9218

References

231. Glomset, J. A., Assmann, G., Gjone, E., and Norum, K. R. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1933–1952, McGraw-Hill, New York
232. Pennisi, E. (1997) *Science* **277**, 180–181
233. Carstea, E. D., and 37 other authors. (1997) *Science* **277**, 228–231
234. Loftus, S. K., Morris, J. A., Carstea, E. D., Gu, J. Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M. A., Tagle, D. A., Pentchev, P. G., and Pavan, W. J. (1997) *Science* **277**, 232–235
234a. Davies, J. P., Ioannou,Y. A. (2000) *J. Biol. Chem.* **275**, 24367–24374
234b. Davies, J. P., Chen, F. W., and Ioannou, Y. A. (2000) *Science* **290**, 2295–2298
234c. Naureckiene, S., Sleat, D. E., Lackland, H., Fensom, A., Vanier, M. T., Wattiaux, R., Jadot, M., and Lobel, P. (2000) *Science* **290**, 2298–2301
235. Assmann, G., von Eckardstein, A., and Brewer, H. B., Jr. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2053–2072, McGraw-Hill, New York
236. Lohse, P., Kindt, M. R., Rader, D. J., and Brewer, H. B., Jr. (1991) *J. Biol. Chem.* **266**, 13513–13518
236a. Gura, T. (1999) *Science* **285**, 814–815
236b. Scott, J. (1999) *Nature (London)* **400**, 816–819
236c. Liscovitch, M., and Lavie, Y. (2000) *Trends Biochem. Sci.* **25**, 530–534
236d. Berge, K. E., Tian, H., Graf, G. A., Yu, L., Grishin, N. V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R., and Hobbs, H. H. (2000) *Science* **290**, 1771–1775
236e. Allayee, H., Laffitte, B. A., and Lusis, A. J. (2000) *Science* **290**, 1709–1711
237. Bloch, K. (1982) *Trends Biochem. Sci.* **7**, 334–336
238. Tanaka, N., Nonaka, T., Tanabe, T., Yoshimoto, T., Tsuru, D., and Mitsui, Y. (1996) *Biochemistry* **35**, 7715–7730
238a. Kramer, W., Sauber, K., Baringhaus, K.-H., Kurz, M., Stengelin, S., Lange, G., Corsiero, D., Girbig, F., König, W., and Weyland, C. (2001) *J. Biol. Chem.* **276**, 7291–7301
238b. Lundell, K., Hansson, R., and Wikvall, K. (2001) *J. Biol. Chem.* **276**, 9606–9612
238c. del Castillo-Olivares, A., and Gil, G. (2000) *J. Biol. Chem.* **275**, 17793–17799
238d. Steinberg, S. J., Mihalik, S. J., Kim, D. G., Cuebas, D. A., and Watkins, P. A. (2000) *J. Biol. Chem.* **275**, 15605–15608
238e. Kotti, T. J., Savolainen, K., Helander, H. M., Yagi, A., Novikov, D. K., Kalkkinen, N., Conzelmann, E., Hiltunen, J. K., and Schmitz, W. (2000) *J. Biol. Chem.* **275**, 20887–20895
238f. Pullinger, C. R., Eng, C., Salen, G., Shefer, S., Batta, A. K., Erickson, S. K., Verhagen, A., Rivera, C. R., Mulvihill, S. J., Malloy, M. J., and Kane, J. P. (2002) *J. Clin. Invest.* **110**, 109–117
239. Berginer, V. M., Salem, G., and Shefer, S. (1984) *N. Engl. J. Med.* **311**, 1649–1652
240. Björkhem, I., and Boberg, K. M. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2073–2099, McGraw-Hill, New York
241. de Vree, J. M. L., Jacquemin, E., Sturm, E., Cresteil, D., Bosma, P. J., Aten, J., Deleuze, J.-F., Desrochers, M., Burdelski, M., Bernard, O., Oude Elferink, R. P. J., and Hadchouel, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 282–287
242. Breuer, O., and Björkhem, I. (1995) *J. Biol. Chem.* **270**, 20278–20284
243. Korytowski, W., Geiger, P. G., and Girotti, A. W. (1996) *Biochemistry* **35**, 8670–8679
244. Javitt, N. B., Kok, E., Burstein, S., Cohen, B., and Kutscher, J. (1981) *J. Biol. Chem.* **256**, 12644–12646
245. Anderson, C. J., Lucas, L. J. H., and Widlanski, T. S. (1995) *J. Am. Chem. Soc.* **117**, 3889–3890
246. Ballabio, A., Parenti, G., Carrozzo, R., Sebastio, G., Andria, G., Buckle, V., Fraser, N., Craig, I., Rocchi, M., Romeo, G., Jobsis, A. C., and Persico, M. G. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4519–4523
247. Beaulieu, M., Lévesque, E., Hum, D. W., and Bélanger, A. (1996) *J. Biol. Chem.* **271**, 22855–22862
248. Porter, J. A., Young, K. E., and Beachy, P. A. (1996) *Science* **274**, 255–258
249. Briggs, M. H., and Brotherton, J. (1970) *Steriod Biochemistry and Pharmacology*, Academic Press, New York
250. Waterman, M. R., and Bischof, L. J. (1997) *FASEB J.* **11**, 419–427
251. Waterman, M. R. (1995) *Science* **267**, 1780–1781
252. Lin, D., Sugawara, T., Strauss, J. F., III, Clark, B. J., Stocco, D. M., Saenger, P., Rogol, A., and Miller, W. L. (1995) *Science* **267**, 1828–1831
253. Heyl, B. L., Tyrrell, D. J., and Lambeth, J. D. (1986) *J. Biol. Chem.* **261**, 2743–2749
254. Rhéaume, E., Sanchez, R., Mébarki, F., Gagnon, E., Carel, J.-C., Chaussain, J.-L., Morel, Y., Labrie, F., and Simard, J. (1995) *Biochemistry* **34**, 2893–2900
255. Bain, P. A., Yoo, M., Clarke, T., Hammond, S. H., and Payne, A. H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8870–8874
256. Borg, W., Shackleton, C. H. L., Pahuja, S. L., and Hochberg, R. B. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1545–1549
256a. Avvakumov, G. V., Muller, Y. A., and Hammond, G. L. (2000) *J. Biol. Chem.* **275**, 25920–25925
256b. Grishkovskaya, I., Avvakumov, G. V., Sklenar, G., Dales, D., Hammond, G. L., and Muller, Y. A. (2000) *EMBO J.* **19**, 504–512
257. Mester, J., and Baulieu, E.-E. (1984) *Trends Biochem. Sci.* **9**, 56–59
258. Baulieu, E.-E. (1989) *Science* **245**, 1351–1357
259. Ulmann, A., Teutsch, G., and Philibert, D. (1990) *Sci. Am.* **262**(Jun), 42–48
260. Beck, C. A., Weigel, N. L., Moyer, M. L., Nordeen, S. K., and Edwards, D. P. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4441–4445
261. Glasier, A., Thong, K. J., Dewar, M., Mackie, M., and Baird, D. T. (1992) *N. Engl. J. Med.* **327**, 1041–1044
262. Grimes, D. A., and Cook, R. J. (1992) *N. Engl. J. Med.* **327**, 1088–1089
263. Brann, D. W., and Mahesh, V. B. (1991) *FASEB J.* **5**, 2691–2698
264. Narasimhulu, S., Eddy, C. R., Dibartolomeis, M., Kowluru, R., and Jefcoate, C. R. (1985) *Biochemistry* **24**, 4287–4294
265. White, P. C. (1994) *N. Engl. J. Med.* **331**, 250–258
266. Phillips, I. R., and Shepard, E. A. (1985) *Nature (London)* **314**, 130–131
267. Goldsworthy, G. J., Robinson, J., and Mordue, W. (1981) *Endocrimology*, Blackie, Glascow
268. Makin, H. L. J., ed. (1984) *Biochemistry of the Steroid Hormones*, 2nd ed., Blackwell, Oxford
269. New, M. I., ed. (1985) *Congenital Adrenal Hyperplasia (Ann. N. York Acad. Sci.)*, Vol. 458, New York Acad. Sci., New York
270. Chrousos, G. P. (1995) *N. Engl. J. Med.* **332**, 1351–1362
271. Hatakeyama, H., Miyamori, I., Fujita, T., Takeda, Y., Takeda, R., and Yamamoto, H. (1994) *J. Biol. Chem.* **269**, 24316–24320
272. McEwen, B. (1981) *Nature (London)* **291**, 610
273. Moore, R. J., Griffin, J. E., and Wilson, J. D. (1975) *J. Biol. Chem.* **250**, 7168–7172
274. Bull, H. G., Garcia-Calvo, M., Andersson, S., Baginsky, W. F., Chan, H. K., Ellsworth, D. E., Miller, R. R., Stearns, R. A., Bakshi, R. K., Rasmusson, G. H., Tolman, R. L., Myers, R. W., Kozarich, J. W., and Harris, G. S. (1996) *J. Am. Chem. Soc.* **118**, 2359–2365
275. Russell, D. W., and Wilson, J. D. (1994) *Ann. Rev. Biochem.* **63**, 25–61
276. Lin, T.-M., and Chang, C. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4988–4993
277. Rittmaster, R. S. (1994) *N. Engl. J. Med.* **330**, 120–125
278. Gormley, G. J., Stoner, E., Bruskewitz, R. C., Imperato-McGinley, J., Walsh, P. C., McConnell, J. D., Andriole, G. L., Geller, J., Bracken, B. R., Tenover, J. S., Vaughan, E. D., Pappas, F., Taylor, A., Binkowitz, B., and Ng, J. (1992) *N. Engl. J. Med.* **327**, 1185–1191
279. Baulieu, E.-E., and Robel, P. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4089–4091
280. Lardy, H., Partridge, B., Kneer, N., and Wei, Y. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6617–6619
281. Berdanier, C. D., Parente, J. A., Jr., and Mcintosh, M. K. (1993) *FASEB J.* **7**, 414–419
282. Nestler, J. E., Clore, J. N., and Blackard, W. G. (1992) *FASEB J.* **6**, 3073–3075
283. Berr, C., Lafont, S., Debuire, B., Dartigues, J.-F., and Baulieu, E.-E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13410–13415
284. Rose, K. A., Stapleton, G., Dott, K., Kieny, M. P., Best, R., Schwarz, M., Russell, D. W., Björkhem, I., Seckl, J., and Lathe, R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4925–4930
285. Wickelgren, I. (1997) *Science* **276**, 675–678
285a. Yen, S. S. C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8167–8169
285b. Baulieu, E.-E., Thomas, G., Legrain, S., Lahlou, N., Roger, M., Debuire, B., Faucounau, V., Girard, L., Hervy, M.-P., Latour, F., Leaud, M.-C., Mokrane, A., Pitti-Ferrandi, H., and 10 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4279–4284
285c. Rose, K., Allan, A., Gauldie, S., Stapleton, G., Dobbie, L., Dott, K., Martin, C., Wang, L., Hedlund, E., Seckl, J. R., Gustafsson, J.-Å., and Lathe, R. (2001) *J. Biol. Chem.* **276**, 23937–23944
286. Cascio, C., Prasad, V. V. K., Lin, Y. Y., Lieberman, S., and Papadopoulos, V. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2862–2867
287. Kochakian, C. D. (1989) *Foundations of Endocrinology*, Univ. of Alabama School of Medicine, Tuscaloosa
288. Kochakian, C. D. (1987) *Trends Biochem. Sci.* **12**, 446–450
289. Bardin, C. W. (1996) *N. Engl. J. Med.* **335**, 52
290. Matsunaga, T., Nomoto, M., Kozak, C. A., and Gonzalez, F. J. (1990) *Biochemistry* **29**, 1329–1341
290a. Yamada, A., Yamada, M., Fujita, Y., Nishigami, T., Nakasho, K., and Uematsu, K. (2001) *J. Biol. Chem.* **276**, 4604–4610
291. Waxman, D. J., Dannan, G. A., and Guengerich, F. P. (1985) *Biochemistry* **24**, 4409–4417
292. MacGeoch, C., Morgan, E. T., Halper, J., and Gustafsson, J.-A. (1984) *J. Biol. Chem.* **259**, 15433–15439
293. Covey, D. F., Hood, W. F., Beusen, D. D., and Carrell, H. L. (1984) *Biochemistry* **23**, 5398–5406
294. Vaz, A. D. N., Kessell, K. J., and Coon, M. J. (1994) *Biochemistry* **33**, 13651–13661

References

295. Harada, N., Ogawa, H., Shozu, M., Yamada, K., Suhara, K., Nishida, E., and Takagi, Y. (1992) *J. Biol. Chem.* **267**, 4781–4785
296. Graham-Lorence, S., Khalil, M. W., Lorence, M. C., Mendelson, C. R., and Simpson, E. R. (1991) *J. Biol. Chem.* **266**, 11939–11946
297. Terashima, I., Suzuki, N., Itoh, S., Yoshizawa, I., and Shibutani, S. (1998) *Biochemistry* **37**, 8803–8807
298. Service, R. F. (1998) *Science* **279**, 1631–1633
299. Henderson, B. E., Ross, R. K., and Pike, M. C. (1993) *Science* **259**, 633–638
299a. Hong, W. K., and Sporn, M. B. (1997) *Science* **278**, 1073–1077
299b. Shimotakahara, S., Gorin, A., Kolbanovskiy, A., Kettani, A., Hingerty, B. E., Amin, S., Broyde, S., Geacintov, N., and Patel, D. J. (2000) *J. Mol. Biol.* **302**, 377–393
300. Marshall, E. (1994) *Science* **264**, 1524–1527
301. Davis, D. L., and Bradlow, H. L. (1995) *Sci. Am.* **273**(Oct), 166–172
302. Kaiser, J. (1996) *Science* **272**, 1418
303. Arnold, S. F., Klotz, D. M., Collins, B. M., Vonier, P. M., Guillette, J., LJ, and McLachlan, J. A. (1996) *Science* **272**, 1489–1492
304. Hess, R. A., Bunick, D., Lee, K.-H., Bahr, J., Taylor, J. A., Korach, K. S., and Lubahn, D. B. (1997) *Nature (London)* **390**, 509–512
305. Sharpe, R. M. (1997) *Nature (London)* **390**, 447–448
306. O'Malley, B. W., and Schrader, W. T. (1976) *Sci. Am.* **234**(Feb), 32–43
307. Fuller, P. J. (1991) *FASEB J.* **5**, 3092–3099
308. La Baer, J., and Yamamoto, K. R. (1994) *J. Mol. Biol.* **239**, 664–688
309. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P., and Chambon, P. (1986) *Nature (London)* **320**, 134–139
310. Kostrouch, Z., Kostrouchova, M., and Rall, J. E. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 156–159
311. Scherrer, L. C., Dalman, F. C., Massa, E., Meshinchi, S., and Pratt, W. B. (1990) *J. Biol. Chem.* **265**, 21397–21400
312. Tsai, M.-J., and O'Malley, B. W. (1994) *Ann. Rev. Biochem.* **63**, 451–486
313. Htun, H., Barsony, J., Renyi, I., Gould, D. L., and Hager, G. L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4845–4850
314. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R., and Sigler, P. B. (1991) *Nature (London)* **352**, 497–505
315. Eriksson, M. A. L., Härd, T., and Nilsson, L. (1995) *Biophys. J.* **68**, 402–426
316. Schwabe, J. W. R., and Rhodes, D. (1991) *Trends Biochem. Sci.* **16**, 291–296
317. Hyder, S. M., Nawaz, Z., Chiappetta, C., Yokoyama, K., and Stancel, G. M. (1995) *J. Biol. Chem.* **270**, 8506–8513
318. Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995) *Nature (London)* **375**, 203–211
319. Lee, M.-O., Dawson, M. I., Picard, N., Hobbs, P. D., and Pfahl, M. (1996) *J. Biol. Chem.* **271**, 11897–11903
320. Vivat, V., Zechel, C., Wurtz, J.-M., Bourguet, W., Kagechika, H., Umemiya, H., Shudo, K., Moras, D., Gronemeyer, H., and Chambon, P. (1997) *EMBO J.* **16**, 5697–5709
321. Brzozowski, A. M., Pike, A. C. W., Dauter, Z., Hubbard, R. E., Bonn, T., Engström, O., Öhman, L., Greene, G. L., Gustafsson, J.-Å., and Carlquist, M. (1997) *Nature (London)* **389**, 753–758
322. Yeh, H., Miyamoto, H., Shima, H., and Chang, C. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5527–5532
323. Paech, K., Webb, P., Kuiper, G. G. J. M., Nilsson, S., Gustafsson, J.-Å., Kushner, P. J., and Scanlan, T. S. (1997) *Science* **277**, 1508–1510
324. Oñate, S. A., Tsai, S. Y., Tsai, M.-J., and O'Malley, B. W. O. (1995) *Science* **70**, 1354–1357
325. Grazzini, E., Guillon, G., Mouillac, B., and Zingg, H. H. (1998) *Nature (London)* **392**, 509–512
326. Kester, H. A., van der Leede, B.-jM., van der Saag, P. T., and van der Burg, B. (1997) *J. Biol. Chem.* **272**, 16637–16643
327. Bowyer, P., Clarke, B. R., Lunness, P., Daniels, M. J., and Osbourn, A. E. (1995) *Science* **267**, 371–374
328. Smith, T. W. (1988) *N. Engl. J. Med.* **318**, 358–365
329. Albuquerque, E. X., Daly, J. W., and Witkop, B. (1971) *Science* **172**, 995–1002
330. Daly, J. W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9–13
330a. Dumbacher, J. P., Spande, T. F., and Daly, J. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12970–12975
331. Coudron, T. A., Law, J. H., and Koeppe, J. K. (1981) *Trends Biochem. Sci.* **6**, 248–252
332. Pongs, O. (1985) in *Interaction of Steroid Hormone Receptors with DNA* (Sluyser, M., ed), pp. 226–240, Ellis Horwood, Chichester
333. Murray, I. A., Cann, P. A., Day, P. J., Derrick, J. P., Sutcliffe, M. J., Shaw, W. V., and Leslie, A. G. W. (1995) *J. Mol. Biol.* **254**, 993–1005
334. Moore, K. S., Wehrli, S., Roder, H., Rogers, M., Forrest, J. N., Jr., McCrimmon, D., and Zasloff, M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1354–1358
335. Stone, R. (1993) *Science* **259**, 1125

Study Questions

1. Outline the sequence and chemical mechanisms of the reactions involved in the conversion of acetyl-CoA into mevalonate.

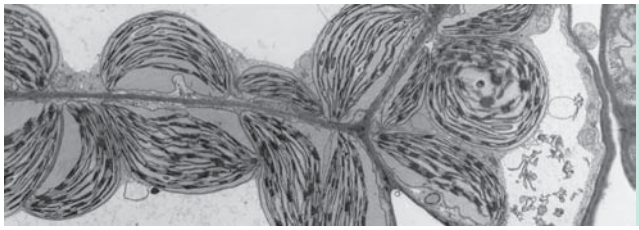
2. a) Show the structures of the reactants for the hydroxymethylglutaryl CoA synthase reaction.
   b) Free coenzyme A is liberated in the above reaction. From which molecule did it come? Explain the metabolic purpose behind the liberation of free CoA.

3. List as many substances as you can that are of polyprenyl origin and are present
   a) in foods
   b) in various commercial products

4. What distinctly different functions do 3-hydroxy-3-methylglutaryl-CoA synthases serve in the cytosol and in mitochondria of the liver?

5. Outline the functions of mitochondrial enzymes in the conversion of fructose into cholesterol in the liver.

6. How do you think that hydroxycitrate, an inhibitor of ATP citrate lyase, would affect the ability of liver to convert dietary fructose into bile acids?



Chloroplasts fill most of the cytoplasm around the junction of three cells of *Arabidopsis thaliana* seen in this micrograph. Both grana stacks and stroma lamellae (pictured in more detail in Fig. 23-19) can be seen. Also present are several small mitochondria. Portions of the large vacuoles, characteristic of plant cells, are seen at top, right, and bottom. Micrograph courtesy of Kenneth Moore.

Contents

1273 **A. Properties of Light**
1275 **B. Absorption of Light**
1275 1. Quantitative Measurement of Light Absorption, Spectroscopy
1276 2. The Energy Levels of Molecules
1276 3. Infrared (IR) and Raman Spectra
1277 *Vibrational frequencies of amide groups*
1278 *Raman spectra*
1280 4. Electronic Spectra
1280 *Shapes of absorption bands*
1282 *Classification of electronic transitions*
1283 *Use of plane polarized light*
1283 *Relationship of absorption positions and intensity to structures*
1284 *Spectra of proteins and nucleic acids*
1285 *Difference spectra and derivative spectra*
1286 5. Circular Dichroism and Optical Rotatory Dispersion
1287 6. Photoacoustic Spectroscopy
1288 7. X-Ray Absorption and Mössbauer Spectroscopies
1288 **C. Fluorescence and Phosphorescence**
1288 1. Excitation and Emission Spectra
1290 *Fluorescence lifetimes*
1290 *Time-resolved fluorescence spectroscopy*
1290 *Triplet states, phosphorescence, and quenching*
1291 *Anisotropy*
1291 2. Fluorescence Resonance Energy Transfer (FRET)
1293 3. Energy-Selective Spectroscopic Methods
1293 4. Analytical Applications of Fluorescence
1294 **D. Photochemistry**
1295 1. Chemical Equilibria in the Excited State
1296 2. Photoreactions of Nucleic Acid Bases
1297 3. Sunburn, Cancer, and Phototherapy
1297 4. Photoaffinity Labels
1297 5. Microphotolysis and Ultrafast Light-Induced Reactions
1298 6. Optical Tweezers, Light-Directed Synthesis, and Imaging
1298 **E. Photosynthesis**
1299 1. Two Photosystems, the Z Scheme, and Reaction Centers
1299 *The Z scheme*
1300 *Electron transport and photophosphorylation*
1301 *Bacterial photosynthesis*
1301 2. Chloroplast Structure
1301 *Chloroplast membranes*
1302 *Photosynthetic pigments and their environments*
1305 3. The Light-Receiving Complexes
1305 *Phycobilinosomes*
1306 *Purple photosynthetic bacteria*
1306 *Green sulfur and nonsulfur bacteria*
1308 *Eukaryotic plants and cyanobacteria*
1310 4. The Reaction Centers and Their Photochemistry
1310 *Reaction centers of purple bacteria*
1312 *Reaction center kinetics*
1314 *Cyclic photophosphorylation in purple bacteria*
1314 *Comparison with other reaction centers*
1314 *PSI of cyanobacteria and green plants*
1317 *PSII and formation of oxygen*
1318 *ATP synthesis in chloroplasts*
1319 *Protection of chloroplasts against radiation and oxygen*
1319 5. Control of Photosynthesis
1320 *Light-induced transcription*
1320 *Light-induced control via the ferredoxin/thioredoxin system*
1321 6. Photorespiration; C_3 and C_4 Plants
1321 *Metabolism of glycolic acid*
1322 *The C_4 cycle for concentration of carbon dioxide*
1323 *Metabolism in the family Crassulaceae*
1324 7. Photosynthetic Formation of Hydrogen
1324 **F. Vision**
1324 1. Visual Pigments
1325 *Transmembrane structure*
1326 *The visual chromophores*
1329 2. The Light-Induced Transformation
1330 3. The Nerve Impulse
1330 *Transducin, cyclic GMP, and phosphodiesterase*
1331 *Rhodopsin kinase, recoverin, and arrestin*
1332 *Some details about cone cells and invertebrate vision*
1332 4. Regeneration of Visual Pigments; the Retinal Cycle
1332 5. Diseases of the Retina
1333 6. Proteins of the Lens
1333 **G. Bacteriorhodopsin and Related Ion Pumps and Sensors**
1335 *Halorhodopsin*
1335 *Sensory rhodopsins*
1336 *The photoactive yellow proteins (xanthopsins)*
1336 *Stentorin*
1337 **H. Phytochrome**
1338 **I. Some Blue Light Responses**
1339 1. Cryptochromes
1339 2. Photolyases
1340 **J. Bioluminescence**

1347 **References**
1357 **Study Questions**

Boxes

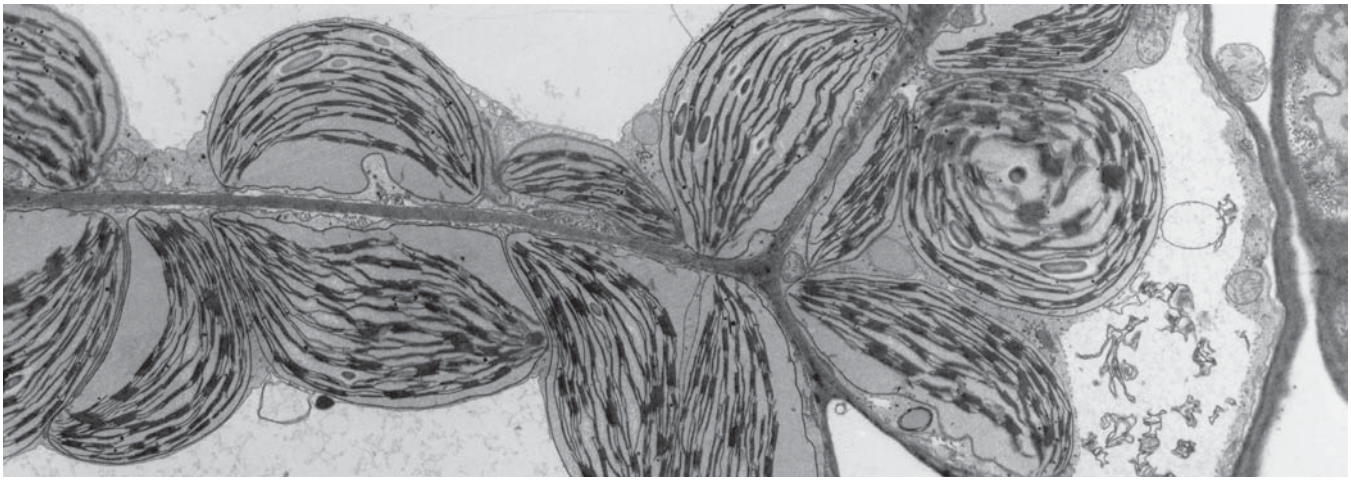
1344 Box 23-A The Green Fluorescent Protein and Other Light-Emitting Antennas

Tables

1273 Table 23-1 Some Properties of Light
1303 Table 23-2 Approximate Composition of Photosynthetic Units in a Spinach Chloroplast
1305 Table 23-3 The Common Bilin Pigments Present in Phycobiliniproteins
1313 Table 23-4 Properties of Various Reaction Centers

Light and Life

23



Light plays a pervasive role in human life. The earth is bathed in light from the sun, and from this light comes not only warmth but also energy for all living organisms. Of the 3×10^4 kJ m^{-2} of light energy falling on the earth each day,[1,2] ~30 kJ m^{-2} are captured by photosynthesis.[3] Light penetrating the atmosphere allows us to see and provides color to our environment. It controls the flowering of plants, the germination of seeds and spores, the greening of seedlings, and the daily cycles of many organisms. High in the stratosphere ultraviolet light reacts with oxygen to create a protective blanket of ozone. The ultraviolet light that is not screened out by the ozone layer kills bacteria, tans our skin, and often mutates our DNA, inducing many cancers.[4,5] Organisms, from bacteria to higher plants, display **phototaxis**, the ability to move toward a source of light or to orient themselves with respect to a source of light. In plants the chloroplasts assume an orientation that maximizes efficiency of light absorption. Plants grow toward light (**phototropism**), and some organisms avoid light. Many organisms emit light.

Many of our most important experimental techniques involve the use of light or of other forms of electromagnetic radiation of a wide range of energies. X-rays, ultraviolet light, infrared light, and microwaves all serve in the study of biomolecules.

A. Properties of Light

Light is a form of electromagnetic radiation and possesses characteristics of both waves and particles (**photons**). The energy of a photon is usually measured by the frequency (or by the wavelength in a vacuum to which it is inversely related, Table 23-1). A portion of the electromagnetic spectrum is shown on a logarithmic scale in Fig. 23-1.[1] At the high-energy end (off the scale of the figure to the right) are cosmic rays and gamma rays, while at the low-energy end radio waves extend to wavelengths of many kilometers. The narrow range of wavelengths from about 100 nm to a few micrometers, which is the subject of this chapter, includes the ultraviolet, visible, and near infrared

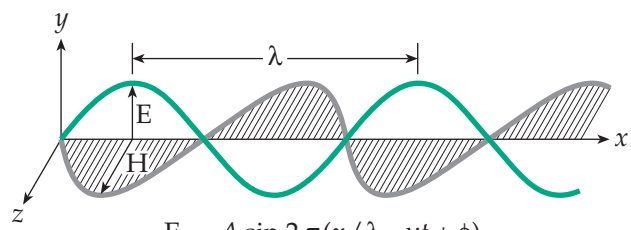
TABLE 23-1
Some Properties of Light

Velocity of light in a vacuum	$c = 2.998 \times 10^8$ m s^{-1}
Velocity of light in a medium	$c' = c/n$ where n = refractive index
Wave number (in nm)	$\bar{\nu} = 1/\lambda$; $\bar{\nu}$ (in cm^{-1}) $= 10^7/\lambda$
Frequency	$\nu = c/\lambda = c\bar{\nu}$ ν (in hertz) $= 2.998 \times 10^{10}\,\bar{\nu}$ (cm^{-1}) in a vacuum
Energy of quantum	$E = h\nu = hc\bar{\nu}$ E (joules) $= 1.986 \times 10^{-23}\,\bar{\nu}$ (cm^{-1}) E (eV) $= 1.240 \times 10^{-4}\,\bar{\nu}$ (cm^{-1})
Energy of einstein	$E = Nh\nu = Nhc\bar{\nu}$ $= 6.023 \times 10^{23}\,hc\bar{\nu}$ E (joules) $= 11.961\,\bar{\nu}$ (cm^{-1}) E (kcal) $= 2.859 \times 10^{-3}\,\bar{\nu}$ (cm^{-1})

ranges. The second line of Fig. 23-1 shows this region expanded. Note that the range of light reaching the earth's surface is narrow, largely being confined to wavelengths of 320–1100 nm. The human eye responds to an even more limited range of 380–760 nm, in which all of the colors of the rainbow can be found. The aromatic rings of proteins and nucleic acids absorb maximally at 280 and 260 nm, respectively. Even though these wavelengths are largely screened out by the ozone layer of the stratosphere, enough light penetrates to cause mutations and to damage the skin of the unwary sunbather.

The energy of a quantum of light is proportional to the **wave number** or **frequency**. The wave number $\bar{\nu}$ is the reciprocal of wavelength and is customarily given in units of cm^{-1} (**reciprocal centimeters**). Most of the absorption spectra in this book are plotted against wave number in cm^{-1}. The frequency ν in **hertz** is equal to $c'\nu$, where c' is the velocity of light in a medium. (The velocity of light in a vacuum is designated c and is equal to 3.00×10^8 m s^{-1}.) The energy of a quantum of light E is equal to $h\nu$, where h is Planck's constant, 6.626×10^{-34} J s^{-1}. From a chemical viewpoint, we are more interested in the energy of one **einstein**, i.e., one "mole" of light (6.023×10^{23} quanta). The energy in kJ per einstein is 11,960 $\bar{\nu}$ (in cm^{-1}, vacuum). Energy relationships are summarized in Table 23-1. The lower three scales of Fig. 23-1 also show the relationships of $\bar{\nu}$, ν, and E to wavelength.

The light wave is characterized by oscillating electrical and magnetic fields.[2,3,6] For propagation of light in the x direction the electric field vector **E**, which is customarily plotted in the y direction, is a function of the wavelength λ and the time (Eq. 23-1).



$$E_y = A \sin 2\pi(x/\lambda - \nu t + \phi) \quad (23\text{-}1)$$

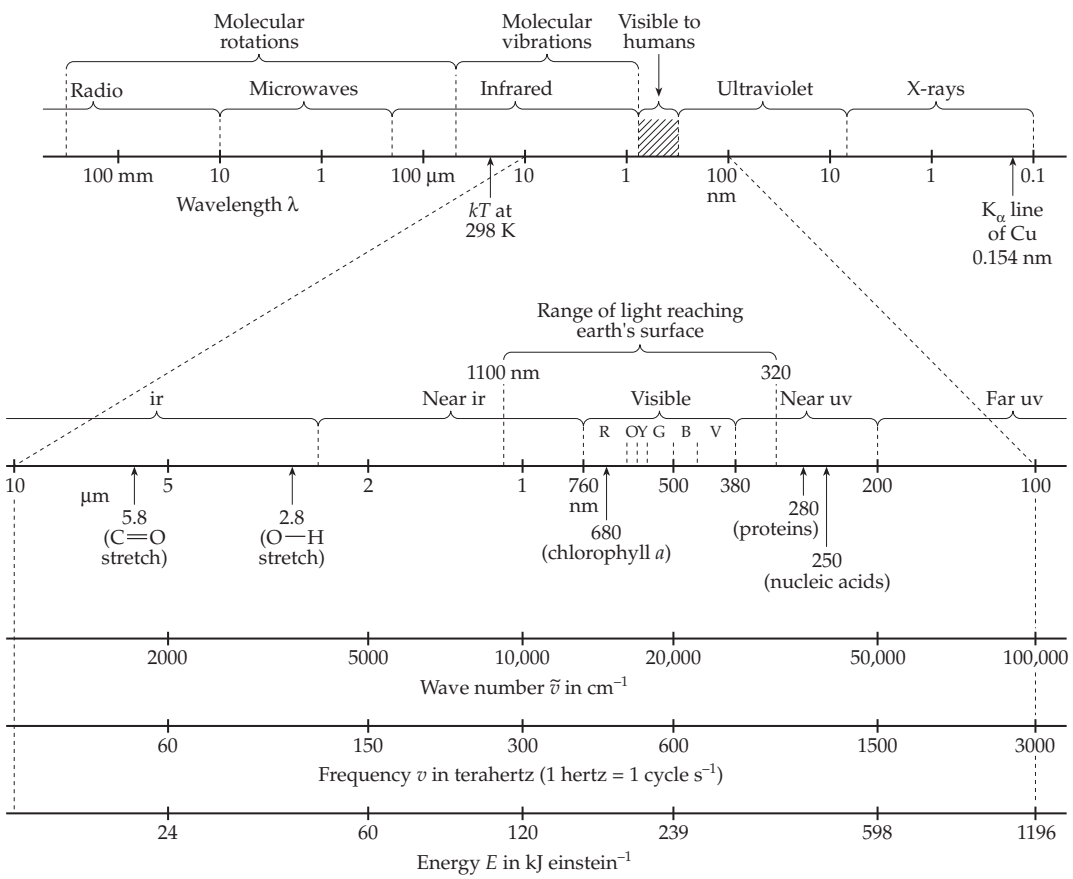


Figure 23-1 A part of the electromagnetic spectrum. The letters V, B, G, Y, O, R over the visible part of the spectrum refer to the colors of the light. The position marked "K_a line of Cu" is the wavelength of X-rays and most widely employed in X-ray diffraction studies of proteins and other organic materials.

The magnetic vector H is at right angles to the electric vector and is given by Eq. 23-2.

$$H_z = (\varepsilon / \mu)^{-} A \sin 2\pi(x/\lambda - \upsilon t + \phi) \qquad (23\text{-}2)$$

The velocity of propagation of light in a medium, c' (Eq. 23-3), depends upon both ε, the **dielectric constant** of the medium, and μ, the **magnetic permeability**.

$$\text{Velocity in a medium: } c' = c / (\varepsilon/\mu)^{-} = c / n \qquad (23\text{-}3)$$

The **refractive index** of a medium relative to a vacuum is given the symbol n. It is the factor by which the velocity of light in a vacuum is diminished in a medium. It is a function of wavelength. For the 589 nm sodium line n is 1.00029 for air and 1.33 for water at 25°C.

The term ϕ in Eqs. 23-1 and 23-2 is a **phase factor**. Most light is called **incoherent** because ϕ varies for the many photons making up the beam. **Coherent light** produced by **lasers** contains photons all with the same phase relationship. If the electric vectors of all the photons in a beam of light are in the same plane (as will be the case for light emerging after passage through certain kinds of crystals), the light is called **plane polarized**. The direction of polarization is that of the electric vector E. **Circularly polarized light**, in which the electric vector rotates and traces out either a left-handed or right-handed helix, can also be generated. A beam of left-handed circularly polarized light, together with a comparable beam polarized in the right-handed direction, is equivalent to a beam of plane polarized light. Conversely, plane polarized light can be resolved mathematically into right- and left-handed circularly polarized components.

B. Absorption of Light

Absorption of light is fundamental to all aspects of photochemistry and provides the basis for absorption spectroscopy.[3,5–11] Light absorption is always **quantized**. It can take place only when the energy $h\nu$ of a quantum is equal to the difference in energy between two energy levels of the absorbing molecule (Eq. 23-4).

$$E_2 - E_1 = h\nu \qquad (23\text{-}4)$$

Not only must the difference $E_2 - E_1$ be correct for absorption but also there must always be a change in the dipole moment of the molecule in going from one energy level to another. Only when this is true can the electric field of the light wave interact with the molecule. A further limitation comes from the symmetry properties of the wave functions associated with each energy level. Quantum mechanical considerations show that transitions between certain energy levels are allowed, while others are forbidden. Consideration of such matters is beyond the scope of this book, but the student should be aware that the quantum mechanical selection rules that express this fact are an important determinant of light absorption.

Many types of light source are used in chemical measurements. Of great importance is the recent development of lasers that deliver very short pulses of light. Pulses as short as five femtoseconds (5 fs)[11a,11b] and even less[11c] are being utilized for very rapid spectroscopy and excitation of fluorescence. Structures are being determined by ultrafast electron diffraction[11d] or X-ray diffraction.[11e,11f] It takes 200 fs or more for a chemical bond to stretch and break during a reaction. The cleavage and formation of bonds during this time can be observed using 5-fs pulses. Lasers with pulses in the attosecond range may soon be used to observe movements of electrons.[11g]

1. Quantitative Measurement of Light Absorption, Spectroscopy

An absorption spectrum is a plot of some measure of the intensity of absorption as a function of wavelength or wave number. The **transmittance** of a sample held in a **cell** (or **cuvette**) is the fraction of incident light that is transmitted, i.e., transmittance = I/I_o where I_o is the intensity of light entering the sample and I is that of the emerging light. The transmittance is usually defined for a single wavelength, i.e., for **monochromatic** light. The absorbance (or optical density) is defined by Eq. 23-5, which also states the **Beer–Lambert law**. The length (in centimeters) of the

$$\text{Absorbance} = A = \log_{10}(I_o / I) = \varepsilon c l \qquad (23\text{-}5)$$

light path through the sample is l, c is the concentration in moles per liter, and ε is the **molar extinction coefficient** (molar absorptivity or molar absorption coefficient), whose units are liter mol^{-1} cm^{-1} (or $\mathbf{M^{-1}}$ $\mathbf{cm^{-1}}$). The reader can derive Eq. 23-5 by assuming that in a thin layer of thickness dx the number of light quanta absorbed is proportional to the number of absorbing molecules in the layer. Integration from x = 0 to l gives the Beer–Lambert law. Equation 23-5 generally holds very well for solutions containing single ionic or molecular forms. However, it is usually valid only for monochromatic light. Furthermore, making precise measurements of absorbance is not easy. At A=1 only 10% of the incident light is transmitted, and the utmost care is required to obtain a value of A good to within ±0.05. At A=2 only 1% of incident light is transmitted, and the value of A will be much less reliable. Very low absorbances are also difficult to measure. In view of the importance that

spectrometry has played in biochemistry, it may seem surprising that the first reliable commercial laboratory ultraviolet–visible spectrophotometers became available in 1940 and the first commercial infrared spectrometer in 1942.[12]

2. The Energy Levels of Molecules

The energy of molecules consists of **kinetic** (translational), **rotational**, **vibrational**, and **electronic** components. The corresponding rotational, vibrational, and electronic energy levels are always quantized. Light quanta of wavelengths 0.2–20 mm (50–0.5 cm^{-1}; frequencies of 1.5 x 10^{12} to 1.5 x 10^{10} s^{-1}) with energies of 0.6–0.006 kJ/einstein are sufficient to excite molecules from a given rotational energy level to a higher one. Spectra in this "far infrared" or "microwave" region often consist of a closely spaced series of lines. For example, the rotational spectrum of gaseous HCl is a series of lines at 20.7 cm^{-1} intervals beginning at that wave number and reaching a maximum at about 186 cm^{-1} (54 µm). The energies involved in absorption of such light are far lower than energies of activation for common chemical reactions and lower than the average translational energy of molecules in solution at ordinary temperatures (3/2 k_BT or 3.7 kJ/mol at 25°C). However, they are still much higher than energies involved in the nuclear transitions of NMR spectra (Chapter 3). Compare 500 Hz for a proton resonating at 1 ppm in a 500 MHz NMR spectrometer with the 10^{10}–10^{12} s^{-1} frequencies of microwave spectra.

Vibrational energies range from about 6 to as much as 100 kJ mol^{-1} with corresponding wave numbers of ~500–8000 cm^{-1}. The resulting absorption spectra are in the infrared region. Excited electronic energy levels are ~120 – 200 kJ mol^{-1}, and the spectral transitions are at 10,000–100,000 cm^{-1} (1000–100 nm wavelengths) in the visible and ultraviolet region.

3. Infrared (IR) and Raman Spectra

Absorption in the near infrared region is dominated by changes in vibrational energy levels. A typical wave number is that of the "amide A" band at 3300 cm^{-1} (3.0 µm wavelength), approximately 10^{14} s^{-1}. First let us consider the stretching vibrations of a diatomic molecule. The two nuclei of the molecule can be thought of as connected with a spring. The energy of oscillation is approximately that of a harmonic oscillator. Application of quantum theory shows that the discrete energy levels that can be assumed by the oscillator are equally spaced. The difference between each pair of successive energy levels is $h\nu$, where ν is the frequency of light that must be absorbed to raise the energy from one level to the next. In the ground state (unexcited state) the molecule still possesses a **zero-point energy,** $E = 1/2\ h\nu_o$, equal to half the energy needed to induce a transition.

While the harmonic oscillator is a good approximation to the behavior of a molecule in the lower vibrational energy states, marked deviations occur at higher energies. At the lower energy levels the change in the distance between the atomic centers during the

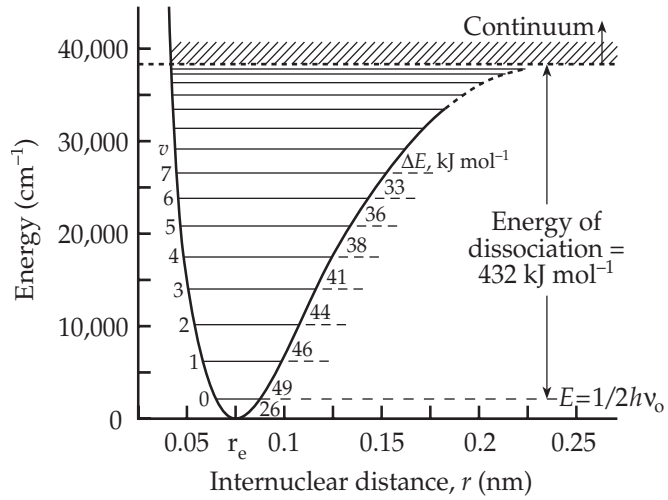


Figure 23-2 The potential energy of the hydrogen molecule as a function of internuclear distance, and the position of its vibrational energy levels. ΔE values are energy differences between successive levels; v designates vibrational quantum numbers. Adapted from Calvert and Pitts,[2] p. 135.

course of the vibration amounts to only ±10% or less, but as the energy becomes greater the bond stretches more and the motion becomes **anharmonic**. The energy states of molecules are often in the form of **Morse curves** in which energy is plotted against internuclear distance (Fig. 23-2). As the internuclear distance becomes very short, the energy rises steeply. As the bond is stretched, there comes a point at which addition of more energy ruptures the bond. A diatomic molecule will dissociate into atoms and more complex molecules into fragments. Vibrational energy levels can be portrayed as horizontal lines at appropriate heights on the Morse curve (Fig. 23-2).

Because there are many rotational energy levels corresponding to each vibrational level, IR spectra contain absorption bands resulting from simultaneous changes in both the vibrational and rotational energy levels of molecules. Instead of single peaks corresponding to single transitions in vibrational energy, progressions of sharp bands at closely spaced intervals are observed. An example is provided by the band corresponding to the stretching frequency of the H–Cl

bond in gaseous HCl at 2886 cm^{-1} (3.46 μm). There is actually no band at this wave number but a series of almost equally spaced bands on either side of the fundamental frequency from ~2600 to ~3100 cm^{-1} at intervals of ~21 cm^{-1}, i.e., the wave number of the rotational frequency seen in the microwave spectrum (Herzberg,[8] p. 55). The effect is to broaden the band as seen in a low-resolution spectrum. This is only one cause of the broadening of IR bands in solution. Another cause is interaction with solvent to provide a heterogeneity in the environments of the absorbing molecules.

The IR spectra of diatomic molecules are relatively easy to interpret, but for more complex substances the infrared absorption bands often cannot be associated with individual chemical bonds. Instead, they correspond to the **fundamental vibrations** (normal vibrations) of the *molecule*. Fundamental vibrations are those in which the center of gravity does not change. For a molecule containing n atoms, there are $3n - 6$ such vibrations. They are sometimes dominated by a vibration of a single bond, but often involve synchronous motion of many atoms. The fundamental vibrations of a molecule are described by such words as *stretching*, *bending* (in-plane and out-of-plane), *twisting*, and *deformation*. Rarely are all $3n - 6$ bands seen in an infrared spectrum. Some of the vibrations, e.g., the symmetric stretching of the linear CO_2 molecule, are not accompanied by any change in dipole moment, while other bands may simply be too weak to be observed clearly.

Vibrations involving many atoms in a molecule, i.e., **skeletal vibrations**, are often found in the region of 700–1400 cm^{-1} (14–7 μm). Vibrational frequencies that are dominantly those of individual functional groups can often be identified in the range 1000–5000 cm^{-1} (10–2 μm). Examples of the latter are the stretching frequencies of C–H, N–H, and O–H bonds, which have wave numbers of ~2900, 3300, and 3600 cm^{-1}, respectively. The energy (and frequency) of the vibrations increases as the difference in electronegativity between the two atoms increases. When a bond connects two heavier atoms, the frequency is lower, e.g., the wave number for C–O in a primary alcohol is ~1053 cm^{-1}. For a double bond it increases; for C=O it is ~1700 cm^{-1}. This C=O stretching frequency usually gives rise to one of the strongest bands observed in IR spectra. Hydrogen bonding has a strong and characteristic effect. Thus, the O–H frequency at ~3600 cm^{-1} is decreased to ~3500 cm^{-1} by hydrogen bonding.

Theory predicts that for a harmonic oscillator only a change from one vibrational energy level to the next higher is allowed, but for anharmonic oscillators weaker transitions to higher vibrational energy levels can occur. The resulting "overtones" are found at approximate multiples of the frequency of the fundamental. Combination frequencies representing sums and differences of frequencies of individual IR bands may also be seen. The intensities of these bands are low, but their presence at relatively high energies in the near IR region (4000–12,500 cm^{-1}) means that they may be easier to observe than the fundamental frequencies in the more crowded IR region.[12a] Development of the very sensitive **Fourier-transform infrared spectroscopy** (FTIR) has made it possible to record the complex IR spectra of macromolecules in dilute aqueous solutions rapidly. The water spectrum is substracted digitally.[13–16] FTIR has been utilized to study amide groups in peptides, carboxyl groups in proteins,[17] conformations of sugar rings in DNA,[18] and the ionization state of phosphate groups.[19] New computational methods involving use of two-dimensional representations provide simplifications in interpretation of IR spectra.[19a] Another variant is **total reflection FTIR**, a technique that records spectra of thin films and has permitted the recording of transient changes in protein spectra with microsecond time resolution.[20,21]

Vibrational frequencies of amide groups. The IR absorption bands of amide groups, which are present in both proteins and in the purine and pyrimidine bases, have attracted a great deal of attention.[13,22–23a] The **amide I** band at ~1680 cm^{-1} is associated with an in-plane normal mode of vibration that involves primarily C=O stretching. (The band is designated I′ if the N–H has been exchanged to form N–^{2}H.)[16] The **amide II** band at ~1500 $^{-1}$ and the **amide III** band at ~1250 cm^{-1} both arise from in-plane modes that involve N–H bending, while the higher frequency **amide A** band at ~3450 cm^{-1} involves N–H stretching. It is shifted to ~3300 cm^{-1} when the N–H is hydrogen bonded. Examples of IR spectra of proteins, including the amide bands A, I, and II, are shown in Fig. 23-3. The band shapes are complex. Those of the amide bands I and III depend upon the conformation of the peptide chain. For example, (Fig 23-3B) amide groups in α helices have an amide I band about 20 cm^{-1} higher than do those in β structures.

If peptide chains can be oriented in a regular fashion, it may be useful to measure **infrared linear dichroism**.[24,25] Absorption spectra are recorded by passing plane polarized light through the protein in two mutually perpendicular directions, with the electric vector either parallel to the peptide chains or perpendicular to the chains. Such a pair of spectra is shown in Fig. 23-3A for oriented fibrils of insulin. In this instance, the insulin molecules are thought to assume a β conformation and to be stacked in such a way that they extend transverse to the fibril axis (a cross-β structure). When the electric vector is parallel to the fibril axis, it is perpendicular to the peptide chains. Since the amide I band is dominated by a carbonyl stretching motion that is perpendicular to the

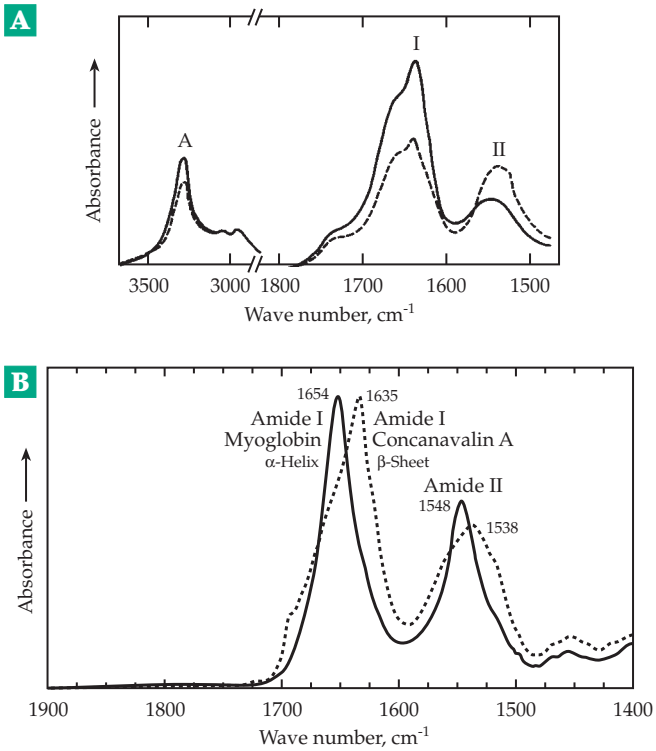


Figure 23-3 Infrared absorbance spectra of the amide regions of proteins. (A) Spectra of insulin fibrils illustrating dichroism. Solid line, electric vector parallel to fibril axis; broken line, electric vector perpendicular to fibril axis. From Burke and Rougvie.[24] Courtesy of Malcolm Rougvie. See also Box 29-E. (B) Fourier transform infrared (FTIR) spectra of two soluble proteins in aqueous solution obtained after subtraction of the background H_2O absorption. The spectrum of myoglobin, a predominantly α-helical protein, is shown as a continuous line. That of concanavalin A, a predominantly β-sheet containing protein, is shown as a broken line. From Haris and Chapman.[14] Courtesy of Dennis Chapman.

peptide chains in the β structure, this band is enhanced when the electric vector is also perpendicular to the peptide chains and is diminished when the electric vector is parallel to the peptide chains (perpendicular to the fibril axis, Fig. 23-3A). The same is true of the amide A band which is dominated by an N–H stretch. On the other hand, the dichroism of the amide II band is the opposite because it tends to be dominated by an N–H bending, which is in the plane of the peptide group but is longitudinal in direction. In **isotope-edited FTIR**, heavy atoms such as ^{13}C are introduced to shift IR bands and assist in their identification. The method can be combined with measurement of linear dichroism of oriented peptides.[25]

The loss of the amide II band in D_2O is one of the major tools for studying protein dynamics.[13,26] (see Chapter 3, Section I,5). In some cases the four main secondary structures, α helix, β sheet, β turn, and random coil, can be distinguished.[27] The amide bands of pyrimidines can also be observed in IR spectra of pyrimidines. Figure 23-4A shows the spectrum of 1-methyluracil in H_2O and also in D_2O. Notice that the amide II band is totally lacking in D_2O. The same figure also shows the IR spectrum of 1-methyluracil contaning ^{18}O in the 4 position. The shift of 9 cm^{-1} in the amide II position is part of the evidence that the NH bending vibration is extensively coupled to C=O and C=C stretching modes.

Raman spectra. In a collision between a photon and a molecule, the photon may undergo **elastic collision** in which the photon loses no energy but changes its direction of travel. Such scattering is known as **Rayleigh scattering** and forms the basis for a method of molecular mass determination. Sometimes **inelastic** collisions occur in which both the

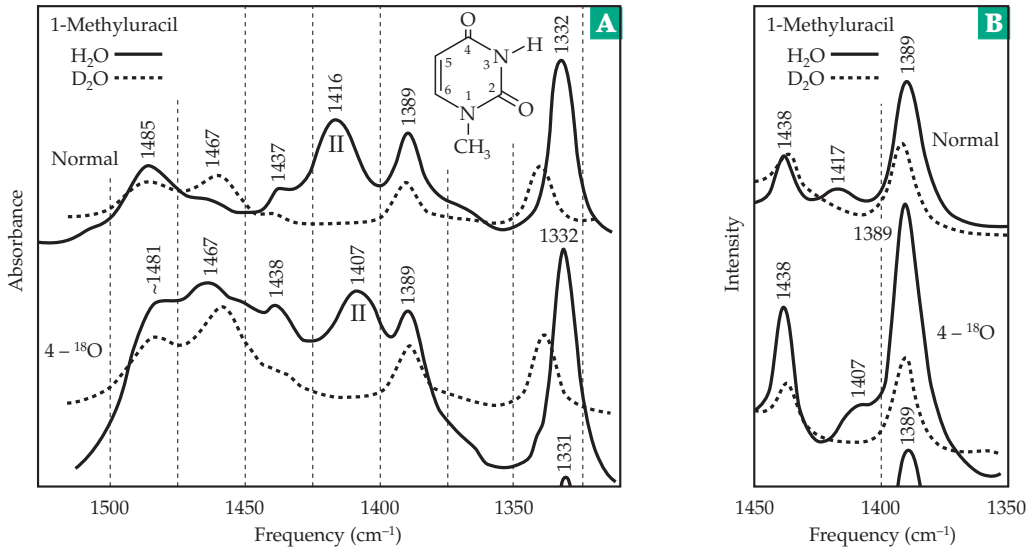


Figure 23-4 (A) Infrared and (B) Raman spectra of 1-methyluracil in H_2O and D_2O. Spectra for normal 1-methyluracil and for the specific isotopic derivatives with ^{18}O in the 4 position are shown. From Miles *et al.*[28]

molecule and the photon undergo changes in energy. Since such changes must be quantized and involve vibrational and rotational levels of the molecule, the spectrum of the scattered light (Raman spectrum) contains much of the same information as an ordinary IR spectrum. However, the selection rules are not the same. Some transitions are "infrared active" and others are "Raman active." IR-active transitions can occur only when the dipole moment varies with time as a molecule vibrates, whereas Raman-active transitions require that the polarizability change with time. For this reason, it, s useful to measure both IR and Raman spectra on the same sample. Until recently, Raman spectroscopy was not used much in biochemistry because of the low intensity of the scattered light. However, with laser excitation the technique is practical[7,13,23,28,29] and can be applied to aqueous or nonaqueous solutions, and to solid or dispersed forms of macromolecules.

Both amide I and amide III bands are seen in Raman spectra of proteins.[30] Lippert *et al.* devised the following method for estimating the fractions of α-helix, β sheet, and random coil conformations in proteins.[31] The amide I Raman bands are recorded at 1632 and 1660 cm^{-1} in D_2O (amide I′). The amide III band, which is weak in D_2O, is measured at 1240 cm^{-1} in H_2O. The intensities of the three bands relative to the intensity of an internal standard (the 1448 cm^{-1} CH_2 deformation) are related to those of standard poly-L-lysine in known conformations. See also Craig and Gaber.[32] The Raman spectrum of 2-methyluracil is shown in Fig. 23-4B. Note the low intensity of the amide II band relative to that of the amide I band, a characteristic of Raman spectra. Linear dichroism observed by polarized Raman microspectrophotometry has provided information about orientation of indole rings of tryptophan in filamentous virus particles.[33,33a]

In **resonance Raman spectroscopy**[34–37] a laser beam of a wavelength that is absorbed in an electronic transition is used. The scattered light is often strongly enhanced at frequencies differing from that of the laser by Raman frequencies of groups within the chromophore or of groups in another molecule adjacent to the chromophore. The resonance effect not only increases the sensitivity of Raman spectroscopy but also allows a person to study specifically the vibrational spectrum of a selected aromatic group or other structure within a macromolecule. Problems associated with the technique are fluorescence, which may be 10^6 times as strong as the Raman emission, and photochemical damage from the intense laser beam. Fluorescence is often quenched with KI (see Section C,1).

If the exciting laser has a frequency ν_0 and the frequency of a vibrationally excitation in a molecule is ν_1 the Raman spectrum will contain a pair of bands,

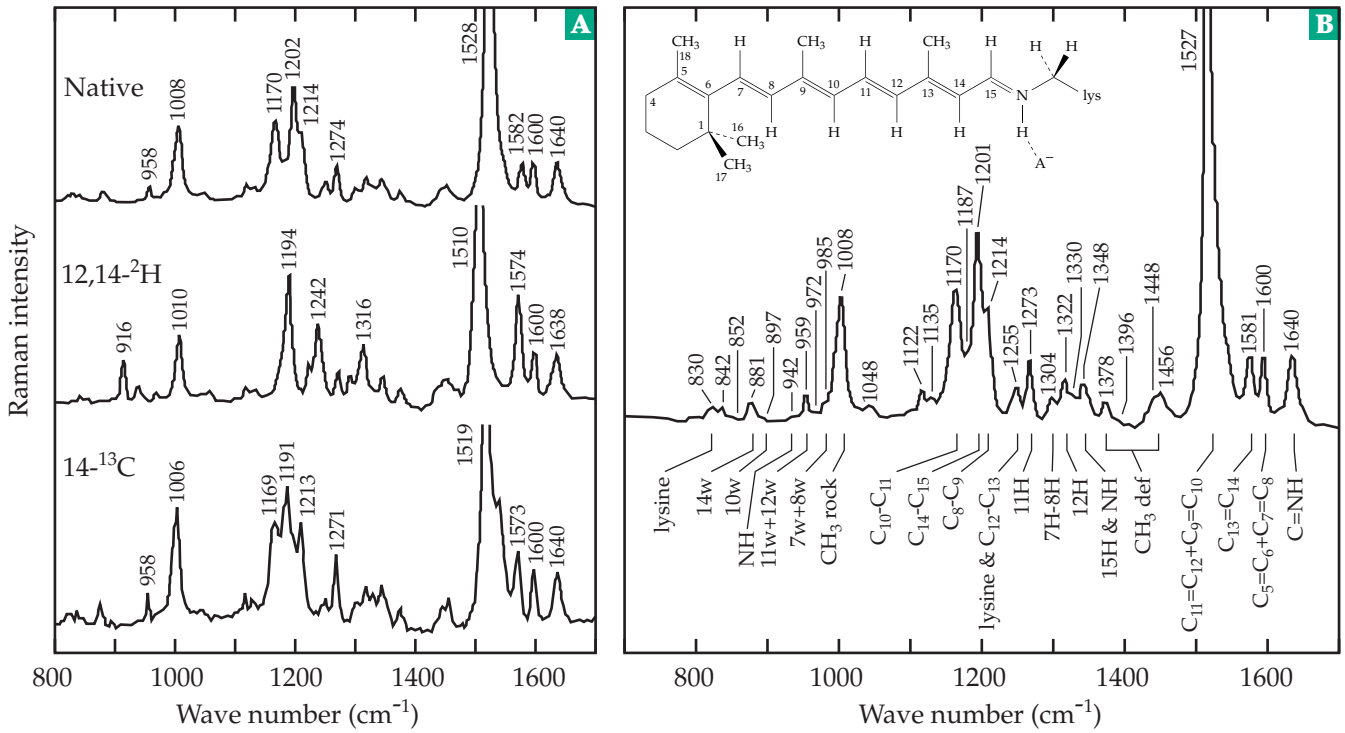


Figure 23-5 Resonance Raman spectra. (A) Of the retinaldehyde-containing bacteriorhodopsin bR_{568} (see Fig. 23-45) and its 12,14-^{2}H and 14-^{13}C isotopic derivatives. (B) Of bR_{568} labeled with the dominant internal coordinates that contribute to the normal modes. From Lugtenburg *et al.*[37]

the stronger one or "stokes" band of energy $h(\nu_0 - \nu_1)$ and a weaker one or "anti-stokes" band of $h(\nu_0 + \nu_1)$. Special techniques such as **coherent anti-stokes** Raman scattering provide a means of getting around the fluorescence problem.[13,36,38] Raman spectroscopy is also possible with excitation of ultraviolet absorption bands. It can be applied to peptides,[39–41] tryptophan or tyrosine residues of proteins,[42] nucleic acid bases,[43–45] heme proteins,[46–48] other metalloproteins,[49] flavin coenzymes,[50–52] pyridoxal phosphate,[53,54] flavoproteins,[54a] carotenoid-containing proteins,[55] and to substrates undergoing cleavage in the active site of papain.[56] Resonance Raman spectroscopy is very useful in the study of adducts of CO, NO, and O_2 with heme proteins because coupled vibrational modes of both the porphyrin rings and axial ligands can be observed.[56a,b] Resonance Raman spectra are strongly amplified for samples adsorbed to specially prepared colloidal silver particles.[57] Examples of resonance Raman spectra are shown in Fig. 23-5.

4. Electronic Spectra

Biochemists make extensive use of spectroscopy in the ultraviolet (UV) and visible range. Visible light begins at the red end at ~12,000 cm^{-1} (800 nm) and extends to 25,000 cm^{-1} (400 nm). The ultraviolet range begins at this point and extends upward, the upper limit accessible to laboratory spectrophotometers being ~55,000 cm^{-1} (180 nm). The energies covered in the visible–UV range are from ~140 to ~660 kJ/mol. The latter is greater than the bond energy of all but the strongest double and triple bonds (Table 6-7). It is understandable that UV light is effective in inducing photochemical reactions. Even the lower energy red light, which is used by plants in photosynthesis, contains enough energy per einstein to make it feasible to generate ATP, to reduce $NADP^+$, and to carry out other photochemical processes. Although the energies of light absorbed in electronic transitions are large, the geometry of molecules in the excited states is often only slightly altered from that in the ground state. The amount of vibration is increased, and the molecule usually expands moderately in one or more dimensions.

The significance of light absorption in biochemical studies lies in the great sensitivity of electronic energy levels of molecules to their immediate environment and to the fact that spectrophotometers are precise and sensitive. The related measurements of circular dichroism and fluorescence also have widespread utility for study of proteins, nucleic acids, coenzymes, and many other biochemical substances that contain intensely absorbing groups or **chromophores**.[58]

Shapes of absorption bands. Electronic absorption bands are usually quite broad, the width of the band at half-height often being 3000–4000 cm^{-1}. The breadth arises largely from the coupling of electronic excitation to changes in the vibrational and rotational energy levels. Inhomogeneity of environments in the solvent also contributes. Shapes of absorption bands are to a large extent determined by the **Franck–Condon principle**, which states that no significant change in the positions of the atomic nuclei of the molecule occurs during the time of the electronic transition. Since the frequency of light absorbed during these transitions is ~10^{15} to 10^{16} s^{-1}, the absorption of light energy occurs within 10^{-15} to 10^{-16} s, the time equivalent to the passage of one wavelength of light. During this period the vibrational motions of the nuclei are almost insignificant because of the much lower frequencies of vibration. Two types of potential energy curves for excited states of molecules are shown in Fig. 23-6.[2] In the first the geometry of the molecule is little changed between ground state and excited state. At room temperature most molecules are in the lowest energy states of at least the most energetic of the various vibrational modes of the molecule ($3/2\, k_B T$ ~300 cm^{-1}). Therefore, the most probable transitions occur from the lowest vibrational states of the ground electronic states. The most probable internuclear distance for a molecule in the ground state is the equilibrium distance r_e (Fig. 23-2). Since that distance is the same in all of the vibrational levels of the electronically excited state, transitions to any of these states may occur. The transition to the first vibrational level of the excited state is most likely. The result is an absorption spectrum in which the sharp band representing the "0–0 transition" is most intense and in which there are progressively weaker bands corresponding to the 0–1, 0–2, 0–3, etc., transitions (Fig. 23-6A). Many organic dyes with long series of conjugated double bonds have spectra of this type.[59,60]

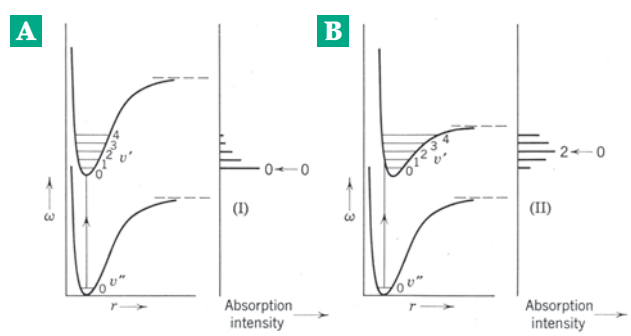


Figure 23-6 Typical potential energy curves for two types of band spectra: (A) For a transition in which the equilibrium internuclear distances r_e are about equal in the ground and excited states. (B) For a transition in which r_e' (excited state) > r_e (ground state). From Calvert and Pitts,[2] p. 179.

A second type of spectrum is illustrated in Fig. 23-6B. In this instance, the molecule has expanded in the excited state, and r_e is greater than in the ground state. The Franck–Condon principle suggests that a transition is likely only to those vibrational levels of the excited state in which the internuclear distance is compressed for a significant fraction of the time, approximately to that of r_e in the ground state. Examination of Fig. 23-6B explains why the resulting absorption spectra tend to have weak 0–0 bands and stronger bands corresponding to transitions to higher levels.

For real spectra of polyatomic molecules the situation is more complex. Some molecules in the ground state do occupy higher vibrational levels of the less energetic modes. Therefore, there will be weaker lines, some of which lie on the low-energy side of the 0–0 transition. Since in polyatomic molecules there are several normal modes of vibration, there will be other progressions of absorption bands paralleling those shown in Fig. 23-6 and filling in the valleys between them. All of the bands are broadened by rotational coupling and by interactions with solvent.

An example of a molecule giving a spectrum of the type shown in Fig. 23-6B is toluene. The vapor phase spectrum contains a large number of sharp lines, some of which can be seen in the low-resolution spectrum of Fig. 23-7. Several progressions can be identified.[61] One begins with the intense 0–0 line at 37.48×10^3 cm^{-1} and in which spacing of ~930 cm^{-1} between lines corresponds to a vibration causing symmetric expansion of the ring (ring breathing frequency), a frequency that can also be observed in the infrared spectrum. Other progressions beginning at the 0–0 line involve additional modes of vibration with frequencies (in the excited electronic state) of 460, 520, and 1190 cm^{-1}. Additional weaker bands are "buried" in the valleys in Fig. 23-6. When the spectrum of toluene is measured in solution, the sharp lines are broadened, but there are still indications of vibrational structure.

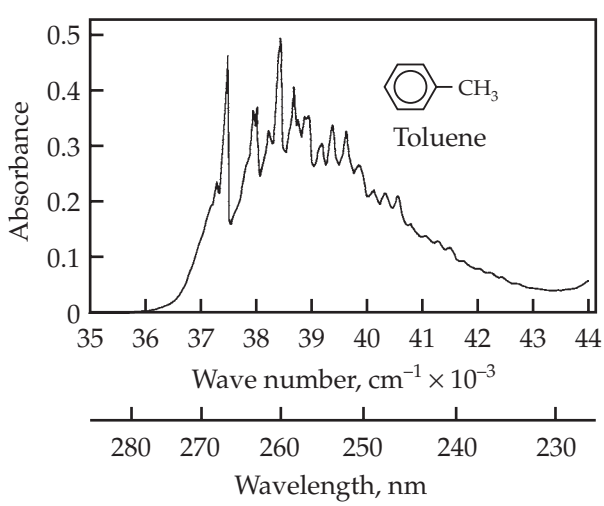


Figure 23-7 The spectrum of the first electronic transition of toluene vapor at low resolution. Cary 1501 spectrophotometer.

Gaussian curves (normal distribution functions) can sometimes be used to describe the shape of the overall envelope of the many vibrationally induced subbands that make up one electronic absorption band, e.g., for the absorption spectrum of the copper-containing blue protein of *Pseudomonas* (Fig. 23-8) Gaussian bands are appropriate. They permit resolution of the spectrum into components representing individual electronic transitions. Each transition is described by a **peak position**, **height** (molar extinction coefficient), and **width** (as measured at the half-height, in cm^{-1}). However, most absorption bands of organic compounds are not symmetric but are skewed

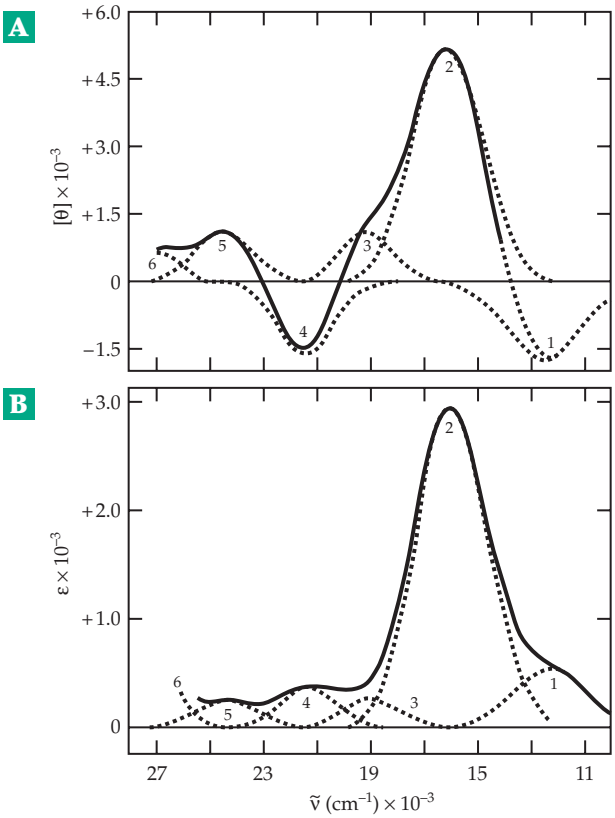


Figure 23-8 Resolution of the visible circular dichroism (ellipticity) spectrum (A) and absorption spectrum (B) of the *Pseudomonas* blue protein into series of overlapping Gaussian bands (—). The numbers 1 to 6 refer to bands of identical position and width in both spectra. Absorption envelopes resulting from the sum of the set of overlapping Gaussian bands (—) correspond within the error of the measurement to the experimental spectra. The *dashed part* of the CD envelope above 700 nm was completed by a curve fitter with the use of a band in the position of *band 1* of the absorption spectrum. From Tang *et al.*[68]

toward the high-energy side. It is best to fit such bands with a skewed function such as the **log normal** distribution curve.[62–64] In addition to position, height, and width, a fourth parameter provides a measure of **skewness**. Computer-assisted fitting with log normal curves gives precise values for the positions, widths, and intensities. In general, the peak position is somewhat to the high-energy side of the 0–0 transition.

Absorption spectra plotted as a linear function of wavelength are sometimes fitted with Gaussian curves. However, Gaussian curves only occasionally give a good fit for such spectra, and it is undesirable to measure bandwidths in nanometers. It is wave number that is proportional to energy. Spectral bands tend to have similar widths across the visible–ultraviolet range when plotted against wave number but not when ploted against wavelength. Another approach to the quantitative description of spectra is to fit the major progressions of vibrational subbands with series of narrow Gaussian curves.[65–67]

Classification of electronic transitions. The intense 600-nm absorption band of the copper blue protein in Fig. 23-8 is attributed to a *d*–*d* transition of an electron in the metal ion from one *d* orbital to another of higher energy.[68] The intensity is thought to arise from transfer of an electron from a cysteine thiolate to the copper (p. 883). The electronic transitions in most organic molecules are of a different type. Transitions lying at energies <55,000 cm^{-1} are classified as either π–π^* or n–π^*. In the π–π^* transitions an electron is moved from a bonding π molecular orbital to an antibonding (π^*) orbital. Such a transition is present in ethylene at 61,540 cm^{-1} (162.5 nm) with a maximum molar extinction coefficient ε_{max} of ~15,000 M^{-1} cm^{-1}. An n–π^* transition results from the raising of an electron in an unshared pair of an oxygen or nitrogen atom into a π^* antibonding orbital. These transitions are invariably weak. For example, acetone in H_2O shows an n–π^* transition at 37,740 cm^{-1} (265 nm). The value of ε_{max} is ~240 and the width is about 6400 cm^{-1}. A characteristic of n–π^* transitions is a strong shift to lower energies as the compound is moved from water into less polar solvents. For example, the peak of the acetone band lies at 36,920 cm^{-1} in methanol and at 35,970 cm^{-1} (278 nm) in hexane. Such a solvent shift is often taken as diagnostic of an n–π^* transition, and it is often stated that the π–π^* bands shift in the opposite direction upon change of solvent character. However, the latter is not true for many of the polar chromophores found in biochemical substances. Thus, the π–π^* bands of tyrosine also shift to lower energies when the molecule is moved from water into hexane. However, the magnitude of the shift is much less than for the n–π^* band of acetone.

A molecule can have several different excited states of increasing energies. In benzene and its derivatives there are three easily detectable π–π^* transitions (see Fig. 3-13). The first is a weak band centered at ~ 260 nm in toluene (Fig. 23-7) and ~275 nm in tyrosine (Fig. 3-13) with $\varepsilon = 10^2$ to 10^3. The second is a band at a higher frequency (at 1.35 0.10 times the frequency of the first band) with ε_{max} often as high as 10^4. The third band is found at still higher energies with ε_{max} reaching 5×10^4. The excited-state energy levels represented by these transitions were labeled ^1L_b, ^1L_a, and ^1B_a by Platt. Other authors described the levels in terms of the symmetries of the molecular orbitals, the ground state being $^1A_{1g}$, and the three excited states $^1B_{2u}$, $^1B_{1u}$, and $^1E_{1u}$. In these symbols the superscript 1 indicates that the excited states are **singlet** in nature; that is, the electrons remain paired in the excited states. Absorption of visible and ultraviolet light almost always leads to singlet excited states initially. For more complex ring systems the number of possible electronic transitions increases, but attempts are often made to relate these transitions back to those of benzene.

The intensities of electronic transitions vary greatly. The area ($\mathcal{A}$) under the absorption band, when ε is plotted against wave number $\bar{\nu}$, is directly proportional (Eq. 23-6) to a dimensionless quantity called the **oscillator strength** *f*.

$$f = \frac{2.303\, m_e c^2}{\pi N e^2} F\mathcal{A} = 4.32 \times 10^{-9}\, F\mathcal{A} \qquad (23\text{-}6)$$

In this equation m_e and e are the mass and charge of the electron, c is the velocity of light, N is Avogadro's number, and $\mathcal{A}$ is the area in a plot of ε vs ν in cm^{-1}; F is a dimensionless correction factor that is related to the refractive index of the medium and is near unity for aqueous solutions. If the area is approximated as that of a triangle of height ε_{max} and width (at half-height) W, we find that for a typical absorption band of $\varepsilon_{max} = 10^4$ and $W = 3000$ cm^{-1}, $f = 0.13$.

The oscillator strength is related to the probability of a transition and can become approximately 1 only for the strongest electronic transitions. However, it is rarely this high. For example, the oscillator strength is ~10^{-4} for Cu^{2+} and ~2×10^{-3} for the toluene absorption band shown in Fig. 23-7. The low intensity of absorption bands of benzene derivatives is related to the fact that these transitions are quantum mechanically forbidden for a completely symmetric molecule. It is only because of coupling with asymmetric vibrations of the ring that the ^1L_b transition of benzene becomes weakly allowed. In the benzene spectrum the 0–0 transition is completely absent, and only those progressions involving uptake of an additional 520 cm^{-1} of a nonsymmetric vibrational energy are observed. In the case of toluene and phenylalanine, the asymmetry of the ring introduced by the substituents permits the 0–0 transition to occur and leads to a higher oscillator

strength than that observed with benzene. The ^1L_a transition of benzenoid derivatives is also partially forbidden by selection rules, and only the third band begins to approach an oscillator strength of one.

Use of plane polarized light. The intensity of a spectral transition is directly related to the **transition dipole moment** (or simply the transition moment), a vector quantity that depends upon the dipole moments of the ground and excited states. For aromatic ring systems, the transition dipole moments of the π–π^* transitions lie in the plane of the ring. However, both the directions and intensities for different π–π^* transitions within a molecule vary.

The transition moment has a dimension of length (usually given in angstroms) and can be thought of as a measure of the extent of the charge migration during the transition. Light is absorbed best when the directions of polarization (i.e., of the electric vector of the light) and of the transition moment coincide. This fact can easily be verified by light absorption measurements on crystals. As with infrared spectra of oriented peptide chains (Fig. 23-3), the electronic spectra of crystals display a distinct dichroism. Crystals of coenzyme-containing proteins (Fig. 23-9) are very appropriate for spectroscopy with polarized light because the chromophores are spaced far enough apart to avoid electronic interaction and have absorbances low enough to record with crystals of the order of 0.1 mm thickness.[69,70]

In contrast to π–π^* transitions, the n–π^* transitions of heterocyclic compounds and carbonyl-containing rings are often polarized in a direction perpendicular to the plane of the ring. Linear dichroism of cytosine, adenine, and other nucleic acid bases has been measured on single crystals and in partially oriented polymer films.[71] Magnetically induced linear dichroism provides a new tool for study of metalloproteins.[72]

Relationship of absorption positions and intensity to structures. While quantum mechanical calculations permit prediction of the correct number and approximate positions of absorption bands, they are imprecise. For this reason, electronic spectroscopy also relies upon a combination of empirical rules and atlases of spectra that can be used for comparison purposes.[74–76] The following may help to orient the student. The position of an absorption band shifts **bathochromically** (to longer wavelength, lower energy) when the number of conjugated double bonds increases. Thus, **butadiene** absorbs at 46,100 cm^{-1} (217 nm) vs the 61,500 cm^{-1} of ethylene. As the number of double bonds increases further, the bathochromic shifts become progressively smaller (but remain more nearly constant in terms of wavelength than wave number). For **lycopene** (Fig. 23-10) with 11

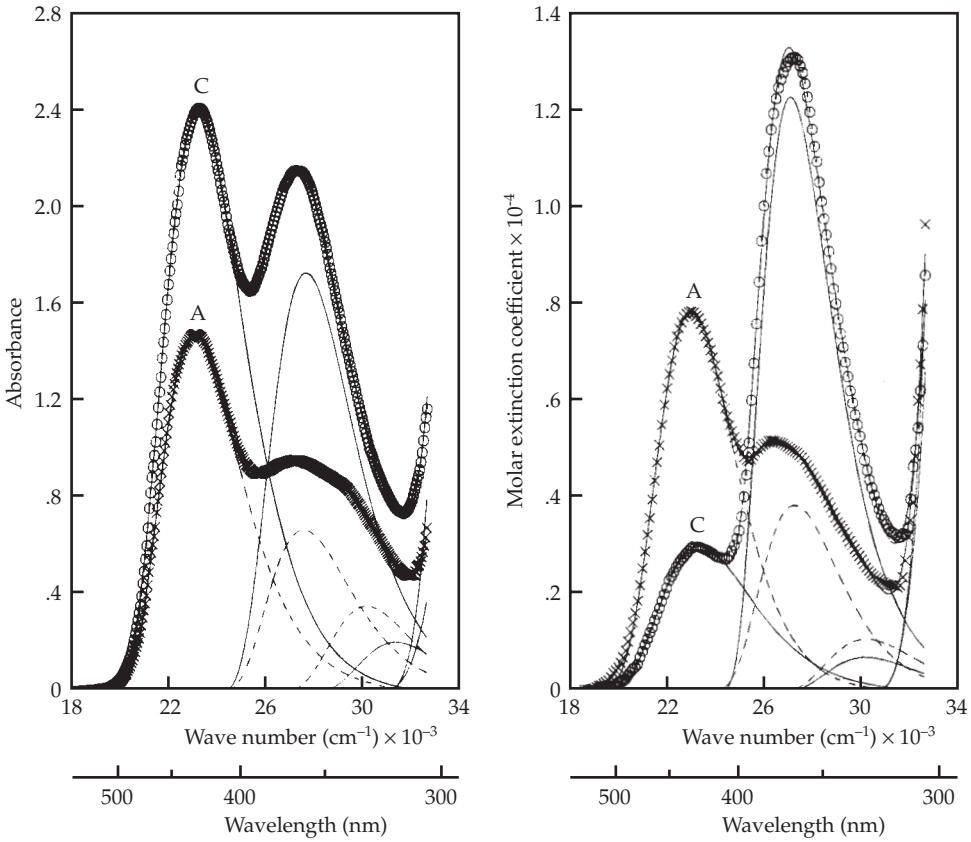


Figure 23-9 Polarized absorption spectra of orthorhombic crystals of cytosolic aspartate aminotransferase. The light beam passed through the crystals along the *b* axis with the plane of polarization parallel to the *a* axis (A) or the *c* axis (C). Left, native enzyme at pH 5.4; right, enzyme soaked with 300 mM 2-methylaspartate at pH 5.9. The band at ~430 nm represents the low pH protonated Schiff base form of the enzyme. Upon soaking with 2-methylaspartate the coenzyme rotates ~30° to form a Schiff base with this quasisubstrate. The result is a large change in the *c*/*a* polarization ratio. The ~364 nm band in the complex represents the free enzyme active site in the second subunit of the dimeric enzyme.[70,73] Courtesy of C. M. Metzler.

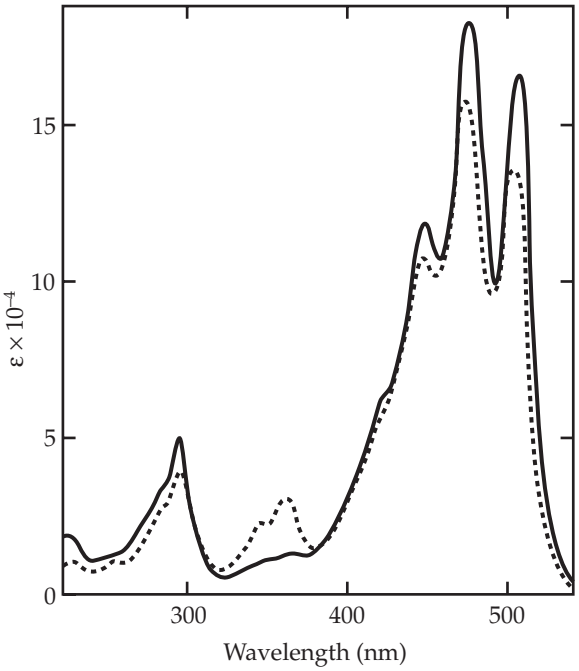


Figure 23-10 The absorption spectrum of lycopene (plotted vs wavelength). Note the vibrational structure, which has a spacing of ~1200–1500 cm^{-1}. The solid line is for all-translycopene while the dashed line is that of the sample after refluxing 45 min in the dark. The new peak at ~360 nm arises from isomers containing some cis double bonds.

conjugated double bonds the absorption band is located at 21,300 cm^{-1} and displays distinct vibrational structure (Fig. 23-10). Certain ring molecules such as the porphyrins and chlorophylls have spectra that can be related back to those of the linear polyenes. Note (Fig. 16-7) that the porphyrin α and β bands represent vibrational structure of a single electronic transition, whereas the intense Soret band results from a different transition.

Substituted benzenes almost invariably absorb at lower energies than the parent hydrocarbon. The stronger the electron withdrawing or donating ability of the substituent, the larger the bathochromic shift. The magnitude of the shift has been correlated with the Hammett σ constants. Thus, the first absorption band of tyrosine in water is shifted 2600 cm^{-1} toward the red from that of benzene, while that of the dissociated tyrosine anion is shifted 4700 cm^{-1}, very roughly in proportion to the σ_p values of Box 6-C. Especially large shifts are observed when functional groups of opposite types (that is, an electron donating group vs an electron accepting group) are both present in the same ring. The effects of ortho and meta substituent pairs are closely similar (in contrast to the differing electronic effects of ortho and meta pairs in chemical reactivity). Substituent pairs in para positions yield somewhat different spectral shifts. When there are more than two substituents, the two strongest groups often dominate in determining the character of the spectrum. Useful empirical rules have been developed.[77,78]

Spectra of proteins and nucleic acids. Most proteins have a strong light absorption band at 280 nm (35,700 cm^{-1}) which arises from the aromatic amino acids tryptophan, tyrosine, and phenylalanine (Fig. 3-14). The spectrum of phenylalanine resembles that of toluene (Fig. 23-7),[61] whose 0–0 band comes at 37.32 x 10^3 cm^{-1}. The vibrational structure of phenylalanine can be seen readily in the spectra of many proteins (e.g., see Fig. 23-11A). The spectrum of tyrosine is also similar (Fig. 3-13), but the 0–0 peak is shifted to a lower energy of ~35,500 cm^{-1} (in water). Progressions with spacings of 1200 and 800 cm^{-1} are prominent.[79] The low-energy band of tryptophan consists of two overlapping transitions ^1L_a and ^1L_b.[65] The ^1L_b transition has well-resolved vibrational subbands, whereas those of the ^1L_a transition are more diffuse. Tryptophan derivatives in hydrocarbon solvents show 0–0 bands for both of these transitions at approximately 289.5 nm (34,540 cm^{-1}). However, within proteins the ^1L_a band may be shifted 3–10 nm (up to 1100 cm^{-1}) toward lower energies, probably as a result of hydrogen bonding to other groups in the protein. The largest shifts can occur when the NH group of the indole ring is hydrogen bonded to COO^-, a ring nitrogen of histidine, or a carbonyl group of amides.[80] In an aqueous medium the ^1L_b band of tryptophan is shifted to higher energies and the ^1L_a band to lower energies than in a hydrocarbon solvent.

In addition to the three aromatic amino acids, disulfide bonds absorb in the near ultraviolet region as indicated in Fig. 3-14. Since the absorption characteristics depend upon the dihedral angles in the disulfide bridges, it is difficult to accurately evaluate the contribution of this chromophore to the 280-nm band.

Tyrosine, tryptophan, and phenylalanine all have additional transitions in the high-energy UV region of the spectrum (Fig. 3-13). Even more intense are the absorption bands of the amide groups, which become significant above 45,000 cm^{-1}.[81] These include a weak $n-\pi^*$ transition at ~45,500 cm^{-1} (210 nm) overlapped by a strong $\pi-\pi^*$ transition at ~52,000 cm^{-1} (192 nm).[82] Histidine also has absorption bands in this region.

As with polypeptides, the light absorption properties of polynucleotides reflect those of the individual components. The spectra of the purine and pyrimidine bases as ribonucleosides are shown in Fig. 5-5. The number of individual electronic transitions and their origins are not immediately obvious, but many measurements in solutions and in crystals, as well as theoretical computations,[7,83,84] have been made. Cytosine has $\pi-\pi^*$ transitions at ~275, 230, 200, and

185 nm,[83] the two highest energy bands being overlapped. Adenine derivatives have seven $\pi-\pi^*$ transitions.[71] Spectra of flavins contain at least four intense transitions (Fig. 15-8).[85]

Whereas proteins have their low energy absorption band at ~280 nm, polynucleotides typically have maxima at ~260 nm (38,500 cm^{-1}). A phenomenon of particular importance in the study of nucleic acids is the **hypochromic effect**. In a denatured polynucleotide the absorption is approximately the sum of that of the individual components. However, when a double helical structure is formed and the bases are stacked together, there is as much as a 34% depression in the absorbance at 260 nm. This provides the basis for optical measurement of DNA melting curves (Fig. 5-45).[45,86] The physical basis for the hypochromic effect is found in dipole–dipole interactions between the closely stacked base pairs.[7,86,87]

Difference spectra and derivative spectra. Changes in light-absorbing properties of proteins and nucleic acids are often measured as a function of some quantity such as pH, temperature, ionic environment, or the presence or absence of another interacting molecule. The induced changes in the spectrum are small but can be seen if the *difference* between the two spectra, one "unperturbed" and the other in the presence of some "perturbant," is recorded. The perturbant might be an additional reagent, an altered solvent (e.g., with added glycerol, D_2O), a change in pH, or temperature. The difference spectrum shown in Fig. 23-11B arises from the binding of an inhibitor succinate together with a substrate carbamoyl phosphate to the catalytic subunit of aspartate carbamoyltransferase (Fig. 7-20).[88] The difference spectrum appears as a pair of peaks and a valley in the aromatic amino acid region. With proper interpretation (caution!) difference spectra can be used to infer something about the change in environment of aromatic amino acids in a protein.[89]

Difference spectra are usually recorded by placing the unperturbed spectrum in the *reference* light beam of a spectrophotometer and the perturbed solution in the *sample* beam in carefully matched cuvettes. However, the spectrum shown in Fig. 23-11B was obtained by recording the two spectra independently and subtracting them with the aid of a computer. The same data have been treated in another way by fitting two log normal curves (p. 1283) to the absorption bands and plotting the differences between the mathematically

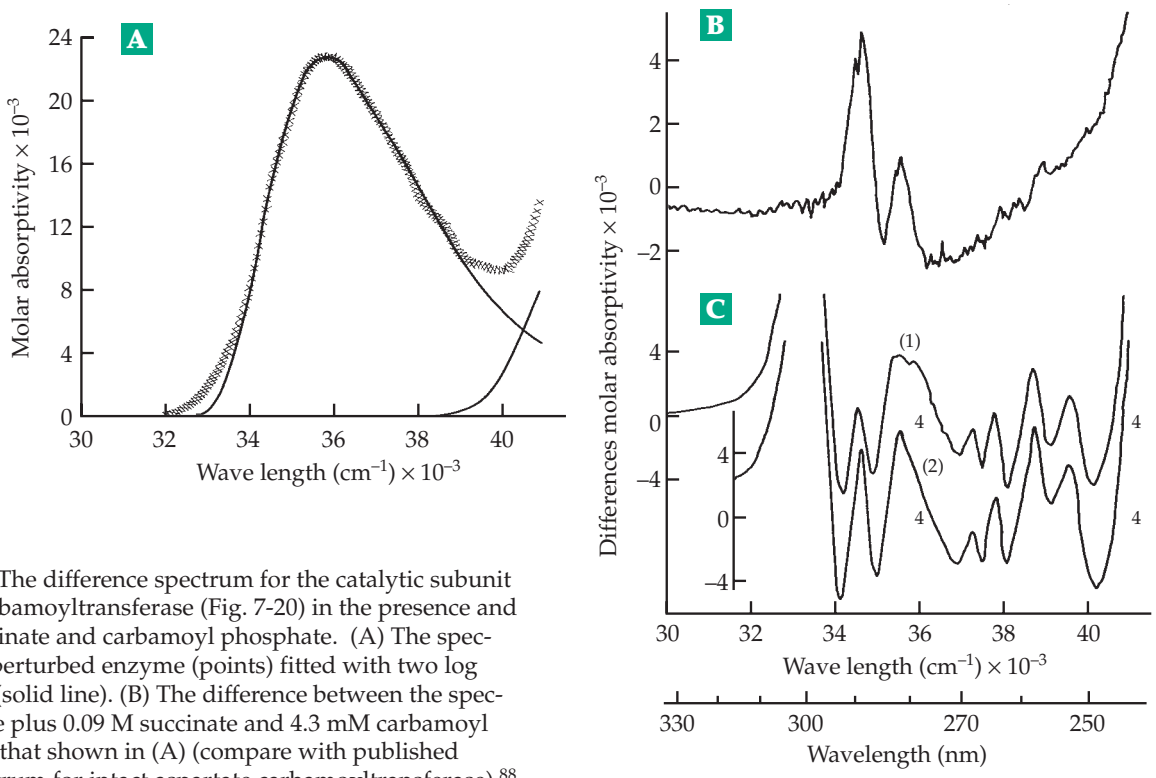


Figure 23-11 The difference spectrum for the catalytic subunit of aspartate carbamoyltransferase (Fig. 7-20) in the presence and absence of succinate and carbamoyl phosphate. (A) The spectrum of the unperturbed enzyme (points) fitted with two log normal curves (solid line). (B) The difference between the spectrum of enzyme plus 0.09 M succinate and 4.3 mM carbamoyl phosphate and that shown in (A) (compare with published difference spectrum for intact aspartate carbamoyltransferase).[88] (C) Curve 1, "fine structure plot" obtained by subtracting the spectrum in A from the smooth curve obtained by summing the two log normal curves. Curve 2, a similar plot for enzyme plus succinate and carbamoyl phosphate. The enzyme was supplied by G. Nagel and H. K. Schachman and the spectra were recorded by I.-Y. Yang.

smooth fitted curve and the experimental points taken at close intervals[90,91] as shown in Fig. 23-11C. The two "fine structure plots" obtained in this way are an alternative way of representing the same data that gave rise to the difference spectrum. The method has the advantage that information about the overall band shape is obtained from the computer-assisted curve fitting process. Thus, the binding of succinate and carbamoyl phosphate caused an almost insignificant shift (of 20 cm^{-1}) in the overall band position and a very slight broadening. The principal effect is an enhancement in the vibrational structure at 34,600 cm^{-1} in the 0–0 band of the two tryptophan residues present in the subunit. The cause of this change is not entirely obvious, a weakness of difference spectroscopy. Another way of examining a spectrum such as that in Fig. 23-11A is to plot a mathematical derivative of the curve. Both second[92] and fourth derivatives[91,93,94] yield curves similar to the difference plots of Fig. 23-11C.

5. Circular Dichroism and Optical Rotatory Dispersion

The circular dichroism of a sample is the difference between the molar extinction coefficients for left-handed and right-handed polarized light (Eq. 23-7) and is observed only for chiral molecules.[7,95–97]

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \quad (\text{units are } M^{-1}\ cm^{-1}) \tag{23-7}$$

The **dichrograph** gives a direct measure of $\Delta\varepsilon$. A circular dichroism (CD) spectrum often resembles an absorption spectrum, the peaks coming at the same positions as the peaks in the absorption spectrum of the same sample. However, the CD can be either positive or negative and may be positive for one transition and negative for another (Fig. 23-8). It is most convenient to plot $\Delta\varepsilon$ directly as a function of wavelength or wave number. However, much of the literature makes use of the **molar ellipticity** (Eq. 23-8):

$$\text{Molar ellipticity} = [\theta] = 3299\ \Delta\varepsilon \quad (\text{units are degrees } cm^2\ \text{decimole}^{-1}) \tag{23-8}$$

The **rotational strength** may also be evaluated (Eq. 23-9):

$$\text{Rotational strength} = \int [(\Delta\epsilon)/\lambda]d\lambda \tag{23-9}$$

The integration is carried out over the entire absorption band for a given transition.

Circular dichroism is closely related to **optical rotatory dispersion**, the variation of optical rotation with wavelength. Optical rotation depends upon the difference in refractive index ($\eta_L - \eta_R$) between left-handed and right-handed polarized light. Rotation α is measured as an angle in degrees or radians. Data are customarily reported in terms of **specific rotation**, that of a hypothetical solution containing 1 g / ml in a 1 dm (decimeter) tube. Specific rotation is calculated (Eq. 23-10) from the observed rotation, the concentration c' in g ml^{-1}, and the length of the tube l' in decimeters.

$$\text{Specific rotation} = [\alpha] = \alpha_{obs} / c'l' \tag{23-10}$$

The **molecular rotation** is defined by Eq. 23-11 in which M_r is the molecular mass and c and l are in moles per liter and cm, respectively.

$$\begin{aligned}\text{Molecular rotation} = [\phi] &= 100\ \alpha_{obs} / cl \\ &= [\alpha]\ M_r / 100\end{aligned} \tag{23-11}$$

It is often multiplied by a factor of $3/(\eta^2 + 2)$ to correct for a minor effect of the polarizability of the field acting on the molecules. The rotation in the radians per centimeter of light path can be related (Eq. 23-12) directly to the wavelength of the light and the refractive indices η_L and η_R.

$$\alpha\ (\text{radians / cm}) = [\alpha]\ c' / 1800 = \pi / \lambda[\eta_L - \eta_R] \tag{23-12}$$

In contrast to circular dichroism, optical rotary dispersion (ORD) extends far from absorption bands into spectral regions in which the compound is transparent. As an absorption band is approached, the optical rotation increases in either the positive or negative direction. Then, within the absorption band it drops abruptly through zero and assumes the opposite sign on the other side of the band (the Cotton effect). Although the occurrence of optical rotation in nonabsorbing regions of the spectrum provides an advantage to ORD measurements, the interpretation of ORD spectra is more complex than that of CD spectra. In principle, the two can be related mathematically and both are able to give the same kind of chemical information.[7]

The CD in the d–d bands of the blue copper protein (Fig. 23-8) arises in part from the fact that within the protein the copper ion is in an asymmetric environment. For a similar reason, the aromatic amino acids of proteins often give rise to circular dichroism. In the case of tyrosine, the sign of the CD bands can be either positive or negative but is the same throughout a given transition. The CD bands are similar in shape to the absorption bands.[36,98] The behavior of phenylalanine is more complex. The progression of vibrational subbands at 930 cm^{-1} intervals above the 0–0 band all have the same sign, and the intensities relative to that of the 0–0 band are similar to those in absorption. However, the vibrations of wave numbers equal to

that of the 0–0 transition plus 180 and 520 cm^{-1} sometimes give rise to CD bands of the opposite sign, and the relative intensity relationships are variable.[61,98]

The binding of a symmetric chromophore to a protein or nucleic acid often induces CD in that chromophore. For example, the bands of enzyme-bound pyridoxal and pyridoxamine phosphates shown in Fig. 14-9 are positively dichroic in CD, but the band of the quinonoid intermediate at 20,400 cm^{-1} (490 nm) displays negative CD. When "transimination" occurs to form a substrate Schiff base (Eq. 14-26), the CD is greatly diminished. While the coenzyme ring is known to change its orientation (Eq. 14-39; Fig. 14-10), it is not obvious how the change in environment is related to the change in CD.

A series of octant rules make it possible to predict the sign and magnitude of CD to be expected for $n-\pi^*$ transitions of simple carbonyl compounds.[99] Theoretical approaches to the CD and ultraviolet absorption of proteins in the high-energy ultraviolet region have also been developed. In a regular β structure, in an α helix or in a crystalline array, the transitions of adjacent amide groups may be **coupled**, the excitation energy being delocalized. This **exciton** delocalization leads to a splitting (Davydov splitting) into two transitions of somewhat different energies and polarized in different directions.[7,9] The amide absorption band at 52,600 cm^{-1} is split in an α helix into components at ~48,500 and 52,600 cm^{-1}. Furthermore, low-energy $\pi-\pi^*$ and $n-\pi^*$ states are close together in energy, a fact that allows mixing of the two states and appearance of rotational strength in the $\pi-\pi^*$ band with a sign opposite to that in the $n-\pi^*$ band.[100]

Both the sign and intensity of the CD bands of peptides also depend upon conformation. Well-defined differences are observed among α helices, β structure, and random-coil conformations. Measurements may be extended into the "vacuum ultraviolet" region—up to 60,000 cm^{-1} in aqueous solutions.[101] A useful empirical approach is to deduce spectra of helices, β structures, and unordered peptide chains from measured spectra together with an examination of actual structures obtained by X-ray crystallography[7,95,97,102,103] (Fig. 23-12). Note that the CD curve for the α helix has a deep minimum at 222 nm, whereas the β form has a shallower minimum. The random structure has almost no CD at the same wavelength. The approximate helix content of a protein is often estimated from the depth of the trough at 222 nm in the CD spectrum. Better predictions can be made by using a computer-assisted comparison of an experimental CD spectrum with those of a series of proteins of known 3-D structure.[7]

The circular dichroism of polynucleotides at 275 nm is a linear function of both the helix winding angle and the base pair twist.[103] Measurement of CD spectra on large polynucleotides or large molecular aggregates

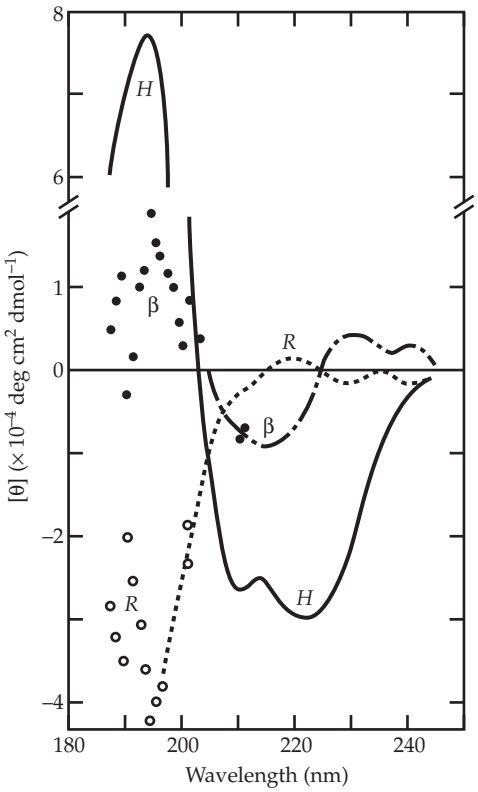


Figure 23-12 Circular dichroism of the helix (H), β, and unordered (R) form computed from the CD of five proteins. Points are plotted when a smooth curve could not be drawn. From Y.-H. Chen *et al.*[102]

is complicated by differential light scattering of right and left circularly polarized light. However, the phenomenon can also provide new structural information.[104] Progress has been made in attempts to predict the optical rotation of molecules from quantitative values for the polarizabilities of individual atoms.[105–109]

Vibrational circular dichroism involves IR absorption bands. The technique has been applied to sugars,[110] oligosaccharides,[111] proteins,[112] and nucleic acids.[113] The related **vibrational Raman optical activity** has also been applied to polyribonucleotides.[114]

6. Photoacoustic Spectroscopy

Photoacoustic spectra are recordings of the energy emitted as heat after absorption of monochromatic light. The sample is placed in a closed photoacoustic cell. The light beam, which is chopped at an audiofrequency, induces a periodic heating and cooling of the gas in contact with the sample in the cell. This is sensed as sound by a sensitive microphone. The

resulting electrical signals are sent to a computer for analysis. The output is an absorption spectrum resembling that measured optically. The samples do not have to be transparent.[5] A related technique is photoacoustic calorimetry.[115,116]

7. X-Ray Absorption and Mössbauer Spectroscopies

The importance of X-ray diffraction (discussed in Chapter 3) to biochemistry is obvious, but techniques related to absorption of X-rays and γ-rays have also come into widespread use.[117] Abbreviations such as **XANES** and **EXAFS** are common in the metalloprotein literature. The names arise from the sharp increase in the absorption coefficient for X-rays as their energy is increased to what is called the **K absorption edge**. At slightly lower energies absorption of an X-ray by an atom leads to expulsion of an electron or the raising of an electron to an excited state. Absorption of X-rays will expel all except the inner 1*s* electrons. As the energy is increased further, the stepwise increase in absorption that constitutes the edge is observed. At higher energies the absorption decreases. However, with a high-resolution instrument distinct oscillations are observed on the high-energy side of the edge, extending for ~20 eV. This is **X-ray absorption near-edge structure** (XANES). When an X-ray absorbing atom in a molecule is surrounded by other atoms, a fine structure that depends upon the nature of these atoms and their distances from the absorbing atom is observed over a range of several hundred electron volts above the edge. This is **extended X-ray absorption fine structure** (EXAFS).[117]

The EXAFS technique has been especially useful for metalloproteins. It has often provided the first clues as to the identity of atoms (O, N, S) surrounding a metal atom and either covalently bonded to it or coordinated with it (Chapter 16). Interpretations are often difficult, and a common approach is to try to simulate the observed spectrum by calculation from a proposed structure.[118] Tautomerism in crystalline Schiff bases (see Eq. 23-24) has been studied by near-edge X-ray absorption fine structure (NEXAFS) employing soft X-rays.[119]

Mössbauer spectroscopy, also called recoil-free nuclear resonance absorption, depends upon resonant absorption of γ-rays emitted by a radioactive source by atomic nuclei.[120] The phenomenon was initially difficult to observe, but the German physicist Mössbauer devised a way in which to record the absorption of a quantum of energy equal to the difference in two energy states of the atomic nucleus. The method depends upon a Doppler effect observed when the sample or source moves. Consequently, Mössbauer spectra, such as that in Fig. 16-18, are plots of absorption versus velocity. Mössbauer spectroscopy has been applied to numerous metalloproteins, especially those containing iron centers. It is a major tool in investigation of Fe–S proteins.[121,122] Since ^{56}Fe is "silent" in Mössbauer spectroscopy, proteins are often enriched with ^{57}Fe for observation.

C. Fluorescence and Phosphorescence

An electronically excited molecule is able to lose its excitation energy and return to the ground state in several ways. One of these is to reemit a quantum of light as fluorescence.[7,123–127] The intensity and spectral properties of fluorescent emission can be measured by illuminating a sample in a cuvette with four clear faces with the measuring photomultiplier set at right angles to the exciting light beam. In absorption spectrophotometry we measure a difference between the light intensity of the beam entering the sample and that emerging from the sample. In fluorescence spectroscopy we measure the absolute intensity of the light emitted. Although this intensity is small, the measurement can be made extremely sensitive, far more so than can light absorption. For this reason, fluorescence is widely used for detection and analysis, e.g., in DNA sequencers. Enzyme kinetics can be studied with fluorescent substrates at very low concentrations.[127a] Fluorescent antibodies, DNA chips, and numerous bioassay and imaging methods are dependent upon measurement of fluorescence. Fluorescence can also yield a wealth of information about the chemical and physical properties of electronically excited states of molecules.

1. Excitation and Emission Spectra

Measurements of the intensity of fluorescence at any wavelength vs the wavelength of monochromatic light used to excite the fluorescence give a fluorescence **excitation spectrum**. The excitation spectrum is an example of an **action spectrum**, which is a measure of any response to absorbed light. At very low concentrations of pure substances, action spectra tend to be identical to absorption spectra. However, since the observed response (fluorescence in this case) is proportional to light absorbed, action spectra should be compared to plots of 1–T (where T = transmittance, Section B,1) vs wavelength rather than to plots of ε vs λ. The two plots are proportional at low concentrations. For a discussion of action spectra see Clayton.[123]

A fluorescence **emission spectrum** is a record of fluorescence intensity vs wavelength for a constant intensity of exciting light. Excitation and emission spectra for a flavin and for the indole ring of tryptophan are both given in Fig. 23-13. The heights of the

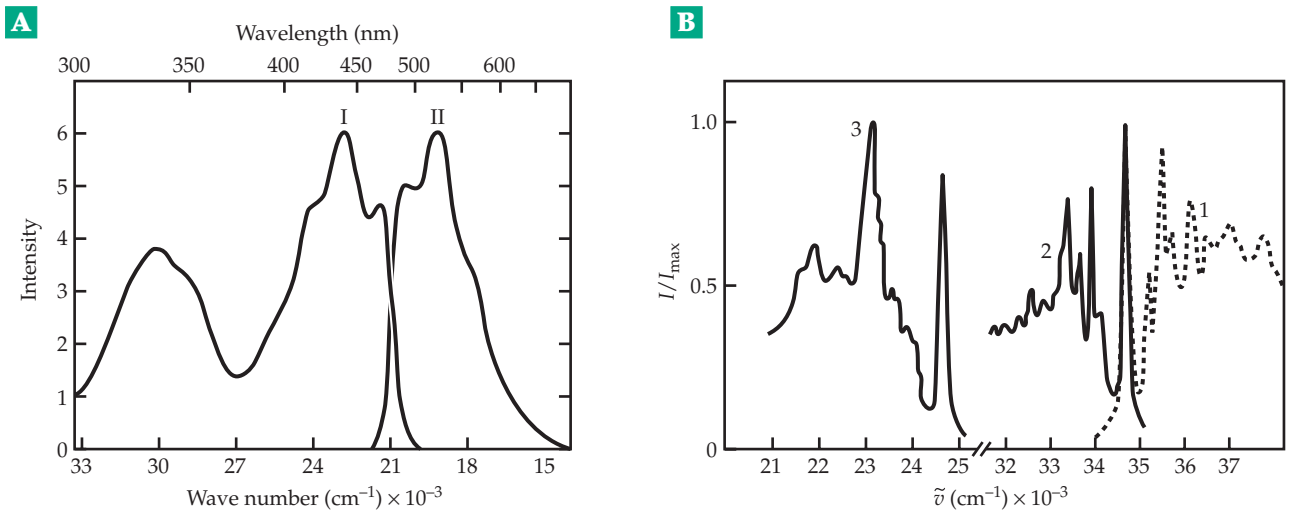


Figure 23-13 (A) Corrected emission and excitation spectra of riboflavin tetrabutyrate in *n*-heptane. Concentration, about 0.4 mg l^{-1}. Curve 1: excitation spectrum; emission at 525 nm. Curve 2: emission spectrum; excitation at 345 nm. From Kotaki and Yagi.[128] (B) Indole in cyclohexane, T = 196 K. 1, Fluorescence excitation spectrum; 2, fluorescence spectrum; and 3, phosphorescence spectrum. From Konev.[125]

emission spectra have been adjusted to the same scale as that of absorption. The fluorescent emission is always at a lower energy than that of the absorbed light. The excitation and emission spectra overlap only slightly, and the emission spectrum is an approximate mirror image of excitation spectrum. To understand this, refer to the diagram in Fig. 23-14. Absorption usually leads to a higher vibrational energy state after light absorption than before. However, most of the excess vibrational energy is dissipated before much fluorescent emission occurs. The excited molecule finds itself in the lowest vibrational state of the upper electronic state, and it is from this state that the bulk of the fluorescent emission takes place. Furthermore, whereas absorption usually occurs from the lowest vibrational state of the ground electronic level, fluorescence can populate many excited vibrational states of the ground electronic state (Fig. 23-14). Consequently, as indicated in the figure, the fluorescent emission spectrum consists of a series of subbands at lower energies than those observed in the absorption spectrum. The two spectra have only the 0–0 transition in common. As can be seen from Fig. 23-14, even the two 0–0 transitions do not coincide exactly. The peak of emission is shifted toward slightly lower energies than that of absorption because during or immediately following absorption of a photon there is some rearrangement of solvent molecules around the absorbing molecule to an energetically more stable arrangement. Just as excess vibrational energy is dissipated in the excited state, so relaxation of these solvent molecules around the excited chromophore leads to a small shift in energy. A similar relaxation

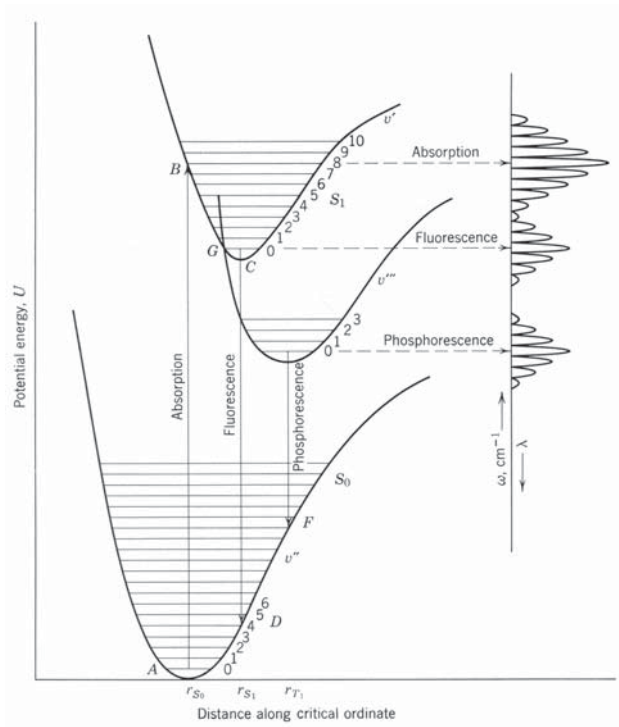


Figure 23-14 Potential energy diagram for the ground state S_0 and the first excited singlet S_1 and triplet T_1 states of a representative organic molecule in solution. G is a point of intersystem crossing $S_1 \rightarrow T_1$. For convenience in representation, the distances r were chosen $r_{S_0} < r_{S_1} < r_{T_1}$; thus, the spectra are spread out. Actually, in complex, fairly symmetric molecules, $r_{S_0} \sim r_{S_1} < r_{T_1}$ and the 0–0 absorption and fluorescence bands almost coincide, but phosphorescence bands are significantly displaced to the lower wavelengths. From Calvert and Pitts,[2] p. 274.

occurs in the ground state of a molecule that has just emitted a photon as fluorescence. This also contributes to the shift in position of the 0–0 band in fluorescence (see Parker,[124] p. 13).

Several molecular properties can be measured using emission and excitation spectra. These include fluorescence lifetime, efficiency, anisotropy of the emitted light, mobility of chromophores, rates of quenching, and energy transfer to other chromophores.

Fluorescence lifetimes. Why are some molecules fluorescent, while others are not? The possibility for fluorescent emission is limited by the radiative lifetime τ_r, which is related by Eq. 23-13 to the first-order rate constant k_f for exponential decay of the excited state by fluorescence.

$$\tau_r = 1/k_f \qquad (23\text{-}13)$$

The radiative lifetime is a function of the wavelength of the light and of the oscillator strength of the transition. For molecules absorbing in the near UV, the approximation of Eq. 23-14 is often made.

$$1/\tau_r \sim 10^4 \varepsilon_{max}) \qquad (23\text{-}14)$$

Thus, if $\varepsilon = 10{,}000$, the radiative lifetime (the time in which the fluorescence decays to $1/e$, its initial value) is $\sim 10^{-8}$ s (10 ns). If the absorption is more intense, the lifetime is shorter, and if it is less intense, it is longer. Other modes of deexcitation compete with fluorescence; therefore, the shorter the radiative lifetime the more likely that fluorescence will be observed.

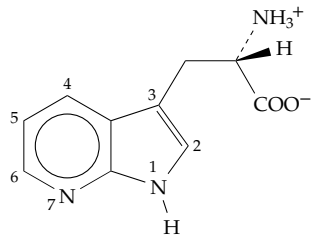
The actual lifetime τ of an excited molecule is usually less than τ_r because of the competing nonradiative processes. The sum of their rate constants can be designated k_{nr}. The **fluorescence efficiency** (or **quantum yield**) ϕ_F is given by Eq. 23-15.

$$\phi_F = k_f/k_f + k_{nr}) = k_r\tau \quad \text{where } \tau^{-1} = k_f + k_{nr} \qquad (23\text{-}15)$$

For a highly fluorescent molecule such as riboflavin, ϕ_f may be 0.25 or more.[129] For tryptophan in water it is about 0.14, and in proteins it varies from near zero to 0.35.[130]

Time-resolved fluorescence spectroscopy. The fluorescence lifetime τ can be measured with either of two different types of fluorometer.[7,127,131–133] **Pulse fluorometers** use pulsed lasers that can deliver pulses of light lasting as little as one picosecond or less. This permits rapid excitation and permits the direct observation of emitted light, using photon countings, over the entire range of time from a few picoseconds to milliseconds required for decay of the fluorescence.[131] The observed value of τ for riboflavin 5'-phosphate (ε_{max} = 12,200 at 450 nm) at 25°C is ~5 ns.[134] That for tryptophan is 3 ns.

Phase fluorometers utilize continuous irradiation by a beam of light that is sinusoidally modulated. If the frequency of the modulation is set correctly, there will be a phase difference in the modulation of the fluorescent emission that will depend upon τ. Phase fluorometry can yield the same information as does pulse fluorometry.[127,132,133] By using two or more modulation frequencies the decay rates and fluorescence lifetimes for various chromophores in a protein can be observed. For example, the protein **colicin A** (Box 8-D) contains three tryptophans W86, W130, and W140. Their fluorescence decays with lifetimes τ_1, τ_2, τ_3 of ~0.6–0.9 ns, 2.0–2.2 ns, and 4.2–4.9 ns at pH 7. While τ_3 originates mainly from W140, both of the other tryptophans contribute to τ_1 and τ_2. Changes in fluorescence intensity with pH reflect a pK_a value of 5.2.[135] Tryptophan, which often occurs at only one or a few places in a protein, is a useful fluorescent probe for study of protein dynamics. The optical properties of 7-azatryptophan, 2-azotryptophan, and 5-hydroxytryptophan are even better because their absorption maxima occur at longer wavelengths. These amino acids can be biosynthetically introduced in place of tryptophan in proteins.[136–138] The maximum fluorescence of tryptophan in one protein is at 350 nm, but for 7-azatryptophan in the same protein it was shifted to 380 nm.[136]



7-Azatryptophan

Triplet states, phosphoresence, and quenching. In addition to emitting fluorescent radiation, molecules can often pass from the excited singlet state to a lower energy **triplet state**, in which two electrons are now unpaired and the molecule assumes something of the character of a diradical (see Fig. 23-14). This process, known as **intersystem crossing**, competes directly with fluorescence and shortens the fluorescence lifetime. The triplet state is long-lived (e.g., for tryptophan in water at 20°C it is 1.2 ms[139]) and is responsible for much of the photochemical behavior of molecules. It also gives rise to the delayed light emission known as **phosphorescence**, as is illustrated in Figure 23-14. Other processes that compete with fluorescence are **photochemical reactions** of the singlet excited state and **internal conversion**. The

latter is the process by which a molecule moves from the lowest vibrational state of the upper electronic level to some high vibrational state of the unexcited electronic level. This is the principal means of depopulating the electronic state and competes directly with fluorescence.

The rate of relaxation by nonradiative pathways can be increased by addition of **quenchers**. Quenching of fluorescence occurs by several mechanisms, many of which involve collision of the excited chromophore with the quenching molecule. Some substances such as iodide ion are especially effective quenchers. The fluorescence efficiency of a substance in the absence of a quencher can be expressed (Eq. 23-16) in terms of the rate constants for fluorescence (k_f), for nonradiative decay (k_{nr}), and for phosphorescence (k_p):

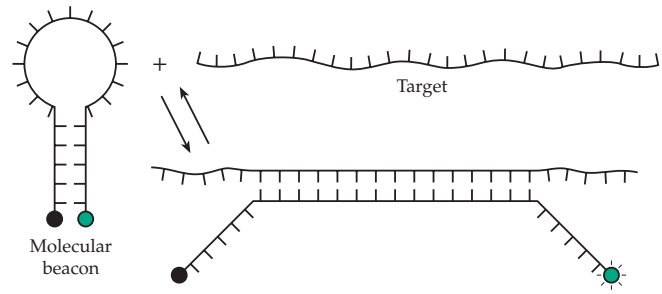
$$\phi_f = k_f / (k_f + k_{nr} + k_p) \qquad (23\text{-}16)$$

In the presence of a quencher, Q, there is an additional rate process for relaxation. The ratio of the fluorescence efficiency in the absence of (ϕ_f^0) and the presence of a quencher is given by the **Stern–Volmer equation**.[7,140]

$$\phi_f^0 / \phi_f = 1 + K[Q] = 1 + k_Q \tau_0 [Q] \qquad (23\text{-}17)$$

The constant K is known as the Stern–Volmer quenching constant; k_Q is the rate constant for the quenching reaction, and τ_0 the lifetime in the absence of quencher. Fluorescence quenching of tryptophan in proteins by acrylamide or O_2 has been used to determine whether tryptophan side chains are accessible to solvent or are "buried" in the protein.[141,142] The long-lived phosphorescence of tryptophan can be studied in a similar way.[143]

A recent application of fluorescence quenching is the development of "**molecular beacons**" for detection of viruses such as the AIDS viruses HIV-1 and HIV-2.[144] A single-stranded oligonucleotide is synthesized with a 25- or 33-nucleotide sequence complementary to a sequence in the target viral RNA. At the ends of this sequence are two 6-nucleotide arms with complementary sequences that will form a stable double-helical stem at the annealing temperature used for PCR amplification of the viral nucleic acid. The end of one arm carries a covalently bonded fluorescent dye, e.g., a fluorescein or rhodamine derivative. The other arm carries a potent covalently linked fluorescence quencher such as 4-(4'-dimethylaminophenylazo)benzoic acid. When the arms form a duplex, the quencher will be next to the fluorophore and no fluorescence will be seen upon irradiation with light of a suitable exciting wavelength. However, if viral DNA is present it will hydridize with the central polynucleotide, keeping the fluorophore and quenchers far apart



(23-18)

and allowing the beacon to signal the presence of a virus (Eq. 23-18). As few as ten retroviral genomes could be detected. By using a series of molecular beacons with different colored fluorescence and specific for different viruses, it is possible to test for more than one virus simultaneously.

Anisotropy. Light emitted from excited molecules immediately after absorption is always partially polarized, whether or not the exciting beam consists of plane polarized light. When light polarized in a vertical plane is used for excitation, part of the emitted light (of intensity I_v) will have its electric vector parallel to that of the exciting light. The remainder of intensity I_h, will be polarized in a horizontal plane. The **polarization *P*** of the emitted radiation is defined by Eq. 23-19 and the **anisotropy *R*** by Eq. 23-20. After excitation by a laser pulse both the fluorescence and its anisotropy decay with time and can be measured. The decay of R (but not of P) can usually be described as the sum

$$P = (I_v - I_h) / (I_v + I_h) \qquad (23\text{-}19)$$

$$R = (I_v - I_h) / (I_v + 2\, I_h) \qquad (23\text{-}20)$$

of simple exponential curves, which are readily obtained by phase fluorometry. These can in turn be related to specific types of motion, such as rotation of the emitting molecule or group.[7,145,146] Rotation of tryptophan rings, both free and restricted, has been studied in a variety of proteins.[145,147] However, interpretation is difficult.[130,148] The rotational rates obtained from anisotropy measurements are strongly affected by the viscosity of the medium (see Eq. 9-35).

2. Fluorescence Resonance Energy Transfer (FRET)

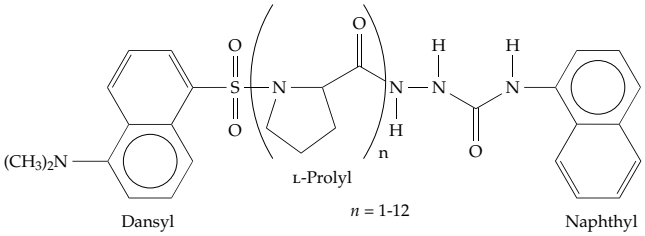
Electronic excitation of one chromophore sometimes elicits fluorescence from a different chromophore that is located nearby. For example, excitation of a monomolecular layer of dye can induce fluorescence in a layer of another dye spaced 5 nm

away. Excitation of tyrosine residues in proteins can lead to fluorescence from tryptophan, and excitation of tryptophan can cause fluorescence in dyes attached to the surface of a protein or in an embedded coenzyme.[134] Such fluorescence resonance energy transfer (**FRET**) is expected for molecules, when the fluorescence spectrum of one overlaps the absorption spectrum of the other. The mechanism is not one of fluorescence emission and absorption but of nonradiative resonant transfer of energy. Resonant transfer of energy is of major biological significance in photosynthesis (Section E). Most of the chlorophyll molecules, which absorb light in a chloroplast, transfer the absorbed energy in a stepwise fashion to a **reaction center**.

Förster[149] calculated that the rate of energy transfer k_t should be proportional to the rate of fluorescence k_f, to an orientation factor K^2, to the spectral overlap interval J, to the inverse fourth power of the refractive index n, and to the inverse sixth power of the distance r separating the two chromophores.

$$k_t \propto k_f K^2 J n^{-4} r^{-6} \qquad (23\text{-}21)$$

Besides predicting the inverse sixth power dependence of energy transfer, Förster provided a formula for calculating R_0, the distance between chromophores at which 50% efficient singlet-singlet energy transfer takes place. R_0 is commonly of the order of 2.0 nm. Making use of these relationships, Stryer proposed a method of measuring distances between chromophores. He calibrated the method by constructing a series of molecules containing various lengths of the rigid threefold polyproline helix to which dansyl groups were attached at one end and naphthyl groups



at the other.[150] By exciting the naphthyl group, which has the higher energy absorption band and is strongly fluorescent, the characteristic lower energy emission of the dansyl group could be observed if energy transfer took place. Since the fluorescent emission band of the naphthyl group overlaps the absorption band of the dansyl group, efficient transfer was expected. The results of a plot of transfer efficiency against distance is shown in Fig. 23-15. The inverse sixth power dependence was followed quite accurately with a value of R_0 ~3.4 nm. Having calibrated his "spectroscopic ruler," Stryer turned his attention to biochemical macromolecules. Attaching the same kinds of fluorescent probe to the visual light receptor rhodopsin, Wu and Stryer were able to estimate distances between specific parts of the molecule and to draw some conclusions about the overall shape.[151]

More recently the FRET technique has been widely applied to a broad range of biochemical problems. Sensitivity has been improved to the extent that fluorescence of single molecules can be detected.[152–154] Use of **terbium** (Tb^{3+}) or **europium** (Eu^{3+}) ions, which can provide luminous labels for metal-binding sites, has provided another advance. These ions absorb light poorly and are therefore only weakly fluorescent. However, they can be excited by resonance energy transfer and become brilliantly luminous. This **luminescence resonance energy transfer** (**LRET**) is a variant of FRET, which allows distances up to ~10 nm to be measured.[155–157] Another advance is the ability to graft into specific proteins fluorescent tags such as the intact **green fluorescent protein** (Section J)[158,159] or an amino acid sequence such as CCXXCC in which the four –SH groups of the cysteines serve to trap an arsenic derivative of fluorescein (see Box 12-B).[159] **Confocal laser scanning microscopy** (Chapter 3) is basic to many applications.[160]

Specific applications of FRET and LRET include observation of myosin movement (Fig. 19-14),[157] measurement of distances between binding sites on tubulin,[161] determining stoichiometry of subunit assembly in a γ-aminobutyrate receptor of brain,[162] association of proteins in peroxisomes,[160] study of hydridization of deoxyribonucleotides,[163] verifying the handedness of various forms of DNA,[164] and other studies of DNA and RNA.[164a,b]

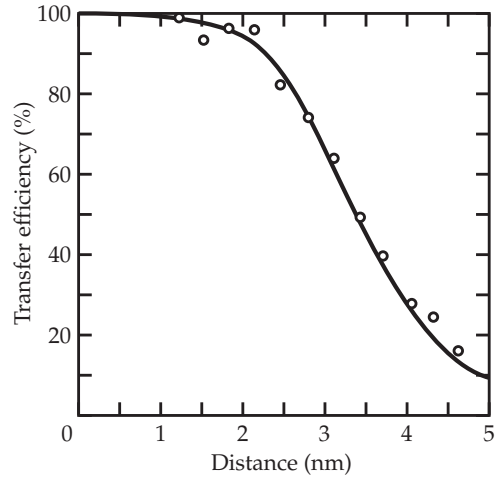


Figure 23-15 Efficiency of energy transfer as a function of distance between α-naphthyl and dansyl groups at the ends of a polyprolyl "rod" (L-prolyl)$_n$. The observed efficiencies of transfer for n = 1 to 12 are shown as points. The solid line corresponds to an r^{-6} distance dependence. From L. Stryer.[150]

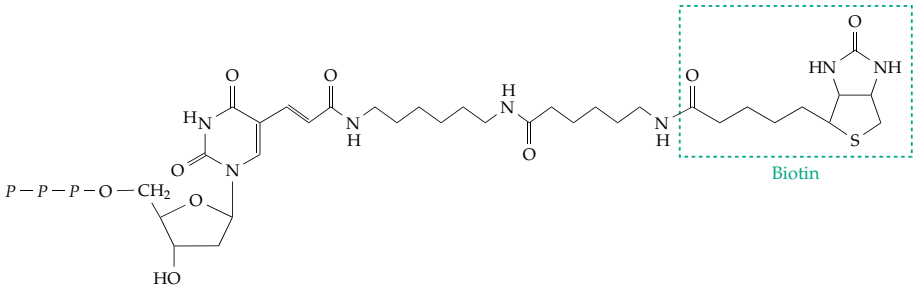
Using the Förster equation the distance between the two calcium-binding sites in parvalbumin (Fig. 6-7) has been estimated by energy transfer from Eu(III) in one site to Tb(III) in the other[165] to within 10–15% of the distance of 1.18 nm based on X-ray crystallography.

3. Energy-Selective Spectroscopic Methods

At low enough temperatures vibrational fine structure of aromatic chromophores may be well resolved, especially if they are embedded in a suitable matrix such as argon or N_2, which is deposited on a transparent surface at 15 K. This **matrix isolation spectroscopy**[77,166] may reveal differences in spectra of conformers or, as in Fig. 23-16, of tautomers. In the latter example the IR spectra of the well-known amino-oxo and amino-hydroxy tautomers of cytosine can both be seen in the matrix isolation IR spectrum. Figure 23-16 is an IR spectrum, but at low temperatures electronic absorption spectra may display sharp vibrational structure. For example, aromatic hydrocarbons dissolved in *n*-heptane or *n*-octane and frozen often have absorption spectra, and therefore fluorescence excitation spectra, which often consist of very narrow lines. A laser can be tuned to excite only one line in the absorption spectrum. For example, in the spectrum of the carcinogen 11-methylbenz(*a*)anthrene in frozen octane three major transitions arise because there are three different environments for the molecule. Excitation of these lines separately yields distinctly different emission spectra.[77] Likewise, in complex mixtures of different hydrocarbons emission can be excited from each one at will and can be used for estimation of amounts. Other related methods of energy-selective laser spectroscopy include **fluorescence line narrowing**[167] and **spectral hole burning**.[167,168]

4. Analytical Applications of Fluorescence

Because of the high sensitivity with which fluorescence can be detected, its measurement is important as an analytical tool. As a result of improved techniques **fluorescence microscopy** has become one of the most important of all tools in biological studies.[168a] New types of microscopes (see also pp. 129–131) have increased resolution beyond what was thought possible.[168b–d] Studies such as those of lipid metabolism in the transparent zebrafish are possible using substrates that carry fluorescent labels.[168e] As mentioned in the preceding paragraph, many aromatic compounds can be detected by their fluorescence. The relatively weak fluorescence of proteins and nucleic acids can be greatly enhanced by the binding of a highly fluorescent dye to the macromolecules. Fluorescent antibodies are widely employed for this purpose. Fluorescent labels are rapidly replacing radioisotopes in analysis of nucleic acids. For example, biotin may be attached to a pyrimidine base of a nucleoside triphosphate by a long spacer arm. The modified base can then be incorporated enzymatically into polynucleotides, e.g., in the synthesis of probes used for hybridization. The attached biotin can be detected by binding to avidin or streptavidin (Box 14-B) and use of fluorescent antibodies to this



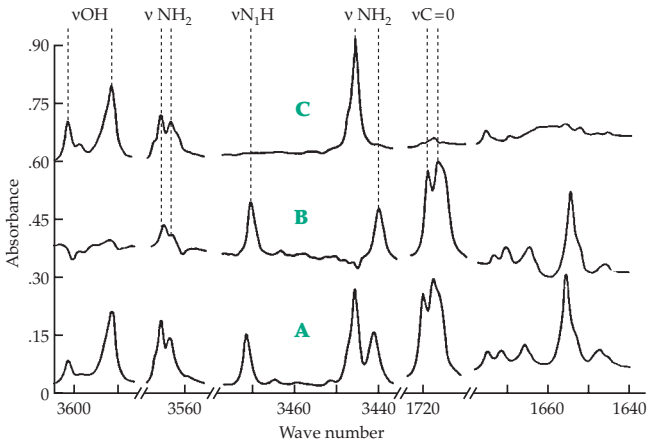
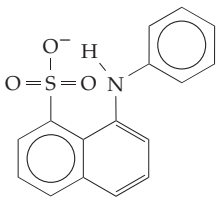


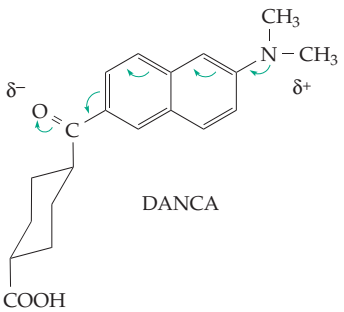
Fig. 23-16 Infrared spectrum of cytosine in the –OH, NH, and C=O stretching region in an argon matrix in 15 K. (A) Initial spectrum of a sample containing both tautomers. (B) Spectrum of the amino-oxo tautomer calculated by substraction of the spectrum of the amino-hydroxy tautomer. (C) Spectrum of the aminohydroxy tautomer obtained by irradiating the initial sample in order to photochemically convert it into the aminohydroxy tautomer. Spectra (B) and (C) were also corrected for residual amounts of the second tautomer. From Szczesniak *et al.*[166] Courtesy of W. B. Person.

protein.[169,170] Fluorescent dyes can also be covalently attached to nucleotides. Fluorescent dideoxynucleoside triphosphates are used as chain terminators in DNA sequencing (Chapter 5). Using a different dye that fluoresces at a different wavelength for each of the four dideoxynucleosides, polynucleotides can be sequenced automatically using a single column rather than four parallel lanes as in Fig. 5-49.

Fluorescent "probes" such as **1-anilinonapthalene-8-sulfonate** or 1,6-diphenyl-1,3,5-hexatriene embedded in membranes, contractile fibers, etc., can reveal changes in mobility that accompany alterations in physiological conditions. For example, molecular changes occurring in membranes during nerve conduction and in mitochondria during electron transport



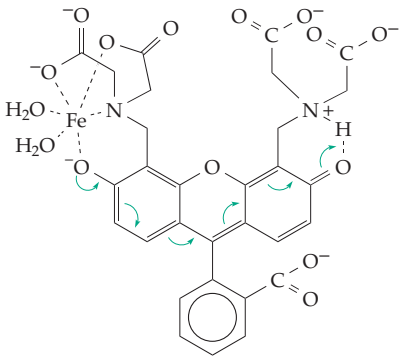
1-Anilinonaphthalene-8-sulfonate



DANCA

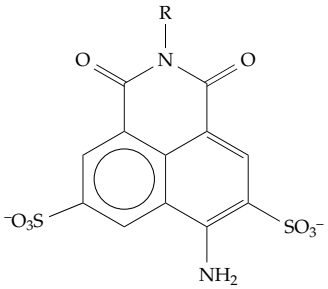
can be observed.[146,171] Another type of probe is exemplified by 2′-(*N*,*N*-dimethylamino)-6-4-*trans*-cyclohexamoic (DANCA). Its emission maximum shifts from 390 nm in cyclohexane to 520 nm in water, presumably as a consequence of increased polarization of the molecule in the excited state, as indicated by the green arrows in the accompanying structural formula.[172] DANCA can be used to obtain some idea of the polarity of sites within macromolecules to which it binds.

Study of calcium ions in living cells has been immensely aided by calcium fluorophores[173] (Box 6-D), which are often derivatives of EDTA (Table 6-9). One of these, **calcein**,[174] is also very specific toward Fe^{3+}. The natural calcium-dependent luminous protein, aequorin (Section J), is also widely used.



Fe^{III} - calcein • H^+

Nontoxic but highly fluorescent dyes are used to study diffusion within and between cells. The so called Lucifer dyes have been employed to trace shapes and branching patterns in neurons.[175]



Structural formula for Lucifer dyes
For Lucifer yellow
R = —NH—CO—NH—NH_2

The recently developed **fluorescence correlation spectroscopy** permits studies of molecular association in one femtoliter of solution using a confocal or two-photon microscope. Two lasers are used to excite two fluorophores of different colors, each one on a different type of molecule. Fluorescence of single molecules can be detected, and molecular associations can be detected by changes in the distribution of the fluctuations in fluorescence intensity caused by Brownian motion.[176–178] A different type of advance is development of computer programs that analyze chromosomes stained with a mixture of dyes with overlapping spectra and display the result as if each chromosome were painted with a specific color.[179–180a] Yet another advance is development of **semiconductor nanocrystals** (or "quantum dots") with narrow absorption bands and intense fluorescence. The wavelength of absorption and fluorescence depends upon the size of the crystals. For example, CdSe crystals of diameter 2–5 nm coated with silica or with a surfactant fluoresce across the visible range. They have a variety of uses in biological staining.[181,182]

D. Photochemistry

Because of their high energy, molecules in either the singlet or triplet photoexcited state undergo a greater variety of chemical reactions than do molecules in the ground state.[5,183,184] Many of these photochemical reactions arise from the triplet state that is formed from the singlet by intersystem crossing. Selection rules forbid transitions between excited triplet state and ground state; therefore, the radiative lifetime of the triplet state is long. The diradical character of the triplet state also makes it unusually reactive. Despite its forbidden character, nonradiative deexcitation of the triplet state is possible, and phosphorescence is observed for most molecules at low temperatures if the solvent is immobilized as a glass. The intense light

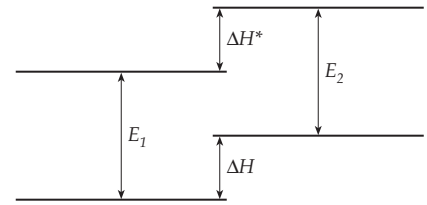
from lasers can also induce a variety of photochemical processes that arise from absorption of two or more photons.

One of the simplest of photochemical processes is the dissociation or uptake of a proton by an excited molecule as a result of a change in the pK_a value of a functional group. Various other bond cleavages may lead to dissociation into ions or radicals. Photoelimination and photoaddition reactions both occur. Molecules may be isomerized, a process of importance in visual receptors. Excited molecules may become strong oxidizing agents able to accept hydrogen atoms or electrons from other molecules. An example is the **photooxidation** of EDTA by riboflavin (which undergoes photoreduction as shown in Fig. 15-8). A biologically more important example is in photosynthesis, during which excited chlorophyll molecules carry out **photoreduction** of another molecule and are themselves transiently oxidized. A frustrating aspect of investigation of photochemical reactions is that the variety of reactions possible often leads to a superabundance of photochemical products, e.g., see the thin layer chromatogram of cleavage products of riboflavin in Fig. 3-5. However, biological photoprocesses are usually much more specific.

1. Chemical Equilibria in the Excited State

When pyridoxamine with a dipolar ionic ring structure (Fig. 14-9) and an absorption peak at 30,700 cm^{-1} (326 nm) is irradiated, fluorescence emission is observed at 25,000 cm^{-1} (400 nm). When basic pyridoxamine with an anionic ring structure and an absorption peak at 32,500 cm^{-1} (308 nm) is irradiated, fluorescence is observed at 27,000 cm^{-1} (370 nm), again shifted ~5500 cm^{-1} from the absorption peak. However, when the same molecule is irradiated in acidic solution, where the absorption peak is at 34,000 cm^{-1} (294 nm), the luminescent emission at 25,000 cm^{-1} is the same as from the neutral dipolar ionic form and abnormally far shifted (9000 cm^{-1}) from the 34,000 cm^{-1} absorption peak.[185,186] The phenomenon, which is observed for most phenols, results from rapid dissociation of a proton from the phenolic group in the photoexcited state. A phenolic group is more acidic in the excited state than in the ground state, and the excited pyridoxamine cation in acid solution is rapidly converted to a dipolar ion.

The variation of fluorescence intensity with pH can provide direct information about the pK_a in the excited state. Förster suggested the following indirect procedure for estimating excited-state pK_a values for phenols. Let E_1 represent the energy of the 0–0 transition (preferably measured as the mean of the observed 0–0 transition energies in absorption and fluorescent emission spectra); let E_2 represent the energy of the 0–0 transition in the dissociated (anionic in the case of a phenol) form, while ΔH and ΔH^* represent the enthalpies of dissociation in the ground and excited states, respectively. It is evident from the diagram that Eq. 23-22 holds.



$$E_1 - E_2 = \Delta H - \Delta H^* \qquad (23\text{-}22)$$

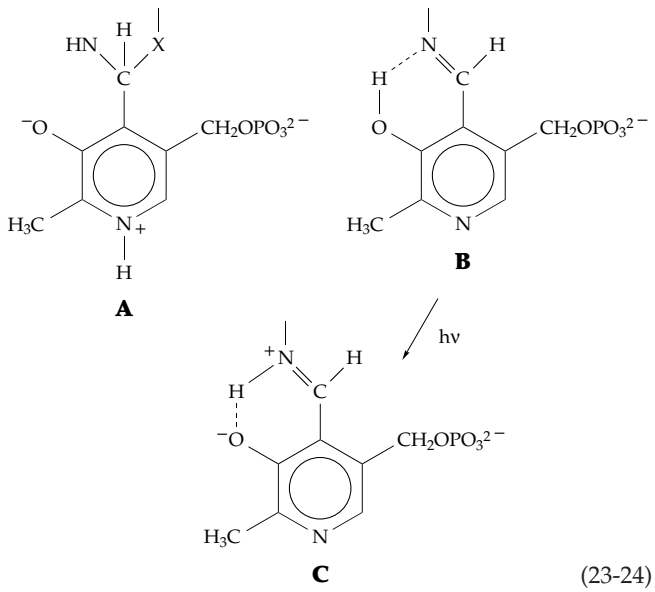
If we assume that the changes in entropy for the reaction are the same in the ground and excited state, Eq. 23-23 follows.

$$\log_{10}(K^*/K) = Nh\,(\Delta\,\bar{\nu})/(2.3RT)$$
$$\text{or}$$
$$pK^* = pK - (2.1 \times 10^3)\Delta\nu\ (\text{cm}^{-1}) \text{ at } 25°\text{C} \qquad (23\text{-}23)$$

For example, a shift in the spectrum of the basic form by 1000 cm^{-1} to a lower wave number compared with the acid form corresponds to a decrease of 2.1 units in pK_a for dissociation of the acid form. Whereas it is desirable to use both absorption and fluorescent measurements to locate the approximate positions of 0–0 bands, absorption measurements alone are often used, and the positions of the band maxima are taken. Thus, for pyridoxamine the shift in absorption maximum from 34,000 cm^{-1} in the protonated form to 30,700 cm^{-1} in the dissociated form suggests that the pK_a of pyridoxamine of 3.4 in the ground state is shifted by 6.9 units to –3.5 in the excited state. Bridges *et al.*[186] evaluated this same pK_a from the pH dependence of fluorescence as pK^* ~ –4.1.

While phenols and amines are usually more acidic in the singlet excited state than in the ground state, some substances, e.g., aromatic ketones, may become more *basic* in the photoexcited state.

Observation of an abnormally large shift in the position of fluorescent emission of pyridoxal phosphate (PLP) in glycogen phosphorylase answered an interesting chemical question.[187,188] A 330 nm (30,300 cm^{-1}) absorption band could be interpreted either as arising from an adduct of some enzyme functional group with the Schiff base of PLP and a lysine side chain (structure A) or as a nonionic tautomer of a Schiff base in a hydrophobic environment (structure B, Eq. 23-24). For structure A, the fluorescent emission would be expected at a position similar to that of pyridoxamine. On the other hand, Schiff bases of the



(23-24)

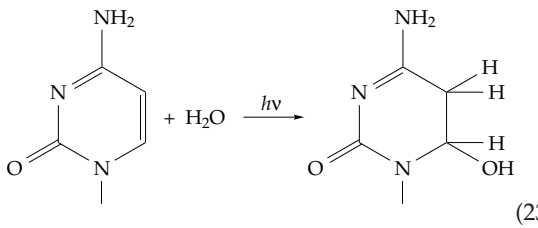
type indicated by structure B would be expected to undergo a photoinduced proton shift (phototautomerization) to form structure C[124,187] with an absorption band at 430 nm (23,300 cm^{-1}) and fluorescent emission at a still lower energy. Since the observed fluorescence was at 530 nm, it was judged that the chromophore does have structure B.

The rate of proton dissociations from the excited states of molecules can be measured directly by nanosecond fluorimetry.[189]

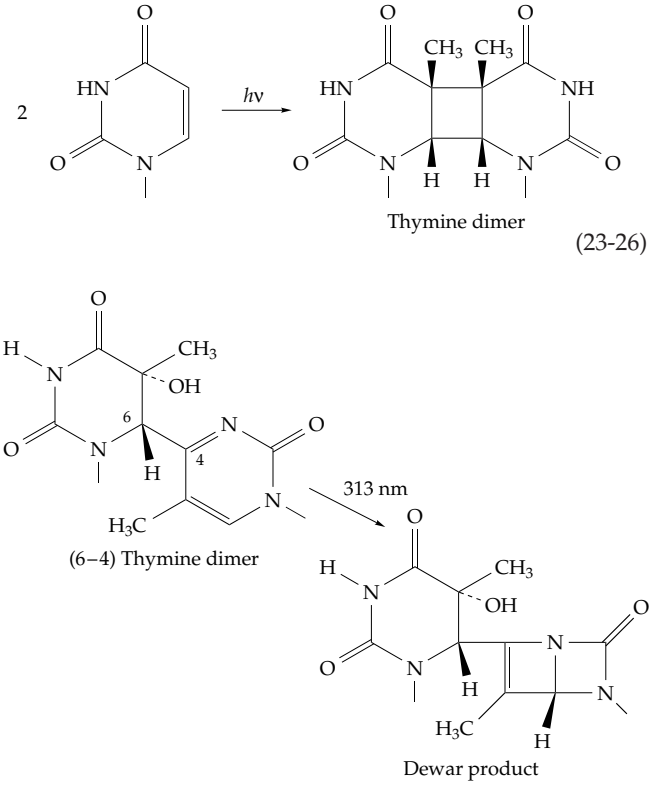
2. Photoreactions of Nucleic Acid Bases

Photochemical reactions of the purines and pyrimidines assume special significance because of the high molar extinction coefficients of the nucleic acids present in cells. Light is likely to be absorbed by nucleic acids and to induce photoreactions that lead to mutations.[190] Both pyrimidines and purines undergo photochemical alterations, but purines are only about one-tenth as sensitive as pyrimidines. **Photohydration** of cytidine (Eq. 23-25) is observed readily. The reaction is the photochemical analog of the hydration of α,β-unsaturated carboxylic acids. Uracil derivatives also undergo photohydration.

A more important reaction is the photodimerization of thymine (Eq. 23-26), a reaction also observed with uracil. A variety of stereoisomers of the resulting



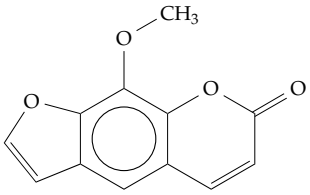
(23-25)



(23-26)

cyclobutane-linked structure are formed. The one shown in the equation predominates after irradiation of frozen thymine. Another important isomer is the 6–4 thymine dimer.[191,191a] Both of these types of **cyclobutane dimers** block DNA replication. This accounts for much of the lethal and mutagenic effect of ultraviolet radiation on organisms. The matter is sufficiently important that a special "excision repair" process is used by cells to cut out the thymine dimers (Chapter 27, Section E). In addition, light-dependent **photolyases**, discussed in Section I, act to reverse the dimerization reactions.

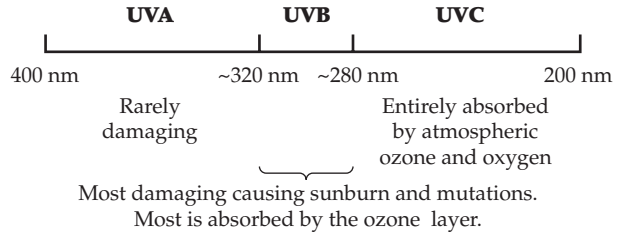
Light can also cause addition and other crosslinking reactions between DNA and proteins or other cell constituents.[192,193] One use of such reactions in the laboratory is DNA "photo footprinting" (Fig. 5-50), a technique which reveals contact regions between DNA and associated proteins.[190] Another type of cellular damage is caused by photosensitization of DNA by a light-absorbing intercalating agent such as **8-methoxypsoralen**.[194,195] DNA as well as adjacent proteins can be damaged.



8-Methoxypsoralen

3. Sunburn, Cancer, and Phototherapy

Ultraviolet light is sometimes classified according to its energy and capacity for damaging cells as follows:[195a]



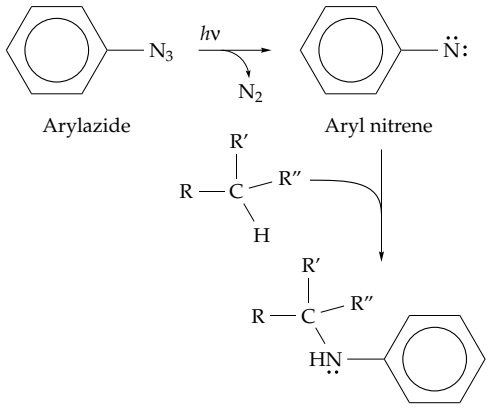
The UVB radiation causes most damage to skin. UVA radiation is at least an order of magnitude less damaging and is usually harmless to human skin. The UVC solar radiation is all absorbed by ozone and dioxygen of the atmosphere. UVC radiation produced by ultraviolet lamps is usually all absorbed in the epidermis.[5]

Ultraviolet light damages proteins as well as DNA. Residues of Trp, Tyr, His, Cys, and Met are especially susceptible to photolysis, or photooxidation by O_2, or by singlet oxygen. Also damaged are unsaturated lipids, porphyrins, flavins, etc. Kynurenic acid (Fig. 25-11) and urocanic acid (Eq. 14-44), an important ultraviolet filter in skin,[196] are also decomposed by light.

Repeated sunburn ages skin and may induce cancer.[4] However, light also has beneficial effects. It allows us to see, provides a source of vitamin D (Box 22-C), induces enzymatic repair of some DNA damage, and provides all of our food energy, directly or indirectly, by photosynthesis.

Light is used in **phototherapy**. This happens most frequently in the irradiation of newborn babies with white light to isomerize bilirubin (Fig. 24-24) from the 4*Z*, 15*Z* form to more readily excreted forms such as the 4*E*, 15*Z* isomer. About half of newborns have some **jaundice** (elevated bilirubin), and if it is severe it must be treated promptly to avoid neurological damage (see Chapter 24). There are sometimes complications, but the treatment is usually effective.[5] **Psoriasis** is frequently treated by irradiation with UVB light, which is thought to inhibit growth of the abnormal skin cells. Ten to 35 treatments are usually required, and the condition may return after some years. An alternative treatment is irradiation with UVA light after ingestion of 8-methoxypsoralen or other psoralen derivative.[5] Another skin condition that may respond to phototherapy is **vitiligo** (Box 25-A). Treatment with UVA and psoralen derivatives may stimulate repigmentation. If depigmentation is extreme, the remaining pigmentation may be reduced by bleaching with the monobenzyl ester of hydroquinone.

Photodynamic therapy is a cancer treatment that involves intravenous injection of a light-absorbing molecule such as a porphyrin, which may be taken up preferentially by cancer cells. Laser irradiation by deeply penetrating red light (650–800 nm wavelength) causes oxygen-dependent photosensitization.[5,197,198] Improvements in lasers, in fiber optics, and in photosensitizers may lead to widespread use of this type of therapy both for cancer and for some other conditions.[5,199]



(23-27)

4. Photoaffinity Labels

Photochemically reactive molecules have often been used as labels for specific sites in proteins and nucleic acids. Psoralen derivatives serve as relatively nonspecific photochemically activated crosslinking agents for DNA and double-stranded RNA.[195] **Aryl azides** are converted by light to aryl nitrenes, which react in a variety of ways including insertion into C–H bonds (Eq. 23-27).[200,201] In some cases UV irradiation can be used to join natural substrates to enzymes or hormones to receptors. For example, progesterone, testosterone, and other steroids have been used for direct photoaffinity labeling of their receptors.[202] Synthetic **benzophenones** have also been used widely as photoactivated probes.[203]

5. Microphotolysis and Ultrafast Light-Induced Reactions

Fluorescence microphotolysis, or photobleaching, has been widely used to study translational mobility of lipids and proteins in membranes. An attenuated laser beam may be focused down to the diameter of a cell or less. Then the intensity can be suddenly increased by several orders of magnitude, bleaching any fluorescent material present. The return of fluorescent material by free diffusion from a neighboring region (**fluorescence recovery after photobleaching**) or by diffusion through a membrane into a cell can then be

observed.[204,205] Diffusion coefficients of labeled biopolymers or of components of cells can be evaluated, and translation and metabolism of lipids and other components can be followed.[206]

Laser-based techniques are being used for ultrafast observation of the results of a photochemical process, e.g., the light-induced dissociation of CO from the hemoglobin • CO complex. A dissociating laser pulse can be as short as 100 fs (0.1 ps) or less. This is shorter than the time of vibrational motion of nuclei in an electronically excited state (~0.3 ps). Using IR spectroscopy, events that follow can be observed at intervals as short as 0.1 ns.[207] X-ray diffraction measurements using 150 ps pulses have allowed direct observation of the CO dissociated from hemoglobin or myoglobin and its recombination with the same protein.[208–210] Femtosecond dynamics of electron transfer along a DNA helix is also being studied.[211,212]

6. Optical Tweezers, Light-Directed Synthesis, and Imaging

The radiation pressure exerted by light is very weak. A bright laser beam of several milliwatts of power can exert only a few piconewtons (pN) of force. However, a force of 10 pN is enough to pull a cell of *E. coli* through water ten times faster than it can swim.[213] In about 1986, it was found that a laser beam focused down to a spot of ~ one λ (~1 μm for an infrared laser) can trap and hold in its focus a refractile bead of ~1 μm diameter. This "optical tweezers" has become an important experimental tool with many uses.[213,214] For example, see Fig. 19-19. Not only are optical tweezers of utility in studying biological motors but also mechanical properties of all sorts of macromolecules can be examined. For example, DNA can be stretched and its extensibility measured.[215] Actin filaments have even been tied into knots![216]

Light-directed solid state synthesis of peptides[217] and oligonucleotides is another new tool.[218] Development of this technology may provide new advances in preparation of DNA chips (Chapter 26) with a higher density of components than are now available.

X-rays and more recently NMR (MRI)-imaging have become well known to us. It might seem impossible to use visible light for a similar type of imaging. However, a laser beam can pass through a person's head. Is it possible, using computer-based techniques, to create an image from the emergent light? Efforts are being made to do exactly this.[219,220] **Optical coherence tomography** using backscattered infrared light and related fast techniques have higher resolution than MRI, computerized tomography (CT), or ultrasound.[220–221a] Ultrabright synchrotron radiation is also being used in **infrared microspectrophotometry**.[222,223]

E. Photosynthesis

The photochemical reduction of CO_2 to organic materials[224–228] is the basic source of energy for the biosphere. Nevertheless, the process is limited to a few genera of photosynthetic bacteria (Table 1-1), eukaryotic algae, and higher green plants. Photosynthetic bacteria include the distinctly different purple, green, and bluegreen (cyanobacteria) groups, each of which has a different array of photosynthetic pigments. However, the basic mechanism of transduction of solar energy into chemical energy is the same in all of the bacteria and in green plants.

As discussed in Chapter 17, photosynthesis involves the incorporation of CO_2 into organic compounds by reduction with NADPH with coupled hydrolysis of ATP. This is most often via the Calvin–Benson cycle of Fig. 17-14. In a few organisms a reductive tricarboxylic acid cycle is employed. The idea that the chloroplasts of plants or the pigmented granules of photosynthetic bacteria generate NADPH or reduced ferredoxin plus ATP (Chapter 17) is now thoroughly accepted. However, it was not always obvious. Consider the overall equation (Eq. 23-28) for formation of glucose by photosynthesis in higher plants:

$$6\,CO_2 + 6\,H_2O \rightarrow 6\,O_2 + C_6H_{12}O_6 \qquad (23\text{-}28)$$

$$6\,CO_2 + 12\,H_2O^* \rightarrow 6\,O_2^* + C_6H_{12}O_6 + 6\,H_2O \qquad (23\text{-}29)$$

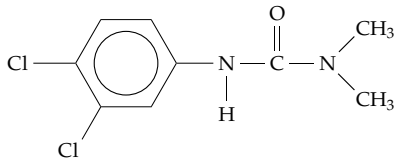
The stoichiometry of the reaction suggests that all 12 of the oxygen atoms of the evolved O_2 might come from CO_2 or that some might come from CO_2 and some from H_2O. In fact, water supplies both of the oxygen atoms needed for formation of O_2, as is indicated by the asterisks in Eq. 23-29. This possibility was suggested by van Niel[229,230] in 1931. He pointed out that in bacterial photosynthesis no O_2 is produced, and that bacteria must have access to a reducing agent to provide hydrogen for the reduction of CO_2 (Eq. 23-30).

$$H_2A \xrightarrow{h\nu} A + 2[H] \qquad (23\text{-}30)$$

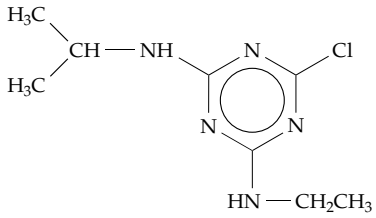
In this equation, H_2A might be H_2S (in the purple sulfur bacteria), elemental H_2, isopropanol, etc. From a consideration of these various reactions, van Niel concluded that in the O_2-producing cyanobacteria and eukaryotic plants water serves as the oxidizable substrate in Eq. 23-30 and is cleaved to form O_2 and to provide hydrogen atoms for reduction. This photochemical cleavage is the only known biological oxidation reaction of H_2O. No oxidizing agents present in living things are powerful enough to dehydrogenate water except for the photochemical **reaction centers** of photosynthetic organisms.

1. Two Photosystems, the Z Scheme, and Reaction Centers

It had long been known that for green plants light of wavelength 650 nm was much more efficient than that of 680 nm. However, Emerson and associates[34] in 1956 showed that a combination of light of 650 nm *plus* that of 680 nm gave a higher rate of photosynthesis than either kind of light alone. This result suggested that there might be two separate photosystems. What is now known as **photosystem I** (PSI) is excited by far red light (~700 nm), while **photosystem II** (PSII) depends upon the higher energy red light of 650 nm. Additional evidence supported the idea. Hill had shown many years before[231] that mild oxidizing agents such as ferricyanide and benzoquinone can serve as substrates for photoproduction of O_2, while Gaffron[232] found that some green algae could be adapted to photooxidize H_2 to protons (Eq. 23-30) and to use the electrons to reduce NADP. Thus, photosystem I could be disconnected from photosystem II. The powerful herbicide **dichlorophenyldimethylurea** (DCMU) was found to block electron transport between the two photosystems. In the presence of DCMU electrons from such artificial donors as ascorbic acid or an indophenol dye could be passed through photosystem I.



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



Atrazine, another PSII inhibitor

The Z scheme. The result of these and other experiments was the development of the series formulation or zigzag scheme of photosynthesis[233] which is shown in Fig. 23-17. Passage of an electron through the system requires two quanta of light. Thus, four quanta are required for each NADPH formed and eight quanta for each CO_2 incorporated into carbohydrate.

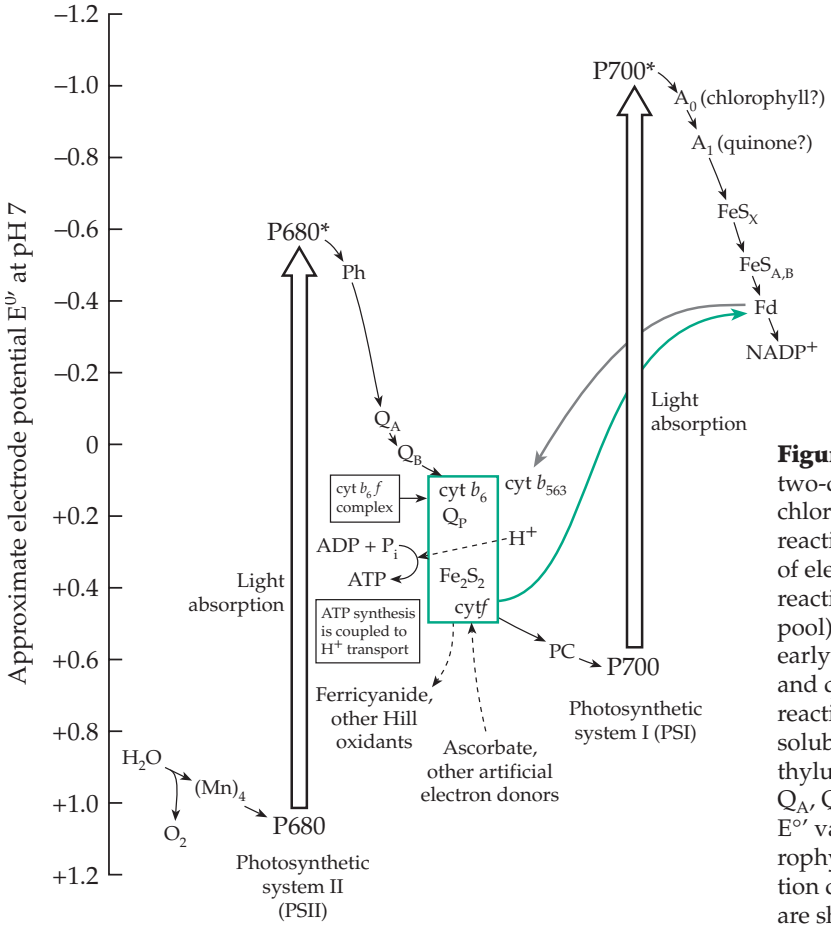


Figure 23-17 The zigzag scheme (Z scheme) for a two-quantum per electron photoreduction system of chloroplasts. Abbreviations are P680 and P700, reaction center chlorophylls; Ph, pheophytin acceptor of electrons from PSII; Q_A, Q_B, quinones bound to reaction center proteins; PQ, plastoquinone (mobile pool); Cyt, cytochromes; PC, plastocyanin; A_0 and A_1, early electron acceptors for PSI, possibly chlorophyll and quinone, respectively; F_x, Fe_2S_2 center bound to reaction center proteins; F_A, F_B, Fe_4S_4 centers; Fd, soluble ferredoxin; and DCMU, dichlorophenyldimethylurea. Note that the positions of P682, P700, Ph, Q_A, Q_B, A_0, and A_1 on the E°′ scale are uncertain. The E°′ values for P682 and P700 should be for the (chlorophyll / chlorophyll cation radical) pair in the reaction center environment. These may be lower than are shown.

An important experiment of Emerson and Arnold[35] employed very short flashes of light and measurement of the quantum efficiency of photosynthesis during those flashes. A striking fact was observed. At most, during a single turnover of the photosynthetic apparatus of the leaf, one molecule of O_2 would be released for each 3000 chlorophyll molecules. However, it could be calculated that for each O_2 released only about eight quanta of light had been absorbed. It followed that about 400 chlorophyll molecules were involved in the uptake of one quantum of light. This finding suggested that a large number of chlorophyll molecules act as a single light receiving unit (usually called a light-harvesting or antenna complex) able to feed energy to one **reaction center**. The concept is now fully accepted.

Electron transport and photophosphorylation. Two molecules of NADPH are required to reduce one molecule of CO_2 via the Calvin–Benson cycle (Fig. 17-14), and three molecules of ATP are also needed. How are these formed? The Z scheme provides part of the answer. There is enough drop in potential between the upper end of PSI and the lower end of PSII to permit synthesis of ATP by electron transport. It is likely that only one molecule of ATP is formed for each pair of electrons passing through this chain. Since, according to Fig. 17-14, one and a half molecules of ATP are needed per NADPH, some other mechanism must exist for the synthesis of additional ATP. Furthermore, many other processes in chloroplasts depend upon ATP so that the actual need for photogenerated ATP may be larger than this.

Arnon[234,235] demonstrated that additional ATP can be formed in chloroplasts by means of **cyclic photophosphorylation**: Electrons from the top of PSI can be recycled according to the dashed lines in Fig. 23-17. An electron transport system, probably that of the Z scheme, is used to synthesize ATP. As isolation of proteins and cloning of their genes progressed, it became clear that a complex of proteins known as **cytochrome b_6f** is closely related to the cytochrome bc_1 of mitochondria (Fig. 18-8).[236–237a] As in that complex, cytochrome b_6 carries two hemes, designated b_h and b_l with E_m of −84 and −158 mv, respectively. E_m values are for *Chlamydomonas* (see Fig. 1-11). Heme b_l is closer to the positively charged membrane surface (lumen side) and heme b_h is closer to the negative surface (stroma side). In maize cytochrome b_6 is a 23-kDa subunit, and the c-type cytochrome *f* is a larger ~ 34-kDa subunit whose heme is close to the lumenal side. Its E_m value is +330 mv. A 20-kDa Rieske Fe–S protein (Chapter 16) and an additional 17-kDa subunit complete the core four-subunit complex, which is found in green plants, green algae,[236] and cyanobacteria.[238,239] Other smaller subunits are also present. It is usually assumed that a Q cycle equivalent to that of mitochondrial complex III (Fig. 18-9) operates in the pumping of protons across the thylakoid membrane (Fig. 23-18).[237b,c] However, at high rates of photosynthesis the electron transfer may bypass cytochrome *f*.[237d] The associated ATP synthase is also subject to complex regulatory mechanisms.[237e] The Rieske protein is encoded by a nuclear gene, but genes for other subunits are chloroplastic. Electrons may be carried from the cytochrome *f* subunit to PSI by plastocyanin[240] or, in many algae and cyanobacteria, by the small **cytochrome c_6**.[241–242b] It is often synthesized when copper is inadequate for synthesis of plastocyanin. Figure 23-18 is a schematic view of PSI, PSII, and the intermediate cyt b_6f complex in a thylakoid membrane.

In spite of the close similarities in structures and function, there are distinct differences between cyto-

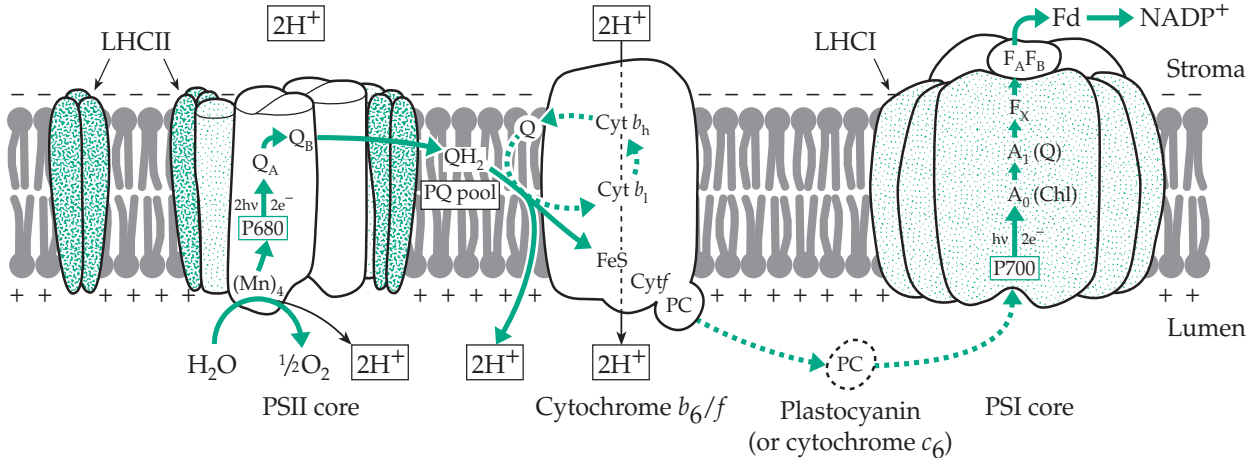


Figure 23-18 Schematic view of photosynthetic reaction centers and the cytochrome b_6f complex embedded in a thylakoid membrane. Plastocyanin (or cytochrome c_6 in some algae and cyanobacteria) carries electrons to the PSI core.

chrome b_6f and the cytochrome bc_1 of mitochondria.[243] Among these are the presence of stoichiometrically bound chlorophyll *a* and β-carotene[238,244] in the photosynthetic complex. The function of the chlorophyll is uncertain, but the carotene is probably there to quench the chlorophyll triplet state,[244] which would probably cause photodamage via formation of singlet O_2.

Bacterial photosynthesis. What is the relationship of the Z scheme of Fig. 23-17 to bacterial photosyntheses? In photoheterotrophs, such as the purple *Rhodospirillum*, organic compounds, e.g., succinate, serve as electron donors in Eq. 23-30. Because they can utilize organic compounds for growth, these bacteria have a relatively low requirement for NADPH or other photochemically generated reductants and a larger need for ATP. Their photosynthetic reaction centers receive electrons via cytochrome *c* from succinate ($E°'$ = + 0.03 V). The centers resemble PSII of chloroplasts and have a high midpoint electrode potential $E°'$ of 0.46 V. The initial electron acceptor is the Mg^{2+}-free bacteriopheophytin (see Fig. 23-20) whose midpoint potential is −0.7 V. Electrons flow from reduced bacteriopheophytin to menaquinone or ubiquinone or both via a cytochrome bc_1 complex, similar to that of mitochondria, then back to the reaction center P870. This is primarily a cyclic process coupled to ATP synthesis. Needed reducing equivalents can be formed by ATP-driven reverse electron transport involving electrons removed from succinate. Similarly, the purple sulfur bacteria can use electrons from H_2S.

In contrast, the reaction centers of green sulfur bacteria resemble PSI of chloroplasts. Their reaction centers also receive electrons from a reduced quinone via a cytochrome *bc* complex.[245] However, the reduced form of the reaction center bacteriochlorophyll donates electrons to iron–sulfur proteins as in PSI (Fig. 23-17). The latter can reduce a quinone to provide cyclic photophosphorylation. Cyanobacteria have a photosynthetic apparatus very similar to that of green algae and higher plants.

2. Chloroplast Structure

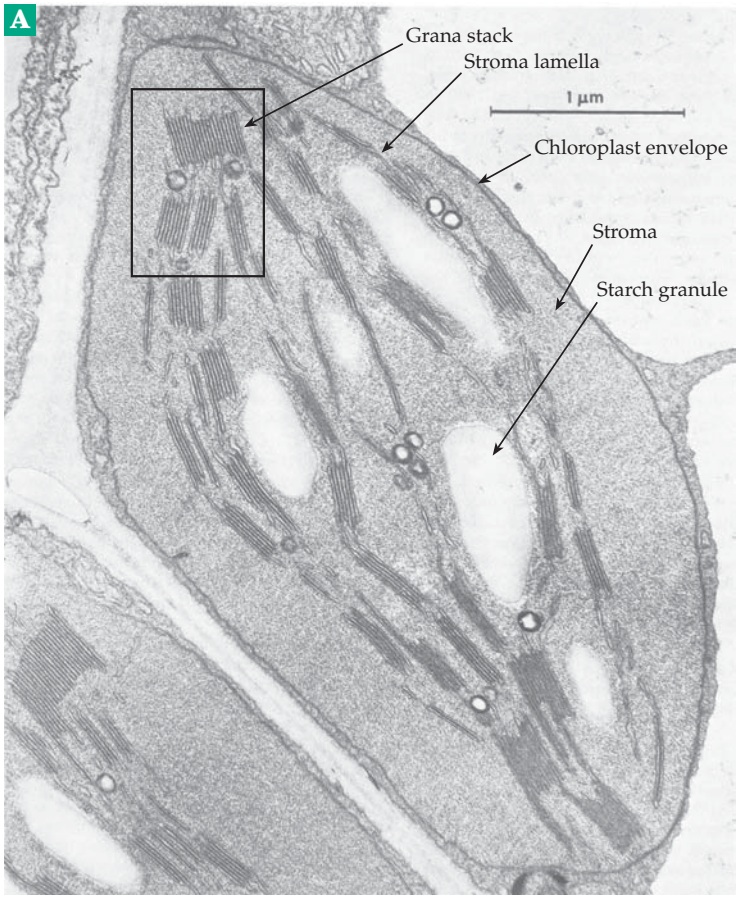
Chloroplasts come in various sizes and shapes, but all contain a small number of DNA molecules ranging in size from 120–160 kb. Complete sequences are known for DNA from chloroplasts of a liverwort (121,025 bp),[246] tobacco (155,844 bp),[247] maize (*Zea mays*),[248] and other plants. The 140,387 bp DNA from maize chloroplasts is a circular molecule containing the genes for 23S, 16S, 5S, and 4.5S RNA, for 30 species of tRNA, and for 70 different proteins. Among them are subunits of RNA polymerase, NADH dehydrogenase, subunits of both PSI and PSII, rubisco (large subunit), cytochromes *b* and *f*, six subunits of ATP synthase, and others. As with mitochondria, some subunits of the enzyme complexes that provide the cell with energy (e.g., ATP synthase) are encoded in the nucleus.

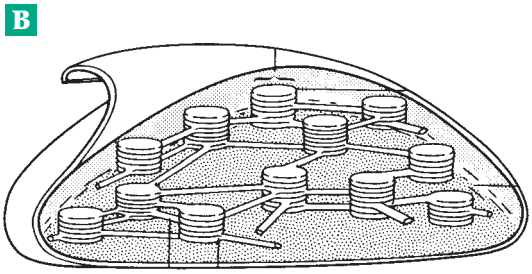
A characteristic of chloroplast genomes is the presence of a pair of large (in maize 22,348 bp) inverted repeat sequences. Since they can form a large hairpin structure with a very large loop, they may stabilize the gene sequence. The mutation rate within the repeat sequence is lower than in the single-stranded regions. The same genes are found in corresponding positions in both maize and rice. Both genomes have a number of sites of departure from the standard genetic code. These "editing sites" give rise to C→ U transitions in the RNA transcripts.

Most land plants have similar chloroplast DNA sequences, but considerable divergence is observed among algae.[249] For example, the red alga *Porphyra purpura* has 70 genes not found in chloroplasts of land plants. Each gene of the chloroplasts of the dinoflagellate *Heterocapsa triquetra* is carried on its own DNA minicircle.[250] However, ~2000 chloroplast proteins are encoded by nuclear DNA. The corresponding proteins are synthesized on cytoplasmic ribosomes and are transported into the chloroplasts.[249a] Some of these proteins must pass through both the double membrane of the envelope and the thylakoid membrane. As in mitochondria (Fig. 18-4) an array of different transport proteins are required. They are distinctly different from the mitochondrial transport proteins and involve their own unique targeting mechanisms.[251–253a]

Chloroplast membranes. Like the other energy-producing organelles, the mitochondria chloroplasts are surrounded by an outer double membrane or **envelope** and also contain an internal membrane system.[225–227,254–255a] Within the colorless **stroma** are stacks of flattened discs known as **grana** (Fig. 23-19). The discs themselves (the **thylakoids**) consist of pairs of closely spaced membranes 9 nm thick, each pair being separated by a thin internal space or loculus (Fig. 23-19). At least 75 different proteins are present in the isolated membranes. There is also a high content of **galactosyl diacylglycerol**, **digalactosyl diacylglycerol**, and **sulfolipid** (Chapter 8, Section A,4). Lipids account for half of the mass of thylakoid membranes.

Through the use of freeze-fracture and freeze-etching techniques of electron microscopy, it is possible to see, embeddded in the thylakoid membranes, particles which may represent individual **photosynthetic units** (also called quantosomes).[227,256–258] They are about 20 nm in diameter, and at least many of them presumably contain a reaction center surrounded by light-collecting chlorophyll–protein complexes. Others may represent the cytochrome b_6f complex and





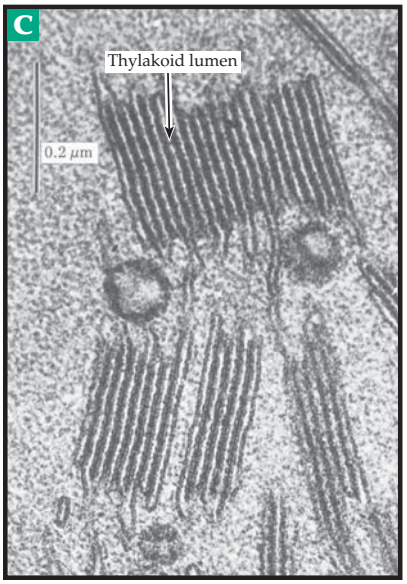


Figure 23-19 (A) Electron micrograph of alfalfa leaf chloroplast. Courtesy of Harry T. Horner, Jr., Iowa State University. (B) Schematic drawing of a chloroplast. From Hall and Rao[227] (C) Enlargement of a portion of (A) to show grana stacks more clearly.

ATP synthase, whose knobs also protrude into the stroma. A photosynthetic unit can also be defined chemically by the number of various types of molecules present in a chloroplast membrane for each four manganese atoms (Table 23-2). Separate units contain PSI and PSII. These reaction centers appear to have a different distribution within the thylakoids, the PSI units being located principally in the unstacked membranes and the PSII units in the grana stacks.[255,259]

Photosynthetic pigments and their environments. The chlorophylls (Fig. 23-20) are related in structure to the hemes (Figs. 16-5, 16-6), but ring IV (D) is not fully dehydrogenated as in the porphyrins. The **chlorin** ring system is further modified in chlorophyll by the addition of a fifth ring (V) containing an oxo group and a methyl ester. Ring V has been formed by crosslinking between the propionic acid side chain of ring III and a methine bridge carbon to give the parent compound **pheoporphyrin**. Chlorophylls contain constituents around the periphery that indicate a common origin with the porphyrins (Fig. 24-23). However, one of the carboxyethyl groups is esterified with the long-chain phytyl group in most of the chlorophylls. Chlorophyll *a* is the major pigment of chloroplasts and is a centrally important chromophore for photosynthesis in green plants. Most of the other chlorophylls, as well as the carotenoids and certain other pigments, are referred to as **accessory pigments**. Many of them have a light-receiving antenna function. Carotenoids are also photoprotectants. The relative numbers of pigment molecules in the photosynthetic units (average of PSI and PSII) of spinach chloroplasts are given in Table 23-2.

While the structure of chlorophyll *a* shown in Fig. 23-20 is the predominant one, other forms exist, e.g., with

$$-CH_2CH_3 \text{ or } -\underset{\underset{OH}{|}}{CH}-CH_3$$

replacing the vinyl group on ring I or with vinyl or hydroxyethyl replacing the ethyl group on ring II. The same kind of variation occurs for chlorophyll *b*.[260] In 80% acetone chlorophyll *a* has a sharp absorption band at 663 nm (15,100 cm^{-1}), but within chloroplasts the absorption maximum is shifted toward the red, the majority of the chlorophyll absorbing at 678 nm. Chlorophyll *b* (Fig. 23-20) is also nearly always present in green leaves.

The absorption peak in acetone is at 635 nm (15,800 cm^{-1}). Chlorophyll *c* found in diatoms, brown algae (Phaeophyta), and dinoflagellates (Fig. 1-9) lacks the phytyl group. Chlorophyll *d* contains a formyl group on ring I.[261]

Photosynthetic bacteria contain **bacteriochlorophylls** in which ring II is reduced (Fig. 23-20). The absorption band is shifted to the red from that of chlorophyll *a* to ~770 nm. The most abundant chlorophylls of green sulfur bacteria, **bacteriochlorophylls *c*, *d***, and ***e*** (or *Chlorobium* chlorophylls), contain a hydroxyethyl group on ring I; ethyl, *n*-propyl, or isobutyl groups on ring II; often an ethyl group instead of methyl on ring III; and a methyl group on the methine carbon linking rings I and IV. A variety of polyprenyl side chains can replace the phytyl group of the chlorophylls of higher plants.[262,263] The **pheophytins**, which are identical to the chlorophylls but lack the central magnesium ion, play an essential role in photosynthetic reaction centers. They can be formed in the laboratory by splitting the Mg^{2+} out from chlorophyll with a weak acid. Other derivatives are the **chlorophyllides** formed by hydrolysis of the methyl ester group and **chlorophyllins** formed by removal of both the methyl and phytyl groups.

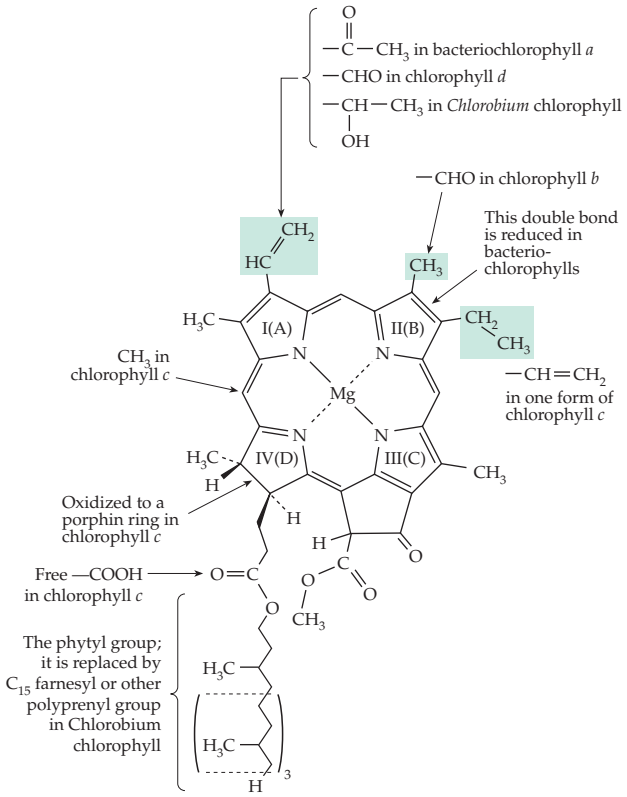
Since chlorophyll can be removed readily from chloroplasts by mild solvent extraction, it might appear that it is simply dissolved in the lipid portion of the membranes. However, from measurements of dichroism (Gregory,[226] p. 111) it was concluded that the chlorophyll molecules within the membranes have a definite orientation with respect to the planes of the thylakoids and are probably bound to fixed structures. The absorption spectrum of chlorophyll in leaves has bands that are shifted to the red by up to 900 cm^{-1} from the position of chlorophyll *a* in acetone. Most green plants contain at least four major chlorophyll bands at ~662, 670, 677, and 683 nm as well as other minor bands[264] (Fig. 23-21). This fact suggested that

TABLE 23-2
Approximate Composition of Photosynthetic Units in a Spinach Chloroplast[a]

Component	Number of molecules[b]
Chlorophyll *a*	160
Chlorophyll *b*	70
Carotenoids	48
Plastoquinone A	16
Plastoquinone B	8
Plastoquinone C	4
α-Tocopherol	10
α-Tocopherylquinone	4
Vitamin K_2	4
Phospholipids	116
Sulfolipids	48
Galactosylglycerides	490
Iron	12 atoms
Ferredoxin	5
Cytochrome b_{563}	1
Cytochrome b_{559}	
Cytochrome *f*	1
Copper	6 atoms
Plastocyanin	1
Manganese	2 atoms
Protein	928 kDa

[a] Averaged for PSI and PSII. After Gregory, R. P. F. (1971) *Biochemistry of Photosynthesis*, Wiley, New York [data of Luchtenthaler, H. K., and Park, R. B. (1963) *Nature* (*London*) **198**, 1070] and White, A., Handler, P., and Smith, E. L. (1973) *Principles of Biochemistry*, 5th ed., p. 528, McGraw-Hill, New York.

[b] Numbers of molecules assuming 2 Mn^{2+} ions per unit (4 for PSII and 0 for PSI).



Name of parent ring system:

without ring V closed:	chlorin
with ring V closed:	pheoporphyrin
with free — COOH on ring V:	chlorophyllides
with free — COOH on both rings II and V:	chlorophyllins
with Mg_{2+} replaced by 2 H^+:	pheophytins

Figure 23-20 Structures of the chlorophylls.

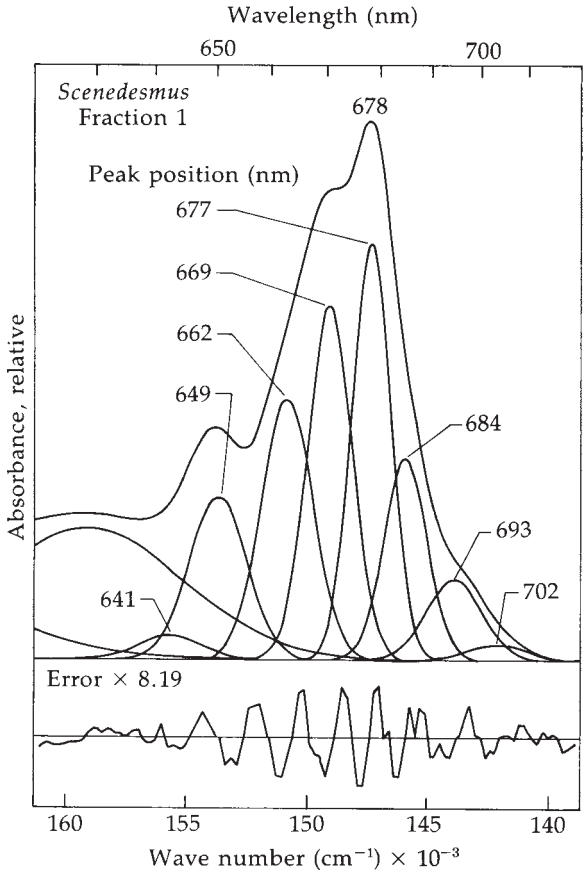


Figure 23-21 Absorption spectrum of chlorophyll in a suspension of chloroplast fragments from the green alga *Scenedesmus* showing the multicomponent nature of the chlorophyll environments. From French and Brown.[264]

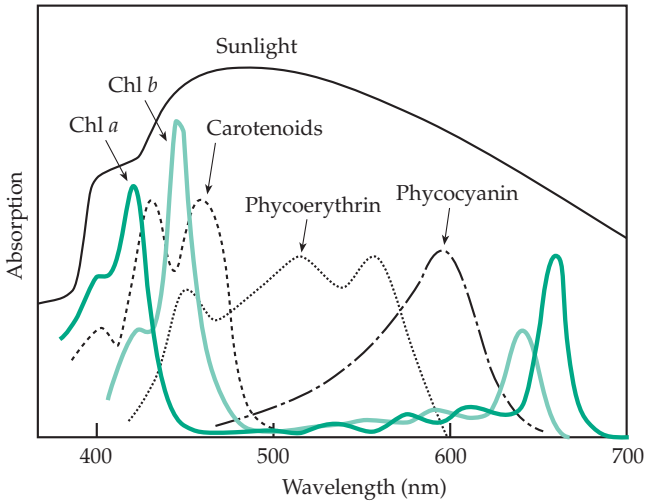


Figure 23-22 Absorption spectra of chlorophylls and accessory pigments compared. Redrawn from G. and R. Govindjee,[265] and from J. J. Wolken.[266]

the chlorophyll exists in a number of different environments. As a result, the absorption is spread over a broader region leading to more efficient capture of light. Only a small fraction of the total chlorophyll is in the reaction centers; that for PSI absorbs at ~700 nm and that for PSII at ~682 nm.

Bacteriochlorophyll in *Chromatium* has three absorption bands with peak positions at 800, 850, and 890 nm. The last includes the reaction center bacteriochlorophyll and is the only form that fluoresces. Recent studies have established that most if not all chlorophyll is bound to specific proteins, a fact that can account for the various overlapping absorption bands.

The **carotenes** and **carotenoids** are very important accessory pigments (Fig. 23-22). The major component in most green plants is β-carotene. Green sulfur bacteria contain γ-carotene in which one end of the molecule has not undergone cyclization and resembles lycopene (Fig. 22-5). Chloroplasts also contain a large variety of oxygenated carotenoids (xanthophylls). Of these, neoxanthin, violaxanthin (Eq. 22-12), and lutein (p. 1240) predominate in higher plants and green algae. *Euglena* and related microorganisms contain much antheraxanthin (Eq. 22-10). A light-collecting protein from marine dinoflagellates contains both chlorophyll *a* and several molecules of the carotenoid peridinin. See Fig. 23-29.[267,268] Brown algae and diatoms contain mostly fucoxanthin and zeaxanthin (Fig. 22-5), while the bacterium *Rhodospirillum rubrum* synthesizes spirilloxanthin (p. 1240).

It is a striking fact that there are no naturally occurring green plants that lack carotenoid pigments.[269] Carotenoidless mutants are used in photosynthesis research, but they apparently cannot survive under natural conditions. Carotenoids not only participate as members of the light-receiving complex but also confer protection to chlorophyll against light-induced destruction by singlet oxygen. This accounts for the fact that carotenoids are usually intimately associated with chlorophyll in the pigment complexes. For example, see Figs. 23-29 and 23-30 and discussion on pp. 1308–1310.

A third class of accessory pigment of more limited distribution are the **open tetrapyrroles**, sometimes called "plant bile pigments" because of their relationship to the pigments of animal bile (Fig. 24-24). Among these are the **phycocyanins**, which provide the characteristic color to cyanobacteria. They are conjugated proteins (biliproteins) containing covalently bound phycocyanobilin (Fig. 23-24).[270] The red **phycoerythrins** of the Rhodophyta contain bound phycoerythrobilin (Fig. 23-23), an isomer of phycoerythrobilin. There are four common isomeric **bilins**, each having a different number of conjugated double bonds.[272,273] Together, they provide for a broad range

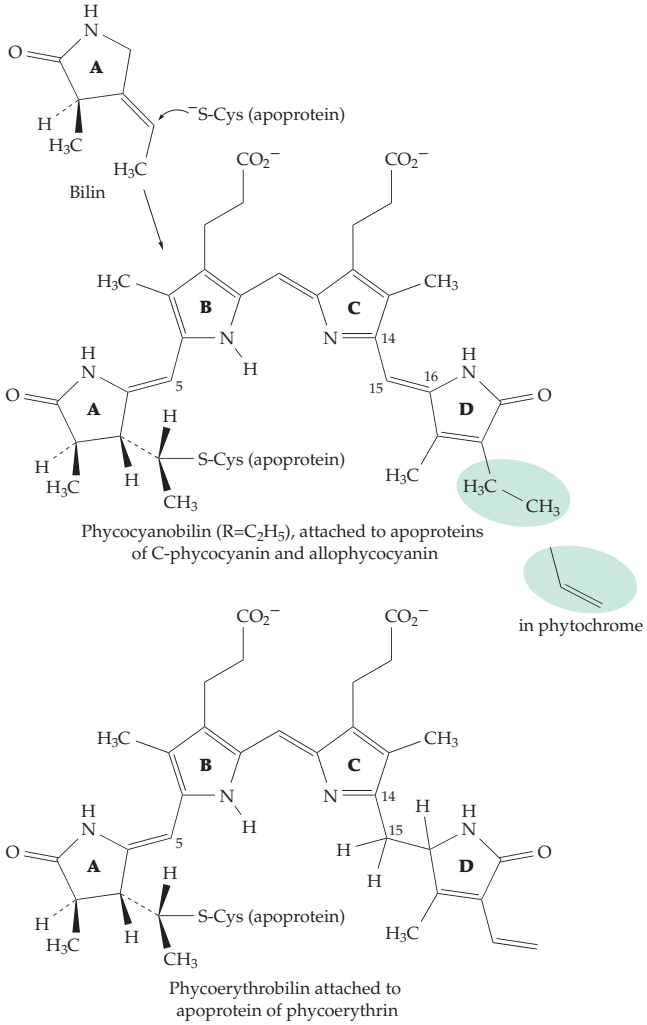


Figure 23-23 Structures of the open tetrapyrroles of plants. See also Fig. 24-24. After Szalontai *et al.*[271]

of colors from blue to red (Table 23-3). The bilins are attached to proteins via addition of a cysteine –SH group to the vinyl group of ring A of the bilin (Fig. 23-23). A double attachment may be formed by addition of –SH groups to both vinyl groups.[273–276] Isolated tetrapyrrole pigments tend to have a helical structure and to absorb light at lower wavelengths than do the protein-bound pigments which assume elongated conformations.

There are three major classes of conjugated phycobiliproteins,[273,277] all of which are αβ heterodimers often associated as $(\alpha\beta)_6$ (Fig. 23-24). The **allophycocyanins** carry one bilin per subunit, the phycocyanins carry one on the α and two on the β subunit, and the phycoerythrins carry two or three on the α subunit and three on the β (Fig. 23-24). Cysteine α-84 is one of the frequent attachment sites.[273] Three-dimensional structures are known for several of these proteins[278–281]

TABLE 23-3
The Common Bilin Pigments Present in Phycobiliniproteins[a]

Isomer	Number of conjugated double bonds	Absorption maximum (nm) when conjugated to proteins
Phycocyanobilin	8	~640
Phycobiliviolin	7	~590
Phycoerythrobilin	6	~550
Phycourobilin	5	~490

[a] Wedemayer, G. J., Kidd, D. G., Wemmer, D. E., and Glazer, A. N. (1992) *J. Biol. Chem.* **267**, 7315–7331.

The bilins are derived biosynthetically rather directly[282] from biliverdin IXα, whose formation is described in Fig. 24-24. The addition of an apoprotein –SH group to a carbon–carbon double bond of the bilin is catalyzed by a specific lyase.[283] **Phytochrome** (Section H) arises in a similar way[284] as does the blue biliprotein **insecticyanin** (Box 21-A).[285]

3. The Light-Receiving Complexes

Irradiation of chloroplasts leads to easily measurable fluorescence from chlorophyll *a*, but no fluorescence is observed from chlorophyll *b* or from other forms of chlorophyll, carotenoids, or other pigments. It appears that the latter all serve as light-collecting or antenna pigments that efficiently transfer their energy to chlorophyll *a* at the reaction centers.[286] As is evident from Fig. 23-22, the light-collecting pigments generally have higher energy absorption bands than do the reaction centers. Thus, a broad range of wavelengths of light are absorbed by an organism, and energy from all of them is funneled into the reaction centers. The light-collecting pigments are bound to specific proteins, which are located close to the reaction centers and are arranged to provide efficient energy-transfer. Distances between adjacent pigment molecules vary from 1 to 7 nm.[287]

Phycobilosomes. Algal and cyanobacterial phycocyanins and phycoerythrins are aggregated in special granules that are on the outsides of the photosynthetic membranes. The granules in the cyanobacteria are known as phycobilosomes (Fig. 23-24).[272,286,286a,288] The $(\alpha\beta)_6$ hexamers form the disks of the phycobilosomes. These are held together by linker proteins,[281,289] which fit asymmetrically into the central cavities. As is indicated in Fig. 23-24C, the disks and linker proteins are assembled into rods which are joined to form the phycobilosomes. The

latter are organized into closely packed parallel arrays on the surface of the photosynthetic membranes.

Purple photosynthetic bacteria. The reaction centers of *Rhodobacter spheroides*, *Rhodospirillum rubrum*, and related purple bacteria are embedded in the plasma membrane. Each center is surrounded by a ring of bacteriochlorophyll *a* molecules bound noncovalently to heterodimeric (αβ) protein subunits made up of ~52- to 54-residue chains. Each αβ dimer binds two molecules of BChl *a*, whose central Mg^{2+} ions are coordinated by conserved histidine imidazole groups, as well as a molecule of spirilloxanthin. About 15–17 αβ subunits form the ring, which is designated LH1 (Fig. 23-25A).[291–293] Most of these bacteria also have smaller rings, designated LH2, floating in the membrane near the LH1 complex. The LH2 rings (Figs. 23-25B, C) consist of about nine αβ subunits with associated BChl *a* and carotenoid.[294–298] Under some conditions a third complex LH3 may be formed.[298a] In *Rhodopseudomonas acidophila* nine of the 27 BChl *a* molecules absorb light maximally at ~800 nm and are designated **B800**. The other 18, designated **B850**, absorb maximally at ~860 nm.[294,299] The B850 BChl *a* molecules have direct contact with the chromophores of neighboring molecules, allowing for easy energy transfer. The B800 chromophores are more isolated.

Low-temperature (1.2 K) single-molecule spectroscopic techniques have been used to obtain the fluorescence-excitation spectra shown in Fig. 23-26C. For an ensemble of LH2 complexes (upper trace) the spectral absorption bands are broad, but for individual LH2 complexes structure can be seen clearly for the B800 chromophores but not for the B850 chromophores. This difference has been interpreted to mean that the excitation energy of an electronically excited B850 molecule is delocalized over the whole ring of 18 BChl *a* molecules as an **exciton**. This permits both fast and efficient energy transfers from B800 to B850 and from B850 of one LH2 ring to another or to an LH1 ring and to the reaction center (Fig. 23-27).[299–300a] Energy transfer may occur by the Förster dipole–dipole mechanism (Section C,2).[298] Many of the antenna are supported by binding through their Mg^{2+} ion to an imidazole group of a protein as can also be seen in reaction center chlorophylls (Fig. 23-31C). Hydrogen-bonding to the C13-oxo groups of the chlorophylls may also be possible.[300b] The orientations of the transition dipole moments of the chlorophyll molecules may be arranged to facilitate rapid energy transfer.[300c]

Green sulfur and nonsulfur bacteria. In these organisms chlorophylls are present in rodlike particles

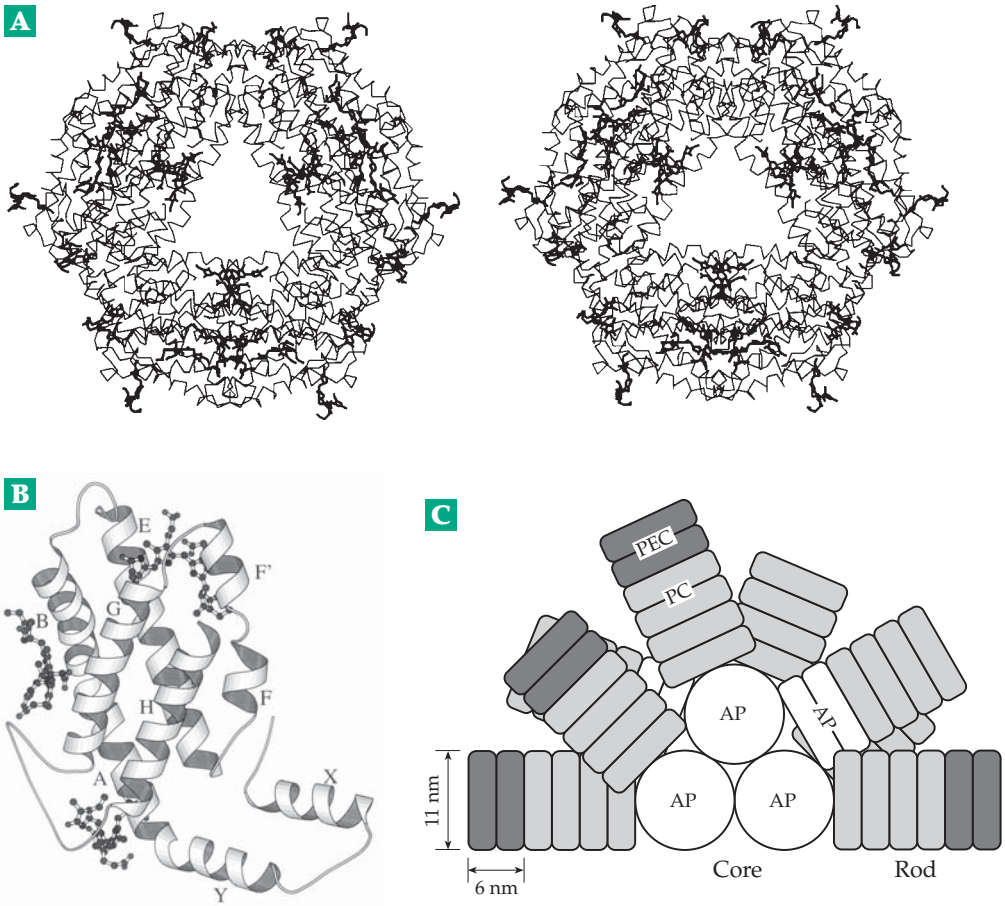


Figure 23-24 (A) Stereoscopic view of a hexameric $(\alpha\beta)_3$ phycobiliprotein. (B) The β subunit of the complex with two molecules of bound phycoerythrobilin and one of phycourobilin. From Chang *et al.*[279] (C) Schematic representation of a phycobilosome of a strain of the cyanobacterium *Anabaena*. Each disk in the structure contains an $(\alpha\beta)_3$ phycobiliprotein. The circles marked AP are cross-sections of rods, each one composed of about four disks of allophycocyanin (AP). The projecting rods contain C-phycocyanin (PC) and phycoerythrocyanin (PEC). From Lao and Glazer.[290]

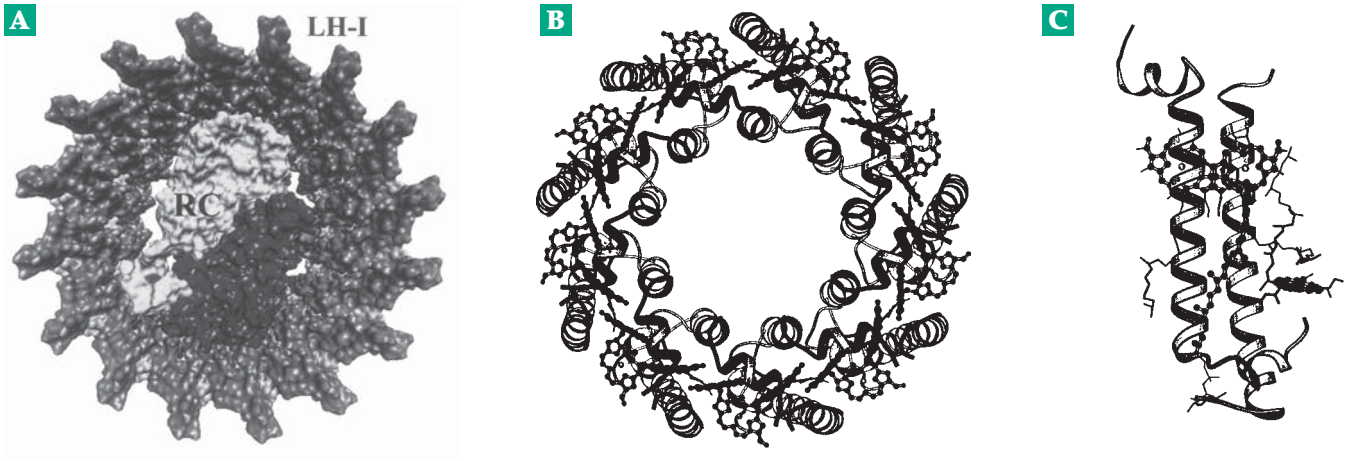


Figure 23-25 (A) The van der Waals contact surface of the periplasmic face of the reaction center and surrounding light-harvesting complex LH1 of *Rhodobacter sphaeroides*. Made with VMD by Theoretical Biophysics Group, UIUC. See also Hu and Schulten.[291] (B) Ribbon drawing of the structure of the circular light-harvesting complex LH2 of the purple photosynthetic bacterium *Rhodopseudomonas acidophila*. The tetrapyrrole rings of the 18 bacteriochlorophyll molecules are also shown. (C) Structure of one of the nine αβ protomers with three associated bacteriochlorophylls. One of these is near the top of the protein, and the other has its chromophoric group protruding at nearly a right angle on the right side. The complete phytyl side chains are also depicted in a stick representation. (B) and (C) are from Prince *et al.*[295] MolScript drawings courtesy of N. W. Isaacs.

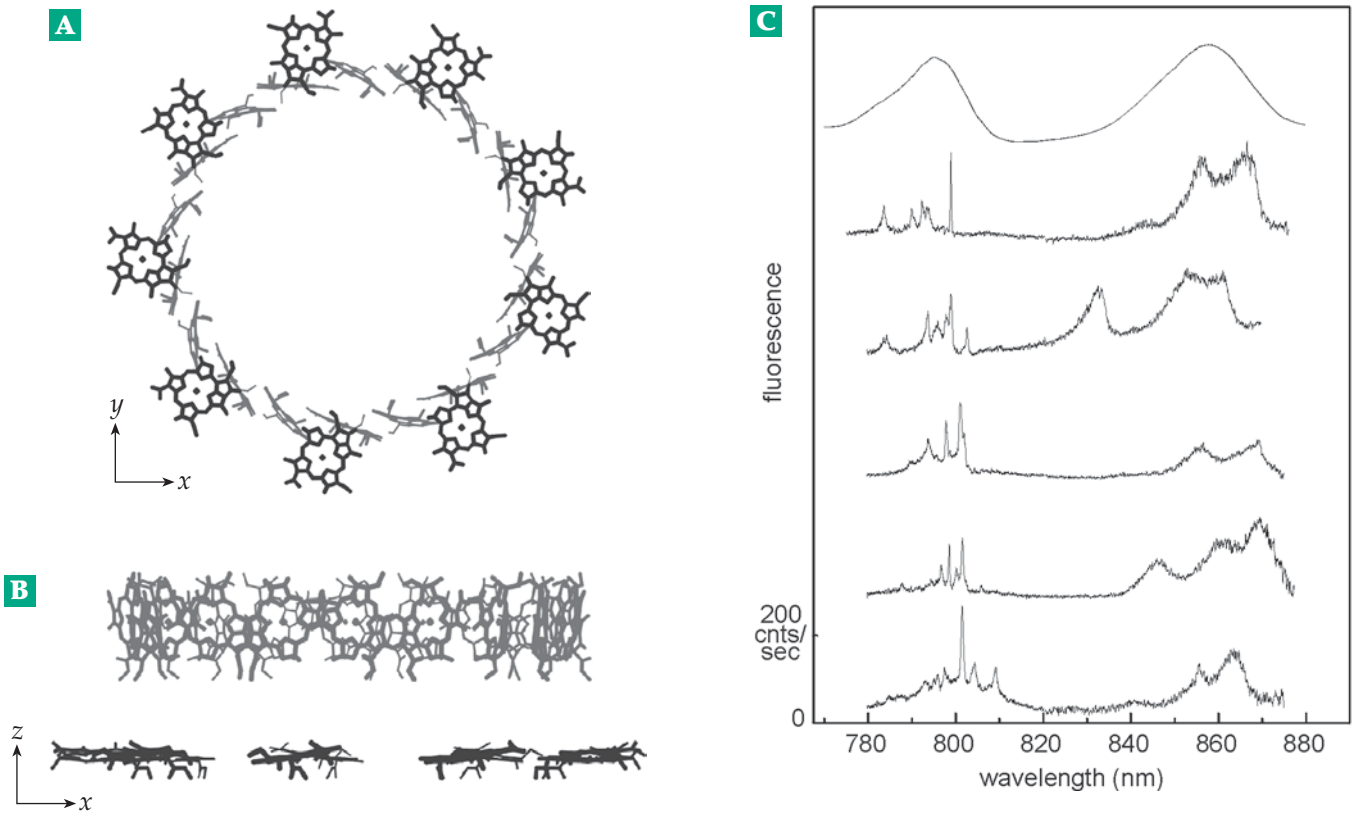


Figure 23-26 (A), (B) Arrangement of bacteriochlorophyll chromophores in the cyclic LH2 array of *Rhodopseudomonas acidophila*. The B850 subunits are gray while the B800 subunits are black. (C) Fluorescence-excitation spectra. Top trace, for an ensemble of LH2 complexes, other traces, for several individual LH2 complexes at 1.2K. Fine structure is evident for the B800 but not for the B850 chromophores. From van Oijen *et al.*[299] with permission.

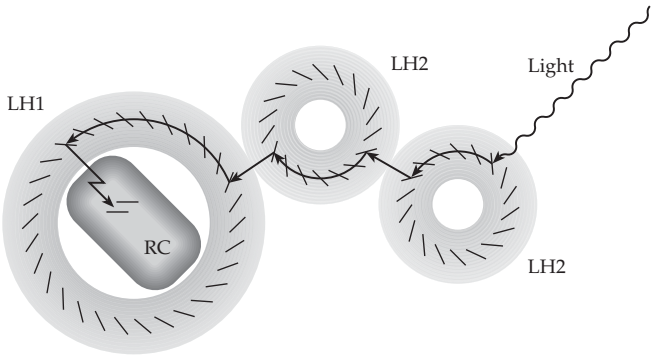


Figure 23-27 Illustration of proposed exciton transfer of the energy of light absorbed by bacteriochlorophyll *a* of purple bacteria. Energy absorbed by the light harvesting complex LH2 is transferred in steps to another LH2, to LH1 and to the reaction center. The short lines within the circles represent the edges of the BChl*a* chromophores. After Kühlbrandt[300] with permission.

of protein present within **chlorosomes**, baglike structures which may be as large as 100 x 260 nm and are attached to the inside of the cytoplasmic membranes,[301–302a] which contain the reaction centers (Fig. 23-28). The over 10,000 light-collecting chlorophylls *c*, *d*, or *e*, which may be present per reaction center, allow these bacteria to live in extremely weak light. The bacteria also contain a smaller "antenna" of ~5000 BChl *a* present as a complex with a water-soluble protein. The three-dimensional structure of this protein is also shown in Fig. 23-28. Each 45-kDa subunit of the trimeric protein contains seven embedded molecules of bacteriochlorophyll *a*.[302,303] Other light-collecting chlorophyll–protein complexes may contain an even higher ratio of chlorophyll to protein.[286]

Eukaryotic plants and cyanobacteria. Photosynthetic dinoflagellates, which make up much of the marine plankton, use both carotenoids and chlorophyll in light-harvesting complexes. The carotenoid **peridinin** (Fig. 23-29), which absorbs blue-green in the 470- to 550-nm range, predominates. The LH complex of *Amphidinium carterae* consists of a 30.2-kDA protein that forms a cavity into which eight molecules of peridinin but only two of chlorophyll a (Chl *a*) and two molecules of a galactolipid are bound (Fig. 23-29).[268]

The allenic carotenoid **fucoxanthin** (Fig. 22-5), which is absent in higher plants, predominates in brown algae, where it occurs in light-harvesting complexes along with Chl *a* and Chl *c*.[306,307]

A family of Chl *a*/*b* binding proteins are found in green plants.[308] These have apparently evolved inde-

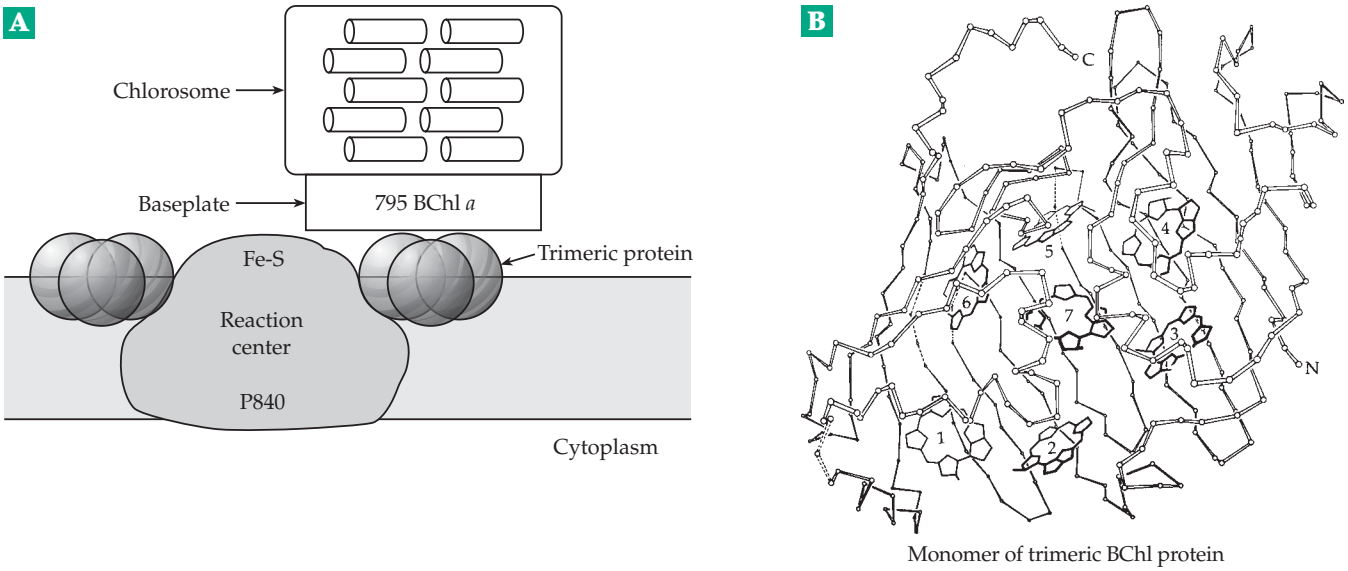


Figure 23-28 (A) Model of a light-harvesting chlorosome from green photosynthetic sulfur bacteria such as *Chlorobium tepidum* and species of *Prosthecochloris*. The chlorosome is attached to the cytoplasmic membrane via a baseplate, which contains the additional antenna bacteriochlorophylls (795 BChl *a*) and is adjacent to the trimeric BChl protein shown in (B) and near the reaction center. After Li *et al.*[302] and Rémigy *et al.*[304] (B) Alpha carbon diagram of the polypeptide backbone and seven bound BChl *a* molecules in one subunit of the trimeric protein from the green photosynthetic bacterium *Prosthecochloris*. For clarity, the magnesium atoms, the chlorophyll ring substituents, and the phytyl chains, except for the first bond, are omitted. The direction of view is from the three-fold axis, which is horizontal, toward the exterior of the molecule. From Fenna and Matthews.[305] See also Li *et al.*[302]

pendently of chlorophyll-binding proteins of green bacteria.[309] Quantitatively most important is the complex known as LHCII, the major Chl *a* / *b* protein associated with PSII and which may also provide energy to PSI. This one protein, whose structure is shown in Fig. 23-30,[310] is thought to bind half of all of the chlorophyll in green plants. The protein is organized as trimers.[311] Each 232-residue monomer binds 5–6 Chl *b*, 7–8 Chl *a*, ~ two molecules of lutein, and one of neoxanthin.[310,311] LHCII also carries all four characteristic thylakoid lipids: mono- and digalactosyl diacylglycerols, phosphatidylglycerol, and sulfoquinovosyl diacylglycerol.[312] The Chl *a* and Chl *b* molecules are in close contact (Fig. 23-30). Subpicosecond transient absorption spectroscopy[311] indicates that half of the Chl *b* to Chl *a* energy transfers occur in < 0.2 ps. Notice the close association of the two luteins in Fig. 23-30 with the chlorophyll rings. The carotenoids are thought to quench chlorophyll triplet states to prevent formation of singlet oxygen.

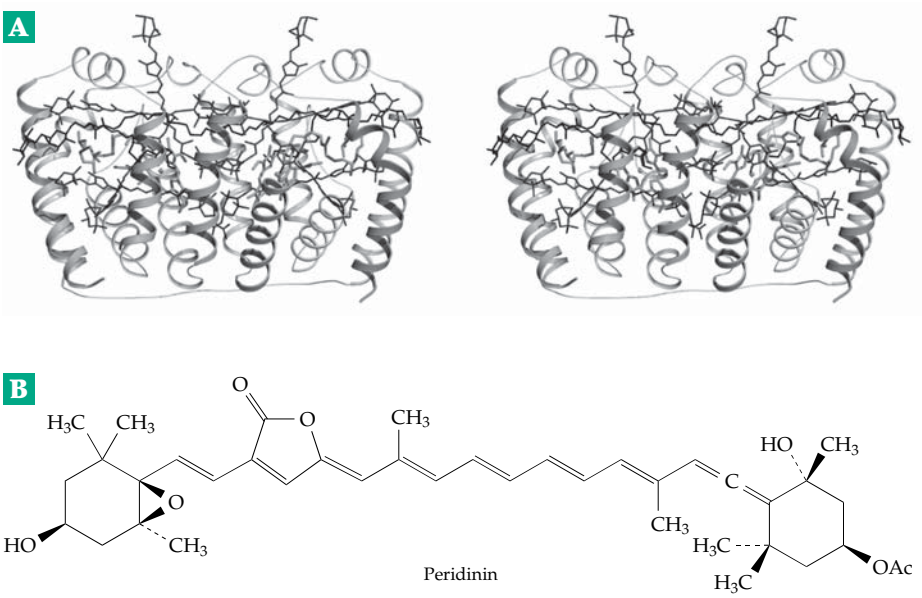


Figure 23-29 (A) Stereoscopic drawing of light-harvesting complex from the dinoflagellate protozoan *Amphidinium carterae*. The central cavity contains eight molecules of peridinin, two of which can be seen protruding from the top. Deeply buried toward the bottom are two molecules of Chl *a*. Also present are two molecules of digalactosyl diacylglycerol. From Hofmann *et al.*[268] Courtesy of Wolfram Welte. (B) Structure of peridinin.

Figure 23-30 Views of light-harvesting protein LHCII of green plants. (A) Side view indicating the approximate position in the lipid bilayer of the thylakoid membrane. Helices are labeled A–D. (B) Stereoscopic top view from the stromal side of the membrane. The structure, at 0.34 nm resolution, was determined by electron crystallography on highly ordered two-dimensional crystals. MolScript drawings from Kühlbrandt *et al.*[310] Courtesy of Werner Kühlbrandt.

PSII also contains several additional chlorophyll-binding proteins, designated CP24, CP26, CP29, CP43, CP47, etc. These lie on both sides of a pair of reaction-center cores.[313–315a] A large fraction of the LHCII complexes are separate from the reaction centers and are mobile, while a smaller fraction are bound to the outer ends of the core complex.[308,314,315] The light-harvesting chlorophylls of PSI are arranged around the core. Some are bound in the N-terminal part of the major core subunits, the products of genes *PsaA* and *PsaB*.[316] They bind ~90 Chl *a* and 14 β-carotene molecules per reaction center. (See Fig. 23-33.) In addition, the peripheral LHCI, composed of four different proteins arranged around the core, binds ~110 Chl *a* and Chl *b* and ~70 molecules of xanthophyll.[317] Some species of cyanobacteria use antenna rings around their PSI trimers instead of phycobilinosomes.[317a]

In every case the light-harvesting complexes are arranged to allow rapid and very efficient transfer of electronic excitation energy from one chromophore to another and finally to the chromophores of the reaction center.[317b] The speed and efficiency appear to depend upon very rigid structures of the proteins and precise orientations of the bound chromophores to allow direct excitonic transfer of energy at distances of less than 2 nm or transfer by the Förster mechanism at distances not exceeding 10 nm.[302] An example of the precision of protein structures was observed when a posttranslationally modified asparagine N^5-methylasparagine at position 72 of the β subunit of many phycobiliniproteins was substituted by aspartate or glutamine. The fluorescence lifetime of the nearby bilin was reduced 7–10% in the mutants, an effect that could cut the >95% efficiency of energy transfer from the phycobilinisomes to the PSII reaction center.[318]

4. The Reaction Centers and Their Photochemistry

The initial or primary processes of photosynthesis occur in the reaction centers in which chlorophyll or bacteriochlorophyll absorbs a photon.[318a] Then, the chlorophyll, in its singlet excited state (Chl*), donates an electron to some acceptor A to form a radical A^- and to leave an oxidized chlorophyll Chl^+ radical (Eq. 23-31).

$$\text{Chl} \xrightarrow{h\nu} \text{Chl}^* \xrightarrow{\text{A}} \text{Chl}^+ + \text{A}^- \qquad (23\text{-}31)$$

In the scheme of Fig. 23-18, acceptor A is Q_A for PSII and A_0 for PSI. The oxidized chlorophyll (Chl^+) quickly reacts further by receiving an electron from some donor.

The photooxidation of chlorophyll indicated in Eq. 23-31 is accompanied by bleaching in the principal light absorption band. However, since there is so much light-gathering chlorophyll for each reaction center, the effect is small. The study of the process has been aided greatly by preparation of isolated bacterial photochemical reaction centers.

Reaction centers of purple bacteria. The exact composition varies, but the properties of reaction centers from several genera of purple bacteria are similar. In *Rhodopseudomonas viridis* there are three peptide chains designated H, M, and L (for heavy, medium and light) with molecular masses of 33, 28, and 24 kDa, respectively. Together with a 38-kDa tetraheme cytochrome (which is absent from isolated reaction centers of other species) they form a 1:1:1:1 complex. This constitutes reaction center P870. The three-dimensional structure of this entire complex has been determined to 0.23-nm resolution[288,319–323] (Fig. 23-31). In addition to the 1182 amino acid residues there are four molecules of bacteriochlorophyll (BChl), two of bacteriopheophytin (BPh), a molecule of menaquinone-9, an atom of nonheme iron, and four molecules of heme in the *c* type cytochrome. In 1984, when the structure was determined by Deisenhofer and Michel, this was the largest and most complex object whose atomic structure had been described. It was also one of the first known structures for a membrane protein. The accomplishment spurred an enormous rush of new photosynthesis research, only a tiny fraction of which can be mentioned here.

The reaction centers are embedded in the cytoplasmic membranes of the bacteria, with the bottom of the structure, as shown in Fig. 23-31, protruding into the cytoplasm and the heme protein at the top projecting out into the periplasm which lies within infoldings of the plasma membrane. Subunits L and M each contain five ~4.0 nm long roughly parallel helices, which span the cytoplasmic membrane. Another membrane-spanning helix is contributed by subunit H, which is located mainly on the cytoplasmic side. An approximate twofold axis of symmetry relates subunits L and M and the molecules of bound chlorophyll and pheophytin.

Spectral measurements suggesting exciton splitting were among early observations that led to the conclusion that the bacteriochlorophyll involved in the initial photochemical process exists as a dimer or **special pair** (Fig. 23-31),[319,324] a conclusion verified by the structure determination. The special pair of BChl *b* lies in the center of the helical bundle that is embedded in the membrane. Nearly perpendicular to the rings of the special pair are two more molecules of BChl *b*. The central magnesium atoms of all four bacteriochlorophylls are held by imidazole groups of histidine side chains.[319,325] Below the chlorophylls are

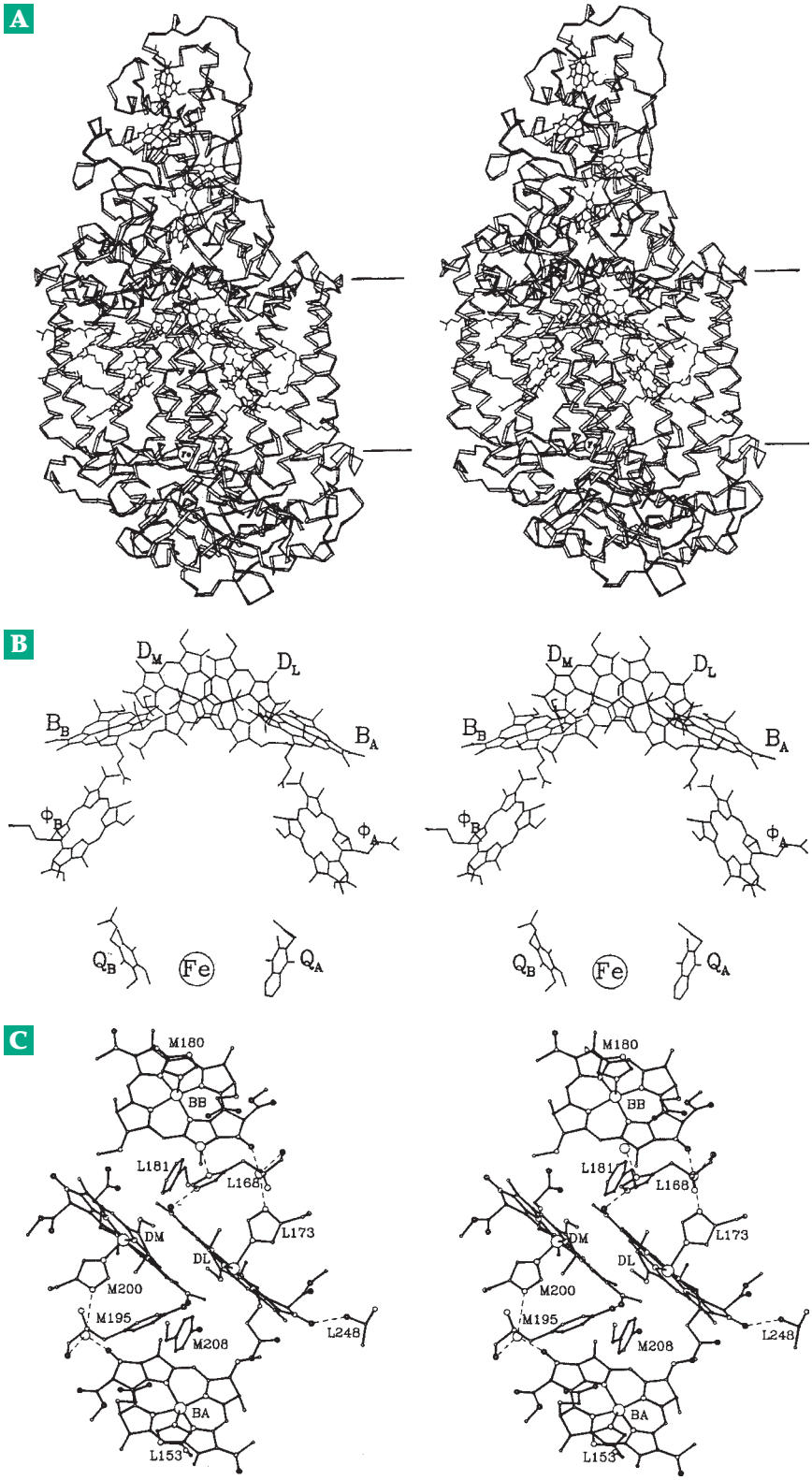


Figure 23-31 (A) Stereoscopic ribbon drawing of the photosynthetic reaction center proteins of *Rhodopseudomonas viridis*. Bound chromophores are drawn as wire models. The H subunit is at the bottom; the L and M subunits are in the center. The upper globule is the cytochrome *c*. The view is toward the flat side of the L, M module with the L subunit toward the observer. (B) Stereo view of only the bound chromophores. The four heme groups He1–He4, the bacteriochlorophylls (Bchl) and bacteriopheophytins (BPh), the quinones Q_A and Q_B, and iron (Fe) are shown. The four hemes of the cytochrome are not shown in (B). From Deisenhofer and Michel.[320] (C) Stereoscopic view of the Bchl *b* molecules along the local twofold axis. The special pair (D_M, D_L) is in the center with its tetrapyrrole rings almost perpendicular to the plane of the paper; the monomeric chlorophylls are labeled B_B and B_A. The four histidine ligands to the magnesium ions of the bacteriochlorophylls as well as two tyrosines (M195 and M208) and three water molecules (large circles) are also shown. From Deisenhofer *et al.*[321] with permission.

the two molecules of bacteriopheophytin and below them the nonheme iron and the menaquinone, the first quinone acceptor Q_A. It corresponds to Q_A of PSII shown in Fig. 23-18.

Isolated reaction centers usually contain or will bind a second quinone, which may be ubiquinone-10 (Q_{10})[325a] and which is usually designated Q_B. Its binding site is to the left of the nonheme iron in Fig. 23-31 in a position symmetrically related to that of Q_A. The reaction centers also contain a carotenoid 1,2-dihydro-

neurosporene.[323] The reaction centers of purple bacterium, *Rhodospirillum rubrum*, each contain one molecule of spirilloxanthin; a variety of carotenoids are present in other species.[326] The reaction centers of a third purple bacterium, *Rhodobacter spheroides*, are closely similar in structure to that in Fig. 23-31 but lack the tetraheme.[327] Reaction centers of these bacteria accept electrons directly from a soluble cytochrome *c*.

Reaction center kinetics. After an 0.8-ps or shorter flash of light the decay of the singlet excited state of the bacteriochlorophyll dimer in isolated reaction centers can be followed by loss of its characteristic fluorescence.[328,329] The lifetime of this excited state in *R. sphaeroides* is only 4 ps indicating a rapid occurrence of the initial electron transfer of Eq. 23-31. A rise in absorbance at 1250 nm is interpreted as formation of the bacteriochlorophyll cation radical $BChl^+$ in the special pair. Other spectral changes support the formation of BPh^- as the first reduction product (A^- in Eq. 23-31). However, this is thought to occur in two steps[323] with the monomeric BChl (B_A in Fig. 23-31B) receiving the electron in ~2.8 ps and passing it to the pheophytin (Phe; ϕ_A in Fig. 23-31B) in ~0.7 ps (Eq. 23-32; corresponding steps for PSII of green plants are also shown).

The quinone Q_A (the secondary acceptor) is next reduced by the BPh^- radical in ~200 ps with development of a characteristic EPR signal[321,330] at $g = 1.82$. Over a much longer period of time (~320 ns) an electron passes from the tetraheme cytochrome subunit to the Chl^+ radical in the special pair.[323,323a] The relatively slow rate of this reaction may be related to the fact that the bacteriochlorophyll of the special pair is 2.1 nm (center-to-center) from the nearest heme in the cytochrome while BPh and Q_A are only 1.4 nm apart (see Fig. 23-31).[288,331] Over a period of ~24 μs after formation of the radical anion Q_A^- an electron from the Q_A^- radical is passed to Q_B, a weakly bound ubiquinone-9, to form the Q_B^- radical. Upon absorption of a second photon by the special BChl pair another electron is passed through the chain to form $Q_A^-Q_B^-$. Uptake of two protons with transfer of the second electron from Q_A^- to Q_B^- yields the ubiquinol QH_2, which dissociates from its binding site (Q_B) into the ubiquinone pool dissolved in the lipids of the membrane bilayer.[331a–g]

Why is this multistep sequence of electron transfers necessary? A variety of techniques such as femtosecond IR[332,333] and electronic[334–335a] spectroscopy, resonance Raman spectroscopy at low temperatures,[336] and study of many mutants[337–338c] have been directed toward an answer to this question. It has been generally accepted that light energy absorbed by any one of the "monomeric" Chl or pheophytins in the reaction centers is funneled "downhill" to the special pair within 0.1–0.2 ps to generate P*.[334,336] The ultrafast ~3 ps electron transfer from P* to the adjacent monomeric BChl or Chl is necessary to prevent loss of energy by fluorescence from P*. The subsequent energetically downhill transfer to a pheophytin and on to Q_A prevents reverse electron transfer, which could also lead to fluorescence. Both the efficiency and the quantum yield are very high.[339]

The rate of the ultrafast proton transfer becomes even higher at cryogenic temperature, suggesting quantum mechanical tunneling.[331,335,340–340b] The transfer is generally treated using Marcus theory (Chapter 16), which indicates a very small reorganization energy for the process. Another aspect of the process is a possible coupling of a vibrational mode of the protein matrix to the electron transfer. Femtosecond near-IR spectra show low-frequency vibrational modes of the excited-state reaction center chromophores which may facilitate electron transfer.[332,336] The transfer of an electron from the tetraheme cytochrome *c* of *R. viridis*[341] or from the small cytochrome c_2 of *Rhodobacter spheroides*[342] to Chl^+ of the special pair has similar characteristics but is slower than the initial electron transfer from P*. On the other hand, electron transfers from Q_A^- to Q_B^- involve two distinct steps and coupled uptake of two protons.[343] An unexplained fact is that photochemical electron transport through the reaction centers always occurs through the L-side to Q_A rather than the M-side.[343a–344] However, rapid electron transfer to the pheophytin on the B-side (M-side) has been observed following excitation with blue light. This may

PSII (Green plants)	*Rhodopseudomonas viridis*
H_2O	
↓	
$(Mn)_4$ Manganese cluster	Cytochrome *c*
↓	↓
Y_Z (Tyr)	Fe (Tetraheme cytochrome *c*)
↓	↓
Chl_2 (P680)	$BChl_2$ Special pair cation radical (P870)
↓	↓
Chl_2*	$BChl_2$* $h\nu$ Electronic excitation of $BChl_2$
~10 ps ↓	2.8 ns ↓
$Chl_A(A_0)$	$BChl_A$
35-200 ps ↓	0.7 ns ↓
Phe (A_1)	$BPhe_A$
↓	200 ps ↓
Q_A (Plastoquinone)	Q_A (Menaquinone)
↓	24 μs ↓ 2 e^- in two steps
Q_B (Plastoquinone)	Q_B (Ubiquinone)

(23-32)

TABLE 23-4
Properties of Various Reaction Centers

	Rhodopseudomonas viridis (now *Blasatochloris*)	Green plants and Cyanobacteria		Green sulfur bacteria, *heliobacteria*[a]
		PSII	PSI	
Subunits	L / M, H	D1 / D2, CP43, CP47	PsaA / PsaB nine others	$(PscA)_2$
Masses, kDa	24 / 28, 33	38 / 39.4	83 / 83	65 / 65
Input	Cyt *c* (4 Fe)	H_2O, $(Mn)_4$, Y_Z, Y_D	Cyt *f*, plastocyanin or cyt *c*	Cyt c_1, c_2
Special pair	P870	P680 $(BChl\ b)_2$	P700 $(Chl\ a)_2$	P840 $(Chl\ a)_2$
Monomeric chlorophyll	$BChl_L$, $BChl_M$		A_0 Both Chl *a*	A_0 Both Chl *a*-like
Pheophytin	$BPhe_L$, $BPhe_M$		A_0 (Phe, Phe)	
Quinone	Q_B, Q_A (Ubiquinone, menaquinone-9)	Q_B, Q_A (Plastoquinone)	Q_B, Q_A (A_1) (Both phylloquinone in cyanobacteria, plastoquinone in chloroplasts)	Q_B, Q_A (Both menaquinone-7)[b]
Iron	Fe^{3+}		F_X (Fe_4S_4)	F_X
Output	Ubiquinone Cyt bc_1, Cyt c_2	Ubiquinone Cyt b_6f	F_A, F_B, Fd $NADP^+$	F_A, F_B, Fd NAD^+

[a] Nitschke, W., and Rutherford, A. W. (1991) *Trends Biochem. Sci.* **16**, 241–245
[b] Kjear, B., Frigaard, N.-U., Yang, F., Zybailov, B., Miller, M., Golbeck, J. H., and Scheller, H. V. (1998) *Biochemistry* **37**, 3237–3242

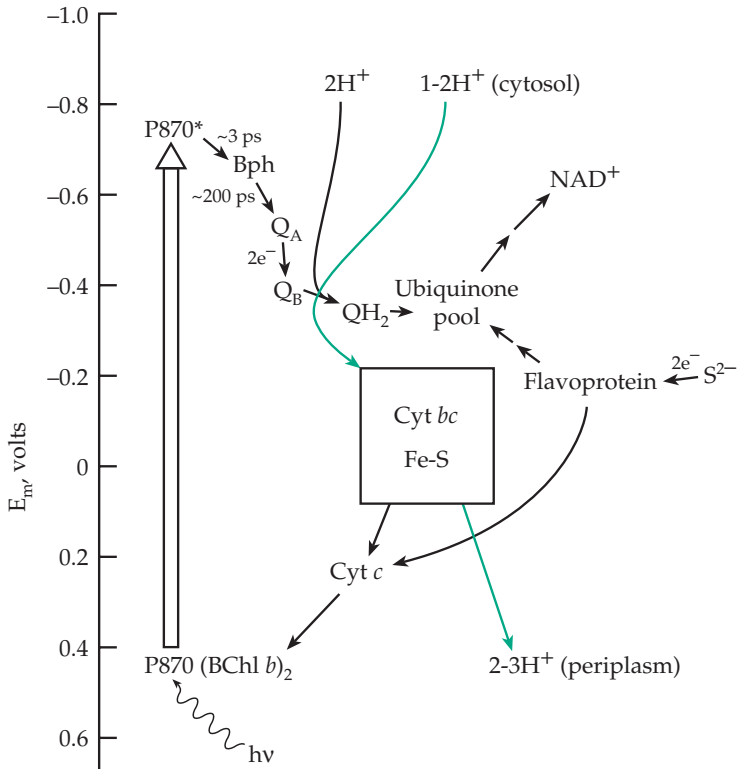


Figure 23-32 Simplified diagram of cyclic electron flow in purple bacteria. Two protons from the cytoplasm bind to Q_B^{2-} in the reaction center to form QH_2 (ubiquinol), which diffuses into the ubiquinone pool. From there it is dehydrogenated by the cytochrome bc_1 complex with expulsion of two protons into the periplasm. A third and possibly a fourth proton may be pumped (green arrows) across the membrane, e.g., via the Q cycle (Fig. 18-9). The protons are returned to the cytoplasm through ATP synthase with formation of ATP. Some electrons may flow to the reaction centers from such reduced substrates as S^{2-} and some electrons may be removed to generate NADPH using reverse electron transport.[345]

represent a photoprotective mechanism.[344a] In the PSI system, in which the two phylloquinones are tightly bound,[344b] both the A-side and the B-side seem to function in electron transfer.[344c]

Cyclic photophosphorylation in purple bacteria. QH_2 is eventually dehydrogenated in the cytochrome bc_1 complex, and the electrons can be returned to the reaction center by the small soluble cytochrome c_2, where it reduces the bound tetraheme cytochrome or reacts directly with the special pair in *Rhodobacter spheroides*. The overall reaction provides for a cyclic photophosphorylation (Fig. 23-32) that pumps 3–4 H^+ across the membrane into the periplasmic space utilizing the energy of the two photoexcited electrons. These protons can pass back into the cytoplasm via ATP synthases located in the same membrane with their catalytic centers in the cytosol (see Figs. 18-5 and 18-14).

Comparison with other reaction centers. Subunit L of the *R. viridis* reaction center was found to have a 25% sequence homology with a quinone-binding protein now known as D1, a component of the reaction center core of PSII of chloroplasts. This protein was identified as the specific target protein for inhibition by herbicides such as DCMU and atrazine (see Section 1 for structures). These compounds act as competitive inhibitors of quinone binding[346] and bind in the Q_B site in the *R. viridis* reaction center. This fact, together with the discovery that the core of PSII of green plants consists of a heterodimer of the related polypeptides D1 and D2, suggested that PSII is very similar to the bacterial center of Fig. 23-31. Both use a quinone as primary acceptor. However, P680 operates at a more positive potential ($E_m \sim 1$ V), consistent with the fact that it must provide an oxidizing agent able to oxidize H_2O to O_2 ($E°' = +0.82$ V). Plastoquinone rather than menaquinone is the primary acceptor in PSII. A chlorophyll *a* dimer is apparently the initial electron donor. The methyl ester carbonyl groups on the edges of rings I (Fig. 23-20) of BChl may coordinate to other groups of the proteins.[288,346a] These ester groups are absent in Chl *a*.

The PSI reaction center (P700) of maize chloroplast was also found to contain a pair of homologous polypeptides with appropriately placed histidyl residues for chlorophyll binding.[347] FTIR spectra also indicate the presence of two Chl *a* molecules. Small shifts in IR frequencies upon electronic excitation suggest that one chlorophyll (P_1) is hydrogen-bonded through its 109-ester and 9-oxo groups (Fig. 23-20) while the chlorophyll (P_2) is free. This may account for the low value of E_m (Fig. 23-17). The charge on $P700^+$ appears to be carried in part on both chlorophylls. However, in the triplet state $^3P700^*$, which may be observed at low temperature, excitation is localized on P_1.[348] Although P700 operates at the low E_m of +0.49 V it produces a powerful reducing agent able to reduce ferredoxin. The first identified acceptors are other Fe–S centers present in integral membrane proteins.[345] The reaction centers of green bacteria[349] and PSI of cyanobacteria[245] have similar characteristics. As more genes have been sequenced and X-ray diffraction, electron crystallography, and electron microscopy with single-particle averaging have advanced, the fundamental similarity of all of the photosynthetic centers has been confirmed.[228,300,350–353] Gene sequences are often not highly conserved, but structures are more conserved. An evolutionary relationship of all of the reaction centers can be seen.[316,354]

PSI of cyanobacteria and green plants. The major reaction-center subunits **PsaA** and **PsaB** each have a C-terminal domain, resembling those of the L and M chains of purple bacteria, and a large N-terminal antenna-chlorophyll-binding domain. Cyanobacterial PSI contains ten other subunits, PsaC to PsaF, PsaI to PsaM, and PsaX. Thirty-one transmembrane helices have been assigned to the various subunits, several of which are in positions corresponding closely to those in the reaction centers of the purple bacteria (Fig. 23-33).[355–357] The PSI of higher plants is somewhat larger than that of cyanobacteria and contains somewhat different subunits.[356a,357,357a]

The electron donor to Chl^+ in PSI of chloroplasts is the copper protein plastocyanin (Fig. 2-16). However, in some algae either plastocyanin or a cytochrome *c* can serve, depending upon the availability of copper or iron.[345] Both Q_A and Q_B of PSI are phylloquinone in cyanobacteria but are plastoquinone-9 in chloroplasts. Mutant cyanobacteria, in which the pathway of phylloquinone synthesis is blocked, incorporate plastoquinone-9 into the A-site.[345a] Plastoquinone has the structure shown in Fig. 15-24 with nine isoprenoid units in the side chain. Spinach chloroplasts also contain at least six other plastoquinones. Plastoquinones C, which are hydroxylated in side-chain positions, are widely distributed. In plastoquinones B these hydroxyl groups are acylated. Many other modifications exist including variations in the number of isoprene units in the side chains.[358,359] There are about five molecules of plastoquinone for each reaction center, and plastoquinones may serve as a kind of electron buffer between the two photosynthetic systems.

Look at the Z scheme of Fig. 23-17. The lower end of each vertical arrow is located at an electrode midpoint potential E_m (or $E°'$) for the couple P^+/P, i.e., for a one-electron reduction of the Chl^+ or $BChl^+$ radical.[360] The top of the arrow is at the estimated value of E_m for the excited state P*. It is more negative than the ground-state value of E_m by the energy (in electron volts) of the light absorbed. This is a little misleading

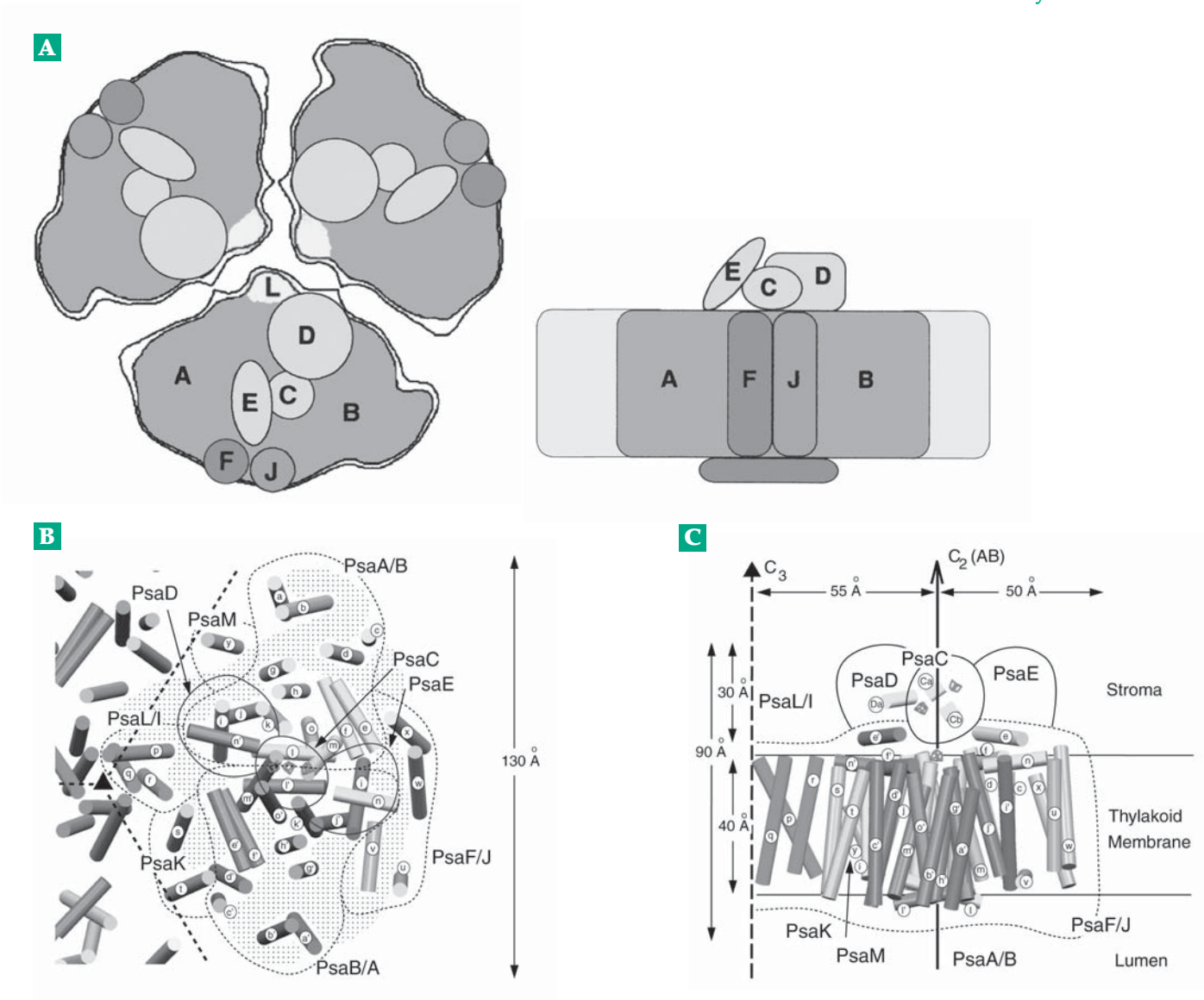
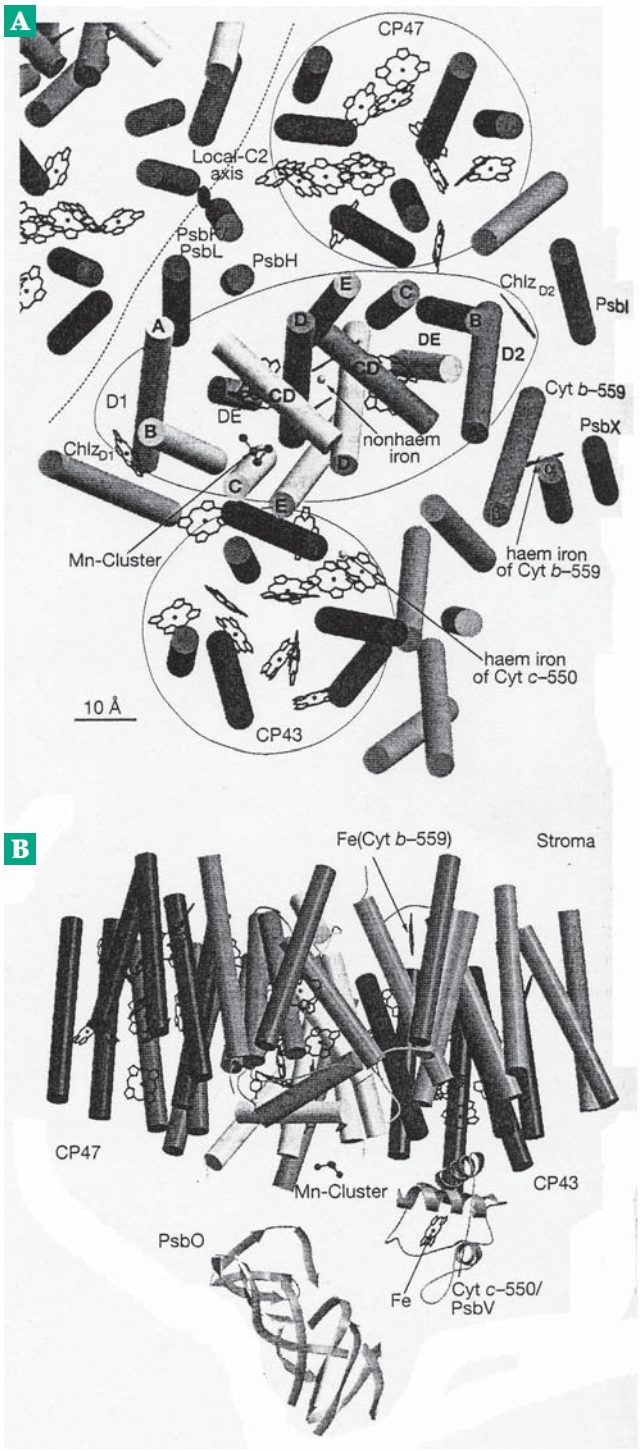


Figure 23-33 Models of the PSI reaction center of the cyanobacterium *Synechocystis* and of higher green plants. (A) Subunit arrangement in trimeric cyanobacterial PSI as gained from EM and crosslinking experiments with wild type and directed mutant forms. From Kruip *et al.*[355] Courtesy of Jochen Kruip: Left, top view; Right, side view. (B) Schematic representation (top view) of the structure of one monomer based on X-ray data at 0.6 nm resolution. The triangle at the left marks the threefold axis of a trimer and the heavy dashed lines the interfaces between monomers. Thirty-one helices, many of which are transmembrane, are shown as shaded cylinders. Fourteen of these are in the central core (light area) while the other 17 are in peripheral regions (light gray) that bind antenna chlorophyll molecules. (C) Side view, stroma at the top showing stromal subunits PsaC, PsaD, and PsaE. (D) Top view of the trimer showing locations of all α helices and core antenna chlorophylls as well as reaction center cofactors. From Schubert *et al.*[356] Courtesy of Nobert Krauss.

for it is not commonly appreciated that light carries entropy as well as energy. An important consequence of this fact is that not all of the energy of sunlight could be harnessed for chemical work. Knox[361] has calculated that at 700 nm at most 78% of the energy could be captured. See Parson[362] for further discussion. Nevertheless, the photoexcited P700* with E_m = – 1.26 V is able to reduce a series of membrane-bound Fe–S centers of $E°'$ ~ –0.5 to –0.6 V. There are three of these designated F_X, F_A, and F_B. Center F_X is an Fe_4S_4 cluster located at almost the same position as the single Fe^{3+} of the *R. viridis* reaction center (Fig. 23-31B). F_A and F_B are also Fe_4S_4 clusters both of which are carried on the small 79-residue PsaC.[363,364] This protein binds to the reaction center on the stromal side as shown in Fig. 23-33. Close to it are two other subunits, PsaD and PsaE, which appear to assist the docking of ferredoxin or flavodoxin to PsaC and cluster F_X.[365,366]



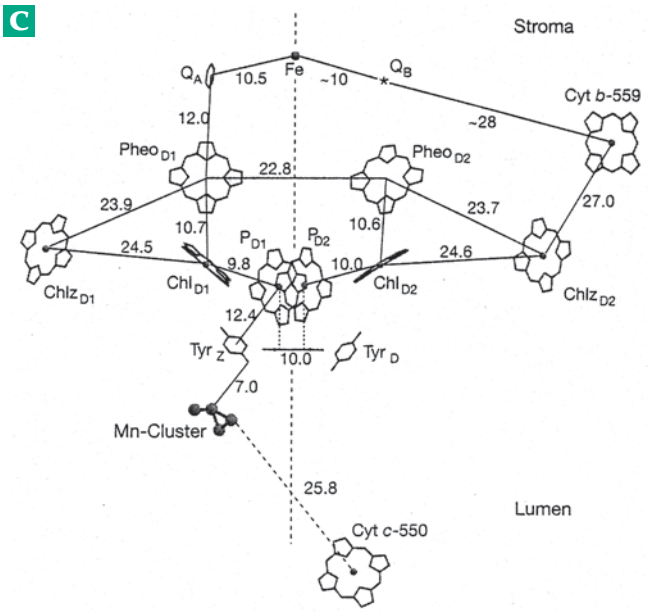


Figure 23-34 Structure of PSII with assignment of protein subunits and cofactors. (A) Arrangement of transmembrane α-helices and cofactors in PSII. One monomer of the dimer is shown completely, with part of the second monomer related by the local-C2 axis (filled ellipse on the dotted interface). Chlorophyll *a* head groups and hemes are indicated by black wire drawings. The view direction is from the luminal side, perpendicular to the membrane plane. The α-helices of D1, D2, and Cyt *b*-559 are labeled. D1/D2 are highlighted by an ellipse and antennae, and CP43 and CP47 by circles. Seven unassigned α-helices are shown in gray. The four prominent landmarks (three irons and the manganese (Mn) cluster) are indicated by arrows. (B) Side view of PSII monomer looking down the long axis of the D1/D2 subunits from the right side in (A), at slightly tilted membrane plane and rotated 180° so that the luminal side is bottom. PsbO (33K protein) is shown as a β-sheet structure, and Cyt *c*-550 as a helical model. (C) Arrangement of cofactors of the electron transfer chain located in subunits D1 and D2. View direction along the membrane plane. Full lines indicate center-to-center distances (nm) between the cofactors (uncertain to about ± 0.1 nm). The pseudo-C2 axis is shown by the vertical dotted line; it runs throught the nonheme iron Fe and is parallel to the local-C2 axis. The asterisk indicates the putative Q_8 binding site. From Zouni *et al.*[371d] Courtesy of Athina Zouni.

The photosynthetic centers of green photosynthetic sulfur bacteria also have centers F_X, F_A, and F_B.

The soluble electron carriers released from the reaction centers into the cytoplasm of bacteria or into the stroma of chloroplasts are reduced single-electron carriers. Bacterial ferredoxin with two Fe_4S_4 clusters is formed by bacteria if enough iron is present. In its absence flavodoxin (Chapter 15), which may carry either one or two electrons, is used. In chloroplasts the carrier is the soluble **chloroplast ferredoxin** (Fig. 16-16,C), which contains one Fe_2S_2 center. Reduced ferredoxin transfers electrons to $NADP^+$ (Eq. 15-28) via **ferredoxin:NADP⁺ oxidoreductase**, a flavoprotein of known three-dimensional structure.[367–369]

PSII and formation of oxygen. The structure of PSII has been difficult to determine directly, but its core has been modeled in atomic detail using bacterial reaction centers as a guide.[370,371] More recently electron crystallography provided a three-dimensional image at 0.8 nm resolution.[314,371a] The resolution has been extended down to 4 nm by X-ray crystallography.[371b–e] The structure of the cyanobacterial PSII (Fig. 23-34) is very similar to that of green plants.[371e,f] PSII contains at least 17 protein subunits, all of which are encoded by chloroplast genes. The large structurally similar D1 and D2 form the core. They are encoded by genes *PsbA* and *PsbD* and have molecular masses of 38.0 and 39.4 kDa, respectively.[315] Both Q_A and Q_B are plastoquinone. It is of historical interest that these cofactors were first designated simply as Q, not for quinone but for *quencher*. This is because Q_A apparently quenched the fluorescence of P680, the reaction-center chlorophyll *a*. If chloroplasts are irradiated with 650-nm light, PSII is activated but PSI is not. Under these conditions Q_A becomes reduced, and the fluorescence of Chl^+ increases, presumably because the electron acceptor Q_A is absent. If PSI is activated by addition of far-red light, Q_A remains more oxidized, and fluorescence is quenched by a mechanism that appears to depend upon one of the additional bound chlorophylls, as well as the chlorophyll pair Chl_2 and the 9-kDa cytochrome b_{559}. The latter is an essential PSII subunit (*PsbE* gene),[372,373] which forms a tight complex with the D1/D2 pair.

Other subunits in PII include the 56- and 50-kDa antenna proteins CP43 (*PsbC* gene) and CP47 (*PsbB* gene). Three **extrinsic proteins**, which bind to the lumenal side of the thylakoid membrane, are the 33-kDa **manganese stabilizing protein** (*PsbO* gene),[374,375] cytochrome c_{550}, and a 12-kDa subunit (Fig. 23-34). The position of the larger two of these proteins are shown in the model in Fig. 23-34, which is based on 0.8-nm resolution data.[315] These extrinsic proteins seem to function together to facilitate binding not only of Mn ions but also of Ca^{2+} and Cl^-, both of which are essential for O_2 evolution.[376] Other smaller subunits are also present.[376a] Together with its antenna complexes PSII may form large supercomplexes with as many as 25 subunits.[376b] The distribution of complexes varies in the different regions of the thylakoid, e.g., the stroma lamellae and grana stacks (Fig. 23-19).[376c]

The four-electron dehydrogenation of two water

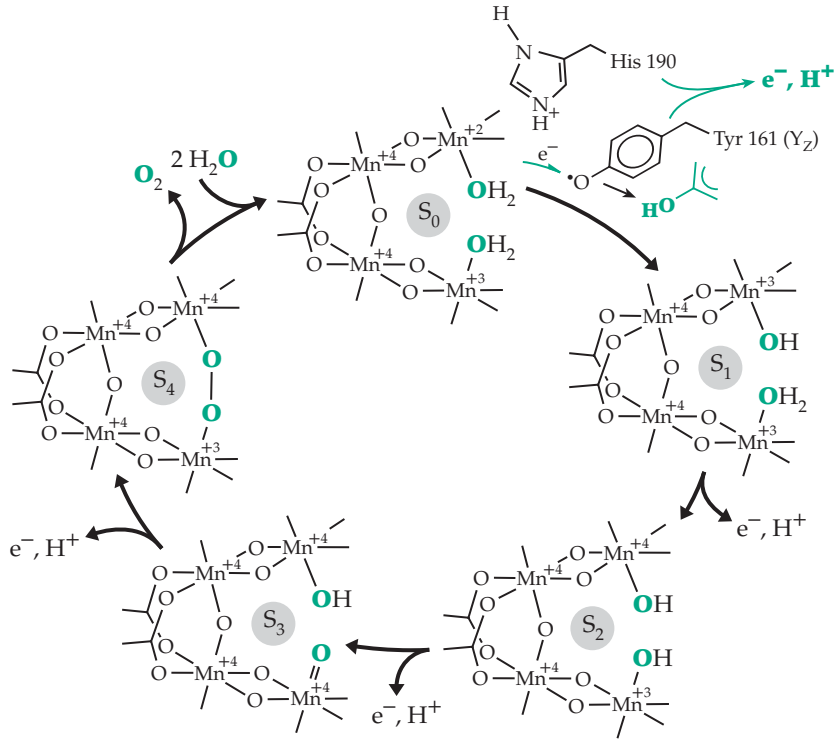


Figure 23-35 Proposed sequence of S-states of the manganese cluster of photosystem II. The successive states as two molecules of H_2O (green oxygen atoms) are converted to O_2 is shown with the successive states S_0–S_4 labeled. To save space and possible confusion tyrosine 161 (Y_Z) and the nearby His 190 are shown only by S_1. The Y_Z radical is thought to remove a hydrogen atom or H^+ from one bound H_2O and an electron from one Mn ion at each of the four S-states S_0–S_3 functioning in each case to eject a proton into the thylakoid lumen and to transfer an electron to P^+ of the reaction center. However, the exact sequence of e^- transfer and H^+ release may not be shown correctly. After Hoganson and Babcock.[392,392a]

molecules to give one molecule of O_2 by PSII is still not well understood. From experiments on oxygen evolution in the presence of repeated short flashes of light it was found that a four-quantum process is required.[376–380] There must be some way of storing oxidizing equivalents until enough are present to snap together an oxygen molecule. There is abundant evidence that manganese is required for this process and that the oxidation of H_2O occurs on a cluster of four atoms of manganese.[379]

The 33-kDa protein PsbO, which is present in all oxygen-forming photosynthetic organisms, is closely associated with the Mn_4 cluster. Removal of this protein leads to a gradual loss of two of the four Mn ions.[375] The structure of the Mn_4 cluster is not yet certain, but on the basis of EXAFS spectroscopy a pair of di-μ-oxo bridged Mn dimers linked by carboxylates and with a fifth μ-oxo bridge as in Fig. 23-35 has been proposed. The distance between Mn ions in the di-μ-oxo-Mn_2 groups is ~0.27 nm, and the planes of these groups are roughly parallel to the surface of the thylakoid membranes.[381,382] The protein groups that bind the Mn atoms include carboxylates, as shown in Fig. 23-35, and probably one or more histidine imidazole groups,[383] perhaps of His 332 and His 337 of the D1 chain.

The immediate donor of an electron to the reaction-center cation P^+ ($ChlZ^+$) of PSII was identified by EPR spectroscopy as a tyrosine radical.[384] On the basis of directed mutations this tyrosine, which is usually designated $\mathbf{Y_Z}$, is Tyr 161 of the D1 chain and is located ~1.2 nm from one of the chlorophylls of the $(Chl)_2$ pair.[371d,376,380,385–387] The two molecules in the pair are not in close contact, their central Mg atoms being ~1.0 nm apart. One of the two probably forms the $P680^{\bullet+}$ intermediate.[371d] Y_Z is also close to the Mn_4 cluster and to the imidazole group of His 190 of subunit D1 as is shown in Fig. 23-34C.[386,388] The $Y_Z^{\bullet}$ radical is able to accept an electron from the Mn_4 cluster within 30–1300 μs depending upon the oxidation state of the cluster (see Fig. 23-35).[388] If a proton is transferred synchronously from a bound H_2O, a neutral –OH group will be created on Y_Z. The proton may then be transferred to His 190, which can eject the proton on its other nitrogen atom into the lumen. Alternatively $Y_Z^{\bullet}$ may accept an electron to become tyrosinate $-O^-$, which then donates an electron to P^+, while His 190 accepts a proton directly from a bound H_2O.

In a mechanism proposed by Hoganson and Babcock (Fig. 23-35) four successive transfers, each of one H^+ + one e^-, leads to a three-electron oxidation of Mn ions, e.g., from the 2^+ and 3^+ oxidation states to all Mn^{4+}, and to joining of the two water oxygens to form a manganese peroxide linkage. Oxidation of the peroxide dianion to O_2 by the adjoining Mn^{4+} and Mn^{3+} ions completes the cycle. This mechanism is hypothetical, and various alternatives have been presented.[387–389b] Most assume a structure similar to that shown in Fig. 23-35. Some are based on nonenzymatic model reactions.[390,391] Chloride ions are essential to O_2 formation[393,394] especially in going from state S_2 to S_3 and S_3 to S_0. This suggests that Cl^- may function in passing electrons between Mn ions.[393,395] Calcium ions are also necessary, but it has been difficult to establish an exact function.[396–397a] A bicarbonate ion may also be an essential ligand in the Mn_4 cluster.[398,398a]

ATP synthesis in chloroplasts. The flow of electrons between PSII and PSI (Fig. 23-18) is of great importance for ATP formation. As previously mentioned, plastocyanin is usually the immediate donor to P700 and serves as a mobile carrier to bring electrons to this reaction center. In this function it is analogous to cytochrome *c* of mitochondrial membranes. The essentiality of plastocyanin was shown by study of copper-deficient *Scenedesmus* (Fig. 1-11). The photoreduction of CO_2 by H_2 is impaired in these cells, but the Hill reaction occurs at a normal rate.

Like mitochondria, chloroplasts (when illuminated) pump protons across their membranes (Fig. 23-18). However, while mitochondria pump protons to the outside, the protons accumulate on the inside of the thylakoids. The ATP synthase heads of coupling factor CF_1 are found on the outside of the thylakoids, facing the stromal matrix, while those of F_1 lie on the insides of mitochondrial membranes. However, the same mechanism of ATP formation is used in both chloroplasts and mitochondria (Chapter 18).

The cleavage of water at PSII also occurs on the inside of the thylakoids. The splitting of one water molecule leaves two protons (one per electron) inside the thylakoids, while the electrons are "photoejected" through the lipid bilayer to acceptor Q_A on the outside. The chlorophyll in PSI is likewise in contact with the inside of the bilayer with acceptor A_0 (Fig. 23-18) on the outside. Since the conversion of NAD^+ to NADH on the outside generates a proton, the overall reaction would be the pumping of one and a half protons per electron passing through the Z scheme.

The pathways involved in cyclic photophosphorylation in chloroplasts are not yet established. Electrons probably flow from the Fe–S centers Fd_x, Fd_a, or Fd_b back to cytochrome b_{563} or to the PQ pool as is indicated by the dashed line in Fig. 23-18. Cyclic flow around PSII is also possible. The photophosphorylation of inorganic phosphate to pyrophosphate (PP_i) occurs in the **chromatophores** (vesicles derived from fragments of infolded photosynthetic membranes) from *Rhodospirillum rubrum*. The PP_i formed in this way may be used in a variety of energy-requiring reactions in these bacteria.[399] An example is formation of NADH by reverse electron transport.

Protection of chloroplasts against radiation and oxygen. Carotenoids often act as accessory light-receiving pigments, but an additional function is protecting photosynthetic organisms against toxic effects of light.[400–403] Carotenoid photoprotection has been demonstrated in photosynthetic bacteria and in the reaction centers and light-harvesting complexes (LHC) of green plants. Excited chlorophyll molecules can pass from the singlet (^{1}Chl*) to the triplet (^{3}Chl*) state by intersystem crossing (Fig. 23-14). The triplet chlorophyll can then react with ordinary oxygen ($^3O_2$) to form singlet oxygen ($^1O_2$). The formation of ^{3}Chl* is favored when the intensity of sunlight is high and energy is absorbed in the LHCs faster than it can be utilized in the reaction centers. Carotenoids are able to quench excitation of both ^{3}Chl* and $^1O_2$. Strains of *Rhodobacter sphaeroides* that lack carotenoids rapidly form both ^{3}BChl* and BChl$^+$ cation radicals in their LH1 and LH2 complexes (Fig. 23-26). However, the presence of carotenoids quenches the triplet bacteriochlorophyll effectively and no formation of BChl$^+$ radicals is seen.[403] Another effect of excessive light energy is the reaction of O_2 with the highly reducing P700* of photoexcited PSI. This interaction can form triplet oxygen, which can react to generate superoxide anion radicals, H_2O_2, and hydroxyl radicals.[224,227,402]

In higher plants the quenching of both ^{3}Chl* and $^1O_2$ depends upon carotenoids[404] and also upon the large transthylakoid membrane potential that is generated by high light intensities.[405,406] High light intensity also induces a rapid reductive deoxygenation of epoxycarotenoids via the **xanthophyll** or **violaxanthin cycle** (Eq. 22-10).[401,407,407a,b] Epoxycarotenoids are found only in photosynthetic O_2 evolving organisms. Although occurring in response to light, the cyclic photodeoxygenation and reoxygenation is not a nonphotochemical process. Violaxanthin contains the epoxy structure at both ends of the molecule. Reduction of one end produces **antheroxanthin** and of both ends **zeaxanthin**. These three carotenoids are found in almost all higher green plants and algae. The deepoxidation is mediated by ascorbic acid, occurs in the lumen of the thylakoids, and is favored by the low pH developed during illumination. Epoxidation (Eq. 22-10) is catalyzed by a monooxygenase located on the stromal side. The significance of the xanthophyll cycle is puzzling. There may be specific binding sites, perhaps in the inner antenna complex on CP29 (Fig. 23-34), that bind zeaxanthin or antheroxanthin. This could alter the antenna structure to form an "exciton trap" in which the fluorescence lifetime would be decreased and excitation energy would be dissipated rapidly as heat.[408] Zeaxanthin is also found in the macular area of the primate retina.[401]

Recent studies, using an *Arabidopsis* mutant defective in the xanthophyll cycle, point to a chlorophyll-binding protein PsbS, which participates in nonphotochemical quenching at high light intensity.[401a,b] Another *Arabidopsis* protein, which is probably a blue light receptor, participates in an avoidance response by which chloroplasts move to the side wall to avoid strong light.[401c]

Yet another carotenoid function in PSII has been proposed. Under some conditions, when electron flow from Tyr_z is blocked, the bound monomeric chlorophyll Z (Chlz) acts as a secondary electron donor to P680$^+$. The cytochrome b_{559} subunits may have a similar function. Both cyt b_{559} and the carotenoid are essential for assembly of PSII, and both may participate in a protective cycle.[409,410] Chloroplasts generate both O_2 and powerful reducing materials such as the membrane-bound FeS centers of PSI, which may form superoxide ions by single-electron donation to O_2. Probably for this reason, chloroplasts are rich in superoxide dismutase which converts superoxide to O_2 and H_2O_2. The latter can diffuse into peroxisomes and react with catalase and peroxidases. It can also be reduced to H_2O within the chloroplasts by ascorbic acid and ascorbate peroxidase.[224,227] The resulting dehydroascorbate (Box 18-D) can be reduced back to ascorbate by glutathione (Box 11-B) and dehydroascorbate reductase, in the following electron transfer sequence:

$$\text{NADPH} \rightarrow \text{G-S-S-G} \rightarrow \text{dehydroascorbate} \rightarrow H_2O_2 \qquad (23\text{-}33)$$

Under extreme conditions of excess light energy **photoinhibition** is observed as a result of damage to the PSII structure.[411–415] The D1 polypeptide is cleaved, probably as a result of oxidation by $^1O_2$ and proteolysis. Damaged proteins are replaced and the PSII structure rebuilt, but the effect is a long-lasting decrease in photosynthetic efficiency. The cyanobacterium *Synechococcus* has three *PsbA* genes and resists UV-B radiation by exchanging a delicate D1 polypeptide with more resistant ones as necessary.[415] Other adaptations to varying light-intensity involve movement of light-harvesting complexes from the thylakoid stacks, which contain much PSII, to the stroma lamellae (Fig. 23-19), which contain more PSI. Some herbicides act by binding into the Q_B site in PSII. They may cause light-induced oxidative stress that kills the plant.[415a]

5. Control of Photosynthesis

The key reaction of the Calvin–Benson cycle of CO_2 reduction is the carboxylation of ribulose bisphosphate to form two molecules of 3-phosphoglycerate (Eq. 13-48). The properties of ribulose bisphosphate carboxylase (**rubisco**, Figs. 13-10 to 13-12), which catalyzes this reaction, are discussed in Chapter 13. It

is controlled in part by CO_2 and by natural inhibitors,[416] but regulation of rubisco starts at the transcriptional level.

Light-induced transcription. Plants depend upon light both as a source of energy and also for control of development. Many genes are activated by light in response to at least three groups of photoreceptors. These are **phytochromes** (Section H) and the blue light responsive **cryptochromes** (Section I) and the ultraviolet light **UV-B photoreceptors**.[417] The synthesis of chlorophyll, of reaction center proteins, and of many enzymes are controlled by light-induced transcription.[418–420] Among these processes are synthesis of both the large and small subunits of rubisco (Fig. 13-10). The small subunits are synthesized in the cytoplasm in a precursor form. After illumination the concentration of the rubisco mRNA may be increased 100-fold.[421,422] On the other hand, the large subunit of the carboxylase is encoded in chloroplast DNA, and stimulation of its synthesis by light appears to be at the translational level.[423]

Light-induced control via the ferredoxin/thioredoxin system. Rubisco is activated by CO_2 (Chapter 13) and by fructose 6-*P* and is inhibited by fructose 1,6-P_2 (Fig. 23-36),[424] whose accumulation is a signal to turn off the carboxylase. Conversely, fructose 6-*P* in high concentrations turns on the Calvin–Benson cycle. Like the reactions of gluconeogenesis (Chapter 17), photosynthetic CO_2 incorporation is dependent on the highly regulated fructose-1,6-bisphosphatase. In chloroplasts it is activated by light through the mediation of reduced ferredoxin and thioredoxin.[424–427] The small mobile thioredoxin (Box 15-C) is reduced to its dithiol form by reduced ferredoxin[428–429a] and then reduces one or more disulfide linkages in the fructose 1,6-bisphosphatase to activate that enzyme (Fig. 23-36). Other light-activated enzymes of the Calvin–Benson cycle include sedoheptulose-1,7-bisphosphatase, the phosphoribulokinase that forms ribulose 1,5-bisphosphate and the $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase. $NADP^+$-dependent malate dehydrogenase, which has a major function in C_4 plants (see Fig. 23-38), is totally inactive in the dark but is activated by the ferredoxin–thioredoxin system in the light.[427] The activity of the LHCII complex is also affected.[429b]

Another aspect of chloroplast metabolism is synthesis of starch. Formation of ADP-glucose from glucose 1-phosphate is induced by 3-phosphoglycerate, a "feed-ahead" type of regulation (Fig. 23-36). Although fructose 2,6-bisphosphate is absent from chloroplasts, it has an important regulatory function in the cytoplasm of plants as it does in animals.[425,430] In the plant cytosol triose phosphates from the chloroplasts are converted to fructose 6-*P*, glucose 6-*P*, UDP-glucose, and sucrose. Inorganic phosphate P_i, which accumulates in plant vacuoles, also has a regulatory function.[431] It activates the kinase that converts fructose 6-*P* to fructose 2,6-P_2 and inhibits the phosphatase that converts the bisphosphate back to fructose 6-*P*. The accumulated fructose 2,6-P_2 inhibits fructose-1,6-bisphosphatase and slows the conversion of triose phosphates to sucrose (Eq. 23-34). Accumulation of fructose 6-*P* due to decreased utilization for sucrose formation will have a similar effect. However, both 3-phosphoglycerate and dihydroxyacetone phosphate have opposite effects and will act to remove the inhibition by lowering the fructose 2,6-P_2 level and to promote rapid sucrose formation (Eq. 23-34).[430]

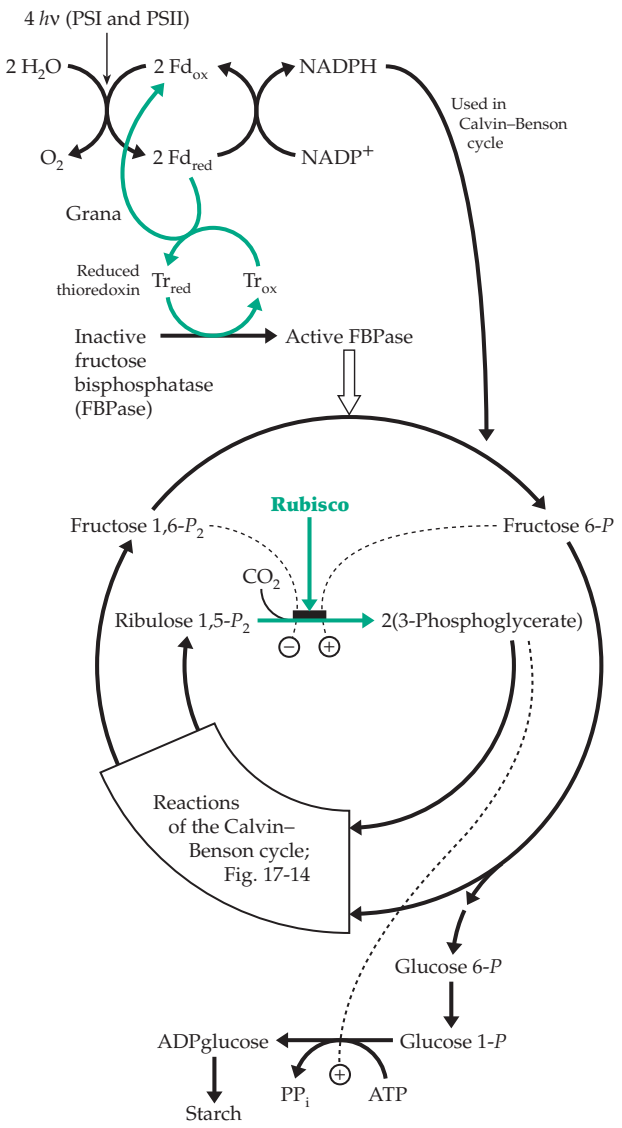
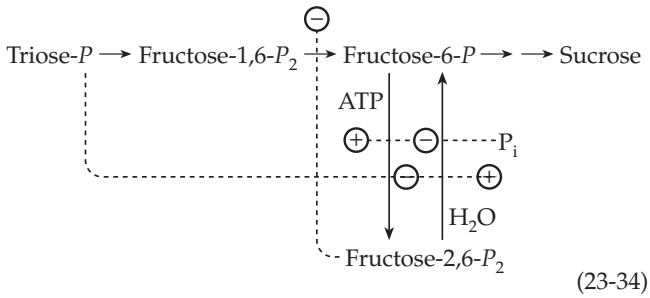


Figure 23-36 Some control mechanisms for photosynthetic assimilation of carbon dioxide. After Buchanan and Schurmann[424] with modifications.



(23-34)

6. Photorespiration; C_3 and C_4 Plants

The first product of incorporation of CO_2 via the Calvin–Benson cycle is 3-phosphoglyceric acid (Box 17-E). It was the rapid appearance of radioactivity from $^{14}CO_2$ in phosphoglycerate and other three-carbon (C_3) compounds that permitted Calvin and associates, using green algae, to work out the complex cycle as it is shown in Fig. 17-14. Green algae, spinach, and many common crop plants are often known as **C_3 plants**. Another group of plants, mostly of tropical origin and capable of extremely fast growth, behave differently.[224,227,432] In these plants, which include sugar cane, maize, and crabgrass, radioactivity from ^{14}C-containing CO_2 is found first in the C_4 compounds oxaloacetate, malate, and aspartate. These **C_4 plants** are characterized by high efficiencies in photosynthesis, a fact that explains the rapid growth of crabgrass and the high yield of corn. Maximum rates of CO_2 incorporation may attain 40–60 mg of CO_2 per square decimeter of leaf surface per hour (~0.3 mmol CO_2 $m^{-1}s^{-1}$ or ~0.10 mol CO_2 per mol of total chlorophyll per second), more than twice that for common crop plants.

Like all other organisms plants respire in the dark, but illumination of C_3 plants markedly increases the rate of oxygen utilization. This light-enhanced respiration (**photorespiration**) may attain 50% of the net rate of photosynthesis. Photorespiration prevents plants from achieving a maximum yield in photosynthesis. For this reason, its understanding and control assume great importance in agriculture. It is difficult to measure the rate of photorespiration, and the literature on the subject often refers instead to the **CO_2 compensation point**. This is the CO_2 concentration (at a given constant light intensity) at which photosynthetic assimilation and respiration balance. (Similarly, the **light compensation point** is the light intensity at which the rate of photosynthetic CO_2 incorporation and that of respiration exactly balance.) Normal air has a CO_2 content of ~0.03% or 300 ppm. For common C_3 crop plants the CO_2 compensation point is ~40–60 ppm at 25°C. The C_4 plants are characterized by a much lower CO_2 compensation point, often less than 10 ppm. In strong sunlight the CO_2 level of air in a field of growing plants drops. Furthermore, as the temperature rises on a hot day, the CO_2 compensation point rises. The result is a serious decrease in efficiency of photosynthesis for the C_3 plants but not for the C_4 plants.

Metabolism of glycolic acid. The 2-carbon glycolic acid is formed in large quantities in the chloroplasts of C_3 plants and moves out into the cytosol.[433,434] The major source of this acid is phosphoglycolate whose formation is catalyzed by rubisco in the chloroplasts through competition of O_2 for the CO_2 binding site of the enzyme (Eq. 13-50). It is easy to understand why an increase in the O_2 pressure in air increases the CO_2 compensation point for a plant. Another less important source of glycolate is transketolase, which may yield glycolaldehyde as a side product (Eq. 17-15). Glycolaldehyde can be oxidized readily to glycolate. Glycolate is metabolized rapidly, some in the chloroplasts,[435] but most in the peroxisomes. There the flavoprotein glycolate oxidase converts it to glyoxylate with formation of H_2O_2 (Fig. 23-37).[436] Some of the hydrogen peroxide formed may react nonenzymatically, decarboxylating glyoxylate to formate and CO_2, but most is probably destroyed by peroxidases or catalase. The latter enzyme is lacking in chloroplasts, one reason why oxidation of glycolate must occur in the peroxisomes.

Glyoxylate undergoes transamination to glycine, which can be oxidatively decarboxylated (Fig. 15-20) in the mitochondria. It can also be converted to serine,[437] some of which returns to the peroxisomes to be oxidized to hydroxypyruvic acid and glyceric acid (Fig. 23-37). The latter can be synthesized into glucose. The net result is the stimulation of a large amount of metabolism that ultimately produces CO_2 and apparently accounts for the light-induced respirations of plants. Because much of the glycine formed in the peroxisomes is oxidatively decarboxylated in the mitochondria, photorespiration also generates large amounts of NH_3. This is recycled into amino acids within the photosynthetic cells (green lines in Fig. 23-37), an energy-requiring process.[438]

Although much metabolism occurs as a result of photorespiration, it appears to waste energy rather than to provide energy to the plant. Why then don't plants avoid this process? Wouldn't a small change in the structure of ribulose bisphosphate carboxylase allow plants to avoid photorespiration and to grow more efficiently? The answer is not clear. It has been difficult to create such modified plants, and there is a possibility that they would not grow well. One theory is that photorespiration protects plants when the CO_2 pressure is low and the absorbed light would damage the chloroplasts if there were not a way to utilize the accumulating reduced Fe–S proteins generated by PSI. Photorespiration provides a mechanism.[434] Most

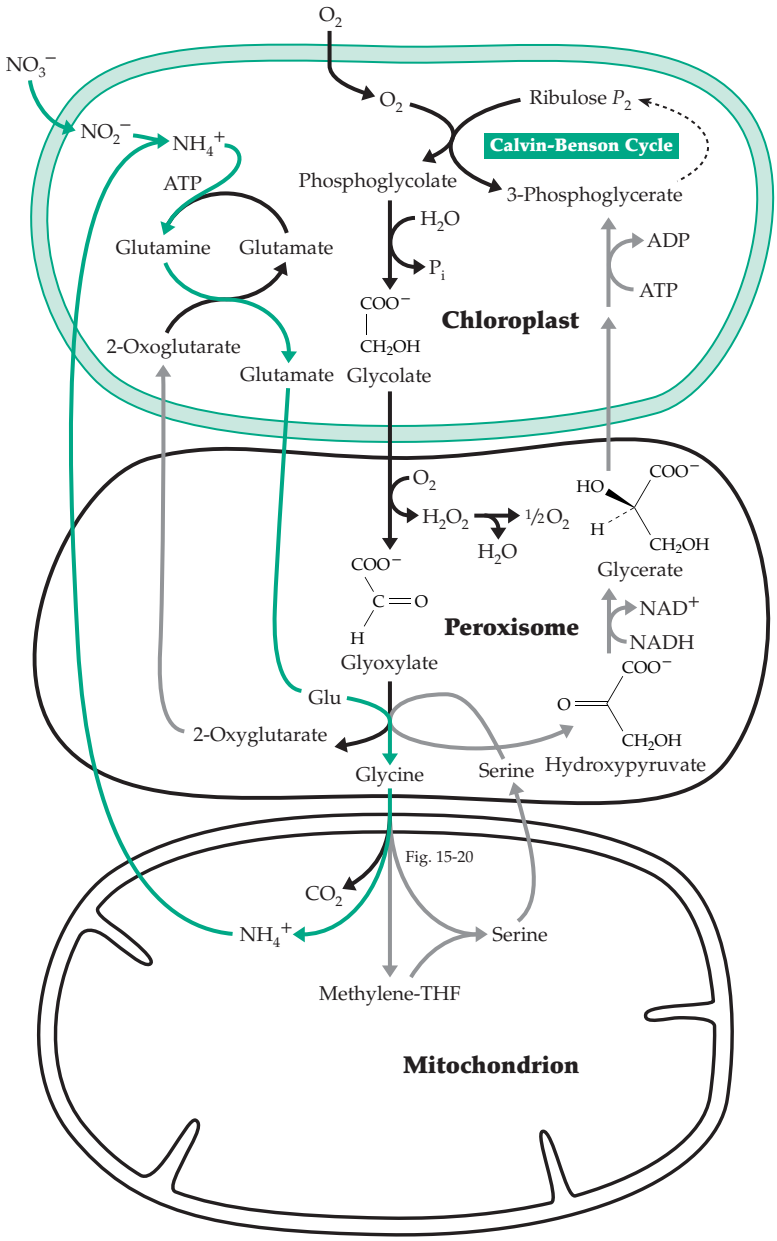
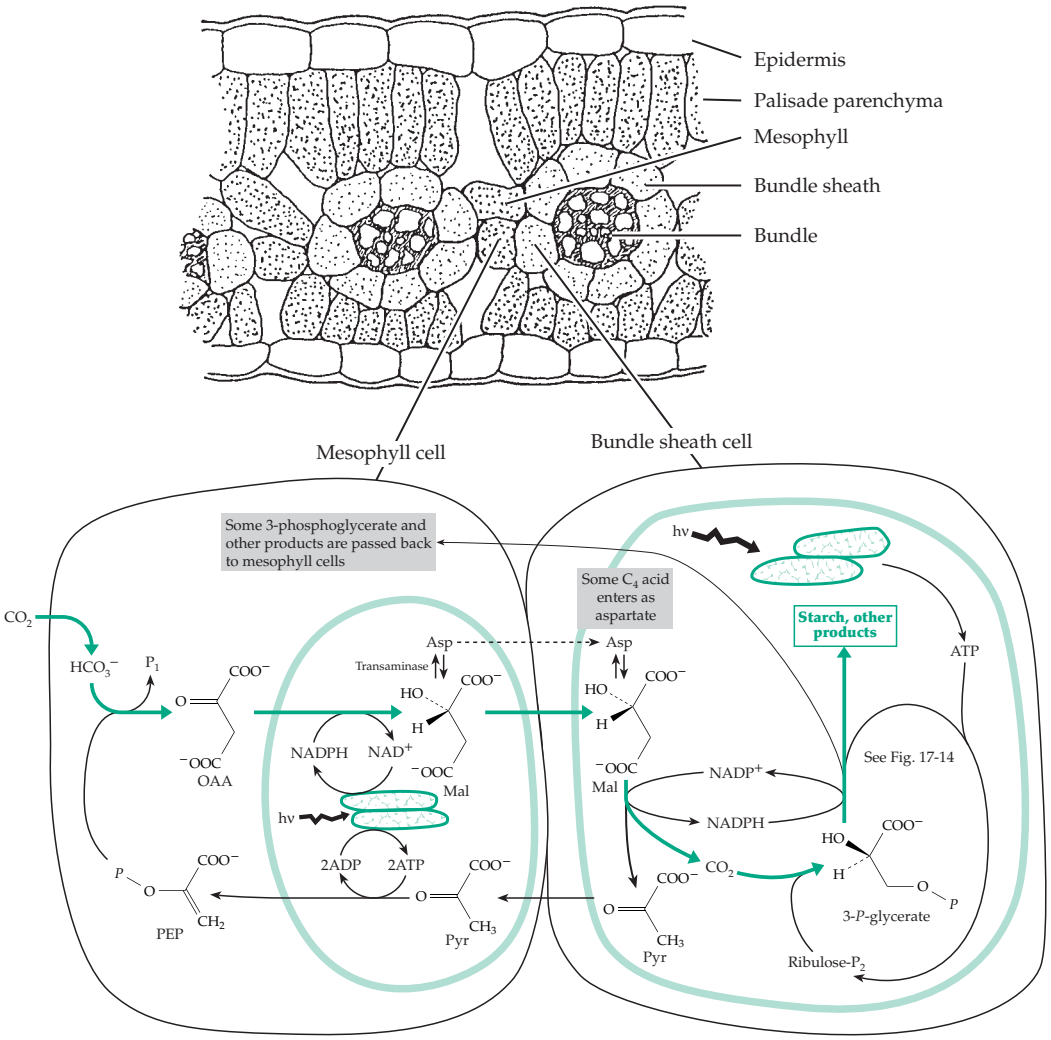


Figure 23-37 Production of glycolate by chloroplasts and some pathways of its metabolism in peroxisomes and in mitochondria. After Tolbert[436] and Givan *et al.*[438]

efforts to breed plants with lower photorespiration rates or to inhibit it chemically have failed.[433,439]

The C_4 cycle for concentration of carbon dioxide. The C_4 plants reduce their rate of photorespiration by using a CO_2 concentrating mechanism that enables them to avoid the competition from O_2. All species of C_4 plants have a characteristic internal leaf anatomy in which a single dense layer of dark green cells surrounds the vascular bundles in the leaves. This **bundle sheath** is surrounded by a loosely packed layer of cells, the **mesophyll**, an arrangement that is sometimes called the "Kranz anatomy." In C_4 plants there is a separation of the chemical reactions between the mesophyll and the bundle sheath cells. The incorporation of CO_2, as bicarbonate ion into oxaloacetate, occurs in the mesophyll cells, principally through the action of PEP carboxylase (Fig. 23-38). Oxaloacetate is reduced to malate by light-generated NADPH. Alternatively, it undergoes transamination to aspartate. Both malate and aspartate then diffuse out of the mesophyll cells and into bundle sheath cells where the malate undergoes oxidative decarboxylation via the malic enzyme (Eq. 14-42) to pyruvate (Fig. 23-38). Aspartate also can be converted to oxaloacetate, malate, and pyruvate in the same cells. The overall effect is to transport CO_2 from the mesophyll cells into the bundle sheath cells along with two reducing equivalents, which appear as NADPH following the action of the malic enzyme. The CO_2, the NADPH, and additional NADPH generated in the chloroplasts of the bundle sheath cells are then used in the Calvin–Benson cycle reactions to synthesize 3-phosphoglycerate and other materials. Of the CO_2 used in the bundle sheath cells, it is estimated that 85% comes via the C_4 cycle and only 15% enters by direct diffusion. The advantage to the cell is a higher CO_2 tension, less competition with O_2, and a marked reduction in photorespiration.

The pyruvate produced in the bundle sheath cell is largely returned to the mesophyll cells where it is acted upon by **pyruvate-phosphate dikinase**.[441] This unusual enzyme (Eq. 17-55) phosphorylates pyruvate to PEP while splitting ATP to AMP and PP_i. The latter is in turn degraded to P_i. In effect, two high-energy linkages are cleaved for each molecule of pyruvate phosphorylated. Because of this extra energy need, it is thought that cyclic photophosphorylation is probably more important in the chloroplasts of the mesophyll cells than in the bundle sheath cells. It also accounts for the fact that C_4 plants are less efficient than C_3 plants under cool or shaded conditions.[442] Other CO_2-concentrating mechanisms exist in plants.[442a] For example, cyanobacteria accumulate HCO_3^- ions in carboxysomes, polyhedral bodies to which rubisco adheres. An ABC type ATP-dependent transporter powers the bicarbonate accumulation.[442b]



(23-35)

Figure 23-38 The C_4 cycle for concentrating CO_2 in the C_4 plants. From Haag and Renger with alterations.[440]

Metabolism in the family Crassulaceae. The crassulacean plants are a large group that includes many ornamental succulents such as *Sedum*. They have a remarkable metabolism by which large amounts of malic and isocitric acids are synthesized at night. During the day when photosynthesis occurs these acids disappear. The stomata in the leaves (Chapter 1) stay closed during the day and open only at night, an adaptation that permits the plants to live with little water. However, this requires that the plant accumulates carbon dioxide by night and incorporates it photosynthetically into organic compounds by day.[443] A possible mechanism is shown in Fig. 23-39. On the left side of the figure are reactions by which starch can be broken down at night to PEP. While it would also be possible to produce that compound by the glycolysis reactions, labeling studies have indicated that the pentose phosphate pathway is more important.[330] The PEP acts as the CO_2 acceptor to create oxaloacetate, which is then reduced to malic acid. A balanced fermentation reaction (Fig. 23-39; Eq. 23-35) can be written by using the NADPH formed in the conversion of glucose 6-*P* to ribulose 5-*P*. During the day when ATP and NADPH are available in abundance from photoreactions, the conversions on the right side of the figure can take place. The initial step, the release of CO_2 from malic acid by the malic enzyme, is the same as that employed by C_4 plants. In this case, it is used to release the CO_2 stored by night, making it available for incorporation via the Calvin–Benson cycle. The remaining pyruvate is reconverted to starch.

$$C_6H_{10}O_5\ (\text{starch}) + 2\ CO_2\ (g) \rightarrow 2\ C_4H_4O_5^{2-}\ (\text{malate}) + 4\ H^+$$
$$\Delta G'\ (\text{pH } 7) = -159\ \text{kJ mol}^{-1} \qquad (23\text{-}35)$$

Many plants store substantial amounts of malate in their cytoplasm and in vacuoles. It apparently

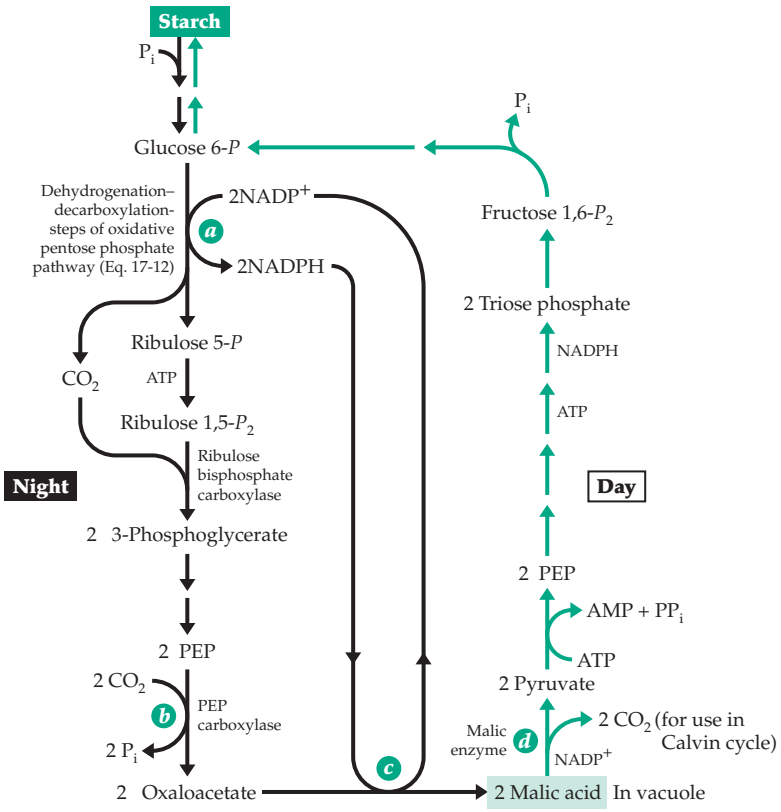


Figure 23-39 A proposed night–day cycle of crassulacean acid metabolism.

serves as a ready reserve for carbohydrate synthesis.

7. Photosynthetic Formation of Hydrogen

A system consisting of chloroplasts, ferredoxin, and hydrogenase has been used to generate H_2 photosynthetically.[444] This may be a prototype of a method of solar energy generation for human use. Another photochemical hydrogen-generating system makes use of both the nitrogen-fixing heterocysts and photosynthetic vegative cells of the cyanobacterium *Anabaena cylindrica*.[445] In this instance hydrogen production is accomplished by nitrogenase (Eq. 24-6). Photogeneration of H_2 by bacteria is just one of many kinds of photometabolism observed among photosynthetic microorganisms.

F. Vision

The light receptors of the eye perform a very different function from those of chloroplasts. Visual receptors initiate nerve impulses, and their primary requirement is a high sensitivity. By the use of stacked membranes containing a high concentration of an intensely absorbing molecule[446,447] the most sensitive visual receptors are able to trap nearly every photon that strikes them. The retina of the human eye contains more than 10^8 tightly packed receptor cells of two types. The **rod cells** are extremely sensitive. Used for night-time vision, they give a "black and white picture" and are concentrated around the periphery of the retina. The retina works as a coincidence detector. An ensemble of ~500 rods must register 5–7 isomerizations within a few tenths of a second in order to trigger a nerve impulse.[448] The less sensitive **cone cells**, which are most abundant in the center of the retina, are of three types with different spectral sensitivities. They provide color vision.

The retinal receptors have a very active metabolism. Human rod cells (Fig. 23-40) may live and function for a hundred years.[190] A self-renewal process leads to a casting off of the older membranous discs from the end of the rod[194] and replacement by new discs at the end nearest to the nucleus. The rod outer segment is surrounded by a plasma membrane. Within the membrane but apparently not attached to it are ~500 stacked discs of ~2 μm diameter and with a repeat distance between centers of ~32 nm. Each disc is enclosed by a pair of membranes each ~7 nm thick with a very narrow space between them. From electron micrographs it appears that this space within the discs is sealed off at the edges. Somewhat larger spaces separate adjacent discs.

The membranes of the rod discs are ~60% protein and 40% lipid (Table 8-3). About 80% of the protein is **rhodopsin** (visual purple), a lipoprotein that is insoluble in water but soluble in detergent solutions. Digitonin is widely used to disperse rhodopsin molecules because it causes no change in optical properties. In addition to rhodopsin, in the outer segment discs of frog retinal rods, there are ~65 molecules of phospholipid and smaller amounts of other materials for each molecule of rhodopsin (Table 8-3). The cone cells have a similar architecture but have a different shape and contain different light receptors. The receptors in the cones are present in deep indentations of the plasma membrane rather than in discs within the cytoplasm.

1. Visual Pigments

The rod pigment rhodopsin is readily available from cattle retinas and has been studied for many

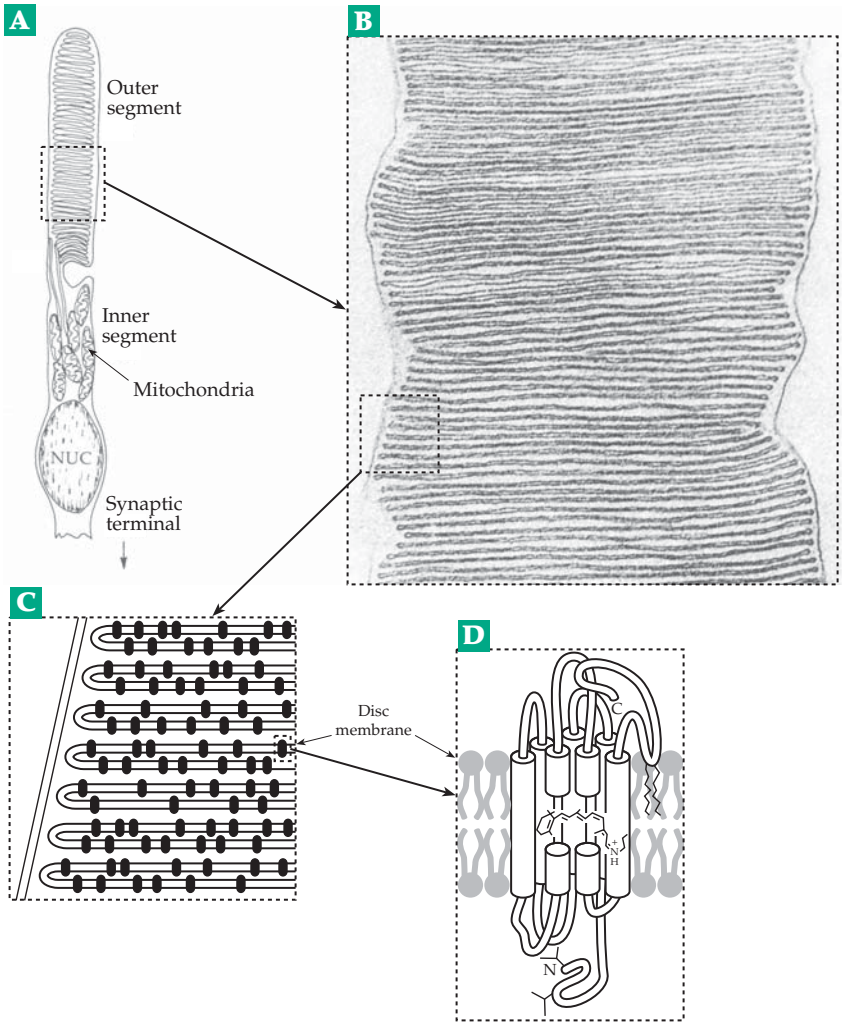
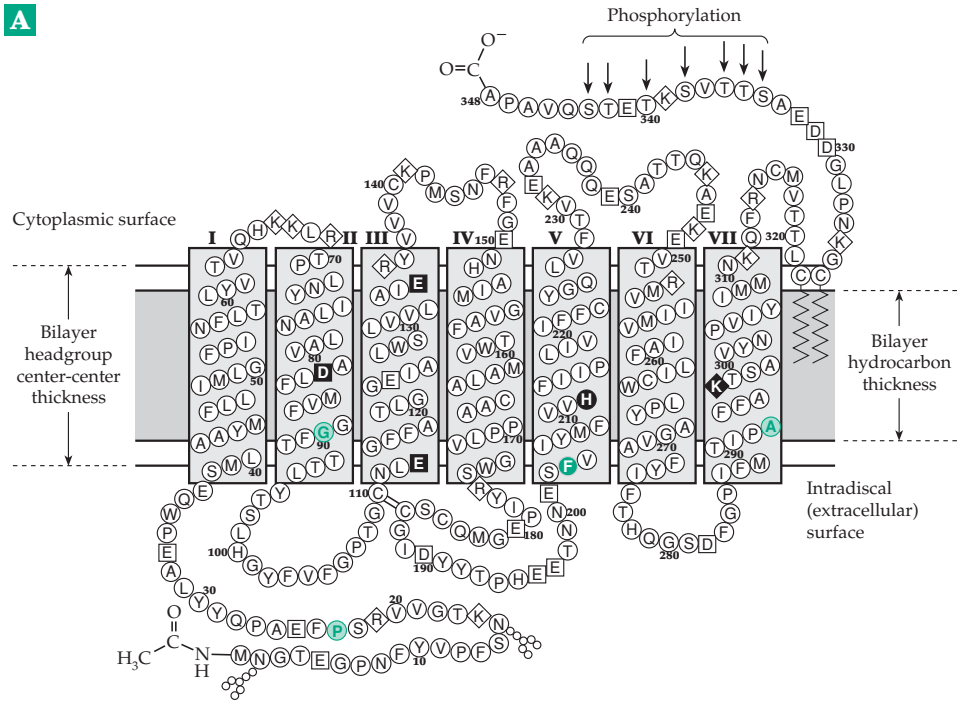


Figure 23-40 (A) Diagram of a vertebrate rod cell. From Abrahamson and Fager.[446] (B) Electron micrograph of a longitudinal section of the outer segment of a retinal rod of a rat. There are 600–2000 discs per rod and 2×10^4 and 8×10^5 rhodopsin molecules per disc. Courtesy of John E. Dowling. (C) Enlarged section; from Dratz and Hargrave.[451] (D) Schematic drawing of rhodopsin. The two α helices in the front have been partly cut away to reveal the 11-*cis* retinal in protonated Schiff base linkage to lysine 296. From Nathans.[448] Courtesy of Jeremy Nathans.

years. It has a molecular mass of ~41 kDa of which ~2 kDa is contributed by two asparagine linked oligosaccharides.[195] Both bovine and human rhodopsins consist of a 348-residue protein known as **opsin** to which is bound a molecule of vitamin A aldehyde, **retinal**. Human and bovine opsins are 93% homologous. A totally synthetic 1057 base pair gene for bovine rhodopsin was made by Khorana and associates.[449,450] The gene was constructed with 28 unique sites for cleavage by specific restriction endonucleases. These have allowed easy specific mutation of the gene and production of a wide variety of mutant forms of opsin. Similar synthetic genes have been constructed for the three human cone pigments[452] and for the related bacterial protein **bacteriorhodopsin**.[453]

Transmembrane structure.

From its circular dichroism rhodopsin was estimated to be 60% helical, and its amino acid sequence suggested that it contains seven parallel membrane-spanning helices (Fig. 23-41)[451] just as does bacteriorhodopsin (Section G; Fig. 23-45). Rhodopsin and other visual pigments are also members of the large family of **G-protein coupled receptors**, which includes the β_2 adrenergic receptor pictured in Fig. 11-6. It has been hard to determine the three-dimensional structure of rhodopsin or other receptors of this family. However, their relationship to bacteriorhodopsin, whose structure was obtained in 1975 by electron crystallography[454] and more recently by X-ray crystallography at 0.15 nm resolution,[455] permitted modeling based on similarities among the proteins.[456–458] New results of electron crystallography[459,460] and mass spectrometry[461] have been combined with studies of mutant forms of rhodopsin and other visual pigments to provide the picture given in Fig. 23-41A,B. Recently a higher resolution structure (Fig. 23-41D) has been obtained by X-ray diffraction.[461a–c] Mutations have been introduced in every part of the rhodopsin molecule, and the effects on photoreception, protein stability, and other properties have been observed.[448,450,462–465] Some of the essential residues identified are indicated in Fig. 23-41A, a schematic diagram showing the seven helices and connecting loops. Figure 23-41B shows an end view of the helix bundle with its retinal prosthetic group buried in the interior of the protein. Rhodopsin is roughly cylindrical with a length of 6–6.5 nm and a diameter of ~2.8 nm and is embedded in the phospholipid bilayer with its long axis perpendicular to the membrane surface (Fig. 23-40). The two oligosaccharide chains, which are attached near the N terminus, project into the intradiscal space on the side away from the cytoplasm. Palmitoyl groups on two cysteine side chains help to anchor the protein. Rhodopsin apparently exists in the mem-



● A few residues essential to the mechanism of vision.

○ Point mutations causing serious diseases. P23H is a frequent cause of retinitis pigmentosa. Mutations G90D and A292E cause congenital night blindness.

● The mutation C203R in the green-sensitive pigment causes red-green color blindness.

Figure 23-41 (A) Model of the topological organization of bovine rhodopsin according to the consensus analysis of Baldwin.[456] Oligosaccharide chains are portrayed on the glycosylation sites Asn 2 and Asn 15. Glutamate 113 provides the counter ion for the *N*-protonated retinal Schiff base. Cysteines 110 and 187 form an essential disulfide linkage. Histidine 211 modulates the interconversion of the metarhodopsin forms I and II. Lysine 296 forms a Schiff base with 11-*cis*-retinal. Cysteines 322 and 323 are sites of palmitoylation, and as many as six serine and threonine hydroxyls (indicated by green arrows) may become phosphorylated during desensitization of rhodopsin. Aspartate 83, glutamate 134, and histidine 211 may be essential for proton movements. The point mutation C203R in the green-sensitive pigment causes red-green color blindness. The mutation

branes as monomers. From the composition it can be calculated[451] that the average center-to-center distance of the cylindrical molecules must be ~5.6 nm.

The visual chromophores. Rhodopsin has been an object of scientific interest for over 100 years.[462] Wald and associates[469,470] established that rhodopsin contains 11-*cis*-retinal bound to the opsin in Schiff base linkage (Eq. 23-36). When native rhodopsin is treated with sodium borohydride, little reduction is observed. However, after the protein is bleached by light, reduction of the Schiff base linkage becomes rapid, and the retinal is incorporated into a secondary amine, which was identified as arising from Lys 296.

In a crystal structure[471] 11-*cis*-retinal has the 12-*s*-*cis* conformation shown at the top in Eq. 23-36 rather than the 12-*s*-*trans* conformation at the center and in which there is severe steric hindrance between the 10–H and 13–CH_3. Nevertheless, ^{1}H– and ^{13}C–NMR spectroscopy suggest that the retinal in rhodopsin is in a twisted 12-*s*-*trans* conformation.[472,472a] The Schiff base of 11-*cis*-retinal with *N*-butylamine has an absorption maximum at ~360 nm but *N*-protonation, as in the structure in Eq. 23-36, shifts the maximum to 440 nm with ε_{max} = 40,600 M^{-1} cm^{-1} (Fig. 23-42). This large shift in the wavelength of the absorption maximum (the *opsin shift*) indicates that binding to opsin stabilizes the light-excited state by almost 33 kJ/mol compared to that in the free *N*-protonated Schiff base. This is evidently the result of a fixed negative charge, that of Glu 113, which is near the polyene chain of the retinal (Fig. 23-41B) and probably separated from it by a hydrogen-bonded water molecule.[473–477]

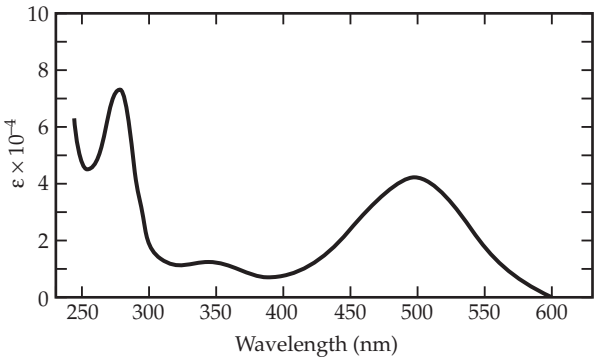
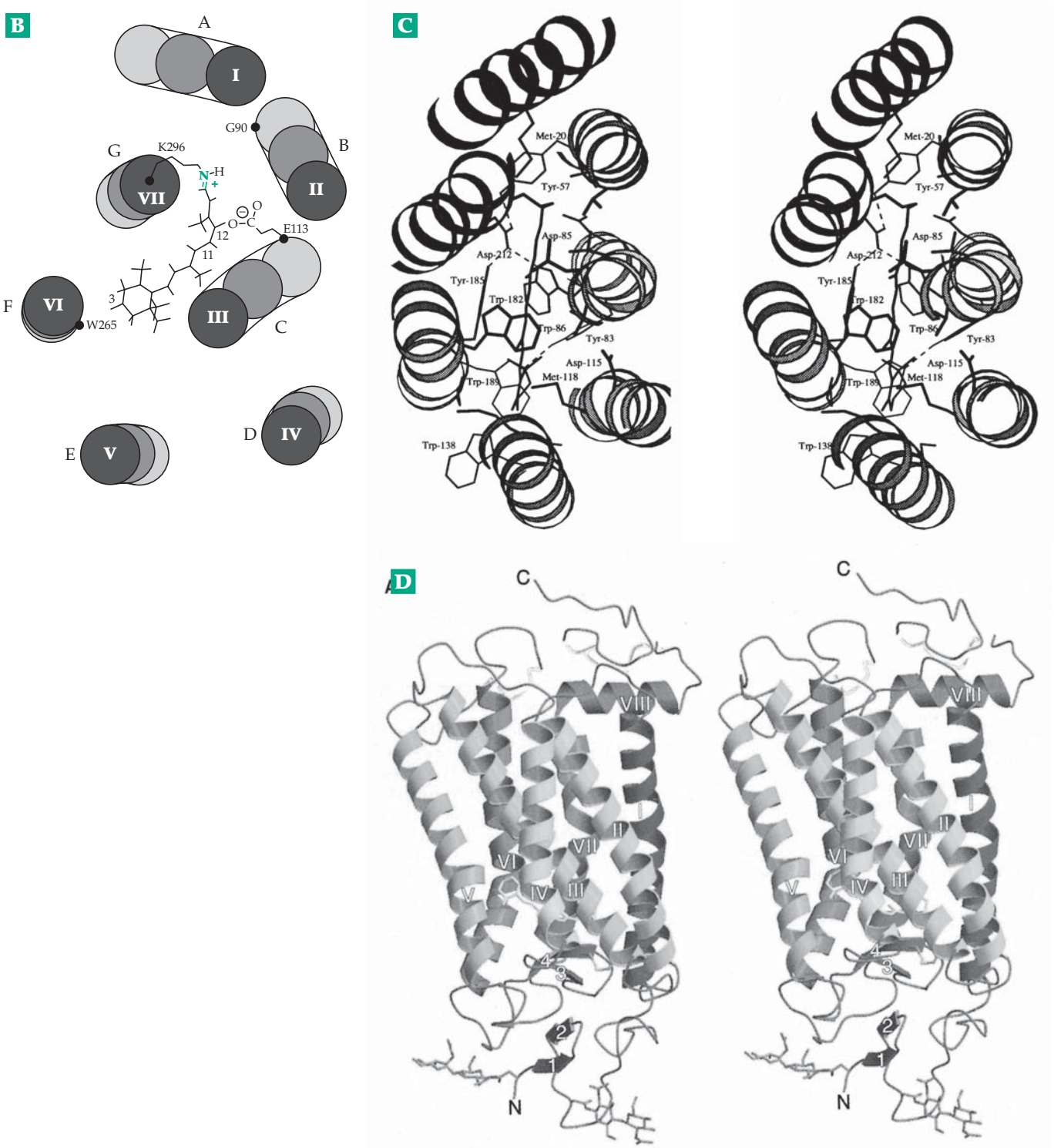
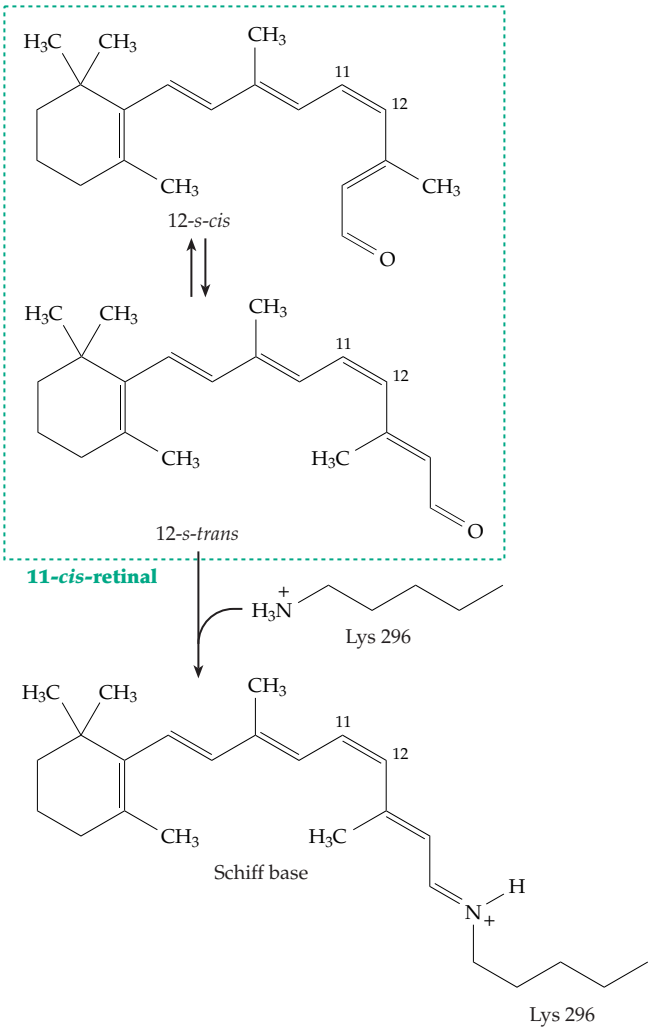


Figure 23-42 Absorption spectrum of cattle rhodopsin in aqueous dispersion with a nonionic detergent. From H. Shichi *et al.*[486,487]

P23H in rhodopsin is one of the most frequent causes of retinitis pigmentosa, and mutations G90D and A292E cause congenital night blindness. From Barnidge *et al.*[461] with modifications. (B) Structural model of rhodopsin based on the helix arrangement of Baldwin[456] and NMR constraints. The seven transmembrane helices, viewed from the cytoplasmic surface, are shown at three levels indicated by differences in shading. The α-carbons of residues of interest are shown in dots of various sizes, indicating the depth from the cytoplasmic domain. Gly 90, Glu 113, and Lys 296 are on the extracellular half of the transmembrane domain and close to each other in space. The 11-*cis*-retinal chromophore has been incorporated into the model using NMR constraints, which require a close interaction between Glu 113 and C_{12} of the chromophore. The relative position of the β-ionone ring and Trp 265 agrees well with crosslinking data of Zhang *et al.*[466] From Han and Smith.[467] (C) Stereoscopic view of the retinal-binding pocket of **bacteriorhodopsin** viewed from the cytoplasmic surface. The retinal, in Schiff base linkage to Lys 216, runs across the central cavity from top to bottom in this view. From Grigorieff *et al.*[468] (D) Ribbon drawing of bovine rhodopsin (stereoview). From Palczewski *et al.*[461a]





(23-36)

Three types of cone cells in the human retina are needed for color vision. Four genes specify the proteins for rhodopsin and for related cone photoreceptors absorbing blue (~425 nm), green (~530 nm), and red (~560 nm) light.[477–479b] All of the cone opsins also bind 11-*cis*-retinal. The rhodopsin gene is located on human chromosome 3, while that of the blue pigment is found on chromosome 7. However, the green and red sensitive pigment genes, whose sequences are 96% identical,[478] are close together on the q arm of the X-chromosome and near the gene for glucose-6-phosphate dehydrogenase. Examination of cloned DNA from persons with inherited red-green color-blindness shows that loss of a functional form of one of these genes is usually responsible for the problem.[480–482] Among Caucasians 8% of males and 1% of females differ from the normal in their color vision. About 30% of affected males are **dichromats** and lack either the red-sensitive pigment (they are protanopes) or the green-sensitive pigment (deuteranopes). They usually have a partial gene deletion. The other 70% often have hybrid genes created by errors in recombination during cell division.[483] A few deuteranopes have the point mutation C203R in the green pigment. John Dalton, of atomic theory fame, reasoned that his red-green confusion resulted from a blue tint in the vitreous humor of his eyes and ordered that they be dissected after his death (in 1884). There was no blue tint but DNA analysis performed more recently showed that Dalton was a deuteranope.[484] Defects in the blue-sensitive receptor are relatively rare affecting about 1 in 500 persons, while only one person in 100,000 has a total lack of color discrimination.[481,485]

All retinal-dependent visual pigments form Schiff bases with lysine side chains of the photoreceptor proteins. How can the same chromophore be "tuned" to absorb across the wavelength range of 360 to 635 nm? Modern techniques such as resonance Raman[477,479] and FTIR spectroscopies and study of mutant forms[488] have shown that interaction of the conjugated double bond system of the chromophores with immediately adjacent dipoles of side chain groups and peptide linkages is sufficient to account for the great variability in absorption maxima.

Visual pigments of many species have been investigated. Most vertebrate animals have a rhodopsinlike pigment plus a variable number of cone pigments. Mammals typically have only two, a short-wavelength pigment absorbing maximally in the ultraviolet, violet, or blue region[488a-c] and a long-wavelength pigment with maximum absorption in the green or red region.[489] The bottlenose dolphin has only a rod pigment of λ_{max} 524 nm.[489] In contrast, the chicken has four cone pigments with maximum absorption for violet, blue, green, and red.[490,490a] The red light receptor, called **iodopsin**, absorbs maximally at 571 nm. However, it binds chloride ions which induce an additional 40-nm red shift. The Cl^--binding site involves His 197 and Lys 200 which are present in an extracellular loop (Fig. 23-41A) and quite far from the bound retinal.[490] Human red and green color vision pigments, and also a green-sensitive pigment of the reptile *Gecko gecko*, undergo spectral shifts upon binding of Cl^- in the same site.[491] However, rhodopsin and most other visual pigments do not share this behavior.

Fishes live in a variety of environments and have a diversity of visual pigments. Goldfish have genes for five opsins, one of which gives rise to an ultraviolet light receptor. They are also sensitive to polarized light.[492] Related visual pigments occur throughout the animal kingdom. Even the eyespot of the alga *Chlamydomonas* (Fig. 1-11) contains rhodopsin with some sequence homology to invertebrate opsins.[493,494] The pineal glands of chickens and probably of reptiles[495–496a] as well as those of fish[497] also contain rhodopsinlike pigments. In a few freshwater marine species the visual pigments (**porphyropsins**) contain **3-dehydroretinal**. The peak positions of light absorption depend both upon the nature of the bound alde-

hyde and on the protein, the latter having the larger effect. Thus, retinal-based pigments absorb in the entire range 467–528 nm (18,900–21,400 cm^{-1}). The fruit fly, *Drosophila*, contains 3-hydroxy-11-*cis*-retinal in its rhodopsin and also contains other related photoreceptors.[498]

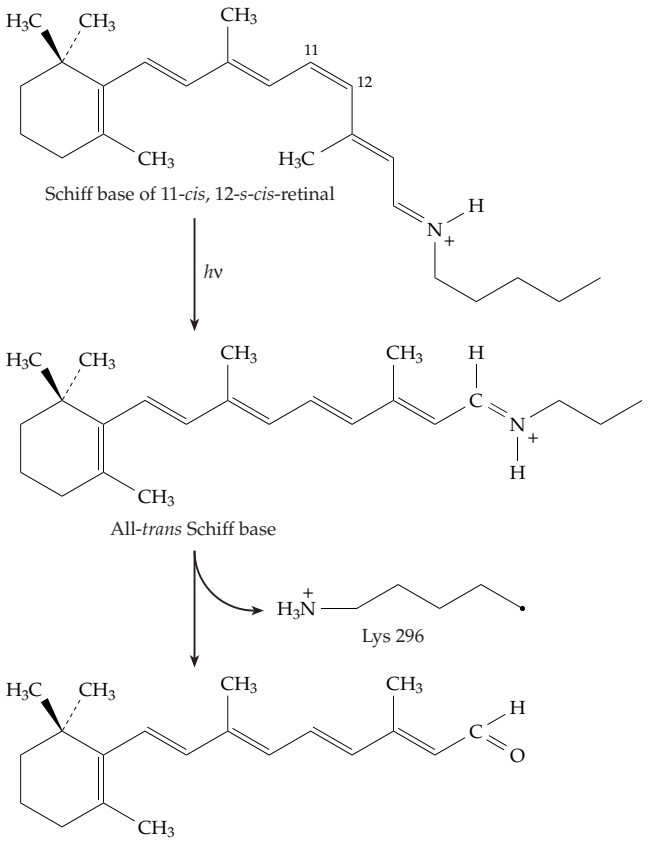
2. The Light-Induced Transformation

The retinal Schiff base chromophore is embedded in rhodopsin with its transition dipole moment parallel to the plane of the discs, i.e., perpendicular to the direction of travel of the incoming photons. Absorption of a photon leads to a sequence of readily detectable spectral changes.[37,461b,499,500] The relaxation times indicated in Eq. 23-37 are for 20°C.

Rhodopsin (498 nm, 20.1 x 10^3 cm^{-1})
$h\nu \downarrow \tau < 6 \times 10^{-12}$s
Bathorhodopsin (prelumirhodopsin; 543 nm, 18.4 x 10^3 cm^{-1})
$\downarrow \tau$ ~50 ns
Lumirhodopsin (497 nm), 20.2 x 10^3 cm^{-1})
$\downarrow \tau$ ~ 50 µs
Metarhodopsin I (478 nm, 20.9 x 10^3 cm^{-1})
$\downarrow\uparrow$ milliseconds
Metarhodopsin II (380 nm, 26.2 x 10^3 cm^{-1})
$\downarrow$ seconds
All-*trans*-retinal (387 nm, 25.8 x 10^3 cm^{-1}) + opsin (23-37)

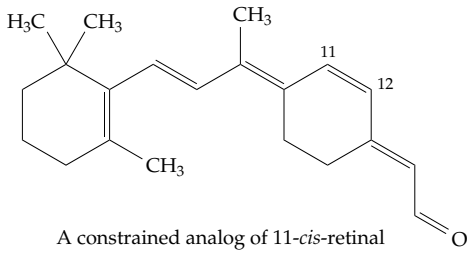
The intermediate chemical species have been named bathorhodopsin, lumirhodopsin, and metarhodopsin I and II. At very low temperatures a transient form **photorhodopsin** with a wavelength maximum at ~580 nm may precede bathorhodopsin.[461b,501–502a] Furthermore, nanosecond photolysis of rhodopsin has revealed a **blue-shifted intermediate** that follows bathorhodopsin within ~40 ns and decays into lumirhodopsin.[500,503,504] The overall result is the light-induced isomerization of the bound 11-*cis*-retinal to all-*trans*-retinal (Eq. 23-38) and free opsin. The free opsin can then combine with a new molecule of 11-*cis*-retinal to complete the photochemical cycle.

What are the chemical structures of the intermediates in Eq. 23-37, and why are there so many of them? The answer to the last question is that the initial photochemical process is very fast. Subsequent conformational rearrangments and movement of protons are slower, occur in distinct steps, and give rise to the observed series of intermediates. To shed light on these processes many experiments have been done with analogs of retinal,[502,505–508] often using very rapid spectroscopic techniques.[37,508] These studies have shown that the isomerization of the Schiff base from 11-*cis* to all-*trans* occurs in the first very rapid step of



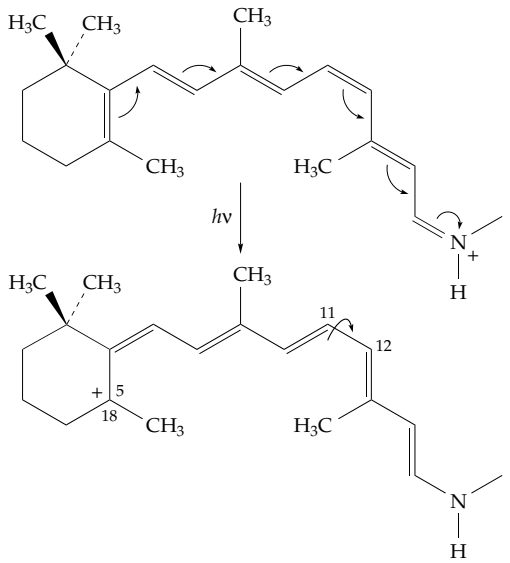
(23-38)

Eq. 23-37.[499,509–510a] Constrained analogs of 11-*cis* retinal also combine with opsin to form rhodopsinlike molecules with absorption maxima near 500 nm.



However, most of these analogs cannot isomerize and illumination does not produce bathorhodopsinlike molecules.[504,509,511]

In the photoexcited state the positive charge on the Schiff base is delocalized. For example, with some of the positive charge located on C5, rotation could occur around the more nearly single C11–C12 bond to give the all-*trans* isomer (Eq. 23-39). However, it seems more likely that simultaneous rotation occurs around two connected single and double bonds.[511a] The mechanism of this photoisomerization, which is among the fastest known chemical reactions, is still being investigated.[511b] In native rhodopsin the conversion to bathorhodopsin occurs with a high quantum



(23-39)

yield of 0.67 within 0.2 ps, a time comparable to the period of torsional vibrations of the retinal. This extreme speed suggests that the isomerization is a concerted process that is **vibrationally coherent**.[511c] Vibrational motion in the electronically excited state is utilized in the isomerization process.[506,512]

The reaction sequence of Eq. 23-37 can be slowed by lowering the temperature. Thus, at 70K illumination of rhodopsin leads to a **photostationary state** in which only rhodopsin, bathorhodopsin, and a third form, **isorhodopsin**, are present in a constant ratio.[510] Isorhodopsin (maximum absorption at 483 nm)[513] contains 9-*cis*-retinal and is not on the pathway of Eq. 23-37. Resonance Raman spectroscopy at low temperature supports a distorted all-*trans* structure for the retinal Schiff base in bathorhodopsin.[510] The same technique suggests the *trans* geometry of the C=N bond shown in Eqs. 23-38 and 23-39. Simple Schiff bases of 11-*cis*-retinal undergo isomerization just as rapidly as does rhodopsin.[514]

Some step in the sequence of Eq. 23-37 must initiate a chemical cascade that sends a nerve impulse out of the rod cell. This is accomplished through a GTP-dependent G protein cascade as outlined in Fig. 23-43. Light-activated rhodopsin initiates the cascade by interacting with the G protein **transducin**. Energy for the activation comes from the quantum of absorbed light. While the primary chemical reaction has long been recognized as the isomerization of the retinal Schiff base,[515] it is not obvious how this generates the signal for transducin to bind and become activated.

The seven helices of rhodopsin form a "box" around the bound retinal. The environment of the retinal is largely hydrophobic. However, there are also buried polar groups, some of which lie in conserved positions in more than 200 G-protein-coupled receptors[458] and internal water molecules whose vibrational spectra can be detected.[516] As in bacteriorhodopsin (Fig. 23-41, C) the buried polar groups and water molecules are doubtless hydrogen-bonded in an internal network. We can anticipate two effects of the isomerization reaction: (1) It will distort the shape of the box in which it occurs. (2) It will break some hydrogen bonds and allow new ones to form and may affect the balance of electrical charges within the protein. This in turn can lead to proton movements and alterations in the internal hydrogen-bonded network. Both of these anticipated effects have been observed.

Conformational changes induced in the rhodopsin protein by the isomerization of the retinal Schiff base include significant movement of the end of helix VI(F) at the cytosolic surface as well as smaller movements of other helices.[517,518] Spectroscopic measurements indicate that the Schiff base nitrogen remains protonated in both the bathorhodopsin and metarhodopsin I forms and in metarhodopsin II, the first long-lived form in the sequence.[518a,b] It seems likely that the proton has jumped via a bridging water molecule and the E113 carboxylate to the external (intradiscal) surface of the molecule. At the same time one or more protons are apparently taken up on the cytoplasmic side.[519] Study of mutant forms suggests that glutamate 134, near the cytosolic surface, and histidine 211 may be involved in proton transport.[520] By analogy with bacteriorhodopsin, aspartate 83 is probably also involved. The combination of conformational change plus altered charge distribution may be needed to create a binding surface with a suitable shape and charge constellation to bind tightly to transducin for the next step.

3. The Nerve Impulse

Which of the intermediates in Eq. 23-37 is responsible for initiation of a nerve impulse? Some evidence favored metarhodopsin I,[326] but its lifetime may be too short. On the other hand, the transformation of metarhodopsin I to metarhodopsin II is the slowest step that could trigger a nerve impulse, which must travel the length of the rod to the synapse in about one ms,[521] and metarhodopsin II is generally believed to be the activated signaling form of rhodopsin.[521a–c]

Transducin, cyclic GMP, and phosphodiesterase. The essential consequence of light absorption is an alteration in the membrane potential in the vicinity of the absorbed photon with the resulting propagation of a nerve impulse down the plasma membrane to the synapse by cable conduction (Chapter 30). The type of potential change that is transmitted differs among vertebrates and invertebrates.[522] In the case of mammalian photoreceptors the rod outer segment is permeable to sodium ions so that a large

dark current of sodium ions flows in through the plasma membrane and is pumped out by sodium pumps in the inner portion of the cell. Visual stimulation causes this permeability to Na^+ to be decreased with an increase in polarization of the membrane. Absorption of a single photon by rhodopsin blocks the outflow of ~10^6 sodium ions.

At one time calcium ions seemed to be the logical internal messenger between rhodopsin and the plasma membrane. If light absorption opened channels from the internal space of the rod discs, calcium ions could be released and diffuse quickly to the plasma membrane and block the entrance of sodium ions.[523] However, light *does not* increase the free [Ca^{2+}] in the cytoplasm but may decrease it from 500 nM to as low as 50 nM.[524,525] Stryer suggested that the essential messenger is **cyclic GMP** (Chapter 11)[526–528] and that a *decrease* in cGMP concentration initiates the nerve response. Cyclic GMP is apparently responsible for keeping the sodium ion channels open. Absorption of a photon in the rod disc (Fig. 23-43) produces activated rhodopsin R* (metarhodopsin II), which acts as an allosteric effector for the heterotrimeric G protein transducin whose structure and properties have been discussed in Chapter 11. Like proteins G_s and G_i of the adenylate cyclase system, transducin contains three subunits: α, 40 kDa, 350 residues; β, 36 kDa; and γ, ~8 kDa.[529] In the resting state they are associated as $T_{\alpha\beta\gamma}$ with a molecule of GDP bound to the α subunit. When activated rhodopsin R* binds to transducin (step *a*, Fig. 23-43) it catalyzes a rapid exchange of GTP for GDP (step *b*). This is followed by dissociation of $T_{\beta\gamma}$ from T_α GTP. The latter serves as an allosteric effector for a **cGMP phosphodiesterase** bound to the disc surface converting it to an active form (step *c*).[529a,b] The activated phosphodiesterase, an $\alpha\beta\gamma_2$ oligomer,[530] hydrolyzes the cGMP (step *d*, Fig. 23-43), reducing its concentration and thereby inhibiting the Na^+ outflow.

Because one molecule of activated cGMP phosphodiesterase can hydrolyze more than 10^5 molecules of cGMP per second the light response is highly amplified. There is also an earlier stage of amplification. Each molecule of light-activated rhodopsin (R*) is able to catalyze the exchange of GTP for GDP on hundreds of molecules of $T_{\alpha\beta\gamma}$ before R* passes on to other intermediates and releases all-*trans*-retinal from opsin (light green lines, Fig. 23-43).

Rhodopsin kinase, recoverin, and arrestin. Metarhodopsin II (R*) can become phosphorylated by rhodopsin kinase on as many as seven serine and threonine side chains on its cytoplasmic surface (Fig. 23-41).[531,532] The 45-kDa protein arrestin binds to such phosphorylated R*,[533–535a] which is rapidly deactivated and desensitized so that it is less likely to be immediately reactivated. This is important in the adaptation of the eye to bright light. At the same time T_α2GTP is hydrolyzed back to T_α2GDP and reforms $T_{\alpha\beta\gamma}$2GDP, and guanylate cyclase regenerates the cGMP.[536] At least four different arrestins are known. Some function in nonvisual tissues. In all cases they seem to serve as "uncouplers" of G protein-coupled receptors.[536a]

Recovery of the inhibited rhodopsin, which occurs most rapidly in dim light, depends upon calcium ions. In dim light both Ca^{2+} and Na^+ enter the visual cells through the cGMP-controlled channels. At the same time Ca^{2+} flows out through a Na^+/Ca^{2+} exchanger. When the channels are blocked by cGMP formed in

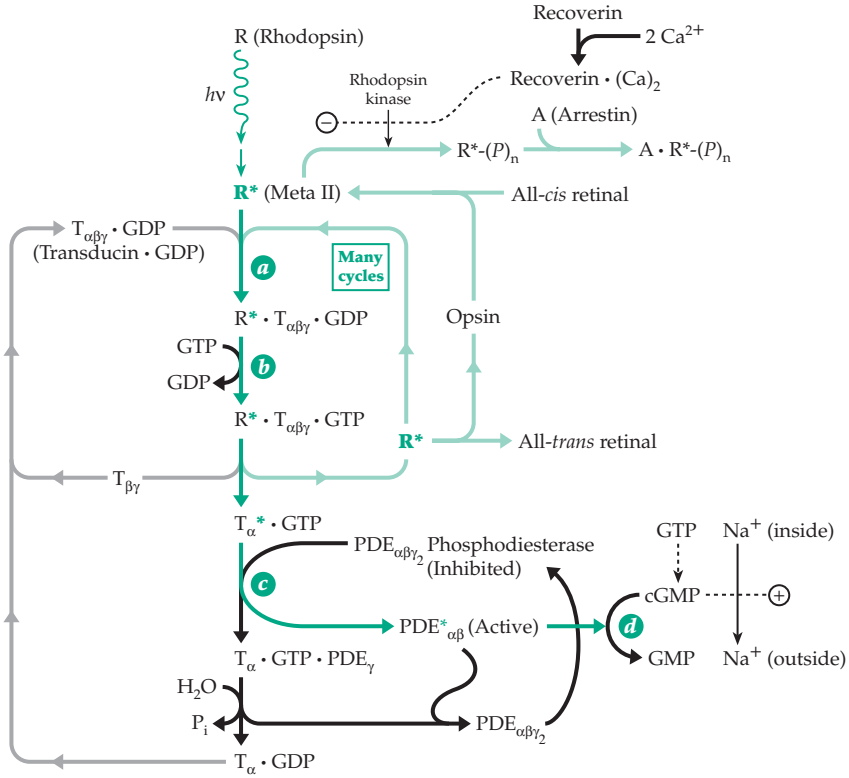


Figure 23-43 The light-activated transducin cycle. In step *a* photoexcited rhodopsin (R*) binds the GDP complex of the heterotrimeric transducin ($T_{\alpha\beta\gamma}$). After GDP–GTP exchange (step *b*) the activated transducin T*•GTP reacts with the inhibited phosphodiesterase ($PDE_{\alpha\beta\gamma2}$) to release the activated phosphodiesterase ($PDE_{\alpha\beta}$). Based on scheme by Stryer[528] and other information.

response to light, [Ca^{2+}] falls as a result of continuing export by exchange with Na^+.[524,537] In the dark the [Ca^{2+}] concentration rises again and binds to a 23-kDa calcium sensor molecule called recoverin. Recoverin, like calmodulin (Fig. 6-8), has four EF-hand Ca^{2+}-binding motifs and also an unexpected feature. A **myristoyl group** attached at the N terminus is bound into a deep hydrophobic pocket in the protein. However, when two Ca^{2+} ions bind, recoverin undergoes a conformational change that allows the myristoyl group to be extruded and to bind to a membrane surface. This allows recoverin to move out of the cytosol to the surface of the disc membrane where it binds to and inhibits the activity of rhodopsin kinase,[538–541] increasing the sensitivity of photodetection.

Some details about cone cells and invertebrate vision. The biochemistry of retinal cones is less well known but is similar to that of rod cells. Cone pigments are present in the plasma membrane rather than in isolated discs (Fig. 23-40C). Different α, β, and γ subunits of transducin are formed in rods and cones.[522] Many differences are seen among various invertebrate visual systems. Inositol triphosphate (IP_3) and Ca^{2+} often serve as signals of photoexcitation. G proteins also play prominent roles.[522]

4. Regeneration of Visual Pigments; the Retinal Cycle

How is all-*trans*-retinal released from photobleached pigments and isomerized to 11-*cis*-retinal for the regeneration of the photopigments? Since new 11-*cis*-retinol is continuously brought in from the bloodstream and oxidized to retinal, isomerization can occur in other parts of the body. However, much of it takes place in the **pigment epithelium** of the retina, the layer of cells immediately behind the rod and cone cells. As indicated in Fig. 23-44, all-*trans*-retinal can leave the photoreceptor cells and, after reduction to retinal, be carried to the pigment epithelial cells by an **interphotoreceptor retinoid-binding** protein.[542,543] There it becomes esterified by the action of **lecithin:retinol acyltransferase**, an enzyme that transfers a fatty acyl group from lecithin to the retinol. The resulting retinyl esters are isomerized, and 11-*cis*-retinol is released.[543a] Some is stored as 11-*cis*-retinyl esters but enough is dehydrogenated to 11-*cis*-retinal to meet the needs of the photoreceptor cells and is transported back to them (Fig. 23-44). In the cephalopods the inner segment of the receptor cells contain a second pigment **retinochrome** that carries out a photochemical conversion of all-*trans*-retinal to 11-*cis*-retinal.[544,544a]

5. Diseases of the Retina

An important cause of blindness is **retinitis pigmentosa**, an inherited disease affecting about one in 3000 persons. Symptoms include progressive night blindness, degeneration of the rod cells, and gradual loss of cone cells and of nerve function in the retina. An autosomal dominant form of the disease arises from deletions or point mutations in the rhodopsin gene. In the United States 15% of cases arise from the mutation P23H.[545] By 1996 ~70 point mutations that cause retinitis pigmentosa had been discovered.[448,546–548] These mutations are found in all three of the rhodopsin domains: intradiscal, transmembrane, and cytosolic. Other rhodopsin point mutations such as G90D and A292E cause congenital night blindness.[549] Retinitis pigmentosa also arises from defects in **peripherin-RDS**, a structural component of rod cells identified originally from the gene *rds* (retinal degeneration slow) of the mouse.[545,550] Another form of congenital night blindness results from mutations in rhodopsin kinase.[551] A dominant **rod–cone dystrophy** is caused by a defect in the photoreceptor guanylate cyclase.[552] The most frequent cause of combined deafness and blindness in adults (**Usher syndrome**) is a defect in a cell adhesion molecule.[553] In **choroideremia**, another X-linked form of retinitis pigmentosa,

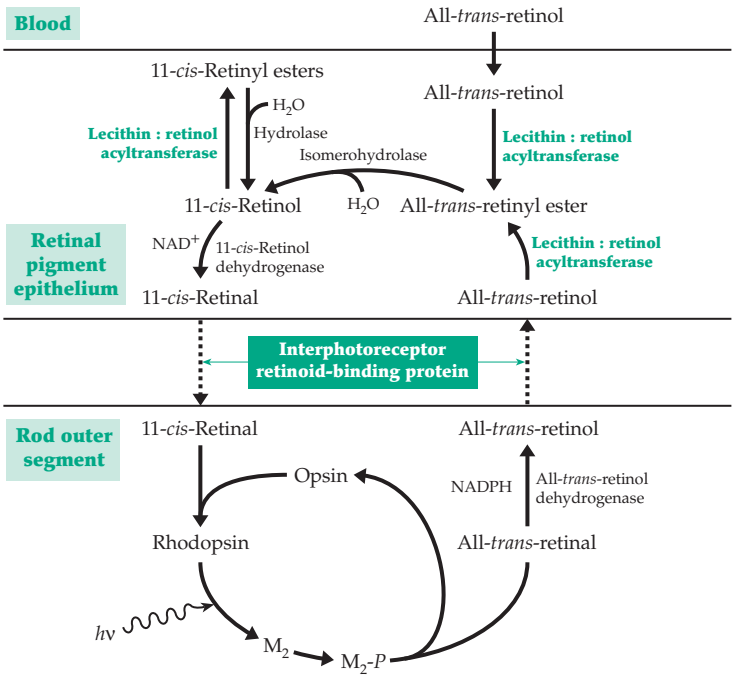


Figure 23-44 Reactions of retinol and the retinal cycle of mammalian rod cells. After Palczewski *et al.*[543]

the choroid layer behind the pigment epithelium also deteriorates. A geranylgeranyl transferase specific for the Rab family of G proteins is defective.[554]

The most frequent cause of vision loss in the elderly is **macular degeneration**. Mild forms of the disease occur in nearly 30% of those over 75 years of age and more serious forms in 7% of that age group. There are many causes, some hereditary.[555–556a] Excessive accumulation of fluorescent lipofuchsin, perhaps arising in part from Schiff base formation between retinal and phosphatidylethanolamine, is sometimes observed.[557]

6. Proteins of the Lens

The lens of the eye encloses cells that cannot be replaced and contains proteins that don't turn over and must last a lifetime.[558] The fiber cells, which make up the bulk of the lens, have no nuclei. They elongate and stretch to cover the central nucleus, the original fetal lens, like the layers of an onion, the edges of the cells interdigitated with the next cell like a piece in a child's construction set.[559] These cells are tightly packed with proteins in aggregates whose size is on the order of the wavelengths of light. The high concentration of proteins is needed to provide transparency and also a high refractive index.[560–562] The membranes of the lens cells acquire increasing amounts of a 28-kDa **major intrinsic protein** as they age.[563] Three classes of soluble lens proteins, called **crystallins**, are found in virtually all lenses. Alpha crystallins, which account for ~40% of the total soluble protein, are heterodimers of ~20-kDa subunits that associate into ~800-kDa complexes.[564] They have a chaperoninlike activity.[565] Beta crystallin, which may account for ~35% of the protein, as well as the γ crystallins are β-sheet proteins with "Greek key" folding motifs.[558]

In addition to the α, β, and γ crystallins many animals have recruited additional "taxon-specific" crystallins δ, ε, λ, etc., that have evolved from prexisting enzymes, chaperonins, or other proteins.[561,566–568] For example, avian and reptilian lenses contain a δ crystallin homologous to argininosuccinate lyase.[569] Many crystallins are derived from dehydrogenases, e.g., for lactate dehydrogenase (duck),[570] hydroxyacyl-CoA dehydrogenase (rabbit),[571] and aldehyde dehydrogenase (squid and octopus).[572] A high concentration of NADH may be present.[568] A crystallin of a diurnal gecko is a retinol-binding protein with bound 3-dehydroretinal (vitamin A_2), which probably acts as an ultraviolet filter that improves visual acuity and protects against ultraviolet damage.[567,572a] Human lenses contain small molecules that act as UV filters, e.g., glucosides of **3-hydroxykynurenine** (Fig. 25-11) and 4-(2-amino-3-hydroxyphenyl)-4-oxobutanoic acid. Lenses tend to discolor and become fluorescent with age, in part by irreversible reactions of crystallins with these compounds.[573–574a]

A common problem with lenses is **cataract**, a term that describes any loss of opacity or excessive coloration. There are many kinds of cataract, most of which develop in older persons.[560,562] Since lens proteins are so long-lived deamidation of some asparagine occurs. However, the reactions are slow. One of the asparagines in α crystalline has a half-life of 15–20 years, and some glutamines are undamaged after 60 years.[575]

G. Bacteriorhodopsin and Related Ion Pumps and Sensors

Under certain conditions, the salt-loving *Halobacterium salinarum* forms a rhodopsinlike protein, which it inserts into patches of **purple membrane** in the surface of the cell. These membranes, which may constitute up to 50% of the cell surface, contain light-operated proton pumps that translocate protons from the inside to the outside of the cells.[454,576–578] In this manner they may provide energy for a variety of cell functions including ion transport and ATP synthesis. The 248-residue retinal-containing **bacteriorhodopsin** makes up 75% of the mass of the membrane. Its molecules aggregate into a two-dimensional crystalline array in the purple patches of the membrane. This allowed determination of the three-dimensional structure to 0.7 nm resolution in 1975 by electron microscopy and neutron diffraction.[454] More recently the structure has been established at progressively higher resolution by electron crystallography[579–581] and X-ray diffraction.[455,582] The most recent studies have been focused on determination of the structural alterations in the protein that accompany the proton pumping.[582a–e] A step-by-step picture is emerging.[582f] Internal changes in the retinal chromophore, movements of protons, and alterations in the shapes of some of the protein helices are involved. The surface loops have been studied both by electron crystallography and by atomic force microscopy.[583] Each bacteriorhodopsin molecule is folded into seven closely packed α-helical segments which extend roughly perpendicular to the membrane. Although 100 residues shorter than rhodopsin, the folding pattern is very similar (Figs. 23-41C; 23-45). The protein molecules form an extremely regular array with phospholipid molecules (mostly of phosphatidylglycerol) filling the spaces between them. The retinal is buried in the interior of the protein and is bound as an *N*-protonated Schiff base with the side chain of lysine 216.

The retinal in bacteriorhodopsin (bR) exists in both all-*trans* and 13-*cis* configurations. The all-*trans* form has an absorption maximum at 568 nm and the 13-*cis*

at 548 nm. The two exist in a slow nonphotochemically mediated equilibrium in the dark.[584,584a] However, in the light the all-*trans* bR_{568} undergoes a rapid photochemical cycle of reactions, which is presented in simplified form in Fig. 23-46. The subscript numbers designate the wavelength of maximum absorption. Many efforts have been made to determine the structures of the intermediates K, L, M, and O and to relate them to a mechanism of proton pumping.[585–585c] Both K and L contain 13-*cis* retinal. Therefore, as with rhodopsin (Eq. 23-39) the very first step is a photochemical isomerization. Intermediate M appears to contain a deprotonated Schiff base, but O is both *N*-protonated and again all-*trans*. It follows logically that the proton bound initially to the Schiff base is pumped out of the cell and is then replaced by a new proton in the O form. However, several questions must be answered if we are to understand this proton pump. Where in the sequence do proton transfers occur? How is the sequence driven by the absorbed light energy? Protons must enter the pump from the cytosol and exit on the exterior (periplasmic) side, flowing out against a concentration gradient. There must be a "gating" or "switch" mechanism that ensures that protons enter and leave the pump in the correct direction.[586]

Some aspects of a possible mechanism for pumping the single proton bound to the retinal Schiff base are included in Fig. 23-46. In bR_{568} the Schiff base bound proton is H-bonded, via a water bridge, to the carboxylate group of Asp 85 (Fig. 23-45B). The charge constellation in the interior of the protein, part of which is shown in this figure, is such that protonation of the Schiff base is stable and the pK_a of the protonated Schiff base is high, with estimates of 16 or above.[587] One cause of the high pK_a is the presence of the nearby negative charges on D85 and D212. Absorption of light and isomerization of the retinal causes a downward movement of the $=NH^+-$ group of the Schiff base[588] and facilitates movement of the Schiff base proton via the water molecule to the D85 carboxylate as indicated by the green arrows in Fig. 23-45B. Loss of the positive charge will instantly substantially raise the pK_a of D85 from a low value, while the loss of the negative charge will lower the pK_a of the Schiff base to closer to 7. The electrostatic interactions of the D212 and E194 carboxylates with the positive charge of R82 may also be altered. At some point in the sequence the interaction of R82 with the E194 carboxylate could cause E204, which is known by spectroscopic measurements to be protonated in the intermediate, to lose its proton to the outside. At some other point, perhaps between M_{412} and another intermediate, M_{408} (Fig. 23-46), a conformational switch must occur to limit flow of a proton back to E204 and to allow a new proton to enter from the cytosol. The D96 carboxylate is thought to accept this proton and to transfer it via a chain of

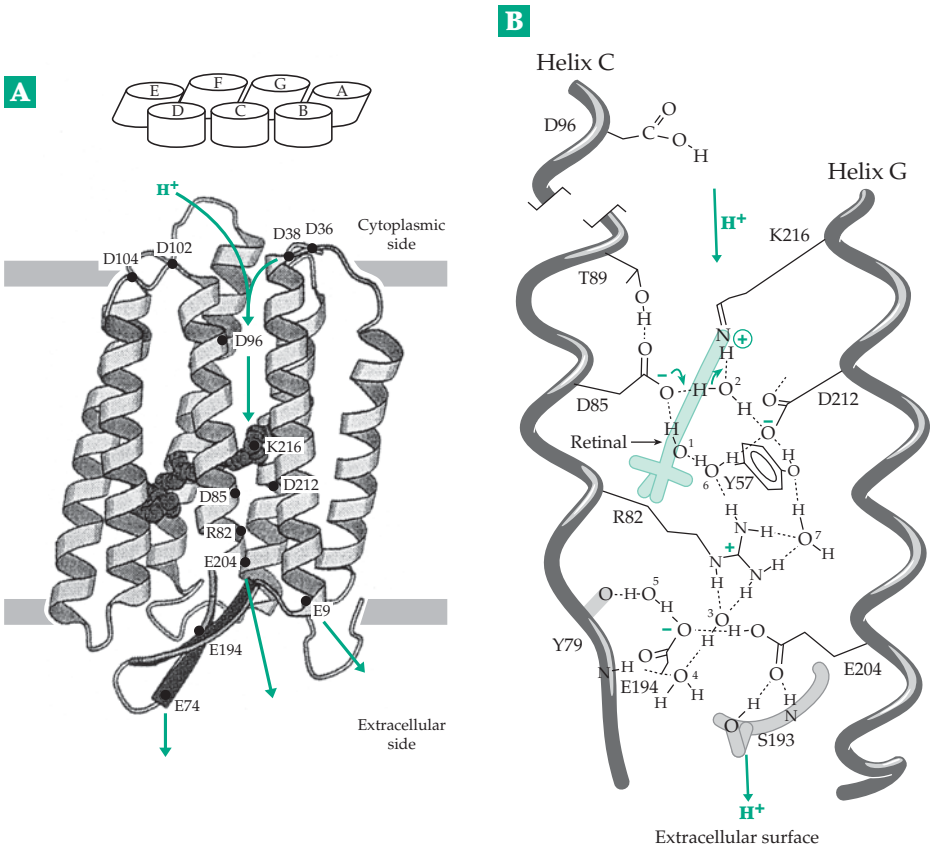


Figure 23-45 (A) Some aspects of the structure of bacteriorhodopsin. Ribbon diagram with the retinal Schiff base in ball-and-stick representation. At the top the helices are labeled as in Fig. 23-41. The locations of aspartate, glutamate, and arginine residues that might carry protons during the proton pumping action are indicated. Retinal is shown attached to lysine 216. From Kimura *et al.*[580] Courtesy of Yoshiaki Kimura. (B) Schematic drawing illustrating hydrogen-binding observed in the three-dimensional structure at 0.14 nm resolution. From Luecke *et al.*[455] The side chains shown are those thought to be involved in proton transport and in a hydrogen-bonded network with bound water molecules, principally between helices C and G. The positions of many of the hydrogen atoms in this network have not been established. They have been placed in reasonable positions in this figure but may be quite mobile. For another view of the hydrogen-bonded network see Fig. 23-41C.

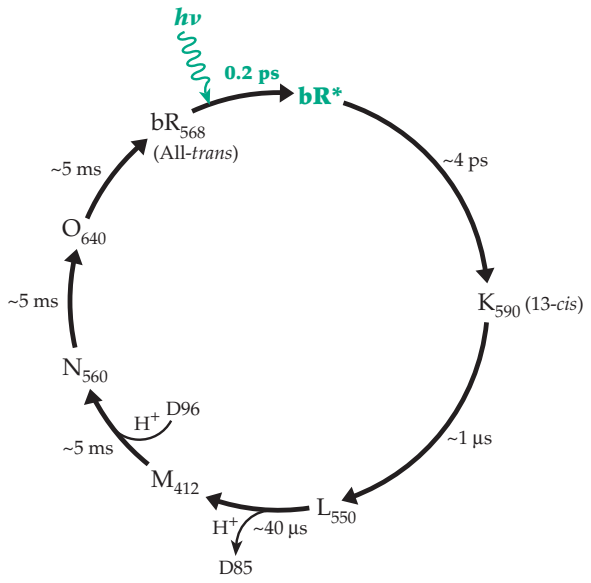
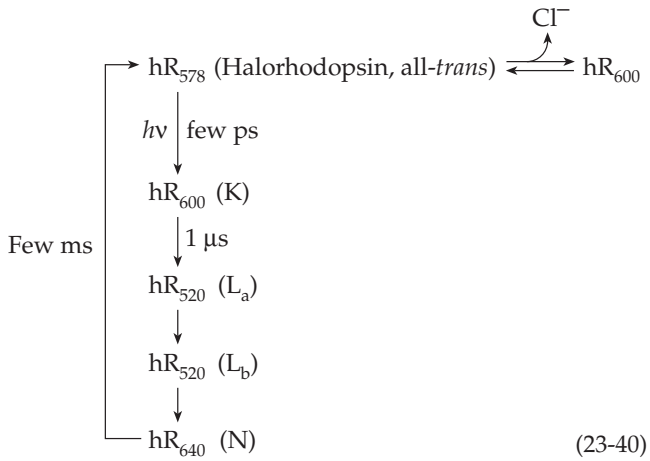


Figure 23-46 The photoreaction cycle of bacteriorhodopsin. After Bullough and Henderson.[585] The subscript numbers indicate the wavelengths of maximum absorption of each intermediate and the approximate lifetimes are given by the arrows. Resting bacteriorhodopsin as well as intermediates J and O have all-*trans* retinal but K through N are thought to all be 13-*cis*. A proton is transferred from L to aspartate 85 and then to the exterior surface of the membrane. A proton is taken up from the exterior surface via aspartate 96 to form N.

water molecules or water molecules plus protein groups to the Schiff base, which may isomerize back to all *trans* in going from intermediate N to O. Isomerization to all-*trans* O is a slow step. Both O_{640} and bR_{568} are all-*trans* and have a 15-*trans* Schiff base linkage. There may be a difference in protein conformation with the chromophore being more distorted in O than in bR_{568}. Many recent studies have provided additional information.[588a–j]

Halorhodopsin. In addition to bacteriorhodopsin there are three other retinal-containing proteins in membranes of halobacteria. From mutant strains lacking bacteriorhodopsin the second protein, **halorhodopsin**, has been isolated. It acts as a light-driven *chloride ion pump*, transporting Cl^- from outside to inside. Potassium ions follow, and the pump provides a means for these bacteria to accumulate KCl to balance the high external osmotic pressure of the environment in which they live.[578] The amino acid sequences of halorhodopsins from several species are very similar to those of bacteriorhodopsin as is the three-dimensional structure.[589] However, the important proton-carrying residues D85 and D96 of bacteriorhodopsin are replaced by threonine and alanine, respectively, in halorhodopsin.[590] Halorhodopsin (hR)



(23-40)

undergoes a light-dependent cycle (Eq. 23-40) that involves an all-*trans* to 13-*cis* photoisomerization with some intermediates resembling those of the bacteriorhodopsin cycle.[590a]

Sensory rhodopsins. The third and fourth light-sensitive proteins of halobacteria are **sensory rhodopsins** (SR)[578,591,591a,b] that are used by the bacteria to control **phototaxis**. These bacteria swim toward long-wavelength light, the maximum in the action spectrum being at ~580 nm. They are repelled by blue or ultraviolet light, the maximum in the action spectrum being at ~370 nm. Evidently the bacteria can detect either a decrease with time in red light intensity or an increase with time in blue light intensity. Either is interpreted as unfavorable and causes the bacteria to tumble and move in a new direction (see Chapter 19). Sensory rhodopsin I (SRI) appears to be able to provide both light responses. Absorption of orange light by SRI_{587}, which contains all-*trans*-retinal, yields SRI_{373}, in which the retinal Schiff bases have been isomerized to 13-*cis* as in bacteriorhodopsin. The red light response is proportional to the fraction of SRI_{373} present. However, this is converted spontaneously back to SR_{587} within seconds. Nevertheless, photoexcitation of SRI_{373} with blue light causes a faster reconversion and induces swimming reversals, the repellent response.[592,593] SRI exists in the bacterial membranes in a complex with a 57-kDa protein designated **halobacterial transducer I** (HtrI), which resembles bacterial chemotaxis receptors (Figs. 11-8 and 19-5) and is modulated by action of a methyltransferase.[591b,c] Interaction of SRI with HtrI depends upon a histidine residue, H166 of SRI. It may be part of a proton transfer pathway.[593]

Sensory rhodopsin II (SRII, also called phoborhodopsin) is specialized for repellant phototaxis.[591a] Blue light converts $SRII_{487}$ in < 1 ms to UV-absorbing $SRII_{360}$. It decays in ~100 ms to $SRII_{540}$ which reverts to the initial $SRII_{487}$ in ~0.5 s. The cycle is accompanied by swimming reversals that result in a repellent

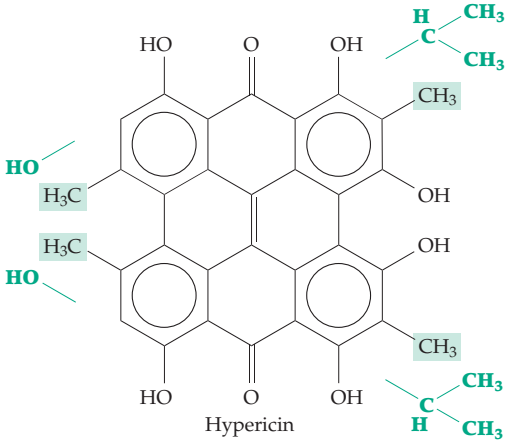
effect of light.[594–596] The three-dimensional structure is known.[593a,b]

Whereas retinal-based proton pumps all have the conserved residues D85 and D96 of bacteriorhodopsin, only the aspartate correspnding to D85 is conserved (as D73) in the sensory rhodopsins. D96 is replaced by tyrosine or phenylalanine.[586] In SRI D73 appears to be protonated and, therefore, does not form a counterion for the Schiff base iminium ion.[597] However, in SRII D73 is apparently unprotonated and available to -serve as a counterion and as a proton acceptor as in bacteriorhodopsin.[597a,b] There is also a corresponding aspartate (Asp 83) in rhodopsin (Fig. 23-41). This suggests a common signaling mechanism for rhodopsin and the sensory rhodopsins. Finally, there are retinal-containing proteins in fungi and in algae. They may serve as blue light receptors.[598,598a]

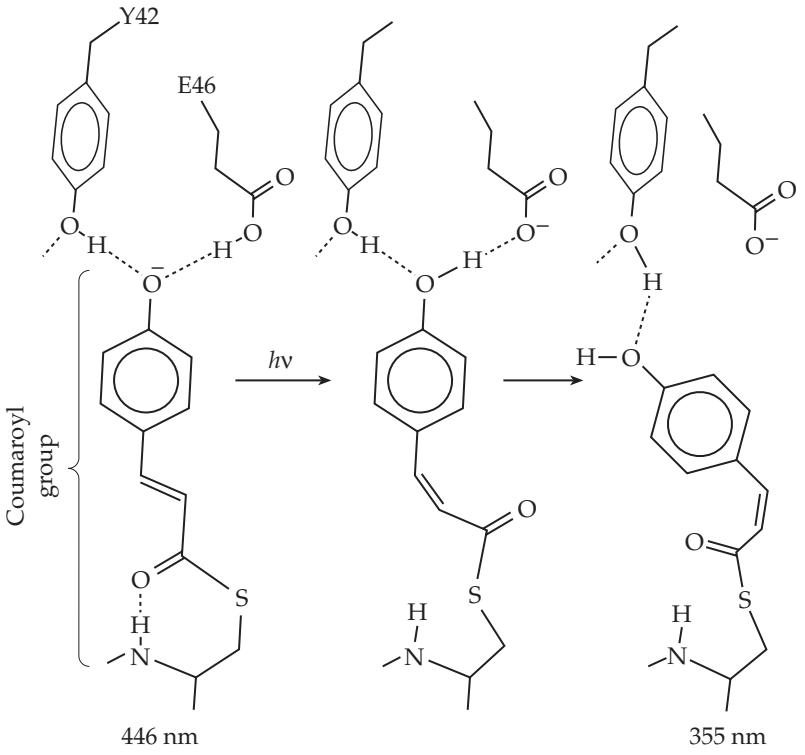
The photoactive yellow proteins (xanthopsins). A 124-residue, 14-kDa yellow protein isolated from a halophilic phototrophic purple bacterium, *Ectothiorhodospira halophila*, was at first thought to be a rhodopsin-type pigment. However, this photoactive yellow protein (PYP) carries a covalently bonded **coumaroyl (4-hydroxycinnamoyl) group** in thioester linkage,[599–601c] which is completely surrounded by the small protein.[601] The coumaroyl group, which was probably derived from coumaroyl-CoA (Fig. 25-8), is bound as a phenolate anion by hydrogen bonds to tyrosine and glutamate side chains (Eq. 23-41). After a laser flash at room temperature a readily observed intermediate I_1 (also called pR) absorbing maximally at 460–465 nm appears within ~3 ns and decays within a few milliseconds to a bleached intermediate I_2 (also called pB or pM)[602,603] with maximum absorption at ~355 nm. This returns to the original 446-nm form within a few seconds. Earlier intermediates I_0 and $I_0^{\ddagger}$ have been identified by picosecond spectroscopy,[601c,604,605] and others have been identified at low temperatures.[602,606]

The structure of PYP is known to 0.1 nm resolution (Fig. 23-47).[601,607] Structures have also been determined for a very early intermediate by trapping at –100°C[607] and for I_1 (pR). The cofactor structures are shown in Eq. 23-41. The light-induced step is apparently the *cis–trans* isomerization,[608] and changes in hydrogen-bonding follow. The hydrogen bond between the phenolate ion of the coumaroyl group and glutamate 46 appears to break, and E46 may donate a proton to the phenolate group to form the 337-nm chromophore of I_2.[602,609,609a] The signaling mechanism may be similar to that in sensory rhodopsins.

Stentorin. A protein with a bound chromophore called stentorin mediates the light-avoidance response of the protozoan *Stentor*. Stentorin,[610] which is found in pigment granules in the cell surface, is a derivative of **hypericin**, a plant compound with antidepressant activity and the active ingredient in the herb St. John's wort.



Stentorin is covalently bonded to a 16-kDa protein in an acid-labile linkage. Its photocycle is not well investigated, but it is thought to initiate a response via



(23-41)

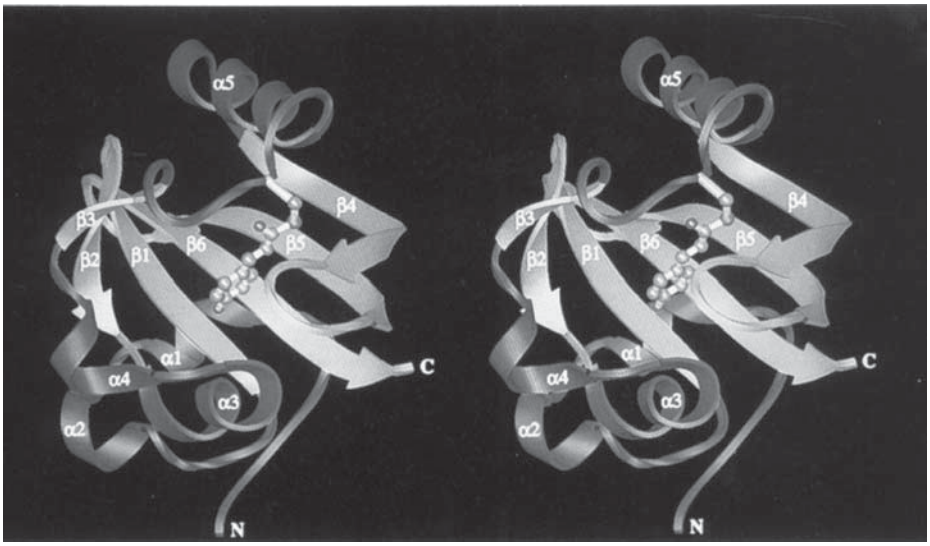


Figure 23-47 Ribbon drawing of the structure of the 125-residue yellow photoactive protein. The 4-hydroxycinnamoyl chromophore, which is attached to cysteine 69, is represented with balls and sticks. From Borgstahl *et al.*[601] Courtesy of Gloria Borgstahl.

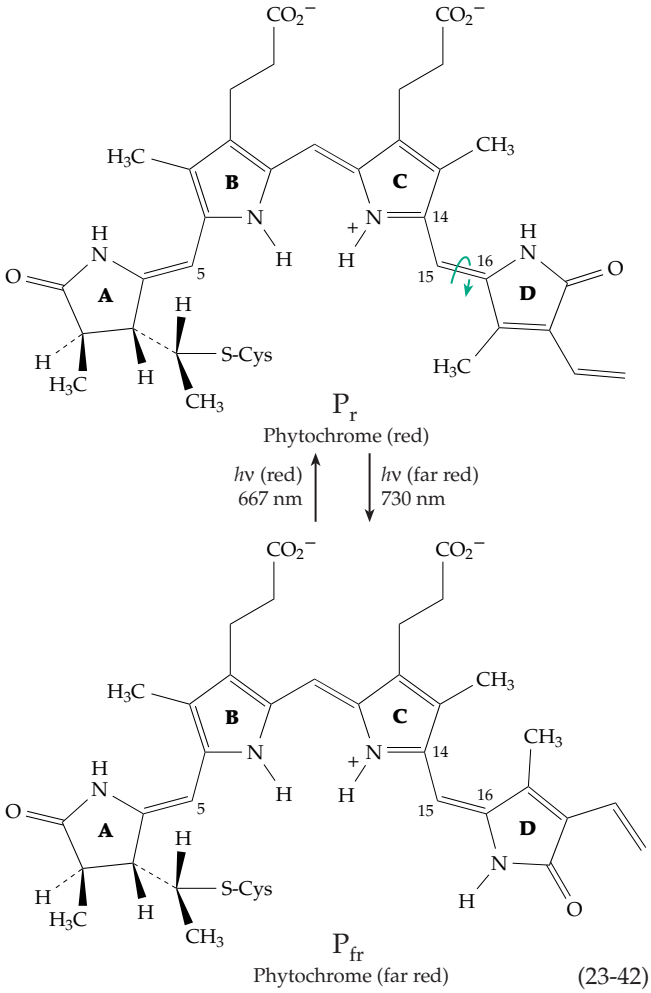
proton transfer.[611] However, stentorin proteins apparently do not belong to the bacteriorhodopsin family.

Hypericin and related compounds have also aroused interest because of their antiviral and antitumor activities.[612,613] Hypericin is a strong photoactivator which produces singlet oxygen with a quantum efficiency of 0.73. However, antiviral activity may involve a radical mechanism.[613] Hypericin is attractive as a possible agent for photodynamic therapy (Section D,3). It can also receive energy from photoexcited firefly luciferin (Section J). A proposed application is to incorporate the gene for the enzyme luciferase of the firefly luminescence system into DNA from the virus HIV. This DNA could be used to promote synthesis of luciferase only in virus-infected cells. Addition of the nontoxic hypericin would lead to photoactivation of hypericin only in virus-infected cells, where the luciferin–luciferase complex would act as a "molecular flashlight" to activate the hypericin and destroy the cell.[613]

H. Phytochrome

In 1951, it was discovered that a flash of red light (maximum activity at 660 nm) during an otherwise dark period promoted a variety of responses in plants.[614] These included flowering, germination of seeds (e.g., those of lettuce), and the expansion of leaves in dark-grown pea seedlings. Interestingly, the effect of the short flash of red light could be *completely reversed* if followed by a flash of *far-red light* (730 nm). This discovery led to the isolation, in 1959, of the chromoprotein phytochrome, a kind of molecular switch that initiates a whole series of far-reaching effects in plants. The phototransformation[615] is completely reversible (Eq. 23-42; Fig. 23-48), and the switch can be thrown in one direction or the other many times in rapid succession by light flashes.

Green plants have a family of phytochromes. There are five genes for the ~125-kDa chains of about 1100 residues each in *Arabidopsis*,[618–619c] and the corres-



(23-42)

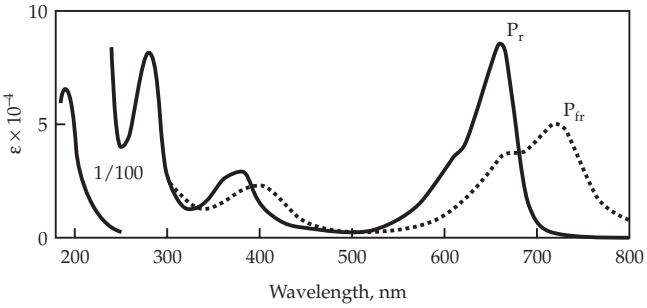


Figure 23-48 Absorption spectra of red (P_r) and far-red (P_{fr}) forms of purified oat phytochrome following saturating irradiations with red and far-red light. See Quail[616] and Anderson *et al*.[617]

sponding phytochromes A – E each appear to have distinct functions. The chromophore is an open tetrapyrrole closely related to phycocyanobilin and covalently attached to the peptide backbone near the N terminus through a cysteine side chain (top structure in Eq. 23-42). The initial photochemical reaction is thought to be the $Z \rightarrow E$ isomerizion around the C15 – C16 double bond, but there may also be rotation about the C14 – C15 single bond. The initial step occurs within a few microseconds and up to four intermediate species have been seen in the $P_r \rightarrow P_{fr}$ conversion and at least two other different ones in the $P_{fr} \rightarrow P_r$ conversion.[620–623c]

Phytochromes exist as two distinct domains, the N-terminal domain bearing the chromophore. However, the first 65 residues at the N terminus as well as the C-terminal domain probably interact with other proteins to transmit a signal. The slow responses to phytochrome are thought to involve regulation of transcription.[619b,623d] Thus, the synthesis of mRNA molecules specific for the small subunit of ribulose bisphosphate carboxylase and for the chlorophyll *a*/*b* binding protein increases in response to formation of P_{fr}. These responses are quite rapid occurring within 15–30 min.[614] Another response to P_{fr} is a decrease in the amount of the specific mRNA for phytochrome itself. That is, light induces a decrease in the concentration of this light-sensing molecule thereby decreasing the sensitivity of the system.[614]

Phytochrome is found not only in higher plants but also in algae, where it controls the movement of chloroplasts,[611] and also in cyanobacteria.[623e,f] Cyanobacterial phytochromes contain histidine kinase domains, which may function in a two-component system with a response regulator similar to protein CheY of the chemotaxis system in *E. coli* (Fig. 19-5).[624,625] Some nonphotosynthetic bacteria also use bacteriophytochromes for light sensing. In some cases biliverdin (Fig. 24-24) is the chromophore.[625a]

Phytochromes of higher plants also have histidine kinase-like N-terminal domains. Searches for associated signaling proteins have revealed a phytochrome-interacting factor in *Arabidopsis*. A possible partner for phytochrome B, it is a nuclear helix–loop–helix protein that may be a transcription factor.[626–627a] Phytochrome A may signal via a WD-repeat protein to control morphogenesis.[628]

One response under phytochrome control is the closing of leaflets of *Mimosa* at the onset of darkness. The response occurs within 5 min, too short a time to be the result of transcriptional control. This and the finding that some phytochrome is tightly bound to membranes have led to the proposal that one primary effect of phytochrome is to alter membrane properties. It is not certain whether it is P_r or P_{fr} that is active in causing a response, but P_{fr} seems to be the most likely candidate for the "active" form. According to one suggestion, phytochrome in plastid membranes may mediate the release of gibberelins stored within the plastids.[629]

I. Some Blue Light Responses

Numerous biological responses to light of wavelength 400–500 nm are known. These include phototropism in higher plants, the phototaxis of *Euglena*, and photorepair of DNA. On the basis of action spectra both carotenoids and flavins were long ago proposed as photoreceptors.[630–632] The action spectrum for opening of the stomates in coleoptiles matches the absorption spectrum of zeaxanthin.[631] On the other hand, genetic evidence[633,634] has strengthened the view that a flavin acts as the photoreceptor in the fungus *Phycomyces*. Recently compelling evidence for a flavoprotein receptor for phototropism in *Arabidopsis thaliana* has been obtained. Deficiency of a gene called *nph*1 (nonphototropic hypocotyl 1) is associated with loss of blue light-dependent phosphorylation of a 120-kDa protein. This protein was identified as the product of *nph*1 gene. The **nph1** protein is a soluble autophosphorylating Ser / Thr protein kinase with an N-terminal flavin-binding region. It apparently binds FMN and is a photoreceptor for phototropism in higher plants.[635–636b]

The complexity of the action spectra suggested the existence of more than one receptor.[634] In higher plants there are not only blue light receptors but also violet receptors and phytochrome. In addition to nph1 and a related protein **npl1**, *Arabidopsis* employs two cryptochromes (next section) and **phototropins**.[636c–e] These are also riboflavin 5'-phosphate (FMN)-dependent proteins. The action of light apparently causes addition of a highly conserved cysteinyl –SH group to the C4a position of the flavin.[636c] Phytochrome absorbs blue and ultraviolet light to some extent (Fig. 23-48) as well as red or far-red. This adds

considerable complexity to the interpretation of light responses in plants.[637] The fern *Adiantum* contains a protein with a phytochrome photosensory domain fused to an NPH1 structure. It may mediate both real far-red and blue-light responses.[636] The protist *Euglena* (Fig. 1-9) makes use of a **photoactivated adenylate cyclase**, also a photoactivated enzyme, in a light avoidance response.[637a]

Fungi such as *Neurospora crassa* provide a simpler system for study of blue-light signaling than do green plants.[637] *Neurospora* contains no phytochromes. However, numerous genes including some involved in carotenoid biosynthesis and some that control the circadian cycle are regulated by blue light. Two mutants defective in riboflavin synthesis show a reduced sensitivity to blue light. A deficiency of either of two other genes *wc*-1 and *wc*-2 results in "blind" *Neurospora* unable to respond to light but able to grow. Proteins WC1 and WC2 are probably transcription factors, which act as a heterodimer. WC1, which contains bound FAD, is the photorecetor.[637,637b,c] Recently a rhodopsinlike protein NOP-1 of *Neurospora* has been identified.[637d]

1. Cryptochromes

The elusive nature of the principal blue-light receptor in plants gave rise to the name cryptochrome.[632] The gene for a cryptochrome in *Arabidopsis thaliana* was isolated by gene tagging and was cloned. It is surprisingly similar in sequence to the gene for the well-known **DNA photolyase** (Section 2).[638] It was soon recognized that cryptochromes, like photolyases, carry a bound flavin and also an antenna chromophore. The latter is probably a 5,10-methylenetetrahydrofolate, as in plant photolyases. It is a better light absorber than the flavin and passes electronic excitation to the flavin.

Two cryptochrome genes, *cry*1 and *cry*2, are present in *Arabidopsis*. The encoded proteins affect many aspects of plant growth. The cry-1 protein, together with NPH1, has a role in controlling phototropisin[639] while cryptochrome cry-2 affects flowering time, apparently via antagonistic signals from cry-2 and phytochrome B.[640–640b] cry-1 is also involved in controlling the daily rhythm of the plant, the **circadian cycle**. The circadian clock, which is discussed in Chapter 30, provides the organisms an oscillator with a period of about 24 hours. However, the oscillator must be **entrained** by the daylight cycle so that it remains in proper synchrony. The nature of the light signal and the mechanism of the entrainment are being investigated in many different organisms from fungi to human beings. In *Arabidopsis* the cycle is controlled by phytochromes A and B and by the cryptochrome cry-1.[641]

Cryptochrome genes have been found in many organisms. In the fly *Drosophila* cryptochrome appears to interact directly with the clock proteins that control the circadian cycle. Most important are products of two genes *per* (period) and *tim* (timeless). They are helix–loop–helix DNA binding proteins that form heterodimers, are translocated to the nucleus, and repress their own transcription. Morning light leads to a rapid disappearance of the TIM protein. The cryptochrome CRY appears to react directly with TIM to inactivate it. However, details remain to be learned.[642] The circadian clock mechanism appears to be universal and the cryptochrome-2 (*mcry*2 gene) appears to function in the mouse.[643,643a] A human cDNA clone was found to have a 48% identity with a relative of cryptochromes, the **(6–4) photolyase** of *Drosophila*. A second related human gene has also been found. The protein products of these two genes (*hcry*1, *hcry*2) lack photolyase activity. They too may encode cryptochromes.[644]

Where in the body is the light sensed for entraining the circadian cycle? Genes for CRY1 and CRY2 are specifically expressed in ganglion cells of the retina in mice. Severing of the optic nerve destroys both vision and light entrainment of mammals. However, in mice with the retinal degeneration (*rd*) syndrome all rod cells and virtually all cone cells are destroyed but the circadian rhythm is normal.[645] Furthermore, many blind persons with no conscious perception of light have normal light entrainment of their circadian cycle. For these reasons the ganglion cells of the retina, which are close to the location of the master circadian clock in the **suprachiasmatic nucleus** of the brain, are the most probable light sensory cells for the cycle[646,647] (see also Chapter 30). Recent evidence points to a retinal-based photoreceptor, **melanopsin**.[647a,b,c] However, vitamin A-deficient mice still display a normal circadian response.[647d]

2. Photolyases

A curious discovery was made many years ago. Bacteria given a lethal dose of ultraviolet radiation can often be saved by irradiating with visible or near ultraviolet light. This **photoreactivation**, which permits many bacteria to survive, results from the action of a **DNA photolyase**,[648,649] which often absorbs light maximally around 380 nm and carries out a photochemical reversal of Eq. 23-26, cutting the pyrimidine–pyrimidine covalent bonds of thymine dimers in DNA. The enzyme is present in cells in such small amounts, only 10–20 molecules per cell, that it was difficult to investigate until the gene had been cloned.[650,651] The significance cannot be doubted, for photoreactivation enzymes appear to be found in most organisms including some mammals. However, there

is some doubt about the presence of a photolyase in the human body.

The *E. coli* DNA photolyase contains a blue flavin radical that arises from **FAD** and absorbs maximally at 580 nm (see also Chapter 15, Section B,6). The enzyme also contains a second chromophore in the form of bound 5,10-methenyltetrahydrofolylpolyglutamate with 3–6 γ-glutamyl residues.[652–653b] as shown in Fig. 23-49. The pterin coenzyme binds near the N terminus in a domain with an α/β folding pattern, while the FAD binds into a larger mostly helical domain. The pterin cofactor is not essential for repair activity, and it is generally agreed that because of its high molar extinction coefficient it acts as an effective **antenna**. It transfers energy in a nonradiative fashion to the $FADH^-$ anion located ~3 nm away.

The enzyme as isolated is in a stable blue radical form (Fig. 23-50; also Fig. 15-13) which must undergo a one-electron light-induced reduction to the anion $FADH^-$ before becoming active. A nearby indole ring of Trp may donate the electron and be reoxidized by a tyrosyl ring.[654] The $FADH^-$ donates an electron to the pyrimidine dimer, initiating the sequence of radical reactions[654a–c] which cleaves both pyrimidine–pyrimidine bonds in the photodimers (Fig. 23-50).

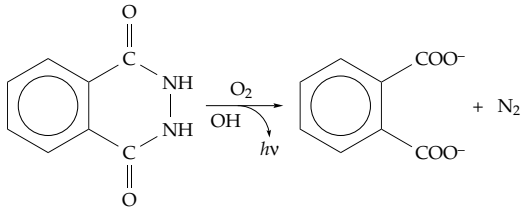
Figure 23-49 Overall view of the DNA photolyase structure from *E. coli*. The ribbon traces the 471-residue chain. The bound cofactors FAD (left) and 5,10-methenyltetrahydrofolate (right) are shown in ball-and-stick representation. From Park *et al.*[652] Courtesy of Johan Deisenhofer.

The structures of all of the photolyases are thought to resemble that in Fig. 23-49. However, in one large group, which includes methanogenic bacteria, **8-hydroxy-5-deazariboflavin** acts as the antenna chromophore.[649] Another light-induced defect in DNA is the so-called 6–4 photoproduct, a different pyrimidine dimer. The 6–4 dimers are normally removed in most organisms by efficient **excision repair** (Chapter 27). However, a 6–4 photolyase was discovered in both *Arabidopsis* and *Drosophila* and has also been found in *Xenopus* and the rattlesnake.[655,655a] It has a structure similar to that of the *E. coli* photolyase and presumably acts by a related mechanism[191,656–657a] that uses the light-excited reduced flavin in an electron donation and return cycle as in Fig. 23-50. A homolog of the *Drosophila* 6–4 photolyase gene has been found in human cells, but there is uncertainty about its function.[658] Is it really a photolyase or is it a cryptochrome involved in the circadian cycle?

J. Bioluminescence

The emission of visible light by living beings is one of the most fascinating of natural phenomena. Luminescent bacteria, glowing toadstools, protozoa that can light up ocean waves, luminous clams, fantastically illuminated railroad worms,[659] and fireflies[660–661a] have all been the objects of the biochemists' curiosity.[662–664] The chemical problem is an interesting one. The firefly's light with a wavelength of 560 nm (17,900 cm^{-1}) has an energy of 214 kJ/einstein. What kind of chemical reaction can lead to an energy yield that high? It is far too great to be provided by the splitting of ATP. Even the oxidation of NADH by oxygen would barely provide the necessary energy.

A clue comes from the fact that chemiluminescence is very common when O_2 is used as an oxidant in nonenzymatic processes. The slow oxidation of alcohols, aldehydes, and many nitrogen compounds (Eqs. 23-43, 23-44) is accompanied by emission of light



Luminol, a synthetic luminescent compound

(23-43)

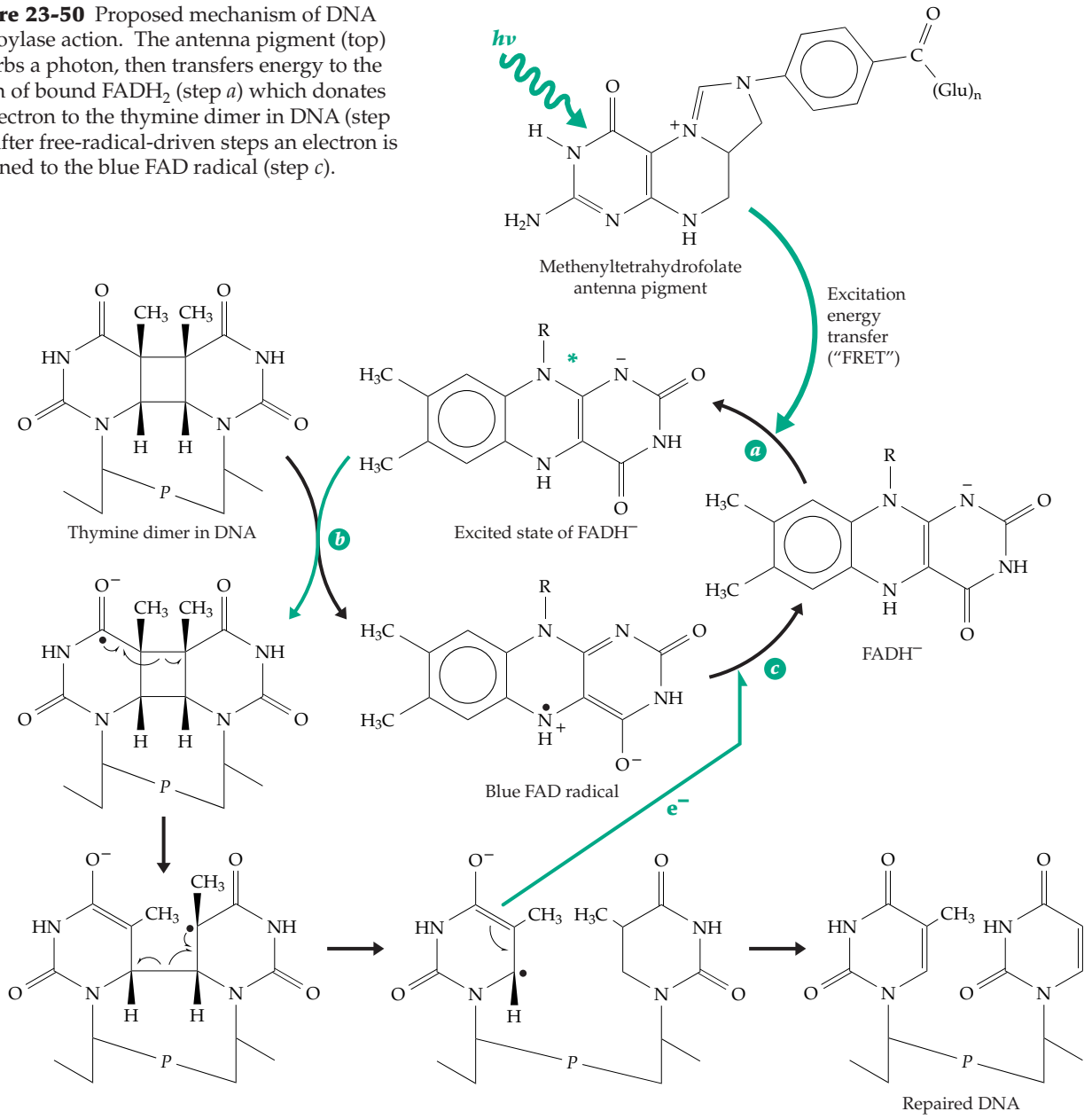
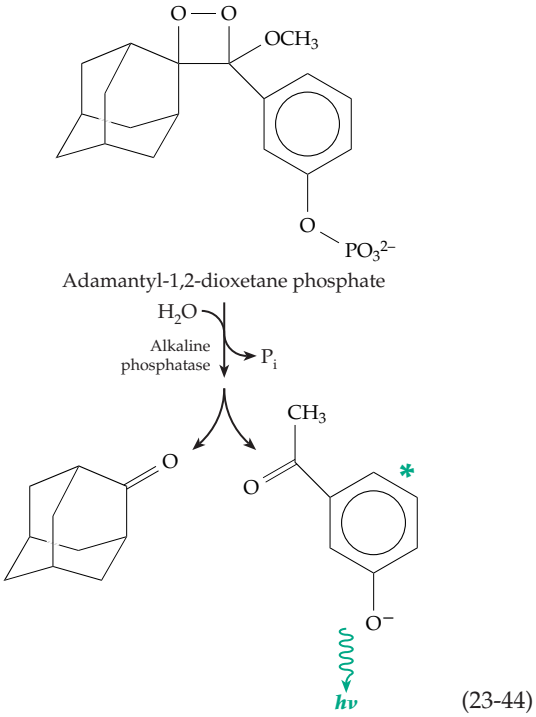


Figure 23-50 Proposed mechanism of DNA photolyase action. The antenna pigment (top) absorbs a photon, then transfers energy to the anion of bound $FADH_2$ (step *a*) which donates an electron to the thymine dimer in DNA (step *b*). After free-radical-driven steps an electron is returned to the blue FAD radical (step *c*).

visible to the eye. Chemiluminescence is especially pronounced in those reactions that are thought to occur by radical mechanisms. The recombination of free radicals provides enough energy to permit the release of visible light. Cleavage of a peroxide linkage, e.g., in a dioxetane (Eq. 23-44),[665] is often involved.[666] For example, the reaction of Eq. 23-44 is used in a sensitive light-detected assay for alkaline phosphatase.

In view of these facts it is perhaps not so surprising that many organisms have mastered the ability to channel the energy released in an oxygenation reaction into light emission. Attempts to extract luminous materials from organisms date from the last century when the French physiologist, DuBois, in 1887 prepared both a cold-water extract and a hot-water extract of luminous clams.[662] He showed that the material in the cold-water extract, which he named **luciferase**, caused emission of light when a heat-stable material (which he called **luciferin**) present in hot-water extract was added. These names have been retained and are now used in a general way. Thus, the luciferins are a family of compounds whose structures have been determined for a number of species (Fig. 23-51). Firefly luciferin is a carboxylic acid, but it must be activated in an ATP-requiring reaction to give **luciferyl adenylate**, whose structure is shown in the figure. The latter emits light in the presence of O_2 and luciferase. It can be seen that the original carboxyl group becomes CO_2, while the ring becomes oxidized. In addition, the acyl adenylate linkage is broken. In

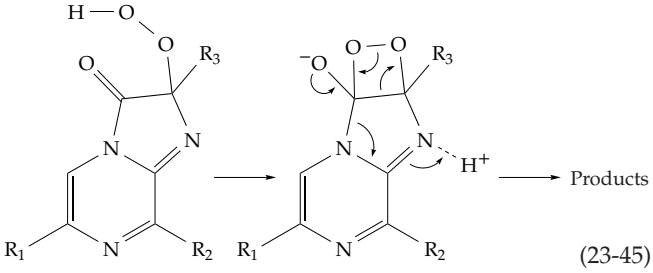


(23-44)

the "sea pansy" *Renilla reniformis* (a coelenterate) the luciferin has quite a different structure.[667,668] However, the reaction with O_2 to produce CO_2 and an oxidized product causes the light emission, just as in the firefly. The luciferin of *Renilla*, which is called **coelenterazine** (Fig. 23-51), is also found in the jellyfish *Aequorea*, the shrimp *Oplaphorus*,[669] the "firefly squid" *Watasenia scintillans*,[670] and other luminous organisms.

In *Renilla* the coelenterazine is stored as a coelenterazine sulfate, possibly having the structure shown. To convert this storage form to the active luciferin the sulfo group is transferred onto adenosine 3',5'-bisphosphate to form 3'-phosphoadenosine 5'-phosphosulfate, the reverse of step *d* of Eq. 17-38. The luciferin of the ostracod crustacean *Vargula hilgendorfii* has a structure (Fig. 23-51) close to that from *Renilla*. In *Vargula* (formerly *Cypridina*) the luciferin and luciferase are produced in separate glands and are secreted into the surrounding water where they mix and produce a bright cloud of light.[671]

In most mechanisms suggested for luciferase action O_2 reacts at the carbon atom that becomes the carbonyl group in the product to form an intermediate peroxide. In the case of *Renilla* luciferin this can easily be visualized as a result of flow of electrons (perhaps one at a time) from the pyrazine nitrogen (at the bottom of the structure in Fig. 23-51) into the O_2. According to one proposal, the peroxide group that is formed adds to the carbonyl to form a four-membered dioxetane ring as shown in Eq. 23-45 for coelenterazine peroxide. The latter opens, as indicated by the arrows, to give the products. This theory was tested using



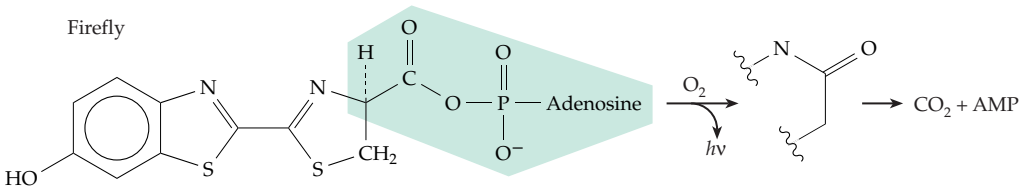
(23-45)

$^{18}O_2$. In the case of *Vargula* luciferin the expected incorporation of one atom of ^{18}O into CO_2 was observed, but with firefly and *Renilla* luciferins no ^{18}O entered the CO_2. In these two cases, a somewhat different mechanism may hold.

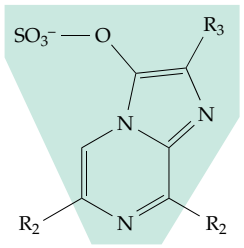
The jellyfish *Aequorea* contains a **photoprotein**, which emits light only when calcium ions are present.[672,673] Since light emission can be measured with great sensitivity (modern photomultipliers can be used to count light quanta) the protein **aequorin** and related photoproteins[674a] are used as a sensitive indicator of calcium ion concentration.[674] (In a similar way the firefly luciferin–luciferase system, which requires ATP for activation, is widely used in an assay for ATP.)

To identify the chromophore in aequorin over 4000 kg of jellyfish were used to obtain 125 mg of electrophoretically pure photoprotein.[674] From this one mg of a chromophoric substance AF-350 (Fig. 23-51) was isolated and characterized as a product. The close relationship to the *Renilla* and *Vargula* luciferins is obvious, and it is thought that coelenterazine is present in aequorin and other photoproteins as a peroxide (as in Eq. 23-45). For this reason no additional oxygen is needed to complete the reaction when Ca^{2+} acts to alter the conformation of the protein.[675–676a] The structure of a photoactive intermediate from the coelenterazine-containing protein obelin (from *Obelia longissima*), however, shows only one oxygen atom attached to C2 (Fig. 23-51)[676a] Although coelenterazine is utilized by many cnidarians they apparently cannot synthesize the compound but must obtain it through their diet. The source of biosynthesis is unknown.[676b] Some dinoflagellates emit light from a 137-kDa luciferin that contains three homologous domains each of which binds a molecule of a tetrapyrrole.[676c] From its structure the latter appears to have arisen from chlorophyll (Fig. 23-20), whose ring has been opened to give a structure somewhat similar to that of phytochrome (Fig. 23-23).[676d]

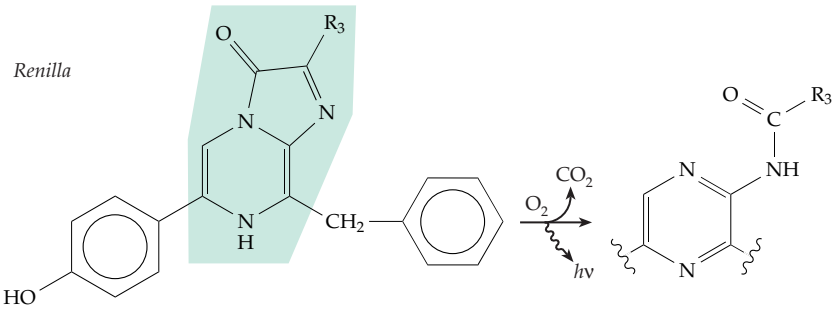
The first step in the formation of light in the firefly is a reaction with ATP to form luciferyl adenylate (Eq. 23-46, step *a*).[676e] The proton on the carbon may then be removed making use of the electron accepting properties of the adjacent ring system and carbonyl group before addition of the O_2. The reactions should be compared to those catalyzed by oxygenases, e.g., Eq. 18-42. The large 62-kDa firefly luciferase has a

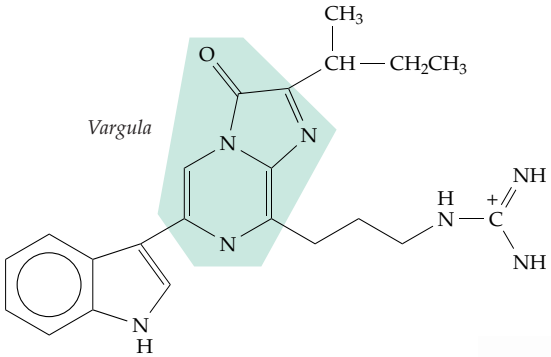


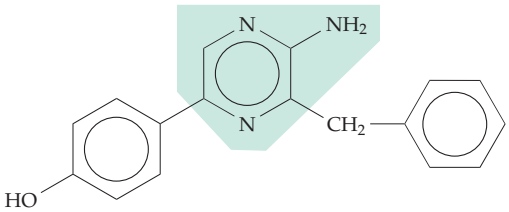
Luciferyl adenylate



Luciferyl sulfate of *Renilla*

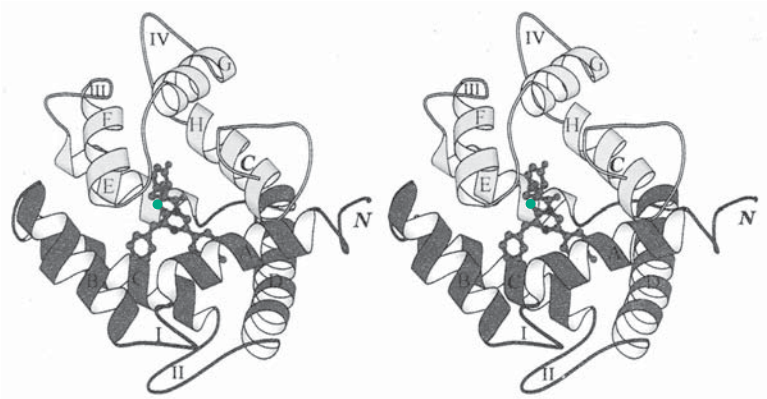






Compound AF-350 from *Aequorea*

Photoprotein **obelin** from *obelia* showing an oxy complex of bound coelenterazina. The oxygen atom is green. From Liu *et al.*[676a]



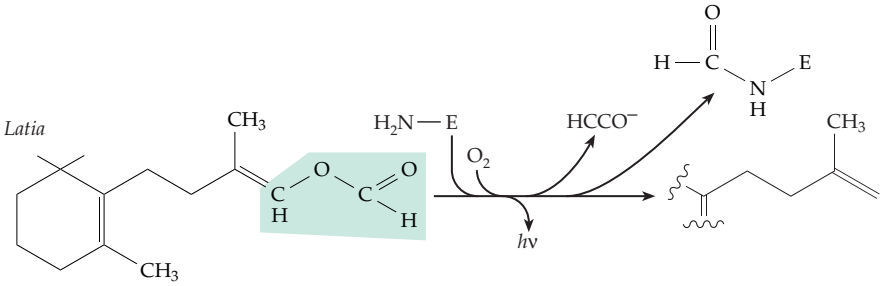
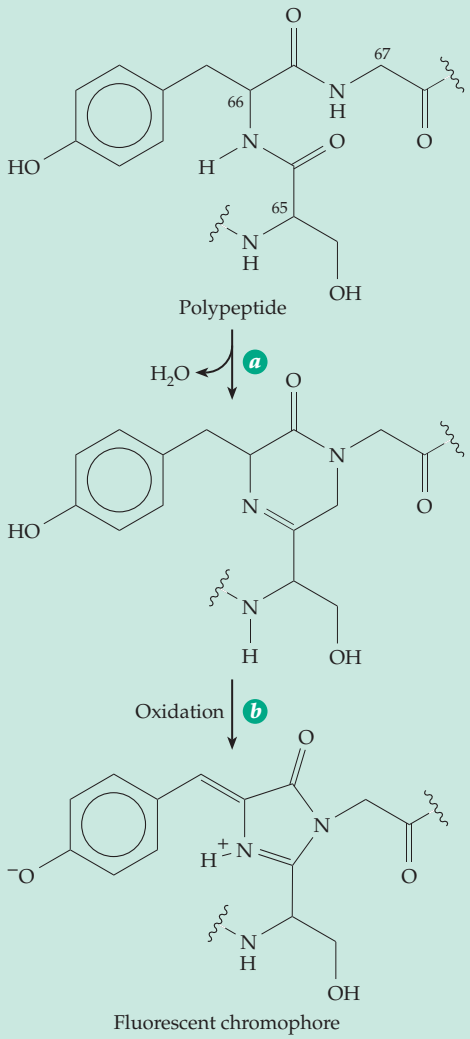


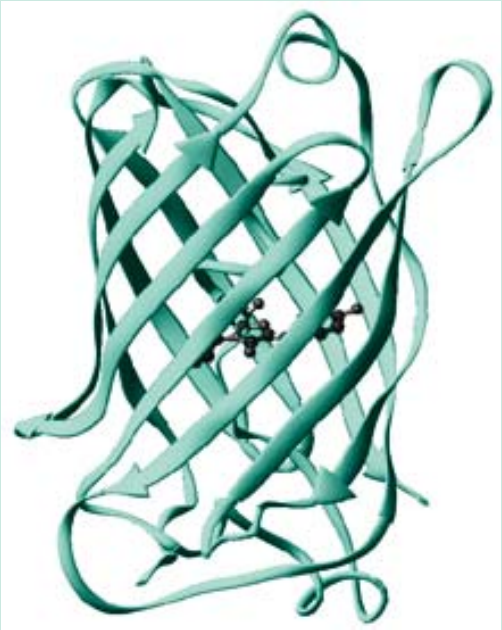
Figure 23-51 Structures of luciferins from several luminous organisms. The forms shown are the "activated" molecules ready to react with O_2. However, compound AF-350 is a breakdown product of the Ca^{2+}-activated luminous protein aequorin.

BOX 23-A THE GREEN FLUORESCENT PROTEIN AND OTHER LIGHT-EMITTING ANTENNAS

The Ca^{2+} -dependent luminescence of the jellyfish *Aquorea*, discussed in the main text, consists of blue light with a maximum intensity at 470 nm. However, the living organism has a more brilliant green luminescence. The excitation energy is transferred in a nonradiationless process to a green fluorescent protein with absorption maxima at 395 and 475 nm and an emission maximum for fluorescence at 508 nm.[a–c] A similar protein is used by *Renilla*.[d] The 238-residue green fluorescent protein (**GFP**) has a compact three-dimensional structure, an 11-stranded antiparallel β barrel with helices at one end and longer loops at the other, a "**β can**." The chromophore lies in the center of the cylinder. Numerous mutant forms have been made but only 15 residues in the terminal regions could be deleted without loss of fluorescence.[e]

The chromophore of GFP is formed spontaneously from -Ser-Tyr-Gly, residues 65–67 of the protein.[a,f,g] The entire protein has been synthesized chemically and forms the fluorescent chromophore just as in the protein produced biologically.[h] The reaction is autocatalytic, requiring only O_2 as the oxidant. Although in the living organism it accepts energy from the aqueorin chromophore a near ultraviolet lamp will elicit the fluorescence in the laboratory.





Ribbon drawing of the 238-residue green fluorescent protein showing the embedded chromophore as a ball-and-stick structure.[i] Courtesy of S. James Remington.

The green fluorescent protein is used widely in molecular biology as a fluorescent tag. Its rugged chemical nature, resistance to degradation by proteases and ability to form the chromophore autocatalytically from its own amino acids have permitted many applications. The entire GFP can be attached covalently to numerous cell components. Its gene can be spliced into the genome of an organism to form green-glowing worms, flies, and plants. Put behind a suitable promoter the fluorescent protein may be synthesized or not depending upon the control mechanism of a particular promoter (Chapter 28).[c,j–m]

The phenolic group of the GFP chromophore is apparently dissociated in the form absorbing at 395 nm and is in a tautomeric equilibrium with the other species. However, some histidine-containing replacement mutants have pH-dependent spectral changes in which the dipolar ionic form shown above, and absorbing at a longer wavelength, loses

BOX 23-A (continued)

a proton to form the anion. Observation of the excitation spectra for fluorescence of such mutant proteins within cells provides a new method for measuring the internal pH of cells and their organelles.[n–q] Some mutants emit blue or yellow light.[r,s] Two different color mutants have been fused with a molecule of calmodulin in such a way that the Ca^{2+}-induced conformational change in calmodulin allows fluorescence resonance energy transfer (FRET) between the two fluorophores. This provides a new type of calcium ion indicator.[s–u] A structurally similar red fluorescent protein, produced by a coral, extends the range of colors available as biological markers and may be useful in applications based on resonance energy transfer.[v,w,x]

Bioluminescent bacteria of the genus *Photobacterium* produce large amounts of a highly fluorescent 189-residue **lumazine protein** which contains bound 6,7-dimethyl-8-ribityllumazine (see Fig. 25-20).[y,z] Like the green fluorescent protein, it serves as a secondary light emitter receiving its energy by transfer from the flavin primary emitter. Its presence shifts the light-emission from the 495 nm of the luciferase to as low as 470 nm. *Vibrio fischeri* synthesizes a **yellow fluorescent protein** with either bound FMN or riboflavin. Its emission is at 542 nm, a longer wavelength than that of the luciferase emission. The value to the bacteria may be the higher quantum yield of fluorescence from the antenna emitter than from the luciferase. The luciferase fluorescence has a lifetime of 10 ns but on addition of the yellow fluorescent protein it is decreased to 0.25 ns with a greatly intensified emission.[u]

[a] Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A., and Tsien, R. Y. (1995) *Trends Biochem. Sci.* **20**, 448–455

[b] Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M., and Remington, S. J. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2306–2311

[c] Chalfie, M., and Kain, S., eds. (1998) *Green Fluorescent Protein: Properties, Applications and Protocols*, Wiley-Liss, New York

[d] Hart, R. C., Matthews, J. C., Hori, K., and Cormier, M. J. (1979) *Biochemistry* **18**, 2204–2205

[e] Li, X., Zhang, G., Ngo, N., Zhao, X., Kain, S. R., and Huang, C.-C. (1997) *J. Biol. Chem.* **272**, 28545–28549

[f] Niwa, H., Inouye, S., Hirano, T., Matsuno, T., Kojima, S., Kubota, M., Ohashi, M., and Tsuji, F. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13617–13622

[g] Reid, B. G., and Flynn, G. C. (1997) *Biochemistry* **36**, 6786–6791

[h] Nishiuchi, Y., Inui, T., Nishio, H., Bódi, J., Kimura, T., Tsuji, F. I., and Sakakibara, S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13549–13554

[i] Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science* **273**, 1392–1395

j Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994) *Science* **263**, 802–805

[k] Ohashi, T., Kiehart, D. P., and Erickson, H. P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2153–2158

[l] Yeh, E., Gustafson, K., and Boulianne, G. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7036–7040

[m] Hampton, R. Y., Koning, A., Wright, R., and Rine, J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 828–833

[n] Elsliger, M.-A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999) *Biochemistry* **38**, 5296–5301

[o] Robey, R. B., Ruiz, O., Santos, A. V. P., Ma, J., Kear, F., Wang, L.-J., Li, C.-J., Bernardo, A. A., and Arruda, J. A. L. (1998) *Biochemistry* **37**, 9894–9901

[p] Tsien, R. Y. (1998) *Ann. Rev. Biochem.* **67**, 509–544

[q] Wachter, R. M., Yarbrough, D., Kallio, K., and Remington, S. J. (2000) *J. Mol. Biol.* **301**, 157–171

[r] Yang, T.-T., Sinai, P., Green, G., Kitts, P. A., Chen, Y.-T., Lybarger, L., Chervenak, R., Patterson, G. H., Piston, D. W., and Kain, S. R. (1998) *J. Biol. Chem.* **273**, 8212–8216

[s] Pozzan, T. (1997) *Nature (London)* **388**, 834–835

[t] Romoser, V. A., Hinkle, P. M., and Persechini, A. (1997) *J. Biol. Chem.* **272**, 13270–13274

[u] Nagai, T., Sawano, A., Park, E. S., and Miyawaki, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3197–3202

[v] Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V., and Remington, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 462–467

[w] Garcia-Parajo, M. F., Koopman, M., van Dijk, E. M. H. P., Subramaniam, V., and van Hulst, N. F. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14392–14397

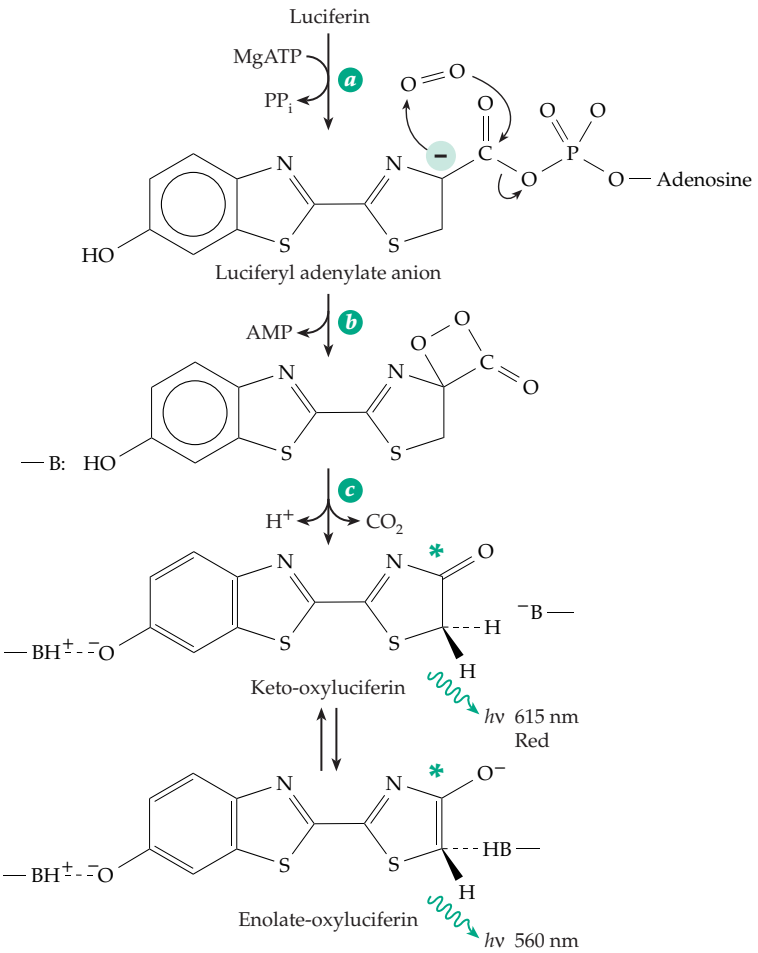
[x] Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877–7882

[y] Hughes, R. E., Brzovic, P. S., Klevit, R. E., and Hurley, J. B. (1995) *Biochemistry* **34**, 11410–11416

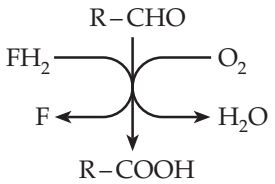
[z] Petushkov, V. N., Gibson, B. G., and Lee, J. (1996) *Biochemistry* **35**, 8413–8418

two-domain structure that suggests that domain movement may be essential to bring reactants together.[660] The structure is homologous to those of acyl-CoA ligases and peptide synthetases which share a similarity in step *a*. Formation of the dioxetane intermediate is assisted by the loss of AMP (Eq. 23-46, step *b*). The electronically excited decarboxylation product interacts with groups in the protein. It apparently exists as an anion bound to acidic and basic groups of the protein. An equilibrium between oxo- and enolate forms is thought to regulate the color of the emitted light which can vary from red to yellow and green in various fireflies, other beetles, and larvae.[661,661a,677,677a] Oxyluciferin can be reconverted to luciferin for the next flash.[677b]

A very different light-producing reaction is used by the limpet *Latia*. The luciferin is an unusual terpene derivative (Fig. 23-51) that lacks any chromophore suitable for light emission.[678] Evidently oxidation of this luciferin causes electronic excitation of some other molecule, presumably a "purple protein" which is also needed for luminescence. A complex of luciferin plus the purple protein is believed to react with the luciferase (abbreviated E-NH_2 in Fig. 23-51). It is



(23-46)



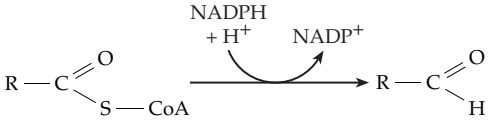
(23-47)

thought that the formyl group is released from its enolic ester linkage in the luciferin. A Schiff base of the resulting aldehyde may form with the enzyme and react with oxygen. Yet another type of luciferin is found in dinoflagellates (Fig. 23-51).[679]

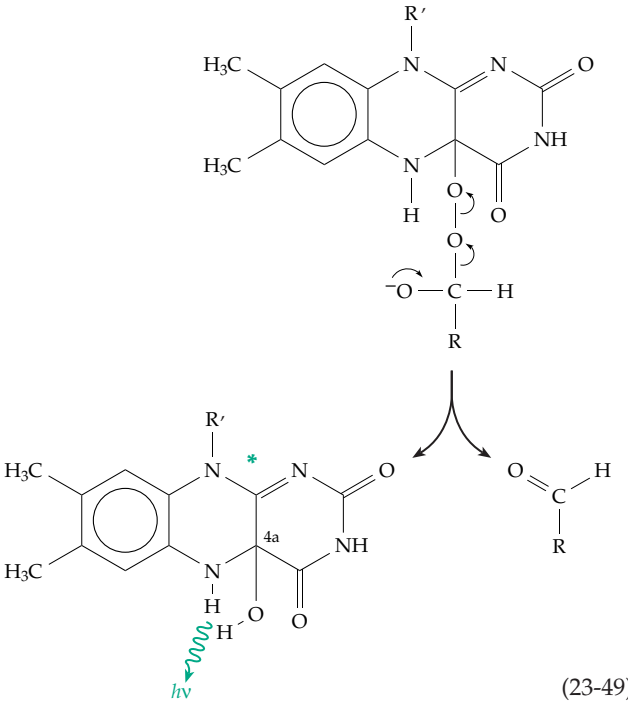
Luminescent bacteria all appear to obtain light from a riboflavin-5'-phosphate dependent oxygenase, which converts a long-chain aldehyde (usually *n*-tetradecanal) to a carboxylic acid (Eq. 23-47). Here FH_2 is the riboflavin 5'-*P*, which is thought to be supplied by a flavin reductase.[679a]

Bacterial luciferases are αβ heterodimers with subunit masses of ~40(α) and 35 (β) kDa.[664,680,681] In *Vibrio harveyi* these are encoded by the *lux A* and *lux B* genes. At least five other genes are essential for light production including two regulatory genes.[682,683] The tetradecanal and other long-chain aldehydes are supplied by reduction of the corresponding acyl-CoA (Eq. 23-48). A special thioesterase releases a myristoyl group from an acyl carrier protein, diverting it for luminescence in *V. harveyi*.[684] There is good evidence from ^{13}C NMR and electronic spectra for an enzyme-bound reduced flavin hydroperoxide as in Eq. 15-31. While this hydroperoxide can decompose slowly to flavin and H_2O_2 in the dark, it can also carry out the oxidation of the aldehyde with emission of light.[685,685a] The luminescent emission spectrum resembles the fluorescence spectrum of the 4a –OH adduct (Eq. 23-49), which is probably the light-emitting species.[686–688]

Cells of *Vibrio fischeri*, from the light organ of the fish *Monocentrus japonicus*, emit light only in dense cultures where a chemical inducer identified as N-(2-oxocaproyl)homoserine lactone[689,690] accumulates.



(23-48)



(23-49)

References

1. Wald, G. (1959) *Sci. Am.* **201**(Oct), 92–108
2. Calvert, J. G., and Pitts, J. N., Jr. (1966) *Photochemistry*, Wiley, New York
3. Suzuki, H. (1967) *Electronic Absorption Spectra and Geometry of Organic Molecules*, Academic Press, New York
4. Kraemer, K. H. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11–14
5. Kohen, E., Santus, R., and Hirschberg, J. G. (1995) *Photobiology*, Academic Press, San Diego
6. Murrell, J. N. (1967) *The Theory of the Electronic Spectra of Organic Molecules*, Academic Press, New York
7. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry Part II*, Freeman, San Francisco, California (pp. 349–480)
8. Herzberg, G. (1950) *Molecular Spectra and Molecular Structure*, 2nd ed., Vol. I, Van Nostrand-Reinhold, Princeton, New Jersey
9. Brown, S. B., ed. (1980) *An Introduction to Spectroscopy for Biochemists*, Academic Press, New York
10. Bell, J. E., ed. (1980, 1981) *Spectroscopy in Biochemistry*, CRC Press, Boca Raton, Florida
11. Campbell, I. D., and Dwek, R. A. (1983) *Biological Spectroscopy*, Addison-Wesley, Reading, Massachusetts

11a. Steinmeyer, G., Sutter, D. H., Gallmann, L., Matuschek, N., and Keller, U. (1999) *Science* **286**, 1507–1512

11b. Hopkins, J.-M., and Sibbett, W. (2000) *Sci. Am.* **283**(Sep), 72–79

11c. Service, R. F. (2001) *Science* **292**, 1627–1628

11d. Ihee, H., Lobastov, V. A., Gomez, U. M., Goodson, B. M., Srinivasan, R., Ruan, C.-Y., and Zewail, A. H. (2001) *Science* **291**, 458–462

11e. Lattman, E. E. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6535–6536

11f. Drescher, M., Hentschel, M., Kienberger, R., Tempea, G., Spielmann, C., Reider, G. A., Corkum, P. B., and Krausz, F. (2001) *Science* **291**, 1923–1927

11g. Bhattacharjee, Y. (2001) *Nature (London)* **412**, 474–476

12. Kochakian, C. D. (1988) *Trends Biochem. Sci.* **13**, 359–362

12a. Burns, D. A., and Ciurczak, E. W., eds. (2001) *Handbook of Near-Infrared Analysis*, 2nd ed., Dekker, New York

13. Parker, F. S. (1983) *Applications of Infrared, Raman and Resonance Raman Spectroscopy in Biochemistry, Biology and Medicine*, Plenum, New York
14. Haris, P. I., and Chapman, D. (1992) *Trends Biochem. Sci.* **17**, 328–333
15. Surewicz, W. K., Mantsch, H. H., and Chapman, D. (1993) *Biochemistry* **32**, 389–394

15a. Gremlich, H.-U., and Yan, B., eds. (2000) *Infrared and Raman Spectroscopy of Biological Materials*, Dekker, New York

16. Reisdorf, W. C., Jr., and Krimm, S. (1996) *Biochemistry* **35**, 1383–1386
17. Wright, W. W., Laberge, M., and Vanderkooi, J. M. (1997) *Biochemistry* **36**, 14724–14732
18. Dagneaux, C., Liquier, J., and Taillandier, E. (1995) *Biochemistry* **34**, 16618–16623
19. Sanchez-Ruiz, J. M., and Martinez-Carrion, M. (1988) *Biochemistry* **27**, 3338–3342

19a. Zhao, W., and Wright, J. C. (1999) *J. Am. Chem. Soc.* **121**, 10994–10998

20. Zscherp, C., Schlesinger, R., Tittor, J., Oesterhelt, D., and Heberle, J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5498–5503
21. Meskers, S., Ruysschaert, J.-M., and Goormaghtigh, E. (1999) *J. Am. Chem. Soc.* **121**, 5115–5122
22. Fraser, R. D. B., and MacRae, T. P. (1973) *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*, Academic Press, New York (pp. 95–106)
23. Grasselli, J. G., Snavely, M. K., and Balkin, B. J. (1981) *Chemical Applications of Raman Spectroscopy*, Wiley, New York

23a. Baumruk, V., Pancoska, P., and Keiderling, T. A. (1996) *J. Mol. Biol.* **259**, 774–791

24. Burke, M. J., and Rougvie, M. A. (1972) *Biochemistry* **11**, 2435–2439
25. Anderson, T. S., Hellgeth, J., and Lansbury, P. T., Jr. (1996) *J. Am. Chem. Soc.* **118**, 6540–6546
26. Reilly, K. E., and Thomas, G. J., Jr. (1994) *J. Mol. Biol.* **241**, 68–82
27. de Jongh, H. H. J., Goormaghtigh, E., and Ruysschaert, J.-M. (1997) *Biochemistry* **36**, 13603–13610
28. Miles, H. T., Lewis, T. P., Becker, E. D., and Frazier, J. (1973) *J. Biol. Chem.* **248**, 1115–1117
29. Hendra, P., Jones, C., and Warnes, G. (1991) *Fourier Transform Raman Spectroscopy*, Ellis Horwood, New York
30. Overman, S. A., and Thomas, G. J., Jr. (1998) *Biochemistry* **37**, 5654–5665
31. Lippert, J. L., Tyminski, D., and Desmeules, P. J. (1976) *J. Am. Chem. Soc.* **98**, 7075–7080
32. Craig, W. S., and Gaber, B. P. (1977) *J. Am. Chem. Soc.* **99**, 4130–4134
33. Tsuboi, M., Overman, S. A., and Thomas, G. J., Jr. (1996) *Biochemistry* **35**, 10403–10410

33a. Wen, Z. Q., Overman, S. A., Bondre, P., and Thomas, G. J., Jr. (2001) *Biochemistry* **40**, 449–458

34. Emerson, R., Chalmers, R., and Cederstrand, C. (1957) *Proc. Natl. Acad. Sci. U.S.A.* **43**, 133–143
35. Emerson, R., and Arnold, W. (1932) *J. Gen. Physiol.* **16**, 191–205
36. Carey, P. R. (1982) *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York
37. Lugtenburg, J., Mathies, R. A., Griffin, R. G., and Herzfeld, J. (1988) *Trends Biochem. Sci.* **13**, 388–393
38. Wright, J. C. (1982) in *Applications of Lasers to Chemical Problems* (Evans, T. R., ed), pp. 35–179, Wiley, New York
39. Chen, X. G., Li, P., Holtz, J. S. W., Chi, Z., Pajcini, V., Asher, S. A., and Kelly, L. A. (1996) *J. Am. Chem. Soc.* **118**, 9705–9715
40. Chi, Z., Chen, X. G., Holtz, J. S. W., and Asher, S. A. (1998) *Biochemistry* **37**, 2854–2864
41. Overman, S. A., and Thomas, G. J., Jr. (1999) *Biochemistry* **38**, 4018–4027
42. Hu, X., and Spiro, T. G. (1997) *Biochemistry* **36**, 15701–15712
43. Miura, T., and Thomas, G. J., Jr. (1994) *Biochemistry* **33**, 7848–7856
44. Miura, T., and Thomas, G. J., Jr. (1995) *Biochemistry* **34**, 9645–9654
45. Chan, S. S., Austin, R. H., Mukerji, I., and Spiro, T. G. (1997) *Biophys. J.* **72**, 1512–1520
46. Hu, X., Rodgers, K. R., Mukerji, I., and Spiro, T. G. (1999) *Biochemistry* **38**, 3462–3467
47. Chi, Z., and Asher, S. A. (1999) *Biochemistry* **38**, 8196–8203
48. Macdonald, I. D. G., Sligar, S. G., Christian, J. F., Unno, M., and Champion, P. M. (1999) *J. Am. Chem. Soc.* **121**, 376–380
49. Dong, S., and Spiro, T. G. (1998) *J. Am. Chem. Soc.* **120**, 10434–10440
50. Kitagawa, T., Nishina, Y., Kyogoku, Y., Yamano, T., Ohishi, N., Takai-Suzuki, A., and Yagi, K. (1979) *Biochemistry* **18**, 1804–1808
51. Kitagawa, T., Nishina, Y., Kyogoky, Y., Yamano, T., Ohishi, N., Takai-Suzuki, A., and Yagi, K. (1979) *Biochemistry* **18**, 1804–1808
52. Clarkson, J., Palfey, B. A., and Carey, P. R. (1997) *Biochemistry* **36**, 12560–12566
53. Benecky, M. J., Copeland, R. A., Hays, T. R., Lobenstine, E. W., Rava, R. P., Pascal, R. A., Jr., and Spiro, T. G. (1985) *J. Biol. Chem.* **260**, 11663–11670
54. Benecky, M. J., Copeland, R. A., Rava, R. P., Feldhaus, R., Scott, R. D., Metzler, C. M., Metzler, D. E., and Spiro, T. G. (1985) *J. Biol. Chem.* **260**, 11671–11678

54a. Altose, M. D., Zheng, Y., Dong, J., Palfey, B. A., and Carey, P. R. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3006–3011

55. Clark, R. J. H., D'Urso, N. R., and Zagalsky, P. F. (1980) *J. Am. Chem. Soc.* **102**, 6693–6698
56. Ozaki, Y., Pliura, D. H., Carey, P. R., and Storer, A. C. (1982) *Biochemistry* **21**, 3102–3108

56a. Vogel, K. M., Kozlowski, P. M., Zgierski, M. Z., and Spiro, T. G. (1999) *J. Am. Chem. Soc.* **121**, 9915–9921

56b. Franzen, S. (2001) *J. Am. Chem. Soc.* **123**, 12578–12589

57. Rospendowski, B. N., Kelly, K., Wolf, C. R., and Smith, W. E. (1991) *J. Am. Chem. Soc.* **113**, 1217–1225
58. Dähne, S. (1978) *Science* **199**, 1163–1167
59. Morley, J. O., Morley, R. M., and Fitton, A. L. (1998) *J. Am. Chem. Soc.* **120**, 11479–11488
60. Marder, S. R., Gorman, C. B., Meyers, F., Perry, J. W., Bourhill, G., Brédas, J.-L., and Pierce, B. M. (1994) *Science* **265**, 632–635
61. Horwitz, J., Strickland, E. H., and Billups, C. (1969) *J. Am. Chem. Soc.* **91**, 184–190
62. Siano, D. B., and Metzler, D. E. (1969) *J. Chem. Phys.* **51**, 1856–1861
63. Metzler, D. E., Harris, C. M., Johnson, R. J., Siano, D. B., and Thomson, J. A. (1973) *Biochemistry* **12**, 5377–5392
64. Metzler, C. M., Cahill, A. E., Petty, S., Metzler, D. E., and Lang, L. (1985) *Appl. Spectroscopy* **39**, 333–339
65. Horwitz, J., Strickland, E. H., and Billups, C. (1970) *J. Am. Chem. Soc.* **92**, 2119–2129
66. Horwitz, J., and Strickland, E. H. (1971) *J. Biol. Chem.* **246**, 3749–3752
67. Zucchelli, G., Dainese, P., Jennings, R. C., Breton, J., Garlaschi, F. M., and Bassi, R. (1994) *Biochemistry* **33**, 8982–8990
68. Tang, S. W., Coleman, J. E., and Myer, Y. P. (1968) *J. Biol. Chem.* **243**, 4286–4297
69. Hofrichter, J., and Eaton, W. A. (1976) *Annu Rev Biophys Bioeng.* **5**, 511–560
70. Arnone, A., Christen, P., Jansonias, J. N., and Metzler, D. E. (1985) in *Transaminases* (Christen, P., and Metzler, D. E., eds), pp. 349–357, Wiley, New York
71. Holmén, A., Broo, A., Albinsson, B., and Nordén, B. (1997) *J. Am. Chem. Soc.* **119**, 12240–12250
72. Peterson, J., Pearce, L. L., and Bominaar, E. L. (1999) *J. Am. Chem. Soc.* **121**, 5972–5980
73. Metzler, C. M., Mitra, J., Metzler, D. E., Makinen, M. W., Hyde, C. C., Rogers, P., and Arnone, A. (1988) *J. Mol. Biol.* **203**, 197–220
74. Morton, R. A. (1975) *Biochemical Spectroscopy*, Wiley, New York
75. Perkampus, H. H., Sandeman, I., and Timmons, C. J. (1966–1977) *UV Atlas of Organic Compounds*, Vol. 1–5, Plenum, New York
76. Lang, L., ed. (1961–1975) *Absorption Spectra in the Ultraviolet and Visible Region*, Academic Press, New York (A serial publication)
77. Petruska, J. (1961) *J. Chem. Phys.* **34**, 1120–1136
78. Stevenson, P. E. (1965) *J. Mol. Spectroscopy* **15**, 220–256
79. Strickland, E. H., Wilchek, M., Billups, C., and Horowitz, J. (1972) *J. Biol. Chem.* **247**, 572–580
80. Strickland, E. H., Billups, C., and Kay, E. (1972) *Biochemistry* **11**, 3657–3662

References

81. Pajcini, V., Chen, X. G., Bormett, R. W., Geib, S. J., Li, P., Asher, S. A., and Lidiak, E. G. (1996) *J. Am. Chem. Soc.* **118**, 9716–9726
82. Clark, L. B. (1995) *J. Am. Chem. Soc.* **117**, 7974–7986
83. Fülscher, M. P., and Roos, B. O. (1995) *J. Am. Chem. Soc.* **117**, 2089–2095
84. Fülscher, M. P., Serrano-Andrés, L., and Roos, B. O. (1997) *J. Am. Chem. Soc.* **119**, 6168–6176
85. Harders, H., Forster, S., Voelter, W., and Bacher, A. (1974) *Biochemistry* **13**, 3360–3364
86. Blackburn, G. M., and Gait, M. J., eds. (1996) *Nucleic Acids in Chemistry and Biology*, 2nd ed., Oxford Univ. Press, Oxford
87. Murrell, J. N. (1963) *The Theory of the Electronic Spectra of Organic Molecules*, Wiley, New York (Chapter 7)
88. Collins, K. D., and Stark, G. R. (1971) *J. Biol. Chem.* **246**, 6599–6605
89. Donovan, J. W. (1973) in *Methods in Enzymology*, Vol. 27, Part D (Hirs, C. H. W., and Timasheff, S. N., eds), pp. 497–525, Academic Press, New York
90. Metzler, D. E., Harris, C., Yang, I.-Y., Siano, D., and Thomson, J. A. (1972) *Biochem. Biophys. Res. Commun.* **46**, 1588–1597
91. Metzler, D. E., Metzler, C. M., and Mitra, J. (1986) *Trends Biochem. Sci.* **11**, 157–159
92. Kornblatt, J. A., Kornblatt, M. J., and Hoa, G. H. B. (1995) *Biochemistry* **34**, 1218–1223
93. Padrós, E., Dunach, M., Morros, A., Sobés, M., and Manosa, J. (1984) *Trends Biochem. Sci.* **9**, 508–510
94. Butler, W. L. (1979) *Methods Enzymol.* **56**, 501–515
95. Fasman, G. D., ed. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum, New York
96. Foss, J. G. (1963) *J. Chem. Educ.* **40**, 592–597
97. Johnson, W. C., Jr. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 145–166
98. Strickland, E. H. (1974) *Crit. Revs. Biochem.* **2**, 113–175
99. Moffitt, W., Woodward, R. B., Moscowitz, A., Klyne, W., and Djerassi, C. (1961) *J. Am. Chem. Soc.* **83**, 4013–4018
100. Bayley, P. M. (1973) *Prog. Biophys. and Mol. Biol.* **27**, 1–76
101. Johnson, W. C., Jr., and Tinoco, I., Jr. (1972) *J. Am. Chem. Soc.* **94**, 4389–4390
102. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) *Biochemistry* **13**, 3350–3359
103. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part II*, Freeman, San Francisco, California (pp. 409–433)
104. Tinoco, I., Jr., Maestre, M. F., and Bustamante, C. (1983) *Trends Biochem. Sci.* **8**, 41–44
105. Applequist, J. (1973) *J. Am. Chem. Soc.* **95**, 8255–8262
106. Applequist, J. (1987) *Am. Scientist* **75**, 58–68
107. Applequist, J., and Bode, K. A. (1999) *J. Phy. Chem. B* **103**, 1767–1773
108. Bode, K. A., and Applequist, J. (1998) *J. Am. Chem. Soc.* **120**, 10938–10946 , Erratum 13545
109. Bode, K. A., and Applequist, J. (1997) *Biopolymers* **42**, 855–860
110. Paterlini, M. G., Freedman, T. B., and Nafie, L. A. (1986) *J. Am. Chem. Soc.* **108**, 1389–1397
111. Bose, P. K., and Polavarapu, P. L. (1999) *J. Am. Chem. Soc.* **121**, 6094–6095
112. Baumruk, V., and Keiderling, T. A. (1993) *J. Am. Chem. Soc.* **115**, 6939–6942
113. Self, B. D., and Moore, D. S. (1997) *Biophys. J.* **73**, 339–347
114. Bell, A. F., Hecht, L., and Barron, L. D. (1997) *J. Am. Chem. Soc.* **119**, 6006–6013
115. Peters, K. S., Watson, T., and Marr, K. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 343–362
116. Hung, R. R., and Grabowski, J. J. (1999) *J. Am. Chem. Soc.* **121**, 1359–1364
117. Meisel, A., Leonhardt, G., and Szargan, R. (1989) *X-Ray Spectra and Chemical Binding*, Springer-Verlag, New York
118. Binsted, N., Strange, R. W., and Hasnain, S. S. (1992) *Biochemistry* **31**, 12117–12125
119. Ito, E., Oji, H., Araki, T., Oichi, K., Ishii, H., Ouchi, Y., Ohta, T., Kosugi, N., Maruyama, Y., Naito, T., Inabe, T., and Seki, K. (1997) *J. Am. Chem. Soc.* **119**, 6336–6344
120. Vértes, A., Korecz, L., and Burger, K. (1979) *Mössbauer Spectroscopy*, Elsevier, Amsterdam
121. Popescu, C. V., Bates, D. M., Beinert, H., Münck, E., and Kiley, P. J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13431–13435
122. Schünemann, V., Trautwein, A. X., Illerhaus, J., and Haehnel, W. (1999) *Biochemistry* **38**, 8981–8991
123. Clayton, R. K. (1970) *Light and Living Matter*, Vol. 1, McGraw-Hill, New York
124. Parker, C. A. (1968) *Photoluminescence of Solutions*, Elsevier, Amsterdam
125. Konev, S. V. (1967) *Fluorescence and Phosphorescence of Proteins and Nucleic Acids*, Plenum, New York
126. Lakowicz, J. R. (1985) *Principles of Fluorescence Spectroscopy*, Plenum, New York
127. Baeyens, W. R. G., De Keukeleire, D., and Korkidis, K., eds. (1991) *Luminescence Techniques in Chemical and Biochemical Analysis*, Dekker, New York

127a. Meyer-Almes, F.-J., and Auer, M. (2000) *Biochemistry* **39**, 13261–13268

128. Kotaki, A., and Yagi, K. (1970) *J. Biochem.* **68**, 509–516
129. Koziol, J., and Knobloch, E. (1965) *Biochim. Biophys. Acta.* **102**, 289–300
130. Chen, Y., and Barkley, M. D. (1998) *Biochemistry* **37**, 9976–9982
131. Beechem, J. M., and Brand, L. (1985) *Ann. Rev. Biochem.* **54**, 43–71
132. Anderson, S. R. (1991) *J. Biol. Chem.* **266**, 11405–11408
133. Hedstrom, J., Sedarous, S., and Prendergast, F. G. (1988) *Biochemistry* **27**, 6203–6208
134. Wahl, P., Auchet, J.-C., Visser, A. J. W. G., and Veeger, C. (1975) *Eur. J. Biochem.* **50**, 413–418
135. Vos, R., Engelborghs, Y., Izard, J., and Baty, D. (1995) *Biochemistry* **34**, 1734–1743
136. Broos, J., ter Veld, F., and Robillard, G. T. (1999) *Biochemistry* **38**, 9798–9803
137. Ross, J. B. A., Senear, D. F., Waxman, E., Kombo, B. B., Rusinova, E., Huang, Y. T., Laws, W. R., and Hasselbacher, C. A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 12023–12027
138. Farah, C. S., and Reinach, F. C. (1999) *Biochemistry* **38**, 10543–10551
139. Strambini, G. B., and Gonnelli, M. (1995) *J. Am. Chem. Soc.* **117**, 7646–7651
140. Zhou, J. S., and Hoffman, B. M. (1994) *Science* **265**, 1693–1696
141. Eftink, M. R., and Ghiron, C. A. (1976) *Biochemistry* **15**, 672–680
142. Calhoun, D. B., Vanderkooi, J. M., and Englander, S. W. (1983) *Biochemistry* **22**, 1533–1539
143. Wright, W. W., Owen, C. S., and Vanderkooi, J. M. (1992) *Biochemistry* **31**, 6538–6544
144. Vet, J. A. M., Majithia, A. R., Marras, S. A. E., Tyagi, S., Dube, S., Poiesz, B. J., and Kramer, F. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6394–6399
145. Hilinski, E. F., and Rentzepis, P. M. (1983) *Nature (London)* **302**, 481–487
146. Radda, G. K. (1971) *Curr. Top. Bioenerg.* **4**, 81–176
147. Lakowicz, J. R., Laczko, G., Gryczynski, I., and Cherek, H. (1986) *J. Biol. Chem.* **261**, 2240–2245
148. Ruggiero, A. J., Todd, D. C., and Fleming, G. R. (1990) *J. Am. Chem. Soc.* **112**, 1003–1014
149. Förster, T. (1948) *Ann Physik* **2**, 55–75
150. Stryer, L. (1968) *Science* **162**, 526–533
151. Wu, C.-W., and Stryer, L. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1104–1108
152. Ha, T., Enderle, T., Ogletree, D. F., Chemla, D. S., Selvin, P. R., and Weiss, S. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6264–6268
153. Weiss, S. (1999) *Science* **283**, 1676–1683
154. Ha, T., Zhuang, X., Kim, H. D., Orr, J. W., Williamson, J. R., and Chu, S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9077–9082
155. Fu, P. K.-L., and Turro, C. (1999) *J. Am. Chem. Soc.* **121**, 1–7
156. Hogue, C. W. V., MacManus, J. P., Banville, D., and Szabo, A. G. (1992) *J. Biol. Chem.* **267**, 13340–13347
157. Xiao, M., Li, H., Snyder, G. E., Cooke, R., Yount, R. G., and Selvin, P. R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15309–15314
158. Gordon, G. W., Berry, G., Liang, X. H., Levine, B., and Herman, B. (1998) *Biophys. J.* **74**, 2702–2713
159. Tsien, R. Y., and Miyawaki, A. (1998) *Science* **280**, 1954–1955
160. Wouters, F. S., Bastiaens, P. I. H., Wirtz, K. W. A., and Jovin, T. M. (1998) *EMBO J.* **17**, 7179–7189
161. Ward, L. D., Seckler, R., and Timasheff, S. N. (1994) *Biochemistry* **33**, 11900–11908
162. Farrar, S. J., Whiting, P. J., Bonnert, T. P., and McKernan, R. M. (1999) *J. Biol. Chem.* **274**, 10100–10104
163. Parkhurst, K. M., and Parkhurst, L. J. (1995) *Biochemistry* **34**, 293–300
164. Jares-Erijman, E. A., and Jovin, T. M. (1996) *J. Mol. Biol.* **257**, 597–617

164a. Norman, D. G., Grainger, R. J., Uhrín, D., and Lilley, D. M. J. (2000) *Biochemistry* **39**, 6317–6324

164b. Tóth, K., Brun, N., and Langowski, J. (2001) *Biochemistry* **40**, 6921–6928

165. Rhee, M.-J., Sudnick, D. R., Arkle, V. K., and Horrocks, W. DeW., Jr. (1981) *Biochemistry* **20**, 3328–3334
166. Szczesniak, M., Szczepaniak, K., Kwiatkowski, J. S., KuBulat, K., and Person, W. B. (1988) *J. Am. Chem. Soc.* **110**, 8319–8330
167. Vanderkooi, J. M., Kaposi, A., and Fidy, J. (1993) *Trends Biochem. Sci.* **18**, 71–76
168. Friedrich, J., Gafert, J., Zollfrank, J., Vanderkooi, J., and Fidy, J. (1994) *Proc. atl. Acad. Sci. U.S.A.* **91**, 1029–1033

168a. Bastiaens, P. I. H., and Pepperkok, R. (2000) *Trends Biochem. Sci.* **25**, 631–637

168b. Frohn, J. T., Knapp, H. F., and Stemmer, A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7232–7236

168c. Klar, T. A., Jakobs, S., Dyba, M., Egner, A., and Hell, S. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8206–8210

168d. Weiss, S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8747–8749

168e. Farber, S. A., Pack, M., Ho, S.-Y., Johnson, I. D., Wagner, D. S., Dosch, R., Mullins, M. C., Hendrickson, H. S., Hendrickson, E. K., and Halpern, M. E. (2001) *Science* **292**, 1385–1388

169. Ruth, J. L. (1984) *DNA* **3**, 123
170. Renz, M., and Kurz, C. (1984) *Nucleic Acids Res.* **12**, 3435–3444
171. Davenport, L., Dale, R. E., Bisby, R. H., and Cundall, R. B. (1985) *Biochemistry* **24**, 4097–4108
172. MacGregor, R. B., and Weber, G. (1986) *Nature (London)* **319**, 70–73

References

173. Tsien, R. Y., and Poenie, M. (1986) *Trends Biochem. Sci.* **11**, 450–455
174. Thomas, F., Serratrice, G., Béguin, C., Saint Aman, E., Pierre, J. L., Fontecave, M., and Laulhère, J. P. (1999) *J. Biol. Chem.* **274**, 13375–13383
175. Stewart, W. W. (1981) *Nature (London)* **292**, 17–21
176. Berland, K. M. (1997) *Biophys. J.* **72**, 1487–1488
177. Schwille, P., Meyer-Almes, F.-J., and Rigler, R. (1997) *Biophys. J.* **72**, 1878–1886
178. Korlach, J., Schwille, P., Webb, W. W., and Feigenson, G. W. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8461–8466
179. Marx, J. (1996) *Science* **273**, 430
180. Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M. A., Ning, Y., Ledbetter, D. H., Bar-Am, I., Soenksen, D., Garini, Y., and Ried, T. (1996) *Science* **273**, 494–497

180a. Pathak, S., Cjoi, S.-K., Arnheim, N., and Thompson, M. E. (2001) *J. Am. Chem. Soc.* **123**, 4103–4104

181. Bruchez, M., Jr., Moronne, M., Gin, P., Weiss, S., and Alivisatos, A. P. (1998) *Science* **281**, 2013–2016
182. Chan, W. C. W., and Nie, S. (1998) *Science* **281**, 2016–2018
183. Turro, N. J., and Schuster, G. (1975) *Science* **187**, 303–312
184. Salem, L. (1976) *Science* **191**, 822–830
185. Bazhulina, N. P., Morozov, Y. V., Karpeisky, M. Y., Ivanov, V. I., and Kuklin, A. I. (1966) *Biofizika* **11**, 42–47
186. Bridges, J. W., Davies, D. S., and Williams, R. T. (1966) *Biochem. J.* **98**, 451–468
187. Johnson, G. F., Tu, J.-I., Bartlett, M. L. S., and Graves, D. J. (1970) *J. Biol. Chem.* **245**, 5560–5568
188. Shaltiel, S., and Cortijo, M. (1970) *Biochem. Biophys. Res. Commun.* **41**, 594–600
189. Loken, M. R., Hayes, J. W., Gohlke, J. R., and Brand, L. (1972) *Biochemistry* **11**, 4779–4786
190. Becker, M. M., and Wang, J. C. (1984) *Nature (London)* **309**, 682–687
191. Sancar, A. (1996) *Science* **272**, 48–49

191a. Lee, J.-H., Bae, S.-H., and Choi, B.-S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4591–4596

192. Shaw, A. A., Falick, A. M., and Shetlar, M. D. (1992) *Biochemistry* **31**, 10976–10983
193. Pashev, I. G., Dimitrov, S. I., and Angelov, D. (1991) *Trends Biochem. Sci.* **16**, 323–326
194. Pearlman, D. A., Holbrook, S. R., Pirkle, D. H., and Kim, S.-H. (1985) *Science* **227**, 1304–1305
195. Cimino, G. D., Gamper, H. B., Isaacs, S. T., and Hearst, J. E. (1985) *Ann. Rev. Biochem.* **54**, 1151–1193

195a. Yoon, J.-H., Lee, C.-S., O'Connor, T. R., Yasui, A., and Pfeifer, G. P. (2000) *J. Mol. Biol.* **299**, 681–693

196. Hanson, K. M., Li, B., and Simon, J. D. (1997) *J. Am. Chem. Soc.* **119**, 2715–2721
197. Edelson, R. L. (1988) *Sci. Am.* **259**(Aug), 68–75
198. Berns, M. W. (1991) *Sci. Am.* **264**(Jun), 84–90
199. Amato, I. (1993) *Science* **262**, 32–33
200. Staros, J. V. (1980) *Trends Biochem. Sci.* **5**, 320–322
201. Jo, E., Blazyk, J., and Boggs, J. M. (1998) *Biochemistry* **37**, 13791–13799
202. Gronemeyer, H. (1985) *Trends Biochem. Sci.* **10**, 264–267
203. Dormán, G., and Prestwich, G. D. (1994) *Biochemistry* **33**, 5661–5673
204. Gribbon, P., and Hardingham, T. E. (1998) *Biophys. J.* **75**, 1032–1039
205. Peters, R. (1985) *Trends Biochem. Sci.* **10**, 223–227
206. Pagano, R. E., and Longmuir, K. J. (1983) *Trends Biochem. Sci.* **8**, 157–161
207. Anfinrud, P., de Vivie-Riedle, R., and Engel, V. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8328–8329
208. Goldbeck, R. A., Paquette, S. J., Björling, S. C., and Kliger, D. S. (1996) *Biochemistry* **35**, 8628–8639
209. Service, R. F. (1997) *Science* **276**, 1986–1987
210. Brunori, M., Cutruzzola, F., Savino, C., Travaglini-Allocatelli, C., Vallone, B., and Gibson, Q. H. (1999) *Trends Biochem. Sci.* **24**, 253–255
211. Wan, C., Fiebig, T., Kelley, S. O., Treadway, C. R., Barton, J. K., and Zewail, A. H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6014–6019
212. Henderson, P. T., Jones, D., Hampikian, G., Kan, Y., and Schuster, G. B. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8353–8358
213. Block, S. M. (1992) *Nature (London)* **360**, 493–495
214. Svoboda, K., and Block, S. M. (1994) *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285
215. Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997) *Biophys. J.* **72**, 1335–1346
216. Arai, Y., Yasuda, R., Akashi, K.-i, Harada, Y., Miyata, H., Kinosita, K., Jr., and Itoh, H. (1999) *Nature (London)* **399**, 446–448
217. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991) *Science* **251**, 767–773
218. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., and Hinsberg, W. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13555–13560
219. Taubes, G. (1997) *Science* **276**, 1991–1993
220. Benaron, D. A., Cheong, W.-F., and Stevenson, D. K. (1997) *Science* **276**, 2002–2003
221. Bappart, S. A., Tearney, G. J., Bouma, B. E., Southern, J. F., Brezinski, M. E., and Fujimoto, J. G. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4256–4261

221a. Sharpe, J., Ahlgren, U., Perry, P., Hill, B., Ross, A., Hecksher-Sorensen, J., Baldock, R., and Davidson, D. (2002) *Science* **296**, 541–545

222. Jamin, N., Dumas, P., Moncuit, J., Fridman, W.-H., Teillaud, J.-L., Carr, G. L., and Williams, G. P. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4837–4840
223. Wetzel, D. L., and LeVine, S. M. (1999) *Science* **285**, 1224–1225
224. Foyer, C. H. (1984) *Photosynthesis*, Wiley, New York
225. Clayton, R. K. (1981) *Photosynthesis Physical Mechanisms and Chemical Patterns*, Cambridge Univ. Press, London
226. Gregory, R. P. F. (1989) *Biochemistry of Photosynthesis*, 3rd ed., Wiley, New York
227. Hall, D. O., and Rao, K. K. (1994) *Photosynthesis*, 5th ed., Cambridge Univ. Press, New York
228. Barber, J., and Andersson, B. (1994) *Nature (London)* **370**, 31–34
229. Kamminga, H. (1981) *Trends Biochem. Sci.* **6**, 164–165
230. van Niel, C. B. (1931) *Adv. Enzymol.* **1**, 263–328
231. Hill, R. (1937) *Nature (London)* **139**, 881–882
232. Gaffron, H. (1960) in *Plant Physiology*, Vol. 1B (Steward, F. C., ed), pp. 176–180, Academic Press, New York
233. Prince, R. C. (1996) *Trends Biochem. Sci.* **21**, 121–122
234. Arnon, D. I., Tsujimoto, H. Y., and McSwai, B. D. (1965) *Nature (London)* **207**, 1357–1372
235. Arnon, D. I. (1984) *Trends Biochem. Sci.* **9**, 258–262
236. Zito, F., Finazzi, G., Joliot, P., and Wollman, F.-A. (1998) *Biochemistry* **37**, 10395–10403

236a. Finazzi, G., Zito, F., Barbagallo, R. P., and Wollman, F.-A. (2001) *J. Biol. Chem.* **276**, 9770–9774

237. Haley, J., and Bogorad, L. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 1534–1538

237a. Roberts, A. G., and Kramer, D. M. (2001) *Biochemistry* **40**, 13407–13412

237b. Deniau, C., and Rappaport, F. (2000) *Biochemistry* **39**, 3304–3310

237c. Sainz, G., Carrell, C. J., Ponamarev, M. V., Soriano, G. M., Cramer, W. A., and Smith, J. L. (2000) *Biochemistry* **39**, 9164–9173

237d. Fernández-Velasco, J. G., Jamshidi, A., Gong, X.-S., Zhou, J., and Ueng, R. Y. (2001) *J. Biol. Chem.* **276**, 30598–30607

237e. Bunney, T. D., van Walraven, H. S., and de Boer, A. H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4249–4254

238. Peterman, E. J. G., Wenk, S.-O., Pullerits, T., Pålsson, L.-O., van Grondelle, R., Dekker, J. P., Rögner, M., and van Amerongen, H. (1998) *Biophys. J.* **75**, 389–398
239. Carrell, C. J., Schlarb, B. G., Bendall, D. S., Howe, C. J., Cramer, W. A., and Smith, J. L. (1999) *Biochemistry* **38**, 9590–9599
240. Xue, Y., Ökvist, M., Hansson, Ö., and Young, S. (1998) *Protein Sci.* **7**, 2099–2105
241. Schnackenberg, J., Than, M. E., Mann, K., Wiegand, G., Huber, R., and Reuter, W. (1999) *J. Mol. Biol.* **290**, 1019–1030
242. De la Cerda, B., Díaz-Quintana, A., Navarro, J. A., Hervás, M., and De la Rosa, M. A. (1999) *J. Biol. Chem.* **274**, 13292–13297

242a. Molina-Heredia, F. P., Hervás, M., Navarro, J. A., and De la Rosa, M. A. (2001) *J. Biol. Chem.* **276**, 601–605

242b. Baymann, F., Rappaport, F., Joliot, P., and Kallas, T. (2001) *Biochemistry* **40**, 10570–10577

243. Kuras, R., Guergova-Kuras, M., and Crofts, A. R. (1998) *Biochemistry* **37**, 16280–16288
244. Zhang, H., Huang, D., and Cramer, W. A. (1999) *J. Biol. Chem.* **274**, 1581–1587
245. Oh-oka, H., Iwaki, M., and Itoh, S. (1998) *Biochemistry* **37**, 12293–12300
246. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-i, Inokuchi, H., and Ozeki, H. (1986) *Nature (London)* **322**, 572–574
247. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Sugiura, M. (1986) *EMBO J.* **5**, 2043–2049
248. Maier, R. M., Neckermann, K., Igloi, G. L., and Kössel, H. (1995) *J. Mol. Biol.* **251**, 614–628
249. Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K., and Sugiura, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5967–5972

249a. Cline, K. (2000) *Nature (London)* **403**, 148–149

250. Zhang, Z., Green, B. R., and Cavalier-Smith, T. (1999) *Nature (London)* **400**, 155–159
251. Kouranov, A., and Schnell, D. J. (1996) *J. Biol. Chem.* **271**, 31009–31012
252. Asai, T., Shinoda, Y., Nohara, T., Yoshihisa, T., and Endo, T. (1999) *J. Biol. Chem.* **274**, 20075–20078
253. Bölter, B., Soll, J., Schulz, A., Hinnah, S., and Wagner, R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15831–15836

References

253a. Dabney-Smith, C., van den Wijngaard, P. W. J., Treece, Y., Vredenberg, W. J., and Bruce, B. D. (1999) *J. Biol. Chem.* **274**, 32351–32359
254. Douce, R., and Joyard, J. (1981) *Trends Biochem. Sci.* **6**, 237–240
255. Hoober, J. K. (1984) *Chloroplasts*, Plenum, New York
255a. von Wettstein, D. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3633–3635
256. Muhlethaler, K. (1971) in *Structure and Function of Chloroplasts* (Gibbs, M., ed), pp. 7–34, Springer-Verlag, Berlin and New York
257. Miller, K. R. (2000) *Sci. Am.*, 102–113
258. Seibert, M., DeWit, M., and Staehelin, L. A. (1987) *J. Cell Biol.* **105**, 2257–2265
259. Anderson, J. M., and Anderson, B. (1982) *Trends Biochem. Sci.* **7**, 288–292
260. Rebeiz, C. A., and Lascelles, J. (1982) in *Photosynthesis*, Vol. I (Govindjee, ed), pp. 699–780, Academic Press, New York
261. Miyashita, H., Ikemoto, H., Kurano, N., Adachi, K., Chihara, M., and Miyachi, S. (1996) *Nature (London)* **383**, 402
262. Caple, M. B., Chow, H., and Strouse, C. E. (1978) *J. Biol. Chem.* **253**, 6730–6737
263. Smith, K. M., Kehres, L. A., and Tabba, H. D. (1980) *J. Am. Chem. Soc.* **102**, 7149–7151
264. French, C. S., and Brown, J. S. (1972) in *Photosynthesis. Two Centuries after Its Discovery by Joseph Priestly*, 2nd ed. (Forti, G., Avron, M., and Melandri, A., eds), pp. 291–306, Junk, The Hague
265. Govindjee, R. (1974) *Sci. Am.* **231**(Dec), 68–82
266. Wolken, J. J. (1975) *Photoprocesses, Photoreceptors, and Evolution*, Academic Press, New York
267. Song, P.-S. (1978) *Trends Biochem. Sci.* **3**, 25–27
268. Hofmann, E., Wrench, P. M., Sharples, F. P., Hiller, R. G., Welte, W., and Diederichs, K. (1996) *Science* **272**, 1788–1791
269. Goodwin, T. W., ed. (1971) *Aspects of Terpenoid Chemistry and Biochemistry*, Academic Press, New York (pp. 346–348)
270. Schirmer, W., Bode, R., Sidler, W., and Zuber, H. (1985) *J. Mol. Biol.* **184**, 257–277
271. Szalontai, B., Gombos, Z., Csizmadia, V., Bagyinka, C., and Lutz, M. (1994) *Biochemistry* **33**, 11823–11832
272. Glazer, A. N. (1989) *J. Biol. Chem.* **264**, 1–4
273. Jung, L. J., Chan, F. C., and Glazer, A. N. (1995) *J. Biol. Chem.* **270**, 12877–12884
274. Killilea, S. D., O'Carra, P., and Murphy, R. F. (1980) *Biochem. J.* **187**, 311–320
275. Bishop, J. E., Nagy, J. O., O'Connell, J. F., and Rapoport, H. (1991) *Chemical Society Special Publication, No. 2* **113**, 8024–8035
276. Fairchild, C. D., and Glazer, A. N. (1994) *J. Biol. Chem.* **269**, 28988–28996
277. MacColl, R., Guard-Friar, D., and Ryan, T. J. (1990) *Biochemistry* **29**, 430–435
278. Wilk, K. E., Harrop, S. J., Jankova, L., Edler, D., Keenan, G., Sharples, F., Hiller, R. G., and Curmi, P. M. G. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8901–8906
279. Chang, W.-r, Jiang, T., Wan, Z.-l, Zhang, J.-p, Yang, Z.-x, and Liang, D.-c. (1996) *J. Mol. Biol.* **262**, 721–731
280. Liu, J.-Y., Jiang, T., Zhang, J.-P., and Liang, D.-C. (1999) *J. Biol. Chem.* **274**, 16945–16952
281. Reuter, W., Wiegand, G., Huber, R., and Than, M. E. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1363–1368
282. Rhie, G., and Beale, S. I. (1992) *J. Biol. Chem.* **267**, 16088–16093
283. Fairchild, C. D., and Glazer, A. N. (1994) *J. Biol. Chem.* **269**, 8686–8694
284. Terry, M. J., McDowell, M. T., and Lagarias, J. C. (1995) *J. Biol. Chem.* **270**, 11111–11118
285. Riley, C. T., Barbeau, B. K., Keim, P. S., Kézdy, F. J., Heinrikson, R. L., and Law, J. H. (1984) *J. Biol. Chem.* **259**, 13159–13165
286. Glazer, A. N. (1983) *Ann. Rev. Biochem.* **52**, 125–157
286a. Grossman, A. R., Bhaya, D., and He, Q. (2001) *J. Biol. Chem.* **276**, 11449–11452
287. Thornber, P. J., and Maxwell, J. P. (1981) *Trends Biochem. Sci.* **6**, 122–124
288. Deisenhofer, J., Michel, H., and Huber, R. (1985) *Trends Biochem. Sci.* **10**, 243–248
289. Brejc, K., Ficner, R., Huber, R., and Steinbacher, S. (1995) *J. Mol. Biol.* **249**, 424–440
290. Lao, K., and Glazer, A. N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5258–5263
291. Hu, X., and Schulten, K. (1998) *Biophys. J.* **75**, 683–694
292. Karrasch, S., Bullough, P. A., and Ghosh, R. (1995) *EMBO J.* **14**, 631–638
293. Walz, T., Jamieson, S. J., Bowers, C. M., Bullough, P. A., and Hunter, C. N. (1998) *J. Mol. Biol.* **282**, 833–845
294. McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J., and Isaacs, N. W. (1995) *Nature (London)* **374**, 517–521
295. Prince, S. M., Papiz, M. Z., Freer, A. A., McDermott, G., Hawthornthwaite-Lawless, A. M., Cogdell, R. J., and Isaacs, N. W. (1997) *J. Mol. Biol.* **268**, 412–423
296. Hu, X., Damjanovic, A., Ritz, T., and Schulten, K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5935–5941
297. Nagarajan, V., Alden, R. G., Williams, J. C., and Parson, W. W. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13774–13779
298. Fraser, N. J., Dominy, P. J., Ücker, B., Simonin, I., Scheer, H., and Cogdell, R. J. (1999) *Biochemistry* **38**, 9684–9692
298a. McLuskey, K., Prince, S. M., Cogdell, R. J., and Isaacs, N. W. (2001) *Biochemistry* **40**, 8783–8789
299. van Oijen, A. M., Ketelaars, M., Köhler, J., Aartsma, T. J., and Schmidt, J. (1999) *Science* **285**, 400–402
300. Kühlbrandt, W. (1995) *Nature (London)* **374**, 479–498
300a. van Grondelle, R., and Novoderezhkin, V. (2001) *Biochemistry* **40**, 15057–15068
300b. Alia, Matysik, J., Soede-Huijbregts, C., Baldus, M., Raap, J., Lugtenburg, J., Gast, P., van Gorkom, H. J., Hoff, A. J., and de Groot, H. J. M. (2001) *J. Am. Chem. Soc.* **123**, 4803–4809
300c. Simonetto, R., Crimi, M., Sandoná, D., Croce, R., Cinque, G., Breton, J., and Bassi, R. (1999) *Biochemistry* **38**, 12974–12983
301. Feick, R. G., and Fuller, R. C. (1984) *Biochemistry* **23**, 3693–3700
302. Li, Y.-F., Zhou, W., Blankenship, R. E., and Allen, J. P. (1997) *J. Mol. Biol.* **271**, 456–471
302a. Vassilieva, E. V., Antonkine, M. L., Zybailov, B. L., Yang, F., Jakobs, C. U., Golbeck, J. H., and Bryant, D. A. (2001) *Biochemistry* **40**, 464–473
303. Matthews, B. W., Fenna, R. E., Bolognesi, M. C., Schmid, M. F., and Olson, J. M. (1979) *J. Mol. Biol.* **131**, 259–285
304. Rémigy, H.-W., Stahlberg, H., Fotiadis, D., Müller, S. A., Wolpensinger, B., Engel, A., Hauska, G., and Tsiotis, G. (1999) *J. Mol. Biol.* **290**, 851–858
305. Fenna, R. E., and Matthews, B. W. (1975) *Nature (London)* **258**, 573–577
306. Douady, D., Rousseau, B., and Caron, L. (1994) *Biochemistry* **33**, 3165–3170
307. Pascal, A. A., Caron, L., Rousseau, B., Lapouge, K., Duval, J.-C., and Robert, B. (1998) *Biochemistry* **37**, 2450–2457
308. Green, B. R., Pichersky, E., and Kloppstech, K. (1991) *Trends Biochem. Sci.* **16**, 181–186
309. La Roche, J., van der Staay, G. W. M., Partensky, F., Ducret, A., Aebersold, R., Li, R., Golden, S. S., Hiller, R. G., Wrench, P. M., Larkum, A. W. D., and Green, B. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15244–15248
310. Kühlbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994) *Nature (London)* **367**, 614–621
311. Kleima, F. J., Gradinaru, C. C., Calkoen, F., van Stokkum, I. H. M., van Grendelle, R., and van Amerongen, H. (1997) *Biochemistry* **36**, 15262–15268
312. Simidjiev, I., Barzda, V., Mustárdy, L., and Garab, G. (1998) *Biochemistry* **37**, 4169–4173
313. Rhee, K.-H., Morris, E. P., Zheleva, D., Hankamer, B., Kühlbrandt, W., and Barber, J. (1997) *Nature (London)* **389**, 522–526
314. Rhee, K.-H., Morris, E. P., Barber, J., and Kühlbrandt, W. (1998) *Nature (London)* **396**, 283–286
315. Barber, J., Nield, J., Morris, E. P., and Hankamer, B. (1999) *Trends Biochem. Sci.* **24**, 43–45
315a. Voigt, B., Irrgang, K.-D., Ehlert, J., Beenken, W., Renger, G., Leupold, D., and Lokstein, H. (2002) *Biochemistry* **41**, 3049–3056
316. Schubert, W.-D., Klukas, O., Saenger, W., Witt, H. T., Fromme, P., and Krauss, N. (1998) *J. Mol. Biol.* **280**, 297–314
317. Croce, R., Zucchelli, G., Garlaschi, F. M., Bassi, R., and Jennings, R. C. (1996) *Biochemistry* **35**, 8572–8579
317a. Bibby, T. S., Nield, J., Partensky, F., and Barber, J. (2001) *Nature (London)* **413**, 590
317b. Rogl, H., Schödel, R., Lokstein, H., Kühlbrandt, W., and Schubert, A. (2002) *Biochemistry* **41**, 2281–2287
318. Thomas, B. A., McMahon, L. P., and Klotz, A. V. (1995) *Biochemistry* **34**, 3758–3770
318a. Heathcote, P., Fyfe, P. K., and Jones, M. R. (2002) *Trends Biochem. Sci.* **27**, 79–87
319. Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985) *Nature (London)* **318**, 618–624
320. Deisenhofer, J., and Michel, H. (1991) *Ann. Rev. Biophys. Biophys. Chem.* **20**, 247–266
321. Deisenhofer, J., Epp, O., Sinning, I., and Michel, H. (1995) *J. Mol. Biol.* **246**, 429–457
322. Deisenhofer, J., and Michel, H. (1989) *Science* **245**, 1463–1473
323. Lancaster, C. R. D., and Michel, H. (1999) *J. Mol. Biol.* **286**, 883–898
323a. Chen, I.-P., Mathis, P., Koepke, J., and Michel, H. (2000) *Biochemistry* **39**, 3592–3602
324. Philipson, K. D., and Sauer, K. (1972) *Biochemistry* **11**, 1880–1885
325. Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1984) *J. Mol. Biol.* **180**, 385–398
325a. McAuley, K. E., Fyfe, P. K., Ridge, J. P., Cogdell, R. J., Isaacs, N. W., and Jones, M. R. (2000) *Biochemistry* **39**, 15032–15043
326. Siefermann-Harms, D. (1985) *Biochim. Biophys. Acta.* **811**, 325–355
327. El-Kabbani, O., Chang, C.-H., Tiede, D., Norris, J., and Schiffer, M. (1991) *Biochemistry* **30**, 5361–5369
328. Govindjee. (1978) *Photochemistry and Photobiology* **28**, 935–938
329. Blankenship, R. E., and Parson, W. W. (1978) *Ann. Rev. Biochem.* **47**, 635–653
330. Dutton, P. L., Leigh, J. S., Jr., and Reed, D. W. (1972) *Biochim. Biophys. Acta.* **292**, 654–664
331. Dutton, P. L., and Mosser, C. C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10247–10250
331a. Li, J., Takahashi, E., and Gunner, M. R. (2000) *Biochemistry* **39**, 7445–7454
331b. Rabenstein, B., Ullmann, G. M., and Knapp, E.-W. (2000) *Biochemistry* **39**, 10487–10496
331c. Paddock, M. L., Ädelroth, P., Chang, C., Abresch, E. C., Feher, G., and Okamura, M. Y. (2001) *Biochemistry* **40**, 6893–6902

References

331d. Tandori, J., Baciou, L., Alexov, E., Maróti, P., Schiffer, M., Hanson, D. K., and Sebban, P. (2001) *J. Biol. Chem.* **276**, 45513–45515
331e. Gerencsér, L., and Maróti, P. (2001) *Biochemistry* **40**, 1850–1860
331f. Ädelroth, P., Paddock, M. L., Tehrani, A., Beatty, J. T., Feher, G., and Okamura, M. Y. (2001) *Biochemistry* **40**, 14538–14546
331g. Xu, Q., and Gunner, M. R. (2002) *Biochemistry* **41**, 2694–2701
332. Vos, M. H., Rappaport, F., Lambry, J.-C., Breton, J., and Martin, J.-L. (1993) *Nature (London)* **363**, 320–325
333. Hamm, P., Zurek, M., Mäntele, W., Meyer, M., Scheer, H., and Zinth, W. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1826–1830
334. Van Brederode, M. E., Jones, M. R., Van Mourik, F., Van Stokkum, I. H. M., and Van Grondelle, R. (1997) *Biochemistry* **36**, 6855–6861
335. Venturoli, G., Drepper, F., Williams, J. C., Allen, J. P., Lin, X., and Mathis, P. (1998) *Biophys. J.* **74**, 3226–3240
335a. van Brederode, M. E., van Stokkum, I. H. M., Katilius, E., van Mourik, F., Jones, M. R., and van Grondelle, R. (1999) *Biochemistry* **38**, 7545–7555
336. Cherepy, N. J., Shreve, A. P., Moore, L. J., Boxer, S. G., and Mathies, R. A. (1997) *Biochemistry* **36**, 8559–8566
337. Cua, A., Kirmaier, C., Holten, D., and Bocian, D. F. (1998) *Biochemistry* **37**, 6394–6401
338. Ivancich, A., Artz, K., Williams, J. C., Allen, J. P., and Mattioli, T. A. (1998) *Biochemistry* **37**, 11812–11820
338a. Takahashi, E., Wells, T. A., and Wraight, C. A. (2001) *Biochemistry* **40**, 1020–1028
338b. Kuglstatter, A., Ermler, U., Michel, H., Baciou, L., and Fritzsch, G. (2001) *Biochemistry* **40**, 4253–4260
338c. Eastman, J. E., Taguchi, A. K. W., Lin, S., Jackson, J. A., and Woodbury, N. W. (2000) *Biochemistry* **39**, 14787–14798
339. Tang, C.-K., Williams, J. C., Taguchi, A. K. W., Allen, J. P., and Woodbury, N. W. (1999) *Biochemistry* **38**, 8794–8799
340. Parson, W. W., Chu, Z. T., and Warshel, A. (1998) *Biophys. J.* **74**, 182–191
340a. Balabin, I. A., and Onuchic, J. N. (2000) *Science* **290**, 114–117
340b. Yakovlev, A. G., Shkuropatov, A. Y., and Shuvalov, V. A. (2002) *Biochemistry* **41**, 2667–2674
341. Dohse, B., Mathis, P., Wachtveitl, J., Laussermair, E., Iwata, S., Michel, H., and Oesterhelt, D. (1995) *Biochemistry* **34**, 11335–11343
342. Lin, X., Williams, J. C., Allen, J. P., and Mathis, P. (1994) *Biochemistry* **33**, 13517–13523
343. Stowell, M. H. B., McPhillips, T. M., Rees, D. C., Soltis, S. M., Abresch, E., and Feher, G. (1997) *Science* **276**, 812–816
343a. de Boer, A. L., Neerken, S., de Wijn, R., Permentier, H. P., Gast, P., Vijgenboom, E., and Hoff, A. J. (2002) *Biochemistry* **41**, 3081–3088
343b. Schulten, E. A. M., Matysik, J., Alia, Kiihne, S., Raap, J., Lugtenburg, J., Gast, P., Hoff, A. J., and de Groot, H. J. M. (2002) *Biochemistry* **41**, 8708–8717
344. Heller, B. A., Holten, D., and Kirmaier, C. (1995) *Science* **269**, 940–945
344a. Lin, S., Katilius, E., Haffa, A. L. M., Taguchi, A. K. W., and Woodbury, N. W. (2001) *Biochemistry* **40**, 13767–13773
344b. Purton, S., Stevens, D. R., Muhiuddin, I. P., Evans, M. C. W., Carter, S., Rigby, S. E. J., and Heathcote, P. (2001) *Biochemistry* **40**, 2167–2175
344c. Guergova-Kuras, M., Boudreaux, B., Joliot, A., Joliot, P., and Redding, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4437–4442
345. Nugent, J. H. A. (1984) *Trends Biochem. Sci.* **9**, 354–357
345a. Johnson, T. W., Zybailov, B., Jones, A. D., Bittl, R., Zech, S., Stehlik, D., Golbeck, J. H., and Chitnis, P. R. (2001) *J. Biol. Chem.* **276**, 39512–39521
346. Sinning, I. (1992) *Trends Biochem. Sci.* **17**, 150–154
346a. Witt, H., Schlodder, E., Teutloff, C., Niklas, J., Bordignon, E., Carbonera, D., Kohler, S., Labahn, A., and Lubitz, W. (2002) *Biochemistry* **41**, 8557–8569
347. Fish, L. E., Kuck, U., and Bogorad, L. (1985) *J. Biol. Chem.* **260**, 1413–1421
348. Breton, J., Nabedryk, E., and Leibl, W. (1999) *Biochemistry* **38**, 11585–11592
349. Francke, C., Permentier, H. P., Franken, E. M., Neerken, S., and Amesz, J. (1997) *Biochemistry* **36**, 14167–14172
350. Nitschke, W., and Rutherford, A. W. (1991) *Trends Biochem. Sci.* **16**, 241–245
351. Golbeck, J. H. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1642–1646
352. Karrasch, S., Typke, D., Walz, T., Miller, M., Tsiotis, G., and Engel, A. (1996) *J. Mol. Biol.* **262**, 336–348
353. Deisenhofer, J., and Norris, J. R., eds. (1993) *The Photosynthetic Reaction Center*, Vol. 1 and 2, Academic Press, San Diego, California
354. Blankenship, R. E., and Hartman, H. (1998) *Trends Biochem. Sci.* **23**, 94–97
355. Kruip, J., Chitnis, P. R., Lagoutte, B., Rögner, M., and Boekema, E. J. (1997) *J. Biol. Chem.* **272**, 17061–17069
356. Schubert, W.-D., Klukas, O., Krauss, N., Saenger, W., Fromme, P., and Witt, H. T. (1997) *J. Mol. Biol.* **272**, 741–769
356a. Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauß, N. (2001) *Nature (London)* **411**, 909–916
356b. Kühlbrandt, W. (2001) *Nature (London)* **411**, 896–899
357. Kitmitto, A., Mustafa, A. O., Holzenburg, A., and Ford, R. C. (1998) *J. Biol. Chem.* **273**, 29592–29599
357a. Boekema, E. J., Jensen, P. E., Schlodder, E., van Breemen, J. F. L., van Room, H., Scheller, H. V., and Dekker, J. P. (2001) *Biochemistry* **40**, 1029–1036
358. Morton, R. A. (1971) *Biol. Rev. Cambridge Philos. Soc.* **46**, 47–96
359. Threlfall, D. R., and Whistance, G. R. (1971) in *Aspects of Terpenoid Chemistry and Biochemistry* (Goodwin, T. W., ed), pp. 372–374, Academic Press, New York
360. Blankenship, R. E., and Prince, R. C. (1985) *Trends Biochem. Sci.* **10**, 382–384
361. Knox, R. S. (1969) *Biophys. J.* **9**, 1351–1362
362. Parson, W. W. (1978) *Photochemistry and Photobiology* **28**, 389–393
363. Yu, L., Zhao, J., Lu, W., Bryant, D. A., and Golbeck, J. H. (1993) *Biochemistry* **32**, 8251–8258
364. Pandini, V., Aliverti, A., and Zanetti, G. (1999) *Biochemistry* **38**, 10707–10713
365. Barth, P., Lagoutte, B., and Sétif, P. (1998) *Biochemistry* **37**, 16233–16241
366. Meimberg, K., Lagoutte, B., Bottin, H., and Mühlenhoff, U. (1998) *Biochemistry* **37**, 9759–9767
367. Bruns, C. M., and Karplus, P. A. (1995) *J. Mol. Biol.* **247**, 125–145
368. Medina, M., Martínez-Júlvez, M., Hurley, J. K., Tollin, G., and Gómez-Moreno, C. (1998) *Biochemistry* **37**, 2715–2728
369. Martínez-Júlvez, M., Hermoso, J., Hurley, J. K., Mayoral, T., Sanz-Aparicio, J., Tollin, G., Gómez-Moreno, C., and Medina, M. (1998) *Biochemistry* **37**, 17680–17691
370. Xiong, J., Subramaniam, S., and Govindjee. (1996) *Protein Sci.* **5**, 2054–2073
371. Svensson, B., Etchebest, C., Tuffery, P., van Kan, P., Smith, J., and Styring, S. (1996) *Biochemistry* **35**, 14486–14502
371a. Büchel, C., Morris, E., Orlova, E., and Barber, J. (2001) *J. Mol. Biol.* **312**, 371–379
371b. Kuhl, H., Kruip, J., Seidler, A., Krieger-Liszkay, A., Bünker, M., Bald, D., Scheidig, A. J., and Rögner, M. (2000) *J. Biol. Chem.* **275**, 20652–20659
371c. Shen, J.-R., and Kamiya, N. (2000) *Biochemistry* **39**, 14739–14744
371d. Zouni, A., Witt, H.-T., Kern, J., Fromme, P., Krauß, N., Saenger, W., and Orth, P. (2001) *Nature (London)* **409**, 739–743
371e. Dismukes, G. C. (2001) *Science* **292**, 447–448
371f. Nield, J., Kruse, O., Ruprecht, J., da Fonseca, P., Büchel, C., and Barber, J. (2000) *J. Biol. Chem.* **275**, 27940–27946
372. Schweitzer, R. H., and Brudvig, G. W. (1997) *Biochemistry* **36**, 11351–11359
373. Merry, S. A. P., Nixon, P. J., Barter, L. M. C., Schilstra, M., Porter, G., Barber, J., Durrant, J. R., and Klug, D. R. (1998) *Biochemistry* **37**, 17439–17447
374. Betts, S. D., Ross, J. R., Pichersky, E., and Yocum, C. F. (1997) *Biochemistry* **36**, 4047–4053
375. Enami, I., Kamo, M., Ohta, H., Takahashi, S., Miura, T., Kusayanagi, M., Tanabe, S., Kamei, A., Motoki, A., Hirano, M., Tomo, T., and Satoh, K. (1998) *J. Biol. Chem.* **273**, 4629–4634
376. Babcock, G. T. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10893–10895
376a. Shi, L.-X., Lorkovic, Z. J., Oelmüller, R., and Schröder, W. P. (2000) *J. Biol. Chem.* **275**, 37945–37950
376b. Boekema, E. J., van Breemen, J. F. L., van Roon, H., and Dekker, J. P. (2000) *J. Mol. Biol.* **301**, 1123–1133
376c. Mamedov, F., Stefansson, H., Albertsson, P.-Å., and Styring, S. (2000) *Biochemistry* **39**, 10478–10486
377. Kok, B., Forbush, B., and McGloin, M. (1970) *Photochem. and Photobiol.* **11**, 457–475
378. Radmer, R., and Kok, B. (1975) *Ann. Rev. Biochem.* **44**, 409–433
379. Murata, N., and Miyao, M. (1985) *Trends Biochem. Sci.* **10**, 122–124
380. Lydakis-Simantiris, N., Dorlet, P., Ghanotakis, D. F., and Babcock, G. T. (1998) *Biochemistry* **37**, 6427–6435
381. Schiller, H., Dittmer, J., Iuzzolino, L., Dörner, W., Meyer-Klaucke, W., Solé, V. A., Nolting, H.-F., and Dau, H. (1998) *Biochemistry* **37**, 7340–7350
382. Chu, H.-A., Gardner, M. T., O'Brien, J. P., and Babcock, G. T. (1999) *Biochemistry* **38**, 4533–4541
383. Noguchi, T., Inoue, Y., and Tang, X.-S. (1999) *Biochemistry* **38**, 10187–10195
384. Barry, B. A., and Babcock, G. T. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7099–7103
385. Diner, B. A., Force, D. A., Randall, D. W., and Britt, R. D. (1998) *Biochemistry* **37**, 17931–17943
386. Hays, A.-M. A., Vassiliev, I. R., Golbeck, J. H., and Debus, R. J. (1998) *Biochemistry* **37**, 11352–11365
387. Ahlbrink, R., Haumann, M., Cherepanov, D., Bögershausen, O., Mulkidjanian, A., and Junge, W. (1998) *Biochemistry* **37**, 1131–1142
388. Mamedov, F., Sayre, R. T., and Styring, S. (1998) *Biochemistry* **37**, 14245–14256

References

388a. Bernát, G., Morvaridi, F., Feyziyev, Y., and Styring, S. (2002) *Biochemistry* **41**, 5830–5843
389. Siegbahn, P. E. M., and Crabtree, R. H. (1999) *J. Am. Chem. Soc.* **121**, 117–127
389a. Schlodder, E., and Witt, H. T. (1999) *J. Biol. Chem.* **274**, 30387–30392
389b. Geijer, P., Morvaridi, F., and Styring, S. (2001) *Biochemistry* **40**, 10881–10891
390. Baldwin, M. J., and Pecoraro, V. L. (1996) *J. Am. Chem. Soc.* **118**, 11325–11326
391. Limburg, J., Vrettos, J. S., Liable-Sands, L. M., Rheingold, A. L., Crabtree, R. H., and Brudvig, G. W. (1999) *Science* **283**, 1524–1527
392. Hoganson, C. W., and Babcock, G. T. (1997) *Science* **277**, 1953–1956
392a. Hillier, W., and Babcock, G. T. (2001) *Biochemistry* **40**, 1503–1509
393. Wincencjusz, H., Yocum, C. F., and van Gorkom, H. J. (1999) *Biochemistry* **38**, 3719–3725
394. Haddy, A., Hatchell, J. A., Kimel, R. A., and Thomas, R. (1999) *Biochemistry* **38**, 6104–6110
395. Kühne, H., Szalai, V. A., and Brudvig, G. W. (1999) *Biochemistry* **38**, 6604–6613
396. Seidler, A., and Rutherford, A. W. (1996) *Biochemistry* **35**, 12104–12110
397. Ädelroth, P., Lindberg, K., and Andréasson, L.-E. (1995) *Biochemistry* **34**, 9021–9027
397a. Vrettos, J. S., Stone, D. A., and Brudvig, G. W. (2001) *Biochemistry* **40**, 7937–7945
398. Klimov, V. V., Hulsebosch, R. J., Allakhverdiev, S. I., Wincencjusz, H., van Gorkom, H. J., and Hoff, A. J. (1997) *Biochemistry* **36**, 16277–16281
398a. Baranov, S. V., Ananyev, G. M., Klimov, V. V., and Dismukes, G. C. (2000) *Biochemistry* **39**, 6060–6065
399. Keister, D. L., and Raveed, N. J. (1974) *J. Biol. Chem.* **249**, 6454–6458
400. Schubert, H., Kroon, B. M. A., and Matthijs, H. C. P. (1994) *J. Biol. Chem.* **269**, 7267–7272
401. Demmig-Adams, B., Gilmore, A. M., and Adams, W. W., III. (1996) *FASEB J.* **10**, 403–412
401a. Demming-Adams, B., and Adams, W. W., III. (2000) *Nature (London)* **403**, 371–374
401b. Li, X.-P., Björkman, O., Shih, C., Grossman, A. R., Rosenquist, M., Jansson, S., and Niyogi, K. K. (2000) *Nature (London)* **403**, 391–395
401c. Kagawa, T., Sakai, T., Suetsugu, N., Oikawa, K., Ishiguro, S., Kato, T., Tabata, S., Okada, K., and Wada, M. (2001) *Science* **291**, 2138–2141
402. Havaux, M., and Niyogi, K. K. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8762–8767
403. Limantara, L., Fujii, R., Zhang, J.-P., Kakuno, T., Hara, H., Kawamori, A., Yagura, T., Cogdell, R. J., and Koyama, Y. (1998) *Biochemistry* **37**, 17469–17486
404. Telfer, A., Dhami, S., Bishop, S. M., Phillips, D., and Barber, J. (1994) *Biochemistry* **33**, 14469–14474
405. Bugos, R. C., and Yamamoto, H. Y. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6320–6325
406. Niyogi, K. K., Björkman, O., and Grossman, A. R. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14162–14167
407. Ruban, A. V., Lee, P. J., Wentworth, M., Young, A. J., and Horton, P. (1999) *J. Biol. Chem.* **274**, 10458–10465
407a. Frank, H. A., Bautista, J. A., Josue, J. S., and Young, A. J. (2000) *Biochemistry* **39**, 2831–2837
407b. Jahns, P., Wehner, A., Paulsen, H., and Hobe, S. (2001) *J. Biol. Chem.* **276**, 22154–22159
408. Gilmore, A. M., Shinkarev, V. P., Hazlett, T. L., and Govindjee. (1998) *Biochemistry* **37**, 13582–13593
409. Hanley, J., Deligiannakis, Y., Pascal, A., Faller, P., and Rutherford, A. W. (1999) *Biochemistry* **38**, 8189–8195
410. Buser, C. A., Diner, B. A., and Brudvig, G. W. (1992) *Biochemistry* **31**, 11449–11459
411. Barber, J., and Andersson, B. (1992) *Trends Biochem. Sci.* **17**, 61–66
412. Sharma, J., Panico, M., Shipton, C. A., Nilsson, F., Morris, H. R., and Barber, J. (1997) *J. Biol. Chem.* **272**, 33158–33166
413. Hagman, Å., Shi, L.-X., Rintamäki, E., Andersson, B., and Schröder, W. P. (1997) *Biochemistry* **36**, 12666–12671
414. Krieger, A., Rutherford, A. W., Vass, I., and Hideg, É. (1998) *Biochemistry* **37**, 16262–16269
415. Campbell, D., Eriksson, M.-J., Öquist, G., Gustafsson, P., and Clarke, A. K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 364–369
415a. Rutherford, A. W., and Krieger-Liszkay, A. (2001) *Trends Biochem. Sci.* **26**, 648–653
416. Kobza, J., and Seemann, J. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3815–3819
417. Puente, P., Wei, N., and Deng, X. W. (1996) *EMBO J.* **15**, 3732–3743
418. Fluhr, R., Kuhlemeier, C., Nagy, F., and Chua, N.-H. (1986) *Science* **232**, 1106–1112
419. Moses, P. B., and Chua, N.-H. (1988) *Sci. Am.* **258**(Apr), 88–93
420. Klein, R. R., and Mullet, J. E. (1990) *J. Biol. Chem.* **265**, 1895–1902
421. Morelli, G., Nagy, F., Fraley, R. T., Roger, S. G., and Chua, N.-H. (1985) *Nature (London)* **315**, 200–204
422. Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A., and Cashmore, A. R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7089–7093
423. Inamine, G., Nash, B., Weissbach, H., and Brot, N. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5690–5694
424. Buchanan, B. B., and Schurmann, P. (1973) *Curr. Top. Cell. Regul.* **7**, 1–20
425. Wolosiuk, R. A., Ballicora, M. A., and Hagelin, K. (1993) *FASEB J.* **7**, 622–637
426. Kelly, G. (1982) *Trends Biochem. Sci.* **7**, 81–82
427. Johansson, K., Ramaswamy, S., Saarinen, M., Lemaire-Chamley, M., Issakidis-Bourguet, E., Miginiac-Maslow, M., and Eklund, H. (1999) *Biochemistry* **38**, 4319–4326
428. Hirasawa, M., Schürmann, P., Jacquot, J.-P., Manieri, W., Jacquot, P., Keryer, E., Hartman, F. C., and Knaff, D. B. (1999) *Biochemistry* **38**, 5200–5205
429. Staples, C. R., Gaymard, E., Stritt-Etter, A.-L., Telser, J., Hoffman, B. M., Schürmann, P., Knaff, D. B., and Johnson, M. K. (1998) *Biochemistry* **37**, 4612–4620
429a. Dai, S., Schwendtmayer, C., Schürmann, P., Ramaswamy, S., and Eklund, H. (2000) *Science* **287**, 655–658
429b. Rintamäki, E., Martinsuo, P., Pursiheimo, S., and Aro, E.-M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11644–11649
430. Cseke, C., Balogh, A., Wong, J. H., Buchanan, B. B., Stitt, M., Herzog, B., and Held, H. W. (1984) *Trends Biochem. Sci.* **9**, 533–535
431. Bligny, R., Gardestrom, P., Roby, C., and Douce, R. (1990) *J. Biol. Chem.* **265**, 1319–1326
432. Zelitch, I. (1975) *Science* **188**, 626–633
433. Sommerville, C. R., and Ogren, W. L. (1982) *Trends Biochem. Sci.* **7**, 171–174
434. Heber, V., and Krause, G. H. (1980) *Trends Biochem. Sci.* **5**, 32–34
435. Goyal, A., and Tolbert, N. E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3319–3324
436. Tolbert, N. E. (1973) *Curr. Top. Cell. Regul.* **7**, 21–50
437. Ho, C.-L., Noji, M., and Saito, K. (1999) *J. Biol. Chem.* **274**, 11007–11012
438. Givan, C. V., Joy, K. W., and Kleczkowski, L. A. (1988) *Trends Biochem. Sci.* **13**, 433–437
439. Kozaki, A., and Takeba, G. (1996) *Nature (London)* **384**, 557–560
440. Haag, E., and Renger, G. (1997) in *Bioenergetics* (Gräber, P., and Milazzo, G., eds), pp. 212–272, Birkhäuser Verlag, Basel
441. Burnell, J. N., and Hatch, M. D. (1985) *Trends Biochem. Sci.* **10**, 289–291
442. Moore, P. D. (1978) *Nature (London)* **272**, 400–401
442a. Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Freitag, H., and Edwards, G. E. (2001) *Nature (London)* **414**, 543–546
442b. Maeda, S.-i, Price, G. D., Badger, M. R., Enomoto, C., and Omata, T. (2000) *J. Biol. Chem.* **275**, 20551–20555
443. Winter, K., and Smith, J. A. C., eds. (1996) *Crassulacean Acid Metabolism Biochemistry, Ecophysiology and Evolution*, Springer, Berlin
444. Benemann, J. R., Berenson, J. A., Kaplan, N. O., and Kamen, M. D. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2317–2320
445. Benemann, J. R., and Weare, N. M. (1974) *Science* **184**, 174–175
446. Abrahamson, E. W., and Fager, R. S. (1973) *Curr. Top. Bioenerg.* **5**, 125–200
447. Wald, G., and Brown, P. K. (1965) *Cold Spring Harb. Symp. on Quant. Biol.* **30**, 346
448. Nathans, J. (1992) *Biochemistry* **31**, 4923–4931
449. Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M., and Oprian, D. D. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 599–603
450. Khorana, H. G. (1992) *J. Biol. Chem.* **267**, 1–4
451. Dratz, E. A., and Hargrave, P. A. (1983) *Trends Biochem. Sci.* **8**, 128–131
452. Oprian, D. D., Asenjo.AB, Lee, N., and Pelletier, S. L. (1991) *Biochemistry* **30**, 11367–11372
453. Nassal, M., Mogi, T., Karnik, S. S., and Khorana, H. G. (1987) *J. Biol. Chem.* **262**, 9264–9270
454. Henderson, R. (1975) *J. Mol. Biol.* **93**, 123–138
455. Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P., and Lanyi, J. K. (1999) *J. Mol. Biol.* **291**, 899–911
456. Baldwin, J. (1993) *EMBO J.* **12**, 1693–1703
457. Baldwin, J. M., Schertler, G. F. X., and Unger, V. M. (1997) *J. Mol. Biol.* **272**, 144–64
458. Pogozheva, I. D., Lomize, A. L., and Mosberg, H. I. (1997) *Biophys. J.* **72**, 1963–1985
459. Schertler, G. F. X., and Hargrave, P. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11578–11582
460. Herzyk, P., and Hubbard, R. E. (1998) *J. Mol. Biol.* **281**, 741–754
461. Barnidge, D. R., Dratz, E. A., Sunner, J., and Jesaitis, A. J. (1997) *Protein Sci.* **6**, 816–824
461a. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) *Science* **289**, 739–745
461b. Okada, T., Ernst, O. P., Palczewski, K., and Hofmann, K. P. (2001) *Trends Biochem. Sci.* **26**, 318–331
461c. Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., and Stenkamp, R. E. (2001) *Biochemistry* **40**, 7761–7772
462. Khorana, H. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1166–1171
462a. Altenbach, C., Klein-Seetharaman, J., Cai, K., Khorana, H. G., and Hubbell, W. L. (2001) *Biochemistry* **40**, 15493–15500
463. Yeagle, P. L., Alderfer, J. L., and Albert, A. D. (1995) *Biochemistry* **34**, 14621–14625
464. Altenbach, C., Klein-Seetharaman, J., Hwa, J., Khorana, H. G., and Hubbell, W. L. (1999) *Biochemistry* **38**, 7945–7949
465. Struthers, M., Yu, H., Kono, M., and Oprian, D. D. (1999) *Biochemistry* **38**, 6597–6603
466. Zhang, H., Lerro, K. A., Yamamoto, T., Lien, T. H., Sastry, L., Gawinowicz, M. A., and Nakanishi, K. (1994) *J. Am. Chem. Soc.* **116**, 10165–10173

References

467. Han, M., and Smith, S. O. (1995) *Biochemistry* **34**, 1425–1432
468. Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M., and Henderson, R. (1996) *J. Mol. Biol.* **259**, 393–421
469. Wald, G. (1968) *Nature (London)* **219**, 800–807
470. Dowling, J. E. (1997) *Nature (London)* **387**, 356
471. Gilardi, R., Sperling, W., Karle, I. L., and Karle, J. (1971) *Nature (London)* **232**, 187–188
472. Shriver, J. W., Mateescu, G. D., and Abrahamson, E. W. (1979) *Biochemistry* **18**, 4785–4792
472a. Singh, D., Hudson, B. S., Middleton, C., and Birge, R. R. (2001) *Biochemistry* **40**, 4201–4204
473. Honig, B., Dinur, U., Nakanishi, K., Balogh-Nair, V., Gawinswicz, M. A., Arnaboldi, M., and Motto, M. G. (1979) *J. Am. Chem. Soc.* **101**, 7084–7086
474. Baasov, T., and Sheves, M. (1985) *J. Am. Chem. Soc.* **107**, 7524–7533
475. Sakmar, T. P., Franke, R. R., and Khorana, H. G. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8309–8313
475a. Lewis, J. W., Szundi, I., Fu, W.–Y., Sakmar, T. P., and Kliger, D. S. (2000) *Biochemistry* **39**, 599–606
476. Creemers, A. F. L., Klaassen, C. H. W., Bovee-Geurts, P. H. M., Kelle, R., Kragl, U., Raap, J., de Grip, W. J., Lugtenburg, J., and de Groot, H. J. M. (1999) *Biochemistry* **38**, 7195–7199
476a. Verhoeven, M. A., Creemers, A. F. L., Bovee-Geurts, P. H. M., De Grip, W. J., Lugtenburg, J., and de Groot, H. J. M. (2001) *Biochemistry* **40**, 3282–3288
477. Kochendoerfer, G. G., Lin, S. W., Sakmar, T. P., and Mathies, R. A. (1999) *Trends Biochem. Sci.* **24**, 300–305
478. Nathans, J., Thomas, D., and Hogness, D. S. (1986) *Science* **232**, 193–202
479. Kochendoerfer, G. G., Wang, Z., Oprian, D. D., and Mathies, R. A. (1997) *Biochemistry* **36**, 6577–6587
479a. Dukkipati, A., Vought, B. W., Singh, D., Birge, R. R., and Knox, B. E. (2001) *Biochemistry* **40**, 15098–15108
479b. Roorda, A., and Williams, D. R. (1999) *Nature (London)* **397**, 520–522
480. Nathans, J., Piantanida, T. D., Eddy, R. L., Shows, T. B., and Hogness, D. S. (1986) *Science* **232**, 203–210
481. Motulsky, A. G., and Deeb, S. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 3 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 4275–4295, McGraw-Hill, New York
482. Nathans, J. (1989) *Sci. Am.* **260**(Feb), 42–49
483. Merbs, S. L., and Nathans, J. (1992) *Science* **258**, 464–466
484. Hunt, D. M., Dulai, K. S., Bowmaker, J. K., and Mollon, J. D. (1995) *Science* **267**, 984–988
485. Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., Klingaman, R., and Fishman, G. (1989) *Science* **245**, 831–838
486. Shichi, H., Lewis, M. S., Irreverre, F. and Stone, A. L. (1969) *J. Biol. Chem.* **244**, 529–536
487. Shichi, H., and Somers, R. L. (1974) *J. Biol. Chem.* **249**, 6570–6577
488. Fasick, J. I., Lee, N., and Oprian, D. D. (1999) *Biochemistry* **38**, 11593–11596
488a. Shi, Y., Radlwimmer, F. B., and Yokoyama, S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11731–11736
488b. Janz, J. M., and Farrens, D. L. (2001) *Biochemistry* **40**, 7219–7227
488c. Dukkipati, A., Kusnetzow, A., Babu, K. R., Ramos, L., Singh, D., Knox, B. E., and Birge, R. R. (2002) *Biochemistry* **41**, 9842–9851
489. Fasick, J. I., and Robinson, P. R. (1998) *Biochemistry* **37**, 433–438
490. Imamoto, Y., Hirano, T., Imai, H., Kandori, H., Maeda, A., Yoshizawa, T., Groesbeek, M., Lugtenburg, J., and Shichida, Y. (1999) *Biochemistry* **38**, 11749–11754
490a. Imai, H., Hirano, T., Kandori, H., Terakita, A., and Shichida, Y. (2001) *Biochemistry* **40**, 2879–2886
491. Wang, Z., Asenjo, A. B., and Oprian, D. D. (1993) *Biochemistry* **32**, 2125–2130
492. Johnson, R. L., Grant, K. B., Zankel, T. C., Boehm, M. F., Merbs, S. L., Nathans, J., and Nakanishi, K. (1993) *Biochemistry* **32**, 208–214
493. Foster, K. W., Saranak, J., Derguini, F., Zarrilli, G. R., Johnson, R., Okabe, M., and Nakanishi, K. (1989) *Biochemistry* **28**, 819–824
494. Deininger, W., Kröger, P., Hegemann, U., Lottspeich, F., and Hegemann, P. (1995) *EMBO J.* **14**, 5849–5858
495. Max, M., Surya, A., Takahashi, J. S., Margolskee, R. F., and Knox, B. E. (1998) *J. Biol. Chem.* **273**, 26820–26826
496. Max, M., McKinnon, P. J., Seidenman, K. J., Barrett, R. K., Applebury, M. L., Takahashi, J. S., and Margolskee, R. F. (1995) *Science* **267**, 1502–1506
496a. Nakamura, A., Kojima, D., Imai, H., Terakita, A., Okano, T., Shichida, Y., and Fukada, Y. (1999) *Biochemistry* **38**, 14738–14745
497. Soni, B. G., Philp, A. R., Foster, R. G., and Knox, B. E. (1998) *Nature (London)* **394**, 27–28
498. Zuker, C. S. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 571–576
499. Peters, K., Applebury, M. L., and Rentzepis, P. M. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3119–3123
500. Strassburger, J. M., Gärtner, W., and Braslavsky, S. E. (1997) *Biophys. J.* **72**, 2294–2303
501. Zhu, Y., and Liu, R. S. H. (1993) *Biochemistry* **32**, 10233–10238
502. Liu, R. S. H., and Shichida, Y. (1991) in *Photochemistry in Organized and Constrained Media* (Ramamurthy, V., ed), VCH Publishers, New York (Chapter 18)
502a. Ishiguro, M. (2000) *J. Am. Chem. Soc.* **122**, 444–451
503. Hug, S. J., Lewis, W. J., Einterz, C. M., Thorgeirsson, T. E., and Kliger, D. S. (1990) *Biochemistry* **29**, 1475–1485
504. Lewis, J. W., Pinkas, I., Sheves, M., Ottolenghi, M., and Kliger, D. S. (1995) *J. Am. Chem. Soc.* **117**, 918–923
505. Shichida, Y., Nakamura, K., Yoshizawa, T., Trehan, A., Denny, M., and Liu, R. S. H. (1988) *Biochemistry* **27**, 6495–6499
506. Wang, Q., Schoenlein, R. W., Peteanu, L. A., Mathies, R. A., and Shank, C. V. (1994) *Science* **266**, 422–424
507. Jäger, F., Fahmy, K., Sakmar, T. P., and Siebert, F. (1994) *Biochemistry* **33**, 10878–10882
508. DeLange, F., Bovee-Geurts, P. H. M., VanOostrum, J., Portier, M. D., Verdegem, P. J. E., Lugtenburg, J., and DeGrip, W. J. (1998) *Biochemistry* **37**, 1411–1420
509. Fukada, Y., Shichida, Y., Yoshizawa, T., Ito, M., Kodama, A., and Tsukida, K. (1984) *Biochemistry* **23**, 5826–5832
510. Bagley, K. A., Balogh-Nair, V., Croteau, A. A., Dollinger, G., Ebrey, T. G., Eisenstein, L., Hong, M. K., Nakanishi, K., and Vittitow, J. (1985) *Biochemistry* **24**, 6055–6071
510a. Borhan, B., Souto, M. L., Imai, H., Shichida, Y., and Nakanishi, K. (2000) *Science* **288**, 2209–2212
511. Ridge, K. D., Bhattacharya, S., Nakayama, T. A., and Khorana, H. G. (1992) *J. Biol. Chem.* **267**, 6770–6775
511a. Liu, R. S. H., and Hammond, G. S. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11153–11158
511b. González-Luque, R., Garavelli, M., Bernardi, F., Merchán, M., and Robb, M. A. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9379–9384
511c. Kim, J. E., Tauber, M. J., and Mathies, R. A. (2001) *Biochemistry* **40**, 13774–13778
512. Chosrowjan, H., Mataga, N., Shibata, Y., Tachibanaki, S., Kandori, H., Shichida, Y., Okada, T., and Kouyama, T. (1998) *J. Am. Chem. Soc.* **120**, 9706–9707
513. Loppnow, G. R., Miley, M. E., Mathies, R. A., Liu, R. S. H., Kandori, H., Shichida, Y., Fukada, Y., and Yoshizawa, T. (1990) *Biochemistry* **29**, 8985–8991
514. Freedman, K. A., and Becker, R. S. (1986) *J. Am. Chem. Soc.* **108**, 1245–1251
515. Yoshizawa, T., and Wald, G. (1963) *Nature (London)* **197**, 1279–1286
516. Nagata, T., Terakita, A., Kandori, H., Kojima, D., Shichida, Y., and Maeda, A. (1997) *Biochemistry* **36**, 6164–6170
517. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L., and Khorana, H. G. (1996) *Science* **274**, 768–770
518. Dunham, T. D., and Farrens, D. L. (1999) *J. Biol. Chem.* **274**, 1683–1690
518a. Bartl, F. J., Ritter, E., and Hofmann, K. P. (2001) *J. Biol. Chem.* **276**, 30161–30166
518b. Vogel, R., Fan, G.-B., Siebert, F., and Sheves, M. (2001) *Biochemistry* **40**, 13342–13352
519. Dickopf, S., Mielke, T., and Heyn, M. P. (1998) *Biochemistry* **37**, 16888–16897
520. Jäger, S., Szundi, I., Lewis, J. W., Mah, T. L., and Kliger, D. S. (1998) *Biochemistry* **37**, 6998–7005
521. Parkes, J. H., and Liebman, P. A. (1984) *Biochemistry* **23**, 5054–5061
521a. Yeagle, P. L., Choi, G., and Albert, A. D. (2001) *Biochemistry* **40**, 11932–11937
521b. Vought, B. W., Salcedo, E., Chadwell, L. V., Britt, S. G., Birge, R. R., and Knox, B. E. (2000) *Biochemistry* **39**, 14128–14137
521c. Pan, D., and Mathies, R. A. (2001) *Biochemistry* **40**, 7929–7936
522. Yarfitz, S., and Hurley, J. B. (1994) *J. Biol. Chem.* **269**, 14329–14332
523. Puckett, K. L., Aronson, E. T., and Goldin, S. M. (1985) *Biochemistry* **24**, 390–400
524. Dizhoor, A. M., Olshevskaya, E. V., Henzel, W. J., Wong, S. C., Stults, J. T., Ankoudinova, I., and Hurley, J. B. (1995) *J. Biol. Chem.* **270**, 25200–25206
525. Yau, K.-W., and Nakatani, K. (1985) *Nature (London)* **313**, 579–583
526. Stryer, L., Hurley, J. B., and Fung, B. K. K. (1981) *Trends Biochem. Sci.* **6**, 245–247
527. Stryer, L. (1987) *Sci. Am.* **257**(Jul), 42–50
528. Stryer, L. (1991) *J. Biol. Chem.* **266**, 10711–10714
529. Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E., and Sigler, P. B. (1996) *Nature (London)* **379**, 311–319
529a. Granovsky, A. E., and Artemyev, N. O. (2001) *Biochemistry* **40**, 13209–13215
529b. Norton, A. W., D'Amours, M. R., Grazio, H. J., Hebert, T. L., and Cote, R. H. (2000) *J. Biol. Chem.* **275**, 38611–38619
530. Shimoda, Y., Hurley, J. B., and Miller, W. H. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 616–619
531. Gibson, S. K., Parkes, J. H., and Liebman, P. A. (1998) *Biochemistry* **37**, 11393–11398
532. Palczewski, K., Buczylko, J., Lebioda, L., Crabb, J. W., and Polans, A. S. (1993) *J. Biol. Chem.* **268**, 6004–6013
533. Gurevich, V. V., Chen, C.-Y., Kim, C. M., and Benovic, J. L. (1994) *J. Biol. Chem.* **269**, 8721–8727

References

534. Smith, W. C., McDowell, J. H., Dugger, D. R., Miller, R., Arendt, A., Popp, M. P., and Hargrave, P. A. (1999) *Biochemistry* **38**, 2752–2761
535. Granzin, J., Wilden, U., Choe, H.-W., Labahn, J., Krafft, B., and Büldt, G. (1998) *Nature (London)* **391**, 918–921
535a. Oakley, R. H., Laporte, S. A., Holt, J. A., Caron, M. G., and Barak, L. S. (2000) *J. Biol. Chem.* **275**, 17201–17210
536. Baylor, D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 560–565
536a. Penn, R. B., Pascual, R. M., Kim, Y.-M., Mundell, S. J., Krymskaya, V. P., Panettieri, R. A., Jr., and Benovic, J. L. (2001) *J. Biol. Chem.* **276**, 32648–32656
537. Laitko, U., and Hofmann, K. P. (1998) *Biophys. J.* **74**, 803–815
538. Klenchin, V. A., Calvert, P. D., and Bownds, M. D. (1995) *J. Biol. Chem.* **270**, 16147–16152
539. Ames, J. B., Ishima, R., Tanaka, T., Gordon, J. I., Stryer, L., and Ikura, M. (1997) *Nature (London)* **389**, 198–202
540. Ames, J. B., Tanaka, T., Stryer, L., and Ikura, M. (1996) *Curr. Opin. Struct. Biol.* **6**, 432–438
541. Lange, C., and Koch, K.-W. (1997) *Biochemistry* **36**, 12019–12026
542. Palczewski, K., Jäger, S., Buczylko, J., Crouch, R. K., Bredberg, D. L., Hofmann, K. P., Asson-Batres, M. A., and Saari, J. C. (1994) *Biochemistry* **33**, 13741–13750
543. Palczewski, K., Van Hooser, J. P., Garwin, G. G., Chen, J., Liou, G. I., and Saari, J. C. (1999) *Biochemistry* **38**, 12012–12019
543a. McBee, J. K., Kuksa, V., Alvarez, R., de Lera, A. R., Prezhdo, O., Haeseleer, F., Sokal, I., and Palczewski, K. (2000) *Biochemistry* **39**, 11370–11380
544. Hara, R., Hara, T., Tokunaga, F., and Yoshizawa, T. (1981) *Photochem. Photobiol.* **33**, 883–891
544a. Terakita, A., Yamashita, T., and Shichida, Y. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14263–14267
545. Humphries, P., Kenna, P., and Farrar, G. J. (1992) *Science* **256**, 804–808
546. Dryja, T. P. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 3 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 4297–4309, McGraw-Hill, New York
546a. Hwa, J., Klein-Seetharaman, J., and Khorana, H. G. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4872–4876
547. Maniloff, J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10004–10006
548. Liu, X., Garriga, P., and Khorana, H. G. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4554–4559
549. Rao, V. R., Cohen, G. B., and Oprian, D. D. (1994) *Nature (London)* **367**, 639–642
550. Kajiwara, K., Berson, E. L., and Dryja, T. P. (1994) *Science* **264**, 1604–1608
551. Khani, S. C., Nielsen, L., and Vogt, T. M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2824–2827
552. Tucker, C. L., Woodcock, S. C., Kelsell, R. E., Ramamurthy, V., Hunt, D. M., and Hurley, J. B. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9039–9044
553. Eudy, J. D., Weston, M. D., Yao, S., Hoover, D. M., Rehm, H. L., Ma-Edmonds, M., Yan, D., Ahmad, I., Cheng, J. J., Ayuso, C., Cremers, C., Davenport, S., Moller, C., Talmadge, C. B., Beisel, K. W., Tamayo, M., Morton, C. C., Swaroop, A., Kimberling, W. J., and Sumegi, J. (1998) *Science* **280**, 1753–1757
554. Seabra, M. C., Brown, M. S., and Goldstein, J. L. (1993) *Science* **259**, 377–381
555. Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., Peiffer, A., Zabriskie, N. A., Li, Y., Hutchinson, A., Dean, M., Lupski, J. R., and Leppert, M. (1997) *Science* **277**, 1805–1807
556. Pennisi, E. (1998) *Science* **281**, 31
556a. Biswas, E. E. (2001) *Biochemistry* **40**, 8181–8187
557. Parish, C. A., Hashimoto, M., Nakanishi, K., Dillon, J., and Sparrow, J. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14609–14613
558. Norledge, B. V., Trinkl, S., Jaenicke, R., and Slingsby, C. (1997) *Protein Sci.* **6**, 1612–1620
559. Koretz, J. F., and Handelman, G. H. (1988) *Sci. Am.* **259**(Jul), 92–99
560. Tardieu, A., and Delaye, M. (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 47–70
561. Wistow, G. (1993) *Trends Biochem. Sci.* **18**, 301–306
562. Hejtmancik, J. F., Kaiser, M. I., and Piatigorsky, J. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 3 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 4325–4349, McGraw-Hill, New York
563. König, N., Zampighi, G. A., and Butler, P. J. G. (1997) *J. Mol. Biol.* **265**, 590–602
564. Surewicz, W. K., and Olesen, P. R. (1995) *Biochemistry* **34**, 9655–9660
565. Clark, J. I., and Huang, Q.-L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15185–15189
566. Piatigorsky, J., and Wistow, G. (1991) *Science* **252**, 1078–1079
567. Röll, B., Amons, R., and de Jong, W. W. (1996) *J. Biol. Chem.* **271**, 10437–10440
568. Zigler, J. S., Jr., and Rao, P. V. (1991) *FASEB J.* **5**, 223–225
569. Abu-Abed, M., Turner, M. A., Vallée, F., Simpson, A., Slingsby, C., and Howell, P. L. (1997) *Biochemistry* **36**, 14012–14022
570. Kraft, H. J., Hendriks, W., de Jong, W. W., Lubsen, N. H., and Schoenmakers, J. G. G. (1993) *J. Mol. Biol.* **229**, 849–859
571. Mulders, J. W. M., Hendriks, W., Blankesteijn, W. M., Bloemendal, H., and de Jong, W. W. (1988) *J. Biol. Chem.* **263**, 15462–15466
572. Zinovieva, R. D., Tomarev, S. I., and Piatigorsky, J. (1993) *J. Biol. Chem.* **268**, 11449–11455
572a. Werten, P. J. L., Röll, B., van Aalten, D. M. F., and de Jong, W. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3282–3287
573. Garner, B., Vazquez, S., Griffith, R., Lindner, R. A., Carver, J. A., and Truscott, R. J. W. (1999) *J. Biol. Chem.* **274**, 20847–20854
574. Tessier, F., Obrenovich, M., and Monnier, V. M. (1999) *J. Biol. Chem.* **274**, 20796–20804
574a. Hood, B. D., Garner, B., and Truscott, R. J. W. (1999) *J. Biol. Chem.* **274**, 32547–32550
575. Takemoto, L., and Boyle, D. (1998) *Biochemistry* **37**, 13681–13685
576. Stoeckenius, W. (1999) *Protein Sci.* **8**, 447–459
577. Racker, E., and Stoeckenius, W. (1974) *J. Biol. Chem.* **249**, 662–663
578. Stoeckenius, W. (1985) *Trends Biochem. Sci.* **10**, 483–486
579. Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y., and Fujiyoshi, Y. (1999) *J. Mol. Biol.* **286**, 861–882
580. Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T., and Fujiyoshi, Y. (1997) *Nature (London)* **389**, 206–211
581. Subramaniam, S., Lindahl, M., Bullough, P., Faruqi, A. R., Tittor, J., Oesterhelt, D., Brown, L., Lanyi, J., and Henderson, R. (1999) *J. Mol. Biol.* **287**, 145–161
582. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., and Landau, E. M. (1997) *Science* **277**, 1676–1681
582a. Subramaniam, S., and Henderson, R. (2000) *Nature (London)* **406**, 653–657
582b. Edman, K., Nollert, P., Royant, A., Belrhali, H., Pebay-Peyroula, E., Hajdu, J., Neutze, R., and Landau, E. M. (1999) *Nature (London)* **401**, 822–826
582c. Royant, A., Edman, K., Ursby, T., Pebay-Peyroula, E., Landau, E. M., and Neutze, R. (2000) *Nature (London)* **406**, 645–648
582d. Luecke, H., Schobert, B., Cartailler, J.-P., Richter, H.-T., Rosengarth, A., Needleman, R., and Lanyi, J. K. (2000) *J. Mol. Biol.* **300**, 1237–1255
582e. Xiao, W., Brown, L. S., Needleman, R., Lanyi, J. K., and Shin, Y.-K. (2000) *J. Mol. Biol.* **304**, 715–721
582f. Kühlbrandt, W. (2000) *Nature (London)* **406**, 569–570
583. Müller, D. J., Sass, H.-J., Müller, S. A., Büldt, G., and Engel, A. (1999) *J. Mol. Biol.* **285**, 1903–1909
584. Smith, S. O., Myers, A. B., Pardoen, J. A., Winkel, C., Mulder, P. P. J., Lugtenburg, J., and Mathies, R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2055–2059
584a. Patzelt, H., Simon, B., terLaak, A., Kessler, B., Kühne, R., Schmieder, P., Oesterhelt, D., and Oschkinat, H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9765–9770
585. Bullough, P. A., and Henderson, R. (1999) *J. Mol. Biol.* **286**, 1663–1671
585a. Herbst, J., Heyne, K., and Diller, R. (2002) *Science* **297**, 822–825
585b. Kandori, H., Belenky, M., and Herzfeld, J. (2002) *Biochemistry* **41**, 6026–6031
585c. Maeda, A., Balashov, S. P., Lugtenburg, J., Verhoeven, M. A., Herzfeld, J., Belenky, M., Gennis, R. B., Tomson, F. L., and Ebrey, T. G. (2002) *Biochemistry* **41**, 3803–3809
586. Haupts, U., Tittor, J., Bamberg, E., and Oesterhelt, D. (1997) *Biochemistry* **36**, 2–7
587. Gat, Y., and Sheves, M. (1993) *J. Am. Chem. Soc.* **115**, 3772–3773
588. Moltke, S., Wallat, I., Sakai, N., Nakanishi, K., Brown, M. F., and Heyn, M. P. (1999) *Biochemistry* **38**, 11762–11772
588a. Sass, H. J., Büldt, G., Gessenich, R., Hehn, D., Neff, D., Schlesinger, R., Berendzen, J., and Ormos, P. (2000) *Nature (London)* **406**, 649–653
588b. Kandori, H., Yamazaki, Y., Shichida, Y., Raap, J., Lugtenburg, J., Belenky, M., and Herzfeld, J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1571–1576
588c. Dioumaev, A. K., Brown, L. S., Needleman, R., and Lanyi, J. K. (2001) *Biochemistry* **40**, 11308–11317
588d. Oka, T., Yagi, N., Fujisawa, T., Kamikubo, H., Tokunaga, F., and Kataoka, M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14278–14282
588e. Vonck, J. (2000) *EMBO J.* **19**, 2152–2160
588f. Rouhani, S., Cartailler, J.-P., Facciotti, M. T., Walian, P., Needleman, R., Lanyi, J. K., Glaeser, R. M., and Luecke, H. (2001) *J. Mol. Biol.* **313**, 615–628
588g. Allen, S. J., Kim, J.-M., Khorana, H. G., Lu, H., and Booth, P. J. (2001) *J. Mol. Biol.* **308**, 423–435
588h. Aharoni, A., Weiner, L., Lewis, A., Ottolenghi, M., and Sheves, M. (2001) *J. Am. Chem. Soc.* **123**, 6612–6616
588i. Spassov, V. Z., Luecke, H., Gerwert, K., and Bashford, D. (2001) *J. Mol. Biol.* **312**, 203–219
588j. Imasheva, E. S., Lu, M., Balashov, S. P., Ebrey, T. G., Chen, Y., Ablonczy, Z., Menick, D. R., and Crouch, R. K. (2001) *Biochemistry* **40**, 13320–13330
589. Havelka, W. A., Henderson, R., and Oesterhelt, D. (1995) *J. Mol. Biol.* **247**, 726–738

References

590. Chon, Y.-S., Kandori, H., Sasaki, J., Lanyi, J. K., Needleman, R., and Maeda, A. (1999) *Biochemistry* **38**, 9449–9455
590a. Sato, M., Kanamori, T., Kamo, N., Demura, M., and Nitta, K. (2002) *Biochemistry* **41**, 2452–2458
591. Seiff, F., Wallat, I., Ermann, P., and Heyn, M. P. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3227–3231
591a. Kandori, H., Furutani, Y., Shimono, K., Shichida, Y., and Kamo, N. (2001) *Biochemistry* **40**, 15693–15698
591b. Swartz, T. E., Szundi, I., Spudich, J. L., and Bogomolni, R. A. (2000) *Biochemistry* **39**, 15101–15109
591c. Wegener, A.-A., Klare, J. P., Engelhard, M., and Steinhoff, H.-J. (2001) *EMBO J.* **20**, 5312–5319
592. Jung, K.-H., and Spudich, J. L. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6557–6561
593. Zhang, X.-N., and Spudich, J. L. (1997) *Biophys. J.* **73**, 1516–1523
593a. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N., and Spudich, J. L. (2001) *Science* **293**, 1499–1503
593b. Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E. M., Pebay-Peyroula, E., and Navarro, J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10131–10136
594. Hirayama, J., Imamoto, Y., Shichida, Y., Kamo, N., Tomioka, H., and Yoshizawa, T. (1992) *Biochemistry* **31**, 2093–2098
595. Spudich, E. N., Zhang, W., Alam, M., and Spudich, J. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4960–4965
596. Chizhov, I., Schmies, G., Seidel, R., Sydor, J. R., Lüttenberg, B., and Engelhard, M. (1998) *Biophys. J.* **75**, 999–1009
597. Rath, P., Olson, K. D., Spudich, J. L., and Rothschild, K. J. (1994) *Biochemistry* **33**, 5600–5606
597a. Bergo, V., Spudich, E. N., Scott, K. L., Spudich, J. L., and Rothschild, K. J. (2000) *Biochemistry* **39**, 2823–2830
597b. Ren, L., Martin, C. H., Wise, K. J., Gillespie, N. B., Luecke, H., Lanyi, J. K., Spudich, J. L., and Birge, R. R. (2001) *Biochemistry* **40**, 13906–13914
598. Bieszke, J. A., Braun, E. L., Bean, L. E., Kang, S., Natvig, D. O., and Borkovich, K. A. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8034–8039
598a. Sineshchekov, O. A., Jung, K.-H, and Spudich, J. L. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8689–8694
599. Hoff, W. D., Devreese, B., Fokkens, R., Nugteren-Roodzant, I. M., Van Beeumen, J., Nibbering, N., and Hellingwerf, K. J. (1996) *Biochemistry* **35**, 1274–1281
599a. Xie, A., Kelemen, L., Hendriks, J., White, B. J., Hellingwerf, K. J., and Hoff, W. D. (2001) *Biochemistry* **40**, 1510–1517
600. Kort, R., Hoff, W. D., Van West, M., Kroon, A. R., Hoffer, S. M., Vlieg, K. H., Crielaard, W., Van Beeumen, J. J., and Hellingwerf, K. J. (1996) *EMBO J.* **15**, 3209–3218
601. Borgstahl, G. E. O., Williams, D. R., and Getzoff, E. D. (1995) *Biochemistry* **34**, 6278–6287
601a. Brudler, R., Meyer, T. E., Genick, U. K., Devanathan, S., Woo, T. T., Millar, D. P., Gerwert, K., Cusanovich, M. A., Tollin, G., and Getzoff, E. D. (2000) *Biochemistry* **39**, 13478–13486
601b. Ren, Z., Perman, B., Srajer, V., Teng, T.-Y., Pradervand, C., Bourgeois, D., Schotte, F., Ursby, T., Kort, R., Wulff, M., and Moffat, K. (2001) *Biochemistry* **40**, 13788–13801
601c. Imamoto, Y., Kataoka, M., Tokunaga, F., Asahi, T., and Masuhara, H. (2001) *Biochemistry* **40**, 6047–6052
602. Imamoto, Y., Mihara, K., Hisatomi, O., Kataoka, M., Tokunaga, F., Bojkova, N., and Yoshihara, K. (1997) *J. Biol. Chem.* **272**, 12905–12908
603. Perman, B., Srajer, V., Ren, Z., Teng, T., Pradervand, C., Ursby, T., Bourgeois, D., Schotte, F., Wulff, M., Kort, R., Hellingwerf, K., and Moffat, K. (1998) *Science* **279**, 1946–1950
604. Genick, U. K., Devanathan, S., Meyer, T. E., Canestrelli, I. L., Williams, E., Cusanovich, M. A., Tollin, G., and Getzoff, E. D. (1997) *Biochemistry* **36**, 8–14
605. Ujj, L., Devanathan, S., Meyer, T. E., Cusanovich, M. A., Tollin, G., and Atkinson, G. H. (1998) *Biophys. J.* **75**, 406–412
606. Imamoto, Y., Kataoka, M., and Tokunaga, F. (1996) *Biochemistry* **35**, 14047–14053
607. Genick, U. K., Soltis, S. M., Kuhn, P., Canestrelli, I. L., and Getzoff, E. D. (1998) *Nature (London)* **392**, 206–209
608. Devanathan, S., Genick, U. K., Canestrelli, I. L., Meyer, T. E., Cusanovich, M. A., Getzoff, E. D., and Tollin, G. (1998) *Biochemistry* **37**, 11563–11568
609. Xie, A., Hoff, W. D., Kroon, A. R., and Hellingwerf, K. J. (1996) *Biochemistry* **35**, 14671–14678
609a. Unno, M., Kumauchi, M., Sasaki, J., Tokunaga, F., and Yamauchi, S. (2002) *Biochemistry* **41**, 5668–5674
610. Tao, N., Orlando, M., Hyon, J.-S., Gross, M., and Song, P.-S. (1993) *J. Am. Chem. Soc.* **115**, 2526–2528
611. Song, P.-S. (1983) *Annu Rev Biophys Bioeng.* **12**, 35–68
612. Smirnov, A., Fulton, D. B., Andreotti, A., and Petrich, J. W. (1999) *J. Am. Chem. Soc.* **121**, 7979–7988
613. Carpenter, S., Fehr, M. J., Kraus, G. A., and Petrich, J. W. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12273–12277
614. Quail, P. H. (1984) *Trends Biochem. Sci.* **9**, 450–453
615. Kelly, J. M., and Lagarias, J. C. (1985) *Biochemistry* **24**, 6003–6010
616. Quail, P. H. (1976) in *Plant Biochemistry*, 3rd ed. (Bonner, J., and Varner, J. E., eds), pp. 683–711, Academic Press, New York
617. Anderson, G. R., Jenner, E. L., and Mumford, F. E. (1970) *Biochim. Biophys. Acta.* **221**, 69–73
618. Quail, P. H., Boylan, M. T., Parks, B. M., Short, T. W., Xu, Y., and Wagner, D. (1995) *Science* **268**, 675–680
619. Ruddat, A., Schmidt, P., Gatz, C., Braslavsky, S. E., Gärtner, W., and Schaffner, K. (1997) *Biochemistry* **36**, 103–111
619a. Fankhauser, C. (2001) *J. Biol. Chem.* **276**, 11453–11456
619b. Nagy, F., and Schäfer, E. (2000) *EMBO J.* **19**, 157–163
619c. Smith, H. (2000) *Nature (London)* **407**, 585–591
620. Foerstendorf, H., Mummert, E., Schäfer, E., Scheer, H., and Siebert, F. (1996) *Biochemistry* **35**, 10793–10799
621. Andel, F., III, Lagarias, J. C., and Mathies, R. A. (1996) *Biochemistry* **35**, 15997–16008
622. Chen, E., Lapko, V. N., Lewis, J. W., Song, P.-S., and Kliger, D. S. (1996) *Biochemistry* **35**, 843–850
623. Kneip, C., Hildebrandt, P., Schlamann, W., Braslavsky, S. E., Mark, F., and Schaffner, K. (1999) *Biochemistry* **38**, 15185–15192
623a. Foerstendorf, H., Benda, C., Gärtner, W., Storf, M., Scheer, H., and Siebert, F. (2001) *Biochemistry* **40**, 14952–14959
623b. Andel, F., III, Murphy, J. T., Haas, J. A., McDowell, M. T., van der Hoef, I., Lugtenburg, J., Lagarias, J. C., and Mathies, R. A. (2000) *Biochemistry* **39**, 2667–2676
623c. Hennig, L., and Schäfer, E. (2001) *J. Biol. Chem.* **276**, 7913–7918
623d. Huq, E., Tepperman, J. M., and Quail, P. H. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9789–9794
623e. Park, C.-M., Kim, J.-I., Yang, S.-S., Kang, J.-G., Kang, J.-H., Shim, J.-Y., Chung, Y.-H., Park, Y.-M., and Song, P.-S. (2000) *Biochemistry* **39**, 10840–10847
623f. Schmitz, O., Katayama, M., Williams, S. B., Kondo, T., and Golden, S. S. (2000) *Science* **289**, 765–768
624. Yeh, K.-C., Wu, S.-H., Murphy, J. T., and Lagarias, J. C. (1997) *Science* **277**, 1505–1508
625. Jiang, Z., Swem, L. R., Rushing, B. G., Devanathan, S., Tollin, G., and Bauer, C. E. (1999) *Science* **285**, 406–409
625a. Bhoo, S.-H., Davis, S. J., Walker, J., Karniol, B., and Vierstra, R. D. (2001) *Nature (London)* **414**, 776–779
626. Smith, H. (1999) *Nature (London)* **400**, 710–712
627. Ni, M., Tepperman, J. M., and Quail, P. H. (1999) *Nature (London)* **400**, 781–784
627a. Sweere, U., Eichenberg, K., Lohrmann, J., Mira-Rodado, V., Bäurle, I., Kudla, J., Nagy, F., Schäfer, E., and Harter, K. (2001) *Science* **294**, 1108–1111
628. Hoecker, U., Tepperman, J. M., and Quail, P. H. (1999) *Science* **284**, 496–499
629. Evans, A., and Smith, H. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 138–142
630. Song, P.-S., and Moore, T. A. (1974) *Photochem. Photobiol.* **19**, 435–441
631. Quiñones, M. A., Lu, Z., and Zeiger, E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2224–2228
632. Cashmore, A. R., Jarillo, J. A., Wu, Y.-J., and Liu, D. (1999) *Science* **284**, 760–765
633. Presti, D., Hsu, W.-J., and Delbrück, M. (1976) *Photochem. Photobiol.* **26**, 403–405
634. Galland, P., and Lipson, E. D. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 104–108
635. Christie, J. M., Reymond, P., Powell, G. K., Bernasconi, P., Raibekas, A. A., Liscum, E., and Briggs, W. R. (1998) *Science* **282**, 1698–1701
636. Nozue, K., Kanegae, T., Imaizumi, T., Fukuda, S., Okamoto, H., Yeh, K.-C., Lagarias, J. C., and Wada, M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15816–15830
636a. Sakai, T., Kagawa, T., Kasahara, M., Swartz, T. E., Christie, J. M., Briggs, W. R., Wada, M., and Okada, K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6969–6974
636b. Motchoulski, A., and Liscum, E. (1999) *Science* **286**, 961–964
636c. Christie, J. M., and Briggs, W. R. (2001) *J. Biol. Chem.* **276**, 11457–11460
636d. Kinoshita, T., Doi, M., Suetsugu, N., Kagawa, T., Wada, M., and Shimazaki, K.-i. (2001) *Nature (London)* **414**, 656–660
636e. Swartz, T. E., Wenzel, P. J., Corchnoy, S. B., Briggs, W. R., and Bogomolni, R. A. (2002) *Biochemistry* **41**, 7183–7189
637. Linden, H., and Macino, G. (1997) *EMBO J.* **16**, 98–109
637a. Iseki, M., Matsunaga, S., Murakami, A., Ohno, K., Shiga, K., Yoshida, K., Sugai, M., Takahashi, T., Hori, T., and Watanabe, M. (2002) *Nature (London)* **415**, 1047–1048
637b. Froehlich, A. C., Liu, Y., Loros, J. J., and Dunlap, J. C. (2002) *Science* **297**, 815–819
637c. Linden, H. (2002) *Science* **297**, 777–778
637d. Bieszke, J. A., Spudich, E. N., Scott, K. L., Borkovich, K. A., and Spudich, J. L. (1999) *Biochemistry* **38**, 14138–14145

References

638. Ahmad, M., and Cashmore, A. R. (1993) *Nature (London)* **366**, 162–166
639. Ahmad, M., Jarillo, J. A., Smirnova, O., and Cashmore, A. R. (1998) *Nature (London)* **392**, 720–723
640. Guo, H., Yang, H., Mockler, T. C., and Lin, C. (1998) *Science* **279**, 1360–1363
640a. Guo, H., Mockler, T., Duong, H., and Lin, C. (2001) *Science* **291**, 487–489
640b. Wang, H., Ma, L.-G., Li, J.-M., Zhao, H.-Y., and Deng, X. W. (2001) *Science* **294**, 151–158
641. Somers, D. E., Devlin, P. F., and Kay, S. A. (1998) *Science* **282**, 1488–1490
642. Ceriani, M. F., Darlington, T. K., Staknis, D., Más, P., Petti, A. A., Weitz, C. J., and Kay, S. A. (1999) *Science* **285**, 553–556
643. Thresher, R. J., Vitaterna, M. H., Miyamoto, Y., Kazantsev, A., Hsu, D. S., Petit, C., Selby, C. P., Dawut, L., Smithies, O., Takahashi, J. S., and Sancar, A. (1998) *Science* **282**, 1490–1494
643a. Hardin, P. E., and Glossop, N. R. J. (1999) *Science* **286**, 2460–2461
644. Hsu, D. S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R.-P., Todo, T., Wei, Y.-F., and Sancar, A. (1996) *Biochemistry* **35**, 13871–13877
645. Lucas, R. J., Freedman, M. S., Munoz, M., Garcia-Fernández, J.-M., and Foster, R. G. (1999) *Science* **284**, 505–507
646. Miyamoto, Y., and Sancar, A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6097–6102
647. Moore, R. Y. (1999) *Science* **284**, 2102–2103
647a. Hattar, S., Liao, H.-W., Takao, M., Berson, D. M., and Yau, K.-W. (2002) *Science* **295**, 1065–1070
647b. Berson, D. M., Dunn, F. A., and Takao, M. (2002) *Science* **295**, 1070–1073
647c. Barinaga, M. (2002) *Science* **295**, 955–957
647d. Thompson, C. L., Blaner, W. S., Van Gelder, R. N., Lai, K., Quadro, L., Colantunoni, V., Gottesman, M. E., and Sancar, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11708–11713
648. Hearst, J. E. (1995) *Science* **268**, 1858–1859
649. Sancar, A. (1994) *Biochemistry* **33**, 2–9
650. Snapka, R. M., and Sutherland, B. M. (1980) *Biochemistry* **19**, 4201–4208
651. Gindt, Y. M., Vollenbroek, E., Westphal, K., Sackett, H., Sancar, A., and Babcock, G. T. (1999) *Biochemistry* **38**, 3857–3866
652. Park, H.-W., Kim, S.-T., Sancar, A., and Deisenhofer, J. (1995) *Science* **268**, 1866–1872
653. Lipman, R. S. A., and Jorns, M. S. (1996) *Biochemistry* **35**, 7968–7973
653a. Durbeej, B., and Eriksson, L. A. (2000) *J. Am. Chem. Soc.* **122**, 10126–10132
653b. MacFarlane, A. W., IV, and Stanley, R. J. (2001) *Biochemistry* **40**, 15203–15214
654. Aubert, C., Mathis, P., Eker, A. P. M., and Brettel, K. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5423–5427
654a. Kay, C. W. M., Feicht, R., Schulz, K., Sadewater, P., Sancar, A., Bacher, A., Möbius, K., Richter, G., and Weber, S. (1999) *Biochemistry* **38**, 16740–16748
654b. Antony, J., Medvedev, D. M., and Stuchebrukhov, A. A. (2000) *J. Am. Chem. Soc.* **122**, 1057–1065
654c. Weber, S., Möbius, K., Richter, G., and Kay, C. W. M. (2001) *J. Am. Chem. Soc.* **123**, 3790–3798
655. Hitomi, K., Kim, S.-T., Iwai, S., Harima, N., Otoshi, E., Ikenaga, M., and Todo, T. (1997) *J. Biol. Chem.* **272**, 32591–32598
655a. Hitomi, K., Nakamura, H., Kim, S.-T., Mizukoshi, T., Ishikawa, T., Iwai, S., and Todo, T. (2001) *J. Biol. Chem.* **276**, 10103–10109
656. Kim, S.-T., Malhotra, K., Smith, C. A., Taylor, J.-S., and Sancar, A. (1994) *J. Biol. Chem.* **269**, 8535–8540
657. Zhao, X., Liu, J., Hsu, D. S., Zhao, S., Taylor, J.-S., and Sancar, A. (1997) *J. Biol. Chem.* **272**, 32580–32590
657a. Joseph, A., Prakash, G., and Falvey, D. E. (2000) *J. Am. Chem. Soc.* **122**, 11219–11225
658. Todo, T., Ryo, H., Yamamoto, K., Toh, H., Inui, T., Ayaki, H., Nomura, T., and Ikenaga, M. (1996) *Science* **272**, 109–112
659. Viviani, V. R., Bechara, E. J. H., and Ohmiya, Y. (1999) *Biochemistry* **38**, 8271–8279
660. Conti, E., Franks, N. P., and Brick, P. (1996) *Structure* **4**, 287–298
661. Branchini, B. R., Magyar, R. A., Murtiashaw, M. H., Anderson, S. M., and Zimmer, M. (1998) *Biochemistry* **37**, 15311–15319
661a. Branchini, B. R., Magyar, R. A., Murtiashaw, M. H., Anderson, S. M., Helgerson, L. C., and Zimmer, M. (1999) *Biochemistry* **38**, 13223–13230
662. McElroy, W. D., and Seliger, H. H. (1962) *Sci. Am.* **207**(Dec), 76–89
663. Campbell, A. K. (1986) *Trends Biochem. Sci.* **11**, 104–108
664. Fisher, A. J., Raushel, F. M., Baldwin, T. O., and Rayment, I. (1995) *Biochemistry* **34**, 6581–6586
665. Bronstein, I., and McGrath, P. (1989) *Nature (London)* **338**, 599–600
666. Adam, W., Bronstein, I., Edwards, B., Engel, T., Reinhardt, D., Schneider, F. W., Trofimov, A. V., and Vasil'ev, R. F. (1996) *J. Am. Chem. Soc.* **118**, 10400–10407
667. Hori, K., Wampler, J. E., Matthews, J. C., and Cormier, M. J. (1973) *Biochemistry* **12**, 4463–4468
668. Matthews, J. C., Hori, K., and Cormier, M. J. (1977) *Biochemistry* **16**, 85–91
669. Shimamura, O., and Johnson, F. H. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2611–2615
670. Tsuji, F. I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4629–4632
671. Thompson, E. M., Nagata, S., and Tsuji, F. I. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6567–6571
672. Ray, B. D., Ho, S., Kemple, M. D., Prendergast, F. G., and Nageswara Rao, B. D. (1985) *Biochemistry* **24**, 4280–4287
673. Inouye, S., Noguchi, M., Sakaki, Y., Takagi, Y., Miyata, T., Iwanaga, S., Miyata, T., and Tsuji, F. I. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3154–3158
674. Shimomura, O., and Johnson, F. H. (1972) *Biochemistry* **11**, 1602–1608
674a. Markova, S. V., Vysotski, E. S., Blinks, J. R., Burakova, L. P., Wang, B.-C., and Lee, J. (2002) *Biochemistry* **41**, 2227–2236
675. Cormier, M. J., and Wampler, J. E. (1975) *Ann. Rev. Biochem.* **44**, 255–272
676. Hart, R. C., Matthews, J. C., Hori, K., and Cormier, M. J. (1979) *Biochemistry* **18**, 2204–2205
676a. Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J., and Wang, B.-C. (2000) *Protein Sci.* **9**, 2085–2093
676b. Haddock, S. H. D., Rivers, T. J., and Robison, B. H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11148–11151
676c. Li, L., Liu, L., Hong, R., Robertson, D., and Hastings, J. W. (2001) *Biochemistry* **40**, 1844–1849
676d. Nakamura, H., Kishi, Y., Shimomura, O., Morse, D., and Hastings, J. W. (1989) *J. Am. Chem. Soc.* **111**, 7607–7611
676e. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A., and Anderson, S. M. (2000) *Biochemistry* **39**, 5433–5440
677. Thompson, J. F., Geoghegan, K. F., Lloyd, D. B., Lanzetti, A. J., Magyar, R. A., Anderson, S. M., and Branchini, B. R. (1997) *J. Biol. Chem.* **272**, 18766–18771
677a. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A., Portier, N. C., Ruggiero, M. C., and Stroh, J. G. (2002) *J. Am. Chem. Soc.* **124**, 2112–2113
677b. Gomi, K., and Kajiyama, N. (2001) *J. Biol. Chem.* **276**, 36508–36513
678. Shimomura, O., and Johnson, F. H. (1968) *Biochemistry* **7**, 2574–2580
679. Nakamura, H., Kishi, Y., Shimomura, O., Morse, D., and Hastings, J. W. (1989) *J. Am. Chem. Soc.* **111**, 7607–7611
679a. Jeffers, C. E., and Tu, S.-C. (2001) *Biochemistry* **40**, 1749–1754
680. Xin, X., Xi, L., and Tu, S.-C. (1994) *Biochemistry* **33**, 12194–12201
681. Tanner, J. J., Miller, M. D., Wilson, K. S., Tu, S.-C., and Krause, K. L. (1997) *Biochemistry* **36**, 665–672
682. Baldwin, T. O., Berends, T., Bunch, T. A., Holzman, T. F., Rausch, S. K., Shamansky, L., Treat, M. L., and Ziegler, M. M. (1984) *Biochemistry* **23**, 3663–3667
683. Cohn, D. H., Mileham, A. J., Simon, M. I., Nealson, K. H., Rausch, S. K., Bonam, D., and Baldwin, T. O. (1985) *J. Biol. Chem.* **260**, 6139–6146
684. Lawson, D. M., Derewenda, U., Serre, L., Ferri, S., Szittner, R., Wei, Y., Meighen, E. A., and Derewenda, Z. S. (1994) *Biochemistry* **33**, 9382–9388
685. Kurfürst, M., Ghisla, S., and Hastings, J. W. (1983) *Biochemistry* **22**, 1521–152
685a. Lin, L. Y.-C., Sulea, T., Szittner, R., Kor, C., Purisima, E. O., and Meighen, E. A. (2002) *Biochemistry* **41**, 9938–9945
686. Eckstein, J. W., Hastings, J. W., and Ghisla, S. (1993) *Biochemistry* **32**, 404–411
687. Francisco, W. A., Abu-Soud, H. M., DelMonte, A. J., Singleton, D. A., Baldwin, T. O., and Raushel, F. M. (1998) *Biochemistry* **37**, 2596–2606
688. Li, H., Ortego, B. C., Maillard, K. I., Willson, R. C., and Tu, S.-C. (1999) *Biochemistry* **38**, 4409–4415
689. Engebrecht, J., and Silverman, M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4154–4158
690. Eberhard, A., Burlingame, A. L., Eberhard, C., Kenyon, G. L., Nealson, K. H., and Oppenheimer, N. J. (1981) *Biochemistry* **20**, 2444–2449

Study Questions

1. Why is the Emerson enhancement effect (i.e., light at 650 nm plus 680 nm gives a higher rate of photosynthesis than either one alone) not observed with photosynthetic bacteria?

2. Agents that uncouple oxidative phosphorylation in mitochondria uncouple photoelectron transport and ATP formation in photosynthesis. Explain.

3. The action spectrum of photosynthesis, which describes the efficiency of photosynthesis as a function of the wavelength of incident light, has a valley around 550 nm. Why?

4. Plants exposed to $C^{18}O_2$ will have the ^{18}O first appear in a) carbohydrate; b) water; c) oxygen gas. (More than one answer may be true.)

5. Plants exposed to $H_2{}^{18}O$ will have the label first appear in a) oxygen gas; b) carbohydrate; c) CO_2. (More than one may be correct.)

6. The general equation describing the photosynthesis of glucose in higher plants is:

$$6\ CO_2 + 6\ H_2O \rightarrow C_6H_{12}O_6$$

   We know that the oxygen gas comes from water, yet the equation shows only six atoms of oxygen in water on the left-hand side vs 12 in oxygen gas on the right. Explain.

7. The fructose bisphosphatase of green plants has an amino acid sequence which is very similar to those of the corresponding enzymes isolated from other sources such as yeast or mammals, except that the plant enzyme has an additional sequence of 20 or so amino acids that has no counterpart in the enzymes found in the other species. What function might this additional sequence have in the plant enzyme?

8. The following substances are either inhibitors or activators of rubisco, the enzyme that catalyzes the condensation of CO_2 with ribulose bisphosphate to yield 3-phosphoglycerate. State whether the substance should be an activator or an inhibitor of the enzyme and succinctly provide the logic supporting your conclusion.
   a) Fructose 1,6-bisphosphate
   b) Visible light
   c) NADPH

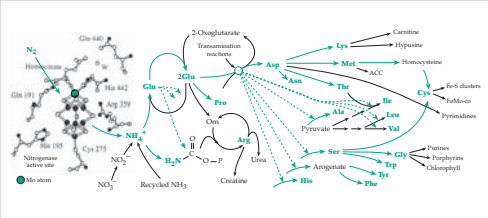
9. The reagent DCMU specifically inhibits electron transfer to plastoquinone in photosystem II. Discuss how the administration of this compound to a suspension of illuminated chloroplasts will affect the production of oxygen, ATP, and NADPH.

10. A chemical reagent is added to a solution of plant chloroplasts which immediately and specifically poisons photosystem II. What is the *short-term* effect of each of the following? Give a one-sentence defense for your conclusion.
    a) Cyclic photophosphorylation
    b) Noncyclic photophosphorylation
    c) Photorespiration
    d) NADPH production

11. If a C_3 and a C_4 plant are placed together in a sealed illuminated box, the C_3 plant withers and dies long before the C_4 plant. Explain.

12. What tricarboxylic acid cycle enzyme is analogous to the malate enzyme of bundle-sheath cells? What is the mechanism of the reaction?

13. There are two different forms of glyceraldehyde-3-phosphate dehydrogenase in higher plant cells.
    a) In which cell compartment is each one found?
    b) What are the reactions catalyzed by these two isozymes?
    c) Why are there two forms?



The air provides an abundant source of nitrogen for living organisms. Nitrogenase present in specialized bacteria utilizes the molybdenum- and iron-containing FeMo-co to reduce N_2 to two molecules of NH_3 (or NH_4^+). NH_3 is incorporated into the side chain of glutamine and much is transferred to the 5-carbon skeleton of 2-oxoglutarate to form glutamate. Nitrogen from glutamate and glutamine moves into the other amino acids via action of transaminases and glutamine amidotransferases. Thousands of compounds, a few of which are indicated here, are formed. (The 20 amino acid constituents of proteins are shown in green.) NH_3 from decaying materials is recycled, often after oxidation to nitrite or nitrate. Nitrates may also be formed by lightening and NH_3 industrially by catalytic reductions of N_2 by H_2 at high temperature and pressure (the Haber process).

Contents

1359 **A. Fixation of N_2 and the Nitrogen Cycle**
1359 1. Reduction of Elemental Nitrogen
1360 2. Nitrogenases
1362 *The mechanism of nitrogenase action*
1365 *Nitrogen fixation genes*
1365 *Legume nodules and cyanobacterial heterocysts*
1366 *Genetic engineering*
1366 *Other nitrogenases*
1366 3. Interconversion of Nitrate, Nitrite, and Ammonium Ions
1367 **B. Incorporation of NH_3 into Amino Acids and Proteins**
1369 1. Uptake of Amino Acids by Cells
1369 2. Glutamate Dehydrogenase and Glutamate Synthase
1370 3. Glutamine Synthetase
1371 4. Catabolism of Glutamine, Glutamate, and Other Amino Acids
1371 *Catabolism initiated by decarboxylation*
1371 *Fermentation of glutamate*
1374 **C. Synthesis and Catabolism of Proline, Ornithine, Arginine, and Polyamines**
1374 1. Synthesis and Catabolism of Proline
1374 2. Synthesis of Arginine and Ornithine and the Urea Cycle
1376 *The urea cycle*
1376 *Carbamoyl phosphate synthetases*
1377 *Citrulline and argininosuccinate*
1378 *Excretion of ammonia*
1378 *Catabolism of arginine*
1379 *Insecticidal analogs of arginine*
1379 3. Amidino Transfer and Creatine Synthesis
1380 4. The Polyamines
1381 *Biosynthesis*
1382 *Breakdown*
1383 **D. Compounds Derived from Aspartate**
1383 1. Control of Biosynthetic Reactions of Aspartate
1383 2. Lysine Diaminopimelate, Dipicolinic Acid, and Carnitine
1386 3. The Catabolism of Lysine
1388 4. Metabolism of Homocysteine and Methionine
1389 *A salvage pathway*
1389 *The plant hormone ethylene*
1391 5. Metabolism of Threonine
1391 **E. Alanine and the Branched-Chain Amino Acids**
1393 1. Catabolism
1397 2. Ketogenic and Glucogenic Amino Acids
1397 **F. Serine and Glycine**
1397 1. Biosynthetic Pathways from Serine
1398 2. Metabolism of Glycine
1399 3. Porphobilinogen, Porphyrins, and Related Substances
1400 *Porphyrins*
1402 *Corrins*
1402 *Chlorophyll*
1403 *The porphyrias*
1404 *The bile pigments*
1406 **G. Cysteine and Sulfur Metabolism**
1407 1. Synthesis and Catabolism of Cysteine
1407 2. Cysteine Sulfinate and Taurine
1408 3. Mercaptopyruvate, Thiosulfate, and Assembly of Iron–Sulfur Centers

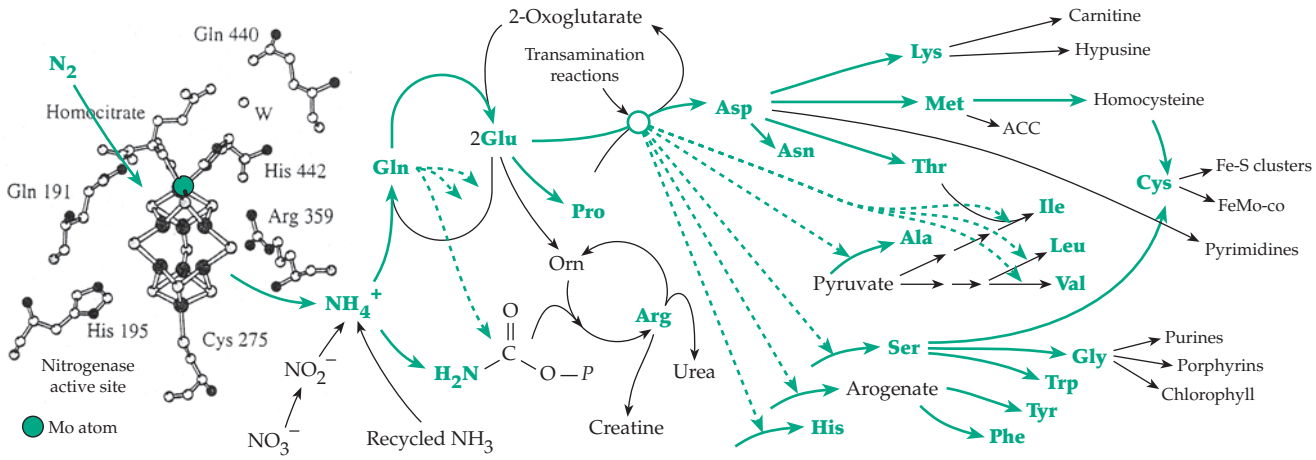
1411 **References**
1418 **Study Questions**

Boxes

1394 Box 24-A Maple Syrup Urine Disease and Jamaican Vomiting Sickness
1408 Box 24-B Sulfur Compounds of Garlic, Onions, Skunks, etc.

The Metabolism of Nitrogen and Amino Acids

24



Because it is found in so many compounds and can exist in several oxidation states, nitrogen has a complex metabolism. The inorganic forms of nitrogen found in our surroundings range from the highly oxidized nitrate ion, in which N has an oxidation state of +5, to ammonia, in which the oxidation state is −3. Living cells both reduce and oxidize these inorganic forms. The organic forms of nitrogen are most often derived by incorporation of **ammonium ions** into amino groups or amide groups. Once it has been incorporated into an organic compound, nitrogen can be transferred into many other carbon compounds. Certain compounds including glutamic acid, aspartic acid, glutamine, asparagine, and carbamoyl phosphate are especially active in these transfer reactions. They constitute a **nitrogen pool** from which nitrogen can be withdrawn and to which it can be returned.

In addition to the pathways for synthesis and degradation of nitrogenous substances, many organisms have specialized metabolism for incorporation of excess nitrogen into relatively nontoxic excretion products. All of these aspects of nitrogen metabolism will be dealt with in this and the following chapter. We will look first at the reactions by which organic nitrogen compounds are formed from inorganic compounds, then at the reactions of the nitrogen pool. After that we will examine the specific reactions of synthesis and catabolism of individual nitrogenous compounds.

A. Fixation of N_2 and the Nitrogen Cycle

Most of the nitrogen of the biosphere exists as the unreactive N_2, which makes up 80% of the molecules of air. The "fixation" of N_2 occurs principally by the action of a group of bacteria known as **diazotrophs** and to a lesser extent by lightning, which forms oxides of nitrogen and eventually nitrate and nitrite. Human beings also contribute a smaller but significant share through production of chemical fertilizer by the Haber process. These reactions are an important part of the **nitrogen cycle**.[1,2] Quantitatively even more important are the biochemical processes of **nitrification**, by which ammonium ions from decaying organic materials are oxidized to NO_2^- and NO_3^- by soil bacteria (Fig. 24-1), and reactions of reduction and **assimilation** of nitrate and nitrite by bacteria, fungi, and green plants. Another reductive process catalyzed by **denitrifying bacteria** returns N_2 to the atmosphere (Fig. 24-1).

1. Reduction of Elemental Nitrogen

One of the most remarkable reactions of nitrogen metabolism is the conversion of dinitrogen (N_2) to ammonia. It was estimated that in 1974 this biological nitrogen fixation added 17 x 10^{10} kg of nitrogen to the earth (compared with 4 x 10^{10} kg fixed by chemical reactions).[3] The quantitative significance can be more easily appreciated by the realization that one square meter of land planted to nodulated legumes such as soybeans can fix 10–30 g of nitrogen per year.

Fixation of N_2 by *Clostridium pasteurianum* and a few other species was recognized by Winogradsky[4] in 1893. Subsequent nutritional studies indicated that both iron and molybdenum were required for the process. Inhibition by CO and N_2O was observed. While ammonia was the suggested product, the possibility remained that more oxidized compounds such

as hydroxylamine were the ones first incorporated into organic substances. When cell-free preparations capable of fixing nitrogen were obtained in 1960 rapid progress became possible.[5] It was discovered that nitrogen-fixing bacteria are invariably able to reduce acetylene to ethylene, a catalytic ability that goes hand in hand with the ability to reduce N_2. A simple, sensitive **acetylene reduction test** permits easy measurement of the nitrogen-fixing potential of cells.

Application of this test revealed that nitrogen fixation is not restricted to a few species, but is a widespread ability of many prokaryotes. Most studied are *Azotobacter vinelandii*, Winogradsky's *C. pasteurianum*, *Klebsiella pneumoniae* (a close relative of *E. coli*), and several species of *Rhizobium*, the symbiotic bacterium of root nodules of legumes. The latter deserves special attention. Although some free-living rhizobia reduce N_2, the reaction usually takes place only in nodules developed by infected roots. Within these nodules the bacteria degenerate into **bacteroids**; and the special hemoglobin **leghemoglobin**,[6,7] whose sequence is specified by a plant gene,[8] is synthesized.

Legumes are not the only plants with nitrogen-fixing symbionts.[9] Some other angiosperms are hosts to nitrogen-fixing actinomycetes and some gymnosperms contain nitrogen-fixing blue-green algae. Leaf nodules of certain plants infected with *Klebsiella* fix nitrogen. While the nutritional significance is uncertain, nitrogen-fixing strains of *Klebsiella* have also been found in the intestinal tracts of humans in New Guinea. Of the free-living nitrogen-fixing organisms, cyanobacteria appear to be of most importance quantitatively. For example, in rice paddy fields cyanobacteria may fix from 2.4 to 10 g of nitrogen per square meter per year. Cyanobacteria in the oceans fix enormous amounts of nitrogen.[9a]

2. Nitrogenases

Cell-free nitrogenases have been isolated from a number of organisms. These enzymes all share the property of being inactivated by oxygen, a fact that impeded early work. Apparently nitrogen fixation occurs in anaerobic regions of cells. Leghemoglobin may protect the nitrogen-fixing enzymes in root nodules from oxygen. It probably also functions to deliver O_2 by facilitated diffusion to the aerobic mitochondria of the bacteroids at a stable, low partial pressure.[6,10] Some bacteria utilize protective proteins to shield the nitrogenase molecules when the O_2 pressure is too high.[10a]

Nitrogenases catalyze the six-electron reduction of N_2 to ammonia (Eq. 24-1) and are also able to reduce

$$N_2 + 6\,H^+ + 6\,e^- \rightarrow 2\,NH_3 \qquad (24\text{-}1)$$

many other compounds. For example, the reduction of acetylene to ethylene (Eq. 24-2) is a two-electron process. Azide is reduced to N_2 and NH_4^+ in another two-electron reduction (Eq. 24-3). Cyanide ions yield methane and ammonia (Eq. 24-4).[11] Alkyl nitriles as well as N_2O and carbonyl sulfide (COS) are also reduced. Carbon dioxide is reduced slowly to CO[12], and nitrogenases invariably catalyze reduction of protons to H_2 (Eq. 24-5).

$$HC{\equiv}CH + 2\,H^+ + 2\,e^- \rightarrow H_2C{=}CH_2 \qquad (24\text{-}2)$$

$$N{=}N{=}N^- + 4\,H^+ + 2\,e^- \rightarrow N_2 + NH_4^+ \qquad (24\text{-}3)$$

$$^-C{\equiv}N + 8\,H^+ + 6\,e^- \rightarrow CH_4 + NH_4^+ \qquad (24\text{-}4)$$

$$2\,H^+ + 2\,e^- \rightarrow H_2 \qquad (24\text{-}5)$$

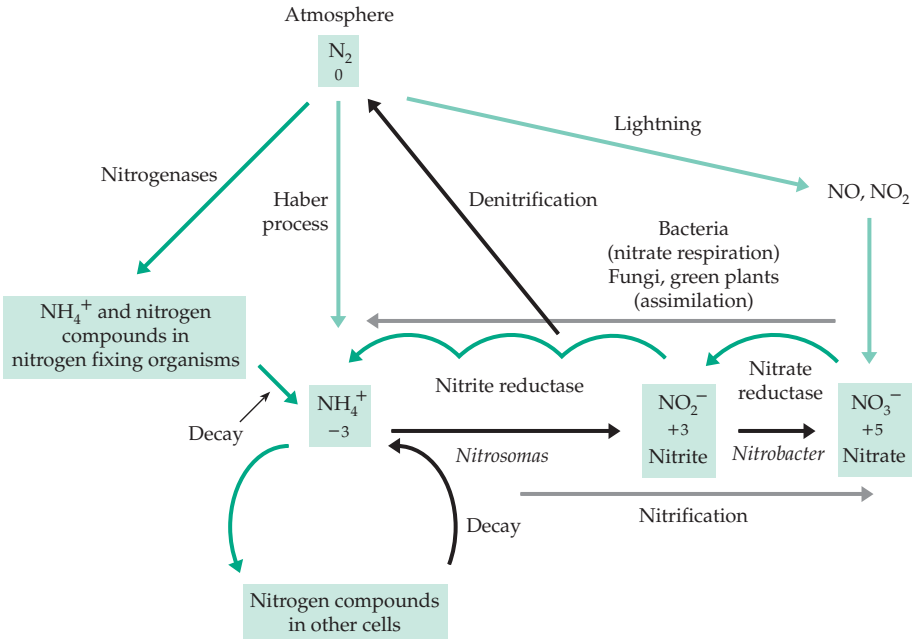


Figure 24-1 The nitrogen cycle. Conversion of N_2 (oxidation state 0) to NH_4^+ by nitrogen-fixing bacteria, assimilation of NH_4^+ by other organisms, decay of organic matter, oxidation of NH_4^+ by the nitrifying bacteria *Nitrosomas* and *Nitrobacter*, reduction of NO_3^- and NO_2^- back to NH_4^+, and release of nitrogen as N_2 by denitrifying bacteria are all part of this complex cycle.[1]

In early experiments it was found that sodium pyruvate was required for fixation of N_2 in cell-free extracts, and that large amounts of CO_2 and H_2 accumulated. Investigation showed that cleavage of pyruvate supplies cells with two important products: ATP and reduced ferredoxin. Pyruvate can be replaced by a mixture of ATP plus Mg^{2+} and reduced ferredoxin (Fd_{red}). Furthermore, the nonbiological reductant dithionite ($S_2O_4^{2-}$) can replace the reduced ferredoxin. Since ADP is inhibitory to the nitrogenase system, it is best in laboratory studies to supply ATP from an ATP-generating system such as a mixture of creatine phosphate, creatine kinase, and a small amount of ADP (Eqs. 6-65, 6-67).

The commonest type of nitrogenase can be separated easily into two components (Fig. 24-2). One of these, the **iron protein** (dinitrogenase reductase, azoferredoxin, or component II), is an extremely oxygen-sensitive iron-sulfur protein. It consists of two identical ~32-kDa peptide chains; those of *A. vinlandii* each contain 189 amino acid residues. The three-dimensional structure of the dimeric protein[13–16] shows that each subunit forms a nucleotide-binding domain with an ATP-binding site. About 2 nm away from this site is *a single Fe_4S_4 cluster which is shared symmetrically by the two subunits of the protein*. Each subunit contributes two thiolate groups from Cys 97 and Cys 132 as well as three N–H--S hydrogen bonds from NH groups at helix ends.[13]

The other component, the **molybdenum–iron protein** (dinitrogenase, molybdoferredoxin, or component I), contains both iron and molybdenum as well as labile sulfide. It is a mixed ($\alpha_2\beta_2$) tetramer of ~240-kDa mass and an analytical metal ion composition ~$Mo_2Fe_{30}S_{26}$. However, the X-ray structure[16–19a] suggests the composition $Mo_2Fe_{34}S_{36}$. The MoFe protein is a symmetric molecule in which each αβ subunit contains two types of complex metal clusters. The active sites for N_2 reduction, which are embedded in the α subunits, contain the **FeMo-coenzyme** molecules, each with the metal composition $MoFe_7S_9$ and also containing a molecule of **homocitrate**.[20–22a] The other clusters, known as **P-clusters**, are shared between the α and β subunits, which for *A. vinlandii* contain 491 and 522

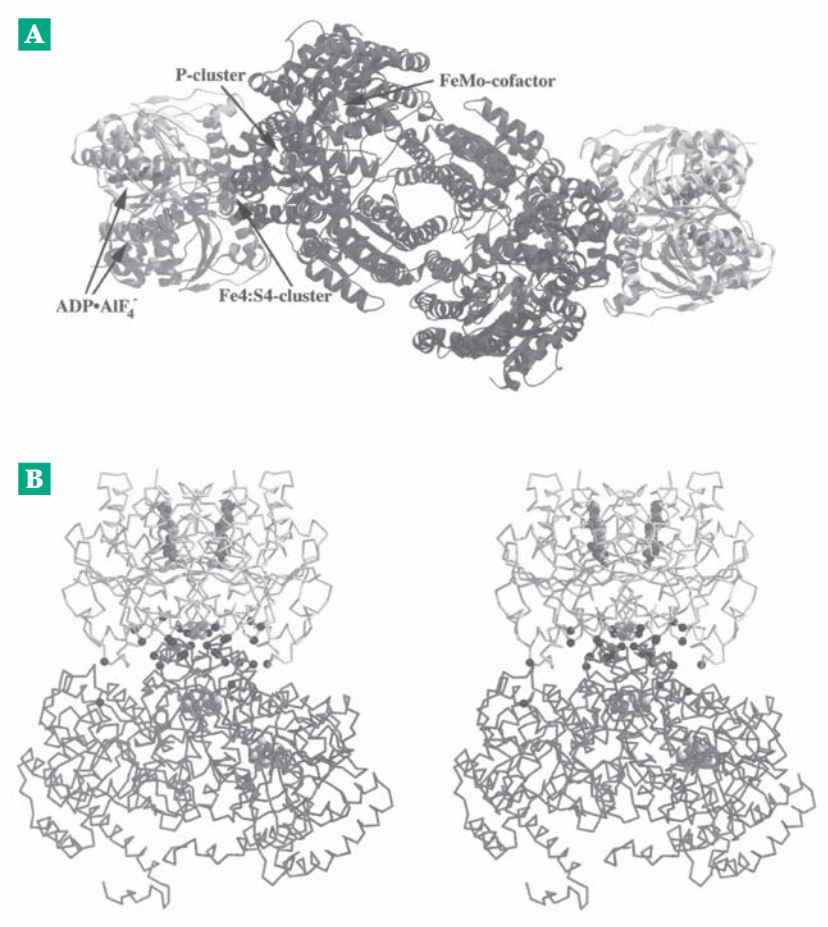
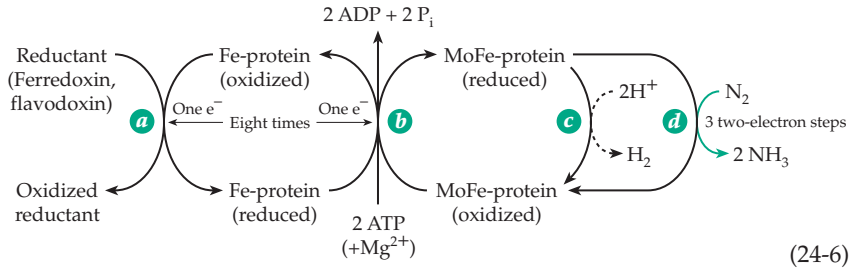


Figure 24-2 (A) Ribbon drawing of the three-dimensional structure of the nitrogenase from *Azotobacter vinlandii* viewed down the twofold axis of the molecule. The αβ subunits of the MoFe-protein are in the center while the Fe-protein subunits are at the outer ends. The Fe_4S_4 cluster of the Fe-protein and the FeMo-co and P-cluster of the MoFe-protein are marked for the left-hand complex. The site of binding of ATP is identified by the bound ADP•AlF_4^- complex. (B) Stereoscopic view of one complete half of the nitrogenase complex at a 90° angle to the view in (A). The Fe-protein is at the top, and the MoFe-protein is below. The ADP•AlF_4^- complex is visible in the two symmetrically located binding sites of the Fe-protein. The shared Fe_4S_4 cluster is in the center above the P-cluster. The small black spheres mark α-carbons of residues that interact in forming the complex of Fe-protein and MoFe-proteins. When the Fe_4S_4 cluster accepts one electron from a molecule of ferredoxin or flavodoxin, the Fe-protein binds to the MoFe-protein and donates an electron to one of the two nearby P-clusters (one of which is in each of the αβ subunits). At the same time both of the molecules of ATP bound to the Fe-protein are hydrolyzed. The oxidized Fe-protein then dissociates from the complex and is replaced by another reduced Fe-protein–ATP complex. The net result is that each electron is "pumped" from the Fe_4S_4 cluster of the Fe-protein into a P-cluster of one of the αβ units of the MoFe-protein. Electrons then move from the P-clusters into the FeMo-coenzyme. From Schindelin *et al.*[19] Courtesy of Douglas C. Rees.

amino acid residues, respectively. Each P-cluster is actually a *joined pair* of cubane-type clusters, one Fe_4S_4 and one Fe_4S_3 with two bridging cysteine –SH groups and one iron atom bonded to three sulfide sulfur atoms (Fig. 24-3).[17,23] The FeMo-coenzyme can be released from the MoFe-protein by acid denaturation followed by extraction with dimethylformamide.[24] While homocitrate was identified as a component of the isolated coenzyme, the three-dimensional structure of FeMo-co was deduced from X-ray crystallography of the intact molybdenum–iron protein.[14,17,18]



(24-6)

When the Fe-protein is reduced an EPR signal at $g = 1.94$, typical of iron–sulfur proteins (Fig. 16-17), is observed. This signal is altered by interaction with Mg and ATP, whereas ATP has no effect on the complex EPR signals produced upon reduction of MoFd. These are among the observations that led to the concept that the Fe-protein is an electron carrier responsible for reduction of the molybdenum in the MoFe-protein. The Mo(IV) or two atoms of Mo(III) formed in this way could then reduce N_2 in three two-electron steps with formation of Mo(VI) (Eq. 24-6). Three successive two-electron steps are required to completely reduce N_2 to two molecules of ammonia. An unexpected feature of nitrogenase action is that there is inevitably what was once regarded as a side reaction, the reduction of protons to H_2.[25,26] The amount of H_2 formed is variable and may be much greater than that of N_2 reduced. However, at high pressures of N_2 the ratio of H_2 formed to N_2 reduced is 1:1. This led to the suggestion[28] that H_2 formation is not a side reaction but an essential step in preparing the active site for the binding of N_2. Two protons that are bound somewhere on the reduced MoFe-protein could be reduced to H_2 in an obligatory step (Eq. 24-6, step *c*) that would, for example, cause a conformational change required for binding of N_2. If no reducible substrate (N_2, C_2H_2, etc.) is present, H_2 would still be formed slowly. Reducible substrates inhibit H_2 formation. However, addition of N_2 or any other reducible substrate causes an initial "burst" of H_2 to be released. This can be measured readily when a slow substrate such as CN^- is used as the inhibitor. The amount of H_2 released in the burst is stoichiometric with one H_2 per Mo being formed.[26] The overall stoichiometry for reduction of one N_2 becomes:

$$N_2 + 8\,e^- + 8\,H^+ + 16\,\text{MgATP} \rightarrow 2\,NH_3 + H_2 + 16\,\text{MgADP} + 16\,P_i \qquad (24\text{-}7)$$

A second remarkable feature of nitrogenase is a requirement for hydrolysis of MgATP that is coupled to reduction of the MoFe-protein (Eq. 24-6). Two molecules of ATP are hydrolyzed to ADP and inorganic phosphate for each electron transferred. This large ATP requirement seems surprising in view of the fact that reduction of N_2 by reduced ferredoxin (Eq. 24-8) is thermodynamically spontaneous:

$$N_2 + 6\,Fd_{red} + 8\,H^+ \rightarrow 2\,NH_4^+ + 6\,Fd_{ox}^+$$
$$\Delta G'\ (\text{pH } 7) = -89.3\ \text{kJ mol}^{-1} \qquad (24\text{-}8)$$

However, N_2 is exceedingly unreactive. In the commercial Haber process high pressure and temperature are needed to cause H_2 and N_2 to combine. Evidently cleavage of 16 molecules of ATP must be coupled to the nitrogenase reduction system to overcome the very high activation energy.

Not only are two molecules of ATP hydrolyzed to pump each electron, but the Fe-protein must receive electrons from a powerful (low $E^{\circ\prime}$) reductant such as reduced ferredoxin, reduced flavodoxin, or dithionite. *Klebsiella pneumoniae* contains a **pyruvate:flavodoxin oxidoreductase** (Eq. 15-35) that reduces either flavodoxin or ferredoxin to provide the low potential electron donor.[29,30] In some bacteria, e.g., the strictly aerobic *Azotobacter*, NADPH is the electron donor for reduction of N_2. The Fe-protein is thought to accept electrons from a chain that includes at least the ordinary bacterial ferredoxin (Fd) and a special one-electron-accepting **azotoflavin,** a flavoprotein that is somewhat larger than the flavodoxins (Chapter 15) and appears to play a specific role in N_2 fixation.[31] In *Clostridium* and *Rhizobium* reduced ferredoxins generated by cleavage of pyruvate reduce nitrogenase directly.[32]

The mechanism of nitrogenase action. The one-electron reduction of the Fe_4S_4 cluster of the Fe-protein (step *a* of Eq. 24-6) initiates the action. This reaction occurs before the Fe-protein forms a complex with the MoFe-protein. Following this initial reduction step the two molecules of ATP required for step *b* of Eq. 24-6 bind to the Fe-protein. One is bound to each subunit of this protein but neither is immediately adjacent to the shared Fe_4S_4 cluster, as can be seen from Fig. 14-3B. The binding to MgATP appears to induce a conformational change that permits the "docking" of the Fe-protein with the MoFe-protein to

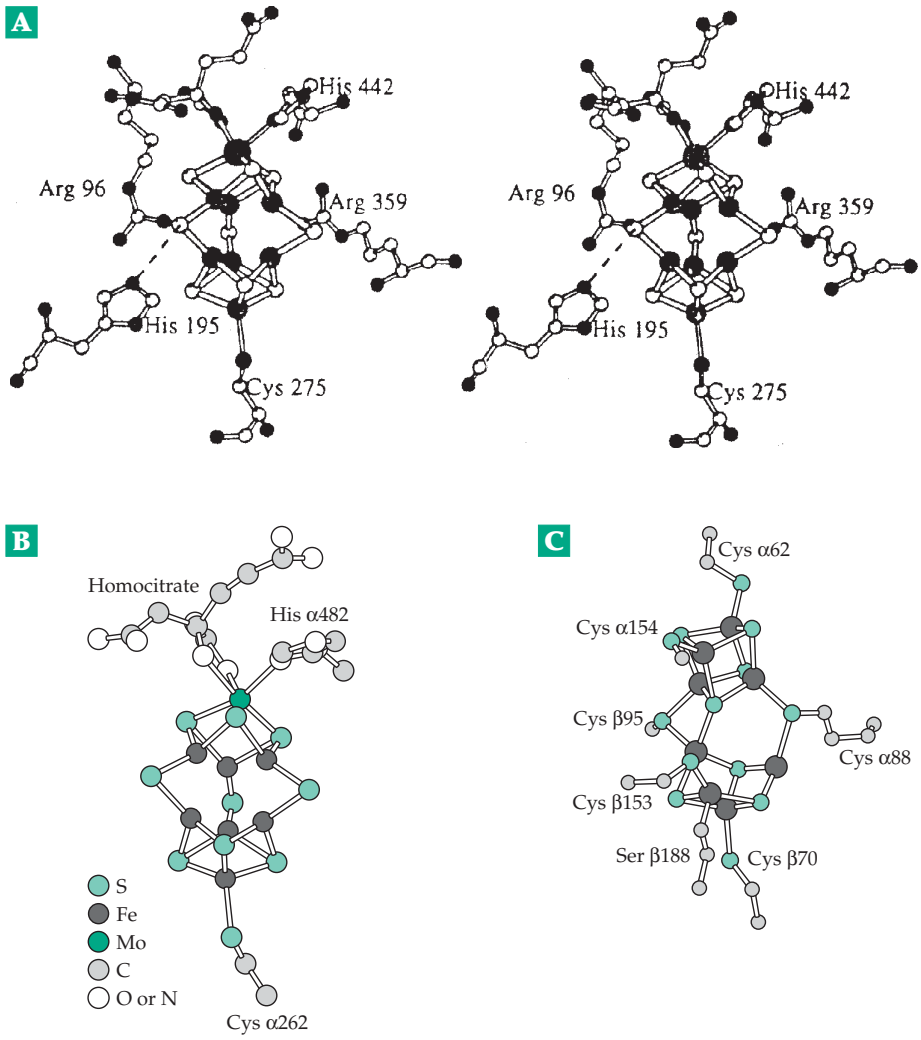


Figure 24-3 Structures of the metal-sulfide clusters of the MoFe-protein. (A) Stereoscopic view of the FeMo-co coenzyme with interacting side chains from the MoFe-protein of *A. vinlandii*. After Kim *et al.*[27] (B) FeMo-co with atom labels. From Kim *et al.*[18] (C) The structure of the oxidized form of the P-cluster. From Peters *et al.*[23] Recent studies by Einsle *et al.*[22a] indicate that the cluster probably also contains a nitrogen atom that is held within the cluster by coordination to six of the iron atoms.

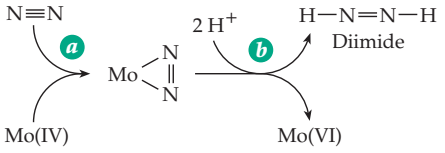
form the complex in which the electron transfer of step *b* (Eq. 24-6) occurs. Abundant evidence indicates that electron transfer does not occur without the binding of MgATP.[33–35d] The electron transfer is coupled to the hydrolysis of the ATP, but the two reactions appear to be consecutive events. In a deletion mutant of the Fe-protein (lacking Leu 127) the hydrolysis of ATP does not occur, but the complex between Fe-protein and MoFe-protein is formed and electron transfer to the MoFe-protein takes place.[36–37a] The binding of the MgATP causes the midpoint redox potential to drop from −0.42 V to −0.62 V, assisting the transfer.[38] X-ray crystallographic studies reveal a distinct conformational change similar to those observed with G-proteins (Chapter 11) and involving movement of the Fe_4S_4 center into a better position for electron transfer.[19,38a]

After electron transfer the complex of the two proteins is thought to be tightly bonded when unhydrolyzed ATP is present.[35] This has allowed the direct observation and imaging of the complex at low resolution (~1.5 nm) using rapid synchrotron X-ray scattering measurements.[39] The ATP is hydrolyzed, and the Fe-proton is released from the complex.

Only one electron is transferred to the MoFe-protein in each catalytic cycle of the Fe-protein. Thus, the cycle must be repeated eight times to accomplish the reduction of $N_2 + 2\ H^+$. Where in the MoFe-protein does a transferred electron go? EPR spectroscopic and other experiments with incomplete and catalytically inactive molybdenum coenzyme[40] have provided a clear answer. The electron is transferred first to one of the two P-clusters, both of which are close to the Fe_4S_4 cluster of the Fe-protein. The transfer causes an observable change both in the spectroscopic properties and in the three-dimensional structure of the P-cluster.[23,40a] Since protons are needed at the active site for the reduction reactions (the FeMo-coenzyme), it is probable that hydrolysis of ATP in the Fe-protein is accompanied by transport of protons across the interface with the MoFe-protein. The electron transfer from the P-cluster on to the FeMo-co center would be assisted by a protic force resulting from ATP cleavage.

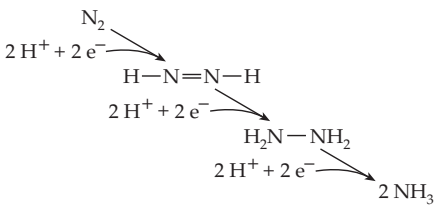
With defective FeMo-co (apparently lacking homocitrate) no reduction of N_2, acetylene, or protons is observed.[40] If intact FeMo-co is present, reduction of the cofactor can be observed. An $s = 3/2$ EPR signal arising from the Mo is seen,[40] and EXAFS measurements reveal decreased Mo–Fe distances as the coenzyme is reduced.[41] The molybdenum is probably present as Mo(VI) in the oxidized state of nitrogenase,[42] but after reduction it isn't clear whether it is Mo(III) or Mo(IV). Isolated FeMo-co exists in three identified oxidation states related by $E^{\circ\prime}$ values of −0.17 and −0.465 V.[43] Only the middle state is EPR-active, but it is the most reduced state that is involved in N_2 reduction.[42,43] With its P-cluster and FeMo-co center each αβ unit of the MoFe-protein could store several electrons. Two or more might be stored in a P-cluster, and Mo(VI) could, in principle, accept three electrons to form Mo(III). However, it is a little hard to imagine storage of the eight electrons needed to reduce both N_2 and H_2 (Eq. 24-8). The reduction of N_2 may begin before all eight electrons have been transferred into the MoFe-protein.

Another uncertainty lies in the mode of binding of N_2 and other substrates. Does N_2 bind end-on to Mo, does it slide between Fe atoms within the coenzyme, or does it bind in some other way? While N_2 is unreactive, it forms nitrides with metals and complexes with some metal chelates. These complexes are generally of an end-on nature, e.g., N≡N–Fe. Stiefel suggested that N_2 first forms a complex of this type with an iron atom of the MoFe-protein.[44] Then an atom of Mo(IV) could donate two electrons to the N_2 (Eq. 24-9, step *a*) to form a complex of N_2 and Mo(VI). Addition of two protons (Eq. 24-9, step *b*) would yield a molecule of **diimide**, which would stay bound at the



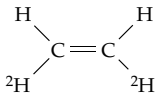
(24-9)

iron site while the molybdenum underwent another round of reduction. The diimide could be reduced to hydrazine and finally to ammonia (Eq. 24-10):



(24-10)

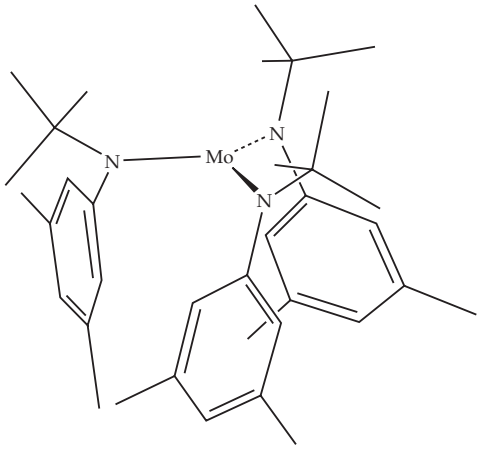
Mo(VI) attracts electrons sufficiently strongly that protons bound to surrounding ligands, such as H_2O, tend to dissociate completely. Thus, the molybdate ion MoO_4^{2-} is not protonated. The same would be true of nitrogenous ligands of a protein that might be coordinated with the bound molybdenum. On the other hand, reduction to Mo(IV) would tend to favor protonation of ligands such as the His 442 imidazole seen in Fig. 24-3A. Concurrently with the electron transfer from molybdenum to N_2 these protons could be transferred to the N_2 molecule (Eq. 24-10). The fact that strictly *cis*-dideuteroethylene is formed from acetylene in the presence of ^2H_2O is in accord with this idea.



However, looking at the FeMo-co molecule and the crowded surroundings of the Mo atom it may be more likely that reduction of N_2 occurs while it is bound to iron. Theoretical calculations as well as experimental data support this possibility.[44a] Recent crystallographic studies at a resolution of 0.12 nm revealed the presence of an atom, probably N, coordinated to six Fe atoms of FeMo-Co. This suggets, as previously proposed by Thorneley and Lowe,[44b,c] that a nitride ion (N_3^-) may be an intermediate in the formation of N_2.

Many mutant forms of nitrogenase have been investigated. Substitutions of His 195, Lys 191, and Gly 69 of the α chain affect reactions with various substrates.[45–45d] For example, the mutant obtained by substitution of His 195, whose imidazole forms an N–H--S hydrogen bond to a central bridging sulfide atom of FeMo-co (Fig. 24-3A), with glutamine (H195Q mutant) reduces N_2 only very slowly.[45] However, it still reduces both acetylene and protons.[27,44a,45b] Thus, it may be that different modes of substrate binding are needed for the individual steps of Eq. 24-10.

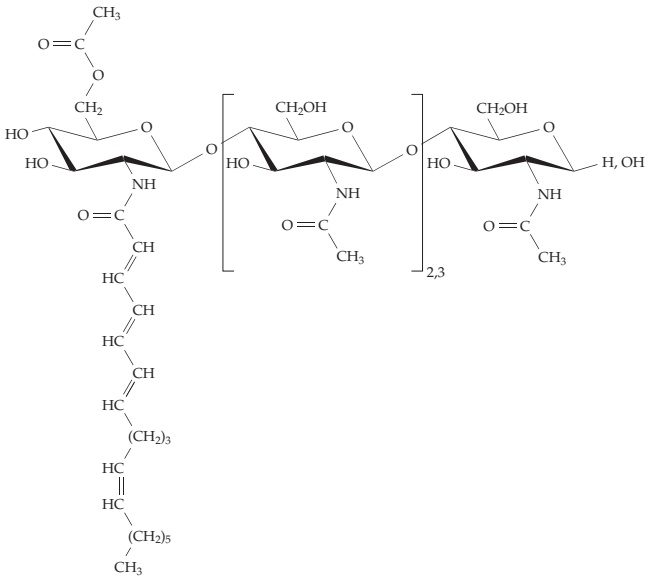
Because of the practical significance to agriculture there is interest in devising better nonenzymatic processes for fixing nitrogen using nitrogenase models that mimic the natural biological reaction.[42,46–49a] One interesting catalyst is the following molybdenum complex Mo(III)(NRAr)$_3$ where R = $C(C_2H_3)_2CH_3$ and Ar = 3,5-dimethylphenyl.[47,50,51]



Many other synthetic complexes have been studied including cubic $MoFe_3S_4$ clusters.[46] However, no exact chemical model for the FeMo-co coenzyme has been developed, and the rates of reaction for all of the model reactions are much slower than those of nitrogenases.

Nitrogen fixation genes. At least 17 genes needed for nitrogen fixation are present in the 23-kb *nif* region of the *Klebsiella* chromosome[52,53] (Fig. 24-4). A similar gene cluster in *A. vinlandii*[54] contains five polygenic transcriptional units and one monogenic unit. The nitrogenase structural genes are *nifK, D*, and *H* as is indicated in Fig. 24-4. The *nifF* and *J* genes encode associated electron-transport proteins. *NifM* is needed to activate the Fe-protein in an unknown fashion. *NifS* encodes a cysteine desulfhydrase needed for assembly of Fe–S clusters in the nitrogenase and the nifU and nifY proteins assist the assembly.[54a,b] The chaperone GroEL is also required.[54c] *Nif Q, B, V, X, N,* E_1, and *H* are needed for synthesis of FeMo-co and for its incorporation into the MoFe-protein.[55,55a] *NifA* is an activator gene for the whole cluster including the *nifL* gene product, which is altered by the presence of O_2 or of glutamine. Accumulation of the latter in cells (see Section B,2) strongly represses transcription of the nitrogenase genes.

Legume nodules and cyanobacterial heterocysts. Nitrogen fixation requires an anaerobic environment. Free-living bacteria fix nitrogen only when anaerobic. However, Rhizobia produce their own anaerobic environment by symbiotic association with the roots of legumes.[10,57–59] Formation of root nodules is a genetically determined process, several nodulation (*nod*) genes of the bacterium being required along with an unknown number of plant genes.[57,58] Initiation of nodulation results from a two-way molecular conversation between root hairs of the plant and bacterial cells.[60,61] The roots secrete **flavonoid compounds** (Chapter 21) which are recognized by bacterial sensors and induce transcription of the *nod* genes. Several of these genes encode enzymes required for synthesis of **Nod factors**,[62,63] small β-linked *N*-acetyl-D-glucosamine oligosaccharides containing 3–5 sugar residues and an *N*-linked long-chain fatty-acyl substituent at the nonreducing terminus (**lipochitooligosaccharides**). See also Box 20-E. Genes *nodA, B, C* specify enzymes needed for synthesis of the oligosaccharide core present in all Nod factors.



Structure of a Nod factor secreted by *Rhizobium leguminosarum*.[64]

Other Nod genes provide for modifications that restrict infection to specific species of legumes. For example, *nodS* encodes a methyltransferase and *nodU* a carbomoyltransferase.[65] NodH is a sulfotransferase.[66] NodD is a transcriptional activator that binds to DNA and induces the synthesis of the other Nod factors needed to initiate nodulation.[67] When an appropriate Nod factor is recognized, the root hairs on the legume

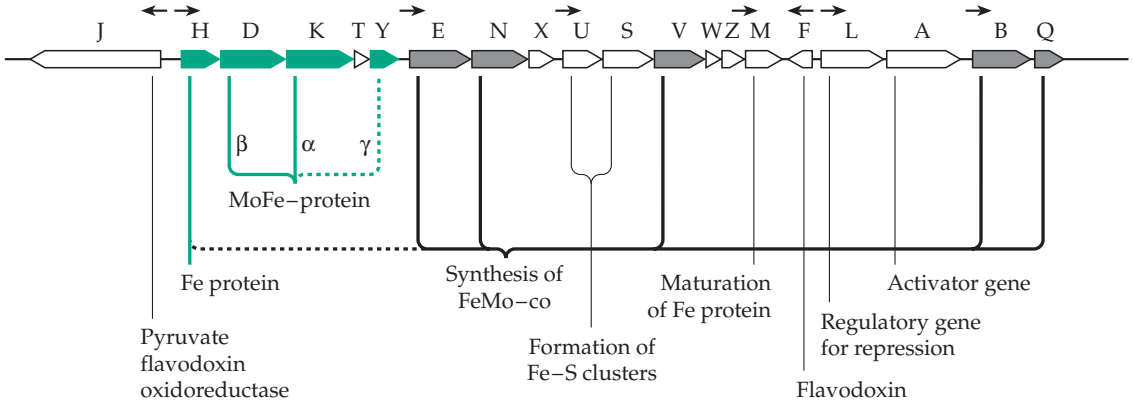


Figure 24-4 Sequence of *nif* genes of *Klebsiella pneumoniae*.[56] These precede the *his* operon directly at the right side. The nitrogenase structural genes are marked with green.

curl around the bacteria to initiate nodulation.[60,61] However, there are other factors. Infecting bacteria must reach a region of low oxygen in the plant. A hemoprotein **FixL** is a sensor kinase that regulates phosphorylation of transcription factor **FixJ**. This two-component system induces transcription of *nifA* (Fig. 24-4) and others.[68–69a] Nitrogen-fixing nodules, which are filled with the bacteroids derived from the infecting bacteria, synthesize leghemoglobin. The polypeptide chain of this protein is encoded by the plant, but its heme may be synthesized by bacteroid enzymes.[10,57] In at least one strain of *Rhizobium* the *nod* genes as well as the *fix* and *nif* genes are all carried on a 536-kb plasmid, which is almost as large as the whole 580-kb genome of *Mycoplasma genitalium* (Table 1-3).[70,71] This arrangement seems to have allowed these bacteria to form an unusually large number of Nod factors and to colonize a wider variety of hosts including a non-leguminous tree.

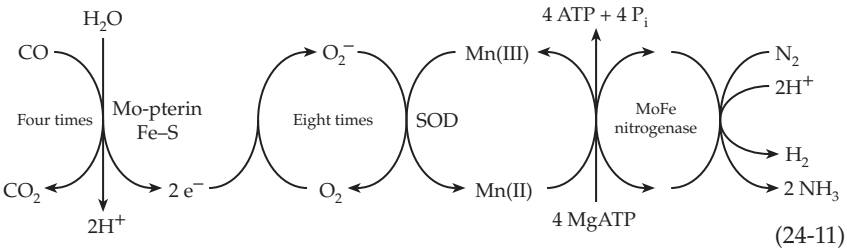
The H_2 that is produced in Eq. 24-6 (step *c*) may be used by bacteroids or by the plant cells. Some nodules evolve H_2, but in others it is utilized by hydrogenases as a source of energy.[57] From Eq. 24-6 it can be seen that up to 1/4 of the ATP utilized can, ideally, be recovered by use of the H_2 in this manner.

In cyanobacteria nitrogen fixation occurs in the **heterocysts**, specialized cells with thickened cell envelopes. They supply NH_4^+ to other cells in the filament of which they are a part. The cell envelopes prevent rapid diffusion of O_2 into the cells but do permit rapid enough entry of N_2 to maintain the observed rate of fixation of N_2.[72] In actinomycetes of the genus *Frankia*, which forms root nodules with woody plants, nitrogen fixation occurs in vesicles that are sheathed by multiple layers of **hopanoid lipids** (see Chapter 22).[73]

Genetic engineering. Because of the high cost of nitrogen fertilizers there is intense interest in improving biological nitrogen fixation. Ideas range from increasing the efficiency of nitrogenase by using fewer molecules of ATP, by limiting excessive evolution of H_2, or transferring the whole *nif* region of a bacterial genome into nonleguminous plants. The last proposal has generated much publicity, but it will probably be difficult because of the need to create an anaerobic environment suitable for nitrogen fixation. A crop plant engineered in this way might not resemble the hoped-for product. It would have an enormous energy requirement for nitrogen fixation, which would have to be met by photosynthesis. At present genetic engineering on *nif* genes to increase efficiency seems most likely to succeed.

Other nitrogenases. Although the well-characterized Mo-containing nitrogenase is responsible for most of their nitrogen fixation, bacteria often have alternative nitrogen fixation systems.[74] *Azotobacter vinlandii* produces three different nitrogenases in response to varying metal compositions in its surroundings.[75–76a] When the molybdenum level is adequate nitrogenase 1 is formed with its FeMo-coenzyme. In a low-molybdenum environment containing vanadium nitrogenase 2 is formed with an FeV-coenzyme.[75–77] If both molybdenum and vanadium are lacking, the bacteria form nitrogenase-3, which has an iron-only FeFe-coenzyme.

An unusual nitrogenase is formed by the chemolithotrophic *Streptomyces thermoautotrophicus*, which obtains energy from reduction of CO_2 or CO by H_2 (Eq. 17-50). These organisms form a MoFe nitrogenase that utilizes a manganese-containing superoxide dismutase to generate superoxide anion radicals. The latter transfer electrons to the MoFe protein in an ATP-dependent process. Electrons for generation of superoxide are formed using another molybdenum enzyme, a CO dehydrogenase containing molybdpterin cytosine dinucleotide (Fig. 16-31) and Fe–S centers.[78] The two systems function together as indicated by Eq. 24-11.



(24-11)

3. Interconversion of Nitrate, Nitrite, and Ammonium Ions

As is indicated in Fig. 24-1, the interconversions of nitrate and nitrite with ammonia and with organic nitrogen compounds are active biological processes. Two genera of nitrifying soil bacteria, which are discussed in Chapter 18, oxidize ammonium ions to nitrate. *Nitrosomas* carries out the six-electron oxidation to nitrite (Eq. 18-17) and *Nitrobacter* the two-electron oxidation of nitrite to nitrate (Eq. 18-18).[79]

The opposite sequence, reduction of nitrate and nitrite ions, provides a major route of acquisition of ammonia for incorporation into cells by bacteria, fungi, and green plants (Fig. 24-1). **Assimilatory** (biosynthetic) **nitrate reductases** catalyze the two-electron reduction of nitrate to nitrite (Eq. 16-61). This is thought to occur at the molybdenum atom of the large ~900-residue highly regulated[79a] molybdopterin-dependent enzyme. In green plants the reductant is

usually NADH while in fungi it is more often NADPH.[80–82] In all cases the cofactors FAD, heme, and molybdopterin are bound to a single polypeptide chain with the molybdopterin domain near the N terminus, and the heme in the middle. The electron-accepting FAD domain is near the C terminus[83–85a] and is thought to transfer the two electrons through the following chain.

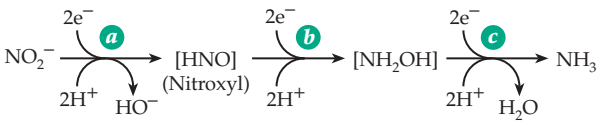
$$\text{NAD(P)H} \rightarrow \text{FAD} \rightarrow \text{Cyt } b \rightarrow \text{Mo-pterin} \rightarrow NO_3^- \qquad (24\text{-}12)$$

Bacterial assimilatory nitrate reductases have similar properties.[86,86a] In addition, many bacteria, including *E. coli*, are able to use nitrate ions as an oxidant for **nitrate respiration** under anaerobic conditions (Chapter 18). The **dissimilatory nitrate reductases** involved also contain molybdenum as well as Fe–S centers.[85] The *E. coli* enzyme receives electrons from reduced quinones in the plasma membrane, passing them through cytochrome *b*, Fe–S centers, and molybdopterin to nitrate. The three-subunit αβγ enzyme contains cytochrome *b* in one subunit, an Fe_3S_4 center as well as three Fe_4S_4 clusters in another, and the molybdenum cofactor in the third.[87] Nitrate reduction to nitrite is also on the pathway of denitrification, which can lead to release of nitrogen as NO, N_2O, and N_2 by the action of **dissimilatory nitrite reductases**. These enzymes[87a] have been discussed in Chapters 16 and 18.

Assimilatory nitrite reductases of plants, fungi, and bacteria carry out the six-electron reduction of nitrite to ammonium ions (Eq. 24-13) using electron donors such as reduced ferredoxins or NADPH.

$$NO_2^- + 6\,e^- + 8\,H^+ \rightarrow NH_4^+ + 2\,H_2O \qquad (24\text{-}13)$$

The enzymes from green plants and fungi are large multifunctional proteins,[80] which may resemble assimilatory sulfite reductases (Fig. 16-19). These contain **siroheme** (Fig. 16-6), which accepts electrons from either reduced ferredoxin (in photosynthetic organisms) or from NADH or NADPH. FAD acts as an intermediate carrier. It seems likely that the nitrite N binds to Fe of the siroheme and remains there during the entire six-electron reduction to NH_3. Nitroxyl (NOH) and hydroxylamine (NH_2OH) may be bound intermediates as is suggested in steps *a*–*c* of Eq. 24-14.



(24-14)

B. Incorporation of NH_3 into Amino Acids and Proteins

Prior to 1935, amino acids were generally regarded as relatively stable nutrient building blocks. That concept was abandoned as a result of studies of the metabolism of $^{15}NH_3$ and of ^{15}N-containing amino acids by Schoenheimer and Rittenberg[88] and more recent studies using ^{13}N by Cooper *et al.*[89,90] These investigations showed that nitrogen could often be shifted rapidly between one carbon skeleton and another. This confirmed proposals put forth earlier by Braunstein, Meister, and others who had pointed out that the C_4 and C_5 amino acids, aspartate and glutamate, which are closely related to the tricarboxylic acid cycle, are able to exchange their amino groups rapidly with those of other amino acids via transamination (Fig. 24-5, step *d*). Since ammonia can be incorporated readily into glutamate (Fig. 24-5, step *a*; see next section), a general means is available for the biosynthesis of amino acids. The citric acid cycle is able to provide any needed amount of 2-oxoglutarate for the synthesis of both glutamate and glutamine.[91–94]

Glutamine, and to a lesser extent asparagine, act as soluble, nontoxic carriers of additional ammonia in the form of their amide groups. An active synthase converts glutamate and ammonia to glutamine (Fig. 24-5, step *d*), and another enzyme transfers the amide nitrogen into aspartate, in an ATP-dependent reaction, to form asparagine (Fig. 24-5, step *e*). The amide nitrogen of glutamine is incorporated in a similar way into a great variety of other biochemical compounds, including carbamoyl phosphate (Fig. 24-5, step *f*; Section C,2), glucosamine (Eq. 20-5), NAD^+, *p*-aminobenzoate,

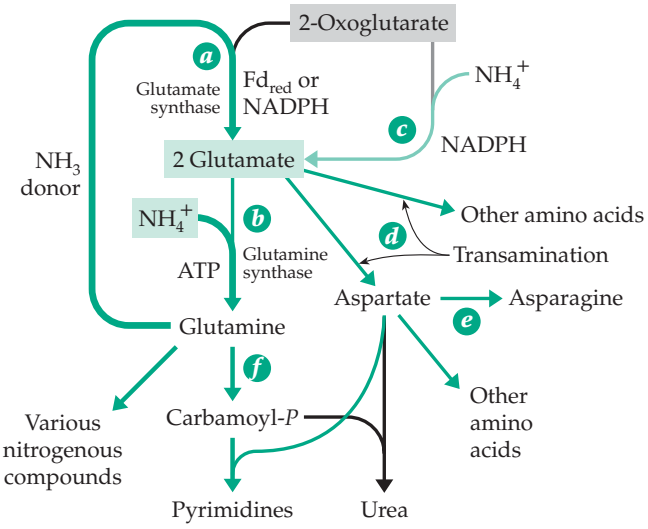


Figure 24-5 Major pathways of incorporation of nitrogen from ammonium ions into organic compounds, traced by green arrows.

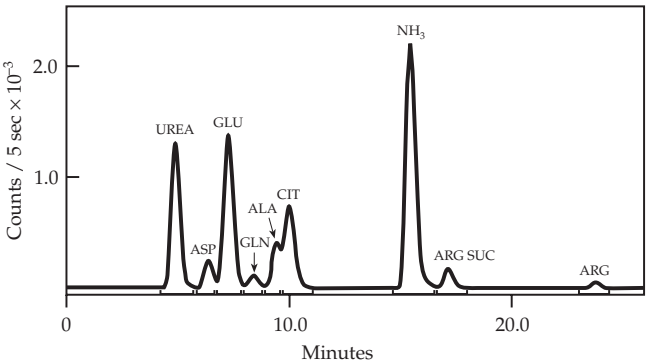
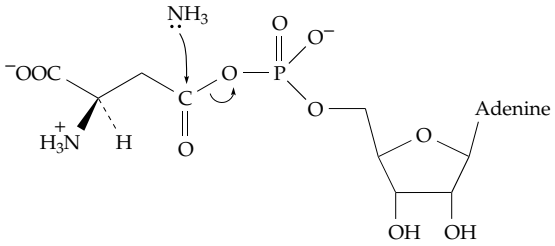


Figure 24-6 Elution profile of ^{13}N-containing metabolites extracted from liver 15 s after injection of $^{13}NH_3$ into the portal vein of an anesthetized adult male rat. CIT, citrulline; ARG SUC, argininosuccinate. From Cooper *et al.*[89]

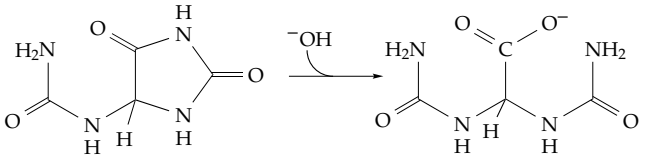
histidine, CTP, and purines (Chapter 25). These reactions are catalyzed by a family of **amidotransferases**,[95–97d] which hydrolyze the glutamine to glutamate and NH_3. The last is entrapped until it reacts with the second substrate. **Asparagine synthetase** apparently first binds ATP and aspartate, which probably react to form β-aspartyl adenylate (β-aspartyl-AMP). Glutamine then binds and is hydrolyzed.[98,98a]



β-Aspartyl adenylate

The liberated NH_3 can attack the β-aspartyl-AMP as indicated in the accompanying diagram to form asparagine. However, it is also possible that NH_3 is transferred via covalently bonded complexes[99] and is never free NH_3. An asparagine synthetase that utilizes free NH_3 as a nitrogen donor is also present in many organisms.

A third mechanism of synthesis, which was only recently recognized, appears to provide the sole source of asparagine for many bacteria.[98b] The asparagine-specific transfer RNA $tRNA^{Asn}$ is "mischarged" with aspartic acid to form Asp-$tRNA^{Asn}$. This compound is then converted to the properly aminoacylated Asn-$tRNA^{Asn}$ by a glutamine-dependent amidotransferase. (The entire ATP-dependent sequence is shown in Eq. 29-6.) The activated asparaginyl group is then transferred from Asn-$tRNA^{Asn}$ into proteins as they are synthesized.



Allantoin Allantoate

(24-15)

Most green plants transport nitrogen from roots to growing shoots as asparagine. However, in peanuts **β-methyleneaspartate** is the major nitrogen carrier,[100] and in some legumes, including soybeans, **allantoin** and **allantoate** (Eq. 24-15) play this role. Allantoin arises from hydrolysis of purines (see Fig. 25-18), which are synthesized in root nodules of nitrogen-fixing plants.[101]

Glutamate, glutamine, and aspartate also play central roles in *removal* of nitrogen from organic compounds.[102] Transamination is reversible and is often the first step in catabolism of excess amino acids. 2-Oxoglutarate is the recipient of the nitrogen, and the glutamate that is formed can be deaminated to form ammonia which can then be incorporated into glutamine. Glutamate can also donate its nitrogen to form aspartate. In the brain glutamate is a major neurotransmitter but is toxic in excess. The astrocyte glial cells take up glutamate from the synaptic clefts between neurons, converting it to glutamine, which is then released into the extracellular space for reuptake by neurons.[103,104] In the animal body both aspartate and glutamine (via carbamoyl phosphate) are precursors of **urea**, the principal nitrogenous excretion product. These relationships are also summarized in Fig. 24-5, and details are provided in later sections.

While reductive amination of glutamate via glutamate synthase appears to be the major pathway for incorporation of nitrogen into amino groups, some direct amination of pyruvate and other 2-oxoacids in reactions analogous to that of glutamate dehydrogenase occurs in bacteria.[105,106] Another bacterial enzyme catalyzes reversible addition of ammonia to fumarate to form aspartate (p. 685).

An initially surprising conclusion drawn from the studies of Schoenheimer and Rittenberg was that proteins within cells are in a continuous steady state of synthesis and degradation. The initial biosynthesis, the processing, oxidative and hydrolytic degradative reactions of peptides, and further catabolism of amino acids all combine to form a series of metabolic loops as discussed in Chapter 17 and dealt with further in Chapters 12 and 29. Within cells some proteins are degraded much more rapidly than others, an important aspect of metabolic control. This is accomplished with the aid of the ubiquitin system (Box 10-C) and proteasomes (Box 7-A).[107] Proteins secreted into extracellular fluids often undergo more rapid turnover than do those that remain within cells.

1. Uptake of Amino Acids by Cells

While cells of autotrophic organisms can make all of their own amino acids, other organisms utilize many preformed amino acids. Human beings and other higher animals require several **essential amino acids** in their diets. Additional amounts of "nonessential" amino acids are also needed. It is true that amino groups can be transferred from one carbon skeleton to another among most of the amino acids. However, the body must take in enough amino groups to supply its need for all of the 20 amino acid components of proteins.[107a,b] Because of an unfavorable equilibrium constant, and the normally low concentration of NH_4^+, glutamate dehydrogenase (step *c* in Fig. 24-5) does not normally synthesize glutamate in the animal body. Its function is to deaminate excess glutamate. Furthermore, cells of some tissues take up amino acids that are made in other tissues. The active transport systems of bacteria have been described in Chapter 8. In mammals the absorption of amino acids takes place through epithelial cells of the intestinal tract, kidney tubules, and the brain (blood–brain barrier). Both Na^+-dependent transport (as for sugars; see Chapter 8) and Na^+-independent processes occur.[107c] Among the latter is the proposed **γ-glutamyl cycle,** which is described in Box 11-B. The cycle makes use of the γ-carboxyl group of glutamate, the same carboxyl that carries ammonia in the form of glutamine. Glutathione supplies the activated γ-glutamyl group. The amino acid to be transported reacts on the membrane surface by **transpeptidation**[108–109a] to form a **γ-glutamylamino acid** which enters the cytoplasm. It releases the free amino acid through an internal displacement by the free amino group of the glutamyl group. The natural tendency of the 5-carbon glutamate to undergo cyclization is used to provide the driving force for release of the bound amino acid. The cyclic product 5-oxoproline is then opened hydrolytically in an ATP-requiring reaction.[110] Cysteinylglycine formed in the initial transpeptidation is hydrolyzed by a peptidase, and glutathione is regenerated in two ATP-dependent steps as indicated in the scheme in Box 11-B.

The significance of the γ-glutamyl cycle is not fully understood. However, the finding of a mentally retarded individual who excretes 25–50 g / day of 5-oxoproline in the urine (possibly because of a defective 5-oxoprolinase) suggests that the pathway is a very active one.[111] A few persons deficient in γ-glutamyl transpeptidase have been found. They excrete glutathione and have a variety of medical problems.[109]

2. Glutamate Dehydrogenase and Glutamate Synthase

In animal tissues and in some bacteria the **glutamate dehydrogenase** reaction (Fig. 24-5, step *c*; see also Chapter 15)[112–115] provides a means of incorporating ammonia reversibly into glutamic acid. In eukaryotic cells the allosteric enzyme is found largely in the mitochondria.[115a] Glutamate dehydrogenase is also found in chloroplasts where it may function in glutamate synthesis when ammonia is present in excess.[116] The action of aminotransferases, both within and without mitochondria, distributes nitrogen from glutamate into most of the other amino acids. Especially active is aspartate aminotransferase (Eq. 14-24; Fig. 24-5, step *d*) which equilibrates aspartate and oxaloacetate with the 2-oxoglutarate–glutamate couple. However, the body obtains glutamate, as well as other amino acids, from foods, the initial source being largely green plants.

In plants as well as in *E. coli* and many other bacteria most glutamate is formed by **glutamate synthase**, which carries out reductive amination of 2-oxoglutarate (Fig. 24-5, reaction *a*). Glutamate synthase (also called **GOGAT**) utilizes the amide side chain of glutamine as the nitrogen donor. It is one of the previously mentioned amidotransferases in which glutamine is hydrolyzed to glutamate and NH_3 within the active site of the enzyme. Formation of a Schiff base and reduction probably occurs as in the reverse of reaction B of Table 15-1. However, one of the two glutamate molecules formed in reaction *a* of Fig. 24-5 must be reconverted by glutamine synthase to glutamine with the utilization of a molecule of ATP (Fig. 24-5, step *b*). Because of this coupling of ATP cleavage to the reaction the equilibrium in reaction *a* lies far toward the synthesis of glutamate. The low value of K_m for NH_4^+ that is characteristic of glutamine synthase favors glutamate synthesis even when little nitrogen is available.

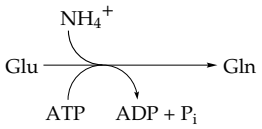
Bacterial glutamate synthases are large oligomeric proteins containing flavin and Fe–S centers. That of *Azospirillum brasilense* consists of αβ units in which the 53-kDa β chains contain FAD and an NADPH binding site.[117–118a] The NADPH evidently transfers electrons to the FAD, which transfers them to an Fe_3S_4 center in the large 162-kDA α subunit. A molecule of bound FMN receives the electrons and reduces the iminoglutarate to glutamate. The site of binding and hydrolysis of glutamine is also present in subunit α. Chloroplasts of higher plants contain two glutamate synthases. One resembles the bacterial enzyme and utilizes NADPH as the reductant. The other requires reduced ferredoxin.[118–120]

Bacteria utilize both D-alanine and D-glutamate in the synthesis of their peptidoglycan layers (Fig. 8-29). Both D-amino acids are formed by racemases. That of

alanine uses PLP (Chapter 15) but **glutamate racemase**[121–122a] does not. It may be able to remove the α-H of glutamate by utilizing the –COOH of the substrate, rather than the PLP ring, as an electron sink. Small amounts of D-amino acids occur also in animals.[123] Animal livers and kidneys contain **D-amino acid oxidase** and **D-aspartate oxidase,** which apparently function to metabolize D-amino acids from foods or those formed by brain activity (Chapter 30) or by aging.

3. Glutamine Synthetase

The formation of glutamine from glutamate (Eq. 24-16) also depends upon a coupled cleavage of ATP:



(24-16)

Glutamine synthase, as isolated from *E. coli*, contains 12 identical 51.6-kDa subunits arranged in the form of two rings of six subunits each with a center-to-center spacing of 4.5 nm. The units in one layer lie almost directly above those in the next,[104,124,125] the center-to-center spacing between the two layers is also 4.5 nm, and the array has 622 dihedral symmetry. The enzyme displays complex regulatory properties,[112,126–129] which are summarized in Fig. 24-7. The enzyme exists in two forms. **Active glutamine synthase** requires Mg^{2+} in addition to the three substrates glutamate, NH_4^+, and ATP. If the glutamate precursor, 2-oxoglutarate, is present in excess, the enzyme tends to remain in the active form because conversion to a modified form is inhibited; when the oxoglutarate concentration falls to a low value and glutamine accumulates, alteration is favored. The modifying enzyme **adenylyltransferase** (AT) in its active form AT_A transfers an adenylyl group from ATP to a tyrosine hydroxyl on glutamine synthase to give an adenylyl enzyme (GS-AMP). This **modified enzyme** requires Mn^{2+} instead of Mg^{2+} and is far more sensitive than the original enzyme to feedback inhibition by a series of end products of glutamine metabolism. All nine of the feedback inhibitors (serine, alanine, glycine, histidine, tryptophan, CTP, AMP, carbamoyl-*P*, and glucosamine 6-*P*) seem to bind to specific sites on the enzyme surface and to exert a cumulative inhibition. Serine, alanine, and glycine appear to be competitive inhibitors at the glutamate binding site.[130]

Relaxation of adenylylated glutamine synthase to the unmodified form is not catalyzed by a separate hydrolase but is promoted by a modified form of the adenylyltransferase AT_D. The active enzyme AT_A is

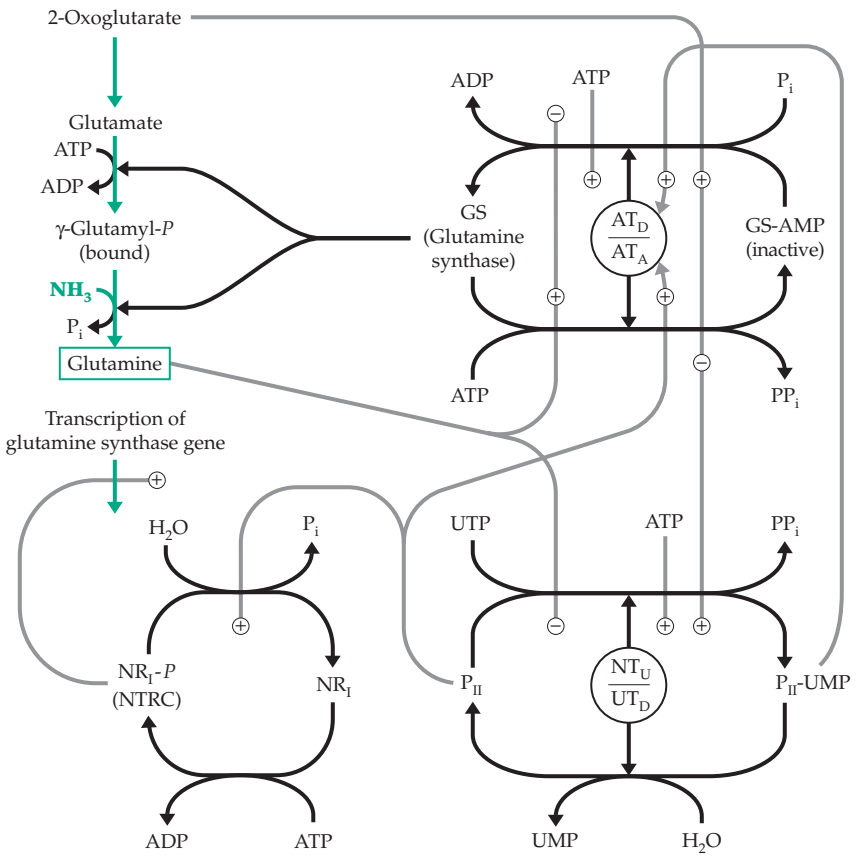
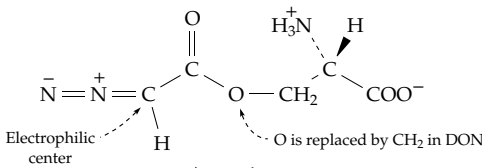


Figure 24-7 Regulation of glutamine synthase of *E. coli* using activation (+) and inhibition (–). Glutamine synthase (GS upper center) is converted by adenylylation of Tyr 397 into an inactive form GS-AMP by the action of an adenylyltransferase AT_A in complex with regulatory protein PII. PII is uridylylated at up to four sites by action of uridylyltransferase UT_U, which resides in the same polypeptide chain as a uridylyl removing enzyme UT_D (or UR). When PII carries a uridylyl group (PII-UMP), AT_A is transformed to AT_D, which reconverts the inactive GS-AMP to active GS by phosphorolytic removal of the adenylyl group. The ratios of AT_A / AT_D and UT_U / UT_D are controlled by the concentrations of the metabolites 2-oxoglutarate, a precursor, and glutamine, the immediate product. The amount of GS formed is controlled at the transcriptional level by an enhancer-binding transcription factor called NRI or NtrC (lower left). It is active when phosphorylated. Dephosphorylation of NRI-*P* is catalyzed by yet another protein and is stimulated by PII. Thus, PII both decreases synthesis of GS and promotes conversion of GS to its inactive form.

actually a complex AT • PII containing the regulatory protein PII. Subunit PII can be uridylylated on a tyrosine side chain by action of a 95-kDa **uridylyltransferase** (UT)[127,131] to form the modified glutamine synthetase AT • PII-UMP orAT_D. This form catalyzes phosphorolytic deadenylylation of glutamine synthetase, P_i displacing the adenylyl group to form ADP. Removal of the uridylyl group from PII-UMP is catalyzed by a fourth enzyme, UT_D (or UR), which is part of the same polypeptide chain as UT_U.[127] The cycle of interconversions of PII catalyzed by the UT_U and UT_D activities is shown at the lower right side of Fig. 24-7. From the allosteric modification reactions indicated by the gray lines, it is seen that glutamine not only promotes the adenylylation of glutamine synthetase but also inhibits the uridylylation of PII, thereby preventing AT_D from removing the adenylyl group from the synthetase. Furthermore, it allosterically inhibits the deadenylylation reaction itself. On the other hand, 2-oxoglutarate acts in the opposite way.

The glutamine synthase regulatory system has another important function. Protein PII stimulates the dephosphorylation of the enhancer-binding transcriptional regulator NRI-*P* (NtrC-*P*).[131,132] This slows transcription of the glutamine synthase gene (see Fig. 24-7) as well as a variety of other genes including those for the nitrogenase proteins in organisms that have them.[133] As a consequence, a deficiency of glutamine turns on a number of genes involved in nitrogen metabolism. Accumulation of glutamine promotes PII accumulation, modification of the synthase, and loss of gene activation.

Nitrogen can be transferred from glutamine into many other substrates.[102] Several antibiotic analogs of glutamine have been useful in studying these processes. Examples are the streptomyces antibiotics **L-azaserine** and 6-diazo-5-oxo-L-norleucine (DON).



L-Azaserine

These compounds act as alkylating agents; N_2 is released and a nucleophilic group from the enzyme becomes attached at the carbon atom indicated.[134] Other inhibitors bind noncovalently to form dead-end complexes.[134a]

4. Catabolism of Glutamine, Glutamate, and Other Amino Acids

Glutamine is hydrolyzed back to glutamate by glutaminases that are found both in eukaryotic tissues and in bacteria.[135,136] Liver contains an isozyme whose function appears to be to release NH_3 from glutamine for urea synthesis.[135,137] Glutamate dehydrogenase deaminates excess glutamate back to 2-oxoglutarate, which is degraded to succinyl-CoA and via β oxidation to malate, pyruvate, and acetyl-CoA. The last can reenter the citric acid cycle and be oxidized to CO_2 (Eq. 24-17). In fact, in mammalian tissues glutamate is essentially in equilibrium with 2-oxoglutarate and other citric acid cycle intermediates (see Box 17-C).

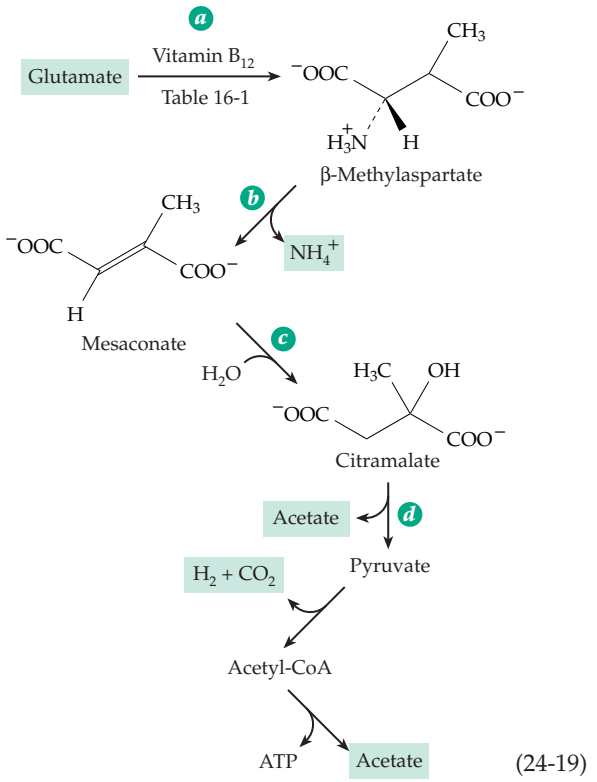
$$\text{Glu} \rightarrow \text{2-oxoglutarate} \rightarrow \text{succinyl-CoA} \rightarrow \rightarrow \rightarrow \text{malate} \rightarrow \text{pyruvate} \rightarrow \text{acetyl-CoA} \rightarrow CO_2 \quad (24\text{-}17)$$

Many other amino acids are degraded in similar ways. In most cases the sequence is initiated by transamination to the corresponding 2-oxoacid. Beta oxidation and breakdown to such compounds as pyruvate and acetyl-CoA follows.

Catabolism initiated by decarboxylation. An alternative pathway for glutamate degradation is through the γ-aminobutyrate shunt (Fig. 17-5). This pathway is initiated by a PLP-dependent decarboxylation rather than by a deamination or transamination. Since decarboxylases are known for most amino acids, there are usually alternative breakdown pathways initiated by decarboxylation. In many cases these pathways lead to important products. For example, γ-aminobutyrate functions in the brain as an important neurotransmitter. Dihydroxyphenylalanine is converted to noradrenaline and adrenaline, tryptophan to serotonin, and histidine to histamine. All of these are neurotransmitters (Chapter 30) and/or have other hormonal functions. A calmodulin-dependent glutamate decarboxylase occurs in higher plants, which accumulate γ-aminobutyrate in response to a variety of stresses.[138] However, the significance of this accumulation is unclear.

Fermentation of glutamate. Special problems face anaerobic bacteria subsisting on amino acids. Their energy needs must be met by balanced fermentations.[138a] For example, glutamate may be converted to CO_2, ammonia, acetate$^-$, and butyrate$^-$ according to the reactions of Fig. 24-8. The end result is described by Eq. 24-18.

$$2\ \text{Glutamate}^- + 2\ H_2O + H^+ \rightarrow 2\ CO_2 + 2\ NH_4^+ + 2\ \text{acetate}^- + \text{butyrate}^- \qquad \Delta G'\ (\text{pH } 7) = -131\ \text{kJ} \quad (24\text{-}18)$$



(24-19)

The sequence begins with the γ-aminobutyrate shunt reactions (Fig. 24-8, steps *a* and *b*), but succinic semialdehyde is reduced to γ-hydroxybutyric acid using the NADH generated in the trans-deamination process of step *c*. With the aid of a CoA-transferase (step *d*) two molecules of the CoA ester of this hydroxy acid are formed at the expense of two molecules of acetyl-CoA. Use is then made of a β,γ elimination of water (step *e*), analogous to that involved in the formation of vaccenic acid (Eq. 21-2). Isomerization (perhaps by the same enzyme that catalyzes elimination) forms crotonyl-CoA (step *f*). The latter undergoes dismutation, one-half being reduced to butyryl-CoA and one-half being hydrated and oxidized to acetoacetyl-CoA in the standard β-oxidation sequence. Acetoacetyl-CoA is cleaved to regenerate the two molecules of acetyl-CoA. The organism can gain one molecule of ATP through cleavage of the butyryl-CoA. Perhaps a second can be gained by oxidative phosphorylation between the NADH produced in the formation of acetoacetyl-CoA and the reduction of crotonyl-CoA to butyryl-CoA. The two processes take place at sufficiently different redox potentials to permit this kind of coupling.

Another fermentation of glutamate is initiated by the vitamin B_{12}-dependent isomerization of glutamate to β-methylaspartate (Eq. 24-19, step *a*).[138b] This rearrangement of structure permits α,β elimination of ammonia (step *b*), a process not possible in the original glutamate. Hydration to **citramalate** (step *c*) and aldol cleavage yields acetate and pyruvate. Acetate is one of the usual end products of the fermentation. The pyruvate can be cleaved to H_2, CO_2, and acetyl-CoA by the pyruvate-formate-lyase system (Fig. 15-16; Eq. 17-25), and cleavage of the acetyl-CoA can provide ATP. Alternatively, two molecules of acetyl-CoA can be coupled and reduced to butyryl-CoA. The reducing power generated in the cleavage of pyruvate is used to reduce crotonyl-CoA rather than being released as H_2. The stoichiometry is identical to that in Fig. 24-8. Still other fermentation mechanisms are used by some Clostridia to degrade glutamate.[138a] See study question number 16 at the end of this chapter.

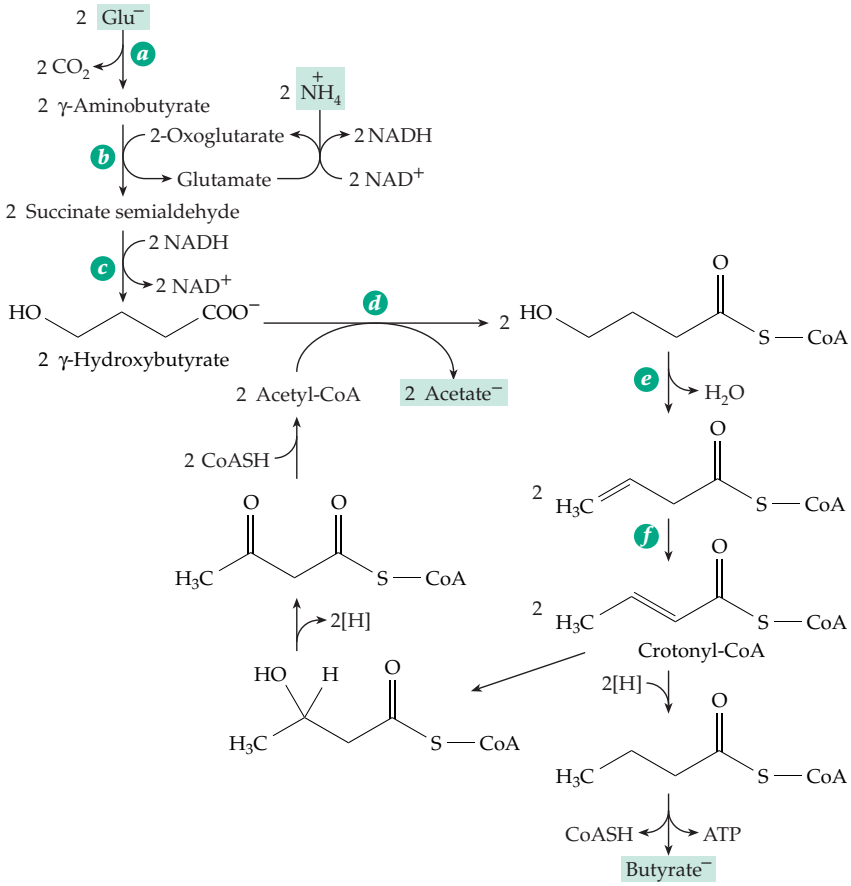


Figure 24-8 Fermentation of glutamate by *Clostridium aminobutylicum.*

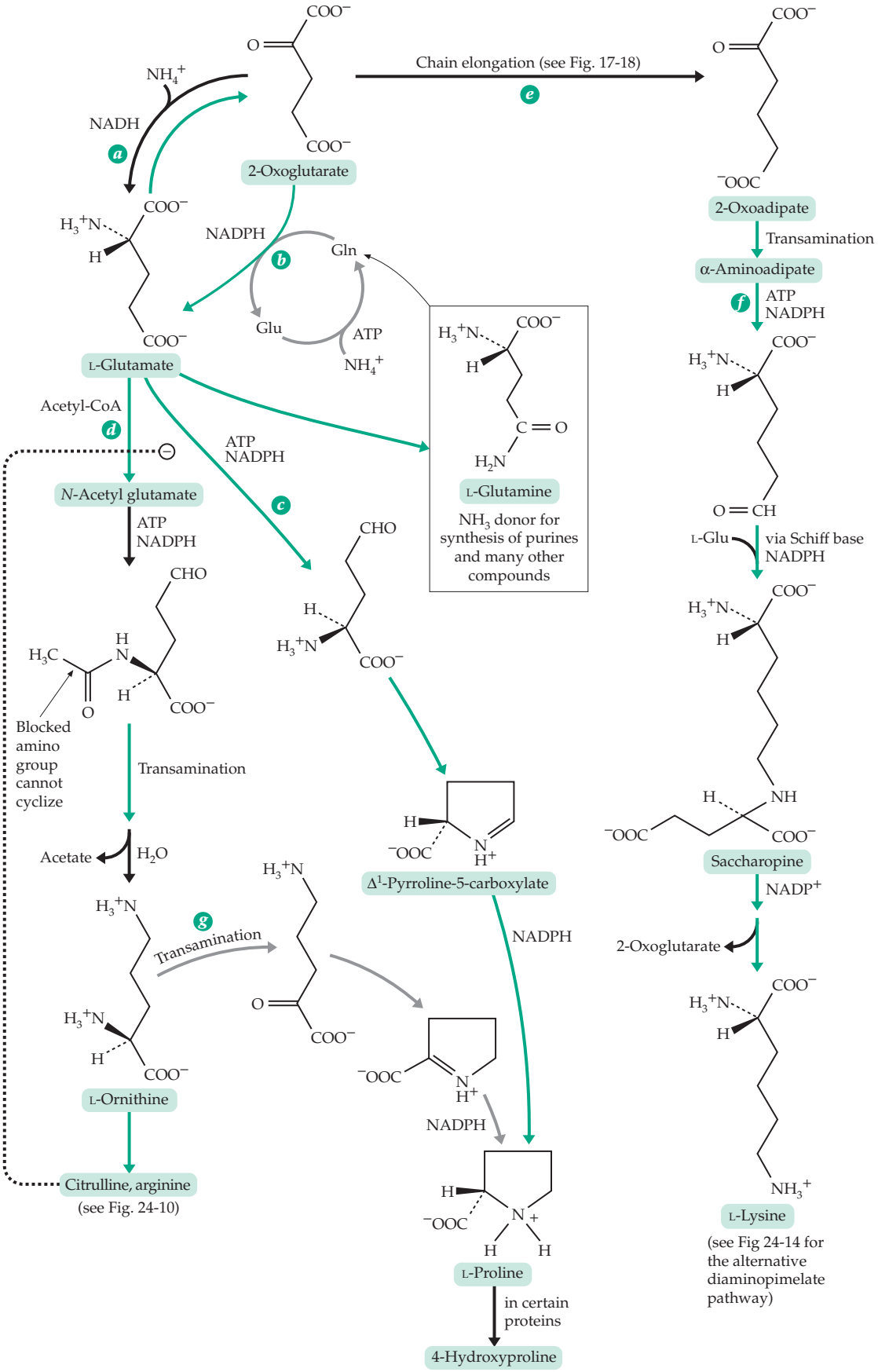


Figure 24-9 Biosynthesis of glutamate, glutamine, proline, and lysine from 2-oxoglutarate.

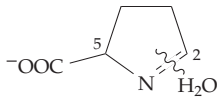
C. Synthesis and Catabolism of Proline, Ornithine, Arginine, and Polyamines

The 5-carbon skeleton of glutamic acid gives rise directly to those of proline, ornithine, and arginine. The reactions are outlined in Fig. 24-9. Arginine, in turn, is involved in the urea cycle, which is shown in detail in Fig. 24-10. Arginine is also a biosynthetic precursor of the polyamines. Another important biosynthetic product of glutamate metabolism is δ-aminolevulinate, a precursor to porphyrins (Eq. 24-44) in some organisms.[139]

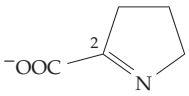
1. Synthesis and Catabolism of Proline

The ATP-dependent reduction of the γ-carboxyl group of glutamate to an aldehyde by NADPH (reaction *c*, Fig. 24-9) is of a standard biosynthetic reaction type, the opposite of the oxidation reaction of Fig. 15-6. Like the latter it is thought to occur via an acyl phosphate intermediate.[140,141] The oxidation product, **glutamate semialdehyde**, cyclizes and can be converted to proline by further reduction (Fig. 24-9, step *c* and subsequent reactions in center of scheme). The pathway has been well established in bacteria and yeast by both biochemical and genetic experiments. In plants both the initial reduction and the cyclization are catalyzed by a bifunctional enzyme.[142,143] An alternative pathway important in animals is initiated by transamination of ornithine to the corresponding 2-oxoacid, spontaneous cyclization, and reduction to proline (Fig. 24-9, step *g*).[140,144] Selected prolines in collagen and in plant glycoproteins[145] are oxygenated to form 4-hydroxyproline (Eqs. 8-6, 18-51).

One route of catabolism of proline is essentially the reverse of its formation from glutamate. **Proline oxidase** yields Δ^1-pyrroline 5-carboxylate.[145a,b]



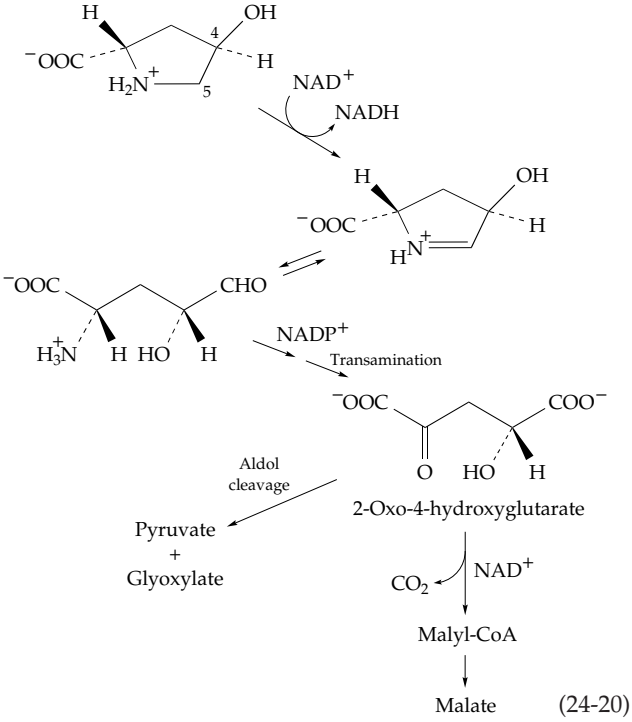
Δ^1-Pyrroline 5-carboxylate



Δ^1-Pyrroline 2-carboxylate

The corresponding open-chain aldehyde, formed by hydrolysis, can be oxidized back to glutamate by pyrroline 5-carboxylate dehydrogenase.[145a–147] Lack of this enzyme is associated with the human genetic deficiency causing **hyperprolinemia**.[147–148a]

Alternatively, degradation can be initiated by oxidation on the other side of the ring nitrogen to form Δ^1-pyrroline 2-carboxylate. The metabolic fate of this compound is uncertain. A corresponding pathway for breakdown of 4-hydroxy-L-proline of collagen yields glyoxylate and pyruvate or malate and CO_2 (Eq. 24-20).[149] Oxidation on the other side of the ring nitrogen



(24-20)

of hydroxyproline is utilized by some pseudomonads to convert the amino acid into 2-oxoglutarate. Anaerobic bacteria may reduce proline to 5-aminovalerate and couple this reaction to the oxidative degradation of another amino acid (Stickland reaction).

2. Synthesis of Arginine and Ornithine and the Urea Cycle

If the amino group of glutamate is blocked by acetylation prior to the reduction to the semialdehyde (Fig. 24-9, step *d*) cyclization is prevented. The γ-aldehyde group can be transaminated to an amino group and the acetyl blocking group removed to form **ornithine**. Ornithine is not usually a constituent of proteins, but it is sometimes formed by hydrolytic modification of arginine at specific sites in a protein. A 67-kDa urate-binding glycoprotein of plasma is reported to contain 43 residues of ornithine.[150,151] It is postulated that a special arginase is needed to form these residues, and that it may be lacking in some cases of gout in which the urate-binding capacity of blood is impaired. Ornithine appears to be present in specific sites in a few other proteins as well.[151] *Neurospora* grown in a minimal medium accumulates large amounts of both ornithine and arginine, over 98% of which is sequestered in vesicles within the cytoplasm.[152,153] This appears to be a way of accumulating a store of arginine that is protected from the active catabolism of that amino acid by the fungus. However, accumulation of ornithine in the human body, as a result of lack of **ornithine aminotransferase**

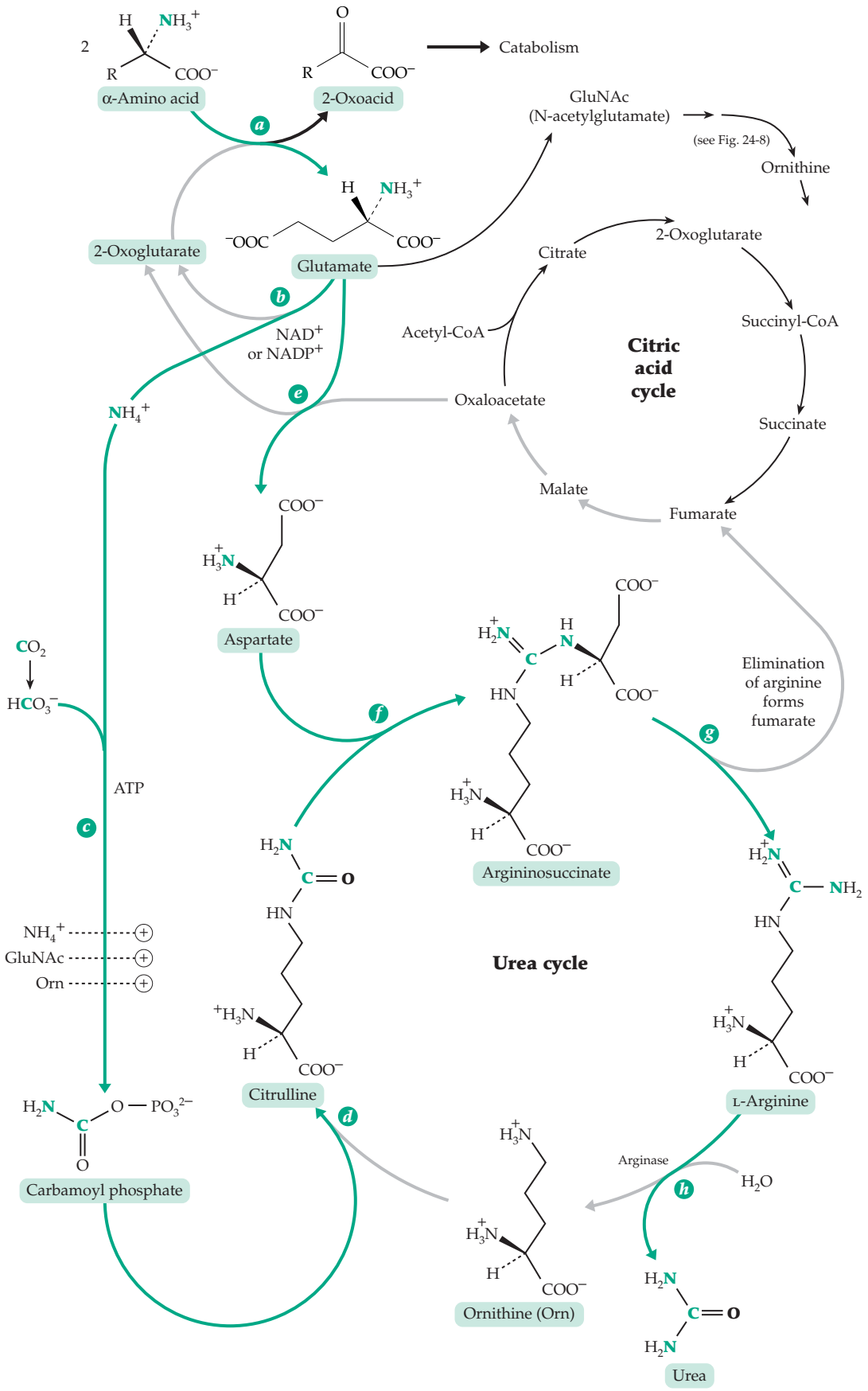


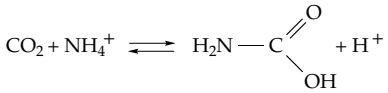
Figure 24-10 Biosynthesis of citrulline, arginine, and urea. The green arrows indicate reactions directly involved in deamination of amino acids and the synthesis of urea. N from amino acids and C from CO_2 are traced in green.

(Fig. 24-9, step *g*), causes gyrate atrophy of the choroid and retina, a disease that results in tunnel vision and blindness.[154,155] A major interest in arginine metabolism asises from its role in formation of urea in the human body. Study of arginine biosynthesis in bacteria has also been important in developing our understanding of regulation of gene expression.[156]

The urea cycle. In 1932, Krebs and Henseleit proposed that urea is formed in the liver by a cyclic process in which ornithine is converted first to **citrulline** and then to arginine.[157,158] The hydrolytic cleavage of arginine produces the urea and regenerates ornithine (Fig. 24-10, bottom). Subsequent experiments fully confirmed this proposal. Urea is the principal nitrogenous end product of metabolism in mammals and many other organisms, but the urea cycle reactions have other functions. As with the citric acid cycle, products other than urea can be withdrawn in any needed quantity. Most notably, the reactions of Fig. 24-10 provides for the biosynthesis of arginine in all organisms.[159,160] Also of physiological importance is the fact that the urea cycle involves both mitochondria and cytosolic enzymes.[161,162] This is illustrated in Fig. 24-11.

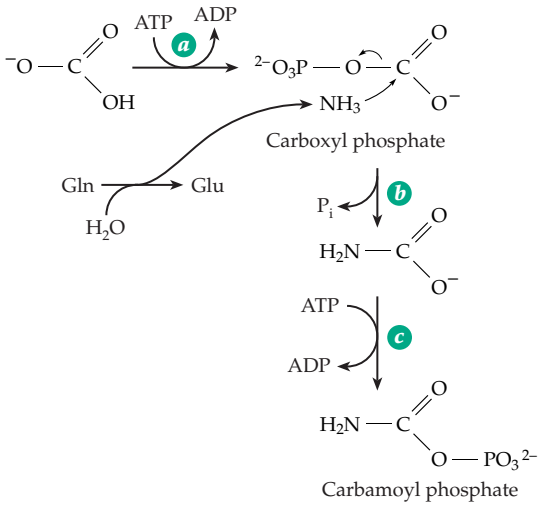
Let us trace the entire route of nitrogen removed by the liver from excess amino acids. Transaminases (step *a*, Fig. 24-10) transfer nitrogen to 2-oxoglutarate to form glutamate. Since urea contains two nitrogen atoms, two molecules of glutamate must donate their amino groups. One molecule is deaminated directly by glutamate dehydrogenase to form ammonia (step *b*). This ammonia is combined with bicarbonate (step *c*) to form carbamoyl phosphate, which transfers its carbamoyl group onto ornithine to form citrulline (step *d*). The second molecule of glutamate transfers its nitrogen by transamination to oxaloacetate (reaction *e*) to form aspartate. The aspartate molecule is incorporated intact into **argininosuccinate** by reaction with citrulline (reaction *f*). Undergoing a simple elimination reaction, the 4-carbon chain of argininosuccinate is converted to fumarate (step *g*) with arginine appearing as the elimination product. Finally, the hydrolysis of arginine (step *h*) yields urea and regenerates ornithine.

Carbamoyl phosphate synthetases. The first of the individual steps in the urea cycle is the formation of carbamoyl phosphate.[163] Carbon dioxide and ammonia equilibrate spontaneously with carbamic acid:



(24-21)

Some bacteria have a kinase able to convert carbamate into carbamoyl phosphate starting with step *a* of Eq. 24-22. However, the equilibrium constant is low (0.04 at pH 9, 10°C), and it is now believed that carbamate kinase functions in the opposite direction, providing a means of synthesis of ATP for bacteria degrading arginine (Section C,5,d). The biosynthetic carbamoyl phosphate synthases harness the cleavage of *two* molecules of ATP to formation of one molecule of carbamoyl phosphate (reaction *c*, Fig. 24-10).[163] In bacteria such as *E. coli*, a single synthase provides carbamoyl phosphate for biosynthesis of both arginine and pyrimidines (Fig. 25-14). However, fungi and higher animals have at least two carbamoyl-*P* synthases. Synthase I provides substrate for formation of citrulline from ornithine (Fig. 24-10), while carbamoyl-*P* synthase II, which is part of a larger multifunctional protein,[164] functions in pyrimidine synthesis. Synthase I is found in mitochondria and synthase II in the cytoplasm. Mammalian carbamoyl phosphate sythase I consists of a single 160-kDa peptide.[163] A powerful allosteric effector for the liver synthase is ***N*-acetylglutamate** (Fig. 24-10), a precursor of ornithine.[165] The enzyme from certain marine elasmobranchs, such as the spiny dogfish *Squalus acanthias*, have carbamoyl-*P* synthase III, an enzyme with somewhat different molecular properties.[166] It probably functions in synthesis of urea, which is used by these animals to regulate osmotic pressure.[167,168] Synthase I utilizes only free NH_3. The others are amidotransferases and prefer glutamine as the ammonia donor. Carbamoyl-*P* synthase from *E. coli* consists of two subunits (~42 and 118 kDa, respectively) and can utilize *either* free ammonia or glutamine.[169,170] The light subunit has **glutaminase** activity; i.e., it is able to hydrolyze glutamine to ammonia. All of these synthatases presumably act by first phosphorylating bicarbonate to an enzyme-bound carboxyl phosphate,[163,171,172] which can then undergo a displacement of phosphate by NH_3 to give enzyme-



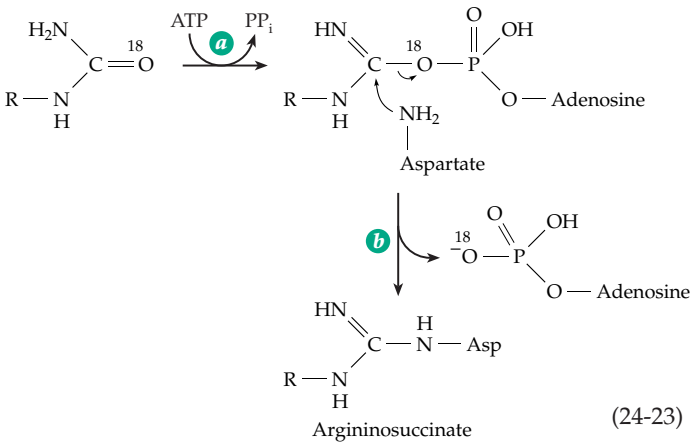
(24-22)

bound carbamate (Eq. 24-22, step *a*). Phosphorylation of the latter by ATP completes the reaction. In the single-chain enzymes the amidotransferase domain is at the N terminus.

Crystallographic study of a mutant form of the *E. coli* enzyme unable to act rapidly on glutamine showed that the latter released its ammonia to form a thioester with cysteine 269,[173,173a] suggesting a mechanism resembling that of serine proteases or papain (Chapter 12) for the glutaminase action. The X-ray crystallography also showed that the released NH_3 must travel 4.5 nm through the interior of the protein to the site of carbamate formation. The carbamate must travel another ~4.5 nm to the site from which carbamoyl phosphate is released.[170,172,173b] The C-terminal regions of the synthases undergo allosteric modification by a number of effectors.[163,173c] Both ornithine and IMP are activators for the *E. coli* enzyme, whereas UMP, a pyrimidine end product, exerts feedback inhibition. Phosphoribosyl pyrophosphate activates synthase II, and *N*-acetylglutamate activates the mammalian liver synthase I by binding near the C terminus.[165]

Citrulline and argininosuccinate. One NH_3 and one HCO_3^- for urea formation are provided by the carbamoyl group, which is transferred from carbamoyl-*P* to ornithine to form citrulline. The second nitrogen atom is transferred from glutamate to aspartate into argininosuccinate (steps *d* and *f*, Fig. 24-10). The equilibrium constant for ornithine transcarbamoylase (reaction *d*) is very high so that ornithine is completely converted to citrulline. The trimeric human enzyme is a trimer of 36-kDa subunits[174–175a] whose structural gene is on the X chromosome. Like many other mitochondrial matrix enzymes it is synthesized as a larger (40 kDa) pre-cursor, which enters the mitochondria in an energy-dependent process.[176] A genetic defect in this sex- linked gene is often lethal to boys, and even girls, heterozygous for the defect, sometimes have serious problems with accumulation of ammonia in the brain.[162,174,177]

The conversion of citrulline to argininosuccinate and the subsequent breakdown to fumarate and arginine take place in the cytosol (Fig. 24-11). The ureido



(24-23)

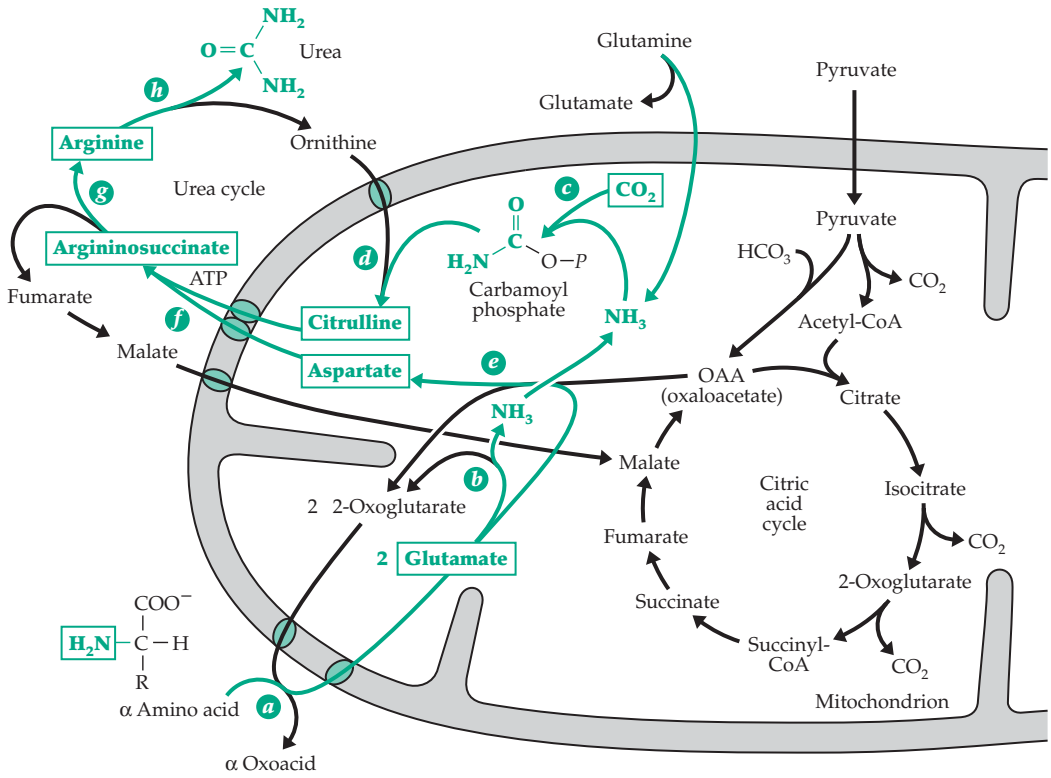


Figure 24-11 Integration of the urea cycle with mitochondrial metabolism. Green lines trace the flow of nitrogen into urea upon deamination of amino acids or upon removal of nitrogen from the side chain of glutamine.

group of citrulline is activated by ATP for the argininosuccinate synthase reaction (Eq. 24-23, step *a*). Thus, ^{18}O present in this group is transferred into AMP. A citrulline adenylate intermediate (center) is likely. **Argininosuccinate lyase** (reaction *g*, Fig. 24-10)[177a] catalyzes the elimination of arginine with formation of fumarate. It is entirely analogous to the bacterial aspartase that eliminates ammonia from aspartate to form fumarate.[178] Like the latter enzyme and fumarate hydratase (Chapter 13), argininosuccinase promotes a trans elimination.[179] The fumarate produced can be reconverted through reactions of the citric acid cycle to oxaloacetate, which can be reaminated to aspartate (Fig. 24-11). Aspartate is used to introduce amino groups in an entirely similar way in other metabolic sequences such as in the formation of adenylic acid from inosinic acid (Fig. 25-16).

The cleavage of arginine to ornithine and urea by the Mn^{2+}-containing **arginase** (Chapter 16)[180,181] converts the biosynthetic route to arginine into a cycle for the synthesis of urea. This cyclic pathway is unique to organisms that excrete nitrogenous wastes as urea, but the biosynthetic path is nearly ubiquitous.[182] Human adults excrete approximately 20 g of urea nitrogen per day. If this rate decreases, ammonia accumulates in the blood to toxic levels. Normally, plasma contains 0.03 mM ammonia, and only 2–3 times this level is required to produce toxic symptoms. Therefore, it is not surprising that five different well-documented hereditary enzyme deficiencies affect the urea cycle.[162,183] One of the most common, **argininosuccinic aciduria**, is a deficiency of the breakdown of argininosuccinic acid.[178] Both lethal and nonlethal variants of this disease are known. Human argininosuccinate lyase consists of two subunits. Defects may occur in either subunit but considerable genetic heterogeneity exists and intragenic complementation between the two subunits accounts for many of the nonlethal forms of the disorder.[177a,178a] A common feature of all of the hereditary defects of the urea cycle is an intolerance to high protein intake and mental symptoms. Toxic accumulation of ammonia in blood is often seen also in **alcoholic liver cirrhosis** as a result of a decreased capacity of the liver for synthesis of urea.

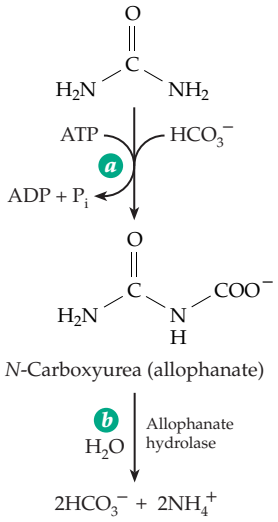
For some urea cycle defects a combination of a low-protein diet together with an arginine supplement prevents the ammonia intoxication while allowing normal growth. In other cases it is necessary to replace the natural dietary protein with a mixture of essential amino acids or with the corresponding 2-oxoacids, which can be converted to amino acids in the body with utilization of endogenous ammonia.[183] A specific treatment for lack of *N*-acetylglutamate synthetase, which forms the carbamoyl phosphate synthase activator *N*-acetylglutamate, is administration of the analog *N*-carbamoylglutamate. This also activates carbamoyl phosphate synthase and is not cleaved by acylases that would prevent the natural activator from being supplied artifically via the blood.[158]

Although the primary function of the urea cycle is usually regarded as the removal of NH_4^+ from the body, it also removes HCO_3^- in equal amounts (Eq. 24-24). This is essential for maintenance of neutral pH,

$$2\,HCO_3^- + 2\,NH_4^+ \longrightarrow H_2NCONH_2 + CO_2 + 3\,H_2O \qquad (24\text{-}24)$$

and Atkinson and Bourke suggested that removal of HCO_3^- is as important a function of the cycle as removal of NH_4^+.[184] However, there are strong arguments against this concept.[162]

Excretion of ammonia. While mammals excrete urea, many invertebrate organisms that live in water as well as some fishes simply excrete NH_3. Other organisms hydrolyze urea to NH_3. Even green plants recycle nitrogen via urea and the Ni^{2+}-dependent urease (Eq. 16-47). Two compounds that can be hydrolyzed by cells to urea and glyoxylate are allantoin and allantoic acid (Eq. 24-15). If cells of *Saccharomyces cerevisiae* are grown on either of these compounds as a sole source of nitrogen, they make a biotin-dependent **urea carboxylase** (Eq. 24-25). This enzyme facilitates the hydrolysis of urea by conversion to the more easily degraded allophanate (Eq. 24-25).

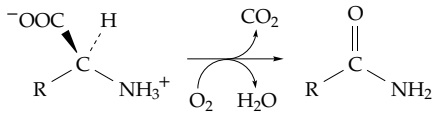


(24-25)

Catabolism of arginine. Arginine can also be converted back to glutamate and 2-oxoglutarate. The initial step is removal of the guanidino group to form ornithine. This occurs in the urea cycle and also in many bacteria[185] by the action of arginase (Fig. 24-5, step *h*). A parallel pathway involving conversion of arginine to N^1-succinylarginine, then on to succinylglutamate, and to free glutamate and succinate is used by some pseudomonads.[186] The alternative **arginine**

dihydrolase pathway, used by some bacteria and a few protozoa such as *Giardia*, is initiated by a different hydrolase that cleaves arginine to citrulline and ammonia.[187] Phosphorolysis of citrulline yields carbamoyl phosphate whose breakdown to CO_2 and ammonia (catalyzed by carbamate kinase, Eq. 24-22) can be utilized for generation of ATP by microorganisms that subsist on arginine.[188]

Degradation of L-arginine by *Streptomyces griseus* is initiated by a hydroxylase that causes decarboxylation and conversion of the amino acid into an amide (Eq. 24-26), a reaction analogous to that catalyzed by the flavin-dependent lysine oxygenase (Eq. 18-41). The

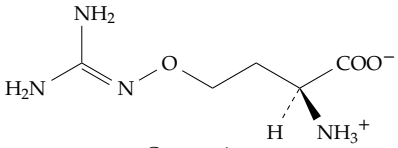


(24-26)

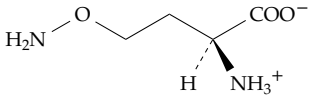
product formed from arginine is γ-guanidinobutyramide, which is further degraded by the hydrolysis of the amide group and cleavage of the guanidino group to form urea and γ-aminobutyrate. *Pseudomonas putida* initiates degradation of arginine by decarboxylation to the corresponding 2-oxoacid and oxidative decarboxylation with a thiamin diphosphate-requiring enzyme to γ-guanidinobutyraldehyde. Dehydrogenation and hydrolysis lead, again, to γ-aminobutyrate.[189]

Specific arginine residues in proteins are methylated on their guanidino groups to give monomethylated and both symmetrically and asymmetrically dimethylated derivatives.[190,191] These methylated arginines also occur free in various mammalian tissues, where they may serve as endogenous regulators of nitric oxide synthases. A Zn^{2+}-containing dimethylarginase hydrolyzes the monomethyl and dimethyl arginines to citrulline and monomethyl or dimethyl amines.[191]

Insecticidal analogs of arginine. The toxic amino acid **L-canavanine** is synthesized by more than 1500 species of legumes including alfalfa and clover.[192–194] It is structurally similar to arginine, the 5-CH_2 group being replaced by O. However, the guanidino group is much less basic than in arginine. Canavanine is a natural insecticide, which in some plants accumulates to a level of 13% of the total dry matter.[193] Plants that store canavanine hydrolyze it to **canaline** and urea, which they use as a nitrogen source. Canaline is a toxic derivative of hydroxylamine and forms oximes with 2-oxoglutarate, other oxoacids, and PLP-containing enzymes. Although canavanine and canaline are effective insecticides, some beetles are adapted to these compounds to the extent that they feed exclusively on canavanine-containing seeds. The tobacco budworm is likewise resistant to these toxins and produces a **canavanine hydrolase** that converts canavanine to L-homoserine, a normal intermediate in



L-Canavanine



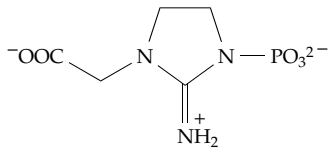
L-Canaline

the threonine, isoleucine, methionine biosynthetic pathway (Fig. 24-13), and **hydroxyguanidine**. The latter undergoes NADH-dependent reduction to guanidine[194] which can be catabolized.

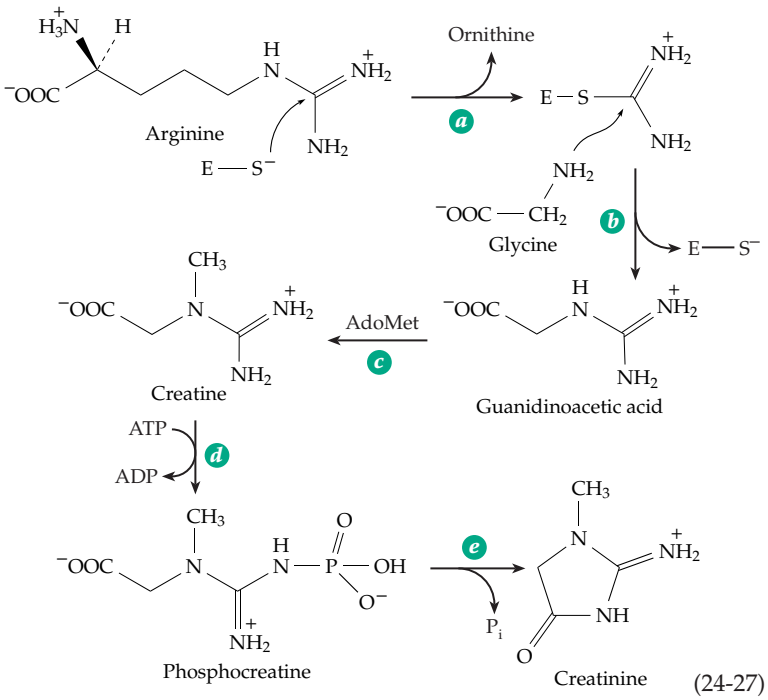
3. Amidino Transfer and Creatine Synthesis

The terminal amidino group of arginine is transferred intact to a number of other substances in simple displacement reactions. An example is the formation of **guanidinoacetic acid** (Eq. 24-27, steps *a* and *b*). The amidino group appears to be transferred first to the SH group of cysteine 407 then to glycine in a double displacement mechanism.[195–197] Transmethylation from *S*-adenosylmethionine (Eq. 24-27, step *c*) converts guanidinoacetic acid to **creatine**, a compound of special importance in muscle. Creatine kinase reversibly transfers the phospho group of ATP to creatine to form the *N*-phosphate (Eq. 24-27, step *d*). **Creatine phosphate**, and in some invertebrates phosphoarginine,[198] serves as an important "energy buffer" for muscular contraction (Chapter 19). Through the reversible action of creatine kinase it is able rapidly to transfer its phospho group back onto ADP as fast as the latter is formed during the hydrolysis of ATP in the contraction process. An end product of creatine phosphate metabolism is the anhydride **creatinine** formed from creatine phosphate as is indicated in Eq. 24-27, step *e* as well as directly from creatine. The urinary creatinine excretion for a given individual is extremely constant from day to day, the amount excreted apparently being directly related to the muscle mass of the person. Another example of the transfer of amidino groups from arginine is found in the synthesis of streptomycin (Box 20-B).

A cyclic analog of creatine, **cyclocreatine**, when fed to animals, accumulates in large amounts in muscle, heart, and brain and is a long-acting phosphagen.[199]



Cyclocreatine phosphate



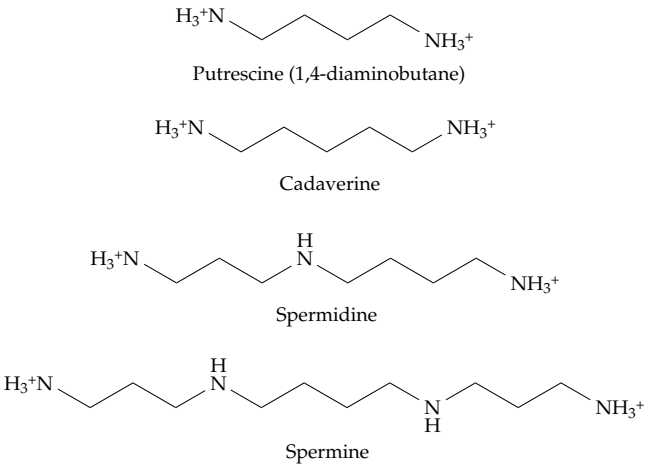
(24-27)

4. The Polyamines

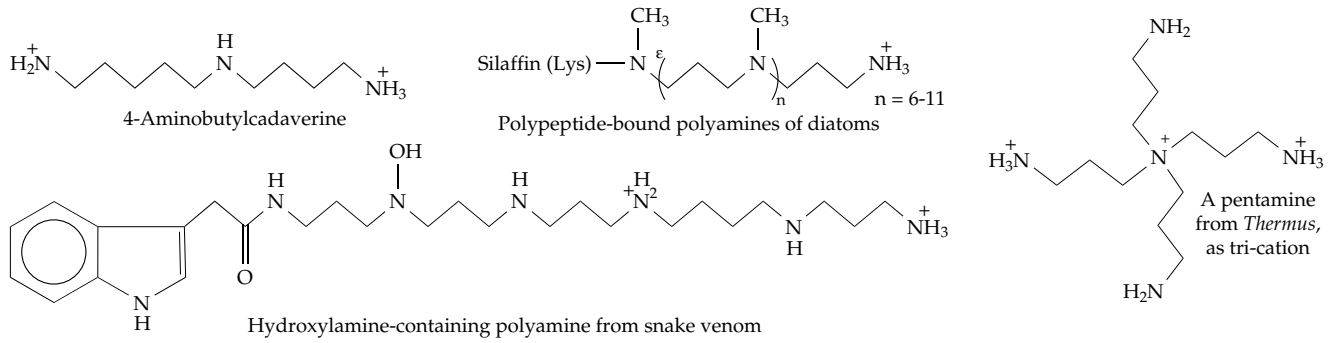
A series of related polyamino compounds, which are derived in part from arginine, are present in all cells in relatively high, often millimolar, concentrations.[200–203] The content of polyamines in cells tends to be stoichiometric with that of RNA, and the polyamines are concentrated in the ribosomes and also in the nucleus. Two moles of polyamine are usually present per mole of any isolated tRNA.[202,204] The first satisfactory crystals of tRNA for X-ray structure determination were obtained in the presence of spermine. Spermidine is associated with RNA in the turnip yellow mosaic virus.[205] The T-even bacteriophage and most bacteria contain polyamines in association with DNA. Polyamines are able to interact with double helical nucleic acids by bridging between strands, the positively charged amino groups interacting with the phosphates of the nucleic acid backbones. Tsuboi suggested that the tetramethylene portion of the polyamine lies in the minor groove bridging three base pairs, and the trimethylene portions (one in spermidine, two in spermine) bridge adjacent phosphate groups in one strand.[206] Polyamines may also stabilize supercoiled or folded DNA.

The structures of polyamines are shown here as di- and tri-cations, but it should be realized that there are multiple positions for protonation and therefore various tautomers. Also, polyamines show extreme anticooperativity in proton binding, i.e., successive pK_a values range from very low to very high for the last proton to leave. Polyamines are thought to have several functions. They can substitute to some extent for cellular K^+ and Mg^{2+}, and they may play essential controlling roles in nucleic acid and protein synthesis. A specific role of spermidine in cell division seems likely.[207,207a] An absolute requirement for polyamines has been demonstrated for some bacteria such as *Hemophilus parainfluenzae*[208] and for mutants of *Aspergillus* and *Neurospora*. Polyamines are also essential for mammalian cells. Polyamines activate some enzymes including the serine/threonine protein kinase CK2.[209]

Mutants of *E. coli* have been constructed in which enzymes of all known biosynthetic pathways for polyamines are blocked by deletion of the genes for arginine decarboxylase (*SpeA*), agmatine ureahydrolase (*SpeB*), ornithine decarboxylase (*SpeC*), and adenosylmethionine decarboxylase (*SpeD*).[210] Even though polyamines cannot be detected in these cells they grow at one-third the normal rate. However, yeast cells require both putrescine and spermidine or spermine for growth.[211,211a] Another effect is seen in strains of yeast carrying the **"killer plasmid,"** a 1500-kDa double-stranded RNA plasmid that encodes a toxic protein, which is secreted and kills other susceptible strains of yeast. Yeast cells carrying the killer plasmid lose it when made deficient in polyamines.[212] The bacterial outer membrane porins OmF and OmC (Fig. 8-20) bind polyamines, especially spermine, and inhibit passage of ions. Polyamines may also modulate ion channels of heart, muscle, and neurons.[213] Both prokaryotic and eukaryotic cells have transporters that allow uptake of polyamines from their surroundings.[214,215]



Major naturally occurring polyamines (as dications)



Biosynthesis. The 4-carbon putrescine arises most directly by decarboxylation of ornithine (Fig. 24-12, step *b*),[216] but it can also be formed by decarboxylation of arginine to agmatine followed by hydrolysis of the latter (Fig. 24-12, steps *c,d*). An alternative pathway utilizes an "agmatine cycle" in which agmatine is first hydrolyzed to ammonium ions and *N*-carbamoylputrescine. The latter transfers its ureido group to ornithine to form citrulline and releases free putrescine (Fig. 24-12, steps *f,g*). The citrulline is reconverted to arginine. This pathway appears to be important in plants.[217] Putrescine is normally present in all cells, and all cells are able to convert it on to spermidine. This is accomplished by decarboxylation of *S*-adenosylmethionine (Fig. 24-12, step *a*) and transfer of the propylamine group from the resulting decarboxylation product onto an amino group of putrescine (Fig. 24-12, step *h*).[218–221]:

The more complex spermine is found only in eukaryotes. It is formed by transfer of a second propylamine group onto spermidine (Fig. 24-12, step *i*). A historical note is that Anthony von Leeuwenhoek with one of his first microscopes observed crystals of the phosphate salt of spermine in human semen in 1678. The 5-carbon diamine cadaverine arises from decarboxylation of lysine (Fig. 24-12, step *d*). The extremely thermophilic bacterium *Thermus thermophilis* produces several additional polyamines including a pentamine (see above), a quaternary nitrogen compound.[222] Many other polyamines are known.[223–224c] Among these are a 4-aminobutylcadaverine isolated from root nodules of the adzuki bean[223] and very long partially aromatic hydroxylamine derivatives from venom of common funnel-web spiders (structures at top of page).[225] Cationic polypeptides called **silaffins**, with masses of ~3 kDa, apparently initiate the growth of the silica cell walls of diatoms (Box 4-B). These peptides contain polyamines consisting of 6 to 11 repeated *N*-methylpropylamine units covalently attached to lysine residues[224a,224b] and also many phosphoserines.[224d]

The synthesis of polyamines is tightly regulated. The PLP-dependent ornithine decarboxylase is present in very low concentrations[226] and apparently has the shortest half-life (~10 min) of any mammalian

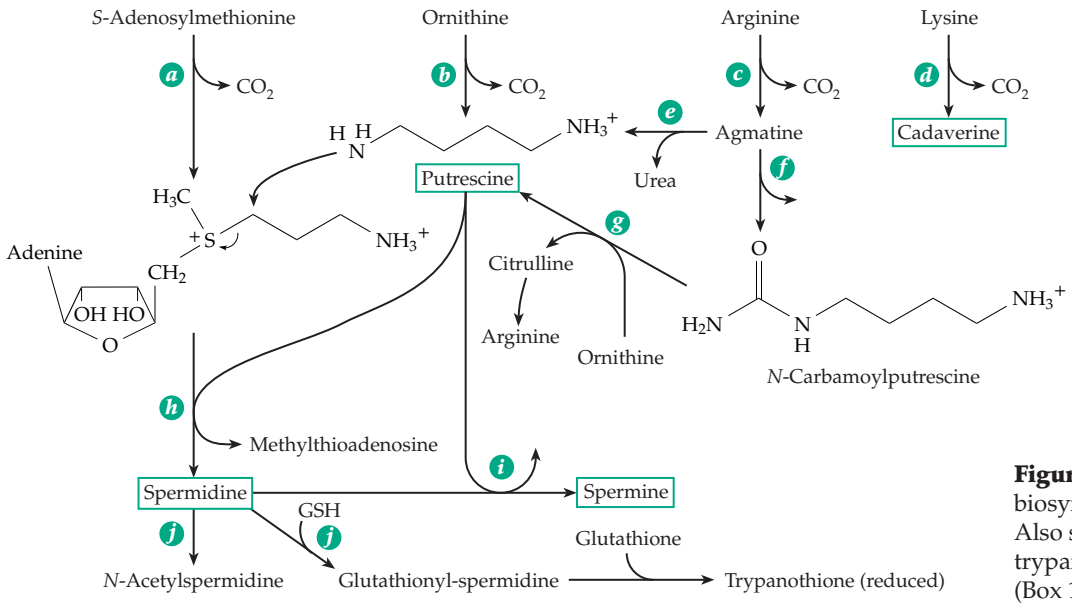
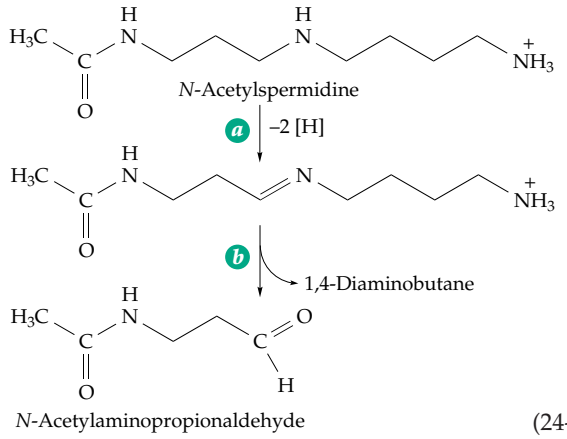


Figure 24-12 Pathways of biosynthesis of polyamines. Also shown is the formation by trypanosomes of trypanothione (Box 11-B).

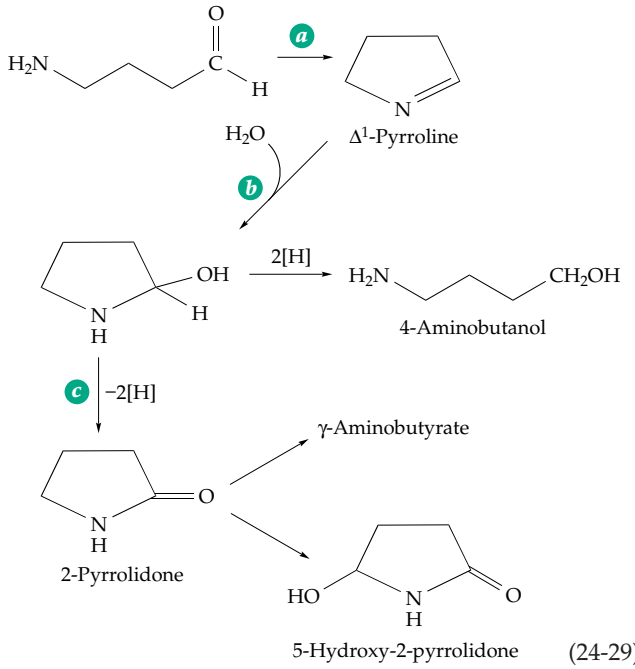
enzyme.[227,228] Its concentration increases rapidly in most species with the onset of rapid growth, transformation to a neoplastic state, or initiation of cell differentiation. The rate of synthesis of the enzyme appears to be regulated by feedback repression by spermidine and by inactivation in response to a buildup of putrescine.[229] One mechanism of inactivation is the synthesis of a 26-kDa specific inhibitor called an **antizyme** in response to the presence of putrescine, spermidine, or spermine. The antizyme is ubiquitous in both prokaryotes and eukaryotes and keeps most of the ornithine decarboxylase bound and inactive and also promotes its degradation by 26S proteosomes.[230–230b] A polyamine-dependent protein kinase in *Physarum* phosphorylates the decarboxylase thereby inhibiting its activity.[231]

Breakdown. The catabolism of polyamines is less well understood than is their biosynthesis. Oxidative cleavages of spermine to spermidine and of the latter to 1,4-diaminobutane appear to occur in the animal body, and a substantial amount of this diamine is excreted in the urine.[201] Spermidine is acetylated on N^1 by acetyl-CoA and a spermidine *N*-acetyltransferase.[232,233] The resulting N^1-acetylspermidine is more readily cleaved by hepatic polyamine oxidase[233a] than is free spermine; again 1,4-diaminobutane is reformed together with an *N*-acetylaminopropionaldehyde (Eq. 24-28). This and other aldehydes formed from polyamines are very toxic but they may play essential roles in regulation of metabolism.[201] Transamination of 1,4-diaminobutane yields γ-aminobutyraldehyde which cyclizes (Eq. 24-29). Diamine oxidases of animal tissues oxidize 1,4-diaminobutane with formation of the same products.[234] Further metabolism of Δ^1-pyrroline yields γ-aminobutyrate, which can undergo transamination and oxidative metabolism as shown in Fig. 17-5. Other products[235] are also indicated in Eq. 24-29. Metabolism of other polyamines also begins by oxidation at the primary amino termini.[236] Formation of β-alanine, needed for synthesis of pantothenic acid, can also occur by oxidation of spermine.[236a]

When *E. coli* cells enter the stationary phase of the growth curve (Box 9-B), most of the spermidine is converted to **glutathionylspermidine** (γ-glutamylcysteinylglycylspermidine) in which glutathione and spermidine are joined by an amide linkage.[237–239] Trypanosomes join a second glutathione at the other end of the spermidine to form reduced **trypta-thione**,[238] a compound also considered in Box 11-B. N^1-γ-Glutamylspermidine and related compounds have been found in proteolytic digests of certain proteins, suggesting that polyamines may be physiological substrates for transglutaminases.[240] Portions of polyamines are incorporated into a variety of products including **nicotine** (Fig. 30-22)[241] and the unusual



(24-28)



(24-29)

amino acid **hypusine** (see p. 1386).[242]

Ornithine decarboxylase is specifically inhibited by the enzyme-activated inhibitor α-difluoromethylornithine, which can cure human infection with *Trypanosoma brucei* (African sleeping sickness) by interfering with polyamine synthesis.[243–244a] In combination with inhibitors of spermidine synthase or *S*-adenosylmethionine decarboxylase,[245] it can reduce polyamine levels and growth rates of cells. Another powerful inhibitor that acts on both ornithine and adenosylmethionine decarboxylases is the hydroxylamine derivative 1-aminooxy-3-aminopropane.[246]

$$H_2N{-}O{-}CH_2CH_2CH_2{-}NH_2$$
1-Aminooxy-3-aminopropane

Like difluoromethylornithine the compound at low concentrations is not toxic to cells but inhibits growth. It is hoped that adequate inhibition of growth of

normal cells may allow more aggressive chemotherapeutic treatment of cancer.

D. Compounds Derived from Aspartate

The 4-carbon aspartate molecule is the starting point for synthesis of **pyrimidines** and of the amino acids **lysine**, **methionine**, **threonine**, **isoleucine**, and **asparagine**.[247,248] The pathways are summarized in Fig. 24-13. There are several branch points, and aspartate can be converted directly to asparagine, to carbamoylaspartate (the precursor of pyrimidines), or to β-aspartyl phosphate and aspartate semialdehyde. The latter can be converted in one pathway to lysine and in another to homoserine. Homoserine can yield either homocysteine and methionine or threonine. Although threonine is one of the end products and a constituent of proteins, it can also be converted further to 2-oxobutyrate, a precursor of isoleucine.

Most of the chemistry has been considered already. The reduction of aspartate via β-aspartyl phosphate[249,249a] and aspartate β-semialdehyde[250] is a standard one. Conversion to methionine can occur in two ways. In *E. coli* homoserine is succinylated with succinyl-CoA. The γ-succinyl group is then replaced by the cysteine molecule in a PLP-dependent γ-replacement reaction (Fig. 24-13). The product **cystathionine** (Eq. 14-33) undergoes elimination to form homocysteine. A similar pathway via *O*-phosphohomoserine occurs in chloroplasts of green plants.[251] A more direct γ replacement of the hydroxyl of homocysteine or *O*-phosphohomoserine by a sulfide ion has also been reported for both *Neurospora* and green plants.[252] Methylation of homocysteine to methionine (Fig. 24-13) has been considered previously, as has the conversion of homoserine to threonine by homoserine kinase[253] and the PLP-dependent **threonine synthase** (p. 746, Fig. 14-7).[254–255a] A standard PLP-requiring β elimination converts threonine to **2-oxobutyrate**, a precursor to isoleucine (Fig. 24-13).[256]

Formation of **asparagine** has been discussed in section B. Asparagine synthase of *E. coli*[98–99] cleaves ATP to AMP and PP_i rather than to ADP via an aspartyladenylate intermediate. In higher animals glutamine serves as the ammonia donor for synthesis of asparagine, but NH_4^+ can also function.[257] **L-Asparaginase**, a bacterial hydrolase, is an experimental antileukemic drug. It acts to deprive fast-growing tumor cells of the exogenous asparagine needed for rapid growth.[136,257a] Tissues with a low asparagine synthase activity are also damaged, limiting the clinical usefulness.

Aspartate can be decarboxylated either to α-alanine by a PLP-dependent enzyme[258] or to β-alanine by a pyruvoyl group-containing enzyme (Chapter 14). Beta-alanine is not only a component of the vitamin pantothenic acid but is found in the dipeptides carnosine (β-alanylhistidine) and anserine (β-alanyl-N^{δ}-methylhistidine) present in vertebrate muscles.[259] It is a crosslinking agent in insect cuticle.

Aspartate can be deaminated to fumarate by bacterial L-**aspartate oxidase**.[259a] This flavoprotein is structurally and mechanistically related to succinate dehydrogenase and can function as a soluble fumarate reductase (p. 1027). However, its main function appears to be to permit the intermediate iminoaspartate to react with dihydroxyacetone-*P* to form quinolinate, which can be converted to **NAD** (see Fig. 25-11).[259b]

1. Control of Biosynthetic Reactions of Aspartate

In *E. coli* there are three **aspartokinases** that catalyze the conversion of aspartate to β-aspartyl phosphate. All three catalyze the same reaction, but they have very different regulatory properties, as is indicated in Fig. 24-13. Each enzyme is responsive to a different set of end products.[247,260] The same is true for the two **aspartate semialdehyde reductases** which catalyze the third step. Both repression of transcription and feedback inhibition of the enzymes are involved. Two of the aspartokinases of *E. coli* are parts of bifunctional enzymes, which also contain the homoserine dehydrogenases that are needed to reduce aspartate semialdehyde in the third step. These aspartokinase–homoserine dehydrogenases I and II (Fig. 24-13) are encoded by *E. coli* genes *thrA* and *metL*, respectively, and have homologous sequences.[247,261–262a] The N-terminal portions are also homologous to the lysine-sensitive aspartokinase III which is encoded by the *lysC* gene.[263] In *Bacillus subtilis* the lysine-sensitive enzyme is known as aspartokinase II. It has an $\alpha_2\beta_2$ oligomeric structure and both α and β chains are encoded within a single gene.[264] There is no associated homoserine dehydrogenase. Both genetic organization and processing of the synthesized protein are thus different in these two bacteria.

2. Lysine Diaminopimelate, Dipicolinic Acid, and Carnitine

Lysine cannot be made at all by animals but is nutritionally essential. There are two distinct pathways for its formation in other organisms. The **α-aminoadipate pathway** (shown in Fig. 24-9) occurs in a few lower fungi, the higher fungi, and euglenids. The 5-carbon 2-oxoglutarate is the starting compound. Bacteria, other lower fungi, and green plants all use the **diaminopimelate** pathway (Fig. 24-14) which originates with the 4-carbon aspartate.

The α-aminoadipate pathway (Fig. 24-9) parallels

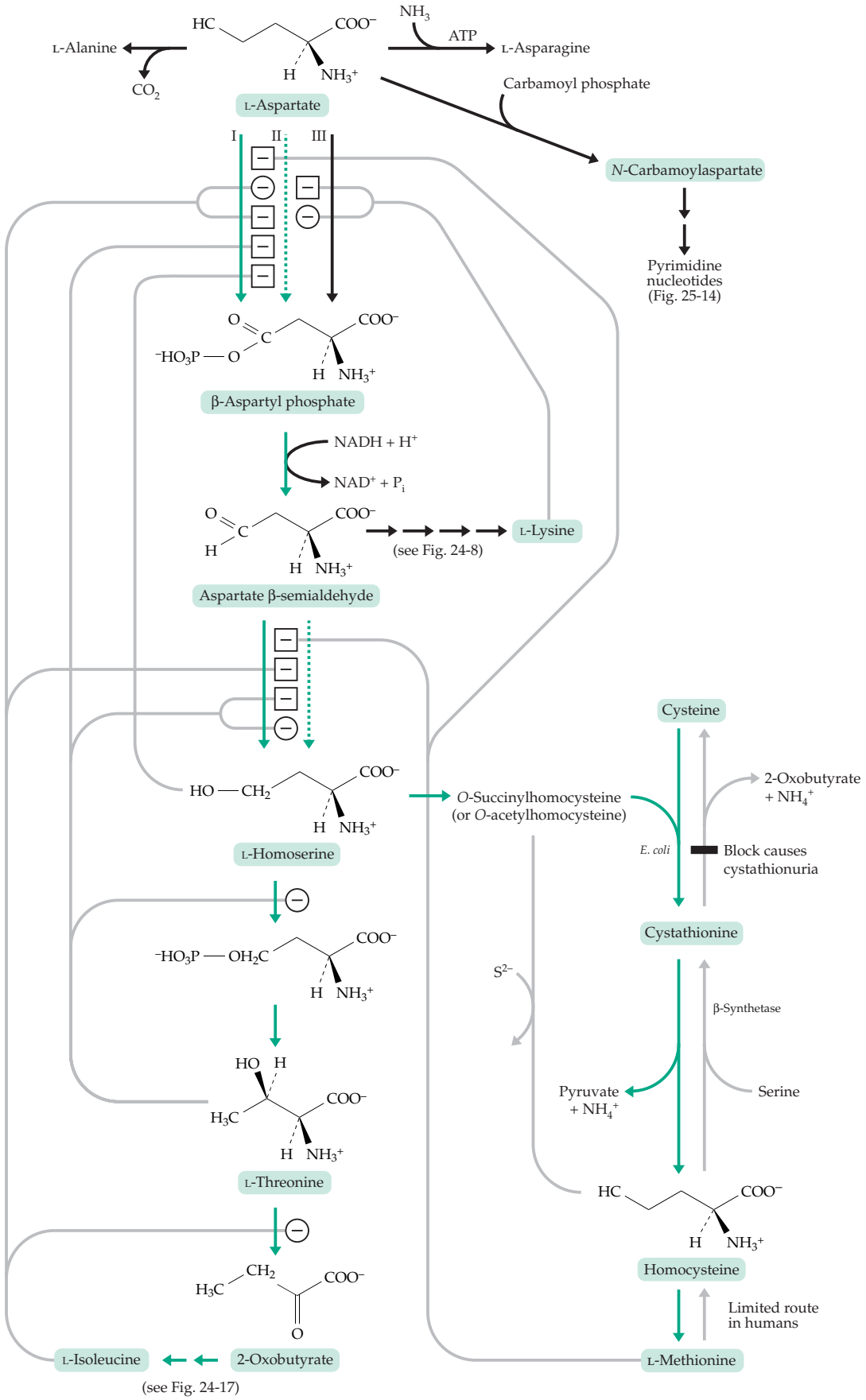
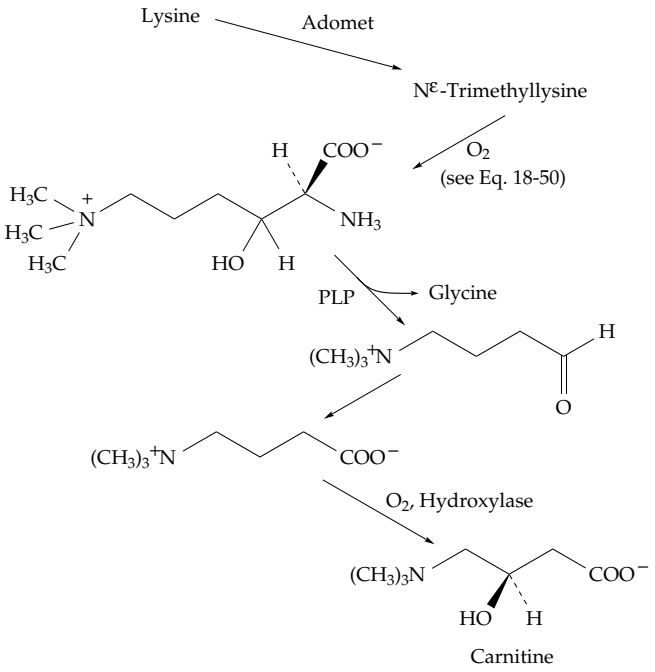


Figure 24-13 Some biosynthetic reactions of aspartate: ⊖, feedback inhibition and ⊟, feedback repression.

that of ornithine biosynthesis, 2-oxoglutarate undergoing chain elongation (Fig. 17-18) to 2-oxoadipate followed by transamination to α-aminoadipate. This is followed by ATP-dependent reduction to the aldehyde.[264a] The final step of transamination is not accomplished in the usual way (with a PLP-dependent enzyme), but through formation of a Schiff base with glutamate and reduction to **saccharopine**.[265] Oxidation now produces the Schiff base of lysine with 2-oxoglutarate.

In the diaminopimelate pathway of lysine synthesis (Fig. 24-14) aspartate is converted to aspartate semialdehyde, and a two-carbon unit is added via aldol condensation with pyruvate.[266–269] Decarboxylation at the end of the sequence yields lysine. A series of cyclic intermediates exist, but it is noteworthy that the initial product of the aldol condensation (bracketed in Fig. 24-14) is converted to diaminopimelic acid by a simple sequence involving α,β elimination of the hydroxy group, reduction with NADPH, and transamination.[267,268,270,271] The process is complicated by the natural tendency for ring closure. All gram-negative and many gram-positive bacteria use the succinylase pathway shown in Fig. 24-14. Succinylation serves to shift the equilibrium back in favor of open-chain compounds.[272–274] Some species of *Bacillus* use acetylation in the same way, while a few bacteria manage to use a dehydrogenase to reductively aminate tetrahydropimelate to diaminopimelate.[275,276] The diaminopimelate pathway is of special significance to prokaryotic organisms for the reason that **dipicolinic acid** is formed as an important side product and because of formation of **diaminopimelic acids**. The cyclic



(24-30)

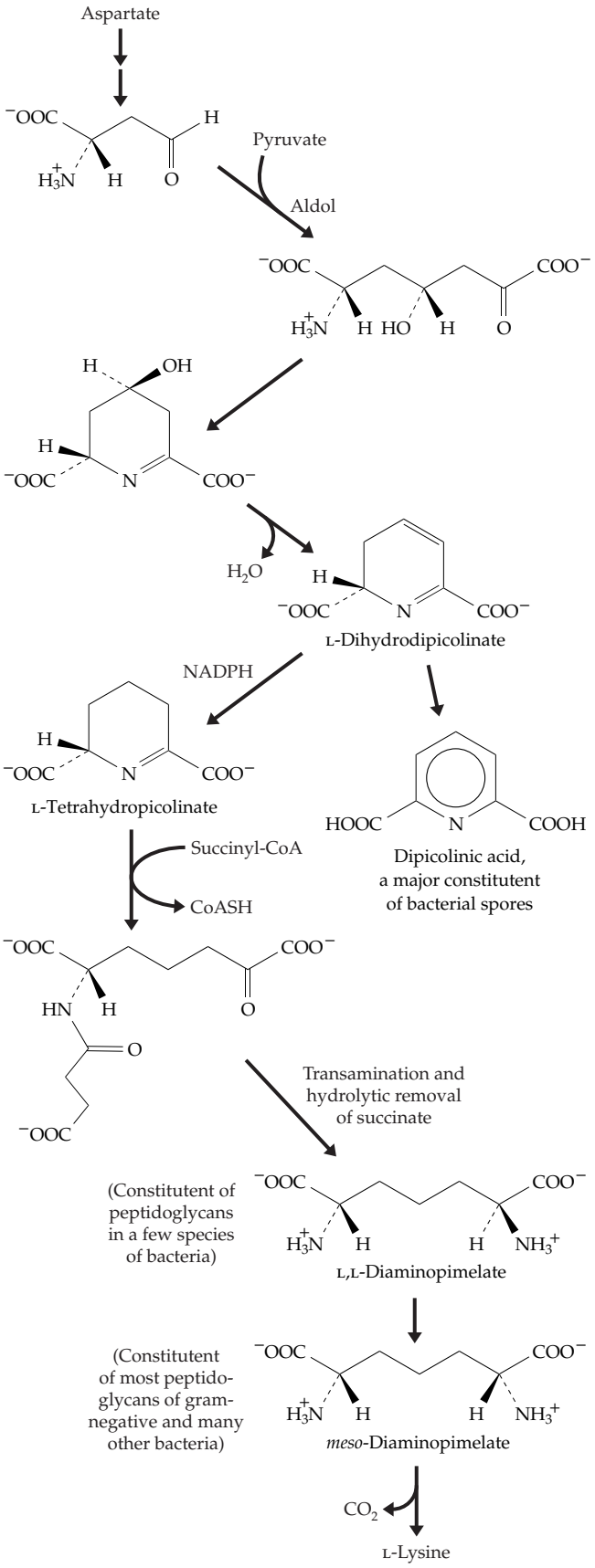
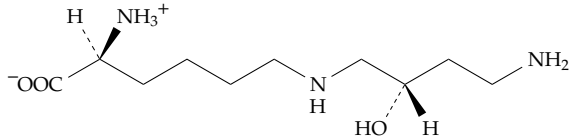


Figure 24-14 The biosynthesis of lysine by the diaminopimelate pathway.

dipicolinic acid is a major constituent of bacterial spores[277] but is rarely found elsewhere in nature. Both L, L- and meso-diaminopimelic acids[276,276a] are constituents of peptidoglycans of bacterial cell walls (Fig. 8-29).

Lysine is not only a constituent of proteins. It can also be trimethylated and converted to **carnitine** (p. 944). In mammals some specific lysyl side chains of proteins undergo *N*-trimethylation and proteolytic degradation with release of free trimethyllysine (Eq. 24-30).[278,279] The free trimethyllysine then undergoes hydroxylation by a 2-oxoglutarate–Fe^{2+}–ascorbate-dependent hydroxylase (Eq. 18-51) to form β-hydroxytrimethyllysine, which is cleaved by a PLP-dependent enzyme (Chapter 14). The resulting aldehyde is oxidized to the carboxylic acid and is converted by a second 2-oxoglutarate–Fe^{2+}–ascorbate-dependent hydroxylase to carnitine (Eq. 24-30; see also Eq. 18-50).

Hypusine (N^{ε}-(4-amino-2-hydroxybutyl)lysine)[242] occurs in mammalian initiation factor 4D, which is utilized in protein synthesis (Chapter 29) and is formed by transfer of the 4-carbon butylamine group from spermidine to a lysine side chain followed by hydroxylation.[280–282a] The lupine alkaloid lupinine[283] is formed from two C_5 units of cadaverine which arises by decarboxylation of lysine. Silaffins (pp. 178, 1381) also contain modified lysines.



Hypusine: N^{ε}-(4-amino-2-hydroxybutyl) lysine

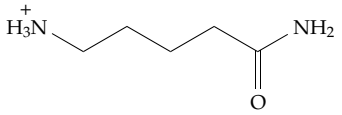
3. The Catabolism of Lysine

An unusual feature of lysine metabolism is that the α-amino group does not equilibrate with the "nitrogen pool." Catabolism is initiated by deamination and proceeds by β oxidation.[284] At least six variations of the β-oxidation process have been proposed. The evolutionary differences concern the manner in which the two amino groups are moved from the carbon skeleton. In the seemingly simplest pathway (A in Fig. 24-15), which is used by *Flavobacterium fuscum*,[285] the ε-amino group is removed in a direct (but atypical) transamination. The resulting α-aminoadipate semialdehyde is oxidized to α-aminoadipate, which *is degraded in a sequence characteristic for the catabolism of amino acids.* Transamination is followed by oxidative decarboxylation of the resulting 2-oxoacid and β oxidation of the coenzyme A derivative. A decarboxylation step by which the terminal carboxyl group is removed is interposed in the β-oxidation sequence for lysine degradation.

Perhaps the initial transamination in pathway A is chemically difficult, for most organisms use more complex sequences to form 2-oxoadipate. In pathway B (which takes place in liver mitochondria and is believed to be the predominant pathway in mammals),[286,287] the ε-amino group is reductively coupled with 2-oxoglutarate to form saccharopine. The latter is in turn oxidized on the opposite side of the bridge nitrogen to form glutamic acid and α-aminoadipate semialdehyde. The overall process is the same as direct transamination and just the opposite of that occurring in the aminoadipate pathway of biosynthesis (Fig. 24-9). Absence of one or both of these dehydrogenases causes familial hyperlysinemia.[287,288]

Pathway C has been established for *Pseudomonus putida*[289] and is also followed to some extent in both plants and animals. In most animal tissues it may be used principally for degradation of D-lysine.[290] However, it is the major L-lysine oxidation pathway in brain.[291] In a fungal parasitic species of *Rhizoctonia* L-lysine is converted to saccharopine via pathway B; then using an $NADP^+$-dependent saccharopine oxidase the sequence is shunted to pathway C.[292] L-Pipicolic acid formed in this way also gives rise to various alkaloids including the α-mannosidase inhibitor swainsonine (Fig. 20-7).[290] Pathway C, like pathway B, makes use of transamination via a reduction-oxidation sequence. It is strictly internal, the oxidizing carbonyl group being formed by transamination of the α-amino group of lysine. Pathway D, apparently used by yeasts,[293] avoids cyclic intermediates by acetylation of the ε-amino group prior to transamination. The 2-oxo group is then effectively blocked by reduction to an alcohol, the blocking group is removed from the ε-amino group, and that end of the molecule is oxidized in a straightforward way to a carboxyl group. Now the hydroxyl introduced at position 2 is presumably oxidized back to the ketone, which again can be converted to give 2-oxoadipate.

Some bacteria, e.g., *Pseudomonas putida*,[294] degrade L-lysine with a flavin-dependent oxygenase (Eq. 18-41) to δ-aminovaleramide:



The product is hydrolyzed and oxidized to **glutaryl-CoA**, rejoining the pathways shown in Fig. 24-15. A remarkable and very different approach to lysine breakdown has been developed by clostridia which obtain energy from the fermentation of Eq. 24-31:

$$\text{L-Lysine} + 2\ H_2O \longrightarrow \text{butyrate}^- + \text{acetate}^- + 2\ NH_4^+ \qquad (24\text{-}31)$$

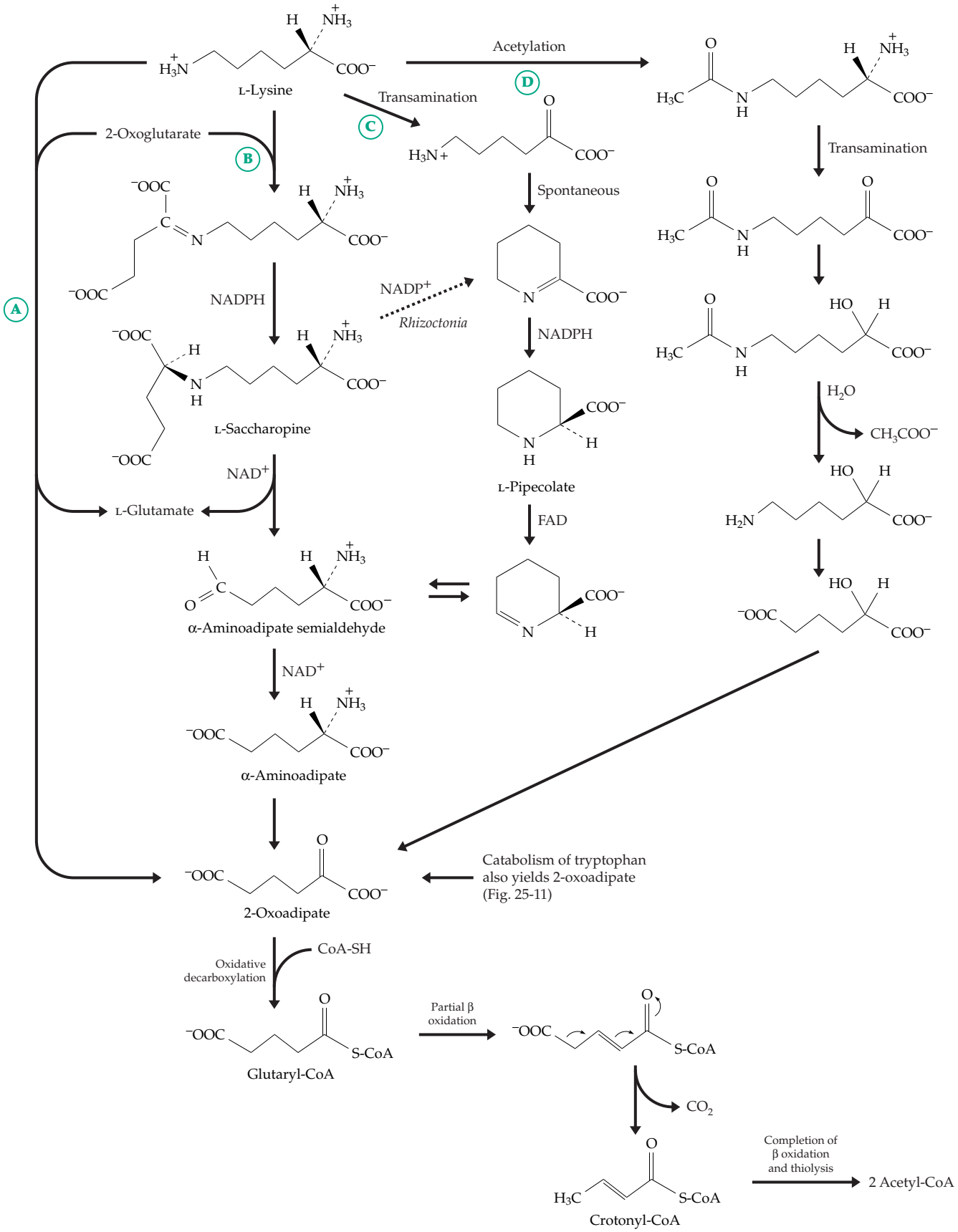
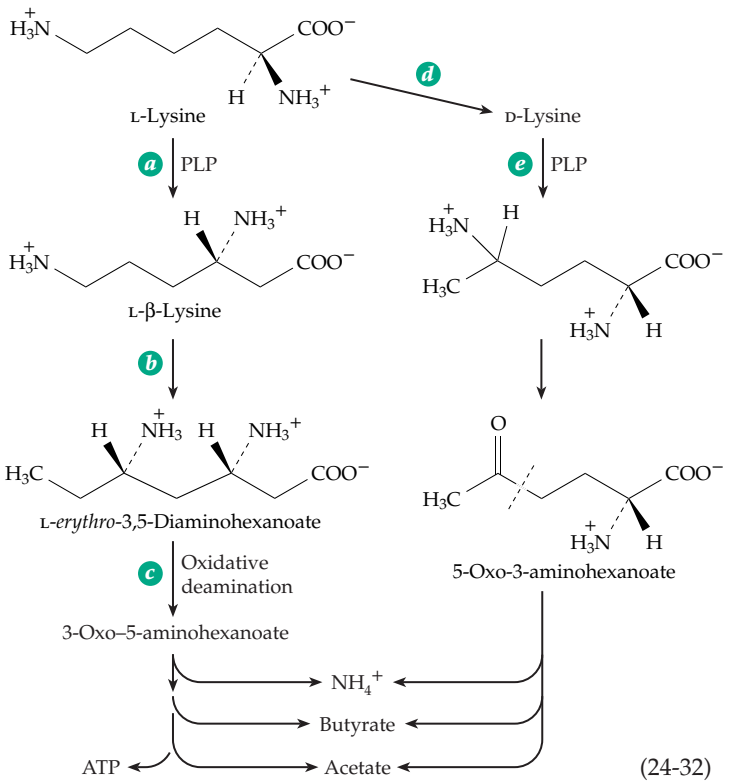


Figure 24-15 Catabolism of lysine.

The reaction is coupled to formation of one molecule of ATP from ADP and P_i. Two pathways have been worked out. In the first lysine is acted upon by a PLP-dependent **L-lysine 2,3-aminomutase** (Eq. 24-32, step *a*) to convert it to β-lysine (3,6-diaminohexanoate). The latter is further isomerized (Eq. 24-32, step *b*) by the vitamin B_{12} and PLP-dependent β-lysine mutase. Oxidative deamination to a 3-oxo compound (Eq. 24-32, step *c*) permits chain cleavage. The reader can easily propose the remaining reactions of chain cleavage, ATP synthesis, elimination of ammonia, and balancing of the redox steps. An alternative pathway begins with a racemase (Eq. 24-32, step *d*) and isomerization of the resulting D-lysine by another B_{12} and PLP-dependent enzyme (Eq. 24-32, step *e*).[294a] Oxidative deamination presumably occurs, but the mechanism for chain cleavage is not so obvious. It does occur between C-4 and C-5 as indicated by the dashed line in Eq. 24-32.

Another variation is used by *Pseudomonas* β4 (Eq. 24-31). Beta-lysine is acetylated on N-6, then undergoes transamination to a 2-oxo acid and removal of the first two carbons as acetyl-CoA. The resulting 4-aminobutyrate is then converted to succinate via succinate semialdehyde.[295]

Why are there so many pathways of lysine breakdown? The answer is probably related to the ease of spontaneous formation of cyclic intermediates as occurs in the pipecolate pathway (pathway C, Fig. 24-15). These intermediates may be too stable for efficient metabolism so the indirect pathways evolved. In the fermentation reactions additional constraints are imposed on the pathways by the need for balanced redox processes and a net Gibbs energy decrease.



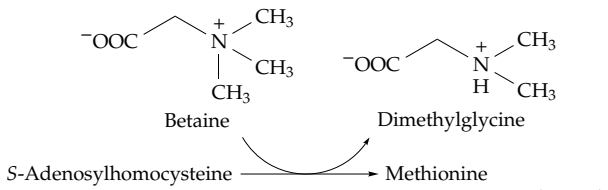
(24-32)

4. Metabolism of Homocysteine and Methionine

Autotrophic organisms synthesize methionine from aspartate as shown in the lower right side of Fig. 24-13. This involves transfer of a sulfur atom from cysteine into homocysteine, using the carbon skeleton of homoserine, the intermediate **cystathionine**, and two PLP-dependent enzymes, **cystathionine γ-synthase**[296,296a] and **cystathionine β-lyase**.[297] This **transsulfuration** sequence (Fig. 24-13, Eq. 14-33) is essentially irreversible because of the cleavage to pyruvate and NH_4^+ by the β-lyase. Nevertheless, this transsulfuration pathway operates in reverse in the animal body, which uses two different PLP enzymes, **cystathionine β-synthase** (which also contains a bound heme)[298–299c] and **cystathionine γ-lyase**[300] (Figs. 24-13, 24-16, steps *h* and *i*), in a pathway that metabolizes excess methionine.

For human beings methionine is nutritionally essential and comes entirely from the diet. However, the oxoacid analog of methionine can be used as a nutritional supplement. Dietary homocysteine can also be converted into methionine to a limited extent. Methionine is incorporated into proteins as such and as ***N*-formylmethionine** at the N-terminal ends of bacterial proteins (steps *a* and *b*, Fig. 24-16). In addition to its function in proteins methionine plays a major role in biological methylation reactions in all organisms. It is converted into ***S*-adenosylmethionine** (AdoMet or SAM; Fig. 24-16, step *e*; see also Eq. 17-37),[301–302b] which is the most widely used methyl group donor for numerous biological methylation reactions (Eq. 12-3). *S*-Adenosylmethionine is also the precursor of the special "wobble base" **queuine** (Fig. 5-33).[312]

The product of transmethylation, ***S*-adenosylhomocysteine**, is converted (step *g*) into homocysteine in an unusual NAD-dependent hydrolytic reaction (Eq. 15-14) by which adenosine is removed (step *g*).[302c] Homocysteine can be reconverted to methionine, as indicated by the dashed line in Fig. 24-16. This can be accomplished by the vitamin B_{12}- and tetrahydrofolate-dependent **methionine synthase**, (Eq. 16-43), which transfers a methyl group from methyl-tetrahydrofolate[303–303b]; by transfer of a methyl group from **betaine**, a trimethylated glycine (Eq. 24-33)[304], or by remethylation with AdoMet (Fig. 24-16).[304a]

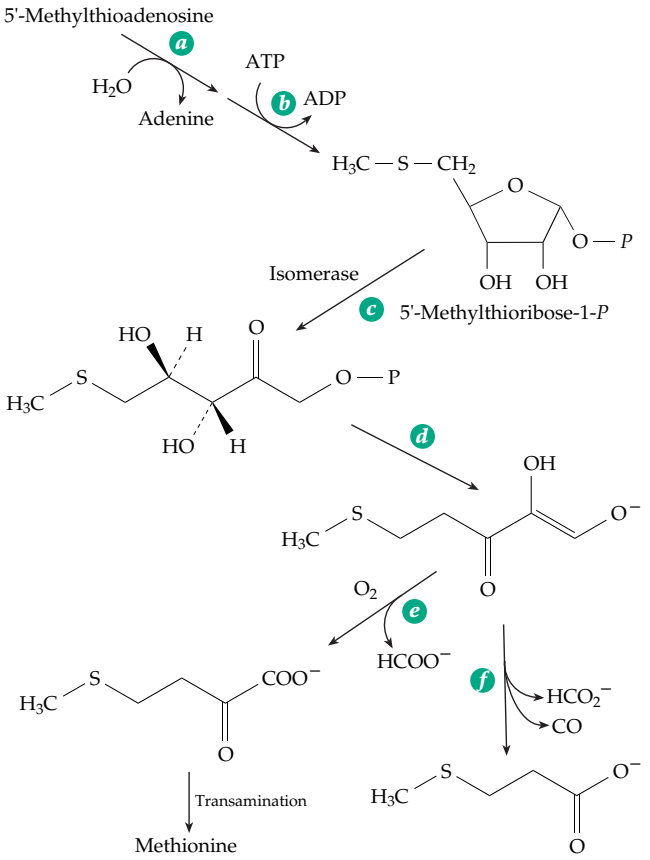


(24-33)

When present in excess methionine is toxic and must be removed. Transamination to the corresponding 2-oxoacid (Fig. 24-16, step *c*) occurs in both animals and plants. Oxidative decarboxylation of this oxoacid initiates a major catabolic pathway,[305] which probably involves β oxidation of the resulting acyl-CoA. In bacteria another catabolic reaction of methionine is γ-elimination of methanethiol and deamination to 2-oxobutyrate (reaction *d*, Fig. 24-16; Fig. 14-7).[306] Conversion to homocysteine, via the transmethylation pathway, is also a major catabolic route which is especially important because of the toxicity of excess homocysteine. A hereditary deficiency of cystathionine β-synthase is associated with greatly elevated homocysteine concentrations in blood and urine and often disastrous early cardiovascular disease.[299,307–309b] About 5–7% of the general population has an increased level of homocysteine and is also at increased risk of artery disease. An adequate intake of vitamin B_6 and especially of folic acid, which is needed for recycling of homocysteine to methionine, is helpful. However, if methionine is in excess it must be removed via the previously discussed transsulfuration pathway (Fig. 24-16, steps *h* and *i*).[310] The products are cysteine and 2-oxobutyrate. The latter can be oxidatively decarboxylated to propionyl-CoA and further metabolized, or it can be converted into leucine (Fig. 24-17) and cysteine may be converted to glutathione.[299a]

Methionine in plants can be converted to the sulfonium compound *S*-methyl-L-methionine, also called vitamin U. It has strong osmoprotectant activity and accumulates in many marine algae and some flowering plants.[311] Other organisms, including mammals, can use *S*-methylmethionine to methylate homocysteine, converting both reactants back to methionine[311a] enabling animals to meet some of their methionine need from this source.

A salvage pathway. Another product of *S*-adenosylmethionine is **5'-methylthioadenosine**, which can be formed by an internal displacement on the γ-methylene group by the carboxylate group (step *l*, Fig. 24-16). Methylthioadenosine also arises during formation of the compounds spermidine (Fig. 24-12) and ACC (Fig. 24-16). Mammalian tissues convert methylthioadenosine back to methionine by the sequence shown in Eq. 24-34. It undergoes phosphorolysis to 5'-methylthioribose whose ring is opened and



(24-34)

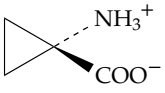
converted to the 2-oxoacid analog of methionine.[313,314]

Step *c* of Eq. 24-34 may occur by ring opening to an enol phosphate which ketonizes to the observed product, but step *e* is a more complex multistep oxidative process.[314a,b] The last step is transamination to methionine with a glutamine-specific aminotransferase. Another enzyme from *Klebsiella* converts the same intermediate anion to methylthiopropionate, formate, and CO (Eq. 24-34, step *f*).[315]

The plant hormone ethylene. A major reaction of *S*-adenosylmethionine in plants is the formation of **ethylene**.[316,317] Ethylene has been recognized since 1858 as causing a thickening of stems of plants and a depression in the rate of elongation. In 1917, it was established that ethylene is formed in fruit and that addition of this gaseous compound hastened ripening. Ethylene is now an established plant hormone having a variety of effects including retardation of mitosis, inhibition of photosynthesis, and stimulation of respiration and of the enzyme phenyalanine ammonia-lyase (Eq. 14-45). These effects are indirectly a result of the action of ethylene on transcription of certain genes. In *Arabidopsis*, with which genetic studies are being made, ethylene binds to the N-terminal part of at least two receptor proteins, which have intracellular histidine kinase domains in the C-terminal parts.[318,319]

DNA-binding proteins specific for **ethylene-responsive elements** (EREs), having the conserved sequence AGCCGCC, are presumably phosphorylated by this kinase[320] and affect transcription of many genes. A protein homologous to one ethylene receptor of *Arabidopsis* has been identified in the tomato. A proline to leucine mutation at position 36, near the N terminus, destroys the sensitivity to ethylene and prevents ripening of this tomato. Another component in the ethylene signaling pathway in *Arabidopsis* is a protein serine / threonine kinase that resembles the mammalian raf kinase involved in the signaling cascade shown in Fig. 11-13.[318]

The formation of ethylene is often induced by the hormone **auxin** (Chapter 30), which stimulates activity of the synthase that forms **1-aminocyclopropane-1-carboxylate** (**ACC**) from *S*-adenosyl methionine (Eq. 14-27, step *j*: Fig. 24-16).[320a,b] Although ACC has



$$+ O_2 + \text{Ascorbate} \rightarrow H_2C = CH_2 + HCN + CO_2 + 2\,H_2O + \text{Dehydroascorbate} \quad (24\text{-}35)$$

been known as a minor plant product for over 25 years, it was much more recently identified as the immediate precursor of ethylene. ACC is often produced in response to stresses such as wounding, drought, or waterlogging of roots.[316,321] In the last of these cases the ACC is transferred through the xylem from the roots upward to shoots, which respond in characteristic ways to the ethylene that is released.

The conversion of ACC to ethylene, HCN, and CO_2 is catalyzed by ACC oxidase, an Fe^{2+}-dependent enzyme of the isopenicillin-*N*-synthase (Eq. 18-52) subfamily of oxygenases. However, unlike most of

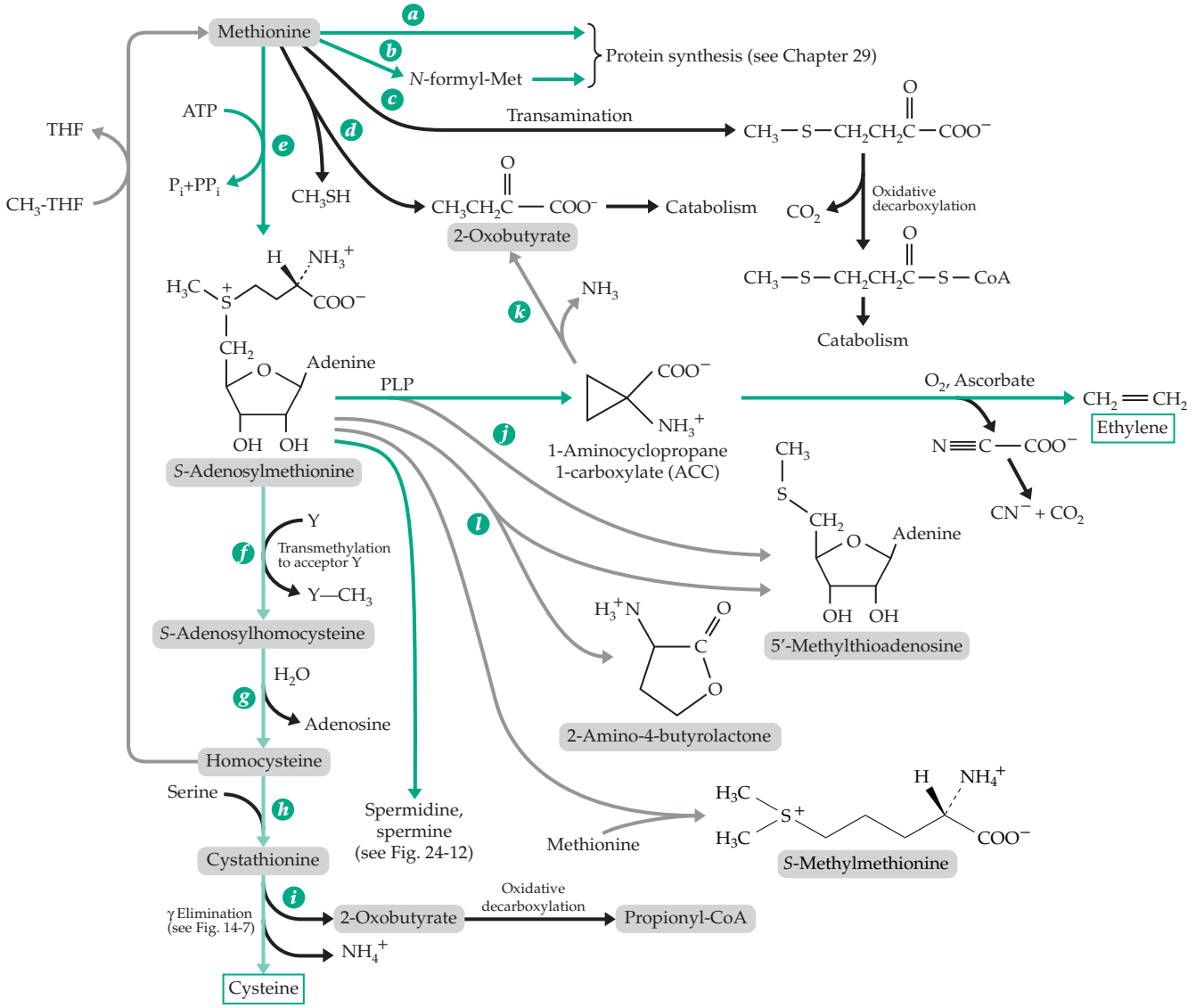
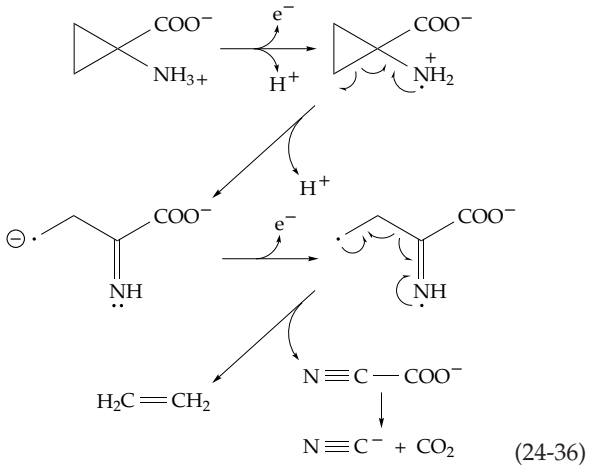


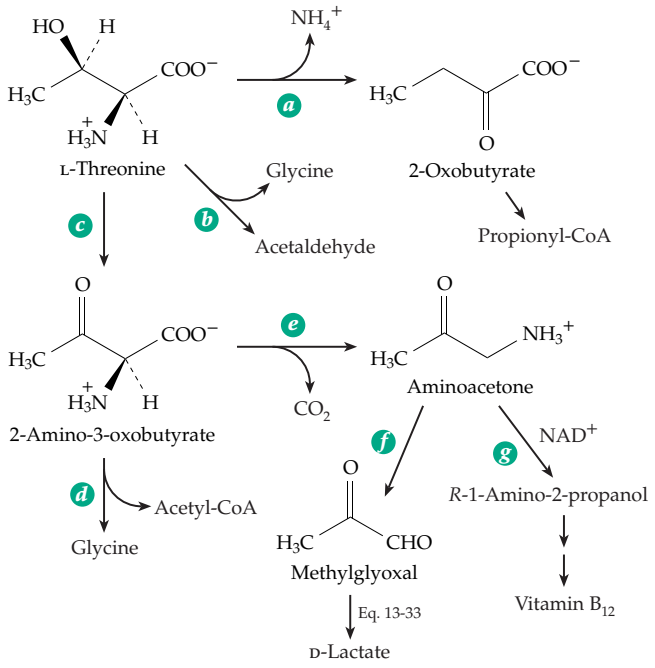
Figure 24-16 Some metabolic reactions of methionine. Biosynthetic reactions are indicated by green arrows.



(24-36)

these enzymes, which utilize 2-oxoglutarate as a co-substrate (Eq. 18-52), ACC oxidase employs ascorbate and forms HCN or cyanide ions.[322–324b] It also requires CO_2 or bicarbonate as an activator.[324,325] A radical mechanism (Eq. 24-36) is probable,[317] with two electrons from ACC and two from ascorbate being utilized to reduce O_2 to 2 H_2O.

Ethylene is rather inert, but it is metabolized slowly, some of it to ethylene glycol.[326] Plants store *N*-malonyl-ACC as a metabolically inert pool. Excess ACC can be deaminated in a PLP-dependent reaction to 2-oxobutyrate (step *k*, Fig. 24-16), a process that also occurs in bacteria able to subsist on ACC.[327,327a] There may also be other mechanisms for ethylene formation, e.g., peroxidation of lipids during scenescence of leaves.[328] See also Chapter 31, Section G.



(24-37)

5. Metabolism of Threonine

Excess threonine is degraded in several ways, one of which is a β elimination reaction catalyzed by L-threonine dehydratase (Eq. 24-37, step *a*). This PLP-requiring enzyme is produced in high amounts in *E. coli* grown on a medium devoid of glucose and oxygen. Under these circumstances the reaction provides a source of propionyl-CoA, which can be converted to propionate with generation of ATP. This **biodegradative threonine dehydratase** (threonine deaminase)[329,330] is allosterically activated by AMP, an appropriate behavior for a key enzyme in energy metabolism. A second **biosynthetic threonine dehydratase** is also produced by *E. coli*[331,332] and is specifically required for production of 2-oxobutyrate needed in the biosynthesis of isoleucine by bacteria, plants,[333] and other autotrophic organisms. In 1956, Umbarger[334] showed that this enzyme is inhibited by isoleucine, the end product of the synthetic pathway. This discovery was instrumental in establishing the concepts of feedback inhibition in metabolic regulation (Chapter 11) and of allostery.

A second catabolic reaction of L-threonine (Eq. 24-37, step *b*) is cleavage to glycine and acetaldehyde. The reaction is catalyzed by serine hydroxymethyltransferase (Eq. 14-30). Some bacteria have a very active D-threonine aldolase.[335] A quantitatively more important route of catabolism in most organisms is dehydrogenation (Eq. 24-37, step *c*)[336] to form 2-amino-3-oxobutyrate. This intermediate can be cleaved by another PLP-dependent enzyme to acetyl-CoA plus glycine (Eq. 24-38, step *d*). It can also be decarboxylated (Eq. 24-38, step *e*) to aminoacetone, a urinary excretion product, or oxidized by amine oxidases to **methylglyoxal** (Eq. 24-37, step *f*).[337] The latter can be converted to D-lactate through the action of glyoxalase (Eq. 13-33). Aminoacetone is also the source of 1-amino-2-propanol for the biosynthesis of vitamin B_{12} (Eq. 24-37, step *g*; Box 16-B).[338,338a]

E. Alanine and the Branched-Chain Amino Acids

As indicated in Fig. 24-17, pyruvate is the starting material for the formation of both L- and D-alanine and also the branched chain amino acids **valine**, **leucine**, and **isoleucine**.[339,340] The chemistry of the reactions has been discussed in the sections indicated in the figure. The first step is catalyzed by the thiamin diphosphate-dependent **acetohydroxyacid synthase** (acetolactate synthase), which joins two molecules of pyruvate or one of pyruvate and one of 2-oxobutyrate (Fig. 24-17; Fig. 14-3).[340a,b] In *E. coli* there are two isoenzymes encoded by genes *ilv B* and *ilv HI*. Both are regulated by feedback inhibition by valine, probably

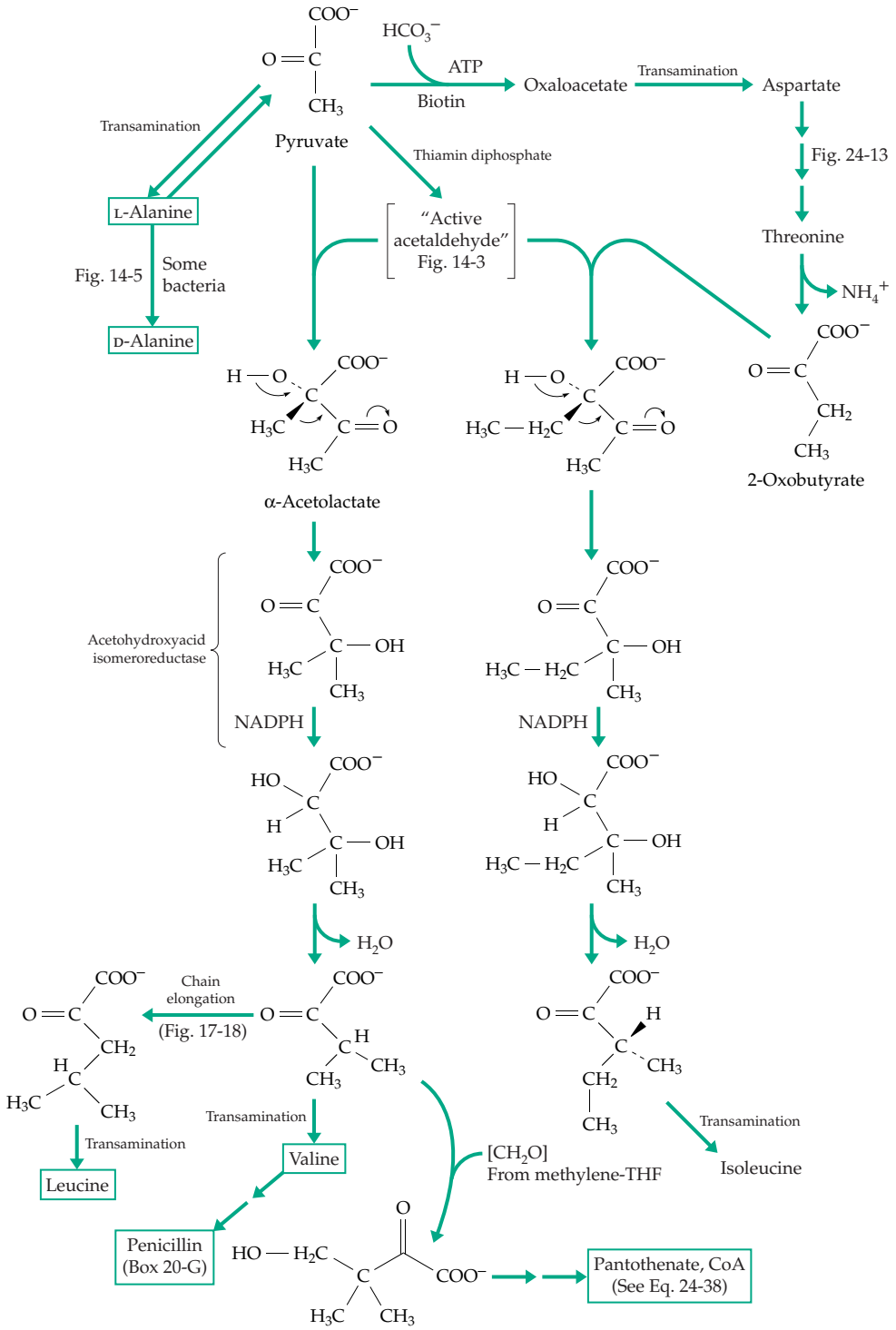
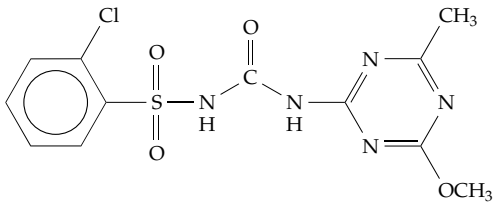
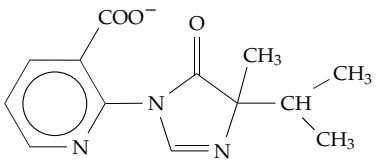


Figure 24-17 Biosynthesis of leucine, isoleucine, valine, and coenzyme A.



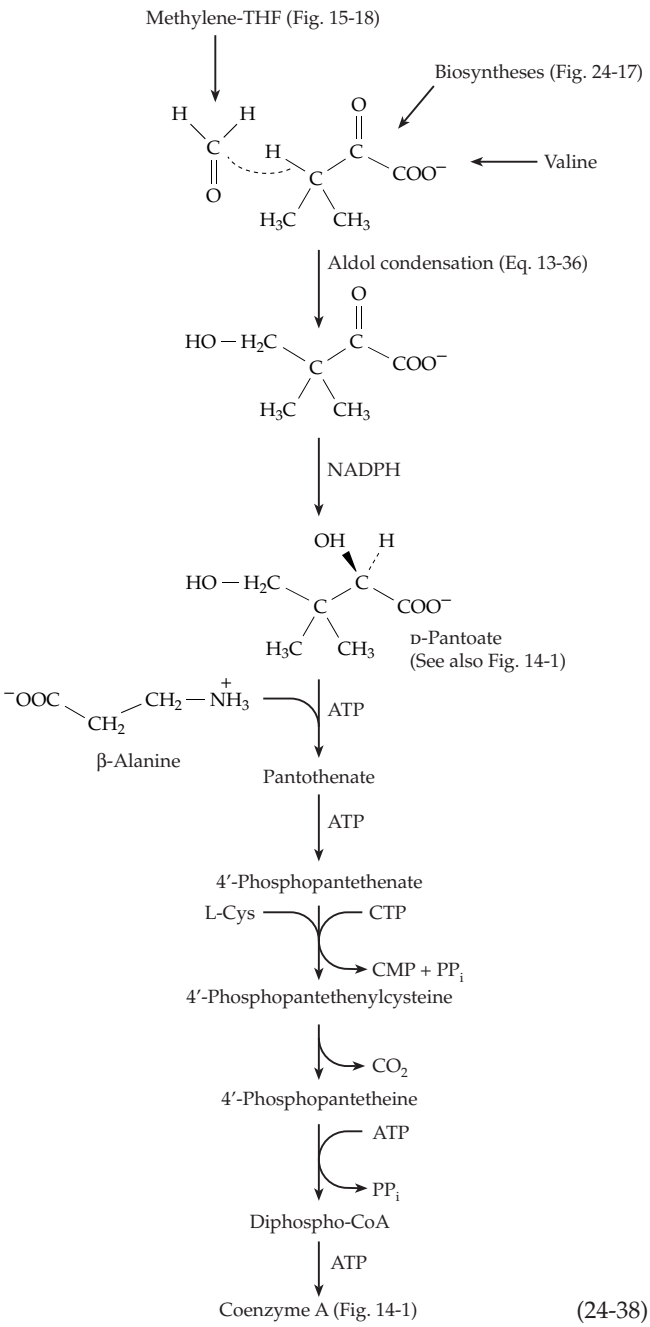
Chlorsulfuron



An imidazolinone herbicide

by an **attenuation** mechanism[341] (explained in Chapter 28). The enzymes are of some practical interest because they are specifically inhibited by two classes of herbicides, the **sulfonylureas**, of which chlorsulfuron is an example, and the **imidazolinones**.[342–345]

The second step in the synthesis, catalyzed by **acetohydroxyacid isomeroreductase**, involves shift of an alkyl group (Fig. 24-17). Neither this reaction nor the preceding one occurs in mammals. For this reason, the enzymes required are both attractive targets for herbicide design.[343,346] The third enzyme, **dihydroxy acid dehydratase**, catalyzes dehydration followed by tautomerization, resembling 6-phosphogluconate dehydratase (Eq. 13-32). The dihydroxyacid dehydratase from spinach contains an Fe_2S_2 cluster and may function by an aconitase type mechanism (Eq. 13-17).[347] In *Neurospora* isoleucine and valine are synthesized in the mitochondria.



(24-38)

While the 2-oxobutyrate needed for isoleucine formation is shown as originating from threonine in Fig. 24-17, bacteria can often make it in other ways,[348] e.g., from glutamate via β-methylaspartate (Fig. 24-8) and transamination to the corresponding 2-oxoacid. It can also be made from pyruvate by chain elongation using acetyl-CoA (Fig. 17-18); citramalate and mesaconate (Fig. 24-8) are intermediates. This latter pathway is used by some methanogens as are other alternative routes.[348] The first step unique to the biosynthetic pathway to leucine is the reaction of the 2-oxo analog of valine with acetyl-Co*A* to form **α-isopropylmalate**, the first step in a chain elongation sequence leading to the oxoacid precursor of leucine (Figs. 17-18; 24-17). The third enzyme required in the chain elongation is a decarboxylating dehydrogenase similar to isocitrate dehydrogenase.[349]

An additional series of reactions,[350] which are shown in Eq. 24-38, leads to **pantoic acid**, **pantetheine**, **coenzyme A**, and related cofactors.[350a–j] The initial reactions of the sequence do not occur in the animal body, explaining our need for pantothenic acid as a vitamin.

Alanine also gives rise to a precursor of the vitamin **biotin** (Eq. 24-39) after a PLP-dependent decarboxylative condensation with the 7-carbon dicarboxylic acid unit of pimeloyl-CoA in a reaction analogous to that of Eq. 14-32.[351] The resulting alcohol is reduced to 7-oxo-8-aminopelargonic acid which is converted by transamination, with *S*-adenosylmethionine as the nitrogen donor,[351a] to 7,8-diaminopelargonic acid. This compound undergoes a two-step ATP-dependent cyclization[352–355] to form **dethiobiotin**. The final step, insertion of sulfur into dethiobiotin, is catalyzed by **biotin synthase**, a free-radical-dependent enzyme related to pyruvate formate lyase (Fig. 15-16). It transfers the sulfur from cysteine via an Fe–S cluster.[355a–c] Biosynthesis of **lipoic acid** involves a similar insertion of two sulfur atoms into octanoic acid.[356] See also p. 1410.

1. Catabolism

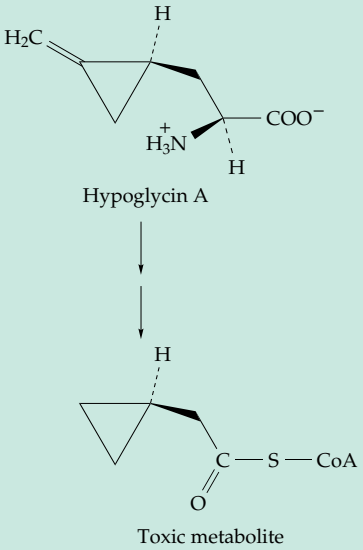
Degradation of amino acids most often begins with conversion, either by transamination[356a] or by NAD^+-dependent dehydrogenation,[357] to the corresponding 2-oxoacid and oxidative decarboxylation of the latter (Fig. 15-16). Alanine, valine, leucine, and isoleucine are all treated this way in the animal body. Alanine gives pyruvate and acetyl-CoA directly, but the others yield CoA derivatives that undergo

BOX 24-A MAPLE SYRUP URINE DISEASE AND JAMAICAN VOMITING SICKNESS

In a rare autosomal recessive condition (discovered in 1954) the urine and perspiration has a maple syrup odor.[a–c] High concentrations of the branched-chain 2-oxoacids formed by transamination of valine, leucine, and isoleucine are present, and the odor arises from decomposition products of these acids. The branched-chain amino acids as well as the related alcohols also accumulate in the blood and are found in the urine. The biochemical defect lies in the enzyme catalyzing oxidative decarboxylation of the oxoacids, as is indicated in Fig. 24-18. Insertions, deletions, and substitutions may be present in any of the subunits (Figs. 15-14, 15-15). The disease which may affect one person in ~200,000, is usually fatal in early childhood if untreated. Children suffer seizures, mental retardation, and coma. They may survive on a low-protein (gelatin) diet supplemented with essential amino acids, but treatment is difficult and a sudden relapse is apt to prove fatal. Some patients respond to administration of thiamin at 20 times the normal daily requirement. The branched-chain oxoacid dehydrogenase from some of these children shows a reduced affinity for the essential coenzyme thiamin diphosphate.[d]

Polled hereford calves in Australia develop maple syrup urine disease relatively often.[a,e] One cause was established as a mutation that introduces a stop codon that causes premature termination within the leader peptide during synthesis of the thiamin diphosphate-dependent E1 subunit. A similar biochemical defect in a mutant of *Bacillus subtilis*[f] causes difficulties for this bacterium, which requires branched-chain fatty acids in its membranes. Branched acyl-CoA derivatives are needed as starter pieces for their synthesis (Chapter 29). With the oxidative decarboxylation of the necessary oxoacids blocked, the mutant is unable to grow unless supplemented with branched-chain fatty acids.

Because persons may be born with defects in almost any gene, a variety of other problems leading to accumulation of organic acids are also known. **Methylmalonic aciduria** and propionic acidemia are discussed in Box 17-B. **Lactic acidemia** (Box 17-F) often results from a defect in pyruvate dehydrogenase. A rare defect of catabolism of leucine is **isovaleric acidemia**, a failure in oxidation of isovaleryl-CoA.[g] The symptoms of this disease are also present in the Jamaican vomiting sickness, caused by eating unripe ackee fruit. Although the ripe fruit is safe to eat, unripe fruit contains a toxin **hypoglycin A** with the following structure.[h–j] It is metabolized to an acyl-CoA derivative as shown.



This is an enzyme-activated inhibitor of the medium-chain fatty acyl-CoA dehydrogenase required for β oxidation of fatty acids.[j,k] The compound also inhibits isovaleryl-CoA dehydrogenase, causing an accumulation of isovaleric acid in the blood. Depression of the central nervous system by isovaleric acid in the blood could be responsible for some symptoms.[h,i] However, death from the highly fatal Jamaican vomiting sickness comes from the hypoglycemic effect. Blood glucose levels may fall as low as 0.5 mM, one-tenth the normal concentration and patients must be treated by infusion of glucose.

---

[a] Patel, M. S., and Harris, R. A. (1995) *FASEB J.* **9**, 1164–1172

[b] Chuang, D. T., and Shih, V. E. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1239–1278, McGraw-Hill, New York

[c] Mamer, O. A., and Reimer, M. L. J. (1992) *J. Biol. Chem.* **267**, 22141–22147

[d] Chuang, D. T., Ku, L. S., and Cox, R. P. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3300–3304

[e] Zhang, B., Healy, P. J., Zhao, Y., Crabb, D. W., and Harris, R. A. (1990) *J. Biol. Chem.* **265**, 2425–2427

[f] Willecke, K., and Pardee, A. B. (1971) *J. Biol. Chem.* **246**, 5264–5272

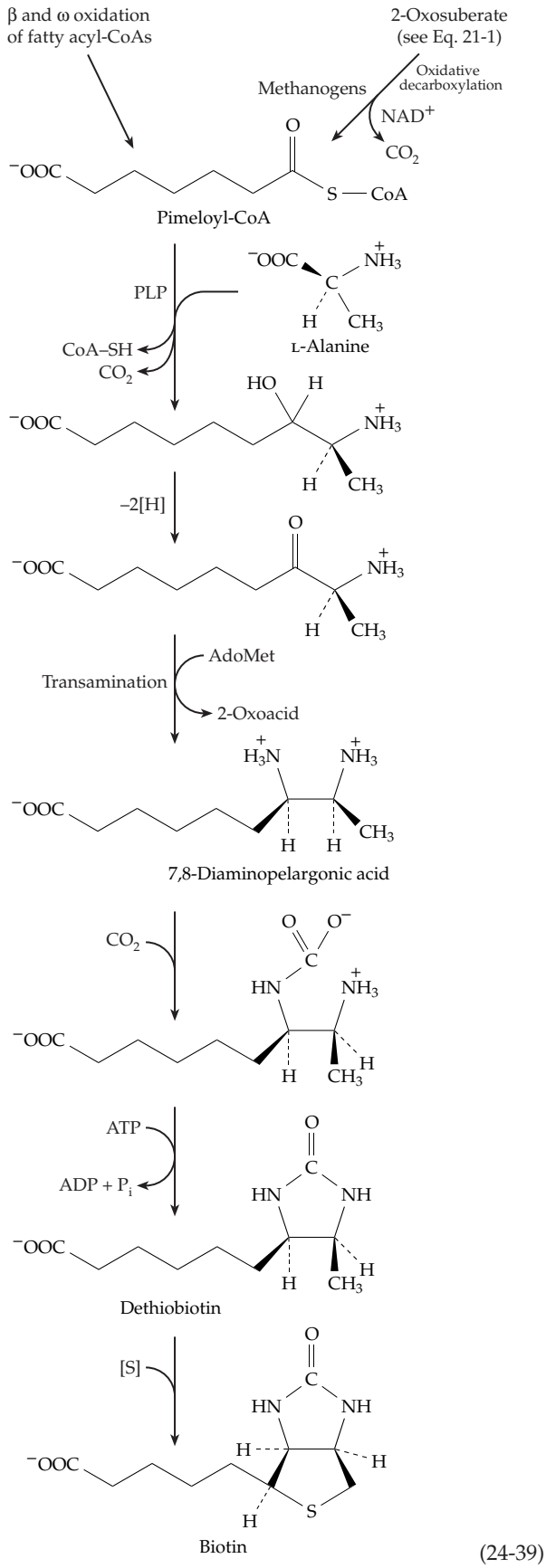
[g] Mohsen, A.-W. A., and Vockley, J. (1995) *Biochemistry* **34**, 10146–10152

[h] Tanaka, K., Isselbacher, K. J., and Shih, V. (1972) *Science* **175**, 69–71

[i] Tanaka, K. (1972) *J. Biol. Chem.* **247**, 7465–7478

[j] Lai, M.-t, Liu, L.-d, and Liu, H.-w. (1991) *J. Am. Chem. Soc.* **113**, 7388–7397

[k] Lai, M.-t, Li, D., Oh, E., and Liu, H.-w. (1993) *J. Am. Chem. Soc.* **115**, 1619–1628

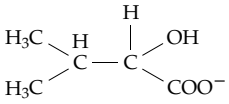


(24-39)

β-oxidation within the mitochondria[358] via the schemes shown in Fig. 24-18. There are some variations from the standard β oxidation sequence for fatty acids shown in Fig. 17-1. In the case of valine the sequence proceeds only to the stage of addition of water to form the β-hydroxy derivative. The latter is converted to free 3-hydroxyisobutyrate, and β oxidation is then completed by oxidation to methylmalonate semialdehyde.[359] The latter is oxidatively decarboxylated to form *S*-methylmalonyl-CoA.[360] Further metabolism of the latter is indicated in Fig. 17-3. However, some methylmalonate semialdehyde may be decarboxylated to propionaldehyde, which could be oxidized to propionate. Either of these compounds could then be metabolized to propionyl-CoA.[361]

In the degradation of isoleucine, β oxidation proceeds to completion in the normal way with generation of acetyl-CoA and propionyl-CoA. However, in the catabolism of leucine after the initial dehydrogenation in the β-oxidation sequence, carbon dioxide is added using a biotin enzyme (Chapter 14). The double bond conjugated with the carbonyl of the thioester makes this carboxylation analogous to a standard β-carboxylation reaction. Why add the extra CO_2? The methyl group in the β position blocks complete β oxidation, but an aldol cleavage would be possible to give acetyl-CoA and acetone. However, acetone is not readily metabolized further. By addition of CO_2 the product becomes acetoacetate, which can readily be completely metabolized through conversion to acetyl-CoA.

An alternative pathway of leucine degradation in the liver is oxidative decarboxylation by a cytosolic oxygenase to form α-hydroxyisovalerate.[362]



α-Hydroxyisovalerate

This compound may be metabolized via the valine catabolic pathway of Fig. 24-18. A third pathway, present in some bacteria, begins with the vitamin B_{12}-dependent isomerization of leucine to β-leucine (Chapter 16), which can undergo transamination to 3-oxoisocaproate. This can be converted to its CoA ester by a CoA transferase and can undergo β cleavage by free CoA-SH to form acetyl-CoA and isobutyryl-CoA. The latter may enter the valine catabolic pathway (Fig. 24-18). Leucine has long been known as a regulator of protein degradation in muscle.[362a–e] Dietary protein deficiency leads to especially rapid degradation of the branched-chain amino acids. The daily turnover of proteins for a 70-kg adult ingesting 70 g of protein per day has been estimated as 280 g, most of which must be reused.[362d] This large turnover can lead to

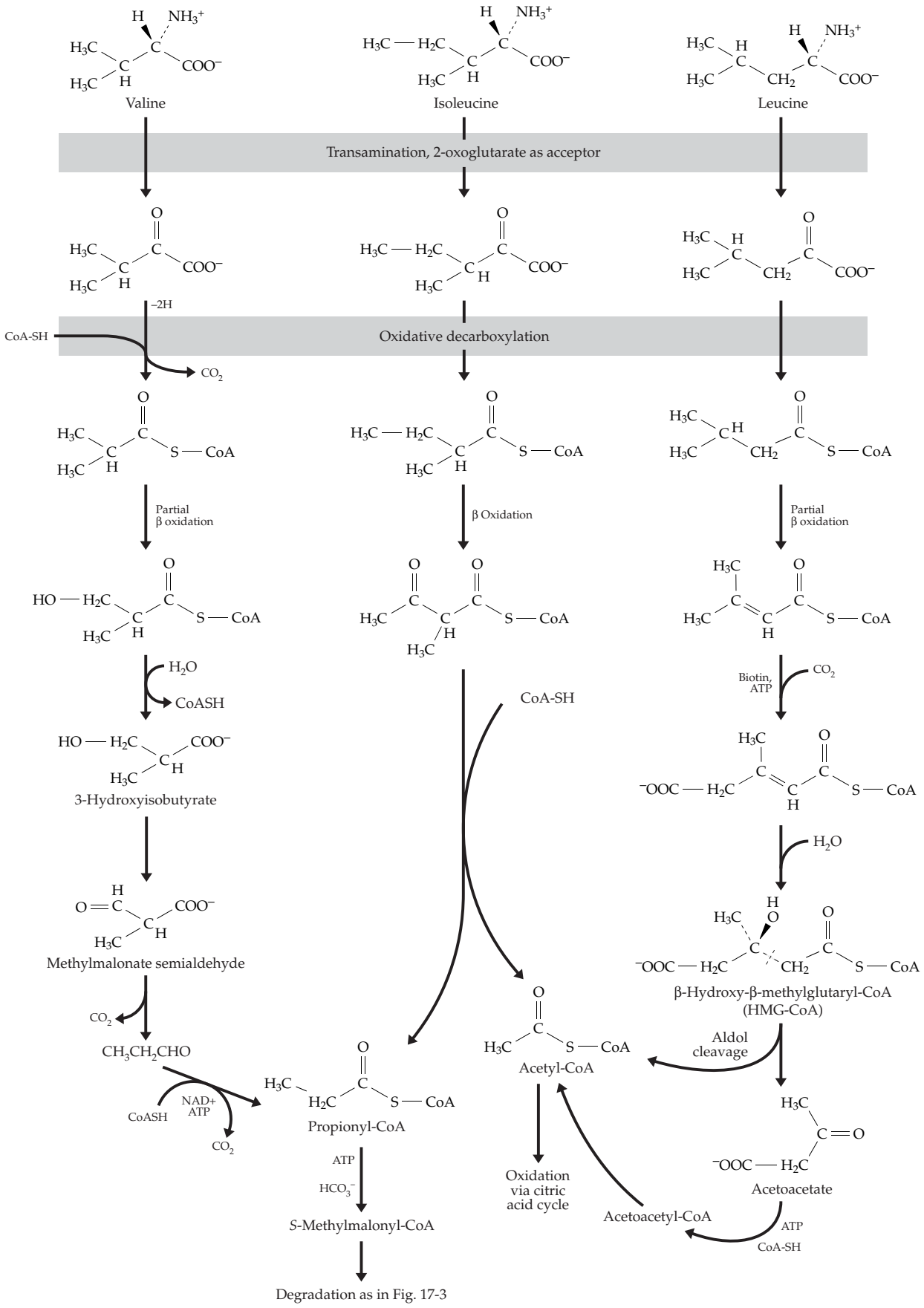


Figure 24-18 Catabolism of valine, leucine, and isoleucine.

excessive muscle wasting in disease states. A minor leucine metabolite found in muscle, β-hydroxy β-methylbutyrate has been proposed as a possible endogenous inhibitor of muscle breakdown.[362e,f] (See study question 17.)

Clostridium propionicum can use alanine as substrate for a balanced fermentation to form ammonium propionate, acetate, and CO_2 (Fig. 24-19).

2. Ketogenic and Glucogenic Amino Acids

According to a long-used classification amino acids are **ketogenic** if (like leucine) they are converted to acetyl-CoA (or acetyl-CoA and acetoacetate). When fed to a starved animal, ketogenic amino acids cause an increased concentration of acetoacetate and other ketone bodies in the blood and urine. On the other hand, **glucogenic** amino acids such as valine, when fed to a starved animal, promote the synthesis of glycogen (in the case of valine via methylmalonyl-CoA, succinate, and oxaloacetate). Examination of Fig. 24-18 shows that isoleucine is both ketogenic and glucogenic, a fact that was known long before the pathway of catabolism was worked out.

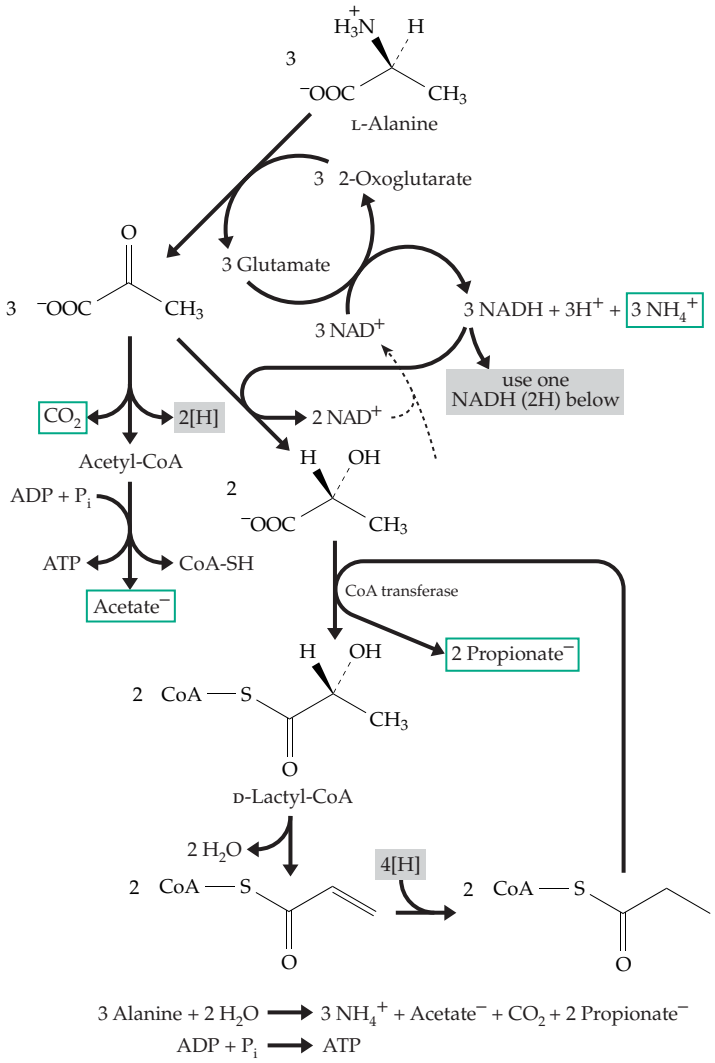


Figure 24-19 Fermentation of L-alanine by *Clostridium propionicum*. After Buckel.[363]

F. Serine and Glycine

Serine originates in a direct pathway from 3-phosphoglycerate (pathway *a*, Fig. 24-20) that involves dehydrogenation, transamination, and hydrolysis by a phosphatase. It can also be formed from glycine by the action of serine hydroxymethyltransferase (Eq. 14-30). This occurs in chloroplasts during photorespiration (Fig. 23-37)[364] and also with some methanogens and other autotropic bacteria and methylotrophs (Fig. 17-15). The glycine decarboxylase cycle shown in Fig. 15-20 provides another mechanism available in bacteria, plants, and animal mitochondria for reversible interconversion of glycine and serine. The principal route of catabolism of serine in many microorganisms is deamination to pyruvate (Fig. 24-20, step *b*),[364a] a reaction also discussed in Chapter 14 (Eq. 14-29). An alternative catabolic pathway is transamination to **hydroxypyruvate,** which as in plants (Fig. 23-37) can be reduced to D-glycerate and back to 3-phosphoglycerate.[365] That this pathway is important in human beings is suggested by the occurrence of a rare metabolic defect **L-glyceric aciduria** (or primary hyperoxaluria type II).[365–367] The biochemical defect may lie in the lack of reduction of hydroxypyruvate to D-glycerate. When hydroxypyruvate accumulates, lactate dehydrogenase effects its reduction to L-glycerate, which is excreted in large amounts (0.3–0.6 g / 24 h) in the urine. Surprisingly, the defect is accompanied by excessive production of oxalate from glyoxylate. This is apparently an indirect result of the primary defect in utilization of hydroxypyruvate. It has been suggested that oxidation of glyoxylate by NAD^+ is coupled to the reduction of hydroxypyruvate by NADH.[366] This and other hyperoxalurias are very serious diseases characterized by the formation of calcium oxalate crystals in tissues and often death from kidney failure before the age of 20.

1. Biosynthetic Pathways from Serine

L-Serine gives rise to many other substances (Fig. 24-20) including **sphingosine** and the **phosphatides**. In many bacteria conversion to

O-acetyl-L-serine (step *c*, Fig. 24-20) provides for the formation of **cysteine** by a β-replacement reaction.[368–369a] Serine is also the major source of glycine (step *d*) and of the single-carbon units needed for the synthesis of methyl and formyl groups. The enzyme **serine hydroxymethyltranferase** (step *d*) also provides the principal route of formation of glycine from serine,[370,371] but a lesser portion comes via phosphatidylserine, phosphatidylethanolamine, phosphatidylcholine, and free choline (step *e*). This pathway includes decarboxylation of phosphatidylserine by a pyruvoyl group-dependent enzyme (pp. 753–755). In contrast, in green plants the major source of ethanolamine is a direct PLP-dependent decarboxylation of serine (Fig. 24-20).[371a] Because the body's capacity to generate methyl groups is limited, **choline** under many circumstances is a dietary essential and has been classified as a vitamin. However, in the presence of adequate amounts of folic acid and vitamin B_{12}, it is not absolutely required. Choline can be used to reform phosphatidylcholine (Fig. 21-5), while an excess can be dehydrogenated to **glycine betaine,** which is one of the osmoprotectant substances in plants (Eq. 24-40).[372,373] This quaternary nitrogen compound is one of a small number of substances that, like methionine, are able to donate methyl groups to other compounds and which are also capable of methylating homocysteine to form methionine. However, the product of transmethylation from betaine, dimethylglycine, is no longer a methylating agent. The two methyl groups are removed oxidatively as formic acid to produce glycine (Eq. 24-40). A third source of glycine is transamination of glyoxylate (step *f*, Fig. 24-20). The equilibrium constant for the reaction favors glycine strongly for almost any amino donor.

2. Metabolism of Glycine

While glycine may be formed from glyoxylate by transamination, the oxidation of glycine by an amino acid oxidase (Table 15-2) permits excess glycine to be converted to glyoxylate. That this pathway, too, is quantitatively important in humans is suggested by the existence of **type 1 hyperoxaluria**.[366] It is thought that a normal pathway for utilization of glyoxylate is blocked in this condition leading to its oxidation to oxalate. The biochemical defect is frequently the absence of a liver-specific alanine:glyoxylate aminotransferase that efficiently converts accumulating glyoxylate back to glycine. In some cases the disease arises because, as a result of

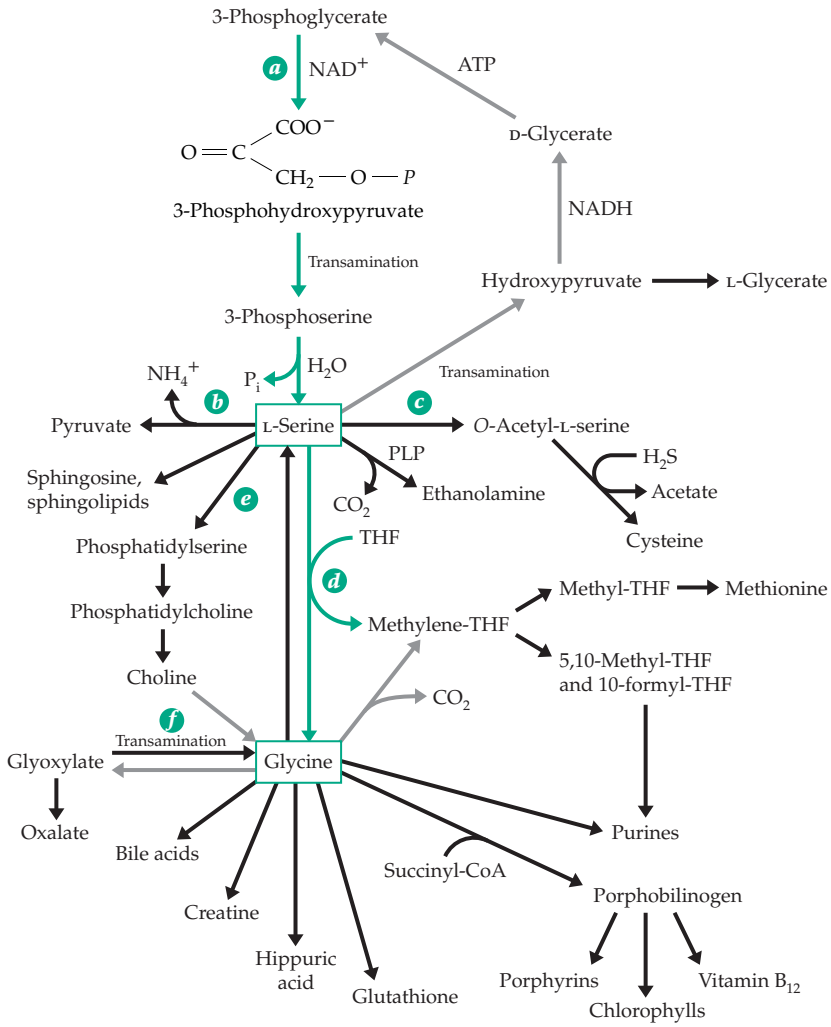
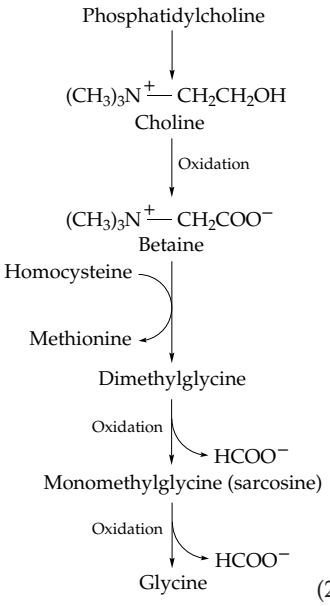


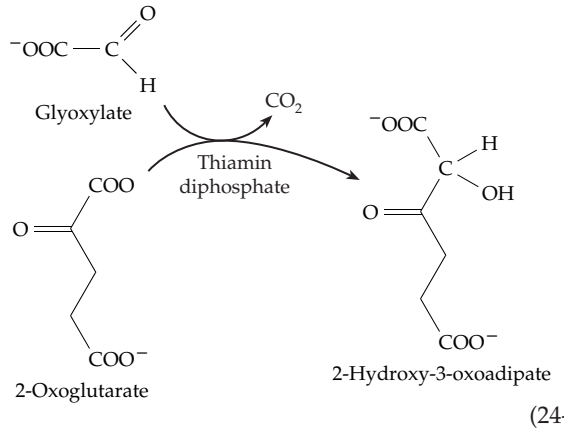
Figure 24-20 Metabolism of serine and glycine.



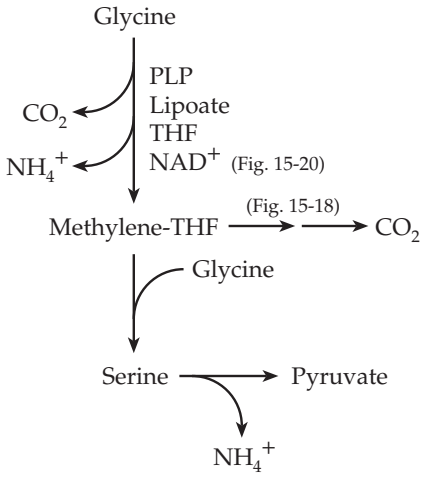
(24-40)

a mutation in its N-terminal targeting sequence the aminotransferase is targeted to mitochondria, where it functions less efficiently.[366a] Another possible defect lies in a thiamin-dependent enzyme that condenses glyoxylate with 2-oxoglutarate to form 2-hydroxy-3-oxoadipate (Eq. 24-41). The function of this reaction is uncertain, but the product could undergo decarboxylation and oxidation to regenerate 2-oxoglutarate. This would provide a cyclic pathway (closely paralleling the dicarboxylic acid cycle, Fig. 17-6) for oxidation of glyoxylate without formation of oxalate. Bear in mind that the demonstrated enzymatic condensation reactions of the rather toxic glyoxylate are numerous, and that its metabolism in most organisms is still not well understood.

An alternative route of catabolism is used by organisms such as *Diplococcus glycinophilus*, which are able to grow on glycine as a sole source of energy, of carbon, and of nitrogen,[374] and is also used in mitochondria of plants and animals.[374a] This glycine cleavage system, depicted in Fig. 15-20, involves decarboxylation, oxidation by NAD^+, release of ammonia, and transfer of the decarboxylated α-carbon of glycine to tetrahydrofolic acid (THF) to form methylene-THF. The C-1 methylene unit of the latter is used primarily for purine biosynthesis but can also be oxidized to CO_2 or can condense with another molecule of glycine (Fig. 15-18, step *c*, reverse) to form serine. This can in turn be converted to pyruvate and utilized for biosynthetic processes (Eq. 24-42).

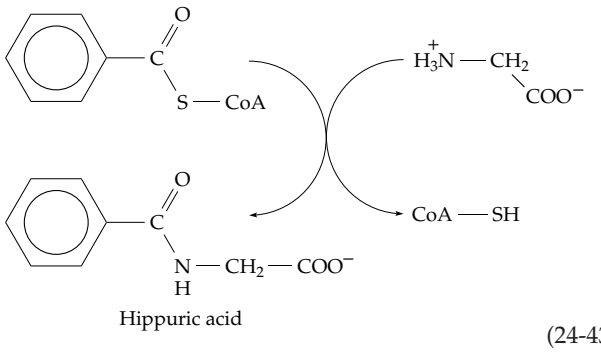


(24-41)



(24-42)

Glycine can be reduced to acetate and ammonia by the selenium-dependent clostridial glycine reductase system (Eq. 15-61). A variety of additional products can be formed from glycine as is indicated in Fig. 24-20. **Hippuric acid** (Box 10-A), the usual urinary excretion product in the "detoxication" of benzoic acid, is formed via benzoyl-CoA (Eq. 24-43):



(24-43)

N-Methylation yields the monomethyl derivative **sarcosine**[375] and also dimethylglycine, compounds that may function as osmoprotectants (Box 20-C). Many bacteria produce **sarcosine oxidase**, a flavoprotein that oxidizes its substrate back to glycine and formaldehyde, which can react with tetrahydrofolate.[376–377a] The formation of porphobilinogen and the various pyrrole pigments derived from it and the synthesis of the purine ring (Chapter 25) represent two other major routes for glycine metabolism.

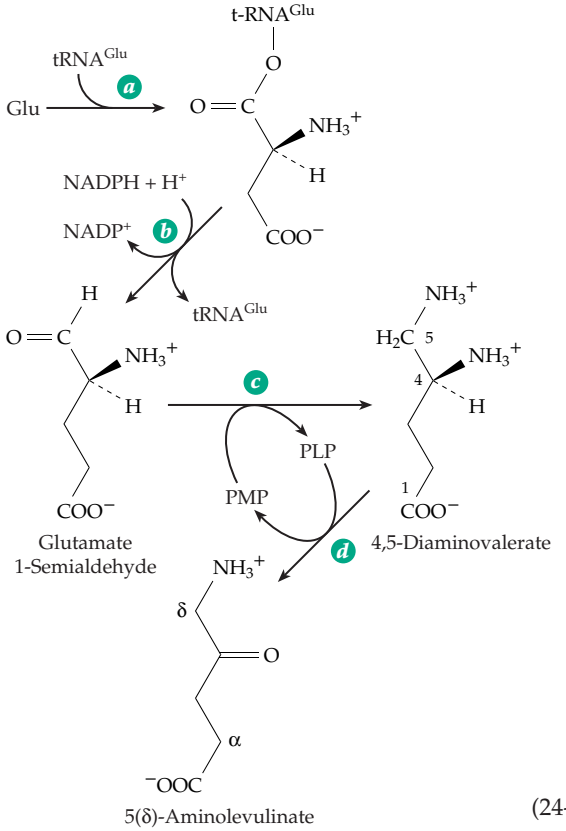
3. Porphobilinogen, Porphyrins, and Related Substances

In 1946, Shemin and Rittenberg[378] described one of the first successful uses of radiotracers in the study of metabolism. They reported that the atoms of the porphyrin ring in heme have their origins in the simple compounds acetate and glycine. As we now know, acetate is converted to succinyl-CoA in the citric acid cycle. Within the mitochondrial matrix of animal cells succinyl-CoA condenses with glycine to form **5(δ)-aminolevulinic acid** (Eq. 14-32),[379–381a] which is converted to **porphobilinogen** (Fig. 24-21), the immediate precursor to the porphyrins. The same pathways lead also to other tetrapyrroles including chlorophyll, the nickel-containing F_{430}, vitamin B_{12}, and other corrins.[382,383]

By degradation of ^{14}C-labeled porphyrins formed from labeled acetate and glycine molecules, Shemin and Rittenberg established the labeling pattern for the

pyrrole ring that is indicated for porphobilinogen in Fig. 24-21. The solid circles mark those atoms that were found to be derived from methyl carbon atoms of acetate (bear in mind that acetyl groups of acetyl-CoA pass around the citric acid cycle more than once to introduce label from the methyl group of acetate into both the 2 and 3 positions of succinyl-CoA). Those atoms marked with open circles in Fig. 24-21 were found to be derived mainly from the methyl carbon of acetate and in small part from the carboxyl carbon. Atoms marked with asterisks came from glycine, while unmarked carbon atoms came from the carboxyl carbon of acetate.[384]

In cyanobacteria and in chloroplasts the intact 5-carbon skeleton of glutamate enters δ-aminolevulinate.[139,385,385a] A surprising finding was that the glutamate becomes coupled to one of the three known glutamate isoacceptor tRNAs that are utilized for protein synthesis. The aminoacyl-tRNA is formed in the usual manner with an ester linkage to the CCA end of the tRNA (Eq. 24-44; see also Eq. 17-36). This ester linkage can be reductively cleaved by NADPH to form glutamate 1-semialdehyde.[386] Isomerization of the glutamate semialdehyde to δ-aminolevulinate is accomplished by an aminomutase that is structurally and functionally related to aminotransferases.[139,387] The enzyme utilizes pyridoxamine phosphate (PMP) to transaminate the substrate carbonyl group to form 4,5-diaminovalerate plus bound PLP. A second transamination step yields the product and regenerates the PMP (Eq. 24-44, steps *c* and *d*).[387]



(24-44)

Porphyrins. As indicated in Fig. 24-21, the conversion of two molecules of 5-aminolevulinate into porphobilinogen is a multistep reaction initiated by **5-aminolevulinate dehydratase** (porphobilinogen synthase).[381,388–390b] The enzyme binds two molecules of substrate in distinct sites known as the A site and the P site (Fig. 24-21). The substrate in the P site forms a Schiff base with a lysine side chain (K247 in the *E. coli* enzyme), while a bound Zn^{2+} is thought to polarize the carbonyl of the substrate in the A site. An aldol condensation (Fig. 24-21, step *a*) ensues and is followed by dehydration to form a carbon–carbon double bond and ring closure (step *b*). Tautomerization step (*c*) leads to porphobilinogen. The enzyme is a sensitive target for poisoning by lead ions.[381,390c] Condensation to form porphyrins requires two enzymes, **porphobilinogen deaminase** (hydroxymethylbilane synthase) and **uroporphyrinogen III cosynthase**. Porphobilinogen deaminase has a bound coenzyme (prosthetic group) consisting of two linked pyrromethane groups, also derived from porphobilinogen.[391] The first step in assembling the porphyrin ring is condensation of porphobilinogen with this coenzyme (Fig. 24-21, step *d*). To initiate this step ammonia is eliminated, probably not by the direct displacements, but by electron flow from the adjacent nitrogen in the same pyrrole ring as indicated in the figure to give an exocyclic double bond. The terminal ring of the coenzyme then adds to the double bond. The condensation process is repeated four times to produce **preuroporphyrinogen** (hydroxymethylbilane).[392–394a] This intermediate is a precursor of the symmetric uroporphyrin I (Fig. 16-5). In the presence of the cosynthase a different ring-closure reaction takes place. The five-membered ring in porphobilinogen has a symmetric arrangement of double bonds. Thus, a condensation reaction can occur at either of the positions α to the ring nitrogen. A sequence of condensation, tautomerization, cleavage, and reformation of the ring as shown in steps *e* to *h* of Fig. 24-21 leads to the unsymmetric uroporphyrinogen III with its characteristic pattern of the carboxymethyl and carboxyethyl side chains. A series of decarboxylation and oxidation reactions then leads directly to protoporphyrin IX.

The first of these decarboxylations is catalyzed by the cytoplasmic **uroporphyrinogen decarboxylase,** which removes the carboxylate groups of the four acetate side chains sequentially from the D, A, B, and C rings.[395–396a] A possible mechanism, utilizing a tautomerized ring, is illustrated in the accompanying structural formula.

The decarboxylated product, coproporphyrinogen (Fig. 16-5), enters the mitochondria and is acted upon by **coproporphyrinogen oxidase,** which oxidatively

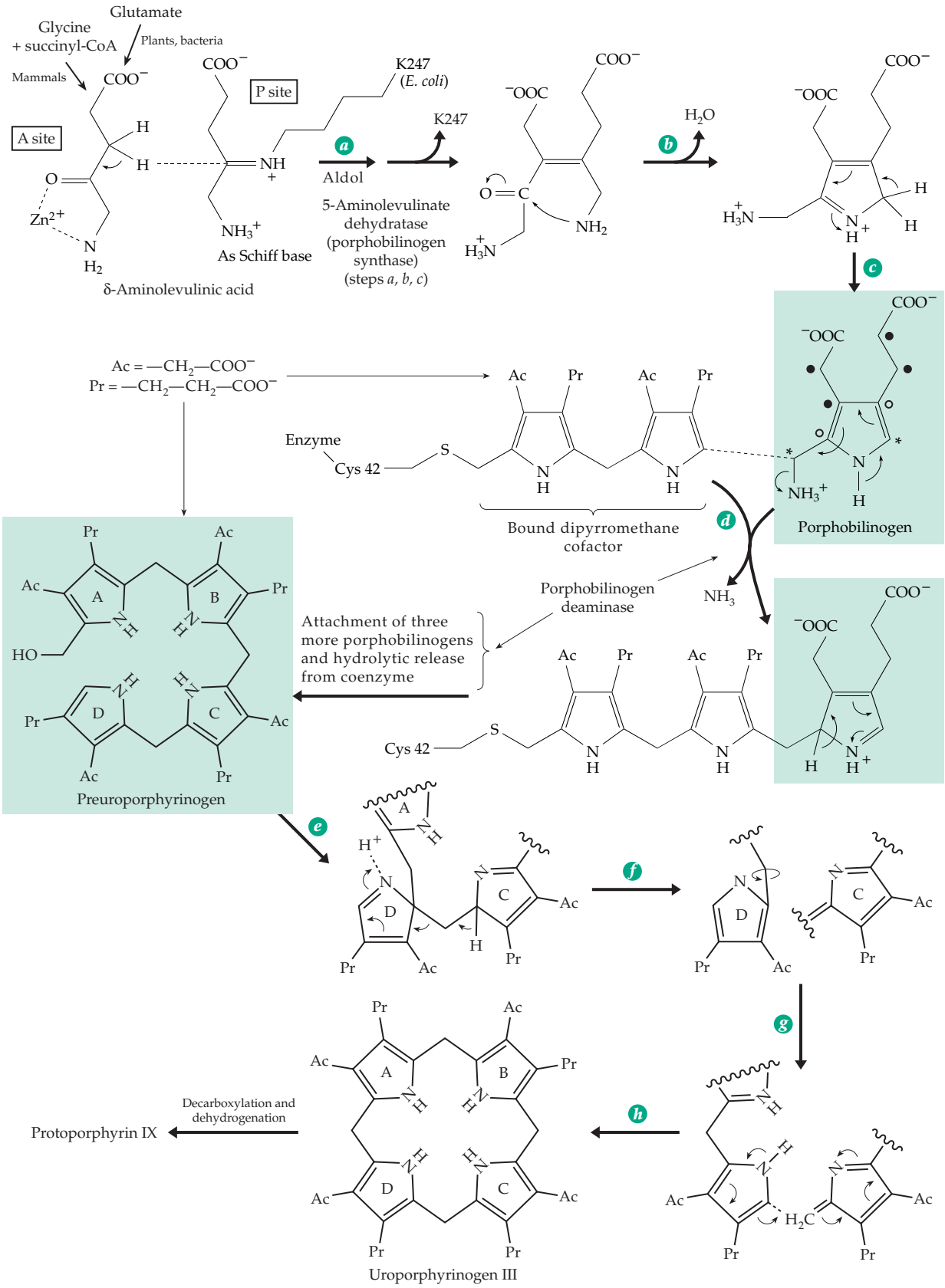
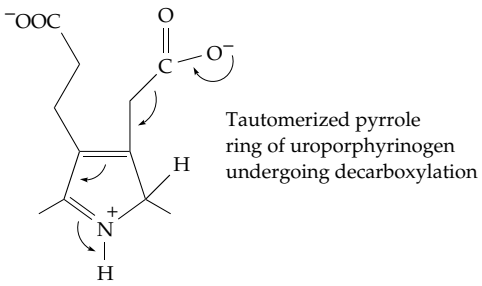


Figure 24-21 Biosynthesis of porphyrins, chlorins, and related compounds.



decarboxylates and oxidizes two of the propyl side chains to vinyl groups.[397] A flavoprotein, **protoporphyrinogen oxidase**, oxidizes the methylene bridges between the pyrrole rings[398,399] to form protoporphyrin IX. A somewhat different pathway from uroporphyrinogen is followed by sulfate-reducing bacteria.[400]

Ferrochelatase (protoheme ferro-lyase)[401–403] inserts Fe^{2+} into protoporphyrin IX to form heme. The enzyme is found firmly bound to the inner membrane of mitochondria of animal cells, chloroplasts of plants, and chromatophores of bacteria. While Fe^{2+} is apparently the only metallic ion ordinarily inserted into a porphyrin, the Zn^{2+} protoporphyrin chelate accumulates in substantial amounts in yeast, and Cu^{2+}–heme complexes are known (p. 843). Ferrochelatase, whose activity is stimulated by Ca^{2+}, appears to be inhibited by lead ions, a fact that may account for some of the acute toxicity of lead.[404]

Heme *b* is utilized for formation of hemoglobin, myoglobin, and many enzymes. It reacts with appropriate protein precursors to form the cytochromes *c*. Heme *b* is converted by prenylation to heme *o*[405] and by prenylation and oxidation to heme *a*.[405a] The porphyrin biosynthetic pathway also has a number of branches that lead to formation of corrins, chlorins, and chlorophylls as shown schematically in Fig. 24-22.

Corrins. The formation of vitamin B_{12}, other corrins, siroheme, and related chlorin chelates[406,407] requires a ring contraction with elimination of the methine bridge between rings A and D of the porphyrins (see Box 16-B). It is natural to assume that the methyl group at C-1 of the corrin ring might arise from the same precursor carbon atom as does the methine bridge in porphyrins, and it is easy to visualize a modified condensation reaction by which ring closure at step *e* in Fig. 24-21 occurs by nucleophilic addition to a C=N bond of ring A. However, ^{13}C-NMR data ruled out this possibility. When vitamin B_{12} was synthesized in the presence of ^{13}C-methyl-containing methionine, it was found that seven methyl groups contained ^{13}C. All of the "extra" methyl groups around the periphery of the molecule as well as the one at C-1 were labeled.[408] Other experiments established uroporphyrinogen III as a precursor of vitamin B_{12}. Therefore, it appeared that the ring first closed in

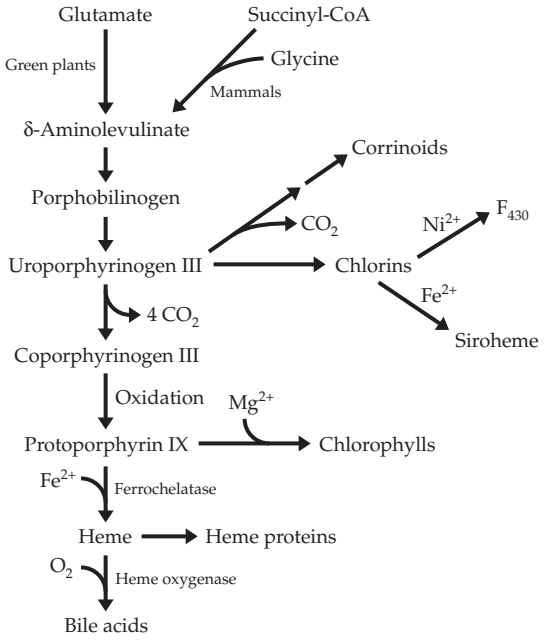


Figure 24-22 Abbreviated biosynthetic pathways from δ-aminolevulinate to heme proteins, corrins, chlorophylls, and related substances.

a normal way and then reopened between rings A and D with removal of the carbon that forms the methylene bridge.[409] This turned out to be true, but with some surprises.

The complex pathways of corrin synthesis have been worked out in detail.[410–415a] This has been possible because of extensive use of ^{13}C and ^{1}H NMR and because the group of ~20 enzymes required has been produced in the laboratory from genes cloned from *Pseudomonas denitrificans*.[410] The first alterations of uroporphyrinogen are AdoMet-dependent methylations on carbon atoms. Surprisingly, one of these is on the bridge atom that is later removed. The details, including the insertion of Co^{2+} by a **cobaltochelatase**, are described by Battersby[410] and portrayed in Michal's *Biochemical Pathways*.[416]

Chlorophyll. The pathway of chlorophyll synthesis has been elucidated through biochemical genetic studies of *Rhodobacter spheroides*[417–418a] which produces bacteriochlorophyll, from studies of cyanobacteria,[419,420] and from investigations of green algae and higher plants,[421] which make chlorophyll *a*. The first step in the conversion of protoporphyrin IX into chlorophyll is the insertion of Mg^{2+} (Fig. 24-23, step *a*). This reaction does not occur readily spontaneously but is catalyzed by an ATP-dependent **magnesium protoporphyrin chelatase**.[419,422] Subsequently, the carboxyethyl side chain on ring C undergoes methylation (Fig. 24-23, step *b*) and β oxidation (step *c*).

Oxidative closure of ring E (step *d*) is followed by reduction of the vinyl group of ring B and of the double bond in ring D to form **chlorophyllide *a***. The latter is coupled with phytol, via phytyl diphosphate, to form chlorophyll *a*.[420] Chlorophyll *b* is derived from chlorophyll *a*, evidently by action of an as yet uncharacterized oxygenase, which converts the methyl group on ring B into a formyl group.[423,424] Bacteriochlorophylls also arise from chlorophyllide *a* and involve reduction of the double bond in ring B.[416,418,420] Most photosynthetic bacteria make bacteriochlorophylls esterified with the C_{20} phytol, but some substitute the unsaturated C_{20} geranylgeranyl group and a variety of other isoprenoid alcohols.

The porphyrias. The human body does not use all of the porphobilinogen produced, and a small amount is normally excreted in the urine, principally as coproporpyrins (Fig. 16-5). In a number of hereditary and acquired conditions blood porphyrin levels are elevated and enhanced urinary excretion (porphyria) is observed.[425–427] Porphyrias may be mild and almost without symptoms, but the intensely fluorescent free porphyrins are sometimes deposited under the skin and cause photosensitivity and ulceration. In extreme cases, in which the excreted porphyrins may color the urine a wine red, patients may have acute neurological attacks and a variety of other symptoms. Lucid accounts of such symptoms, experienced by King George III of England, have been written.[426,428] However, there are doubts about the conclusion that the king suffered from porphyria.[425]

Porphyria may result from several different enzyme deficiences in the porphyrin biosynthetic pathway. The condition is often hereditary but may be induced by drugs or other xenobiotic substances and may be continuous or intermittent.[425,426,429] In one type of congenital porphyria uroporphyrin I is excreted in large quantities. The biochemical defect appears to be a deficiency of the cosynthase that is required for formation of protoporphyrin IX. Another type of porphyria results from overproduction in the liver of

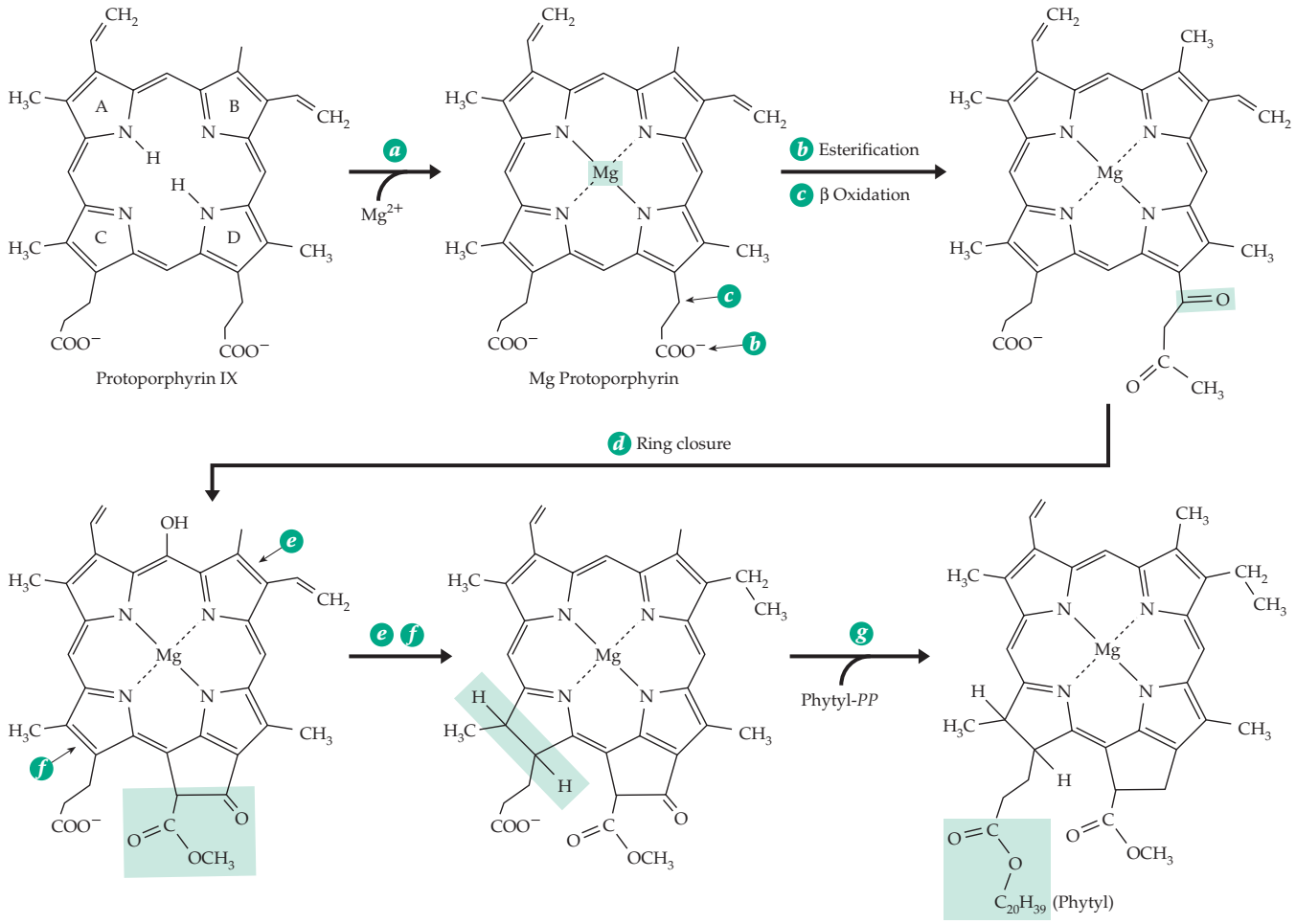
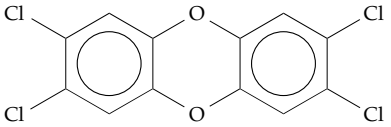


Figure 24-23 Outline of the biosynthetic pathways for conversion of protoporphyrin IX into the chlorophylls and bacteriochlorophylls. After Bollivar *et al.*[417]

δ-aminolevulinic acid, a compound with neurotoxic properties, possibly as a result of its similarity to the neurotransmitter γ-aminobutyrate.[426,427] This may account for some of the neurological symptoms of porphyria.

Some mild forms of intermittent porphyria may go unrecognized. However, ingestion of drugs can precipitate an acute attack, probably by inducing excessive synthesis of δ-aminolevulinate synthase. Among compounds having this effect are hexachlorobenzene and tetrachlorodibenzodioxin.



2,3,7,8-Tetrachlorodibenzo-*p*-dioxin

The latter is one of the most potent inducers of the synthase known.[430] The tendency for this dioxin to be present as an impurity in the herbicide 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) has caused concern. For rodents this dioxin may be the most toxic small molecule known, the oral LD_{50} for guinea pigs being only 1 μg / kg body weight.[430] However, it is over 1000 times less toxic to humans.[430a] It is also a potent teratogenic agent. Synthesis of porphyrins in the liver is controlled by δ-aminolevulinate synthase. This key enzyme is sensitive to feedback inhibition by heme, but the increased synthesis of the enzyme induced by drugs can override the inhibition. Several times as much heme is synthesized in the erythroid cells of bone than in liver, but this is not subject to feedback regulation or to stimulation by drugs.[427] Heme is a potent and toxic regulator. Malaria mosquitos, which utilize blood for food but do not have a heme oxygenase (Fig. 24-24), detoxify heme by inducing its aggregation into an insoluble hydrogen-bonded solid known as β-hematin.[431]

The bile pigments. The enzymatic degradation of heme is an important metabolic process if only because it releases iron to be reutilized by the body. Some of the pathways are illustrated in Fig. 24-24. The initial oxidative attack is by the microsomal **heme oxygenases,**[432–434a] which catalyze the uptake of three molecules of O_2, formation of CO, and release of the chelated Fe. The electron transport protein NADPH-cytochrome P450 reductase brings electrons from NADPH to the oxygenase. An enzyme–substrate heme complex is formed with the oxygenase. Then the Fe^{3+} is reduced to Fe^{2+} which binds O_2 as in myoglobin or hemoglobin. The complex hydroxylates its own heme α-carbon (Fig. 24-24), the other oxygen atom being reduced to OH^- by the Fe^{2+} and an additional electron from NADPH. The same enzyme catalyzes the next steps in which the α carbon is split out as CO by reaction with two molecules of O_2 to form the open tetrapyrrole dicarbonyl compound **biliverdin**, one of the bile pigments (Fig. 24-24).[435–435a] When $^{18}O_2$ was used, it was found that the biliverdin contains two atoms of ^{18}O, and that the CO contains one. Heme from the cytochromes *c* appears to be degraded by the same enzymes after proteolytic release from the proteins to which it is bound.[436]

There are two human heme oxygenases. The first (HO-1) is synthesized principally in the liver and spleen. Its formation is strongly induced by heme. The second heme oxygenase (HO-2) is distributed widely among tissues, but it is most abundant in certain neurons in the brain.[437,437a] Its major function may be to generate CO, which is now recognized as a probable neurohormone (Chapter 30). Bacteria, such as *Corynebacterium diphtheriae,* employ their own heme oxygenase as a means of recovering iron that they need for growth.[438]

A large number of other open tetrapyrroles can be formed from biliverdin by reduction or oxidation reactions. Within our bodies biliverdin is reduced to **bilirubin**, which is transported to the liver as a complex with serum albumin. In the liver bilirubin is converted into glucuronides (Eq. 20-16), glycosylation occurring on the propionic acid side chains.[439] A variety of these bilirubin conjugates are excreted into the bile. In the intestine they are hydrolyzed back to free bilirubin, which is reduced by the action of intestinal bacteria to urobilinogen, stercobilinogen, and *meso*-bilirubinogen. These compounds are colorless but are readily oxidized by oxygen to **urobilin** and **stercobilin**. Some of the urobilin and other bile pigments is reabsorbed into the blood and excreted into the urine where it provides the familiar yellow color.

The yellowing of the skin known as **jaundice** can occur if the heme degradation system is overburdened (e.g., from excessive hemolysis), if the liver fails to conjugate bilirubin, or if there is obstruction of the flow of heme breakdown products into the intestinal tract. Bilirubin is toxic, and continued exposure to excessive bilirubin levels can cause brain damage.[434,439] Bilirubin has a low water solubility and tends to form complexes with various proteins, perhaps partly because it assumes folded conformations rather than the linear one shown in Fig. 24-24.[440] These properties make it difficult to excrete. Thousands of newborn babies are treated for jaundice every year by prolonged irradiation with blue or white light which isomerizes 4Z,15Z bilirubin to forms that are more readily transported, metabolized, and excreted.[441] A more difficult problem is posed by the fatal deficiency of the glucuronosyl transferase responsible for formation of bilirubin glucuronide. Efforts are being made to develop a genetic therapy.[439]

The open tetrapyrroles of algae and the chromophore of phytochrome (Chapter 23) are all derived from **phycoerythrobilin**, which is formed from biliverdin, as indicated in Fig. 24-24. The animal bile pigments have not been found in prokaryotes. However, *Clostridium tetanomorphum*, which accumulates uroporphyrinogen III, a precursor to vitamin B_{12}, and does not synthesize protoporphyrin IX, makes a blue bile pigment **bactobilin**. This is a derivative of uroporphyrin rather than of protoporphyrin.[442]

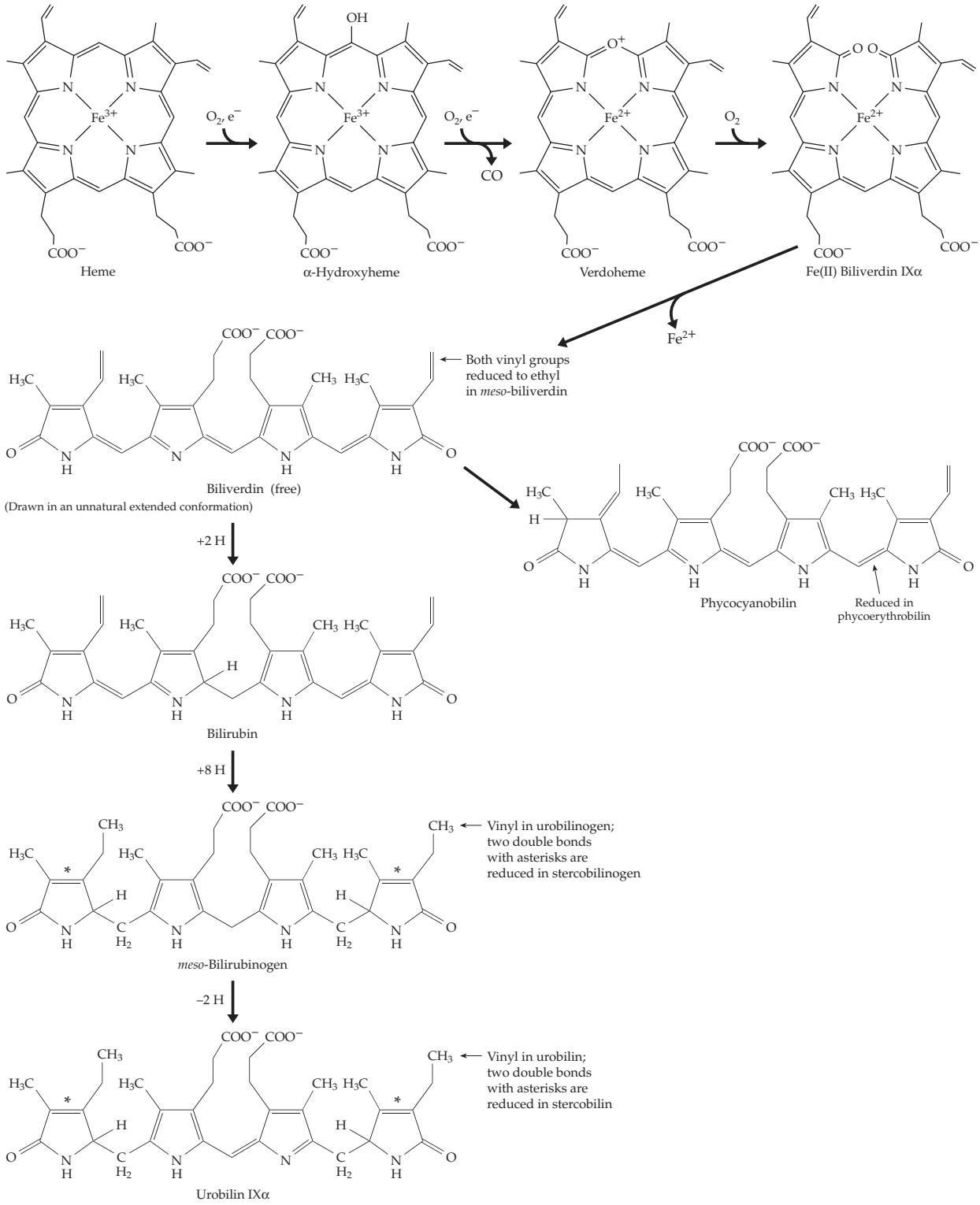


Figure 24-24 The degradation of heme and the formation of open tetrapyrrole pigments.

G. Cysteine and Sulfur Metabolism

Cysteine not only is an essential constituent of proteins but also lies on the major route of incorporation of inorganic sulfur into organic compounds.[443] Autotrophic organisms carry out the stepwise reduction of sulfate to sulfite and sulfide (H_2S). These reduced sulfur compounds are the ones that are incorporated into organic substances. Animals make use of the organic sulfur compounds formed by the autotrophs and have an active oxidative metabolism by which the compounds can be decomposed and the sulfur reoxidized to sulfate. Several aspects of cysteine metabolism are summarized in Fig. 24-25. Some of the chemistry of inorganic sulfur metabolism has been discussed in earlier chapters. Sulfate is reduced to H_2S by sulfate-reducing bacteria (Chapter 18). The initial step in *assimilative* sulfate reduction, used by autotrophs including green plants and *E. coli*, is the formation of adenosine 5'-phosphosulfate (APS) (step *a*, Fig. 24-25; see also Eq. 17-38).[444–444c] The sulfate-reducing bacteria reduce adenylyl sulfate directly to sulfite (Eq. 18-32, step *b*), but the assimilative pathway of reduction in *E. coli* proceeds through 3'-phospho-5'-adenylyl sulfate (PAPS), a compound whose function as "active sulfate" has been considered in Chapter 17. Reduction of PAPS to sulfite (Fig. 24-25, step *d*) is accomplished by an NADPH-dependent enzyme.

The same pathway is found in the alga *Chlorella*, but a second route of sulfate reduction occuring in green plants may be more important.[445] Adenylyl sulfate transfers its sulfo group to a thiol group of a carrier (Eq. 24-43, step *a*). The resulting thiosulfonate is reduced by a ferredoxin-dependent reductase. Finally, a sulfide group is transferred from the –S—S⁻ group of the reduced carrier directly into cysteine in a

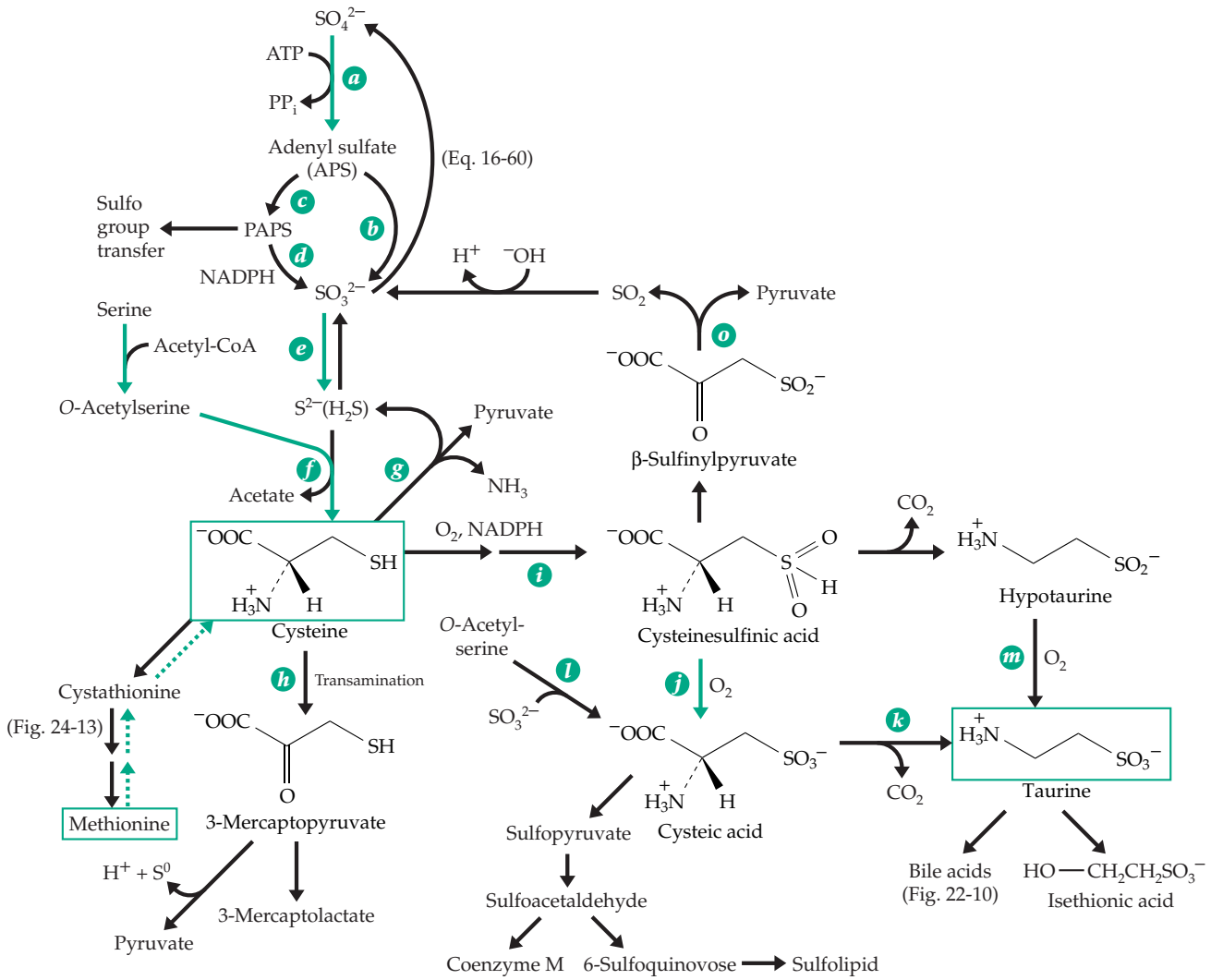
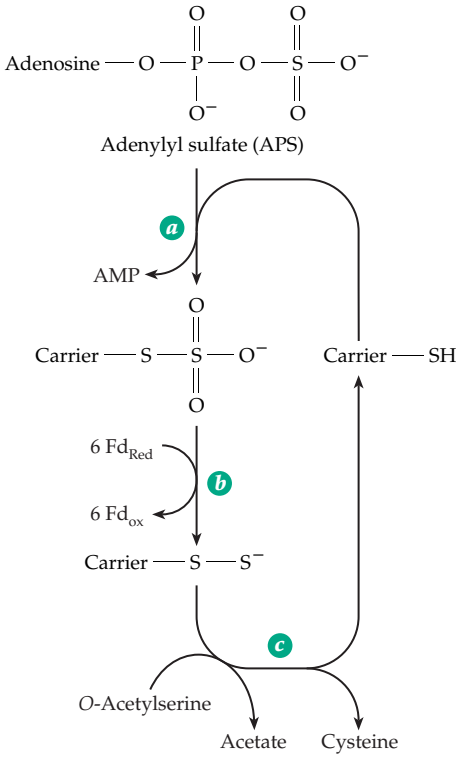


Figure 24-25 Pathways of biosynthesis (green arrows) and catabolism of cysteine as well as other aspects of sulfur metabolism. Solid arrows are major biosynthetic pathways. The dashed arrows represent more specialized pathways; they also show processes occurring in the animal body to convert methionine to cysteine and to degrade the latter.



(24-43)

β-substitution reaction analogous to that described in the next paragraph.

1. Synthesis and Catabolism of Cysteine

Cysteine is formed in plants and in bacteria from sulfide and serine after the latter has been acetylated by transfer of an acetyl group from acetyl-CoA (Fig. 24-25, step *f*). This standard PLP-dependent β replacement (Chapter 14) is catalyzed by **cysteine synthase** (*O*-acetylserine sulfhydrase).[446,447] A similar enzyme is used by some cells to introduce sulfide ion directly into homocysteine, via either *O*-succinyl homoserine or *O*-acetyl homoserine (Fig. 24-13). In *E. coli* cysteine can be converted to methionine, as outlined in Eq. 16-22 and as indicated on the right side of Fig. 24-13 by the green arrows. In animals the converse process, the conversion of methionine to cysteine (gray arrows in Fig. 24-13, also Fig. 24-16), is important. Animals are unable to incorporate sulfide directly into cysteine, and this amino acid must be either provided in the diet or formed from dietary methionine. The latter process is limited, and cysteine is an essential dietary constituent for infants. The formation of cysteine from methionine occurs via the same transsulfuration pathway as in methionine synthesis in autotrophic organisms. However, the latter use cystathionine γ-synthase and β-lyase while cysteine synthesis in animals uses cystathionine β-synthase and γ-lyase.

Some bacteria degrade L-cysteine or D-cysteine[447a] via the PLP-dependent α,β elimination to form H_2S, pyruvate, and ammonia (reaction *g*, Fig. 24-25, Eq. 14-29). Another catabolic pathway is transamination (Fig. 24-25, rection *h*) to **3-mercaptopyruvate**.[448] The latter compound can be reductively cleaved to pyruvate and sulfide. Cysteine can also be oxidized by NAD^+ and lactate dehydrogenase to 3-mercaptopyruvate. An interesting PLP-dependent β-replacement reaction of cysteine leads to **β-cyanoalanine**, the lathyritic factor (Box 8-E) present in some plants.[449] This reaction also detoxifies the HCN produced during the biosynthesis of ethylene from ACC.

Cysteine and cystine are relatively insoluble and are toxic in excess.[450] Excretion is usually controlled carefully. However, in **cystinuria**, a disease recognized in the medical literature since 1810,[451] there is a greatly increased excretion of cystine and also of the dibasic amino acids.[451,452] As a consequence, stones of cystine develop in the kidneys and bladder. Patients may excrete more than 1 g of cystine in 24 h compared to a normal of 0.05 g, as well as excessive amounts of lysine, arginine, and ornithine. The defect can be fatal, but some persons with the condition remain healthy indefinitely. Cystinuria is one of several human diseases with altered membrane transport and faulty reabsorption of materials from kidney tubules or from the small intestine. Substances are taken up on one side of a cell (e.g., at the bottom of the cell in Fig. 1-6) and discharged into the bloodstream from the other side of the cell. In another rare hereditary condition, **cystinosis**, free cystine accumulates within lysosomes.[453]

2. Cysteine Sulfinate and Taurine

A quantitatively important pathway of cysteine catabolism in animals is oxidation to **cysteine sulfinate** (Fig. 24-25, reaction *i*),[450] a two-step hydroxylation requiring O_2, NADPH or NADH, and Fe^{2+}. Cysteine sulfinic acid can be further oxidized to **cysteic acid** (cysteine sulfonate),[454] which can be decarboxylated to **taurine**. The latter is a component of bile salts (Fig. 22-16) and is one of the most abundant free amino acids in human tissues.[455–457] Its concentration is high in excitable tissues, and it may be a neurotransmitter (Chapter 30). Taurine may have a special function in retinal photoreceptor cells. It is an essential dietary amino acid for cats, who may die of heart failure in its absence,[458] and under some conditions for humans.[459] In many marine invertebrates, teleosts, and amphibians taurine serves as a regulator of osmotic pressure, its concentration decreasing in fresh water and increasing in salt water. A similar role has been suggested for taurine in mammalian hearts. A chronically low concentration of Na^+ leads to increased taurine.[460] Taurine can be reduced to **isethionic acid**

(Fig. 24-25), another component of nervous tissue. Cysteic acid can arise in an alternative way from *O*-acetylserine and sulfite (reaction l, Fig. 24-25), and taurine can also be formed by decarboxylation of cysteine sulfinic acid to **hypotaurine** and oxidation of the latter (reaction *m*). Cysteic acid can be converted to the sulfolipid of chloroplasts (p. 387; Eq. 20-12).

Another route of metabolism for cysteine sulfinic acid is transamination to 3-sulfinylpyruvate, a compound that undergoes ready loss of SO_2 in a reaction analogous to the decarboxylation of oxaloacetate (reaction *o*, Fig. 24-25). This probably represents one of the major routes by which sulfur is removed from organic compounds in the animal body. However, before being excreted the sulfite must be oxidized to sulfate by the Mo-containing sulfite oxidase. The essentiality of sulfite oxidase is evidenced by the severe neurological defect observed in its absence (Chapter 16).

Most of the sulfate generated in the body is excreted unchanged in the urine, but a significant fraction is esterified with oligosaccharides and phenolic compounds. These sulfate esters are formed by sulfo transfer from PAPS (Eq. 17-38).

3. Mercaptopyruvate, Thiosulfate, and Assembly of Iron–Sulfur Centers

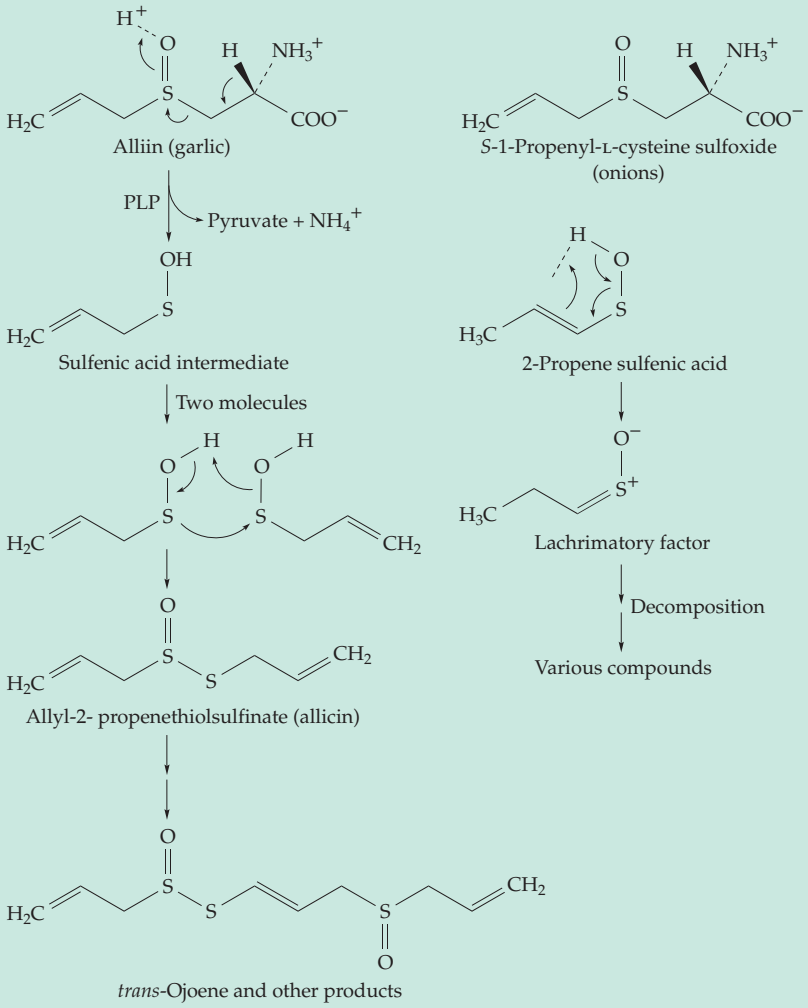
An important property of 3-mercaptopyruvate arises from electron withdrawal by the carbonyl group. This makes the SH group electrophilic and able to be transferred as SH^+, S^0, to a variety of nucleophiles (Eq. 24-44). Thus sulfite yields thiosulfate ($S_2O_3^{2-} + H^+$, Eq. 24-45, step *a*), cyanide yields thiocyanate (Eq. 24-45, step *b*), and cysteine sulfinate yields alanine thiosulfonate.[448,461] The reactions are catalyzed by **mercaptopyruvate sulfurtransferase**, an enzyme very similar to **thiosulfate sulfurtransferase**. The latter is a liver enzyme often called by the traditional

BOX 24-B SULFUR COMPOUNDS OF GARLIC, ONIONS, SKUNKS, ETC.

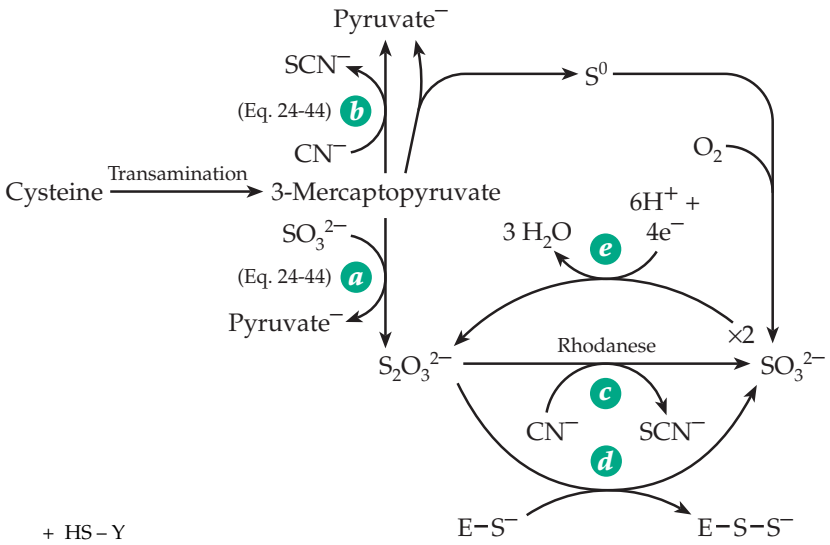
Many familiar odors and tastes come from sulfur-containing compounds. Crushing onions or garlic releases the pyridoxal phosphate-dependent enzyme **alliinase**. In garlic it acts upon the amino acid **alliin** (accompanying scheme) releasing, by β elimination, a sulfenic acid that dimerizes to form **allicin**, a chemically unstable molecule that accounts for the odor of garlic.[a,b] Among the breakdown products of allicin is the nonvolatile **ajoene**, a compound with anticoagulant activity and perhaps accounting for one aspect of the purported medical benefits of garlic. Another is an antibacterial activity.

Onions contain an amino acid that is a positional isomer of alliin. When acted upon by alliinase it produces 2-propene sulfenic acid, which isomerizes to the **lachrimatory factor** that brings tears to the eyes of onion cutters.[c] This, too, decomposes to form many other compounds.[a]

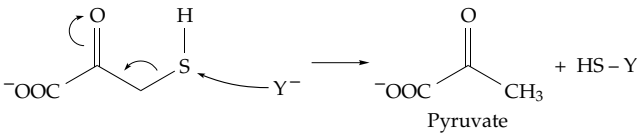
The defensive secretion of the striped skunk has intrigued chemists for over 100 years. The components were shown to contain



name **rhodanese**.[462–463b] It acts by a double displacement mechanism, a thiolate anion in the active site of the enzyme serving as a carrier for the S atom being transferred. Equation 24-46 illustrates the transfer of S^0 from thiosulfate to CN^-, converting that ion to the less toxic thiocyanate (also shown as step *c* in Eq. 24-45). Crystallographic studies show that the negative charge on the thiol and dithiol anions of rhodanese is balanced by the partial



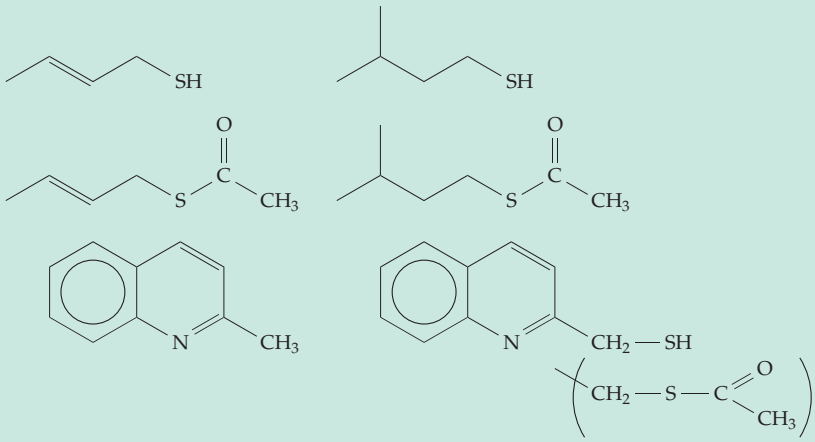
(24-45)



(24-44)

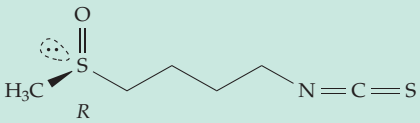
BOX 24-B (continued)

sulfur, and one was incorrectly identified and was long accepted as being butyl mercaptan. Modern capillary gas chromatography by Wood[d] has revealed the presence of seven major components, with the indicated structures. Two are simple volatile mercaptans, but three are thioacetates which hydrolyze in water only slowly, releasing smell for days or weeks from sprayed animals. Washing with mildly basic soap hastens the hydrolysis.



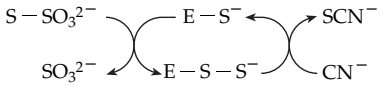
Many readers (~40%) may be aware that after eating asparagus a strong odor appears in their urine. These genetic "stinkers" secrete *S*-methyl thioacrylate, and related compounds, derived from a plant constituent.[e]

The sulfur compound sulfuraphane, extracted from fresh broccoli, has received attention in recent years because of its strong action in inducing synthesis of quinone reductase and glutathione *S*-transferases that help detoxify xenobiotics and may have significant anticancer activity.[f]



Sulforaphane;
1-isothiocyanate-(4*R*)-(methylsulfinyl)butane

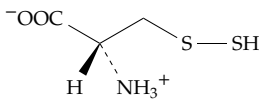
[a] Block, E. (1985) *Sci. Am.* **252**(Mar), 114–119
[b] Jain, M. K., and Apitz-Castro, R. (1987) *Trends Biochem. Sci.* **12**, 252–254
[c] Imai, S., Tsuge, N., Tomotake, M., Nagatome, Y., Sawada, H., Nagata, T., and Kumagai, H. (2002) *Nature (London)* **419**, 685
[d] Wood, W. F. (1990) *J. Chem. Ecol.* **16**, 2057
[e] White, R. H. (1975) *Science* **189**, 810
[f] Zhang, Y., Talalay, P., Cho, C.-G., and Posner, G. H. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2399–2403



(24-46)

positive charges at the N termini of two helices (see Fig. 2-20A) and by hydrogen bonds to protons of several side chains.[462] This evidently explains how the negative thiosulfate anion can react with another anion, E−S⁻. Another interesting feature of this enzyme is that the monomer has a nearly perfect twofold axis of symmetry with respect to the protein folding pattern. However, the symmetry is lacking in the sequence and only one-half of the molecule contains an active site.[462]

Yet another enzyme able to release or transfer sulfur in the S^0 oxidation state is the PLP-dependent **cysteine desulfurase** that is encoded by the *nifS* gene of the nitrogenase gene cluster shown in Fig. 24-4. This enzyme releases S^0 from cysteine with formation of alanine[464–466] as is shown in Eq. 14-34 for release of Se^0 from selenocysteine. As with rhodanese an active site cysteine accepts the departing S^0 of cysteine to form an enzyme-bound persulfide. This protein may in turn transfer the sulfur into the forming Fe–S or Fe–S–Mo clusters.[464] Three PLP-dependent persulfide-forming sulfurtransferases related to the NifS protein have been found in *E. coli*. Similar enzymes are present in other organisms.[466a–d] A sulfur atom may be transferred from the bound persulfide anion to acceptor proteins involved in metal cluster formation. Some members of the nifS-like family act on cystine to release free thiocysteine (cysteine peroxide), which may also serve as a sulfur atom donor.[466e]

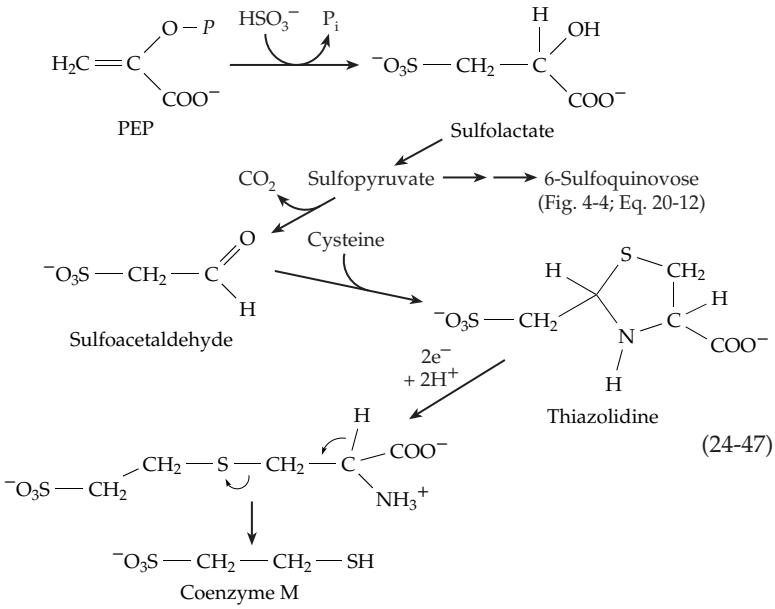


Thiocysteine (cysteine persulfide)

Thiocysteine can also arise in a similar manner through action of cystathionine β lyase on cystine. Thiocysteine is eliminated with production of pyruvate and ammonia from the rest of the cystine molecule.[467] One of the nifS-like proteins of *E. coli* is thought to transfer a selenium atom from selenocysteine (pp. 823–827) into **selenophosphate**.[466a,f] The latter can be formed by transfer of a phospho group from ATP to selenide HSe^-. The other products of ATP cleavage are AMP and P_i. Reduction of Se^0 to HSe^- is presumably necessary.

Several additional proteins identified as necessary for metal-sulfide cluster formation are present in bacteria and in eukaryotes, both in the cytosol and in mitochondria. They may serve as intermediate sulfur carriers, as scaffolds or templates for cluster formation, or for insertion of intact Fe–S, Fe–S–Mo, or other types of clusters into proteins[468–473f] and into 2-selenouridine[473g] (see also p. 1617). Sulfurtransferases are also thought to be involved in insertion of sulfur atoms into organic molecules such as biotin, lipoic acid, or methanopterin.[474]

A reaction that is ordinarily of minor consequence in the animal body but which may be enhanced by a deficiency of sulfite oxidase is the reductive coupling of two molecules of sulfite to form thiosulfate (Eq. 24-45, step *e*). Several organic hydrodisulfide derivatives such as thiocysteine, thioglutathione, and thiotaurine occur in animals in small amounts. Another biosynthetic pathway, outlined in Eq. 24-47 converts sulfite and PEP into coenzyme M (Fig. 15-22).[475,475a] This cofactor is needed not only for methane formation (Fig. 15-2) but also for utilization of alkenes by soil bacteria.[475b]



(24-47)

References

1. Conn, E. E., Stumpf, P. K., Bruening, G., and Doi, R. H. (1987) *Outlines of Biochemistry*, 5th ed., Wiley, New York
2. Smil, V. (1997) *Sci. Am.* **277**(Jul), 76–81
3. Hardy, R. W. F., and Havelka, U. D. (1975) *Science* **188**, 633–643
4. Winogradsky, S. (1893) *C.R. Acad. Sci.* **116**, 1385–1388
5. Orme-Johnson, W. H. (1985) *Ann. Rev. Biophys. Biophys. Chem.* **14**, 419–459
6. Appleby, C. A., Nicola, N. A., Hurrell, J. G. R., and Leach, S. J. (1975) *Biochemistry* **14**, 4444–4450
7. Lee, H. C., Wittenberg, J. B., and Peisach, J. (1993) *Biochemistry* **32**, 11500–11506
8. Stouggard, J., Petersen, T. E., and Marcker, K. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5754–5757
9. Stacey, G., Burris, R. H., and Evans, H. J., eds. (1992) *Biological Nitrogen Fixation*, Chapman and Hall, London

9a. Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., and Carpenter, E. J. (1997) *Science* **276**, 1221–1229

10. Appleby, C. A. (1984) *Ann. Rev. Plant Physiol.* **35**, 443–478

10a. Lou, J., Moshiri, F., Johnson, M. K., Lafferty, M. E., Sorkin, D. L., Miller, A.-F., and Maier, R. J. (1999) *Biochemistry* **38**, 5563–5571

11. Lowe, D. J., Fisher, K., Thorneley, R. N. F., Vaughn, S. A., and Burgess, B. K. (1989) *Biochemistry* **28**, 8460–8466
12. Rasche, M. E., and Seefeldt, L. C. (1997) *Biochemistry* **36**, 8574–8585
13. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) *Science* **257**, 1653–1659
14. Kim, J., and Rees, D. C. (1994) *Biochemistry* **33**, 389–397
15. Howard, J. B., and Rees, D. C. (1994) *Ann. Rev. Biochem.* **63**, 235–264
16. Peters, J. W., Fisher, K., and Dean, D. R. (1995) *Ann. Rev. Microbiol.* **49**, 335–366
17. Kim, J., and Rees, D. C. (1992) *Nature (London)* **360**, 553–560
18. Kim, J., Woo, D., and Rees, D. C. (1993) *Biochemistry* **32**, 7104–7115
19. Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B., and Rees, D. C. (1997) *Nature (London)* **387**, 370–376

19a. Mayer, S. M., Lawson, D. M., Gormal, C. A., Roe, S. M., and Smith, B. E. (1999) *J. Mol. Biol.* **292**, 871–891

20. Hoover, T. R., Imperial, J., Ludden, P. W., and Shah, V. K. (1989) *Biochemistry* **28**, 2768–2771
21. Hoover, T. R., Imperial, J., Liang, J., Ludden, P. W., and Shah, V. K. (1988) *Biochemistry* **27**, 3647–3652
22. Imperial, J., Hoover, T. R., Madden, M. S., Ludden, P. W., and Shah, V. K. (1989) *Biochemistry* **28**, 7796–7799

22a. Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B., and Rees, D. C. (2002) *Science* **297**, 1696 – 1700

23. Peters, J. W., Stowell, M. H. B., Soltis, S. M., Finnegan, M. G., Johnson, M. K., and Rees, D. C. (1997) *Biochemistry* **36**, 1181–1187
24. McLean, P. A., Wink, D. A., Chapman, S. K., Hickman, A. B., McKillop, D. M., and Orme-Johnson, W. H. (1989) *Biochemistry* **28**, 9402–9406
25. Jensen, B. B., and Burris, R. H. (1985) *Biochemistry* **24**, 1141–1147
26. Liang, J., and Burris, R. H. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9446–9450
27. Kim, C.-H., Newton, W. E., and Dean, D. R. (1995) *Biochemistry* **34**, 2798–2808
28. Cleland, W. W., Guth, J., and Burris, R. H. (1983) *Biochemistry* **22**, 5111–5122
29. Wahl, R. C., and Orme-Johnson, W. H. (1987) *J. Biol. Chem.* **262**, 10489–10496
30. Hoover, D. M., and Ludwig, M. L. (1997) *Protein Sci.* **6**, 2525–2537
31. Yoch, D. C. (1972) *Biochem. Biophys. Res. Commun.* **49**, 335–342
32. Carter, K. R., Rawlings, J., Orme-Johnson, W. H., Becker, R. R., and Evans, H. J. (1980) *J. Biol. Chem.* **255**, 4213–4223
33. Renner, K. A., and Howard, J. B. (1996) *Biochemistry* **35**, 5353–5358
34. Duyvis, M. G., Wassink, H., and Haaker, H. (1996) *J. Biol. Chem.* **271**, 29632–29636
35. Lanzilotta, W. N., Fisher, K., and Seefeldt, L. C. (1997) *J. Biol. Chem.* **272**, 4157–4165

35a. Chan, J. M., Wu, W., Dean, D. R., and Seefeldt, L. C. (2000) *Biochemistry* **39**, 7221–7228

35b. Jang, S. B., Seefeldt, L. C., and Peters, J. W. (2000) *Biochemistry* **39**, 14745–14752

35c. Clarke, T. A., Maritano, S., and Eady, R. R. (2000) *Biochemistry* **39**, 11434–11440

35d. Nyborg, A. C., Johnson, J. L., Gunn, A., and Watt, G. D. (2000) *J. Biol. Chem.* **275**, 39307–39312

36. Lanzilotta, W. N., Fisher, K., and Seefeldt, L. C. (1996) *Biochemistry* **35**, 7188–7196
37. Lanzilotta, W. N., and Seefeldt, L. C. (1996) *Biochemistry* **35**, 16770–16776

37a. Chiu, H.-J., Peters, J. W., Lanzilotta, W. N., Ryle, M. J., Seefeldt, L. C., Howard, J. B., and Rees, D. C. (2001) *Biochemistry* **40**, 641–650

38. Lanzilotta, W. N., and Seefeldt, L. C. (1997) *Biochemistry* **36**, 12976–12983

38a. Jang, S. B., Seefeldt, L. C., and Peters, J. W. (2000) *Biochemistry* **39**, 641–648

39. Grossman, J. G., Hasnain, S. S., Yousafzai, F. K., Smith, B. E., and Eady, R. R. (1997) *J. Mol. Biol.* **266**, 642–648
40. Ma, L., Brosius, M. A., and Burgess, B. K. (1996) *J. Biol. Chem.* **271**, 10528–10532

40a. Chan, J. M., Christiansen, J., Dean, D. R., and Seefeldt, L. C. (1999) *Biochemistry* **38**, 5779–5785

41. Christiansen, J., Tittsworth, R. C., Hales, B. J., and Cramer, S. P. (1995) *J. Am. Chem. Soc.* **117**, 10017–10024
42. Venters, R. A., Nelson, M. J., McLean, P. A., True, A. E., Levy, M. A., Hoffman, B. M., and Orme-Johnson, W. H. (1986) *J. Am. Chem. Soc.* **108**, 3487–3498
43. Shah, V. K., Ugalde, R. A., Imperial, J., and Brill, W. J. (1985) *J. Biol. Chem.* **260**, 3891–3894
44. Stiefel, E. I. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 988–992

44a. Rod, T. H., and Noskov, J. K. (2000) *J. Am. Chem. Soc.* **122**, 12751–12763

44b. Thorneley, R. N., and Lowe, D. J. ((1984) *Biochem. J.* **224**, 887 – 894

44c. Durrant, M. C. (2002) *Biochemistry* **41**, 13934 – 13945; 13946 – 13955

45. Dilworth, M. J., Fisher, K., Kim, C.-H., and Newton, W. E. (1998) *Biochemistry* **37**, 17495–17505

45a. Fisher, K., Dilworth, M. J., Kim, C.-H., and Newton, W. E. (2000) *Biochemistry* **39**, 2970–2979

45b. Fisher, K., Dilworth, M. J., Kim, C.-H., and Newton, W. E. (2000) *Biochemistry* **39**, 10855–10865

45c. Christiansen, J., Seefeldt, L. C., and Dean, D. R. (2000) *J. Biol. Chem.* **275**, 36104–36107

45d. Sorlie, M., Christiansen, J., Lemon, B. J., Peters, J. W., Dean, D. R., and Hales, B. J. (2001) *Biochemistry* **40**, 1540–1549

46. Palermo, R. E., Singh, R., Bashkin, J. K., and Holm, R. H. (1984) *J. Am. Chem. Soc.* **106**, 2600–2612
47. Leigh, G. J. (1995) *Science* **268**, 827–828
48. Ferguson, R., Solari, E., Floriani, C., Osella, D., Ravera, M., Re, N., Chiesi-Villa, A., and Rizzoli, C. (1997) *J. Am. Chem. Soc.* **119**, 10104–10115
49. Nishibayashi, Y., Iwai, S., and Hidai, M. (1998) *J. Am. Chem. Soc.* **120**, 10559–10560

49a. Verma, A. K., and Lee, S. C. (1999) *J. Am. Chem. Soc.* **121**, 10838–10839

50. Laplaza, C. E., Johnson, A. R., and Cummins, C. C. (1996) *J. Am. Chem. Soc.* **118**, 709–710
51. Laplaza, C. E., and Cummins, C. C. (1995) *Science* **268**, 861–863
52. Ow, D. W., and Ausubel, F. M. (1983) *Nature (London)* **301**, 307–313
53. Thiel, T., Lyons, E. M., Erker, J. C., and Ernst, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9358–9362
54. Pulakat, L., Hausman, B. S., Lei, S., and Gavini, N. (1996) *J. Biol. Chem.* **271**, 1884–1889

54a. Yuvaniyama, P., Agar, J. N., Cash, V. L., Johnson, M. K., and Dean, D. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 599–604

54b. Ribbe, M. W., Bursey, E. H., and Burgess, B. K. (2000) *J. Biol. Chem.* **275**, 17631–17638

54c. Ribbe, M. W., and Burgess, B. K. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5521–5525

55. Shah, V. K., Imperial, J., Ugalde, R. A., Ludden, P. W., and Brill, W. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1636–1640

55a. Rangaraj, P., Rüttimann-Johnson, C., Shah, V. K., and Ludden, P. W. (2001) *J. Biol. Chem.* **276**, 15968–15974

56. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) *J. Bacteriol.* **171**, 1017–1027
57. Dilworth, M., and Glenn, A. (1984) *Trends Biochem. Sci.* **9**, 519–523
58. Marx, J. L. (1985) *Science* **230**, 157–158
59. Albrecht, C., Geurts, R., and Bisseling, T. (1999) *EMBO J.* **18**, 281–288
60. Fisher, R. F., and Long, S. R. (1992) *Nature (London)* **357**, 655–660
61. Vijn, I., das Neves, L., van Kammen, A., Franssen, H., and Bisseling, T. (1993) *Science* **260**, 1764–1765
62. Cedergren, R. A., Lee, J., Ross, K. L., and Hollingsworth, R. I. (1995) *Biochemistry* **34**, 4467–4477
63. Dunn, S. M., Moody, P. C. E., Downie, J. A., and Shaw, W. V. (1996) *Protein Sci.* **5**, 538–541
64. Spaink, H. P., Sheeley, D. M., van Brussel, A. A. N., Glushka, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N., and Lugtenberg, B. J. J. (1991) *Nature (London)* **354**, 125–130
65. Mergaert, P., D'Haeze, W., Geelen, D., Promé, D., Van Montagu, M., Geremia, R., Promé, J.-C., and Holsters, M. (1995) *J. Biol. Chem.* **270**, 29217–29223
66. Schultze, M., Staehelin, C., Röhrig, H., John, M., Schmidt, J., Kondorosi, E., Schell, J., and Kondorosi, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2706–2709
67. Fisher, R. F., and Long, S. R. (1993) *J. Mol. Biol.* **233**, 336–348
68. Monson, E. K., Ditta, G. S., and Helinski, D. R. (1995) *J. Biol. Chem.* **270**, 5243–5250
69. Soupène, E., Foussard, M., Boistard, P., Truchet, G., and Batut, J. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3759–3763

69a. Miyatake, H., Mukai, M., Park, S.-Y., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T., and Shiro, Y. (2000) *J. Mol. Biol.* **301**, 415–431

70. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A., and Perre, X. (1997) *Nature (London)* **387**, 394–401

References

71. Downie, A. (1997) *Nature (London)* **387**, 352–354
72. Minchin, F. (1986) *Nature (London)* **320**, 483–484
73. Berry, A. M., Harriott, O. T., Moreau, R. A., Osman, S. F., Benson, D. R., and Jones, A. D. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6091–6094
74. Bishop, P. E. (1986) *Trends Biochem. Sci.* **11**, 225–227
75. Blanchard, C. Z., and Hales, B. J. (1996) *Biochemistry* **35**, 472–478
76. Chatterjee, R., Ludden, P. W., and Shah, V. K. (1997) *J. Biol. Chem.* **272**, 3758–3765
76a. Rüttimann-Johnson, C., Rangaraj, P., Shah, V. K., and Ludden, P. W. (2001) *J. Biol. Chem.* **276**, 4522–4526
77. Pau, R. N. (1989) *Trends Biochem. Sci.* **14**, 183–186
78. Ribbe, M., Gadkari, D., and Meyer, O. (1997) *J. Biol. Chem.* **272**, 26627–26633
79. DiSpirito, A. A., and Hooper, A. B. (1986) *J. Biol. Chem.* **261**, 10534–10537
79a. Kanamaru, K., Wang, R., Su, W., and Crawford, N. M. (1999) *J. Biol. Chem.* **274**, 4160–4165
80. Campbell, W. H., and Kinghorn, J. R. (1990) *Trends Biochem. Sci.* **15**, 315–319
81. Lu, G., Lindqvist, Y., Schneider, G., Dwivedi, U., and Campbell, W. (1995) *J. Mol. Biol.* **248**, 931–948
82. Garde, J., Kinghorn, J. R., and Tomsett, A. B. (1995) *J. Biol. Chem.* **270**, 6644–6650
83. Hyde, G. E., Crawford, N. M., and Campbell, W. H. (1991) *J. Biol. Chem.* **266**, 23542–23547
84. Ratnam, K., Shiraishi, N., Campbell, W. H., and Hille, R. (1995) *J. Biol. Chem.* **270**, 24067–24072
85. Cramer, S. P., Solomonson, L. P., Adams, M. W. W., and Mortenson, L. E. (1984) *J. Am. Chem. Soc.* **106**, 1467–1471
85a. Skipper, L., Campbell, W. H., Mertens, J. A., and Lowe, D. J. (2001) *J. Biol. Chem.* **276**, 26995–27002
86. Lin, J. T., and Stewart, V. (1996) *J. Mol. Biol.* **256**, 423–435
86a. Butler, C. S., Charnock, J. M., Bennett, B., Sears, H. J., Reilly, A. J., Ferguson, S. J., Garner, C. D., Lowe, D. J., Thomson, A. J., Berks, B. C., and Richardson, D. J. (1999) *Biochemistry* **38**, 9000–9012
87. Augier, V., Guigliarelli, B., Asso, M., Bertrand, P., Frixon, C., Giordano, G., Chippaux, M., and Blasco, F. (1993) *Biochemistry* **32**, 2013–2023
87a. Zhao, Y., Lukoyanov, D. A., Toropov, Y. V., Wu, K., Shapleigh, J. P., and Scholes, C. P. (2002) *Biochemistry* **41**, 7464–7474
88. Bentley, R. (1985) *Trends Biochem. Sci.* **10**, 171–174
89. Cooper, A. J. L., Nieves, E., Coleman, A. E., Filc-DeRicco, S., and Gelbard, A. S. (1987) *J. Biol. Chem.* **262**, 1073–1080
90. Cooper, A. J. L., Nieves, E., Rosenspire, K. C., Filc-DeRicco, S., Gelbard, A. S., and Brusilow, S. W. (1988) *J. Biol. Chem.* **263**, 12268–12273
91. Chatham, J. C., Forder, J. R., Glickson, J. D., and Chance, E. M. (1995) *J. Biol. Chem.* **270**, 7999–8008
92. Martin, G., Chauvin, M.-F., and Baverel, G. (1997) *J. Biol. Chem.* **272**, 4717–4728
93. Chauvin, M.-F., Mégnin-Chanet, F., Martin, G., Lhoste, J.-M., and Baverel, G. (1994) *J. Biol. Chem.* **269**, 26025–26033
94. Ikeda, T. P., Shauger, A. E., and Kustu, S. (1996) *J. Mol. Biol.* **259**, 589–607
95. Zhalkin, H. (1993) *Adv. Enzymol.* **66**, 203–309
96. Boehlein, S. K., Richards, N. G. J., Walworth, E. S., and Schuster, S. M. (1994) *J. Biol. Chem.* **269**, 26789–26795
97. Muchmore, C. R. A., Krahn, J. M., Kim, J. H., Zalkin, H., and Smith, J. L. (1998) *Protein Sci.* **7**, 39–51
97a. Raushel, F. M., Thoden, J. B., and Holden, H. M. (1999) *Biochemistry* **38**, 7891–7899
97b. Bera, A. K., Smith, J. L., and Zalkin, H. (2000) *J. Biol. Chem.* **275**, 7975–7979
97c. Teplyakov, A., Obmolova, G., Badet, B., and Badet-Denisot, M.-A. (2001) *J. Mol. Biol.* **313**, 1093–1102
97d. Chittur, S. V., Klem, T. J., Shafer, C. M., and Davisson, V. J. (2001) *Biochemistry* **40**, 876–887
98. Boehlein, S. K., Stewart, J. D., Walworth, E. S., Thirumoorthy, R., Richards, N. G. J., and Schuster, S. M. (1998) *Biochemistry* **37**, 13230–13238
98a. Larsen, T. M., Boehlein, S. K., Schuster, S. M., Richards, N. G. J., Thoden, J. B., Holden, H. M., and Rayment, I. (1999) *Biochemistry* **38**, 16146–16157
98b. Min, B., Pelaschier, J. T., Graham, D. E., Tumbula-Hansen, D., and Söll, D. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2678–2683
99. Raghunathan, G., Miles, H. T., and Sasisekharan, V. (1993) *Biochemistry* **32**, 455–462
100. Winter, H. C., and Dekker, E. E. (1986) *J. Biol. Chem.* **261**, 11189–11193
101. Kohl, D. H., Schubert, K. R., Carter, M. B., Hagedorn, C. H., and Shearer, G. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2036–2040
102. Cooper, A. J. L., and Meister, A. (1985) in *Transaminases* (Christen, P., and Metzler, D. E., eds), pp. 533–563, Wiley, New York
103. Sibson, N. R., Dhankhar, A., Mason, G. F., Behar, K. L., Rothman, D. L., and Shulman, R. G. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2699–2704
104. Liaw, S.-H., Kuo, I., and Eisenberg, D. (1995) *Protein Sci.* **4**, 2358–2365
105. Kagan, Z. A., Kretovich, V. L., and Polyakov, V. A. (1966) *Biokhim.* **31**, 355–364
106. Baker, P. J., Waugh, M. L., Wang, X.-G., Stillman, T. J., Turnbull, A. P., Engel, P. C., and Rice, D. W. (1997) *Biochemistry* **36**, 16109–16115
107. Peters, J.-M., Harris, J. R., and Finley, D., eds. (1998) *Ubiquitin and the Biology of the Cell*, Plenum, New York
107a. Katagiri, M., and Nakamura, M. (1999) *Biochem. Edu.* **27**, 83–85
107b. Katagiri, M., and Nakamura, M. (2002) *IUBMB Life* **53**, 125–129
107c. Wipf, D., Ludewig, U., Tegeder, M., Rentsch, D., Koch, W., and Frommer, W. B. (2002) *Trends Biochem. Sci.* **27**, 139–147
108. Rajagopalan, S., Wan, D.-F., Habib, G. M., Sepulveda, A. R., McLeod, M. R., Lebovitz, R. M., and Lieberman, M. W. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6179–6183
109. Harding, C. O., Williams, P., Wagner, E., Chang, D. S., Wild, K., Colwell, R. E., and Wolff, J. A. (1997) *J. Biol. Chem.* **272**, 12560–12567
109a. Stein, R. L., DeCicco, C., Nelson, D., and Thomas, B. (2001) *Biochemistry* **40**, 5804–5811
110. Guo-jie, Breslow, E., and Meister, A. (1996) *J. Biol. Chem.* **271**, 32293–32300
111. Van Der Werf, P., Griffith, O. W., and Meister, A. (1975) *J. Biol. Chem.* **250**, 6686–6692
112. Reitzer, L. J., and Magasanik, B. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 302–318, Am. Soc. for Microbiology, Washington,DC
113. Kanamori, K., Weiss, R. L., and Roberts, J. D. (1988) *J. Biol. Chem.* **263**, 2817–2823
114. Knapp, S., de Vos, W. M., Rice, D., and Ladenstein, R. (1997) *J. Mol. Biol.* **267**, 916–932
115. Wang, X.-G., and Engel, P. C. (1995) *Biochemistry* **34**, 11417–11422
115a. Herrero-Yraola, A., Bakhit, S. M. A., Franke, P., Weise, C., Schweiger, M., Jorcke, D., and Ziegler, M. (2001) *EMBO J.* **20**, 2404–2412
116. Melo-Oliveira, R., Oliveira, I. C., and Coruzzi, G. M. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4718–4723
117. Vanoni, M. A., Edmondson, D. E., Zanetti, G., and Curti, B. (1992) *Biochemistry* **31**, 4613–4623
118. Vanoni, M. A., Fischer, F., Ravasio, S., Verzotti, E., Edmondson, D. E., Hagen, W. R., Zanetti, G., and Curti, B. (1998) *Biochemistry* **37**, 1828–1838
118a. Morandi, P., Valzasina, B., Colombo, C., Curti, B., and Vanoni, M. A. (2000) *Biochemistry* **39**, 727–735
119. Sakakibara, H., Watanabe, M., Hase, T., and Sugiyama, T. (1991) *J. Biol. Chem.* **266**, 2028–2035
120. Knaff, D. B., Hirasawa, M., Ameyibor, E., Fu, W., and Johnson, M. K. (1991) *J. Biol. Chem.* **266**, 15080–15084
121. Ho, H.-T., Falk, P. J., Ervin, K. M., Krishnan, B. S., Discotto, L. F., Dougherty, T. J., and Pucci, M. J. (1995) *Biochemistry* **34**, 2464–2470
122. Gallo, K. A., and Knowles, J. R. (1993) *Biochemistry* **32**, 3981–3990
122a. Glavas, S., and Tanner, M. E. (2001) *Biochemistry* **40**, 6199–6204
123. D'Aniello, A., D'Onofrio, G., Pischetola, M., D'Aniello, G., Vetere, A., Petrucelli, L., and Fisher, G. H. (1993) *J. Biol. Chem.* **268**, 26941–26949
124. Almassy, R. J., Janson, C. A., Hamlin, R., Xuong, N.-H., and Eisenberg, D. (1986) *Nature (London)* **323**, 304–309
125. Liaw, S.-H., and Eisenberg, D. (1994) *Biochemistry* **33**, 675–681
126. Mura, U., Chock, P. B., and Stadtman, E. R. (1981) *J. Biol. Chem.* **256**, 13022–13029
126a. Stadtman, E. R. (2001) *J. Biol. Chem.* **276**, 44357–44364
127. Garcia, E., and Rhee, S. G. (1983) *J. Biol. Chem.* **258**, 2246–2253
128. Son, H. S., and Rhee, S. G. (1987) *J. Biol. Chem.* **262**, 8690–8695
129. Jiang, P., Peliska, J. A., and Ninfa, A. J. (1998) *Biochemistry* **37**, 12802–12810
130. Liaw, S.-H., Pan, C., and Eisenberg, D. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4996–5000
131. Jiang, P., Peliska, J. A., and Ninfa, A. J. (1998) *Biochemistry* **37**, 12795–12801
132. Flashner, Y., Weiss, D. S., Keener, J., and Kustu, S. (1995) *J. Mol. Biol.* **249**, 700–713
133. Rosenfeld, S. A., and Brenchley, J. E. (1983) in *Amino Acids, Biosynthesis and Genetic Regulation* (Herrmann, K. M., and Somerville, R. L., eds), pp. 1–17, Addison-Wesley, Reading, Massachusetts
134. Buchanan, J. M. (1973) *Adv. Enzymol.* **38**, 91–183
134a. Gill, H. S., and Eisenberg, D. (2001) *Biochemistry* **40**, 1903–1912
135. Neu, J., Shenoy, V., and Chakrabarti, R. (1996) *FASEB J.* **10**, 829–837
136. Jakob, C. G., Lewinski, K., LaCount, M. W., Roberts, J., and Lebioda, L. (1997) *Biochemistry* **36**, 923–931
137. Watford, M. (1993) *FASEB J.* **7**, 1468–1474
138. Snedden, W. A., Koutsia, N., Baum, G., and Fromm, H. (1996) *J. Biol. Chem.* **271**, 4148–4153
138a. Locher, K. P., Hans, M., Yeh, A. P., Schmid, B., Buckel, W., and Rees, D. C. (2001) *J. Mol. Biol.* **307**, 297–308
138b. Chih, H.-W., and Marsh, E. N. G. (2001) *Biochemistry* **40**, 13060–13067
139. Smith, M. A., King, P. J., and Grimm, B. (1998) *Biochemistry* **37**, 319–329
140. Adams, E., and Frank, L. (1980) *Ann. Rev. Biochem.* **49**, 1005–1061

References

141. Leisinger, T. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 345–351, Am. Soc. for Microbiology, Washington,DC
142. Zhang, C.-s, Lu, Q., and Verma, D. P. S. (1995) *J. Biol. Chem.* **270**, 20491–20496
143. García-Ríos, M., Fujita, T., LaRosa, P. C., Locy, R. D., Clithero, J. M., Bressan, R. A., and Csonka, L. N. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8249–8254
144. Mestichelli, L. J. J., Gupta, R. N., and Spenser, I. D. (1979) *J. Biol. Chem.* **254**, 640–647
145. Ashford, D., and Neuberger, A. (1980) *Trends Biochem. Sci.* **5**, 245–248
145a. Becker, D. F., and Thomas, E. A. (2001) *Biochemistry* **40**, 4714–4721
145b. Farrant, R. D., Walker, V., Mills, G. A., Mellor, J. M., and Langley, G. J. (2001) *J. Biol. Chem.* **276**, 15107–15116
146. Small, W. C., and Jones, M. E. (1990) *J. Biol. Chem.* **265**, 18668–18672
147. Hu, C.-aA., Lin, W.-W., and Valle, D. (1996) *J. Biol. Chem.* **271**, 9795–9800
148. Phang, J. M., Yeh, G. C., and Scriver, C. R. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1125–1146, McGraw-Hill, New York
148a. Farrant, R. D., Walker, V., Mills, G. A., Mellor, J. M., and Langley, G. J. (2000) *J. Biol. Chem.* **276**, 15107 – 15116
149. Gupta, S. C., and Dekker, E. E. (1980) *J. Biol. Chem.* **255**, 1107–1112
150. Sletten, K., Aakesson, I., and Alvsaker, J. O. (1971) *Nature New Biol.* **231**, 118–119
151. Akers, H. A., and Dromgoole, E. V. (1982) *Trends Biochem. Sci.* **7**, 156–157
152. Umbarger, H. E. (1978) *Ann. Rev. Biochem.* **47**, 533–606
153. Davis, R. H., and Weiss, R. L. (1988) *Trends Biochem. Sci.* **13**, 101–104
154. Shah, S. A., Shen, B. W., and Brünger, A. T. (1997) *Structure* **5**, 1067–1075
155. Valle, D., and Simell, O. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1147–1185, McGraw-Hill, New York
156. Glansdorff, N. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 321–339, Am. Soc. for Microbiology, Washington, DC
157. Krebs, H. A., and Henseleit, K. (1932) *Hoppe-Seyler's Z. Physiol. Chem.* **210**, 33–66
158. Krebs, H. A. (1982) *Trends Biochem. Sci.* **7**, 76–78
159. Wakabayashi, Y., Yamada, E., Yoshida, T., and Takahashi, H. (1994) *J. Biol. Chem.* **269**, 32667–32671
160. Gessert, S. F., Kim, J. H., Nargang, F. E., and Weiss, R. L. (1994) *J. Biol. Chem.* **269**, 8189–8203
161. Watford, M. (1989) *Trends Biochem. Sci.* **14**, 313–314
162. Brusilow, S. W., and Horwich, A. L. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1187–1232, McGraw-Hill, New York
163. Lim, A. L., and Powers-Lee, S. G. (1996) *J. Biol. Chem.* **271**, 11400–11409
164. Guy, H. I., and Evans, D. R. (1996) *J. Biol. Chem.* **271**, 13762–13769
165. McCudden, C. R., and Powers-Lee, S. G. (1996) *J. Biol. Chem.* **271**, 18285–18294
166. Hong, J., Salo, W. L., Lusty, C. J., and Anderson, P. M. (1994) *J. Mol. Biol.* **243**, 131–140
167. Devaney, M. A., and Powers-Lee, S. G. (1984) *J. Biol. Chem.* **259**, 703–706
168. Anderson, P. M. (1981) *J. Biol. Chem.* **256**, 12228–12238
169. Stapleton, M. A., Javid-Majd, F., Harmon, M. F., Hanks, B. A., Grahmann, J. L., Mullins, L. S., and Raushel, F. M. (1996) *Biochemistry* **35**, 14352–14361
170. Miles, B. W., Banzon, J. A., and Raushel, F. M. (1998) *Biochemistry* **37**, 16773–16779
171. Powers, S. G., and Meister, A. (1978) *J. Biol. Chem.* **253**, 1258–1265
172. Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M., and Rayment, I. (1997) *Biochemistry* **36**, 6305–6316
173. Thoden, J. B., Miran, S. G., Phillips, J. C., Howard, A. J., Rauschel, F. M., and Holden, H. M. (1998) *Biochemistry* **37**, 8825–8831
173a. Thoden, J. B., Huang, X., Raushel, F. M., and Holden, H. M. (1999) *Biochemistry* **38**, 16158–16166
173b. Kim, J., Howell, S., Huang, X., and Raushel, F. M. (2002) *Biochemistry* **41**, 12575 – 12581
173c. Fresquet, V., Mora, P., Rochera, L., Ramón-Maiques, S., Rubio, V., and Cervera, J. (2000) *J. Mol. Biol.* **299**, 979–991
174. Shi, D., Morizono, H., Ha, Y., Aoyagi, M., Tuchman, M., and Allewell, N. M. (1998) *J. Biol. Chem.* **273**, 34247–34254
175. Tricot, C., Villeret, V., Sainz, G., Dideberg, O., and Stalon, V. (1998) *J. Mol. Biol.* **283**, 695–704
175a. Langley, D. B., Templeton, M. D., Fields, B. A., Mitchell, R. E., and Collyer, C. A. (2000) *J. Biol. Chem.* **275**, 20012–20019
176. Rosenberg, L. E., Kalousek, F., and Orsulak, M. D. (1983) *Science* **222**, 426–428
177. Maestri, N. E., Brusilow, S. W., Clissold, D. B., and Bassett, S. S. (1996) *N. Engl. J. Med.* **335**, 855–859
177a. Sampaleanu, L. M., Vallée, F., Thompson, G. D., and Howell, P. L. (2001) *Biochemistry* **40**, 15570–15580
178. Turner, M. A., Simpson, A., McInnes, R. R., and Howell, P. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9063–9068
178a. Yu, B., Thompson, G. D., Yip, P., Howell, P. L., and Davidson, A. R. (2001) *Biochemistry* **40**, 15581–15590
179. Garrard, L. J., Mathis, J. M., and Raushel, F. M. (1983) *Biochemistry* **22**, 3729–3735
180. Kanyo, Z. F., Scolnick, L. R., Ash, D. E., and Christianson, D. W. (1996) *Nature (London)* **383**, 554–557
181. Marathe, S., Yu, Y. G., Turner, G. E., Palmier, C., and Weiss, R. L. (1998) *J. Biol. Chem.* **273**, 29776–29785
182. Scolnick, L. R., Kanyo, Z. F., Cavalli, R. C., Ash, D. E., and Christianson, D. W. (1997) *Biochemistry* **36**, 10558–10565
183. Smith, I. (1981) *Nature (London)* **291**, 378–380
184. Atkinson, D. E., and Bourke, E. (1984) *Trends Biochem. Sci.* **9**, 297–300
185. Gargan, R., Rapoport, G., and Débarbouillé, M. (1995) *J. Mol. Biol.* **249**, 843–856
186. Jann, A., Stalon, V., Vander Wauven, C., Leisinger, T., and Haas, D. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4937–4941
187. Weickmann, J. L., Himmel, M. E., Squire, P. G., and Fahrney, D. E. (1978) *J. Biol. Chem.* **253**, 6010–6015
188. Abdelal, A. T. (1979) *Ann. Rev. Microbiol.* **33**, 139–168
189. Vanderbilt, A. S., Gaby, N. S., and Rodwell, V. W. (1975) *J. Biol. Chem.* **250**, 5322–5329
190. Gary, J. D., Lin, W.-J., Yang, M. C., Herschman, H. R., and Clarke, S. (1996) *J. Biol. Chem.* **271**, 12585–12594
191. Bogumil, R., Knipp, M., Fundel, S. M., and Vasák, M. (1998) *Biochemistry* **37**, 4791–4798
192. Rosenthal, G. A. (1984) *Sci. Am.* **250**(Jun), 164–171
193. Rosenthal, G. A. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1780–1784
194. Melangeli, C., Rosenthal, G. A., and Dalman, D. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2255–2260
195. Fritsche, E., Humm, A., and Huber, R. (1999) *J. Biol. Chem.* **274**, 3026–3032
196. Takata, Y., Konishi, K., Gomi, T., and Fujioka, M. (1994) *J. Biol. Chem.* **269**, 5537–5542
197. Humm, A., Fritsche, E., Steinbacher, S., and Huber, R. (1997) *EMBO J.* **16**, 3373–3385
198. Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R., and Chapman, M. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8449–8454
199. Turner, D. M., and Walker, J. B. (1987) *J. Biol. Chem.* **262**, 6605–6609
200. Cohen, S. S. (1971) *Introduction to the Polyamines*, Prentice-Hall, Englewood Cliffs, New Jersey
201. Tabor, C. W., and Tabor, H. (1976) *Ann. Rev. Biochem.* **45**, 285–306
202. Cohen, S. S. (1978) *Nature (London)* **274**, 209–210
203. Abraham, A. K., and Pihl, A. (1981) *Trends Biochem. Sci.* **6**, 106–107
204. Igarashi, K., Saisho, T., Yuguchi, M., and Kashiwagi, K. (1997) *J. Biol. Chem.* **272**, 4058–4064
205. Cohen, S. S., and Greenberg, M. L. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5470–5474
206. Tsuboi, M. (1964) *Bull. Chem. Soc. Jap.* **37**, 1514–1522
207. Balkema, G. W., Mangini, N. J., and Pinto, L. H. (1983) *Science* **219**, 1083–1087
207a. Chattopadhyay, M. K., Tabor, C. W., Tabor, H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10330 — 10334
208. Herbst, E. J. et al. (1955) *J. Biol. Chem.* **214**, 175–
209. Leroy, D., Filhol, O., Delcros, J. G., Pares, S., Chambaz, E. M., and Cochet, C. (1997) *Biochemistry* **36**, 1242–1250
210. Hafner, E. W., Tabor, C. W., and Tabor, H. (1979) *J. Biol. Chem.* **254**, 12419–12426
211. Balasundaram, D., Tabor, C. W., and Tabor, H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5872–5876
211a. Chattopadhyay, M. K., Tabor, C. W., and Tabor, H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10330 – 10334
212. Cohn, M. S., Tabor, C. W., Tabor, H., and Wickner, R. B. (1978) *J. Biol. Chem.* **253**, 5225–5227
213. Iyer, R., and Delcour, A. H. (1997) *J. Biol. Chem.* **272**, 18595–18601
214. Kaouass, M., Gamache, I., Ramotar, D., Audette, M., and Poulin, R. (1998) *J. Biol. Chem.* **273**, 2109–2117
215. Antognoni, F., Del Duca, S., Kuraishi, A., Kawabe, E., Fukuchi-Shimogori, T., Kashiwagi, K., and Igarashi, K. (1999) *J. Biol. Chem.* **274**, 1942–1948
216. Momany, C., Ernst, S., Ghosh, R., Chang, N.-L., and Hackert, M. L. (1995) *J. Mol. Biol.* **252**, 643–655
217. Srivenugopal, K. S., and Adiga, P. R. (1981) *J. Biol. Chem.* **256**, 9532–9541
218. Bowman, W. H., Tabor, C. W., and Tabor, H. (1973) *J. Biol. Chem.* **248**, 2480–2486
219. Pegg, A. E., Wechter, R. S., Clark, R. S., Wiest, L., and Erwin, B. G. (1986) *Biochemistry* **25**, 379–384
219a. Tolbert, W. D., Ekstrom, J. L., Mathews, I. I., Secrist, J. A., III, Kapoor, P., Pegg, A. E., and Ealick, S. E. (2001) *Biochemistry* **40**, 9484–9494
220. Heby, O., and Persson, L. (1990) *Trends Biochem. Sci.* **15**, 153–158
221. Orr, G. R., Danz, D. W., Pontoni, G., Probhakaran, P. C., Gould, S. J., and Coward, J. K. (1988) *J. Am. Chem. Soc.* **110**, 5791–5799
222. Oshima, T., Hamasaki, N., Senshu, M., Kakinuma, K., and Kuwajima, I. (1987) *J. Biol. Chem.* **262**, 11979–11981

References

223. Fujihara, S., Abe, H., and Yoneyama, T. (1995) *J. Biol. Chem.* **270**, 9932–9938
224. Cacciapuoti, G., Porcelli, M., Bertoldo, C., De Rosa, M., and Zappia, V. (1994) *J. Biol. Chem.* **269**, 24762–24769
224a. Kröger, N., Deutzmann, R., and Sumper, M. (1999) *Science* **286**, 1129–1132
224b. Kröger, N., Deutzmann, R., Bergsdorf, C., and Sumper, M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14133–14138
224c. Ober, D., and Hartmann, T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14777–14782
224d. Kröger, N., Lorenz, S., Brunner, B., and Sumper, M. (2002) *Science* **298**, 584 – 585
225. Jasys, V. J., Kelbaugh, P. R., Nason, D. M., Phillips, D., Rosnack, K. J., Saccomano, N. A., Stroh, J. G., and Volkmann, R. A. (1990) *J. Am. Chem. Soc.* **112**, 6696–6704
226. McConlogue, L., Gupta, M., Wu, L., and Coffino, P. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 540–544
227. Rom, E., and Kahana, C. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3959–3963
228. Li, R.-S., Abrahamsen, M. S., Johnson, R. R., and Morris, D. R. (1994) *J. Biol. Chem.* **269**, 7941–7949
229. Pena, A., Reddy, C. D., Wu, S., Hickok, N. J., Reddy, E. P., Yumet, G., Soprano, D. R., and Soprano, K. J. (1993) *J. Biol. Chem.* **268**, 27277–27285
230. Hayashi, S.-i, Murakami, Y., and Matsufuji, S. (1996) *Trends Biochem. Sci.* **21**, 27–30
230a. Zhu, C., Lang, D. W., and Coffino, P. (1999) *J. Biol. Chem.* **274**, 26425–26430
230b. Coffino, P. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4421–4423
231. Atmar, V. J., and Kuehn, G. D. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5518–5522
232. Coleman, C. S., Huang, H., and Pegg, A. E. (1995) *Biochemistry* **34**, 13423–13430
233. Coleman, C. S., and Pegg, A. E. (1997) *J. Biol. Chem.* **272**, 12164–12169
233a. Bina, C., Angelini, R., Federico, R., Ascenzi, P., and Mattevi, A. (2001) *Biochemistry* **40**, 2766–2776
234. Richards, J. C., and Spenser, I. D. (1978) *J. Am. Chem. Soc.* **100**, 7402–7404
235. Lundgren, D. W., and Fales, H. M. (1980) *J. Biol. Chem.* **255**, 4481–4486
236. Lee, Y., and Sayre, L. M. (1998) *J. Biol. Chem.* **273**, 19490–19494
236a. White, W. H., Gunyuzlu, P. L., Toyn, J. H. (2001) *J. Biol. Chem.* **276**, 10794 – 10800
237. Tabor, H., and Tabor, C. W. (1975) *J. Biol. Chem.* **250**, 2648–2654
238. Smith, K., Nadeau, K., Bradley, M., Walsh, C., and Fairlamb, A. H. (1992) *Protein Sci.* **1**, 874–883
239. Bollinger, J. M., Jr., Kwon, D. S., Huisman, G. W., Kolter, R., and Walsh, C. T. (1995) *J. Biol. Chem.* **270**, 14031–14041
240. Folk, J. E., Park, M. H., Chung, S. I., Schrode, J., Lester, E. P., and Cooper, H. L. (1980) *J. Biol. Chem.* **255**, 3695–3700
241. Leete, E., and McDonnell, J. A. (1981) *J. Am. Chem. Soc.* **103**, 658–662
242. Park, M. H., Cooper, H. L., and Folk, J. E. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2869–2873
243. Osterman, A. L., Brooks, H. B., Rizo, J., and Phillips, M. A. (1997) *Biochemistry* **36**, 4558–4567
244. Coleman, C. S., Stanley, B. A., and Pegg, A. E. (1993) *J. Biol. Chem.* **268**, 24572–24579
244a. Marali, S. (1999) *J. Biol. Chem.* **274**, 21017–21022
245. Pegg, A. E., Jones, D. B., and Secrist, J. A., III. (1988) *Biochemistry* **27**, 1408–1415
246. Hyvönen, T., Alakuijala, L., Andersson, L., Khomutov, A. R., Khomutov, R. M., and Eloranta, T. O. (1988) *J. Biol. Chem.* **263**, 11138–11144
247. Cohen, G. N., and Saint-Girons, I. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 429–442, Am. Soc. for Microbiology, Washington, DC
248. Shames, S. L., Ash, D. E., Wedler, F. C., and Villafranca, J. J. (1984) *J. Biol. Chem.* **259**, 15331–15339
249. Angeles, T. S., Hunsley, J. R., and Viola, R. E. (1992) *Biochemistry* **31**, 799–805
249a. Hadfield, A., Kryger, G., Ouyang, J., Petsko, G. A., Ringe, D., and Viola, R. (1999) *J. Mol. Biol.* **289**, 991–1002
250. Ouyang, J., and Viola, R. E. (1995) *Biochemistry* **34**, 6394–6399
251. Ravanel, S., Gakière, B., Job, D., and Douce, R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7805–7812
252. Giovanelli, J., Mudd, S. H., and Datko, A. H. (1978) *J. Biol. Chem.* **253**, 5665–5677
253. Huo, X., and Viola, R. E. (1996) *Biochemistry* **35**, 16180–16185
254. Laber, B., Gerbling, K.-P., Harde, C., Neff, K.-H., Nordhoff, E., and Pohlenz, H.-D. (1994) *Biochemistry* **33**, 3413–3423
255. Curien, G., Job, D., Douce, R., and Dumas, R. (1998) *Biochemistry* **37**, 13212–13221
255a. Thomazeau, K., Curien, G., Dumas, R., and Biou, V. (2001) *Protein Sci.* **10**, 638–648
256. Chinchilla, D., Schwarz, F. P., and Eisenstein, E. (1998) *J. Biol. Chem.* **273**, 23219–23224
257. Horowitz, B., and Meister, A. (1972) *J. Biol. Chem.* **247**, 6708–6719
257a. Aghaiypour, K., Wlodawer, A., and Lubkowski, J. (2001) *Biochemistry* **40**, 5655–5664
258. Tate, S. S., and Meister, A. (1971) *Adv. Enzymol.* **35**, 503–543
259. Funkhouser, J. D., Abraham, A., Smith, V. A., and Smith, W. G. (1974) *J. Biol. Chem.* **249**, 5478–5484
259a. Tedeschi, G., Ronchi, S., Simonic, T., Treu, C., Mattevi, A., and Negri, A. (2001) *Biochemistry* **40**, 4738–4744
259b. Bossi, R. T., Negri, A., Tedeschi, G., and Mattevi, A. (2002) *Biochemistry* **41**, 3018–3024
260. Rafalski, J. A., and Falco, S. C. (1988) *J. Biol. Chem.* **263**, 2146–2151
261. Zakin, M. M., Duchange, N., Ferrara, P., and Cohen, G. N. (1983) *J. Biol. Chem.* **258**, 3028–3031
262. Wedler, F. C., and Ley, B. W. (1993) *J. Biol. Chem.* **268**, 4880–4888
262a. James, C. L., and Viola, R. E. (2002) *Biochemistry* **41**, 3726–3731
263. Cassan, M., Parsot, C., Cohen, G. N., and Patte, J.-C. (1986) *J. Biol. Chem.* **261**, 1052–1057
264. Chen, N.-Y., and Paulus, H. (1988) *J. Biol. Chem.* **263**, 9526–9532
264a. Ehmann, D. E., Gehring, A. M., and Walsh, C. T. (1999) *Biochemistry* **38**, 6171–6177
265. Fujioka, M., Takata, Y., Ogawa, H., and Okamoto, M. (1980) *J. Biol. Chem.* **255**, 937–942
266. Bhattacharjee, J. K. (1992) in *The Evolution of Metabolic Function* (Mortlock, R. P., ed), pp. 47–80, CRC Press, Boca Raton, Florida
267. Mirwaldt, C., Korndörfer, I., and Huber, R. (1995) *J. Mol. Biol.* **246**, 227–239
268. Blickling, S., Beisel, H.-G., Bozic, D., Knäblein, J., Laber, B., and Huber, R. (1997) *J. Mol. Biol.* **274**, 608–621
269. Blickling, S., Renner, C., Laber, B., Pohlenz, H.-D., Holak, T. A., and Huber, R. (1997) *Biochemistry* **36**, 24–33
270. Scapin, G., Blanchard, J. S., and Sacchettini, J. C. (1995) *Biochemistry* **34**, 3502–3512
271. Reddy, S. G., Scapin, G., and Blanchard, J. S. (1996) *Biochemistry* **35**, 13294–13302
272. Simms, S. A., Voige, W. H., and Gilvarg, C. (1984) *J. Biol. Chem.* **259**, 2734–2741
273. Beaman, T. W., Binder, D. A., Blanchard, J. S., and Roderick, S. L. (1997) *Biochemistry* **36**, 489–494
274. Beaman, T. W., Blanchard, J. S., and Roderick, S. L. (1998) *Biochemistry* **37**, 10363–10369
275. Scapin, G., Cirilli, M., Reddy, S. G., Gao, Y., Vederas, J. C., and Blanchard, J. S. (1998) *Biochemistry* **37**, 3278–3285
276. Cirilli, M., Zheng, R., Scapin, G., and Blanchard, J. S. (1998) *Biochemistry* **37**, 16452–16458
276a. Koo, C. W., Sutherland, A., Vederas, J. C., and Blanchard, J. S. (2000) *J. Am. Chem. Soc.* **122**, 6122–6123
277. Daniel, R. A., and Errington, J. (1993) *J. Mol. Biol.* **232**, 468–483
278. Henderson, L. M., Nelson, P. J., and Henderson, L. (1982) *Fed. Proc.* **41**, 2843–2847
279. Dunn, W. A., Rettura, G., Seifter, E., and Englard, S. (1984) *J. Biol. Chem.* **259**, 10764–10770
280. Golebiewski, W. M., and Spenser, I. D. (1984) *J. Am. Chem. Soc.* **106**, 1441–1442
281. Kang, K. R., Wolff, E. C., Park, M. H., Folk, J. E., and Chung, S. I. (1995) *J. Biol. Chem.* **270**, 18408–18412
282. Park, M. H., Joe, Y. A., and Kang, K. R. (1998) *J. Biol. Chem.* **273**, 1677–1683
282a. Wolff, E. C., Wolff, J., and Park, M. H. (2000) *J. Biol. Chem.* **275**, 9170–9177
283. Luckner, M. (1972) *Secondary Metabolism in Plants and Animals*, Academic Press, New York (p. 275)
284. Byron, C. M., Stankovich, M. T., and Husain, M. (1990) *Biochemistry* **29**, 3691–3700
285. Soda, K., Misono, H., and Yamamoto, T. (1968) *Biochemistry* **7**, 4102–4109
286. Grove, J. A., Linn, T. G., Willett, C. J., and Henderson, L. M. (1970) *Biochim. Biophys. Acta.* **215**, 191–194
287. Markovitz, P. J., and Chuang, D. T. (1987) *J. Biol. Chem.* **262**, 9353–9358
288. Markovitz, P. J., Chuang, D. T., and Cox, R. P. (1984) *J. Biol. Chem.* **259**, 11643–11646
289. Fangmeier, N., and Leistner, E. (1980) *J. Biol. Chem.* **255**, 10205–10209
290. Wickwire, B. M., Harris, C. M., Harris, T. M., and Broquist, H. P. (1990) *J. Biol. Chem.* **265**, 14742–14747
291. Mihalik, S. J., McGuinness, M., and Watkins, P. A. (1991) *J. Biol. Chem.* **266**, 4822–4830
292. Wickwire, B. M., Wagner, C., and Broquist, H. P. (1990) *J. Biol. Chem.* **265**, 14748–14753
293. Rothstein, M. (1965) *Arch. Biochem. Biophys.* **111**, 467–476
294. Chang, Y. F., and Adams, E. (1977) *J. Biol. Chem.* **252**, 7987–7991
294a. Tang, K.-H., Harms, A., and Frey, P. A. (2002) *Biochemistry* **41**, 8767 – 8776
295. Ohsugi, M., Kahn, J., Hensley, C., Chew, S., and Barker, H. A. (1981) *J. Biol. Chem.* **256**, 7642–7651
296. Brzovic, P., Holbrook, E. L., Greene, R. C., and Dunn, M. F. (1990) *Biochemistry* **29**, 442–451
296a. Steegborn, C., Laber, B., Messerschmidt, A., Huber, R., and Clausen, T. (2001) *J. Mol. Biol.* **311**, 789–801
297. Clausen, T., Huber, R., Laber, B., Pohlenz, H.-D., and Messerschmidt, A. (1996) *J. Mol. Biol.* **262**, 202–224
298. Kery, V., Bukovska, G., and Kraus, J. P. (1994) *J. Biol. Chem.* **269**, 25283–25288
299. Taoka, S., Ohja, S., Shan, X., Kruger, W. D., and Banerjee, R. (1998) *J. Biol. Chem.* **273**, 25179–25184

References

299a. Mosharov, E., Cranford, M. R., and Banerjee, R. (2000) *Biochemistry* **39**, 13005–13011
299b. Jhee, K.-H., Niks, D., McPhie, P., Dunn, M. F., and Miles, E. W. (2001) *Biochemistry* **40**, 10873–10880
299c. Taoka, S., Lepore, B. W., Kabil, Ö., Ojha, S., Ringe, D., and Banerjee, R. (2002) *Biochemistry* **41**, 10454 – 10461
300. Nishi, N., Tanabe, H., Oya, H., Urushihara, M., Miyanaka, H., and Wada, F. (1994) *J. Biol. Chem.* **269**, 1015–1019
301. Takusagawa, F., Kamitori, S., Misaki, S., and Markham, G. D. (1996) *J. Biol. Chem.* **271**, 136–147
302. Takusagawa, F., Kamitori, S., and Markham, G. D. (1996) *Biochemistry* **35**, 2586–2596
302a. McQueney, M. S., Anderson, K. S., and Markham, G. D. (2000) *Biochemistry* **39**, 4443–4454
302b. Sánchez del Pino, M. M., Corrales, F. J., and Mato, J. M. (2000) *J. Biol. Chem.* **275**, 23476–23482
302c. Elrod, P., Zhang, J., Yang, X., Yin, D., Hu, Y., Borchardt, R. T., and Schowen, R. L., (2002) *Biochemistry* **41**, 8134 – 8142
303. Chen, Z., Crippen, K., Gulati, S., and Banerjee, R. (1994) *J. Biol. Chem.* **269**, 27193–27197
303a. Bandarian, V., and Matthews, R. G. (2001) *Biochemistry* **40**, 5056–5064
303b. Olteanu, H., and Banerjee, R. (2001) *J. Biol. Chem.* **276**, 35558–35563
304. Garrow, T. A. (1996) *J. Biol. Chem.* **271**, 22831–22838
304a. Thomas, D., Becker, A., and Surdin-Kerjan, Y. (2000) *J. Biol. Chem.* **275**, 40718–40724
305. Livesey, G. (1984) *Trends Biochem. Sci.* **9**, 27– 29
306. Soda, K., Tanaka, H., and Esaki, N. (1983) *Trends Biochem. Sci.* **8**, 214–217
307. Welch, G. N., and Loscalzo, J. (1998) *N. Engl. J. Med.* **338**, 1042–1050
308. Durand, P., Lussier-Cacan, S., and Blache, D. (1997) *FASEB J.* **11**, 1157–1168
309. Watanabe, M., Osada, J., Aratani, Y., Kluckman, K., Reddick, R., Malinow, M. R., and Maeda, N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1585–1589
309a. Jakubowski, H. (1999) *FASEB J.* **13**, 2277–2283
309b. Ragone, R. (2002) *FASEB J.* **16**, 401–404
310. Finkelstein, J. D., and Martin, J. J. (1986) *J. Biol. Chem.* **261**, 1582–1587
311. James, F., Nolte, K. D., and Hanson, A. D. (1995) *J. Biol. Chem.* **270**, 22344–22350
311a. Ranocha, P., Bourgis, F., Ziemak, M. J., Rhodes, D., Gage, D. A., and Hanson, A. D. (2000) *J. Biol. Chem.* **275**, 15962–15968
312. Slany, R. K., Bösl, M., Crain, P. F., and Kersten, H. (1993) *Biochemistry* **32**, 7811–7817
313. Furfine, E. S., and Abeles, R. H. (1988) *J. Biol. Chem.* **263**, 9598–9606
314. Wray, J. W., and Abeles, R. H. (1995) *J. Biol. Chem.* **270**, 3147–3153
314a. Dai, Y., Pochapsky, T. C., and Abeles, R. H. (2001) *Biochemistry* **40**, 6379–6387
314b. Al-Mjeni, F., Ju, T., Pochapsky, T. C., and Maroney, M. J. (2002) *Biochemistry* **41**, 6761 – 6769
315. Wray, J. W., and Abeles, R. H. (1993) *J. Biol. Chem.* **268**, 21466–21469
316. Adams, D. O., and Yang, S. F. (1981) *Trends Biochem. Sci.* **6**, 161–164
317. Pirrung, M. C. (1983) *J. Am. Chem. Soc.* **105**, 7207–7209
318. Chang, C. (1996) *Trends Biochem. Sci.* **21**, 129–133
319. Gamble, R. L., Coonfield, M. L., and Schaller, G. E. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7825–7829
320. Hao, D., Ohme-Takagi, M., and Sarai, A. (1998) *J. Biol. Chem.* **273**, 26857–26861
320a. Tatsuki, M., and Mori, H. (2001) *J. Biol. Chem.* **276**, 28051–28057
320b. McCarthy, D. L., Capitani, G., Feng, L., Gruetter, M. G., and Kirsch, J. F. (2001) *Biochemistry* **40**, 12276–12284
321. Boller, T., and Kende, H. (1980) *Nature (London)* **286**, 259–260
322. Peiser, G. D., Wang, T.-T., Hoffman, N. E., Yang, S. F., Liu, H.-w, and Walsh, C. T. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3059–3063
323. Barlow, J. N., Zhang, Z., John, P., Baldwin, J. E., and Schofield, C. J. (1997) *Biochemistry* **36**, 3563–3569
324. Zhang, Z., Barlow, J. N., Baldwin, J. E., and Schofield, C. J. (1997) *Biochemistry* **36**, 15999–16007
324a. Brunhuber, N. M. W., Mort, J. L., Christoffersen, R. E., and Reich, N. O. (2000) *Biochemistry* **39**, 10730–10738
324b. Thrower, J. S., Blalock, R., III, and Klinman, J. P. (2001) *Biochemistry* **40**, 9717–9724
325. Pirrung, M. C., Kaiser, L. M., and Chen, J. (1993) *Biochemistry* **32**, 7445–7450
326. Blomstrom, D. C., and Beyer, E. M., Jr. (1980) *Nature (London)* **283**, 66–68
327. Liu, H.-W., Auchus, R., and Walsh, C. T. (1984) *J. Am. Chem. Soc.* **106**, 5335–5348
327a. Yao, M., Ose, T., Sugimoto, H., Horiuchi, A., Nakagawa, A., Wakatsuki, S., Yokoi, D., Murakami, T., Honma, M., and Tanaka, I. (2000) *J. Biol. Chem.* **275**, 34557–34565
328. Bousquet, J.-F., and Thimann, K. V. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1724–1727
329. Rabinowitz, K. W., Niederman, R. A., and Wood, W. A. (1973) *J. Biol. Chem.* **248**, 8207–8215
330. Hirose, K., Kawata, Y., Yumoto, N., and Tokushige, M. (1991) *J. Biochem.* **110**, 971–975
331. Eisenstein, E. (1991) *J. Biol. Chem.* **266**, 5801–5807
332. Eisenstein, E., Yu, H. D., Fisher, K. E., Iacuzio, D. A., Ducote, K. R., and Schwarz, F. P. (1995) *Biochemistry* **34**, 9403–9412
333. Samach, A., Hareven, D., Gutfinger, T., Ken-Dror, S., and Lifschitz, E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2678–2682
334. Umbarger, H. E. (1956) *Science* **123**, 848
335. Liu, J.-Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S., and Yamada, H. (1998) *J. Biol. Chem.* **273**, 16678–16685
336. Epperly, B. R., and Dekker, E. E. (1991) *J. Biol. Chem.* **266**, 6086–6092
337. Ray, M., and Ray, S. (1985) *J. Biol. Chem.* **260**, 5913–5918
338. Kelley, J. J., and Dekker, E. E. (1984) *J. Biol. Chem.* **259**, 2124–2129
338a. Cheong, C.-G., Escalante-Semerena, J. C., and Rayment, I. (2002) *Biochemistry* **41**, 9079 – 9089
339. Barak, Z., Chipman, D. M., and Schloss, J. V., eds. (1990) *Biosynthesis of Branched Chain Amino Acids*, VCH Publ., Basel
340. Umbarger, H. E. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 352–366, Am. Soc. Microbiology, Washington, DC
340a. Lee, Y.-T., and Duggleby, R. G. (2001) *Biochemistry* **40**, 6836–6844
340b. Mendel, S., Elkayam, T., Sella, C., Vinogradov, V., Vyazmensky, M., Chipman, D. M., and Barak, Z. (2001) *J. Mol. Biol.* **307**, 465–477
341. Lawther, R. P., Calhoun, D. H., Adams, C. W., Hauser, C. A., Gray, J., and Hatfield, G. W. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 922–925
342. Ray, T. B. (1986) *Trends Biochem. Sci.* **11**, 180–183
343. Halgand, F., Vives, F., Dumas, R., Biou, V., Andersen, J., Andrieu, J.-P., Cantegril, R., Gagnon, J., Douce, R., Forest, E., and Job, D. (1998) *Biochemistry* **37**, 4773–4781
344. Roy, T. B. (1986) *Trends Biochem. Sci.* **11**, 180–183
345. Schloss, J. V., van Dyk, D. E., Vasta, J. F., and Kutny, R. M. (1985) *Biochemistry* **24**, 4952–4959
346. Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., and Pebay-Peyroula, E. (1997) *EMBO J.* **16**, 3405–3415
347. Flint, D. H., and Emptage, M. H. (1988) *J. Biol. Chem.* **263**, 3558–3564
348. Ekiel, I., Smith, I. C. P., and Sprout, G. D. (1984) *Biochemistry* **23**, 1683–1687
349. Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D., and Petsko, G. A. (1997) *J. Mol. Biol.* **266**, 1016–1031
350. Powers, S. G., and Snell, E. E. (1976) *J. Biol. Chem.* **251**, 3786–3793
350a. Stolz, J., and Sauer, N. (1999) *J. Biol. Chem.* **274**, 18747–18752
350b. Yun, M., Park, C.-G., Kim, J.-Y., Rock, C. O., Jackowski, S., and Park, H.-W. (2000) *J. Biol. Chem.* **275**, 28093–28099
350c. Matak-Vinkovic, D., Vinkovic, M., Saldanha, S. A., Ashurst, J. L., von Delft, F., Inoue, T., Miguel, R. N., Smith, A. G., Blundell, T. L., and Abell, C. (2001) *Biochemistry* **40**, 14493–14500
350d. Zheng, R., and Blanchard, J. S. (2000) *Biochemistry* **39**, 16244–16251
350e. Zheng, R., and Blanchard, J. S. (2001) *Biochemistry* **40**, 12904–12912
350f. Strauss, E., Kinsland, C., Ge, Y., McLafferty, F. W., and Begley, T. P. (2001) *J. Biol. Chem.* **276**, 13513–13516
350g. Strauss, E., and Begley, T. P. (2001) *J. Am. Chem. Soc.* **123**, 6449–6450
350h. Mootz, H. D., Finking, R., and Marahiel, M. A. (2001) *J. Biol. Chem.* **276**, 37289–37298
350i. Hoenke, S., Schmid, M., and Dimroth, P. (2000) *Biochemistry* **39**, 13233–13240
350j. Schneider, K., Dimroth, P., and Bott, M. (2000) *Biochemistry* **39**, 9438–9450
351. Alexeev, D., Alexeeva, M., Baxter, R. L., Campopiano, D. J., Webster, S. P., and Sawyer, L. (1998) *J. Mol. Biol.* **284**, 401–419
351a. Eliot, A. C., Sandmark, J., Schneider, G., and Kirsch, J. F. (2002) *Biochemistry* **41**, 12582 – 12589
352. Gibson, K. J., Lorimer, G. H., Rendina, A. R., Taylor, W. S., Cohen, G., Gatenby, A. A., Payne, W. G., Roe, D. C., Lockett, B. A., Nudelman, A., Marcovici, D., Nachum, A., Wexler, B. A., Marsilii, E. L., Turner, I. M., Sr., Howe, L. D., Kalbach, C. E., and Chi, H. (1995) *Biochemistry* **34**, 10976–10984
353. Käck, H., Sandmark, J., Gibson, K. J., Schneider, G., and Lindqvist, Y. (1998) *Protein Sci.* **7**, 2560–2566
354. Gibson, K. J. (1997) *Biochemistry* **36**, 8474–8478
355. Käck, H., Gibson, K. J., Lindqvist, Y., and Schneider, G. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5495–5500
355a. Birch, O. M., Hewitson, K. S., Fuhrmann, M., Burgdorf, K., Baldwin, J. E., Roach, P. L., and Shaw, N. M. (2000) *J. Biol. Chem.* **275**, 32277–32280
355b. Ugulava, N. B., Gibney, B. R., and Jarrett, J. T. (2001) *Biochemistry* **40**, 8343–8351
355c. Ollagnier-de-Choudens, S., Mulliez, E., Hewitson, K. S., and Fontecave, M. (2002) *Biochemistry* **41**, 9145 – 9152
356. Hayden, M. A., Huang, I., Bussiere, D. E., and Ashley, G. W. (1992) *J. Biol. Chem.* **267**, 9512–9515
356a. Yennawar, N. H., Conway, M. E., Yennawar, H. P., Farber, G. K., and Hutson, S. M. (2002) *Biochemistry* **41**, 11592 – 11601
357. Turnbull, A. P., Baker, P. J., and Rice, D. W. (1997) *J. Biol. Chem.* **272**, 25105–25111
358. Ikeda, Y., and Tanaka, K. (1983) *J. Biol. Chem.* **258**, 9477–9487

References

359. Hawes, J. W., Crabb, D. W., Chan, R. M., Rougraff, P. M., and Harris, R. A. (1995) *Biochemistry* **34**, 4231–4237
360. Kamoun, P. (1992) *Trends Biochem. Sci.* **17**, 175–176
361. Wolf, D. A., and Akers, H. A. (1986) *Trends Biochem. Sci.* **11**, 390–392
362. Sabourin, P. J., and Bieber, L. L. (1982) *J. Biol. Chem.* **257**, 7460–7467
362a. Hutson, S. M., Cree, T. C., and Harper, A. E. (1978) *J. Biol. Chem.* **253**, 8126–8133
362b. Tischler, M. E., Desautels, M., and Goldberg, A. L. (1982) *J. Biol. Chem.* **257**, 1613–1621
362c. Mordier, S., Deval, C., Béchet, D., Tassa, A., and Ferrara, M. (2000) *J. Biol. Chem.* **275**, 29900–29906
362d. Mitch, W. E., and Goldberg, A. L. (1996) *N. Engl. J. Med.* **335**, 1897–1905
362e. Nissen, S., Sharp, R., Ray, M., Rathmacher, J. A., Rice, D., Fuller, J. C., Jr., Connelly, A. S., and Abumrad, N. (1996) *J. Appl. Physiol.* **81**, 2095–2104
362f. Nissen, S. L., and Abumrad, N. N. (1997) *J. Nutr. Biochem.* **8**, 300–311
363. Buckel, W. (1992) *FEMS Microbiol. Rev.* **88**, 211–232
364. Ho, C.-L., Noji, M., Saito, M., and Saito, K. (1999) *J. Biol. Chem.* **274**, 397–402
364a. Xue, H.-H., Fujie, M., Sakaguchi, T., Oda, T., Ogawa, H., Kneer, N. M., Lardy, H. A., and Ichiyama, A. (1999) *J. Biol. Chem.* **274**, 16020–16027
365. Snell, K. (1986) *Trends Biochem. Sci.* **11**, 241–243
366. Williams, H. E., and Smith, L. H., Jr. (1971) *Science* **171**, 390–391
366a. Lumb, M. J., Drake, A. F., and Danpure, C. J. (1999) *J. Biol. Chem.* **274**, 20587–20596
367. Danpure, C. J., and Purdue, P. E. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2385–2424, McGraw-Hill, New York
368. Noji, M., Inoue, K., Kimura, N., Gouda, A., and Saito, K. (1998) *J. Biol. Chem.* **273**, 32739–32745
369. Tai, C.-H., Yoon, M.-Y., Kim, S.-K., Rege, V. D., Nalabolu, S. R., Kredich, N. M., Schnackerz, K. D., and Cook, P. F. (1998) *Biochemistry* **37**, 10597–10604
369a. Hindson, V. J., Moody, P. C. E., Rowe, A. J., and Shaw, W. V. (2000) *J. Biol. Chem.* **275**, 461–466
370. Renwick, S. B., Snell, K., and Baumann, U. (1998) *Structure* **6**, 1105–1116
371. Vatcher, G. P., Thacker, C. M., Kaletta, T., Schnabel, H., Schnabel, R., and Baillie, D. L. (1998) *J. Biol. Chem.* **273**, 6066–6073
371a. Rontein, D., Nishida, I., Tashiro, G., Yoshioka, K., Wu, W.-I., Voelker, D. R., Basset, G., and Hanson, A. D. (2001) *J. Biol. Chem.* **276**, 35523–35529
372. Rathinasabapathi, B., Burnet, M., Russell, B. L., Gage, D. A., Liao, P.-C., Nye, G. J., Scott, P., Golbeck, J. H., and Hanson, A. D. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3454–3458
373. Glaasker, E., Konings, W. N., and Poolman, B. (1996) *J. Biol. Chem.* **271**, 10060–10065
374. Klein, S. M., and Sagers, R. D. (1967) *J. Biol. Chem.* **242**, 297–300, 301–305
374a. Piper, M. D., Hong, S.-P., Ball, G. E., and Dawes, I. W. (2000) *J. Biol. Chem.* **275**, 30987–30995
375. Fujioka, M., Takata, Y., Konishi, K., and Ogawa, H. (1987) *Biochemistry* **26**, 5696–5702
376. Chlumsky, L. J., Sturgess, A. W., Nieves, E., and Jorns, M. S. (1998) *Biochemistry* **37**, 2089–2095
376a. Eschenbrenner, M., Chlumsky, L. J., Khanna, P., Strasser, F., and Jorns, M. S. (2001) *Biochemistry* **40**, 5352–5367
376b. Zhao, G., and Jorns, M. S. (2002) *Biochemistry* **41**, 9747 – 9750
377. Reuber, B. E., Karl, C., Reimann, S. A., Mihalik, S. J., and Dodt, G. (1997) *J. Biol. Chem.* **272**, 6766–6776
377a. Harris, R. J., Meskys, R., Sutcliffe, M. J., and Scrutton, N. S. (2000) *Biochemistry* **39**, 1189–1198
378. Shemin, D., and Rittenberg, D. (1946) *J. Biol. Chem.* **166**, 621–625
379. Gong, J., Hunter, G. A., and Ferreira, G. C. (1998) *Biochemistry* **37**, 3509–3517
379a. Tan, D., Barber, M. J., and Ferreira, G. C. (1998) *Protein Sci.* **7**, 1208–1213
380. Whiting, M. J., and Granick, S. (1976) *J. Biol. Chem.* **251**, 1340–1346
381. Warren, M. J., Cooper, J. B., Wood, S. P., and Shoolingin-Jordan, P. M. (1998) *Trends Biochem. Sci.* **23**, 217–221
381a. Ruiz de Mena, I., Fernández-Moreno, M. A., Bornstein, B., Kaguni, L. S., and Garesse, R. (1999) *J. Biol. Chem.* **274**, 37321–37328
382. Warren, M. J., and Scott, A. I. (1990) *Trends Biochem. Sci.* **15**, 486–491
383. Jordan, P. M., ed. (1991) *Biosynthesis of Tetrapyrroles*, Elsevier Science Publ., New York
384. Bevan, D. R., Bodlaender, P., and Shemin, D. (1980) *J. Biol. Chem.* **255**, 2030–2035
385. Jahn, D., Verkamp, E., and Söll, D. (1992) *Trends Biochem. Sci.* **17**, 215–218
385a. Moser, J., Schubert, W.-D., Beier, V., Bringemeier, I., Jahn, D., and Heinz, D. W. (2001) *EMBO J.* **20**, 6583–6590
386. Rieble, S., and Beale, S. I. (1991) *J. Biol. Chem.* **266**, 9740–9745
387. Hennig, M., Grimm, B., Contestabile, R., John, R. A., and Jansonius, J. N. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4866–4871
388. Dent, A. J., Beyersmann, D., Block, C., and Hasnain, S. S. (1990) *Biochemistry* **29**, 7822–7828
389. Jaffe, E. K., Volin, M., and Myers, C. B. (1994) *Biochemistry* **33**, 11554–11562
390. Cheung, K.-M., Spencer, P., Timko, M. P., and Shoolingin-Jordan, P. M. (1997) *Biochemistry* **36**, 1148–1156
390a. Erskine, P. T., Newbold, R., Brindley, A. A., Wood, S. P., Shoolingin-Jordan, P. M., Warren, M. J., and Cooper, J. B. (2001) *J. Mol. Biol.* **312**, 133–141
390b. Kervinen, J., Jaffe, E. K., Stauffer, F., Neier, R., Wlodawer, A., and Zdanov, A. (2001) *Biochemistry* **40**, 8227–8236
390c. Jaffe, E. K., Volin, M., Bronson-Mullins, C. R., Dunbrack, R. L. J., Kervinen, J., Martins, J., Quinlan, J. F., Jr., Sazinsky, M. H., Steinhouse, E. M., and Yeung, A. T. (2000) *J. Biol. Chem.* **275**, 2619–2626
391. 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. (1992) *Nature (London)* **359**, 33–39
392. Woodcock, S. C., and Jordan, P. M. (1994) *Biochemistry* **33**, 2688–2695
393. Porcher, C., Picat, C., Daegelen, D., Beaumont, C., and Grandchamp, B. (1995) *J. Biol. Chem.* **270**, 17368–17374
394. Warren, M. J., Gul, S., Aplin, R. T., Scott, A. I., Roessner, C. A., O'Grady, P., and Shoolingin-Jordan, P. M. (1995) *Biochemistry* **34**, 11288–11295
394a. Mathews, M. A. A., Schubert, H. L., Whitby, F. G., Alexander, K. J., Schadick, K., Bergonia, H. A., Phillips, J. D., and Hill, C. P. (2001) *EMBO J.* **20**, 5832–5839
395. Roméo, P.-H., Raich, N., Dubart, A., Beaupain, D., Pryor, M., Kushner, J., Cohen-Solal, M., and Goossens, M. (1986) *J. Biol. Chem.* **261**, 9825–9831
396. Whitby, F. G., Phillips, J. D., Kushner, J. P., and Hill, C. P. (1998) *EMBO J.* **17**, 2463–2471
396a. Martins, B. M., Grimm, B., Mock, H.-P., Huber, R., and Messerschmidt, A. (2001) *J. Biol. Chem.* **276**, 44108–44116
397. Proulx, K. L., Woodard, S. I., and Dailey, H. A. (1993) *Protein Sci.* **2**, 1092–1098
398. Camadro, J.-M., Thome, F., Brouillet, N., and Labbe, P. (1994) *J. Biol. Chem.* **269**, 32085–32091
399. Dailey, T. A., and Dailey, H. A. (1996) *Protein Sci.* **5**, 98–105
400. Ishida, T., Yu, L., Akutsu, H., Ozawa, K., Kawanishi, S., Seto, A., Inubushi, T., and Sano, S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4853–4858
401. Dailey, H. A., Sellers, V. M., and Dailey, T. A. (1994) *J. Biol. Chem.* **269**, 390–395
402. Ferreira, G. C., Franco, R., Lloyd, S. G., Pereira, A. S., Moura, I., Moura, J. J. G., and Huynh, B. H. (1994) *J. Biol. Chem.* **269**, 7062–7065
402a. Ferreira, G. C., Franco, R., Mangravita, A., and George, G. N. (2002) *Biochemistry* **41**, 4809–4818
402b. Lu, Y., Sousa, A., Franco, R., Mangravita, A., Ferreira, G. C., Moura, I., and Shelnutt, J. A. (2002) *Biochemistry* **41**, 8253 – 8262
403. Smith, A. G., Santana, M. A., Wallace-Cook, A. D. M., Roper, J. M., and Labbe-Bois, R. (1994) *J. Biol. Chem.* **269**, 13405–13413
404. Daily, H. A., and Fleming, J. E. (1986) *J. Biol. Chem.* **261**, 7902–7905
405. Saiki, K., Mogi, T., Hori, H., Tsubaki, M., and Anraku, Y. (1993) *J. Biol. Chem.* **268**, 26927–26934
405a. Brown, K. R., Allan, B. M., Do, P., and Hegg, E. L. (2002) *Biochemistry* **41**, 10906 – 10913
406. Warren, M. J., Gonzalez, M. D., Williams, H. J., Stolowich, N. J., and Scott, A. I. (1990) *J. Am. Chem. Soc.* **112**, 5343–5345
407. Yap-Bondoc, F., Bondoc, L. L., Timkovich, R., Baker, D. C., and Hebbler, A. (1990) *J. Biol. Chem.* **265**, 13498–13500
408. Brown, C. E., Shemin, D., and Katz, J. J. (1973) *J. Biol. Chem.* **248**, 8015–8021
409. Scott, A. I., Townsend, C. A., Okada, K., and Kajiwara, M. (1974) *J. Am. Chem. Soc.* **96**, 8054–8069 and 8069–8080
410. Battersby, A. R. (1994) *Science* **264**, 1551–1557
411. Spencer, J. B., Stolowich, N. J., Santander, P. J., Pichon, C., Kajiwara, M., Tokiwa, S., Takatori, K., and Scott, A. I. (1994) *J. Am. Chem. Soc.* **116**, 4991–4992
412. Thibaut, D., Debussche, L., and Blanche, F. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8795–8799
413. Eisenreich, W., and Bacher, A. (1991) *J. Biol. Chem.* **266**, 23840–23849
414. Scott, A. I., Stolowich, N. J., Wang, J., Gawatz, O., Fridrich, E., and Müller, G. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14316–14319
415. Wang, J., Stolowich, N. J., Santander, P. J., Park, J. H., and Scott, A. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14320–14322
415a. Thompson, T. B., Thomas, M. G., Escalante-Semerena, J. C., and Rayment, I. (1999) *Biochemistry* **38**, 12995–13005
416. Michal, G., ed. (1999) *Biochemical Pathways*, Wiley–Spektrum Academischer Verlag, New York–Heidelberg, Germany
417. Bollivar, D. W., Suzuki, J. Y., Beatty, J. T., Dobrowolski, J. M., and Bauer, C. E. (1994) *J. Mol. Biol.* **237**, 622–640
418. Bollivar, D. W., Wang, S., Allen, J. P., and Bauer, C. E. (1994) *Biochemistry* **33**, 12763–12768

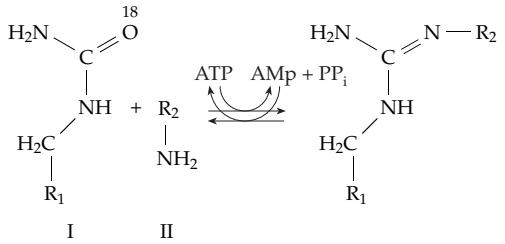
References

418a. Fujita, Y., and Bauer, C. E. (2000) *J. Biol. Chem.* **275**, 23583–23588
419. Jensen, P. E., Gibson, L. C. D., Henningsen, K. W., and Hunter, C. N. (1996) *J. Biol. Chem.* **271**, 16662–16667
420. Oster, U., Bauer, C. E., and Rüdiger, W. (1997) *J. Biol. Chem.* **272**, 9671–9676
421. Porra, R. J., and Meisch, H.-U. (1984) *Trends Biochem. Sci.* **9**, 99–104
422. Gibson, L. C. D., Willows, R. D., Kannangara, C. G., von Wettstein, D., and Hunter, C. N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1941–1944
423. Tanaka, A., Ito, H., Tanaka, R., Tanaka, N. K., Yoshida, K., and Okada, K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12719–12723
424. Scheumann, V., Schoch, S., and Rüdiger, W. (1998) *J. Biol. Chem.* **273**, 35102–35108
425. Warren, M. J., Jay, M., Hunt, D. M., Elder, G. H., and Röhl, J. C. G. (1996) *Trends Biochem. Sci.* **21**, 229–234
426. Brownlie, P. D., Lambert, R., Louie, G. V., Jordan, P. M., Blundell, T. L., Warren, M. J., Cooper, J. B., and Wood, S. P. (1994) *Protein Sci.* **3**, 1644–1650
427. Kappas, A., Sassa, S., Galbraith, R. A., and Nordmann, Y. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2103–2159, McGraw-Hill, New York
428. Macalpine, I., and Hunter, R. (1969) *Sci. Am.* **221**(Jul), 38–46
429. Lee, J.-S., and Anvret, M. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10912–10915
430. Poland, A., and Glover, E. (1973) *Science* **179**, 476–477
430a. Bate, R., ed. (1999) *What Risk*, Butterworth-Heinemann, Oxford; Boston
431. Bohle, D. S., Dinnebier, R. E., Madsen, S. K., and Stephens, P. W. (1997) *J. Biol. Chem.* **272**, 713–716
432. Maines, M. D. (1988) *FASEB J.* **2**, 2557–2568
433. Matera, K. M., Takahashi, S., Fujii, H., Zhou, H., Ishikawa, K., Yoshimura, T., Rousseau, D. L., Yoshida, T., and Ikeda-Saito, M. (1996) *J. Biol. Chem.* **271**, 6618–6624
434. Liu, Y., Moënne-Loccoz, P., Loehr, T. M., and Ortiz de Montellano, P. R. (1997) *J. Biol. Chem.* **272**, 6909–6917
434a. Sakamoto, H., Omata, Y., Palmer, G., and Noguchi, M. (1999) *J. Biol. Chem.* **274**, 18196–18200
435. Gray, C. H. (1983) *Trends Biochem. Sci.* **8**, 381–384
435a. Cunningham, O., Dunne, A., Sabido, P., Lightner, D., and Mantle, T. J. (2000) *J. Biol. Chem.* **275**, 19009–19017
436. Yoshinaga, T., Sassa, S., and Kappas, A. (1982) *J. Biol. Chem.* **257**, 7803–7807
437. Zakhary, R., Gaine, S. P., Dinerman, J. L., Ruat, M., Flavahan, N. A., and Snyder, S. H. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 795–798
437a. Doré, S., Takahashi, M., Ferris, C. D., Hester, L. D., Guastella, D., and Snyder, S. H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2445–2450
438. Wilks, A., and Schmitt, M. P. (1998) *J. Biol. Chem.* **273**, 837–841
439. Takahashi, M., Ilan, Y., Chowdhury, N. R., Guida, J., Horwitz, M., and Chowdhury, J. R. (1996) *J. Biol. Chem.* **271**, 26536–26542
440. Nogales, D., and Lightner, D. A. (1995) *J. Biol. Chem.* **270**, 73–77
441. McDonagh, A. F., Palma, L. A., and Lightner, D. A. (1980) *Science* **208**, 145–151
442. Brumm, P. J., Fried, J., and Friedmann, H. C. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3943–3947
443. Cooper, A. J. L. (1983) *Ann. Rev. Biochem.* **52**, 187–222
444. Klaassen, C. D., and Boles, J. W. (1997) *FASEB J.* **11**, 404–418
444a. Ullrich, T. C., and Huber, R. (2001) *J. Mol. Biol.* **313**, 1117–1125
444b. MacRae, I. J., Segel, I. H., and Fisher, A. J. (2001) *Biochemistry* **40**, 6795–6804
444c. Beynon, J. D., MacRae, I. J., Huston, S. L., Nelson, D. C., Segel, I. H., and Fisher, A. J. (2001) *Biochemistry* **40**, 14509–14517
445. Schmidt, A., Abrams, W. R., and Schiff, J. A. (1974) *Eur. J. Biochem.* **47**, 423–434
446. Benci, S., Vaccari, S., Mozzarelli, A., and Cook, P. F. (1997) *Biochemistry* **36**, 15419–15427
447. Cook, P. F., and Wedding, R. T. (1976) *J. Biol. Chem.* **251**, 2023–2029
447a. Soutourina, J., Blanquet, S., and Plateau, P. (2001) *J. Biol. Chem.* **276**, 40864–40872
448. Cooper, A. J. L., Haber, M. T., and Meister, A. (1982) *J. Biol. Chem.* **257**, 816–826
449. Akopyan, T. N., Braunstein, A. E., and Goryachenkova, E. V. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1617–1621
450. Weinstein, C. L., Haschemeyer, R. H., and Griffith, O. W. (1988) *J. Biol. Chem.* **263**, 16568–16579
451. Scriver, C. R. (1986) *N. Engl. J. Med.* **315**, 1155–1156
452. Colonge, M. J., Volpini, V., Bisceglia, L., Rousaud, F., de Sanctis, L., Beccia, E., Zelante, L., Testar, X., Zorzano, A., Estivill, X., Gasparini, P., Nunes, V., and Palacín, M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9667–9671
453. Schneider, J. A. (1985) *N. Engl. J. Med.* **313**, 1473–1474
454. Weinstein, C. L., and Griffith, O. W. (1988) *J. Biol. Chem.* **263**, 3735–3743
455. Wright, C. E., Tallan, H. H., and Lin, Y. Y. (1986) *Ann. Rev. Biochem.* **55**, 427–453
456. Uchida, S., Kwon, H. M., Preston, A. S., and Handler, J. S. (1991) *J. Biol. Chem.* **266**, 9605–9609
457. Lombardini, J. B., Schaffer, S. W., and Azuma, J., eds. (1992) *Taurine Nutritional Value and Mechanisms of Action*, Plenum, New York
458. Pion, P. D., Kittleson, M. D., Rogers, Q. R., and Morris, J. G. (1987) *Science* **237**, 764–768
459. Geggel, H. S., Ament, M. E., Heckenlively, J. R., Martin, D. A., and Kopple, J. D. (1985) *N. Engl. J. Med.* **312**, 142–146
460. Thurston, H. H., Hauhart, R. E., and Naccarato, E. F. (1981) *Science* **214**, 1373–1374
461. Jarabak, R., and Westley, J. (1980) *Biochemistry* **19**, 900–904
462. Falany, C. N. (1997) *FASEB J.* **11**, 206–216
463. Miller, D. M., Delgado, R., Chirgwin, J. M., Hardies, S. C., and Horowitz, P. M. (1991) *J. Biol. Chem.* **266**, 4686–4691
463a. Gliubich, F., Gazerro, M., Zanotti, G., Delbono, S., Bombieri, G., and Berni, R. (1996) *J. Biol. Chem.* **271**, 21054–21061
463b. Bordo, D., Deriu, D., Colnaghi, R., Carpen, A., Pagani, S., and Bolognesi, M. (2000) *J. Mol. Biol.* **298**, 691–704
464. Zheng, L., Cash, V. L., Flint, D. H., and Dean, D. R. (1998) *J. Biol. Chem.* **273**, 13264–13272
465. Mihara, H., Kurihara, T., Yoshimura, T., Soda, K., and Esaki, N. (1997) *J. Biol. Chem.* **272**, 22417–22424
466. Nakai, Y., Yoshihara, Y., Hayashi, H., and Kagamiyama, H. (1998) *FEBS Lett.* **433**, 143–148
466a. Lacourciere, G. M., Mihara, H., Kurihara, T., Esaki, N., and Stadtman, T. C. (2000) *J. Biol. Chem.* **275**, 23769–23773
466b. Lange, H., Kaut, A., Kispal, G., and Lill, R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1050–1055
466c. Clausen, T., Kaiser, J. T., Steegborn, C., Huber, R., and Kessler, D. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3856–3861
466d. Smith, A. D., Agar, J. N., Johnson, K. A., Frazzon, J., Amster, I. J., Dean, D. R., and Johnson, M. K. (2001) *J. Am. Chem. Soc.* **123**, 11103–11104
466e. Krupka, H. I., Huber, R., Holt, S. C., and Clausen, T. (2000) *EMBO J.* **19**, 3168–3178
466f. Ogasawara, Y., Lacourciere, G., and Stadtman, T. C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9494–9498
467. Gentry-Weeks, C. R., Spokes, J., and Thompson, J. (1995) *J. Biol. Chem.* **270**, 7695–7702
468. Urbina, H. D., Silberg, J. J., Hoff, K. G., and Vickery, L. E. (2001) *J. Biol. Chem.* **276**, 44521–44526
469. Krebs, C., Agar, J. N., Smith, A. D., Frazzon, J., Dean, D. R., Huynh, B. H., and Johnson, M. K. (2001) *Biochemistry* **40**, 14069–14080
470. Garland, S. A., Hoff, K., Vickery, L. E., and Culotta, V. C. (1999) *J. Mol. Biol.* **294**, 8–
471. Tong, W.-H., and Rouault, T. (2000) *EMBO J.* **19**, 5692–5700
472. Kaut, A., Lange, H., Diekert, K., Kispal, G., and Lill, R. (2000) *J. Biol. Chem.* **275**, 15955–15961
473. Lutz, T., Westermann, B., Neupert, W., and Herrmann, J. M. (2001) *J. Mol. Biol.* **307**, 815–825
473a. Nakai, Y., Nakai, M., Hayashi, H., and Kagamiyama, H. (2001) *J. Biol. Chem.* **276**, 8314–8320
473b. Schwartz, C. J., Djaman, O., Imlay, J. A., and Kiley, P. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9009–9014
473c. Voisine, C., Cheng, Y. C., Ohlson, M., Schilke, B., Hoff, K., Beinert, H., Marszalek, J., and Craig, E. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1483–1488
473d. Wu, G., Mansy, S. S., Wu, S.-P., Surerus, K. K., Foster, M. W., and Cowan, J. A. (2002) *Biochemistry* **41**, 5024–5032
473e. Wu, G., Mansy, S. S., Wu, S-p., Surerus, K. K., Foster, M. W., and Cowan, J. A. (2002) *Biochemistry* **41**, 5024 – 5032
473f. Nuth, M., Yoon, T., and Cowan, J. A. (2002) *J. Am. Chem. Soc.* **124**, 8774 – 8775
473g. Mihara, H., Kato, S-i, Lacourciere, G. M., Stadtman, T. C., Kennedy, R. A. J. D., Kurihara, T., Tokumoto, U., Takahashi, Y., and Esaki, N. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6679 – 6683
474. Leimkühler, S., and Rajagopalan, K. V. (2001) *J. Biol. Chem.* **276**, 22024–22031
475. White, R. H. (1988) *Biochemistry* **27**, 7458–7462
475a. Tallant, T. C., Paul, L., and Krzycki, J. A. (2001) *J. Biol. Chem.* **276**, 4485–4493
475b. Krum, J. G., Ellsworth, H., Sargeant, R. R., Rich, G., and Ensign, S. A. (2002) *Biochemistry* **41**, 5005 – 5014
476. Barrios, A. M., and Lippard, S. J. (2000) *J. Am. Chem. Soc.* **122**, 9172–9177
477. Rutherfurd, K. J., Rutherfurd, S. M., Moughan, P. J., and Hendriks, W. H. (2002) *J. Biol. Chem.* **277**, 114–119
478. Hans, M., Bill, E., Cirpus, I., Pierik, A. J., Hetzel, M., Alber, D., and Buckel, W. (2002) *Biochemistry* **41**, 5873 – 5882

Study Questions

1. Bacterial glutamine synthetase is feedback inhibited by serine, glycine, and alanine. Explain specifically the connection between these amino acids and glutamine that would account for the logic of this inhibition.

2. Argininosuccinate synthetase catalyzes the following reaction:



a) The reaction as shown is reversible. What metabolic stratagem is employed to drive the reaction from left to right?
b) It has been shown that the oxygen-18 from I (see structure) is transferred to the phosphate group of AMP. Propose a biochemical reaction mechanism to account for the transfer.

3. An organism has been discovered whose urea cycle does not include any reaction with aspartate. Both of the urea nitrogen atoms come directly from ammonia. All other components of the cycle are present and the net reaction is

$$CO_2 + 2\ NH_3 + 3\ ATP \rightarrow \text{urea} + 2\ ADP + 2\ P_i + AMP + PP_i$$

Explain how this is accomplished. Include plausible reaction mechanisms for any *new* steps proposed.

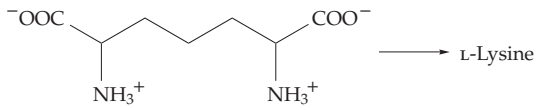
4. Which of the following compounds, if added to an active tissue preparation, might be expected to yield the greatest increase in urea production in terms of moles of urea produced per mole of added compound?

   a) Ammonia
   b) Bicarbonate
   c) Aspartate
   d) Ornithine

More than one answer may be correct. Explain how you decided.

5. On a given diet yielding 2500 kcal per day, a 70-kg man excretes 30.0 g of urea daily. What percentage of his daily energy requirement is met by protein? Assume that 1.0 g of protein yields 4.0 kcal and 0.16 g of nitrogen as urea.

6. In many organisms the immediate biosynthetic precursor of L-lysine is α,ε-diaminopimelic acid (structure below). What type of enzyme would catalyze this reaction; what coenzyme would be required; and what type of enzyme–substrate complex would be formed?



α,ε-Diaminopimelic acid

7. A possible mechanism for the action of urease is pictured in Fig. 16-25 and Eq. 16-47. Carbamate is thought to be one intermediate. Can you suggest an alternative possibility for the initial nickel ion-dependent steps. See Barrios and Lippard.[476]

8. Leucine is known as a "ketogenic" amino acid. Explain what this means.

9. In some organisms leucine is not ketogenic. Why?

10. Here is a possible metabolic reaction for a fungus.

$$\text{L-Leucine} + \text{2-oxoglutarate}^{2-} + 2\ \tfrac{1}{2}\ O_2 \rightarrow \text{L-glutamate}^{-} + \text{citrate}^{3-} + H_2O + 2\ H^+$$

$\Delta G^{\circ\prime}$ (pH 7) = –1026 kJ/mol

Suggest a metabolic pathway for this reaction. Is it thermodynamically feasible?

11. To be practical, the fungus should convert the L-glutamate back to 2-oxoglutarate using a glutamate dehydrogenase. Here are some values for Gibbs energies of formation from the elements under standard conditions (pH = 0).

2-Oxoglutarate^{2-}	–798.0 kJ/mol
L-Glutamate^{-}	–696.8 kJ/mol
NH_4^+	–79.5 kJ/mol

a) Calculate the apparent Gibbs energy change $\Delta G^{\circ\prime}$ (pH 7) for the following reaction:

$$\text{L-Leucine} + 3\ O_2 \rightarrow \text{citrate}^{3-} + NH_4^+ + H_2O + 2\ H^+$$

b) Could this fungal reaction be used as a commercial process for making citric acid?

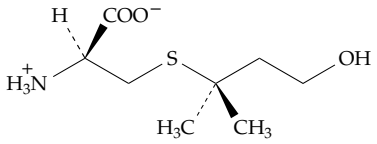
12. Two molecules of pyruvate can react to give a common precursor to valine, leucine, and pantoic acid. An isomerization step involving shift of a methyl group from one carbon to another is involved.

    a) Indicate as much as you can of the pathway for formation of valine showing coenzymes and mechanisms.
    b) Outline the reaction sequence by which the immediate oxoacid precursor of valine is converted to leucine.
    c) Outline a sequence by which the same oxoacid may be converted to pantoic acid using serine as an additional carbon atom source.

13. Explain how carboxylation can be coupled to cleavage of ATP and how this can be used to drive a metabolic sequence.

14. L-Serine is converted to pyruvate + NH_3 by serine dehydratase (deaminase) in a PLP-dependent reaction. However, using the same coenzyme selenocysteine is converted by selenocysteine lyase into L-alanine + elemental selenium Se^0. L-Cysteine may be converted by PLP-dependent enzymes into wither H_2S or into S^0 for transfer into metal clusters. Compare the chemical mechanisms.

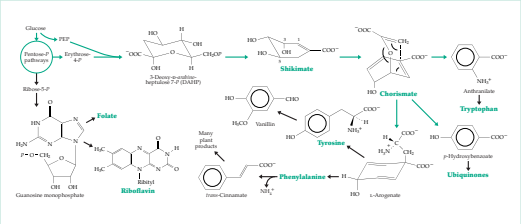
15. Felinine is found in urine of cats, the highest amounts in males. The compound arises from a reaction of glutathione. Propose a route of synthesis. See Rutherfurd *et al.*[477]



Felinine

16. Some clostridia ferment glutamate to ammonia, carbon dioxide, acetate, butyrate, and molecular hydrogen. Write a balanced equation and compare with Eq. 24-18 and Fig. 24-8. See Hans *et al.*[478]

17. How could β-hydroxy-β-methylbutyrate be formed in muscle? Could it be a physiologically important precursor to cholesterol?



Aromatic compounds arise in several ways. The major route utilized by autotrophic organisms for synthesis of the aromatic amino acids, quinones, and tocopherols is the **shikimate pathway**. As outlined here, it starts with the glycolysis intermediate phosphoenolpyruvate (PEP) and erythrose 4-phosphate, a metabolite from the pentose phosphate pathway. Phenylalanine, tyrosine, and tryptophan are not only used for protein synthesis but are converted into a broad range of hormones, chromophores, alkaloids, and structural materials. In plants phenylalanine is deaminated to cinnamate which yields hundreds of secondary products. In another pathway ribose 5-phosphate is converted to pyrimidine and purine nucleotides and also to flavins, folates, molybdopterin, and many other pterin derivatives.

Contents

1421 ..... **A. The Shikimate Pathway of Biosynthesis**
1423 ..... 1. The Enzymes
1424 ..... 2. From Chorismate to Phenylalanine and Tyrosine
1425 ..... 3. Anthranilate, Tryptophan, *para*-Aminobenzoate, and Dihydroxybenzoate
1427 ..... 4. Ubiquinones, Plastoquinones, Vitamin K, and Tocopherols
1428 ..... **B. Metabolism of Phenylalanine and Tyrosine**
1428 ..... 1. Catabolism of Tyrosine in Animals
1430 ..... 2. The Thyroid Hormones
1432 ..... 3. The Catecholamines
1434 ..... 4. The Melanins
1434 ..... 5. Microbial Catabolism of Phenylalanine, Tyrosine, and Other Aromatic Compounds
1438 ..... 6. Quinic and Gallic Acids
1438 ..... 7. The Metabolism of Phenylalanine and Tyrosine in Plants
1438 ..... *Benzoic and salicylic acids*
1440 ..... *Vanilla and other plant products*
1440 ..... *Lignin, lignols, lignans, and phenolic coupling*
1443 ..... 8. Alkaloids
1443 ..... **C. Metabolism of Tryptophan and Histidine**
1443 ..... 1. The Catabolism of Tryptophan
1446 ..... *Formation of NAD^+ and $NADP^+$*
1446 ..... *Auxin*
1447 ..... *Alkaloids from tryptophan*
1449 ..... 2. The Metabolism of Histidine
1449 ..... *Regulation of histidine synthesis*
1450 ..... *Catabolism of histidine*
1450 ..... **D. Biosynthesis and Catabolism of Pyrimidines**
1451 ..... 1. Synthesis of Pyrimidine Nucleotides
1452 ..... *Deoxyribonucleotides*
1452 ..... *Bacteriophage-induced alterations in metabolism*
1452 ..... 2. Catabolism of Pyrimidine Nucleotides and Nucleosides
1453 ..... 3. Reuse or Salvage of Nucleic Acid Bases
1454 ..... **E. Biosynthesis and Metabolism of Purines**
1454 ..... 1. The Enzymatic Reactions of Purine Synthesis
1456 ..... 2. The Purine Nucleotide Cycle and Salvage Pathways for Purines
1459 ..... 3. Oxidative Metabolism of Purines
1460 ..... **F. Pterins, Flavins, Dimethylbenzimidazole, Thiamin, and Vitamin B_6**
1462 ..... *Thiamin*
1463 ..... *Pyridoxol (vitamin B_6)*

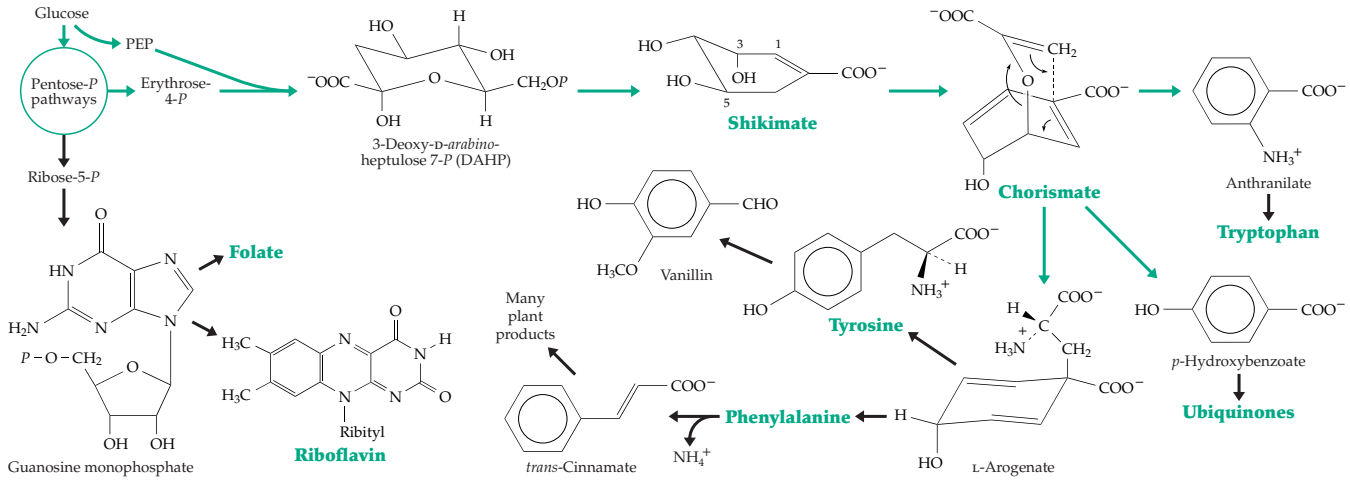
1465 ..... **References**
1471 ..... **Study Questions**

Boxes

1435 ..... Box 25-A Skin Color
1442 ..... Box 25-B The Cyanogenic Glycosides
1448 ..... Box 25-C Royal Purple and Blue Denim
1459 ..... Box 25-D Gout

Metabolism of Aromatic Compounds and Nucleic Acid Bases

25



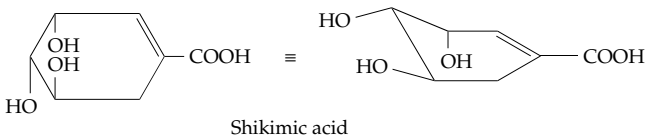
Aromatic rings are found in amino acids, purine and pyrimidine bases, vitamins, antibiotics, alkaloids, pigments of flowers and of the human skin, the lignin present in wood, and in many other substances. There are several biosynthetic pathways. One, which we met in Chapter 21, is the polyketide pathway. However, more important in most autotrophic organisms is the **shikimate pathway** which gives rise to phenylalanine, tyrosine, tryptophan, ubiquinone, plastoquinones, tocopherols, vitamin K, and other compounds.[1] The entire pathway, which is outlined in Fig. 25-1, is present in most bacteria and plants. However, animals are unable to synthesize the ring systems of the aromatic amino acids. Phenylalanine and tryptophan are dietary essentials. Tyrosine can also be formed in the animal body by the hydroxylation of phenylalanine. However, green plants lack phenylalanine hydroxylase and make tyrosine directly through the shikimate pathway (green arrows in Fig. 25-1).

There is interest in applying genetic engineering to increase the output of the shikimate pathway for production of industrially important aromatic compounds, e.g., the dye **indigo**, which is used in manufacture of blue denim (Box 25-C). Study of the enzymes involved has led to the development of potent inhibitors of the shikimate pathway which serve as widely used herbicides.[2,3]

A variety of pathways give rise to the nitrogen- and oxygen-containing heterocyclic rings of nature. All cells must be able to make pyrimidine and purine bases to be used in synthesis of nucleic acids and coenzymes. The pathway for synthesis of pyrimidine begins with aspartic acid and that for purines with glycine. In many organisms the pathway of purine formation is further enhanced because uric acid or a related substance is the major excretory product derived from excess nitrogen. This is true for both birds and reptiles, which excrete uric acid rather than urea, and for spiders which excrete guanine. In some plants, such as soy beans, the transport form of nitrogen is **allantoin** or **allantoic acid**, both of which are produced from uric acid.

A. The Shikimate Pathway of Biosynthesis

The shikimate pathway was identified through the study of ultraviolet light-induced mutants of *E. coli*, *Aerobacter aerogenes*, and *Neurospora*. In 1950, using the penicillin enrichment technique (Chapter 26), Davis obtained a series of mutants of *E. coli* that would not grow without the addition of aromatic substances.[4,5] A number of the mutants required five compounds: tyrosine, phenylalanine, tryptophan, *p*-aminobenzoic acid, and a trace of *p*-hydroxybenzoic acid. It was a surprise to find that the requirements for all five compounds could be met by the addition of shikimic acid, an aliphatic compound that was then regarded as a rare plant acid. Thus, shikimate was implicated as an intermediate in the biosynthesis of the three aromatic amino acids and of other essential aromatic substances.[6,7]



Shikimic acid

The mutants that grew in the presence of shikimic acid evidently had the biosynthetic pathway blocked

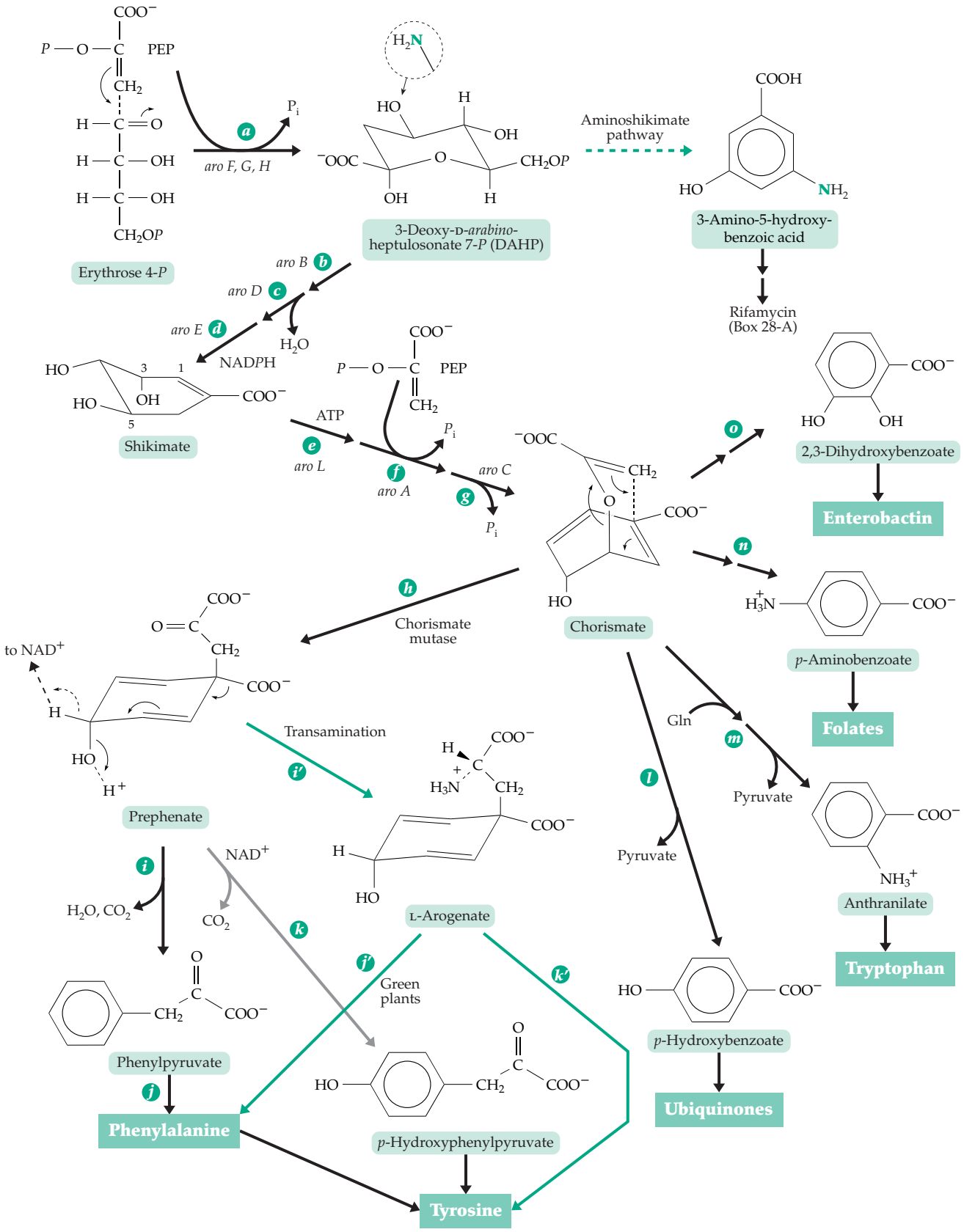
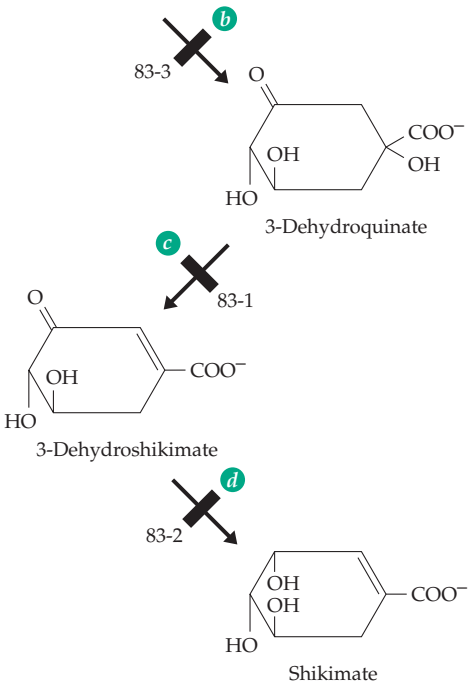


Figure 25-1 Aromatic biosynthesis by the shikimate pathway. The symbols for several of the genes coding for the required enzymes are indicated. Their locations on the *E. coli* chromosome map are shown in Fig. 26-4. The aminoshikimate pathway which is initiated through 4-aminoDAHP leads to rifamycin and many other nitrogen-containing products.

at one or more earlier stages. Among these mutants, certain pairs were found that could not grow alone but that grew when plated together. The phenomenon is called **syntropism**. Mutant 83-2, which we now know to be blocked in the conversion of 5-dehydroshikimate to shikimate, accumulated dehydroshikimate and permitted mutant 83-1 or 83-3 to grow by providing it with a precursor that could be converted on to the end products (Eq. 25-1; the steps in this equation are lettered to correspond to those in Fig. 25-1). Eventually, the entire pathway was traced. The enzymes have all been isolated and studied[1] and the locations of the genes in the *E. coli* chromosome have been mapped[7–10] and are marked in Fig. 26-4.



(25-1)

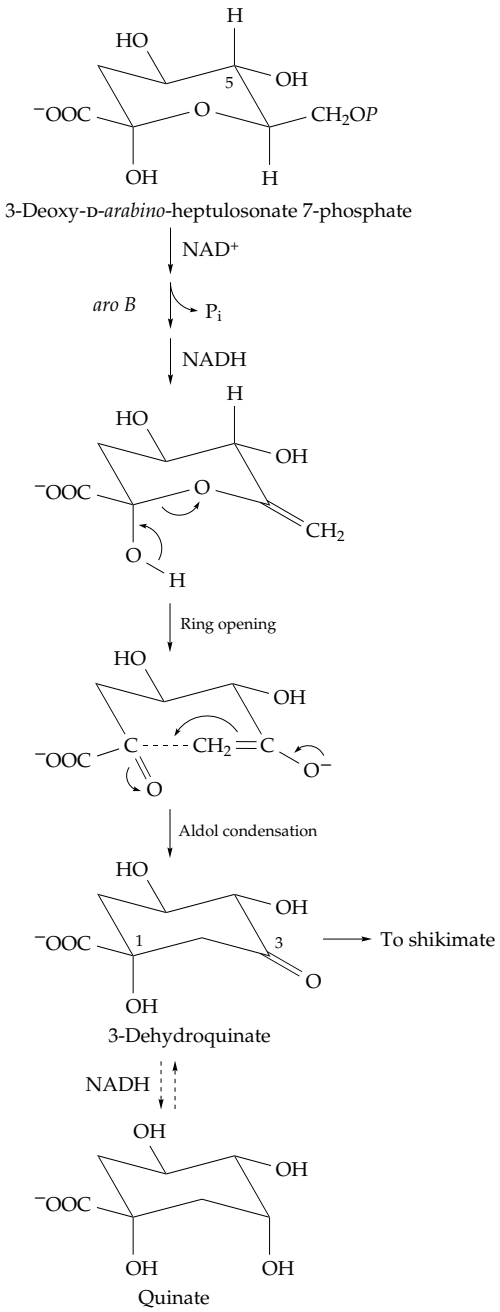
A variety of nitrogen-containing products, including rifamycin (Box 28-A), arise via the **aminoshikimate** pathway,[10a–d] which is also indicated in Fig. 25-1.

1. The Enzymes

The six carbons of the benzene ring of the aromatic amino acids are derived from the four carbons of erythrose 4-phosphate and two of the three carbons of phosphoenolpyruvate (PEP). The initial step in the pathway (Fig. 25-1, step *a*) is the condensation of erythrose 4-*P* with PEP and is catalyzed by 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAHP) synthase. Closely analogous to an aldol condensation, the mechanism provides a surprise.[10] When PEP containing ^{18}O in the oxygen bridge to the phospho group reacts, the ^{18}O is retained in the eliminated phosphate; biochemical intuition would suggest that it should stay in the carbonyl group of the product. See Chapter 20, Section A,5.

Most bacteria and fungi have three isozymes of DAHP synthase, each controlled by feedback inhibition by one of the three products tyrosine, phenylalanine, or tryptophan. In *E. coli* these are encoded by genes *aro F*, *aro G*, and *aro H*, respectively.[11–12a] All of the enzymes contain one atom of iron per molecule and show spectral similarities to hemerythrin.[13]

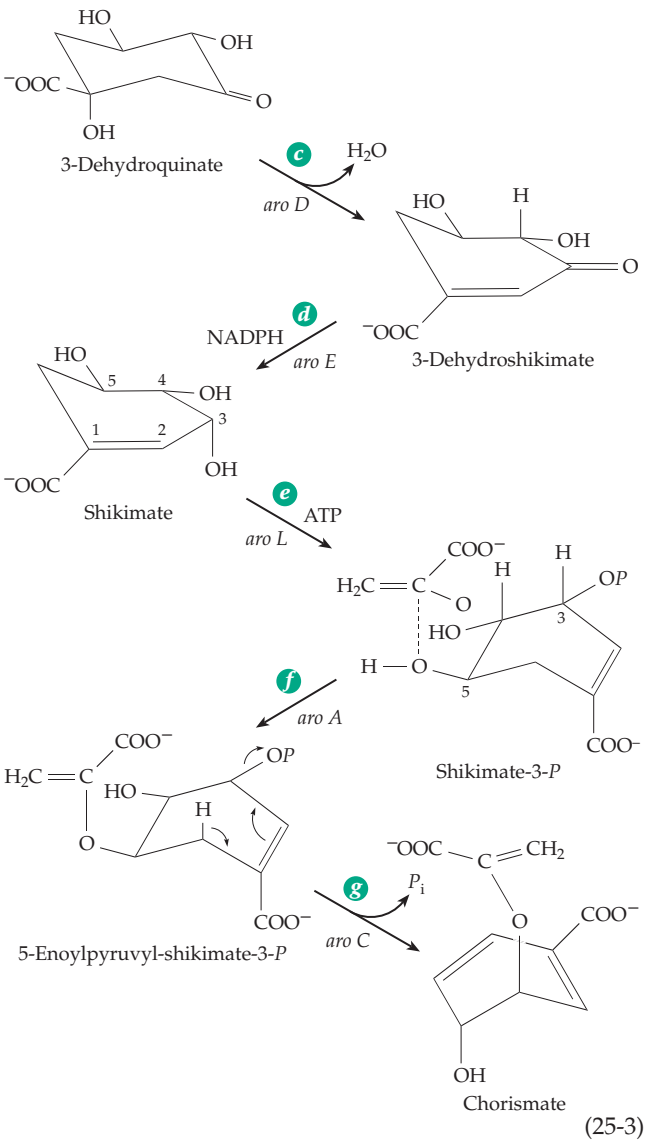
The product of the DAHP synthase, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate, is shown in its cyclic hemiacetal form at the beginning of Eq. 25-2. Its conversion to 3-dehydroquinate is a multistep process that is catalyzed by a single enzyme,[14,15] which is the product of *E. coli* gene *aro B*. The elimination of



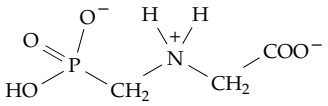
(25-2)

inorganic phosphate in the second step of the sequence is assisted by a transient oxidation of the hydroxyl group at C-5 to a carbonyl group.[6,16,17] The enzyme contains bound NAD^+ and is activated by Co^{2+}. The last step in the dehydroquinate synthase sequence is the addition of an enolic intermediate to a carbonyl group, an aldol condensation, which forms the 6-carbon ring (Eq. 25-2). Also indicated in this equation, with dashed lines, is the reversible conversion of dehydroquinate to quinate. Although it is a "side product" quinic acid accumulates to high concentrations in many plants.[18]

Dehydration of 3-dehydroquinate (step *c*), the first step in Eq. 25-3, is the first of three elimination reactions needed to generate the benzene ring of the end products. This dehydration is facilitated by the presence of the carbonyl group. After reduction of the product to shikimate (step *d*)[19] a phosphorylation reaction (step *e*)[20,21] sets the stage for the future elimination of P_i. In step *f*, condensation with PEP adds three carbon atoms that will become the α, β, and carboxyl carbon atoms of phenylalanine and tyrosine. The reaction occurs by displacement of P_i from the α-carbon atom of PEP and resembles a reaction (Eq. 20-6, step *a*) in the synthesis of *N*-acetylmuramic acid.[22,23] When the reaction is carried out in ^3H-containing water, tritium enters the methylene group,[10,24] suggesting an addition–elimination mechanism (Eq. 25-4).[25] The enzyme 5-enoylpyruvylshikimate 3-phosphate synthase (**EPSP synthase**), which catalyzes this reaction, is strongly inhibited by the commercially important herbicide ***N*-(phosphonomethyl)glycine** (glyphosate).[3,26–27a]



(25-3)



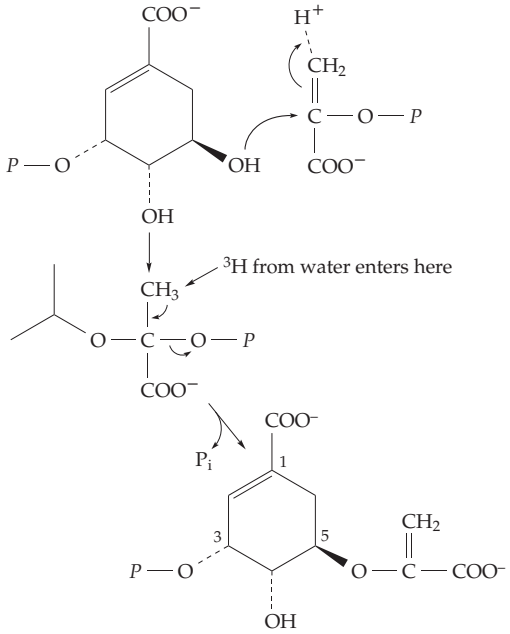
N-(Phosphonomethyl) glycine (glyphosate)

A single mutation (Pro 101 → Ser) in the 427-residue protein from *E. coli* makes the enzyme more resistant to the herbicide.[28] Other mutations affect binding and catalysis.[29]

Elimination of P_i from 5-enolpyruvylshikimate 3-*P* (Eq. 25-3 and Fig. 25-1, step *g*) produces chorismate.[30] The 24-kDa chorismate synthase, which catalyzes this reaction, requires for activity a reduced flavin. Although there is no obvious need for an oxidation–reduction coenzyme, there is strong evidence that the flavin may play an essential role in catalysis, perhaps via a radical mechanism.[31–33b]

2. From Chorismate to Phenylalanine and Tyrosine

Chemical properties appropriate to a compound found at a branch point of metabolism are displayed by chorismic acid. Simply warming the compound in acidic aqueous solution yields a mixture of **prephenate** and ***para*-hydroxybenzoate** (corresponding to reactions *h* and *l* of Fig. 25-1). Note that the latter reaction is a simple elimination of the enolate anion of pyruvate. As indicated in Fig. 25-1, these reactions correspond to only two of several metabolic reactions of the chorismate ion. In *E. coli* the formation of **phenylpyruvate** (steps *h* and *i*, Fig. 25-1) is catalyzed by a single protein molecule with two distinctly different enzymatic activities: **chorismate mutase** and **prephenate dehydratase**.[34–36] However, in some organisms the enzymes are separate.[37] Both of the reactions catalyzed by these enzymes also occur spontaneously upon warming chorismic acid in acidic solution. The chorismate mutase reaction, which is unique in its mechanism,[37a] is discussed in Box 9-E. Stereochemical studies indicate that the formation of phenylpyruvate in Fig. 25-1, step *i*, occurs via a

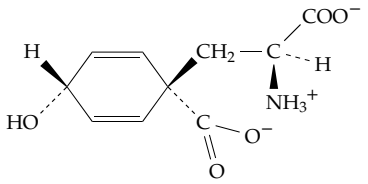


5-Enoylpyruvylshikimate 3-phosphate (25-4)

chair-like transition state.[38] Phenylpyruvate is transaminated to phenylalanine to complete the biosynthesis of that amino acid. Regulation in *E. coli* is accomplished in part by feedback inhibition of the bifunctional chorismate mutase–prephenate dehydratase.[38a]

In *E. coli* and many other bacteria a second bifunctional enzyme, **chorismate mutase-prephenate dehydrogenase** causes the isomerization of chorismate and the oxidative decarboxylation of prephenate to *p*-hydroxyphenylpyruvate (steps *h* and *k*, Fig. 25-1).[39] The latter can be converted by transamination to tyrosine.[40–42]

A slightly different pathway for tyrosine formation was found initially in cyanobacteria but has now been identified in a variety of organisms including higher green plants. In this pathway prephenate undergoes transamination to **L-arogenate** (pretyrosine), step *i*, Fig. 25-1.[43–45] In bacteria L-arogenate is oxidatively



L-Arogenate (pretyrosine)

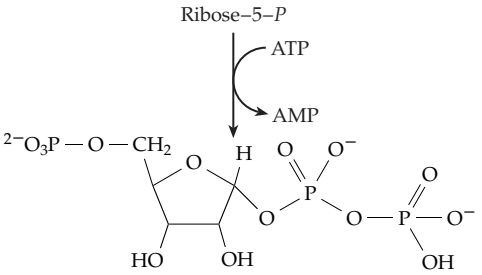
decarboxylated to tyrosine (step *k*, Fig. 25-1). However, in green plants L-arogenate undergoes decarboxylative elimination (step *j*) to give L-phenylalanine. This is a major reaction in green plants, which cannot form tyrosine by hydroxylation of phenylalanine[46,47] but which form a variety of additional products from phenylalanine by a pathway characteristic for plants (Fig. 25-8).

3. Anthranilate, Tryptophan, *para*-Aminobenzoate, and Dihydroxybenzoate

The three chemically similar reactions designated *m*, *n*, and *o* in Figs. 25-1 and 25-2 give rise to a variety of products. Step *m* leads to anthranilate and **L-tryptophan** and step *n* to the vitamin **folic acid.** Each of the three reactions *m*, *n*, and *o* involves addition of either NH_3 or HO^- at a position *ortho* or *para* to the carboxyl group of chorismate with elimination of the 5-OH group. The structures of the three intermediate products are shown in Fig. 25-2.[48–50] The three enzymes have significant similarities in amino acid sequence. Anthranilate synthase and *p*-aminobenzoate synthase are both two-subunit enzymes consisting of a 20-kDa subunit glutamine amidotransferase which is presumed to generate NH_3 (see Chapter 24,B).[49,51–53a] The second 50-kDa subunit of anthranilate synthase catalyzes the remaining steps in the reaction. However, *p*-aminobenzoate synthesis in *E. coli* requires an additional enzyme to catalyze the elimination of pyruvate in the final step[54] of synthesis.

The product of step *o* is known as **isochorismate**.[55,56] Isochorismate gives rise to a variety of products including **vitamin K**, salicylic acid,[56a] the iron chelator **enterobactin** (Fig. 16-1), and other siderophores. These are formed in *E. coli* via 2,3-dihydroxybenzoate as indicated in Fig. 25-2.[57–59a] The genes (*ent*) for the requisite enzymes are clustered at 14 min on the *E. coli* chromosome map (Fig. 26-4).

During the conversion of anthranilate to tryptophan, two additional carbon atoms must be incorporated to form the indole ring. These are derived from **phosphoribosyl pyrophosphate (PRPP)** which is formed from ribose 5-phosphate by transfer of a *pyrophospho* group from ATP.[60,61] The –OH group on the anomeric carbon of the ribose phosphate displaces AMP by attack on P_β of ATP (Eq. 25-5). In many organisms the enzyme that catalyzes this reaction is fused to subunit II of anthranilate synthase.[62] PRPP is also the donor of phosphoribosyl groups for biosynthesis of histidine (Fig. 25-13) and of nucleotides (Figs.



5-Phosphoribosyl 1-pyrophosphate (PRPP) (25-5)

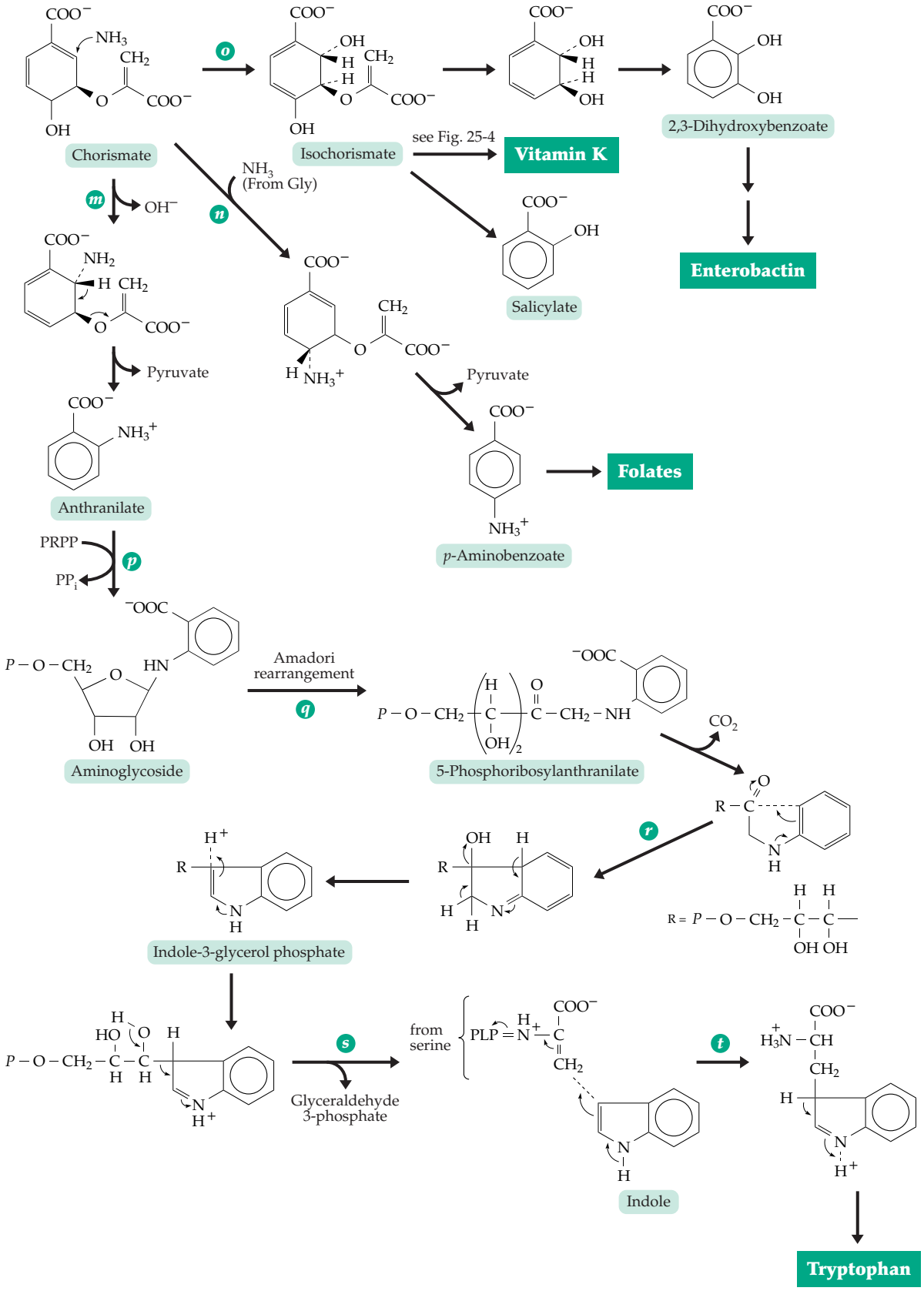


Figure 25-2 The biosynthesis of tryptophan from chorismate and related synthetic reactions.

25-14, 15). In tryptophan biosynthesis PRPP is converted into an aminoglycoside of anthranilic acid by displacement of its pyrophospho group by the amino group of anthranilate (Fig. 25-2, step *p*). The aminoglycoside then undergoes an **Amadori rearrangement** (Eq. 4-8; Fig. 25-2, step *q*). The product has an open chain. Decarboxylation and ring closure, as indicated in this figure, yields **indoleglycerol phosphate**.[63,64]

A β replacement reaction catalyzed by the PLP-dependent **tryptophan synthase** converts indoleglycerol phosphate and serine to tryptophan. Tryptophan synthase from *E. coli* consists of two subunits associated as an $\alpha_2\beta_2$ tetramer (Fig. 25-3). The α subunit catalyzes the cleavage (essentially a reverse aldol) of indoleglycerol phosphate to glyceraldehyde 3-phosphate and free indole (Fig. 25-2, step *s*).[67] The β subunit contains PLP. It presumably generates, from serine, the Schiff base of aminoacrylate, as indicated in Fig. 25-2 (step *t*). The enzyme catalyzes the addition of the free indole to the Schiff base to form tryptophan. The indole must diffuse for a distance of 2.5 nm through a tunnel to the active site where it condenses with the aminoacrylate Schiff base.[65–68c]

The genes encoding the seven enzymes of the tryptophan biosynthetic pathway are organized as a single operon in some bacteria.[69] Its regulation in enteric bacteria is discussed in Chapter 28, Section A,5. The α subunit of tryptophan synthase and the enzymes catalyzing the preceding two steps in tryptophan synthesis are all $(\beta\alpha)_8$-barrel proteins similar to the one shown in Fig. 2-28.[69,70] The biosynthetic pathway for tryptophan in the green plant *Arabidopsis* is the same as in bacteria. The enzymes appear to be present in the chloroplasts.[71]

While the tryptophan synthase of *E. coli* is made up of two different subunits, that of *Neurospora* is a single polypeptide chain. This is one of the first proteins for which it was proposed that there were originally two separate genes, as in *E. coli*, but that they became fused during the course of evolution. After this proposal was made, gene fusion was demonstrated experimentally in *Salmonella* by introduction of two consecutive "frame shift mutations" between two genes of histidine biosynthesis (Chapter 26, Section B,1). Because of the frame shift, the stop signal for protein synthesis is no longer read, with the result that the organism makes a single long protein corresponding to both genes. Gene fusion evidently occurs in nature frequently.[63] There are many instances known in which the two distinctly different catalytic activities are possessed by the same protein in some organisms but by separate enzymes in others. The gene for the α subunit of trytophan synthase in *Salmonella* was of historical importance as it was used to establish the colinearity of the genetic code and its amino acid sequence (Chapter 26, Section B,5).

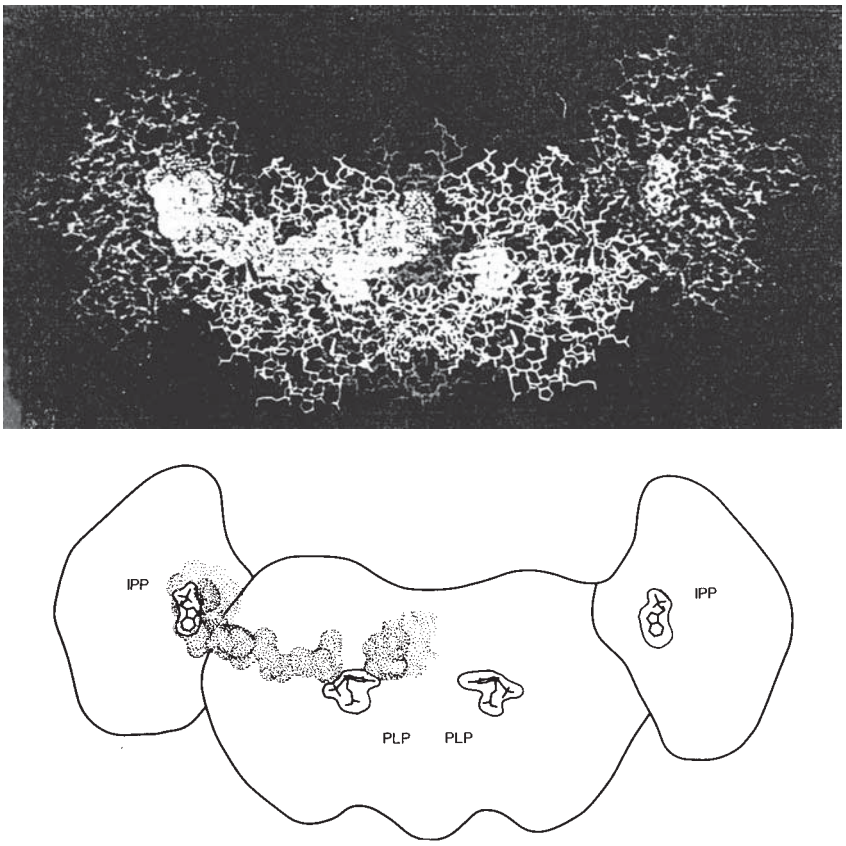


Figure 25-3 The structure of the two-enzyme $\alpha_2\beta_2$ complex tryptophan synthase.[65,66] The view is with the twofold axis of the $\alpha_2\beta_2$ complex vertical with the two α subunits at the ends and the β subunits in the center. The tunnel through which indole molecules released from indole propanol phosphate (IPP) in the α subunits to the pyridoxal phosphate (PLP) in the β subunits is shaded. Courtesy of C. Craig Hyde and Edith Wilson Miles.

4. Ubiquinones, Plastoquinones, Vitamin K, and Tocopherols

Radioactive carbon of [^{14}C]shikimate is efficiently incorporated into the quinones, vitamin K, and tocopherols. These chemically related redox agents (Fig. 15-24) also have a related biosynthetic origin, which has been elucidated in greatest detail for ubiquinone. In bacteria *p*-hydroxybenzoate is formed

directly from chorismate (Fig. 25-1), but in plants it may originate from phenylalanine, *trans*-cinnamate, or *p*-coumarate as indicated in Fig. 25-4. The conversion of *p*-hydroxybenzoate on to the ubiquinones is also shown in this figure. A polyprenyl group is transferred onto a position ortho to the hydroxyl (see Chapter 22). Then a series of consecutive hydroxylation and *S*-adenosylmethionine-dependent transmethylation reactions lead directly to the ubiquinones.[72–73b] Several qui-nones that can serve as precursors to ubiquinones have been isolated from bacteria. Two of the corresponding quinols are shown as intermediates in Fig. 25-4. Chemical considerations suggest that both the methylations and hydroxylations occur on the reduced dihydroxy derivatives. A closely similar pathway is used for synthesis of ubiquinone[74–76] in mitochondria and in the membranes of the endoplasmic reticulum of fungal, plant, and animal cells.[77,78] In bacteria the decarboxylation step occurs early, as shown in Fig. 25-4, whereas in eukaryotes it occurs later.[75] Ubiquinone is poorly absorbed from the blood and it is apparently made in all aerobic tissues.[78] Ubiquinones are thought to serve as important antioxidant compounds in cell membranes. Dietary supplementation may be of value.[73a] Curiously, the nematode *C. elegans* slows its metabolism and lives longer if it has a defect in the hydroxylase catalyzing the next to last step in biosynthesis (Fig. 25-4).[73c] Mutants of *C. elegans* that cannot form their own ubiquinone-9 (containing nine prenyl units in the side chain) are unable to grow on bacteria that make ubiquinone-8. The worms appear to have both essential mitochondrial and nonmitochondrial requirements for ubiquinone.[73d] A ubiquinone deficiency with serious consequences can sometimes be caused in humans by inhibitors of 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase) taken to lower blood cholesterol.[79]

Labeling experiments have shown that the plastoquinones of chloroplasts as well as the tocopherols each bear one methyl group (marked with an asterisk in Fig. 25-4) that originates from chorismate. The dihydroxy compound **homogentisate** is probably an intermediate.[80–83] It is a normal catabolite of tyrosine in the animal body (Fig. 25-5, Eq. 18-49). Both prenylation and methylation by AdoMet are required to complete the synthesis of the plastoquinones and tocopherols. Possible biosynthetic intermediates with one or more double bonds in the polyprenyl side chain have been found in plants and also in fish oils.[83a]

The vitamins K and other naphthoquinones arise from ***O*-succinylbenzoate**[84–86] whose synthesis from chorismate and 2-oxoglutarate depends upon a thiamine diphosphate-bound intermediate, as indicated in Fig. 25-4. Elimination of pyruvate yields *O*-succinylbenzoate. The remaining reactions of decarboxylation, methylation, and prenylation (Fig. 25-4) resemble those of ubiquinone synthesis.

B. Metabolism of Phenylalanine and Tyrosine

Figure 25-5 shows the principal catabolic pathways, as well as a few biosynthetic reactions, of phenylalanine and tyrosine in animals. Transamination to phenylpyruvate (reaction *a*) occurs readily, and the product may be oxidatively decarboxylated to phenylacetate. The latter may be excreted after conjugation with glycine (as in Knoop's experiments in which phenylacetate was excreted by dogs after conjugation with glycine, Box 10-A). Although it does exist, this degradative pathway for phenylalanine must be of limited importance in humans, for an excess of phenylalanine is toxic unless it can be oxidized to tyrosine (reaction *b*, Fig. 25-5). Formation of phenylpyruvate may have some function in animals. The enzyme **phenylpyruvate tautomerase**, which catalyzes interconversion of enol and oxo isomers of its substrate, is also an important immunoregulatory cytokine known as **macrophage migration inhibitory factor**.[86a]

The pterin-dependent hydroxylation of phenylalanine to tyrosine (Eq. 18-45)[87,87a] has received a great deal of attention because of the occurrence of the metabolic disease **phenylketonuria (PKU)**,[88–91b] in which this reaction is blocked. Infants born with phenylketonuria appear normal but mental retardation sets in rapidly. However, if these infants are identified promptly and are reared on a low-phenylalanine diet which supplies only enough of the amino acid for essential protein synthesis, most brain damage can be prevented. Throughout most of the world every infant born is now given a simple urine test to identify phenylketonuria. Tolerance to phenylalanine increases with age, and adults may return to a near normal diet. However, there may still be problems with increased phenylalanine levels during fever and infections. A high phenylalanine level during pregnancy may damage the unborn child. Temporary insertion of multitubular reactors containing phenylalanine-ammonia lyase (Eq. 14-45) can be of value[92] as is administration of the enzyme in encapsulated form.[93] The mechanism by which phenylalanine damages the brain is uncertain.

1. Catabolism of Tyrosine in Animals

The major route of degradation of tyrosine in animals begins with transamination (Fig. 25-5, reaction *c*) to ***p*-hydroxyphenylpyruvate**. The enzyme tyrosine aminotransferase[94] is induced in the liver in response to the action of glucorticoid hormones (Chapter 22). The synthesis of the enzyme is also controlled at the translational level, release of the newly formed protein from liver ribosomes being stimulated by cyclic AMP. The enzyme is subject to posttranscriptional

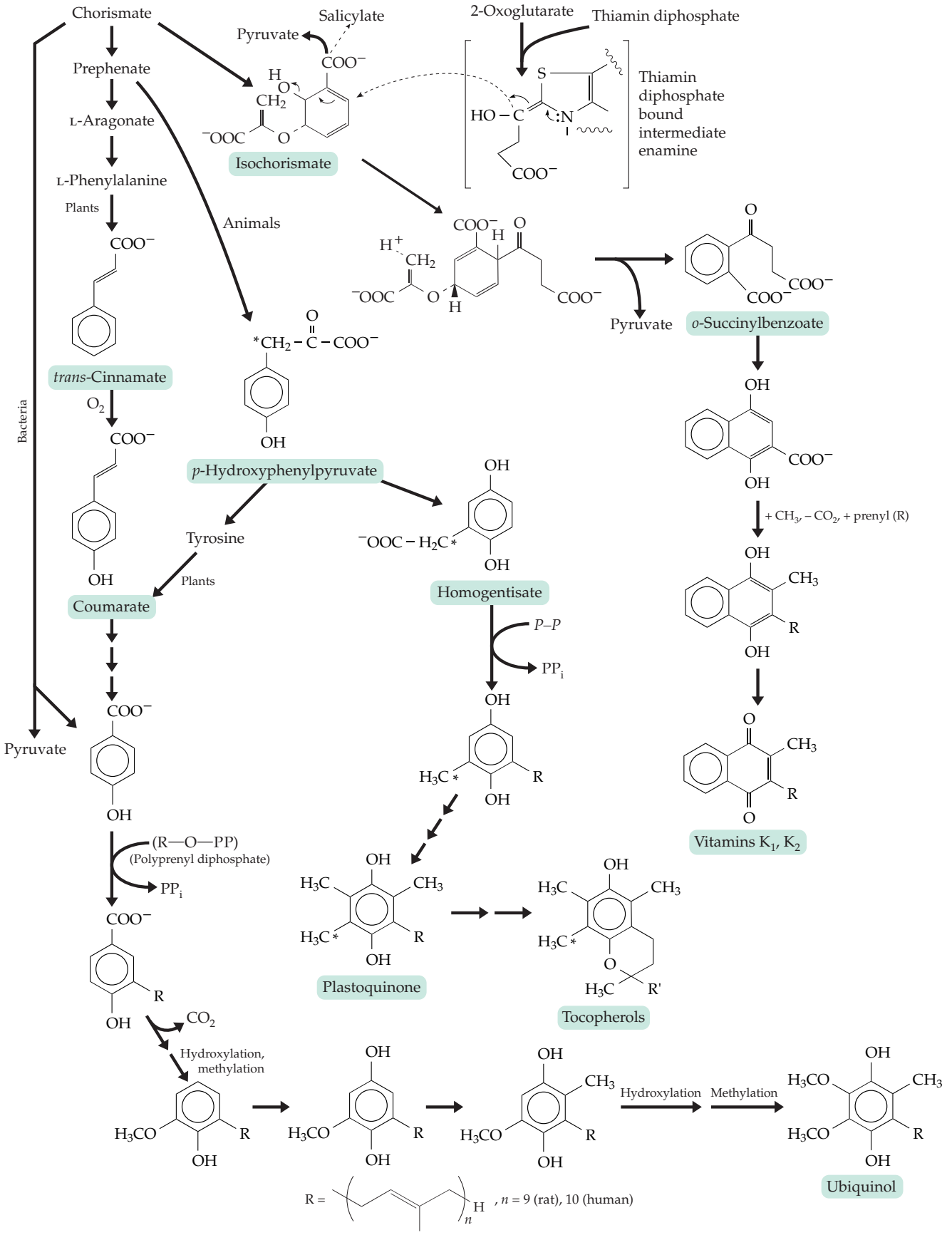
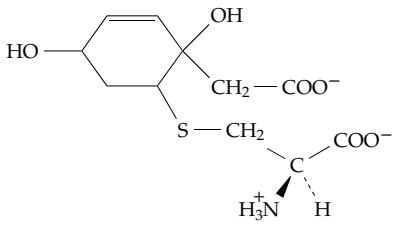


Figure 25-4 Pathways of biosynthesis of ubiquinones, plastoquinones, tocopherols, and vitamin K.

modification including phosphorylation and it undergoes unusually rapid turnover.[95,96]

The 2-oxoacid *p*-hydroxyphenylpyruvate is decarboxylated by the action of a dioxygenase (Eq. 18-49). The product **homogentisate** is acted on by a second dioxygenase, as indicated in Fig. 25-5, with eventual conversion to fumarate and acetoacetate. A rare metabolic defect in formation of homogentisate leads to tyrosinemia and excretion of **hawkinsin**[97] a compound postulated to arise from an epoxide (arene oxide) intermediate (see Eq. 18-47) which is detoxified by a glutathione transferase (Box 11-B).

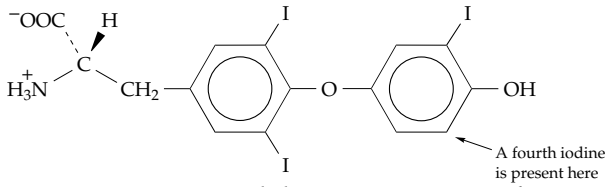


Hawkinsin

One of the first "inborn errors of metabolism" to be recognized was **alkaptonuria**, a lack of the oxygenase that cleaves the ring of homogentsic acid.[98] The condition is recognized by a darkening of the urine upon long standing (caused by oxidation of the homogentisate). Alkaptonuria was correctly characterized by Garrod (Box 1-D) in 1909 as a defect in the catabolism of tyrosine. Although relatively mild and not reducing the life span of individuals, it is nearly always accompanied by arthritis in later years and also by gray to bluish-black pigmentation of connective tissues, which may be visible through the shin or in the scleras (whites) of the eyes.[98] Absence of the next enzyme in the pathway, **maleylacetoacetate isomerase** causes one type of tyrosinemia.[99,99a] Absence of **fumarylacetoacetate hydrolase**, which acts on the product of the isomerase action causes the severe **type 1 hereditary tyrosinemia** which leads to accumulation of the toxic fumarylacetoacetate and its decarboxylation product succinylacetoacetate.[99,100]

2. The Thyroid Hormones

An important product of tyrosine metabolism in vetebrates is the thyroid hormone[101] of which the principal and most active forms are **thyroxine** (T_4) and **triiodothyronine** (T_3).[102] The thyroid gland is rich in iodide ion, which is actively concentrated from the plasma to ~1 μM free I^-.[103] This iodide reacts under the influence of a peroxidase (see Fig. 16-14 and accompanying discussion)[104] to iodinate tyrosyl residues of the very large ~660-kDa dimeric **thyroglobulin**, which is stored in large amounts in the lumen of the thyroid follicles.[105] Several of the tyrosine side chains (up to 15–25) are iodinated to form **mono-** and **di-iodotyrosine** residues (Eq. 25-6), but only between four and eight of these, which are located at specific positions, are converted on to the hormones.[106,107]



Triiodothyronine

The coupling reaction by which the aromatic group from one residue of mono- or diiodotyrosine is joined in ether linkage with a second residue is also catalyzed readily by peroxidases. One dehydroalanine residue is formed for each molecule of hormone released.[108] A possible mechanism involves formation of an electron-deficient radical, which can undergo β elimination to produce a dehydroalanine residue and an aromatic radical. The latter could couple with a second radical to form triiodothyronine or thyroxine. However, as depicted in Eq. 25-6, the radical coupling may occur prior to chain cleavage. While β elimination (pathway A) has been favored, recent evidence suggests hydroxylation and cleavage to form a residue of aminomalonic semialdehyde in the thyroglobulin chain (pathway B).[108a] Alternatively, a PLP-dependent elimination of the radical could be used. Another possibility is oxidative attack on the 2-oxoacids derived from the iodotyrosines.

Thyroxine and triiodothyronine are released from thyroglobulin through the action of a series of proteases. Both the protease action and the release of the thyroid hormones into the bloodstream are stimulated by pituitary **thyrotropin (TSH)**.[109,110] Like glucagon thyrotropin is released from the pituitary in response to **thyrotropin-releasing hormone.**[111] Thyrotropin probably acts through cAMP-mediated mechanisms.[112] The hormones are carried throughout the body while bound to **thyroxine-binding globulin,** which serves as a carrier.[113] Some hormone is carried by other serum proteins such as **transthyretin** (thyroxine-binding prealbumin).[113,114] Both thyroxine and triiodothyronine have powerful hormonal effects on tissues, but the lag time for a response is shortest for triiodothyronine. Thus, it is thought that thyroxine undergoes loss of one iodine atom to form the more active triiodo form of the hormone within the target cells. Three **iodothyronine deiodinases**, all of which are selenoproteins, have been identified (Eq. 15-60).[115–116a]

Organically bound iodine is found in various invertebrates, but with one possible exception

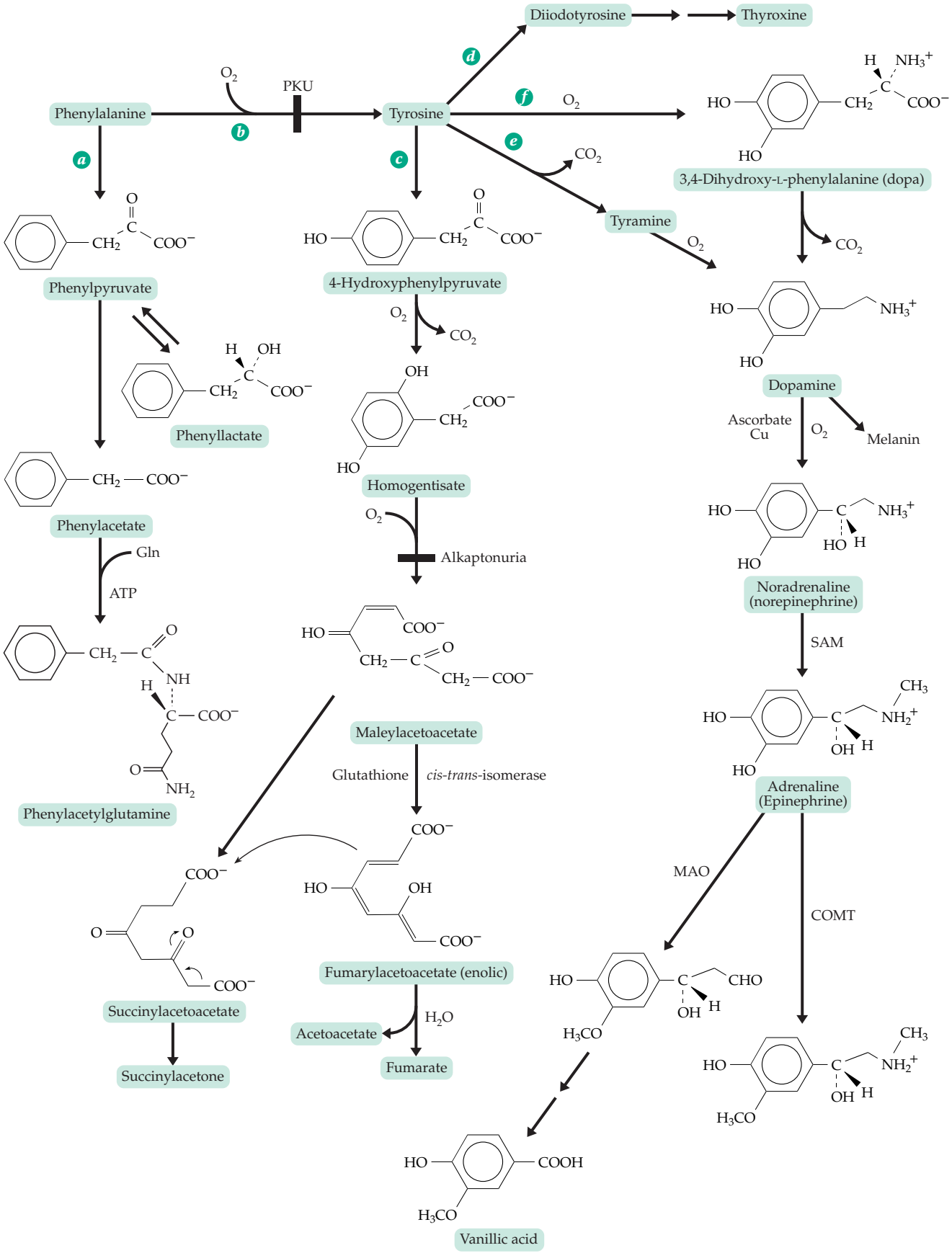
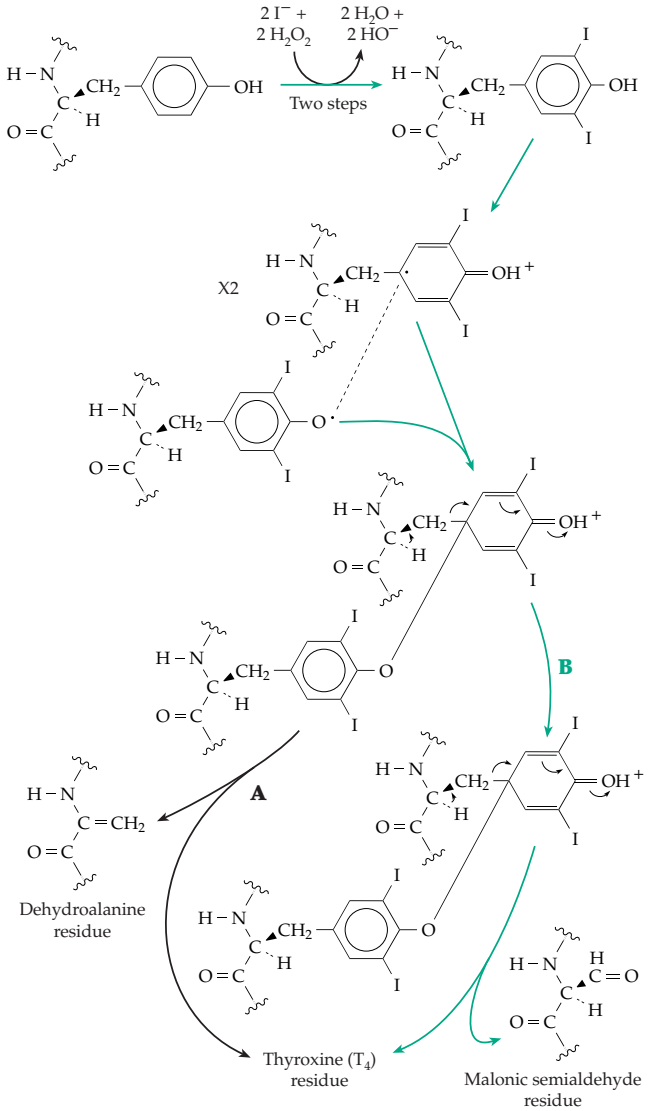


Figure 25-5 Some routes of metabolism of phenylalanine and tyrosine in animals.



(25-6)

thyroxine is present only in vertebrates.[101] Why do we need this iodinated hormone? Halogen-free analogs in which the iodine atoms of thyroxine and triiodothyronine have been replaced by methyl or other alkyl groups are biologically active. Frieden[101] concluded that the role of iodine is related more to the evolution of biosynthetic and catabolic pathways than to chemical properties of the hormones themselves.

Thyroxine and triiodothyronine have many effects, a major one in mammals and birds being stimulation of energy metabolism in tissues. It has long been recognized that a deficiency of thyroid hormone is reflected in an overall lower basal metabolic rate (Chapter 6). Maley and Lardy observed that thyroxine uncouples oxidative phosphorylation (Chapter 18) in isolated mitochondria.[117] When mitochondria from animals receiving extra thyroxine were compared with those from control animals, an increased rate of electron transport was observed. However, there was little or no change in the P / O ratio. Thus, the hormone apparently increased the rate of electron transport without decreasing the overall efficiency of ATP synthesis.

Thyroid hormones also have a general effect on growth and development in all vertebrates. This is especially striking in amphibia in which thyroid hormones control the metamorphosis from the tadpole to adult stages.[101,118] Zebrafish, and presumably also other fishes, require thyroid hormone to complete their life cycles.[119] At present it is thought that most, if not all, effects of thyroid hormones are a result of their action on the transcription of genes.[120–123] Nuclear **thyroid hormone receptors** belong to a family of ligand-regulated transcription factors that respond to steroid, retinoid, and thyroid hormones (Table 22-1).[123–124] These proteins control many metabolic functions, often forming heterodimers with other receptors and also being activated by coactivators[125] or corepressors.[124,125a,b] Transcription of the genes for the thyroxine-synthesis proteins **thyroperoxidase**, **thyroglobulin**, and **iodide transporter**[125c] is regulated by a **thyroid transcription factor**.[126]

A number of thyroid-related diseases are known. Thyroid deficiency is often evident by enlargement of the thyroid gland (**goiter**). The deficiencies may involve inadequacy in dietary intake of iodine, transport of iodide into the thyroid, poor formation of iodinated thyroglobulin, inefficient coupling to form the iodinated thyronine residues,[127] or mutations in thyroid hormone receptors.[122,128] A major cause of goiter is a deficiency in the content of iodine in soil, a condition affecting about one billion (10^9) persons. A more severe effect of thyroid deficiency is the fetal brain damage called **cretinism**.[125c,129] Victims are mentally retarded, deaf-mute, and often with motor rigidity. In Grave disease, the commonest type of **hyperthyroidism**, the blood contains specific thyroid-stimulating autoantibodies.[130,131] These bind to the thyrotropin (TSH) receptors of the thyroid plasma membrane and stimulate excessive formation of thyroid hormone.

3. The Catecholamines

A combination of decarboxylation and hydroxylation of the ring of tyrosine produces derivatives of *o*-dihydroxybenzene (catechol), which play important roles as neurotransmitters and are also precursors to **melanin**, the black pigment of skin and hair. Catecholamines may be formed by decarboxylation of tyrosine into tyramine (step *e*, Fig. 25-5) and subsequent oxidation. However, the quantitatively more important route is hydroxylation by the reduced pterin-dependent tyrosine hydroxylase (Chapter 18) to 3,4-dihydroxyphenylalanine, better known as **dopa**. The latter is decarboxylated to **dopamine**.[131a] Hydroxylation of dopamine by an ascorbic acid and

copper-requiring enzyme (Eq. 18-53) produces the important hormone **noradrenaline** (norepinephrine), which is methylated to form **adrenaline** (epinephrine).

There are two principal catabolic routes for destruction of these catecholamines as is illustrated for adrenaline in Fig. 25-5. **Monoamine oxidase** (MAO)

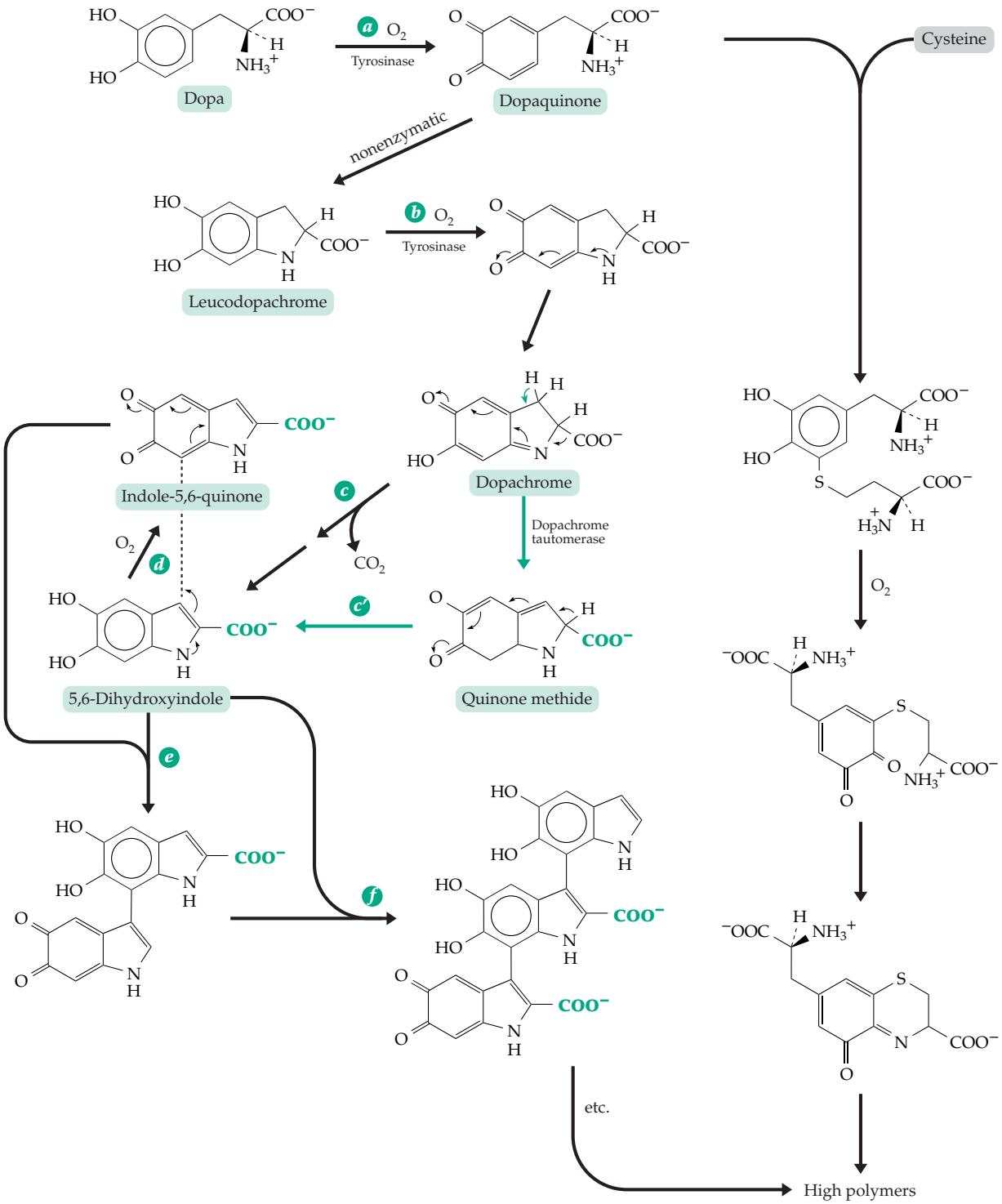


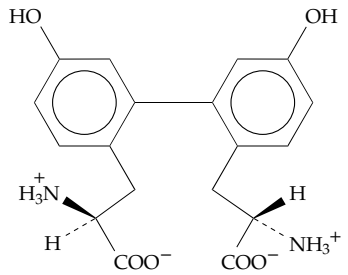
Figure 25-6 Postulated pathways for synthesis of the black pigment melanin and pigments (phaeomelanins) of reddish hair and feathers. Dopachrome reacts in two ways, with and without decarboxylation. The pathway without decarboxylation is indicated by green arrows. To the extent that this pathway is followed the green carboxylate groups will remain in the polymer. The black eumelanin is formed by reactions at the left and center while the reddish phaeomelanin is derived from polymers with cysteine incorporated by reactions at the right.

causes oxidative cleavage with deamination. Subsequent oxidative fission of the side chain together with methylation yields such end products as vanillic acid, which is excreted in the urine. The second catabolic route is immediate O-methylation by **catecholamine *O*-methyltransferase** (COMT), an active enzyme in neural tissues. The metabolites are relatively inactive physiologically and may be secreted as such or may undergo further oxidative degradation.

4. The Melanins

Dihydroxyphenylalanine (dopa) darkens rapidly when exposed to oxygen. The process is hastened greatly by tyrosinase (Chapter 16), which also catalyzes reaction *f* of Fig. 25-5, the oxidation of tyrosine to dopa. Tyrosinase is found in animals only in the organelles known as **melanosomes**, which are present in the melanin-producing **melanocytes** (Boxes 8-F; 25-A).[132] A series of enzymatic and nonenzymatic oxidation, decarboxylation, and coupling reactions forms the pigments. The initial steps for one pathway are indicated in Fig. 25-6. Oxidation of dopa to dopaquinone (step *a*)[133,134] is followed by an intramolecular addition reaction, together with tautomerization to the indole derivative, leucodopachrome. A second oxidation by tyrosinase (step *b*) is followed by decarboxylation and tautomerization to 5,6-dihydroxyindole (step *c*).[135] Alternatively, the tautomerization steps may take place without decarboxylation (green arrows, step *c′*).[133,136] In either case the dihydroxyindole that is formed can undergo a third oxidation step, also catalyzed by tyrosinase, to form indole-5,6-quinone (step *d*). Coupling of the last two products as indicated in step *e* yields a dimer, which is able to continue the addition of dihydroxyindole units (step *f*, etc.) with oxidation to form a high polymer of the black true melanin (**eumelanin**). However, the structure is not regular and is crosslinked. A related series of red polymers, the **phaeomelanins** found in red hair and feathers, are formed by addition of cysteine to dopaquinone.[137,138] Addition is possible at more than one position. The resulting adducts (only one is shown) can undergo oxidative ring closure in the manner indicated. Control of melanin formation is also complex. For example, more than 50 genetic loci affect the coat color of the house mouse.[139] Melanin in some fungi is formed by oxidative reactions of tetrahydroxynaphthalene formed via the polyketide pathways (Chapter 21).[140,140a] The melanin "inks" produced by cuttlefish and other cephalopods are formed in much the same way as melanins of skin.[140a]

Dopa is converted by at least some insects into *N*-β-alanyldopamine, which is a preferred substrate for the *o*-diphenol oxidase of the insect pupal cuticle. Oxidation of this substrate plays a crucial role in the hardening and darkening of the cuticle during pupal tanning.[141–142] There are many other oxidative reactions of tyrosine side chains within proteins. These include coupling of free radicals formed by peroxidases[143] or ultraviolet light[144,144a] to form dityrosines and other products. The walls of yeast ascospores,[145,146]



Dityrosine
Other isomers are also formed

the cements formed by reef-building annelids,[147,148] and adhesive plaques of marine mussels[149,149a] all contain polyphenolic proteins. The 120-kDa "foot protein" of the mussel *Mytilus edulis* consists of tandemly repeated decapeptides, each containing 2 residues of lysine, 1–2 residues of dopa, 1–2 residues of *trans*-4-hydroxyproline, and 1 residue of *trans*-2,3, *cis*-3,4-dihydroxyproline.[149]

5. Microbial Catabolism of Phenylalanine, Tyrosine, and Other Aromatic Compounds

Bacteria and fungi play an essential role in the biosphere by breaking down the many aromatic products of plant metabolism.[150–153] These include vast amounts of lignin, alkaloids, flavonoid compounds, and other biochemically "inert" substances. Lignin is a major constituent of wood and a plant product second only to cellulose in abundance.

The chemical reactions used to degrade these aromatic compounds are numerous and complex. As was mentioned in Chapter 16, some fungi initiate the attack on lignin with peroxidases and produce soluble compounds that can be attacked by bacteria. In other cases elimination reactions may be used to initiate degradation. For example, some bacteria release phenol from tyrosine by β elimination (Fig. 14-5). However, more often hydroxylation and oxidative degradation of side chains lead to derivatives of benzoic acid or of the various hydroxybenzoic acids.[150,151,154–155a]

After the initial reactions many of the compounds are channeled into one of the major pathways illustrated in Fig. 25-7.[151,156,157] Dominant in aerobic bacteria is the conversion to **3-oxodipate** by one of the two convergent pathways shown. The products succinate and acetyl-CoA are readily oxidized by the citric

BOX 25-A SKIN COLOR

The principal pigment of human skin, hair, and eyes is **melanin**, which is synthesized in specialized cells, the **melanocytes**. They lie between the epidermis (outer layer) and the dermis (inner layer) as shown in Box 8-F. Melanocytes originate from embryonic nervous tissue and migrate into the skin by the third month of fetal life. They retain the highly branched morphology of neurons. Persons of different races all have the same numbers of melanocytes but the numbers and sizes of the pigmented melanosomes (Box 8-F) vary as does the content and chemical composition of the melanin.[a–d] Melanosomes not only are found in the dendrites of the melanocytes but are transferred from them into adjacent epithelial cells.[e,f]

Nevi (moles) are clusters of melanocytes that start to appear in the third year of life. They gradually increase in numbers but disappear in old age. **Freckles** appear beginning at about age six in genetically susceptible individuals. They are regions in which a higher concentration of melanin is formed.[a]

Both hair and the iris of the eye are also pigmented by melanin. Although dark eyes and dark hair are more prevalent among persons with dark skin there is no direct correlation. This is only one piece of evidence that the genetics of skin, eye, and hair coloration is complex. In mice over 150 different mutations occurring at more than 50 distinct genetic loci affect pigmentation.[f] Melanin formation begins with the action of tyrosinase. The human genome contains at least three genes for tyrosinase and related proteins.[b,g] The *Tyr* gene is absent in **oculocutaneous albinism**, the lack of pigment in eyes, hair, and skin. The tyrosinase-related protein 2 (TRP2), which has been identified as **dopachrome tautomerase** (see Fig. 25-6), is also a member of the tyrosinase family. Although a key enzyme in pigment synthesis, the amount of tyrosinase or of tyrosinase mRNA is the same in all skin types and colors.[h] Thus, differences in skin color must arise from differences in regulation.

Regulation of melanin formation is achieved in part by hormones, the **melanocyte-stimulating hormone** (MSH or melanotropin) being the most important.[a,f,g] The 13-residue pituitary hormone greatly increases pigmentation and stimulates differentiation of melanocytes. Other regulatory influences arise from interleukins, prostaglandins, interferons, tetrahydrobiopterins,[h] and protein kinase C.[i] Light also has a major effect, causing rapid tanning, especially in darker skin. Release of NO and cyclic GMP may be involved.[j]

Melanin and phaeomelanins have an important role in protecting skin from sunlight. This includes protection of light-sensitive vitamins, proteins, and DNA and RNA. The correlation of high pigmentation with the high intensity of light in tropical regions may reflect this property. Light-skinned persons of northern and southern latitudes, where light intensity is weaker, are less pigmented, allowing more adequate synthesis of vitamin D in the skin (Box 22-C).

A total lack of melanin as a result of a defective *Tyr* gene is seen in oculocutaneous albinism. Lacking protection from sunlight by melanin, albino individuals must shield their skin and eyes carefully. A second type of albinism results from mutations in the P gene, known in the mouse as the "pink-eyed dilution locus." In this condition synthesis of phaeomelanin is not impaired. Mutations in the *KIT* gene, which encodes a tyrosine kinase receptor lead to **piebaldism**, with white and dark splotched skin (or fur in animals).[b,k] While piebaldism is hereditary, **vitiligo** is an acquired autoimmune disease involving spotty loss of pigment and affecting 0.5 to 4% of the world's population. Melanocytes may be present in the affected areas but are unable to make melanin.[h,l]

---

[a] Lerner, A. B. (1961) *Sci. Am.* **205**(Jul), 98–108

[b] King, R. A., Hearing, V. J., Creel, D. J., and Oetting, W. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 4353–4392, McGraw-Hill, New York

[c] Molnar, S. (1998) *Human Variation, Races, Types and Ethnic Groups*, 4th ed., Prentice Hall, Upper Saddle River, New Jersey (pp. 230–247)

[d] Robins, A. H. (1991) *Biological Perspectives on Human Pigmentation*, Cambridge Univ. Press, Cambridge

[e] Potterf, S. B., Muller, J., Bernardini, I., Tietze, F., Kobayashi, T., Hearing, V. J., and Gahl, W. A. (1996) *J. Biol. Chem.* **271**, 4002–4008

[f] Hearing, V. J., and Tsukamoto, K. (1991) *FASEB J.* **5**, 2902–2909

[g] Aroca, P., Urabe, K., Kobayashi, T., Tsukamoto, K., and Hearing, V. J. (1993) *J. Biol. Chem.* **268**, 25650–25655

[h] Schallreuter, K. U., Wood, J. M., Pittelkow, M. R., Gütlich, M., Lemke, K. R., Rödl, W., Swanson, N. N., Hitzemann, K., and Ziegler, I. (1994) *Science* **263**, 1444–1446

[i] Park, H.-Y., Russakovsky, V., Ohno, S., and Gilchrest, B. A. (1993) *J. Biol.Chem.* **268**, 11742–11749

[j] Roméro-Graillet, C., Aberdam, E., Biagoli, N., Massabni, W., Ortonne, J.-P., and Ballotti, R. (1996) *J. Biol. Chem.* **271**, 28052–28056

[k] Schmidt, A., and Beermann, F. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4756–4760

[l] Overwijk, W. W., Lee, D. S., Surman, D. R., Irvine, K. R., Touloukian, C. E., Chan, C.-C., Carroll, M. W., Moss, B., Rosenberg, S. A., and Restifo, N. P. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2982–2987

acid cycle and associated reactions. Many different compounds can be converted into the starting compounds shown at the top of the figure. Both D- and L-mandelate, toluene, benzyl alcohol, L-tryptophan, phenanthrene, naphthalene, and benzene can be converted to catechol and be metabolized via the catechol branch of the pathway. Benzoate, *p*-toluate, shikimate, and quinate can be metabolized via the protocatechuate branch. Halogenated compounds, e.g., 3-chlorocatechol, may sometimes be degraded via

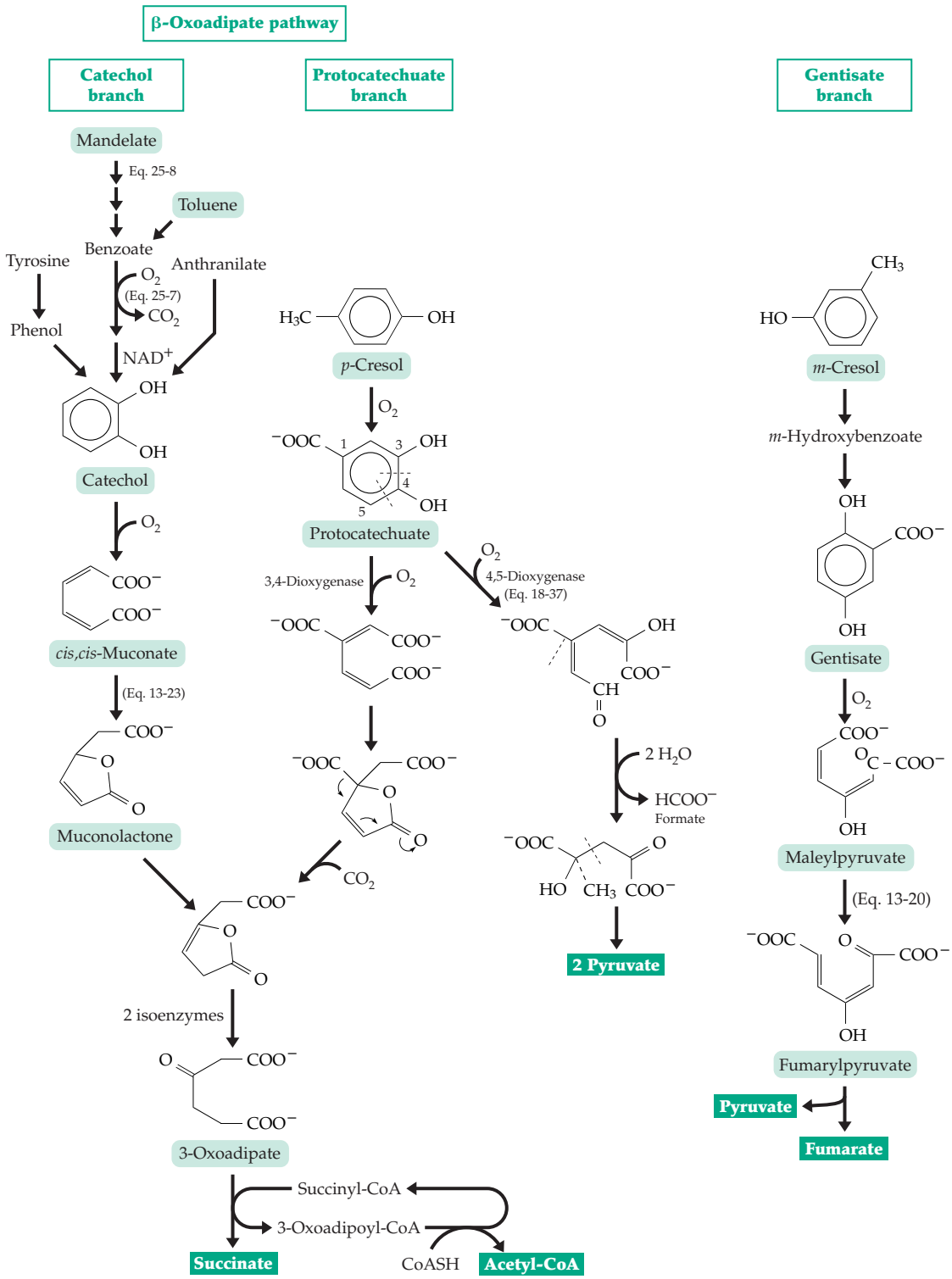
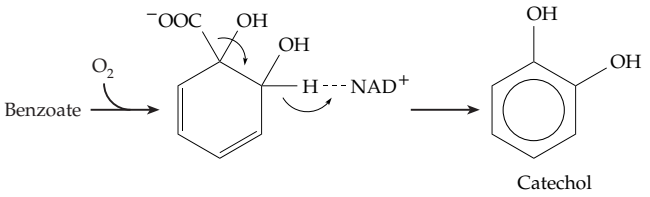


Figure 25-7 A few pathways of catabolism of aromatic substances by bacteria.

the same pathways[151,156] or in parallel or related pathways.[157–158c] However, chlorine atoms are sometimes eliminated as Cl^- at some point in the process.[159–160a] Other substrates, including *m*-hydroxybenzoate and sometimes anthranilate, are degraded via the gentisate pathway (Fig. 25-7).[157,161] Both benzoate and phenylacetate are sometimes degraded after conversion to coenzyme A thioesters.[161a,b]

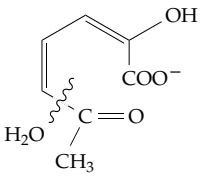
Dioxygenases play a major role in all of these pathways of aromatic catabolism. In most cases a dioxygenase (Chapter 18) is required for the opening of the benzene ring. The pathways contain interesting isomerization steps, some of which have been discussed in Chapter 13, Section B. Microorganisms often have alternative choices in the chemistry of their attack. For example, tyrosine can be converted by one bacterium to homogentisate, as in animals (Fig. 25-5), or by other bacteria to protocatechuate, homoprotocatechuate, or gentisate (Fig. 25-7) before the ring is opened.[150] A single compound can be acted upon by more than one dioxygenase. Thus protocatechuate can be opened by a 3,4-dioxygenase or by a 4,5-dioxygenase (Fig. 18-22) leading to the branch point at protocatechuate in Fig. 25-7.

The initial hydroxylation of benzene, toluene, and other alkylbenzenes is accomplished by multicomponent aromatic ring dioxygenases that introduce two oxygen atoms to form diols.[158] Dioxygenation of benzoate yields a diol that can be **oxidatively decarboxylated** by reaction with NAD^+ (Eq. 25-7) to form catechol.[157,162] Toluene gives 3-methylcatechol



(25-7)

whose ring is, however, opened by an extradiol 2,3-dioxygenase, a so called *meta*-cleavage.[163,164] The product, 2-hydroxy-6-oxohepta-2,4-dienoate, is cleaved hydrolytically as indicated on the structure



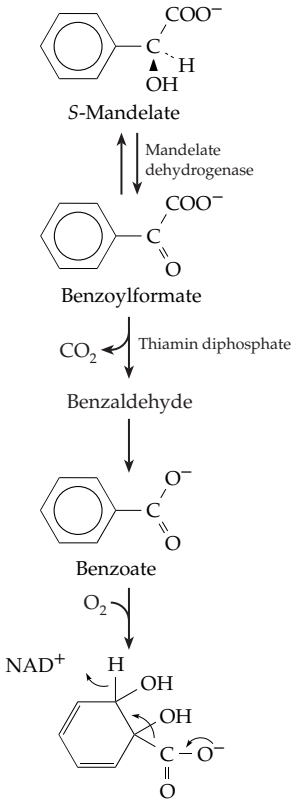
2-Hydroxy-6-oxohepta-2,4-dienoate

to give acetic acid and 2-hydroxypent-2,4-dienoate, which can be further metabolized. The hydrolytic cleavage is unusual[164] but is a β cleavage similar to the C–C bond cleavage by ribulose bisphosphate carboxylase (Eq. 13-48). Toluene can also be oxidized via benzoate through the β-oxoadipate pathway.

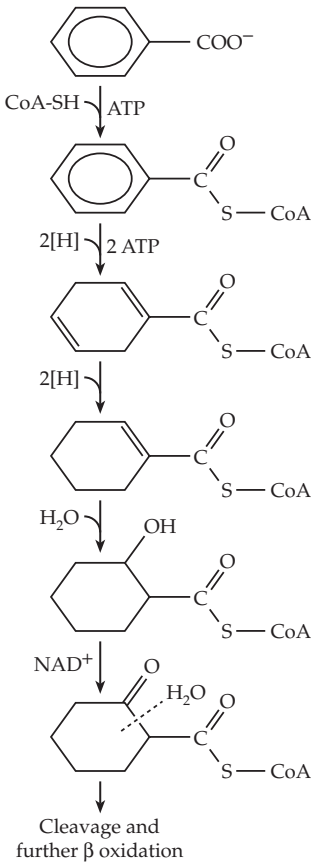
The plant acid *S*-mandelate must undergo conversion to *R*-mandelate by action of a racemase (Fig. 13-5) dehydrogenation, and side-chain cleavage as shown in Eq. 25-8 to form benzoate before it can be metabolized further.[165]

Although pseudomonads are well known for aerobic decomposition of aromatic compounds, some strains of *Pseudomonas*, as well as many other bacteria, are able to degrade aromatic compounds under completely anerobic conditions.[166,167] Benzoate can be converted to benzoyl-CoA and the ring can be partially reduced in two ATP- and NADH-dependent reactions (Eq. 25-9). The first of these reduction steps is unusual because ATP is apparently needed to drive the reaction.[166,166a,b] This is analogous to the need for ATP in nitrogen fixation (Eq. 24-6, step *b*).

Toluene, 3-chlorobenzoate, cinnamate, and 2-aminobenzoate can all be converted to benzoyl-CoA and be metabolized via the pathway of Eq. 25-9. Phenol, cresol, coumarate, protocatechuate, and vanillate can be converted to 4-hydroxybenzoyl-CoA and degraded in a similar fashion.[166] The breakdown of various forms of vitamin B_6 by bacteria is described in Section F (Eq. 25-24).



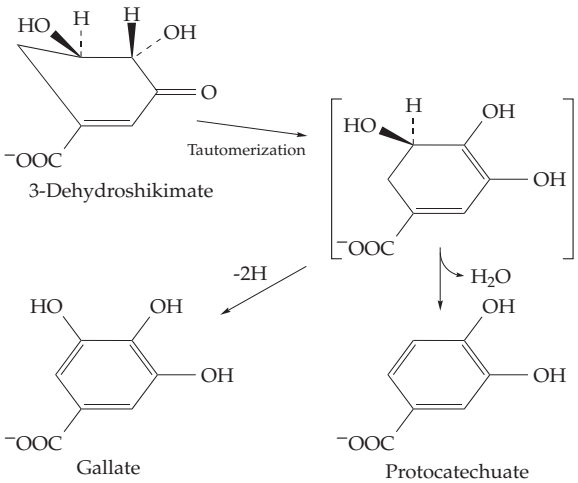
(25-8)



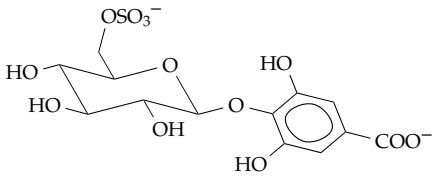
(25-9)

6. Quinic and Gallic Acids

Quinic acid, a compound accumulated by many green plants, can be formed by reduction of 3-dehydroquinate (Eq. 25-2) in both plants and bacteria. Quinic acid can be converted into useful industrial products such as benzoquinone and hydroquinone, and its production by bacteria provides a convenient route to these compounds.[168] In the main shikimate pathway 3-dehydroquinate is dehydrated to 3-dehydroshikimate (Eq. 25-3). The latter can be dehydrated further to **protocatechuate** (Eq. 25-10) either nonenzymatically[169] or by enzymatic action in bacteria.[170] Protocatechuate can be decarboxylated enzymatically to catechol, another compound of industrial value.[170] Nonenzymatic oxidation of 3-dehydroshikimate (Eq. 25-10) yields gallate.[169] Gallic acid derivatives are important plant constituents, but the biosynthetic origin has been obscure.[171] Gallate is probably formed from 3-dehydroshikimate as indicated in Eq. 25-10.[172] Esters and other derivatives of gallic acid constitute the **hydrolyzable tannins**. These materials accumulate in the vacuoles of the plants and are also deposited in the bark along with the **condensed tannins**, which are polymeric flavonoid compounds (Box 21-E).



(25-10)



Gallic acid 4-*O*-(β-D-glucopyranosyl-6′-sulfate), the periodic leaf movement factor from *Mimosa*

7. The Metabolism of Phenylalanine and Tyrosine in Plants

Some of the pathways of animal and bacterial metabolism of aromatic amino acids also are used in plants. However, quantitatively more important are the reactions of the **phenylpropanoid pathway**,[173–174a] which is initiated by **phenylalanine ammonia-lyase** (Eq. 14-45).[175] As is shown at the top of Fig. 25-8, the initial product from phenylalanine is ***trans*-cinnamate**. After hydroxylation to 4-hydroxycinnamate (*p*-coumarate) and conversion to a coenzyme A ester,[175a] the resulting *p*-coumaryl-CoA is converted into mono-, di-, and trihydroxy derivatives including anthocyanins (Box 21-E) and other flavonoid compounds.[176] The dihydroxy and trihydroxy methylated products are the starting materials for formation of lignins and for a large series of other plant products, many of which impart characteristic fragrances. Some of these are illustrated in Fig. 25-8.

Benzoic and salicylic acids. Two of the simplest plant acids arising from *trans*-cinnamate are **benzoic acid**, accumulated in plums and cranberries, and **salicylic acid**, present in all green plants and accumulated as methyl esters or glycosides in some plants, e.g., those of the willow family. Salicylic acid is made by hydroxylation of benzoic acid,[177] which can be formed from *trans*-cinnamate by β oxidation as depicted in Fig. 25-8, but it may also arise from isochorismate as shown in Fig. 25-2.[178] Salicylic acid plays a central role in resistance of plants to a variety of

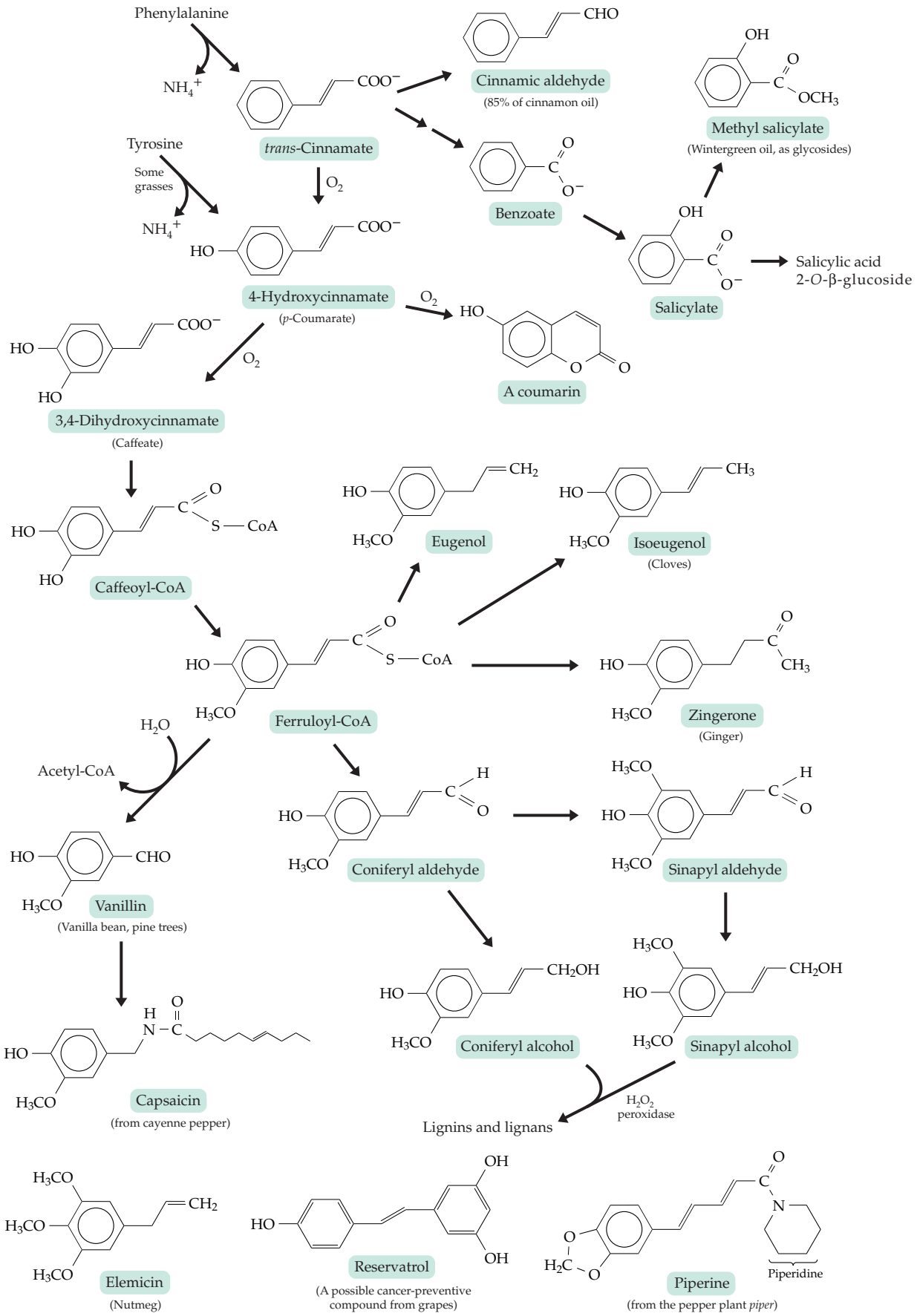
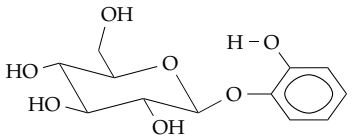


Figure 25-8 Formation of some plant metabolites from phenylalanine and tyrosine via the phenylpropanoid pathway.

diseases, a phenomenon called **aquired systemic resistance**.[179–181] A large number of defense-related genes are induced by salicylate leading to increased synthesis of phytoalexins, proteins, and lignins. The mechanism may be to inhibit catalases allowing the level of H_2O_2 to rise. Hydrogen peroxide not only is a precursor to potent antimicrobial compounds as HOCl (Eqs. 16-12, 16-13) and a participant in lignin synthesis, but it may directly activate transcription of disease resistance genes.[182] This mode of action appears to parallel an **acute-phase response** of the vertebrate immune system through which H_2O_2 activates the transcription factor NF-kB (Fig. 5-40; Chapter 28).[182] It is possible that salicylate also has an effect on transcription in the human body.[183]

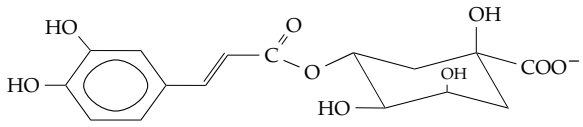
Salicylic acid derivatives, such as aspirin (acetylsalicylic acid), have long been known as pain relievers for the human body. The major effect is thought to be inhibition of a cyclooxygenase (Eq. 21-16). Willow bark has been known since the 18th century to contain a pain reliever, which was identified as salicylate esters and salicyl alcohol derivatives such as salicin.



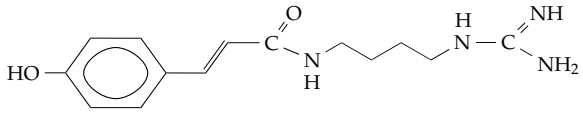
Salicin, a glucoside of salicyl alcohol

Vanilla and other plant products. One of the most widely used natural plant products is **vanilla extract**, which is obtained from cured, unripe fruit of the orchid *Vanilla planifolia*. The curing process releases **vanillin** and related compounds such as vanillic acid and 4-hydroxybenzaldehyde from glycosides. Because the flowers must be hand pollinated every day, natural vanilla extract is extremely expensive.[184,185] Most vanillin used in flavoring is obtained by hydrolysis of lignin, but production from glucose using bacterially produced enzyme reactors is possible.[185] Prince and Gunson, in an interesting article,[184] described the use of mass spectrometry and ^{13}C-enriched synthetic vanillin in the battles to distinguish natural vanilla extract from artificial mixtures of vanillin and other compounds and to camouflage the latter. Conversion of ferulic acid to vanillin in plants is apparently accomplished by β oxidation of acyl-CoA derivatives[186] as indicated in Fig. 25-8.

The ubiquitous plant compound **chlorogenic acid** (isolated from green coffee beans) is formed by transesterification with the glycoside cinnamoyl-glucose.[187] Coumaroyl-CoA is converted into monomeric and dimeric amides with **agmatine**, which provides barley plants with resistance to mildew.[188] Similar compounds with various polyamines and derived from *p*-coumaric, caffeic, ferulic, or sinapic acid appear to function in plant development. For example, caffeoylputrescine and caffeoyl-γ-aminobutyrate are accumulated specifically in sex organs of tobacco flowers.[189]

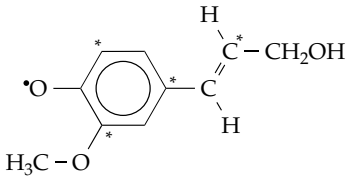


Chlorogenic acid
(3-*O*-caffeoylquinic acid)

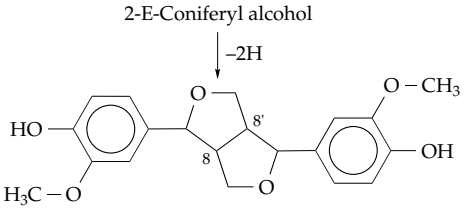


P-Coumaroylagmatine

Lignin, lignols, lignans, and phenolic coupling. Lignin is a complex material of relative molecular mass greater than 10,000. It is remarkably stable, being insoluble in hot 70% sulfuric acid. Lignin may be described as a "statistical polymer of oxyphenylpropane units." It arises from oxidative coupling of **coniferyl and sinapyl alcohols** (Figs. 25-8, 25-9) and related monomers known as **lignols**.[190–190c] The enzyme responsible for the polymerization may be a peroxidase, which catalyzes formation of lignin from the monomeric alcohols and H_2O_2. A radical generated by loss of an electron from a phenolate anion of coniferyl alcohol consists of a number of resonance forms in which the unpaired electron may be present not only on the oxygen but also at the positions marked by asterisks in the following structure:



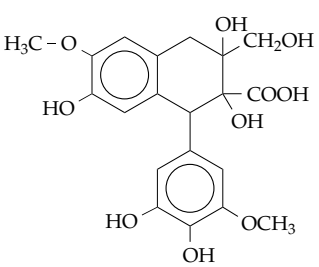
Coupling of such radicals yields a great variety of products. One type of dimerization gives the stable ether linked **pinoresorcinol** (Eq. 25-11). Through a complex sequence of reactions, it can be converted into other plant compounds including the phytoalexin **plicatic acid**, a major component of western



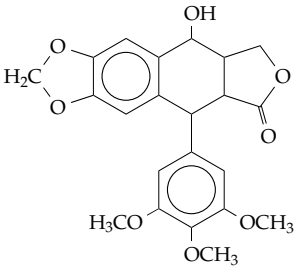
(25-11)

Figure 25-9 Proposed structure of beech lignin. There are 25 different C_9 units, of which several can, to some extent, be replaced by the three dimeric units in brackets. Redrawn from Nimz,[194] p. 317.

red cedar heartwood, and **podophyllotoxin**, found in the poisonous roots of the May apple (*Podophyllum peltatum*).[191] The little yellow "apple" is edible. Podophyllotoxin is used in cancer treatment. These dimeric 8,8'-carbon linked derivatives of lignols are called **lignans**[191a] while oligomers linked in other ways are **neolignans**.[192] The lignols are also incorporated covalently into **suberin**, a waxy layer of plant cell walls.[193]



Plicatic acid



Podophyllotoxin

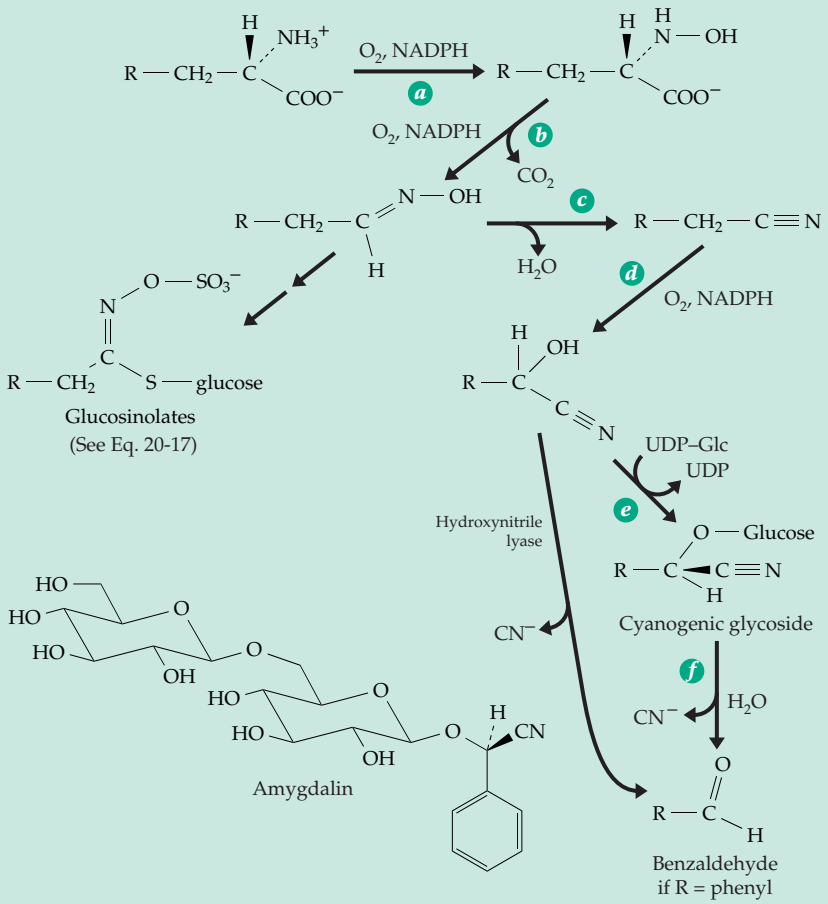
The lignols are synthesized within cells and are thought to move out into cell walls, possibly as phenolic glycosides.[190c,195] Cell wall peroxidases or laccases initiate polymerization.[173] The previously discussed lignans do not tend to polymerize, but other dimeric forms do. The dimers still contain hydroxyl groups capable of radical formation and addition to other units. At least ten types of intermonomer linkage other than that in the lignans are shown in Fig. 25-9. Lignin represents an enormous potentially valuable

BOX 25-B THE CYANOGENIC GLYCOSIDES

Cyanide-containing glycosides are synthesized by many higher plants including such crop plants as sorghum, cassava, and white clover.[a] The starting compounds are L-amino acids, most often phenylalanine, tyrosine, valine, or isoleucine.[b] The following sequence was proposed by Conn and others.[b–f] The conversion to an *N*-hydroxyamino acid in the first reaction (step *a*) is catalyzed by a cytochrome P450 hydroxylase system requiring NADPH and O_2. The same enzyme catalyzes a second hydroxylation that is followed by dehydration and decarboxylation (step *b*) to form an oxime.[f] The oxime is dehydrated to form a nitrile (step *c*) and a third hydroxylation (step *d*) produces an α-hydroxynitrile (cyanohydrin). Glycosylation by transfer from UDP-Glc or other sugar nucleotide (step *e*) forms the cyanogenic glycoside. If R = *p*-hydroxyphenyl in the foregoing equation and the sugar is glucose the product is **dhurrin**, present in **sorghum**. In **amygdalin**, present in bitter almonds and in pits of apricots, peaches, cherries, etc., two glucosyl units in β-1,6 linkage (gentiobiose) are attached to mandelonitrile.

Cyanogenic glycosides generate free cyanide by elimination when the glycosidic linkage is hydrolyzed.[a] This occurs with dhurrin at high pH and with others at pH=1, 70–100°C. Elimination of cyanide from the hydroxynitriles is catalyzed enzymatically.[g,h] Another cytochrome P450 dependent process utilizes oxidation to an oxime, as in the foregoing scheme, but converts the oxime to a **glucosinolate** in a two-step process.[i]

At one time amygdalin, sold as **Laetrile**, was promoted as a treatment for cancer, presumably based on the hope that the cancer cells would be poisoned by the released cyanide.[j] The tubers and leaves of the cassava plant provide a major source of food in many tropical countries. However, unless the cyanogenic glycosides are removed by boiling the tubers and pulping the leaves cassava is very toxic.[e,k]



a Vennesland, B., Castric, P. A., Conn, E. E., Solomonson, L. P., Volini, M., and Westley, J. (1982) *Fed. Proc.* **41**, 2639–

b Conn, E. E. (1979) *Nateuwissenschaften* **66**, 28–34

c Moller, B. L., and Conn, E. E. (1980) *J. Biol. Chem.* **255**, 3049–3056

d Moller, B. L., and Conn, E. E. (1979) *J. Biol. Chem.* **254**, 8575–8583

e Andersen, M. D., Busk, P. K., Svendsen, I., and Moller, B. L. (2000) *J. Biol. Chem.* **275**, 1966–1975

f Sibbesen, O., Koch, B., Halkier, B. A., and Moller, B. L. (1995) *J. Biol. Chem.* **270**, 3506–3511

g Lauble, H., Miehlich, B., Förster, S., Wajant, H., and Effenberger, F. (2001) *Protein Sci.* **10**, 1015–1022

h Lauble, H., Miehlich, B., Förster, S., Wajant, H., and Effenberger, F. (2002) *Biochemistry* **41**, 12043 – 12050

i Du, L., Lykkesfeld, J., Olsen, C. E., and Halkier, B. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12505–12509

j Newmark, J., Brady, R. O., Grimley, P. M., Gal, A. E., Waller, S. G., and Thistlethwaite, J. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6513–6516

k Ononogbu, I. C. (1980) *Trends Biochem. Sci.* **5**, X

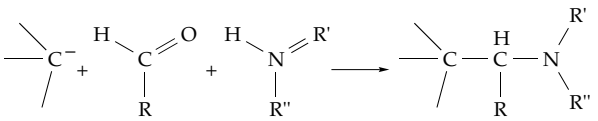
industrial source of aromatic raw materials, whose utilization has proved difficult. Oxidative degradation of lignin produces **humic acid**, an important organic constituent of soils.[196]

Oxidative coupling of radicals derived from phenols has a much wider role in nature than in lignin formation. Many alkaloids and other plant and fungal metabolites are synthesized using this reaction.[197] Tyrosine radicals are thought to be involved in formation of thyroxine (Eq. 25-6), melanin (Fig. 25-6), crosslinkages of ferulic acid with polysaccharides of plant cell walls,[197a] dityrosine, and other protein crosslinkages (Section B,4). Autocatalytic reaction of tyrosine-derived aminoquinone radicals in the Cu^{2+}-containing active site of amine oxidases apparently generates the mature prosthetic group topaquinone (Eq. 15-53).[198–198b] Oxidative coupling of tryptophanyl or cysteinyl side chains generates tryptophan tryptophylquinone (p. 817) or **cysteine tryptophylquinone** (CTA).[198b,c] As discussed on pp. 885–886 the tyrosine-cysteine thioether-bridged prosthetic groups of galactose oxidase (Fig. 16-29) and several other enzymes are also self-processing.[198d,e]

8. Alkaloids

More than 12,000 miscellaneous nitrogen-containing compounds, known as alkaloids, are produced by plants.[174a,199] Alkaloids are often thought of simply as end products of nitrogen metabolism in plants. However, most plants do not make alkaloids, whereas certain families of plants make many. There are probably ecological reasons.[200] Alkaloids often have potent physiological effects on animals, and many have been used as medicines from ancient times. Some have been prized through centuries as hallucinogens and intoxicants.

There are several classes of alkaloids. Among these are purines such as xanthine and caffeine, terpenes (Chapter 22), polyketides (Chapter 21), and alkaloids derived from amino acids. The basic amino acids ornithine, arginine, histidine, and lysine as well as the aromatic amino acids, anthranilate, and nicotinate are some of the starting materials.[199,201] Robinson[202,203] in 1917 recognized that many alkaloids are derived directly from aromatic amino acids. He proposed that alkaloids arise from **Mannich reactions** (Eq. 25-12) in which an amine and an aldehyde (probably through a Schiff base) react with a nucleophilic carbon such as that of an enolate anion. Many of the



(25-12)

amines are formed by decarboxylation of amino acids, and the aldehydes may arise by oxidative decarboxylation (transamination and decarboxylation) of amino acids. Thus, amino acids can provide both of the major reactants for alkaloid synthesis. Furthermore, nucleophilic centers in the aromatic rings, e.g., in positions para to hydroxyl substituents, are frequent participants in the proposed Mannich condensations. While Robinson's ideas on alkaloid biosynthesis were initially speculative, they have been confirmed by isotopic labeling experiments and more recently by isolation of the enzymes involved. Nevertheless, many questions remain. The postulated aldehydes are not proved intermediates. The condensations with 2-oxo acids may occur prior to decarboxylation.

An example is shown in Fig. 25-10. Dopa is decarboxylated to dopamine and is oxidized to 3,4-dihydroxybenzaldehyde. A Mannich reaction (via the Schiff base as shown) leads to ring closure. Oxidation of the ring produces an **isoquinoline** ring, a structural characteristic of a large group of alkaloids. Methylation produces **papaverine**, found in the opium poppy. A related alkaloid **morphine** (Fig. 25-10), at first glance, appears dissimilar. However, the biosynthetic route is similar. The initial Schiff base is formed from tyramine. Closure of the third ring together with hydroxylation and methylation yield ***R*-reticuline**, a precursor to many alkaloids. Its two rings are then oxidatively coupled through a C–C bond and an ether linkage.[204,204a] *S*-Reticuline is the precursor to another large family of alkaloids.[205] The six-membered ring of another alkaloid, **colchicine** (Box 7-D), originates from phenylalanine, while the seven-membered tropolone ring is formed from tyrosine by ring expansion.

C. Metabolism of Tryptophan and Histidine

The biosynthesis of tryptophan is outlined in Fig. 25-2. This amino acid not only assumes great importance in the structure and functioning of proteins but is converted into hormones, both in animals and plants, and into alkaloids in some plants. Some of the pathways are indicated in Figs. 25-11 and 25-12.

1. The Catabolism of Tryptophan

The primary catabolic pathway for tryptophan in animal cells is initiated (step *a*, Fig. 25-11) by **tryptophan 2,3-dioxygenase** (tryptophan pyrrolase; Eq. 18-38).[206] The enzyme is induced both by glucocorticoids and by tryptophan.[207] The related **indolamine 2,3-dioxygenase** catalyzes the same reaction of L-tryptophan but also acts on D-tryptophan and other substrates. It has different tissue distribution and reg- ulatory properties[208] and may play a role in

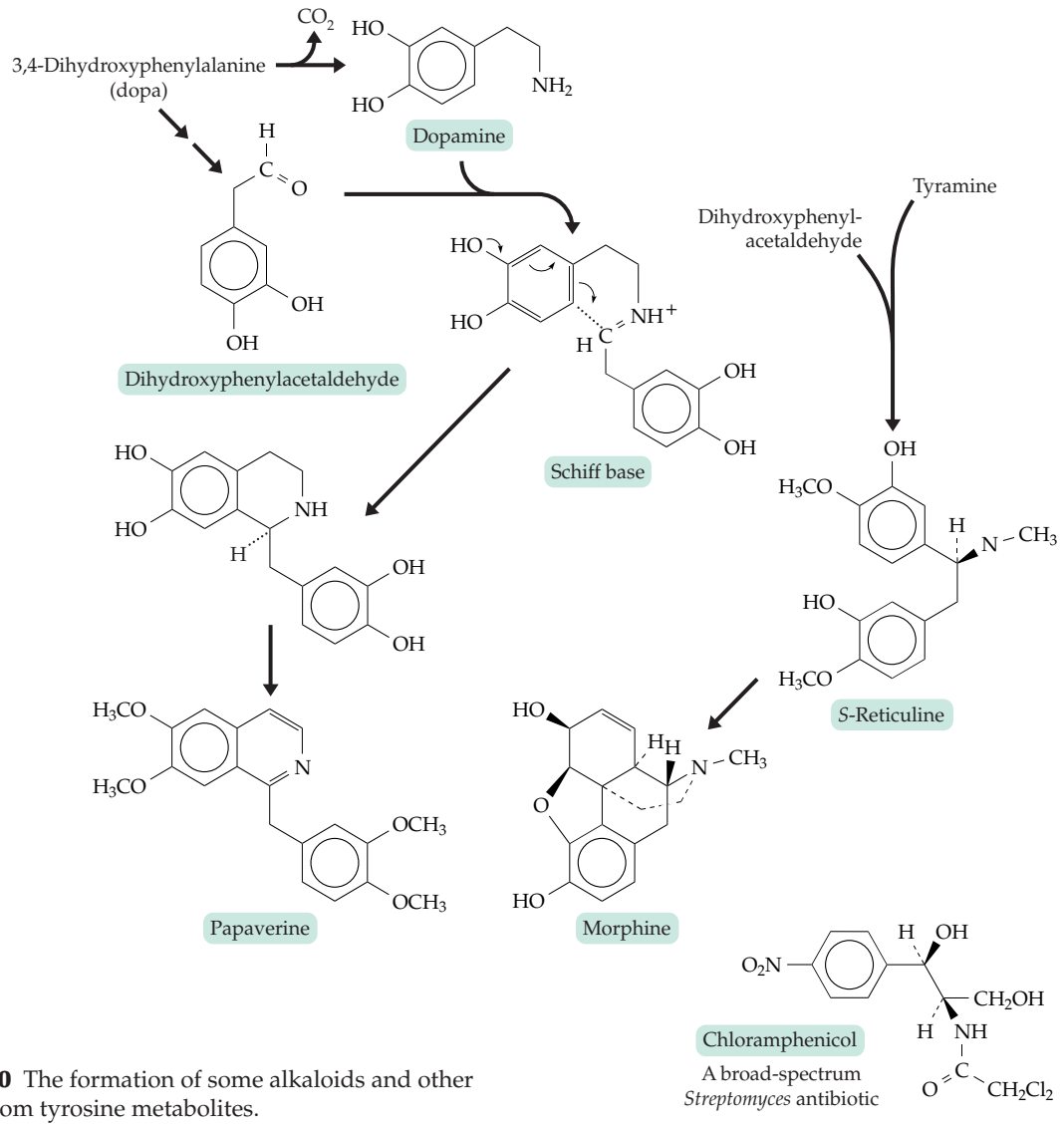


Figure 25-10 The formation of some alkaloids and other substances from tyrosine metabolites.

inflammatory responses.[206,209] An alternative pathway of tryptophan breakdown takes place in intestinal bacteria, which utilize tryptophan indolelyase (tryptophanase) to eliminate indole (step *b*, Fig. 25-11).[209a] The indole is hydroxylated to **indoxyl**, some of which is absorbed into the bloodstream and excreted in the urine as indoxyl sulfate.

Returning to the major tryptophan catabolic pathway, marked by green arrows in Fig. 25-11, formate is removed hydrolytically (step *c*) from the product of tryptophan dioxygenase action to form **kynurenine**, a compound that is acted upon by a number of enzymes. Kynureninase (Eq. 14-35) cleaves the compound to anthranilate and alanine (step *d*), while transamination leads to the cyclic **kynurenic acid** (step *e*). The latter is dehydroxylated in an unusual reaction to **quinaldic acid**, a prominent urinary excretion product.

Another major pathway of kynurenine metabolism (step *f*, Fig. 25-11) is hydroxylation to **3-hydroxykynurenine**, which in turn can undergo transamination to the cyclic **xanthurenic acid**. Xanthurenic acid is excreted from the human body, but in the malaria mosquito *Anopheles gambia* it acts as a mating factor for the malaria parasite *Plasmodium*.[210,211] In many insects, including *Anopheles*, 3-hydroxykynurenine is a precursor of insect eye pigments or "**omnochromes**."[210,212–213a] 3-Hydroxykynurenine also has neurotoxic properties.[213a] (See p. 1796.)

Cleavage of 3-hydroxykynurenine by kynureninase (step *g*, Fig. 25-11) forms 3-hydroxyanthranilate, which is opened under the action of another dioxygenase (step *h*) with eventual degradation to acetyl-CoA, as indicated. In insects the reactive 3-hydroxyanthranilate is utilized in "tanning" reactions, e.g., coupling to tyrosine residues to toughen insect cuticles and walls of cocoons.[214]

Tryptophan is hydroxylated to 5-hydroxytryptophan[213b] which is decarboxylated to **serotonin** (5-hydroxytryptamine), an important neurotransmitter

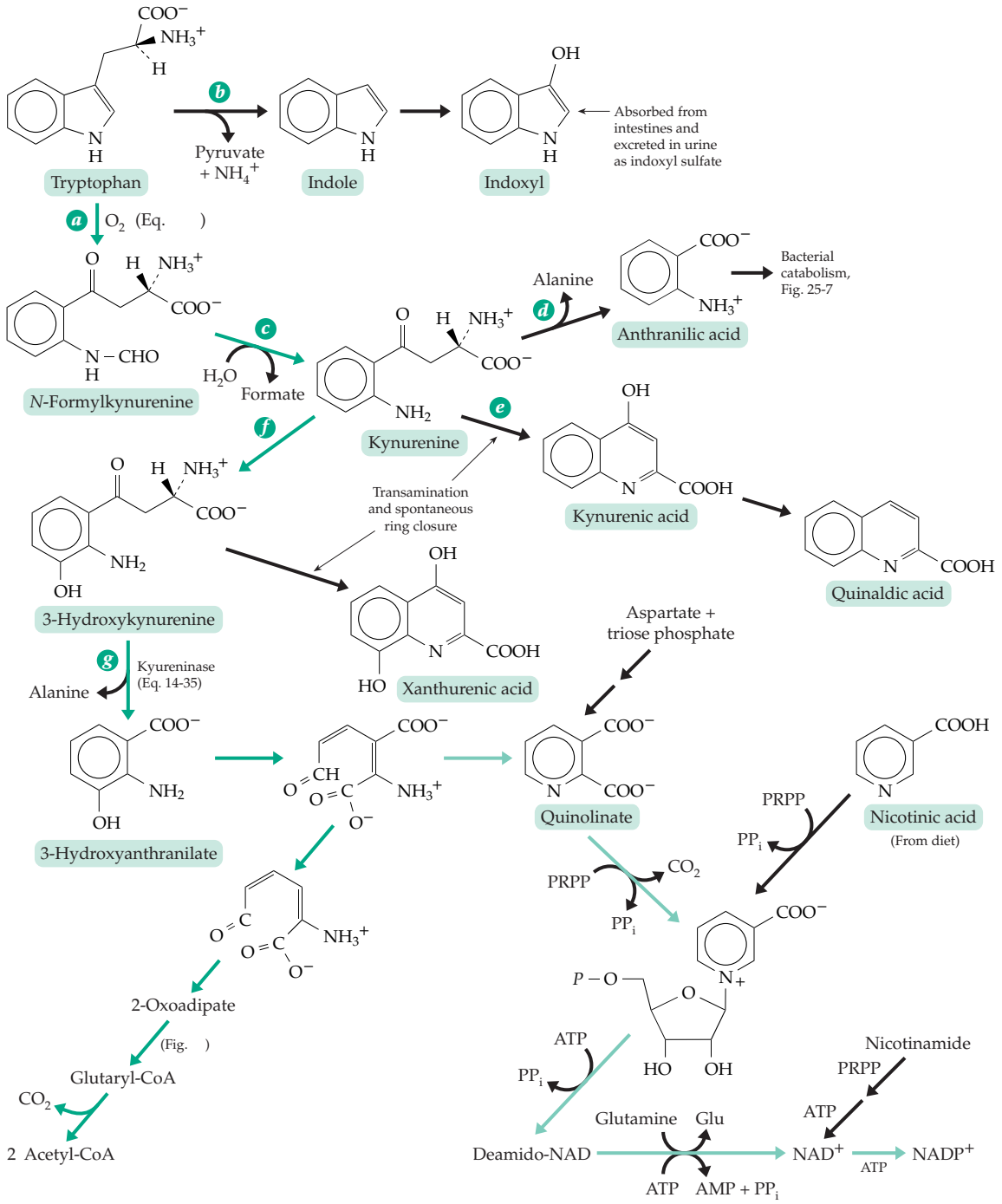
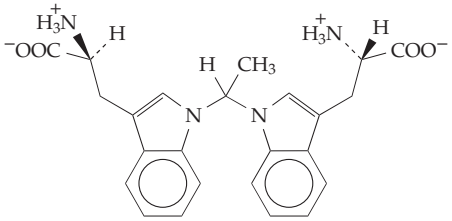


Figure 25-11 Some catabolic reactions of tryptophan and synthetic reactions leading to NAD and NADP.

substance[215,215a] and a regulatory component of plants and animals alike.[216] In the pineal gland serotonin is methylated and acetylated to **melatonin**, the pineal hormone[217–221] (Fig. 25-12).

The following dangerous tryptophan derivative was evidently formed in a fermentation used to produce tryptophan sold as a food supplement in 1990. More than 1,500 persons became ill and 27 died, perhaps as a direct result of toxicity of this compound.[222,223]



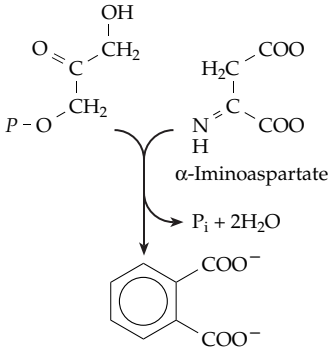
1,1'-Ethylidenebis[tryptophan]

Formation of NAD⁺ and NADP⁺. An alternative pathway, marked by shaded green arrows in Fig. 25-11, allows animals to form the nicotinamide ring of NAD^+ and $NADP^+$ from tryptophan.[224] The aldehyde produced by the ring opening reaction of step *h* can reclose (step *i*) to a pyridine ring in the form of **quinolinic acid**.[225] The latter, in a reaction that is also accompanied by decarboxylation, is coupled with a phosphoribosyl group of PRPP to form **nicotinate mononucleotide**.[225a] Adenylylation produces deamido NAD, which is converted to **NAD** by a glutamine- and ATP-dependent amination of the carboxyl group.[226]

As indicated in Fig. 25-11, free nicotinic acid can also be used to form NAD. Not surprisingly, nicotinic acid, an essential vitamin, is about 60 times more efficient than tryptophan as a source of NAD. Nevertheless, a high-tryptophan diet partially overcomes a deficiency in dietary intake of nicotinic acid. The effectiveness of a diet containing only maize as a source of protein in inducing the deficiency disease pellagra (Box 15-A) is in part a result of the low tryptophan content of maize protein. Nicotinic acid is rapidly converted in the liver to an amide with glycine, **nicotinuric acid**. Nicotinurate can be oxidatively cleaved by peptidylglycine monoxygenase to nicotinamide[227] in another alternative synthetic route to NAD.

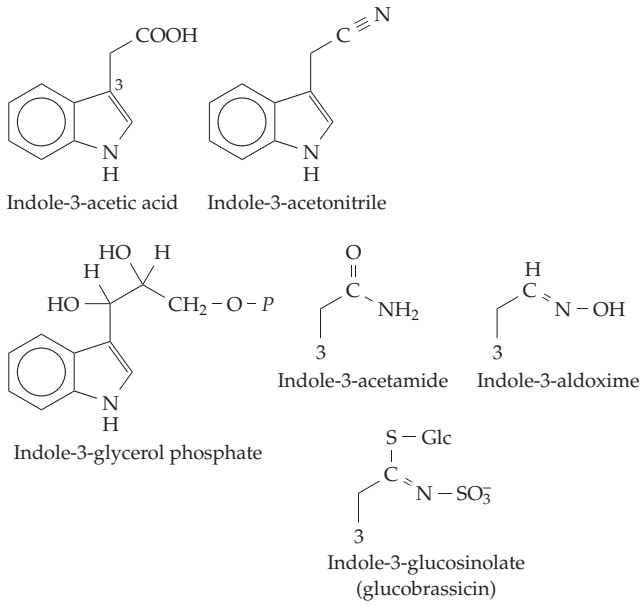
An alternative pathway for synthesis of quinolinate from aspartate and a triose phosphate exists in bacteria and in plants and provides the major route of nicotinic acid synthesis in nature. In *E. coli* the reaction is catalyzed by two enzymes, one an FAD-containing L-aspartate oxidase which oxidizes aspartate to α-iminoaspartate.[228] The latter condenses with dihydroxyacetone-*P* to form quinolinate (Eq. 25-13).[229] There are at least two other pathways for synthesis of quinolinic acid as well as five or more salvage pathways for resynthesis of degraded pyridine nucleotide coenzymes.[224,230,231]

Although quinolinic acid provides an important source of nicotinamide coenzymes, in excess it is a neurotoxic **excitotoxin** (Chapter 30) that has been associated with epilepsy and with inflammatory neuropathalogical conditions resulting from encephalitis.[213,232,233]



Quinolinate (25-13)

Auxin. The important plant hormone **indole-3-acetic acid** (IAA; often called by the more general name **auxin**) is partially derived by oxidative decarboxylation (Fig. 25-12, top) catalyzed by tryptophan-2-monooxygenase,[234–234b] a flavoprotein similar to lysine monooxygenase (Eq. 18-41). The reduction product indole-3-ethanol also occurs in plants and is metabolically active.[235] However, most IAA in plants is not formed from tryptophan but from some precursor,[236,236a] perhaps indole-3-glycerol phosphate, which immediately precedes tryptophan in its biosynthesis (Fig. 25-2). Routes of synthesis from indole-3-acetonitrile, indoleacetaldoxime,[236b] indole-3-glucosinolate (glucobrassicin), indole-3-pyruvate, and tryptamine have also been reported.[237–238a] Nitrilases are found in



plants such as *Arabidopsis thaliana*[238] and also in symbiotic bacteria such as *Rhizobium*. Together with amidases they convert the acetonitrile, acetaldoximes, or acetamide derivatives to IAA,[237] which is transported throughout the plant.[238b]

About 95% of the IAA within plants is stored as conjugated forms that include amides with various amine acids and peptides[239] and glycosyl derivatives.[240] The gall-forming *Pseudomonas savastonoi* forms both IAA and conjugates such as N^{ε}-(indole-3-acetyl)-L-lysine, which aid these bacteria in colonizing olive and oleander plants.[241]

As a hormone IAA has a broad range of effects on plants, altering tissue differentiation, root growth, cell elongation, and cell division.[241a] The fastest observed response is an effect on cell elongation, which can be observed within 15–20 minutes.[242] In *A. thaliana* IAA

causes very rapid transcription of at least five genes, one of which encodes 1-aminocyclopropane-carboxylase (ACC) synthase (Eq. 14-27).[243]

Alkaloids from tryptophan. The alkaloid **harmine**, which is found in several families of plants, can be formed from tryptophan and acetaldehyde (or pyruvate) in the same manner as is indicated for the formation of **papaverine** in Fig. 25-10. Some other characteristic plant metabolites such as **psilocybine**, an hallucinogenic material from the mushroom *Psilocybe aztecorum*, are formed directly from serotonin (Fig. 25-12). For many years **gramine** from barley was regarded as a curiosity because only one carbon atom separates the nitrogen atom from the indole ring. It is now believed that tryptophan is cleaved in a PLP-dependent reaction analogous to that of serine trans-hydroxymethylase (Eq. 14-30; Fig. 25-12). Other alkaloids arise in a more conventional fashion. Condensation of an isopentenyl group on the indole ring of tryptamine (Fig. 25-12) initiates the formation of **lysergic acid** and other ergot alkaloids.[244] The indole ring

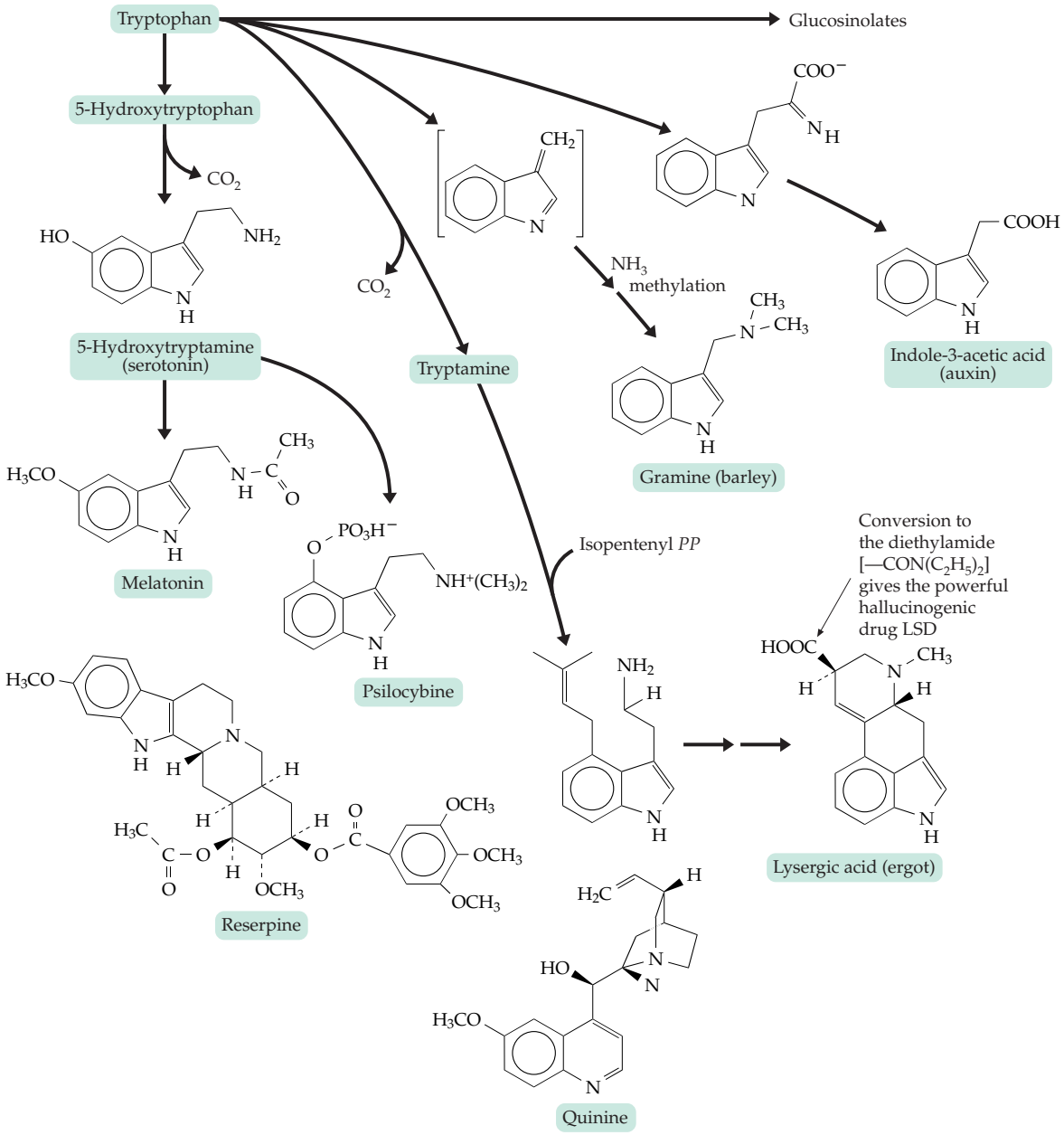
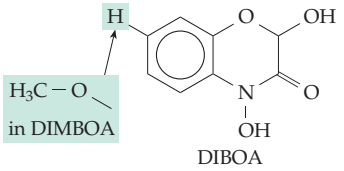


Figure 25-12 Structures and some biosynthetic pathways for some hormones, indole alkaloids, and other metabolites of tryptophan.

of tryptophan is clearly visible in the structure of **reserpine** (Fig. 25-12). This compound from *Rauwolfia* is of medical interest because of its effect in lowering blood pressure and in depleting nervous tissues of serotonin, dopamine, and noradrenaline. Reserpine also contains a benzene ring, which is derived from tryptophan by a ring expansion. The periwinkle alkaloids, including the antitumor drug **vincristine** (see Box 7-D), are formed by condensation of tryptamine with the complex glycosidic aldehyde secologamin. Additional reactions form vincristine and more than 100 other indole alkaloids.[245,245a]

Another group of plant metabolites derived from tryptophan are cyclic hydroxamic acids whose names



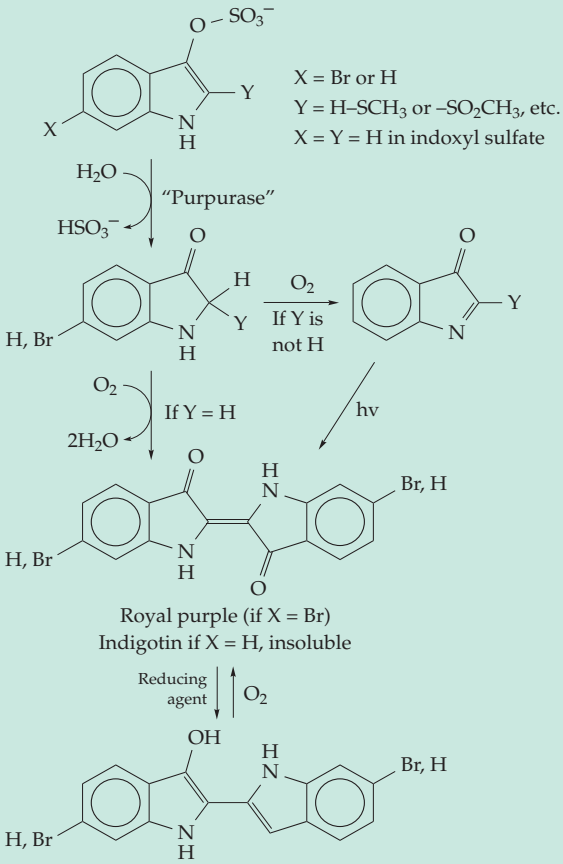
BOX 25-C ROYAL PURPLE AND BLUE DENIM

Ancient people, including the Phoenicians, Chinese, and Peruvians, discovered a dyeing process that utilized secretions of certain marine molluscs, animals that were also eaten as food.[a] In processes that were perhaps closely guarded secrets, the molluscan secretions were heated for days in vats with water, salt, and additional additives including human urine, honey, etc. When the mixture was right, wool was dipped and allowed to air dry in sunlight to give the famous royal purple colors. In this ancient process as many as 10,000 molluscs were used to produce one gram of the dye.[a] In other parts of the world blue dyes were generated by fermentation of plants of the genus *Indigofera* and also of the European woad plant.[b,c]

By 1909 the chemical nature of the royal purple pigment dibromoindogotin (see scheme) had been established and by 1897 synthetic indigo production had already begun. Ancient indigo dyeing utilized the precursors, sulfate esters of **indoxyl** or of substituted indoxyl, metabolites of tryptophan (see Fig. 25-11). The sulfate esters were hydrolyzed by the sulfatase "purpurase" to give the tautomer of indoxyl that is shown in the accompanying scheme. Atmospheric oxygen converts these compounds to the corresponding oxidized dyes, the **indigotins**. However, they are very insoluble and unsuitable for dyeing. In modern indigo dyeing to form such fabrics as blue denim, the indigotin is reduced with sodium dithionite, about 2 kg of the latter being used to reduce 1 kg of the dye to the reduced leuco form.[c] Either wool or cotton can by dyed with this reduced form, air oxidation returning the dye to the blue oxidized form.

Ancient dyers also had to maintain the dye in the reduced form. In fermentations of the woad plant a species of thermophilic *Clostridium* apparently supplied the reducing agent. Padden *et al.* suggested that such bacterial reduction might be used today to avoid pollution by the by-products of dithionite reduction.[c] Use of engineered bacteria to form high yields of indole and indoxyl as a source of indigoid precursors has also been suggested.[b]

One more complexity needs to be considered. Some of the precursors are adducts ($Y = -SCH_3$, $-SO_2CH_3$ in the structures) and cannot be oxidized directly to the indigotins. Use of sunlight in a photochemical process was required in these cases.[a]



[a] McGovern, P. E., and Michel, R. H. (1990) *Acc. Chem. Res.* **23**, 152

[b] Ensley, B. D., Ratzkin, B. J., Osslund, T. D., Simon, M. J., Wackett, L. P., and Gibson, D. T. (1983) *Science* **222**, 167–169

[c] Padden, A. N., Dillon, V. M., John, P., Edmonds, J., Collins, M. D., and Alvarez, N. (1998) *Nature (London)* **396**, 225

are often abbreviated to DIBOA and DIMBOA. They are part of the defense system of grasses against insects and fungi. DIBOA is the main hydroxamic acid in rye while DIMBOA predominates in wheat and maize. The compounds arise from indole generated from indole-3-glycerol-*P* followed by action of four cytochrome P450 enzymes.[246]

2. The Metabolism of Histidine

The biosynthesis of histidine, which might be regarded as the "super catalyst" of enzyme active centers, begins with a remarkable reaction of ATP, the "super coenzyme" of cells. The reaction is a displacement by N-1 of the adenine ring on C-1 of PRPP (step *a*, Fig. 25-13). The resulting product undergoes a ring opening reaction, step *b*, followed by an Amadori rearrangement (step *c*). The rearrangement product is cleaved via reaction with ammonia released from glutamine with formation of **5-aminoimidazole-4-carboxamide**, whose ribotide is an established intermediate in the synthesis of ATP and other purines. Here it is recycled via ATP (Fig. 25-13). The other product of the cleavage contains the five carbons of the original ribosyl group of PRPP, together with one nitrogen and one carbon split out from the ATP molecule and the nitrogen donated by glutamine.[246a] Ring closure (step *e*) forms the imidazole group, which is attached to a glycerol phosphate molecule. The glycerol-*P* end of the molecule undergoes dehydration[247] and ketonization of the resulting enol to a product, which can be transaminated[247a,b] and dephosphorylated to histidinol. Dehydrogenation of this alcohol forms histidine.[248–249a]

Regulation of histidine synthesis. In all, ten different genes code for the enzymes of histidine biosynthesis in *Salmonella typhimurium*. They are clustered as the **histidine operon**, a consecutive series of genes which are transcribed into messenger RNA as a unit.[250,251] The gene symbols *HisA*, *HisB*, etc., are indicated in Fig. 25-13, and their positions on the *E. coli* gene map are indicated in Fig. 26-4. The gene *HisB* codes for a complex protein with two different enzymatic activities as shown in Fig. 25-13.

The presence of an excess of histidine in a bacterial cell brings about repression of synthesis of all of the

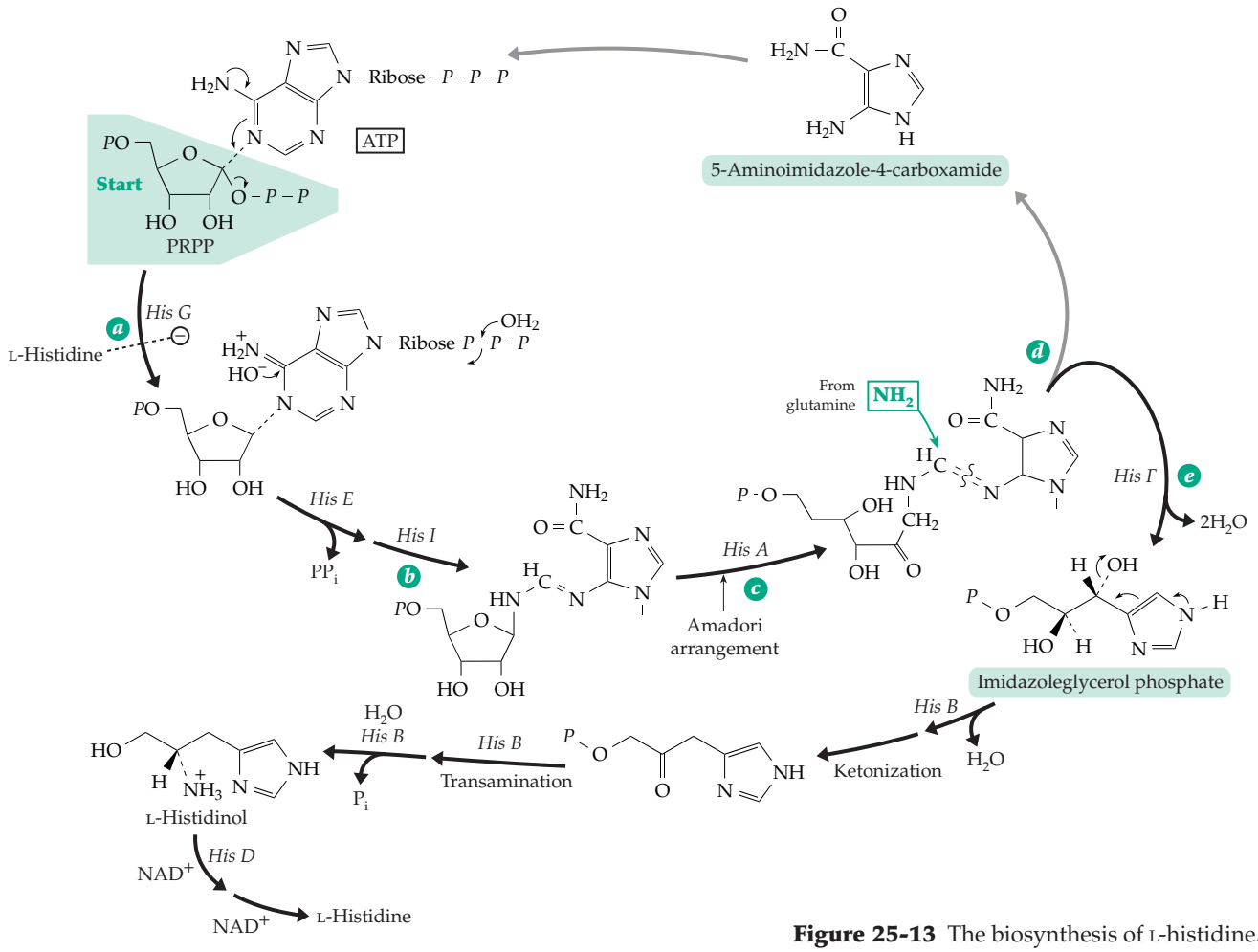
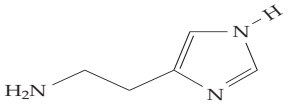


Figure 25-13 The biosynthesis of L-histidine.

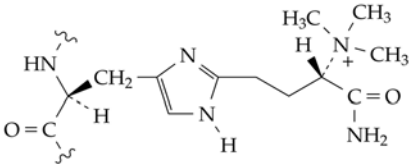
enzymes encoded in the histidine operon. Details of the functioning of this and other operons are considered in Chapter 28. Histidine is also an allosteric inhibitor for the first enzyme of the biosynthetic sequence, i.e., step *a* of Fig. 25-13. Thus, instantaneous inhibition of the biosynthesis occurs if an excess of histidine accumulates. Similar patterns of both repression and feedback inhibition exist for many of the pathways of amino acid biosynthesis (Chapter 28).

Catabolism of histidine. The first steps of the major degradative pathway for histidine metabolism have already been discussed. Elimination of ammonia, followed by hydration and ring cleavage to **formiminoglutamate**, involves unusual reactions (Eq. 25-14)[252] which have been discussed earlier. Transfer of the formimino group to tetrahydrofolic acid and its further metabolism have also been considered (Chapter 15).

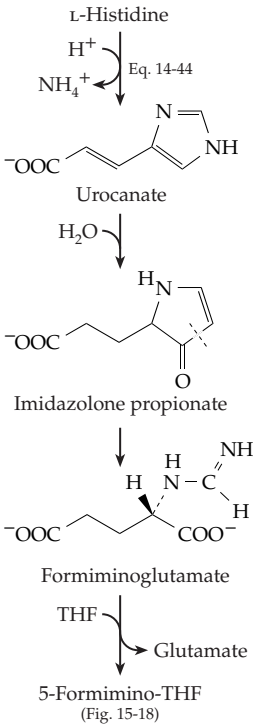


Histamine

Other products from histidine include the hormonal substance **histamine** formed by decarboxylation, the oxidation product, imidazole acetic acid, and N^{δ}- and N^{ε}-methylhistidines. Histamine plays a role in release of gastric secretions and allergic responses (Chapter 5). Drugs (antihistamines) that inhibit its release are in widespread use. The unusual amino acid **diphthamide** has an unknown function in protein synthesis, occurring within the peptide chain of eukaryotic elongation factor 2 (Chapter 29).[253] Its biosynthesis from a specific histidine in EF-2 of all eukaryotes and archaebacteria requires four molecules of *S*-adenosylmethionine. The first transfers the four-carbon backbone of AdoMet to $C^{\varepsilon 1}$ of the histidyl group, a nucleophilic displacement resembling that of AdoMet-dependent C-methylation (Eq. 12-4). This is followed by transfer of three methyl groups, each from AdoMet, and finally an ATP-dependent amidation of the carboxyl group.[253] Diphthamide is the target for attack by diphtheria toxin (Box 29-A).



Dipthamide residue within a peptide chain.



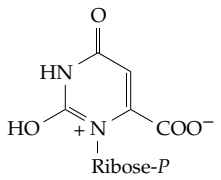
(25-14)

D. Biosynthesis and Catabolism of Pyrimidines

L-Aspartate contributes four of the six ring atoms of pyrimidines including the nitrogen. The α-carboxylate group is eventually lost as CO_2, the decarboxylation helping to drive the synthetic sequence. Six enzymatic steps are required to form the product uridine 5'-phosphate (UMP) as shown in Fig. 25-14, steps *a*–*f*. UMP is then converted on to the cytidine, uridine, and thymidine nucleotides as shown. This pathway of pyrimidine synthesis has been conserved throughout evolution and is used by all but a few specialized organisms.[254–258] The first step is synthesis of carbamoyl phosphate by the glutamine-dependent carbamoyl phosphate synthetase, an allosteric enzyme discussed in Chapter 24 (Eq. 24-22).[259–260a] The next step is transfer of the carbamoyl group to aspartate (Fig. 25-14, step *b*). The product is able to cyclize immediately (step *c*) by elimination of water to form **dihydroorotate**.[260a] The highly controlled aspartate carbamoyltransferase has been discussed in Chapters 7 and 11. Although this is a monofunctional enzyme in bacteria, it is fused with two other proteins in mammalian cells. The resulting multifunctional enzyme (with 240-kDa subunits) catalyzes three consecutive steps: the synthesis of carbamoyl phosphate, the carbamoyltransferase reaction, and the Zn^{2+}-dependent ring-closing reaction that forms dihydroorotate.[261,262] This cyclic product is oxidized to **orotate** (Fig. 25-14, step *d*) by a flavoprotein oxidase, which in mammals utilizes ubiquinones as electron acceptors.[263–265a] A displacement reaction with PRPP (Fig. 25-14, step *e*; see also Fig. 25-13, step *a*) converts it into **orotidine 5'-phosphate**[266,267] with release of PP_i.

1. Synthesis of Pyrimidine Nucleotides

Orotidine 5'-phosphate undergoes an unusual decarboxylation (Fig. 25-14, step *f*), which apparently is not assisted by any coenzyme or metal ion but is enhanced over the spontaneous decarboxylation rate 10^{17}-fold. No covalent bond formation with the enzyme has been detected.[268] It has been suggested that the enzyme stabilizes a dipolar ionic tautomer of the substrate. Decarboxylation to form an intermediate ylid would be assisted by the adjacent positive charge.[269,270] Alternatively, a concerted mechanism may be assisted by a nearby lysine side chain.[270a–d] Hereditary absence of this decarboxylase is one cause of orotic aciduria. Treatment with uridine is of some value.[271]



Orotidine 5'-phosphate
(dipolar ionic tautomer)

We see that **uridine 5'-phosphate** (UMP) is formed from aspartate in a relatively direct and simple way. Phosphorylation with ATP in two steps produces UDP and UTP. The **cytosine nucleotides** are formed from UTP, the initial step being amination to CTP (step *h*, Fig. 25-14). This reaction resembles in many respects the conversion of citrulline to arginine, which depends upon ATP and involves transfer of the nitrogen of aspartate (Eq. 24-23). However, in the formation of CTP glutamine serves as the nitrogen donor (NH_4^+ can substitute). Observation of positional

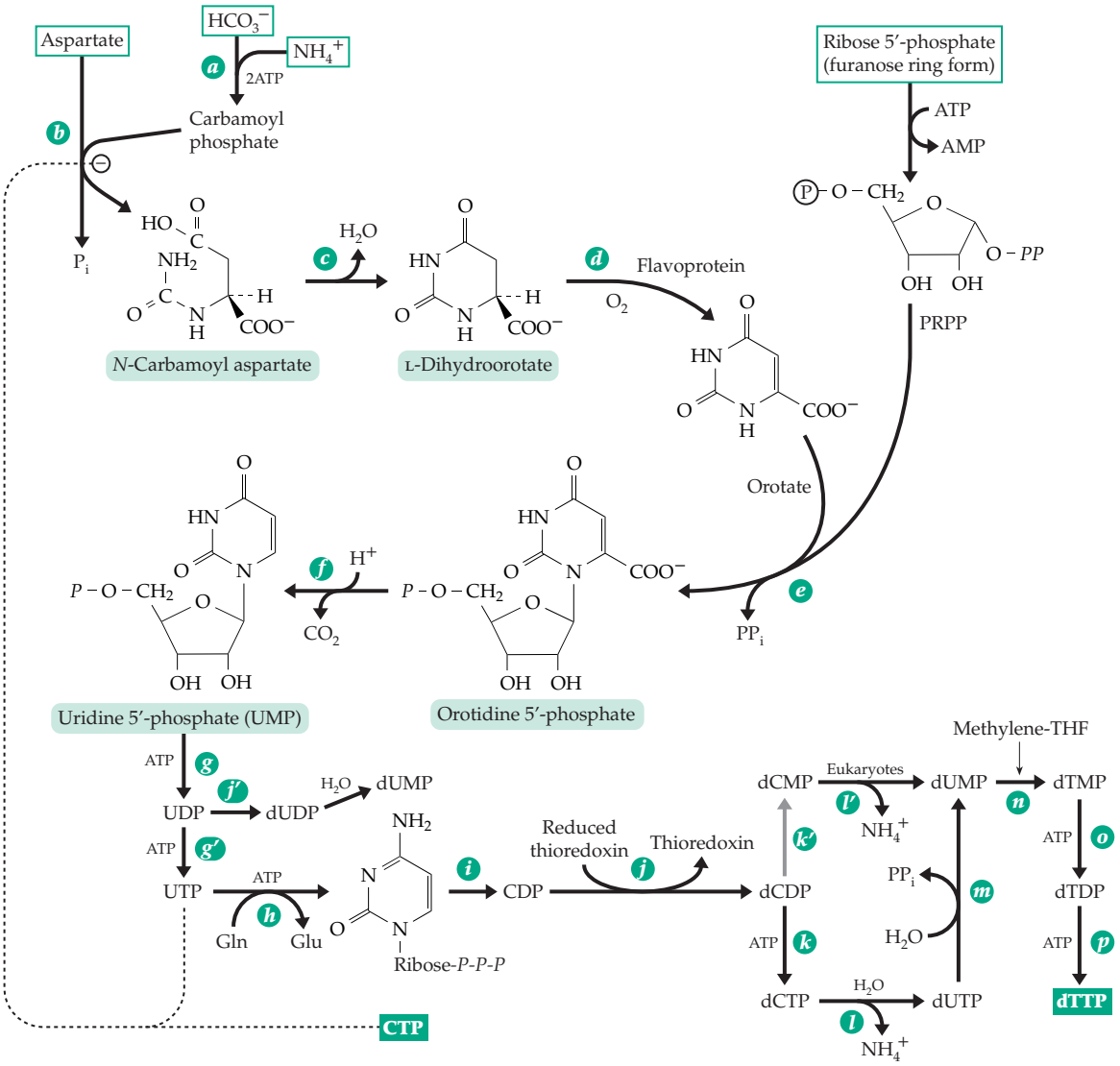
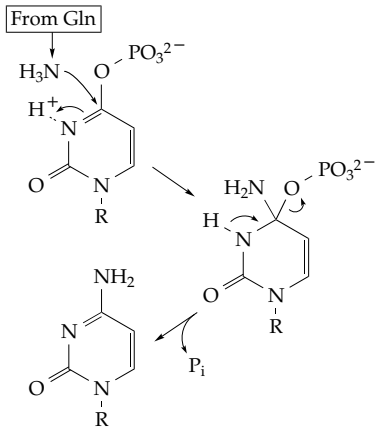


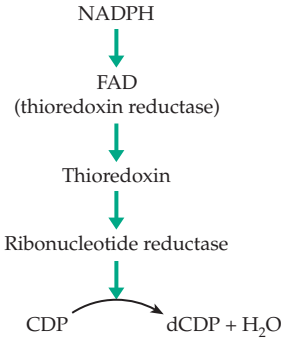
Figure 25-14 Assembly of the pyrimidine ring and biosynthesis of the pyrimidine nucleotide precursors of RNA and DNA.

isotope exchange with [γ-$^{18}O_4$]ATP suggested the occurrence of an enolic phosphate intermediate (Eq. 25-15)[272] as had also been suggested for adenylosuccinate synthase (Fig. 25-15, step *k*). CTP is incorporated directly into RNA and into such metabolic intermediates as CDP- choline, or it can be dephosphorylated to CDP (Fig. 25-14, step *i*). It is CDP that serves as the principal precursor for the deoxyribonucleotides **dCDP** and **thymidine diphosphate** (dTTP).



(25-15)

Deoxyribonucleotides. A chain involving NADPH, a flavoprotein, thioredoxin, and ribonucleotide reductase converts either the ribonucleoside diphosphates or triphosphates to the corresponding 2-deoxy forms (step *j*, Fig. 25-14) as indicated in Eq. 25-16.



(25-16)

Ribonucleotide reductases are discussed in Chapter 16. Some are iron-tyrosinate enzymes while others depend upon vitamin B_{12}, and reduction is at the nucleoside *triphosphate* level. Mammalian ribonucleotide reductase, which may be similar to that of *E. coli*, is regarded as an appropriate target for anticancer drugs. The enzyme is regulated by a complex set of feedback mechanisms, which apparently ensure that DNA precursors are synthesized only in amounts needed for DNA synthesis.[273] Because an excess of one deoxyribonucleotide can inhibit reduction of all ribonucleoside diphosphates, DNA synthesis can be inhibited by deoxyadenosine or by high levels of thymidine, despite the fact that both compounds are precursors of DNA.

Phosphorylation of dCDP to dCTP (step *k*, Fig. 25-14) completes the biosynthesis of the first of the pyrimidine precursors of DNA. The uridine nucleotides arise in two ways. Reduction of UDP yields dUDP (step *j′*, Fig. 25-14). However, the deoxycytidine nucleotides are more often hydrolytically deaminated (reactions *l* and *l′*).[274] Methylation of dUMP to form **thymidylate**, dTMP (step *n*, Fig. 25-14), is catalyzed by thymidylate synthase. The reaction involves transfer of a 1-carbon unit from methylene tetrahydrofolic acid with subsequent reduction using THF as the electron donor. A probable mechanism is shown in Fig. 15-21. See also Box 15-E. Some bacterial transfer RNAs contain 4-thiouridine (Fig. 5-33). The sulfur atom is introduced by a sulfurtransferase (the *ThiI* gene product in *E. coli*). The same protein is essential for thiamin biosynthesis (Fig. 25-21).[274a]

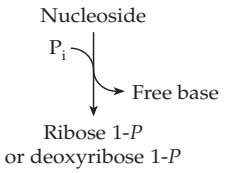
Formation of dUMP in eukaryotes may occur by hydrolytic removal of phosphate from dUDP or from the conversions dCDP → dCMP → dUMP (steps *k′* and *l′*, Fig. 25-14). A more roundabout pathway is employed by *E. coli*: dCDP → dCTP → dUTP → dUMP (steps *k*, *l*, and *m*, Fig. 25-14). One of the intermediates is dUTP. DNA polymerases are able to incorporate dUMP from this compound into polynucleotides to form uracil-containing DNA. The only reason that this does not happen extensively within cells is that dUTP is rapidly converted to dUMP by a pyrophosphatase (step *m*, Fig. 25-14). The uracil that is incorporated into DNA is later removed by a repair enzyme (Chapter 27). The presence of dUTP in DNA provides the basis for one of the most widely used methods of directed mutation of DNA (Chapter 26).

Bacteriophage-induced alterations in metabolism. Interesting alterations in nucleotide metabolism occur in cells of *E. coli* infected by T-even bacteriophage. Genes carried by the phage are transcribed, and the corresponding proteins are synthesized by the host cell.[273] A number of these viral gene products are enzymes affecting nucleotide metabolism. One enzyme catalyzes the hydrolytic conversion of dCTP to dCMP, and another promotes the synthesis of 4-hydroxymethyl-dCMP. Such virus-specified enzymes may be appropriate target sites for antiviral drugs.

2. Catabolism of Pyrimidine Nucleotides and Nucleosides

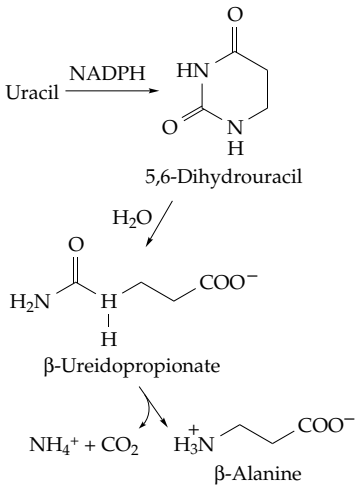
Nucleic acids within cells, as well as in the digestive tract, are continually under attack by many **nucleases**. Messenger RNA is degraded, often quite

rapidly, as an essential part of the control of protein synthesis. Although DNA is very stable, nucleases are called upon to cut out damaged segments of single strands as part of essential repair processes (Chapter 27). Thus, there is an active breakdown of polynucleotides to mononucleotides, which are hydrolyzed to nucleosides by phosphatases. Nucleosides are converted to free bases by the action of **nucleoside phosphorylases** (Eq. 25-17). The further degradation of



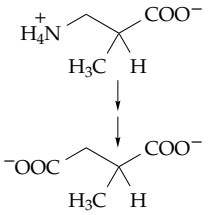
(25-17)

cytosine is initiated by deamination to uracil.[274b,c] Catabolism of uracil starts with reduction by NADPH according to Eq. 25-18 to form **β-alanine**.[275,275a] The latter can be oxidatively degraded to malonic semialdehyde and malonyl-CoA (see Fig. 17-3),[276] but it also serves as a biosynthetic precursor of pantothenic acid and coenzyme A (Eq. 24-38) and of the peptides carnosine and anserine.



(25-18)

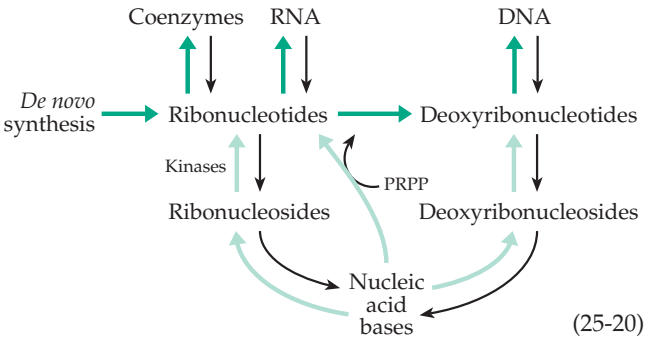
Thymine undergoes degradation in a pathway analogous to that of Eq. 25-18, but with the formation of **3-aminoisobutyrate**. The latter can be oxidatively converted to methylmalonate (Eq. 25-19), which can enter the methylmalonyl pathways (Fig. 17-3).



(25-19)

3. Reuse or Salvage of Nucleic Acid Bases

Almost all organisms except for protozoa synthesize nucleotides via the *de novo* pathways of Figs. 25-14 and 25-15. However, they also receive nucleotides, nucleosides, and free nucleic acid bases from catabolism of RNA and DNA. Both the synthetic and degradative pathways are carefully controlled by cells to ensure that they can grow and repair their nucleic acids and not be inhibited by accumulation of an excess of any component. Furthermore, animals receive an additional supply of preformed bases from their foods. Protozoa receive all of their bases in this way. A large set of enzymes is needed to break down the nucleotides and nucleosides. Another group of enzymes resynthesize nucleotides through **salvage pathways**.[273,277] This keeps levels of inhibitory compounds low, and at the same time produces a small constant pool of nucleotide triphosphates ready for use in nucleic acid synthesis. These pathways are summarized in a general way in Eq. 25-20, which uses dark green arrows for biosynthetic pathways, light green for salvage, and black for degradative reactions (after Kornberg and Baker[273]). The very active salvage pathways of protozoa have provided a variety of targets for inhibitors aimed at parasitic protozoa such as trypanosomes and *Toxoplasma gondii*. The target enzymes often have structures sufficiently different from the corresponding human enzymes to allow for design of selective inhibitors that can serve as drugs.[278,279]



(25-20)

Just as orotic acid is converted to a ribonucleotide in step *e* of Fig. 25-14, other free pyrimidine and purine bases can react with PRPP to give monoribonucleotides plus PP_i. The reversible reactions, which are catalyzed by **phosphoribosyltransferases** (ribonucleotide pyrophosphorylases), are important components of the salvage pathways by which purine and pyrimidine bases freed by the degradation of nucleic acids are recycled.[273] However, thymine is usually *not* reused. Thymine will react with deoxribose 1-*P* to form thymidine plus inorganic phosphate (thymidine phosphorylase), and thymidine is rapidly

phosphorylated by the action of successive kinases to dTTP; a substrate for DNA polymerases.[273] This has allowed biochemists to introduce radioactive thymine or thymidine into the DNA of an organism, an important experimental tool. Another important reaction of the salvage pathways for pyrimidines is the conversion of cytosine to uracil, the same kind of hydrolytic deamination represented by step *l* in Fig. 25-14.

E. Biosynthesis and Metabolism of Purines

The first decisive experiments shedding light on the biosynthetic origins of purines were done with pigeons, which form large amounts of uric acid. Labeling experiments established the complex pattern indicated in the box in the upper left-hand corner of Fig. 25-15. Two carbon atoms were found derived from glycine, one from CO_2, and two from formate. One nitrogen came from glycine, two from glutamine, and one from aspartate. In the case of adenine, the 6-NH_2 group was also found derived from aspartate.

1. The Enzymatic Reactions of Purine Synthesis

The detailed biosynthetic pathway, for which enzymes have now been isolated and studied, is indicated in Fig. 25-15. The first "committed step" in purine synthesis is the reaction of PRPP with glutamine to form **phosphoribosylamine** (step *a*).[280,281] This is another glutamine-dependent amination, pyrophosphate being displaced by ammonia generated *in situ* from glutamine. The amino group of the intermediate so-formed is coupled with glycine in a standard manner (step *b*),[282] and the resulting product is formylated by 10-formyltetrahydrofolate (step *c*).[283–285b] The latter can be generated from free formate,[286] accounting for the labeling pattern indicated in the box in Fig. 25-15. For many years it was accepted, incorrectly, that 5,10-methenyl-THF was the formyl donor for this reaction.

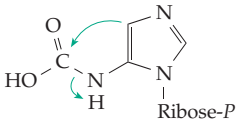
In step *d* of Fig. 25-15 a second glutamine-dependent amination takes place, possibly through aminolysis of an intermediate enol phosphate. An ATP-requiring ring closure and tautomerization (step *e*) serve to complete the formation of the imidazole ring.[287,287a] In many, perhaps all, eukaryotes a single multifunctional enzyme catalyzes steps *b*, *c*, and *e* of Fig. 25-15. The chicken enzyme has ~110-kDa subunits.[288] The product is a ribonucleotide of 5-amino-4-imidazole-4-carboxamide, (AIR), a compound that was isolated in 1945 from cultures of *E. coli* treated with sulfonamides. The latter are important drugs which are antagonists of *p*-aminobenzoate (Box 9-C) and interfere with completion of purine synthesis. This deprives the bacterial cells of essential folic acid derivatives. Its structure immediately suggested that 5-aminoimidazole-4-carboxamide might be a purine precursor. Later it was shown that it is actually the corresponding ribonucleotide (AIR or ZMP) that lies on the main route of purine synthesis. It also is an intermediate in the biosynthesis of thiamin (Fig. 25-21). Free 5-aminoimidazole-4-carboxamide participates in formation of histidine (Fig. 25-13).

The trinucleotide ZTP also accumulates, not only in bacteria but also in many eukaryotic cells. Bochner and Ames suggested it may be an **alarmone** signaling a deficit of folate coenzymes in the cell and causing a shutdown of protein synthesis. ZTP is synthesized by an unusual reaction, transfer of a pyrophosphate group from PRPP (phosphoribosyl pyrophosphate).

$$\text{ZMP} + \text{PRPP} \rightarrow \text{ZTP} + \text{Ribose 5'-}P \qquad (25\text{-}21)$$

This is similar to the reaction by which guanosine 5'-diphosphate 3'-diphosphate (ppGpp) is formed from GDP and ATP (Eq. 29-11).

In the next step of purine synthesis (Fig. 25-15, step *f*) a molecule of CO_2 is incorporated in an unusual type of carboxylation. It is shown in Fig. 25-15 as a single-step direct reaction of CO_2 with AIR. However, in many organisms it is a two-step ATP-dependent reaction to form a compound carboxylated on the 5-amino group. This rearranges[289–290a] to the product shown in Fig. 25-15.



While the arrows on the foregoing structure suggest a possible mechanism of rearrangement, the implied four-membered ring transition state makes it unlikely. The reaction resembles biotin-dependent carboxylations, suggesting the possibility that the carboxylate releases CO_2, which moves and rebinds while trapped within a closed active site.[290] In a two-step amination reaction (steps *g* and *h*) nitrogen is transferred from aspartate in a manner strictly comparable to that in urea synthesis in which argininosuccinic acid is an intermediate (Fig. 24-10). As in urea formation, the carbon skeleton of the aspartate molecule is eliminated as fumarate (step *h*), leaving the nitrogen in the purine precursor. The final carbon atom is added from 10-formyltetrahydrofolic acid (step *i*).[291–292b] Spontaneous ring closure is followed by dehydration to **inosine 5'-phosphate** (IMP, inosinic acid), step *j*. Steps *i* and *j* are catalyzed by a single bifunctional enzyme.

IMP is converted via two different pathways to either AMP or GTP. Conversion to AMP (Fig. 25-15, steps *k* and *l*) occurs via another two-step aspartate-dependent amination.[292c,d] The intermediate is

adenylosuccinate.[293–294b] Positional isotope exchange studies using [γ-^{18}O]GTP suggested that the enolic 6-phospho-IMP is an intermediate comparable to that in Eq. 24-23.[295,296] X-ray studies have confirmed the prediction.[297] An NAD$^+$-dependent oxidation converts IMP to the corresponding **xanthine** ribonucleotide

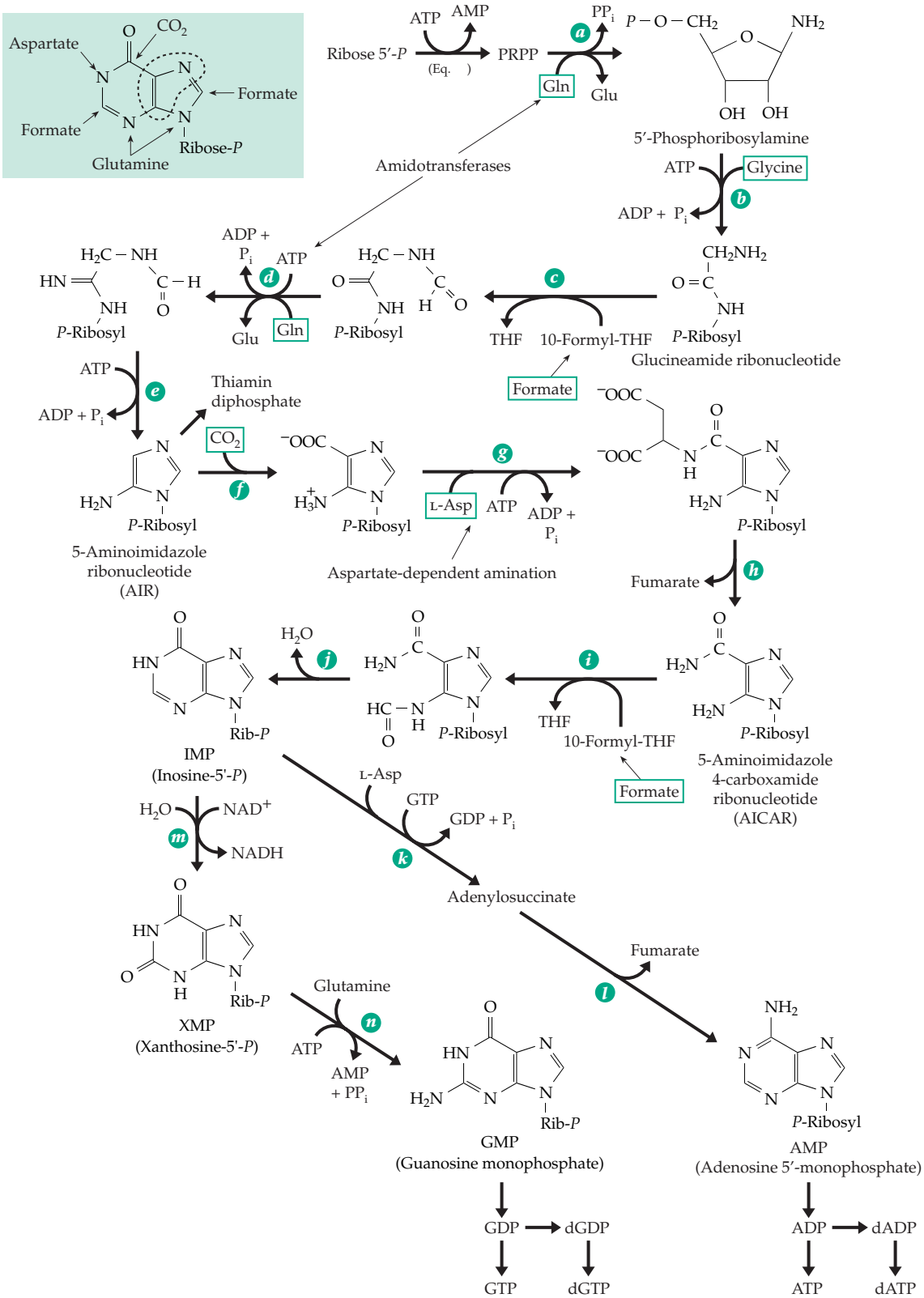


Figure 25-15 Biosynthesis of purine nucleotides from ribose 5'-phosphate.

(step *m*),[297a–d] which is aminated in a glutamine-dependent process,[298–300] as indicated (step *n*).

Synthesis of purines is under complex control.[273] Some of the mechanisms found in bacteria are outlined in Fig. 25-16. Both feedback inhibition and activation are involved. Very important is the fact that GTP is needed in the synthesis of ATP, and that ATP is needed for synthesis of GTP. This kind of control ensures that an excess of either nucleotide will not be formed for long. In bacteria all of the final end product nucleotides inhibit the initial reaction of step *a* in Fig. 25-15.

Modified purine nucleosides are important constituents of transfer RNAs (Fig. 5-33; Chapter 28, Section A,7). Among them are the 7-deazaguanosines **queuosine** (Fig. 5-33) and **archaeosine**, which contains a 7-formamidino group. These nucleosides are incorporated into tRNAs by an exchange mechanism catalyzed by **tRNA-guanine** transglycosylase. This exchanges a precursor of queuine (7-aminomethyl-7-deazaguanine) for guanine in selected residues in the tRNAs. In eukaryotic tRNAs free queuine is exchanged into the tRNAs whereas in Archaea the archaeosine precursor is incorporated.[301a,b] The conversion of the precursors to the final modified bases occurs in the tRNAs (Eq. 28-4). Many other purine derivatives are found in nature, e.g., **puromycin** (Box 29-B), which is formed from adenosine.[301c]

2. The Purine Nucleotide Cycle and Salvage Pathways for Purines

Muscular work is accompanied by the production of ammonia, the immediate source of which is adenosine 5'-phosphate (AMP).[301,302] This fact led to the recognition of another substrate cycle (Chapter 11) that functions by virtue of the presence of a biosynthetic pathway and of a degradative enzyme in the same cells (cycle A, Fig. 25-17). This **purine nucleotide cycle** operates in the brain[303,304] as well as in muscle. The key enzyme 5'-AMP aminohydrolase (AMP deaminase; step *a*, Fig. 25-17) also occurs in erythrocytes and many other tissues.[304,305] Persons having normal erythrocyte levels but an absence of this enzyme in muscles suffer from muscular weakness and cramping after exercise.[306]

Purine bases from ingested foods, or formed by catabolism of nucleic acids, are able to react with PRPP under the influence of phosphoribosyltransferases.[306a] Two such enzymes are known to act on purines. One converts adenine to AMP (Fig. 25-17, step *b*) and also acts upon 5-aminoimidazole-4-carboxamide. This enzyme may be especially important to parasitic protozoa such as *Leishmania*, which lack the *de novo* pathway of purine synthesis (Fig. 25-15).[278,306b]

AMP can be converted by the action of AMP 5'-nucleotidase to adenosine (step *c*, Fig. 25-17), which is thought to be an important local hormone or second

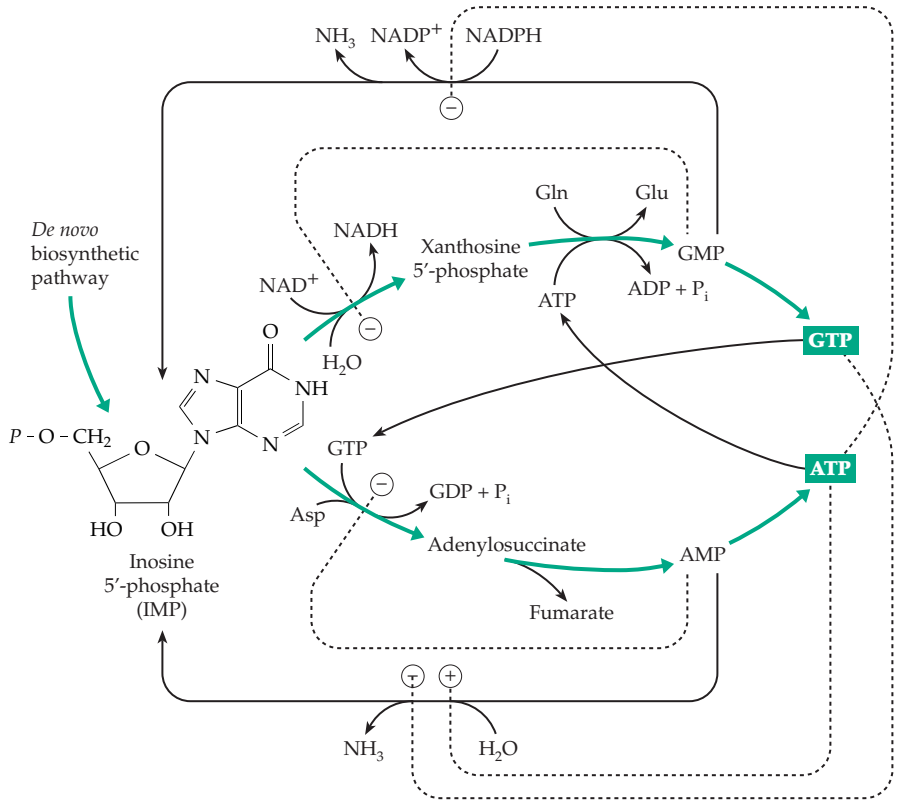


Figure 25-16 Control of the conversion of inosine 5'-phosphate to the adenine and guanine ribonucleotides and deoxyribonucleotides in bacteria by feedback inhibition and activation.

messenger.[307–309] Adenosine has a variety of effects on all organs of an animal. It affects heart rate, smooth muscle tone, and white blood cell function. It modulates the catabolic effects of hormones such as catecholamines and stimulation of the anabolic hormone insulin.[309] Adenosine receptors of at least three types are known.[310,311] Binding of adenosine to A_1 receptors inhibits adenylate cyclase, while binding to A_2 receptors stimulates this enzyme.[307] However, effects of adenosine on K^+ transport are probably more important.

Deamination of adenosine (step *d*) together with reconversion of the resulting inosine to IMP (steps *e* and *f*) completes a second cycle (cycle B, Fig. 25-17). Intense interest has been focused on adenosine deaminase because hereditary lack of this enzyme is linked to a severe immunodeficiency in which the numbers of B and T lymphocytes are inadequate to combat infection.[312] Until recently bone marrow transplantation in infancy was the only possible treatment for this otherwise fatal disease. Regular injections of adenosine deaminase covalently attached to polyethylene glycol (to delay removal from the bloodstream) have been used for some patients. Since 1990 **gene therapy**, transfer of an adenosine deaminase gene into white blood cells, has also been used with apparent success.[313–315] This topic is discussed in Chapter 31. Adenosine deaminase is a 40-kDa protein,[316,317] which exists as a complex with a large 200-kDa binding protein[315] which anchors the deaminase to cell membranes.

The basic cause of the severe immunodeficiency symptoms is uncertain. However, adenosine deaminase also catalyzes hydrolysis of 2'-deoxyadenosine, and in the absence of the enzyme both this compound and its trinucleotide precursor 2'-deoxy-ATP (dATP) accumulate in tissues.[312] Ribonucleotide reductase is allosterically inhibited by dATP, and this inhibition may interfere with DNA synthesis and with the rapid growth of lymphocytes needed in response to infections. Since T lymphocytes are more severely affected than B lymphocytes, it is necessary to postulate a difference in the extent to which these two cell types accumulate dATP.

The conversions of inosine to hypoxanthine (Fig. 25-17, step *e*), of guanosine to guanine (step *g*), and of other purine ribonucleosides and deoxyribonucleosides to free purine bases are catalyzed by **purine nucleoside phosphorylase**.[318–321b] Absence of this enzyme also causes a severe immune deficiency which involves the T cells. However, B cell function is not impaired.[312,315,322]

The last enzyme in cycle B of Fig. 25-17 (catalyzing step *f* and also step *b*) is the X-linked **hypoxanthine-guanine phosphoribosyltransferase** (HGPRT or HPRT).[322a,b] Its absence causes the **Lesch–Nyhan** syndrome characterized not only by overproduction of uric acid but by a serious disorder of the central nervous system. It causes both mental retardation and a compulsive form of self-mutilation of the gums and hands by biting.[323–325] The excessive production of uric acid is easy to understand because the accumulating hypoxanthine and guanine are both readily converted to uric acid by the reactions of Fig. 25-18. Patients with a partial deficiency in HGPRT escape the worst neurological symptoms but may have severe gouty arthritis (Box 25-D).[326] Efforts are being made to treat the disease by gene transfer.[327]

Trypanosomes and other parasitic protozoa are unable to synthesize purines and must obtain them from their hosts using salvage pathway. Selective inhibition of their HGPRT or of nucleoside hydrolases, which are absent from mammalian cells, are goals of drug development.[327a,b]

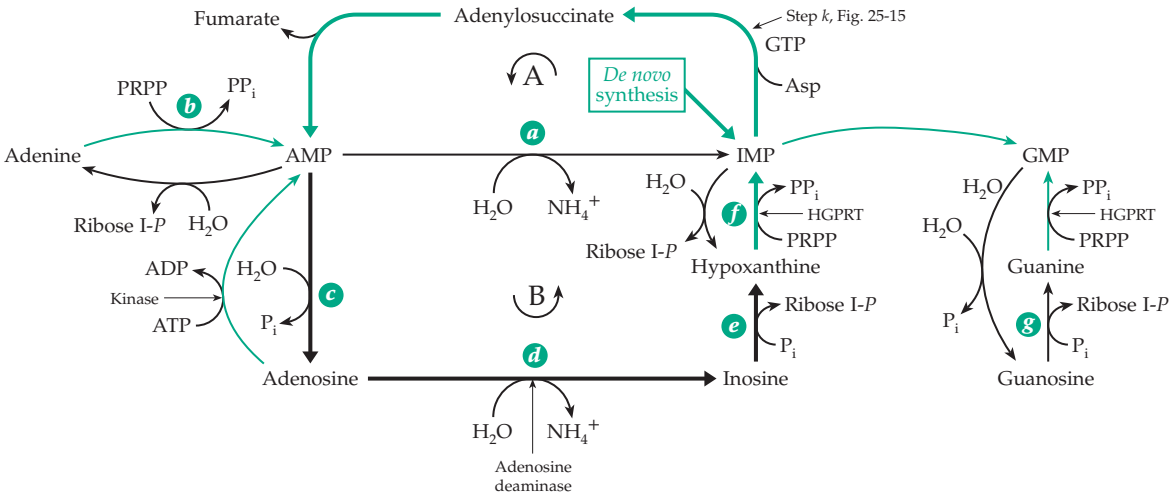


Figure 25-17 Some purine salvage pathways and related reactions. Green lines indicate biosynthetic pathways.

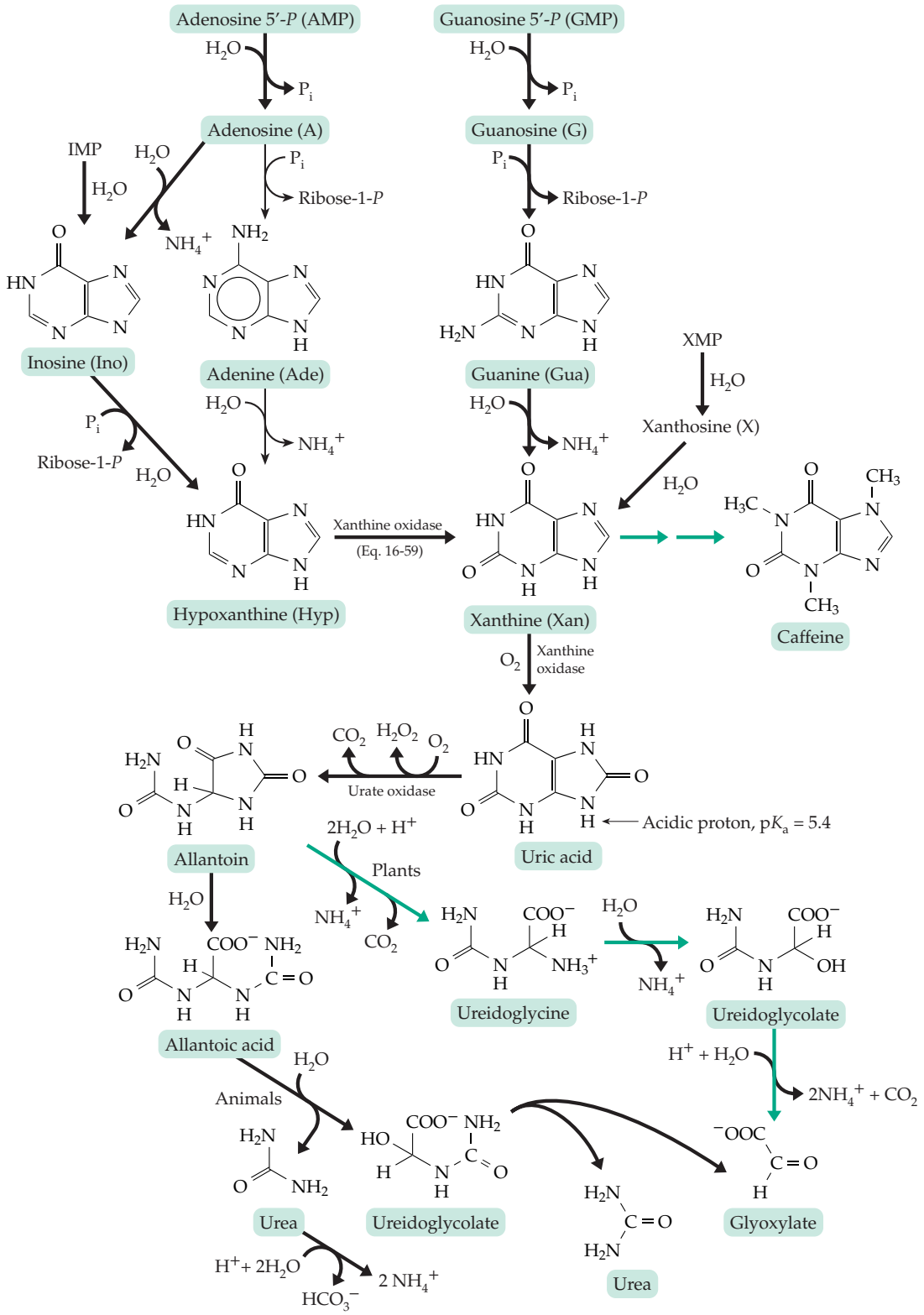


Figure 25-18 Pathways of catabolism of purine nucleotides, nucleosides, and free bases. Spiders excrete xanthine while mammals and birds excrete uric acid. Spiders and birds convert all of their excess nitrogen via the *de novo* pathway of Fig. 25-15 into purines. Many animals excrete allantoin, urea, or NH_4^+. Some legumes utilize the pathway marked by green arrows in their nitrogen transport via ureides.

3. Oxidative Metabolism of Purines

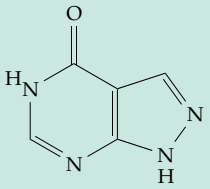
As indicated in Fig. 25-18, free adenine released from catabolism of nucleic acids can be deaminated hydrolytically to hypoxanthine, and guanine can be deaminated to xanthine.[328] The molybdenum-containing xanthine oxidase (Chapter 16) oxidizes hypoxanthine to xanthine and the latter on to uric acid. Some clostridia convert purine or hypoxanthine to xanthine by the action of a selenium-containing purine hydroxylase.[328a] Another reaction of xanthine occurring in some plants is conversion to the trimethylated derivative **caffeine**.[328b] One of the physiological effects of caffeine in animals is inhibition of pyrimidine synthesis.[329] However, the effect most sought by coffee drinkers may be an increase in blood pressure caused by occupancy of adenosine receptors by caffeine.[330]

Uric acid is the end product of purine metabolism in the human. Spiders excrete **xanthine**,[331,332] but in most animals **urate oxidase** converts uric acid to **allantoin**. Although urate oxidase contains no coenzymes or metal cofactors,[333] it catalyzes the reaction with O_2 to form a peroxide (Eq. 25-22),[334,335] a reaction resembling that of a reduced flavin (Eq. 15-31) or a reduced pterin. As in these other cases there may be an initial electron transfer between O_2 and urate to form a radical pair which couple. Elimination of H_2O_2, a feature also of flavoprotein oxidases, accomplishes oxidation of the urate ring. Hydration, ring opening, and decarboxylation complete the conversion to allantoin. The ease of formation of urate radicals permits uric acid to act as an effective oxidant. This may account for the fact that we long-lived human beings retain a high internal urate concentration.[335a,b]

BOX 25-D GOUT

A common metabolic derangement with an incidence of ~3 per 1000 persons is **hyperuricemia** or **gout**.[a,b] As with most metabolic defects, there is a family of diseases ranging from mild to severe. In acute gouty arthritis, a sudden attack occurs, usually in the night, when sodium urate crystals precipitate in one or more joints. In half the cases the victim is awakened by a terrible pain in the big toe. The disease most often strikes adult males. The heredity is apparently complex and not fully understood. The primary biochemical defect in gout is usually an overproduction of uric acid which, in some cases, may result from an overactive PRPP synthase.[c] In other cases a kidney defect interferes with excretion. The less severe saturnine gout, which occurs in relatively young persons of both sexes, is a result of chronic lead poisoning. It may involve deposition of guanine in the joints as a result of inhibition by Pb^{2+} of guanine aminohydrolase, the enzyme that hydrolyzes guanine to xanthine (Fig. 25-18).[d]

If properly controlled, simple gout may have few adverse effects. However, the severe neurological symptoms of Lesch–Nyhan syndrome (Section E,2 of text)[e] cannot be corrected by medication. Colchicine (Box 7-D), in a manner which is not understood, alleviates the painful symptoms of gout caused by the deposits of sodium urate in joints and tissues. It is also important to keep the dietary purine intake low and it is often necessary to inhibit xanthine oxidase. A widely used and effective inhibitor is the isomer of hypoxanthine known as **allopurinol**, which is taken daily in amounts of 100–600 mg or more.



Allopurinol

Allopurinol and its oxidation product **oxypurinol**, a xanthine analog, both inhibit xanthine oxidase and patients receiving allopurinol excrete much of their purines as xanthine and hypoxanthine. Nucleotide derivatives of oxypurinol also inhibit the *de novo* purine biosynthetic pathway. The accumulating hypoxanthine is reused to a greater extent than normal, decreasing the total purine excretion. A number of other drugs stimulate increased excretion of uric acid.[e] Although many patients tolerate allopurinol for many years, some experience dangerous side effects.

---

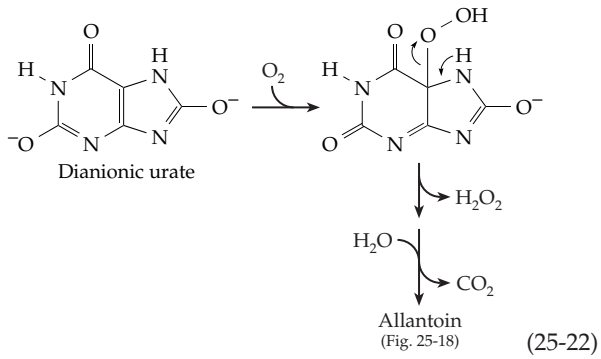
[a] Becker, M. A., and Roessler, B. J. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1655–1677, McGraw-Hill, New York

[b] Kelley, W. N., and Wyngaarden, J. B. (1972) in *The Metabolic Basis of Inherited Disease*, 3rd ed. (Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S., eds), pp. 969–1002, McGraw-Hill, New York

[c] Becker, M. A., Kostel, P. J., and Meyer, L. J. (1975) *J. Biol. Chem.* **250**, 6822–6830

[d] Farkas, W. T., Stanawitz, T., and Scheider, M. (1978) *Science* **199**, 786–787

[e] Rossiter, B. J. F., and Caskey, C. T. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1679–1706, McGraw-Hill, New York



(25-22)

Allantoin is the excretory product in most mammals other than primates. Most fish hydrolyze allantoin to **allantoic acid,** and some excrete that compound as an end product. However, most continue the hydrolysis to form urea and glyoxylate using peroxisomal enzymes.[336] In some invertebrates the urea may be hydrolyzed further to ammonia. In organisms that hydrolyze uric acid to urea or ammonia, this pathway is used only for degradation of purines from nucleotides. Excess nitrogen from catabolism of amino acids either is excreted directly as ammonia or is converted to urea by the urea cycle (Fig. 24-10).

Plants also form the **ureides** allantoin and allantoic acid, and in some legumes, such as soy beans, these compounds account for 70–80% of the organic nitrogen in the xylem. They appear to function in nitrogen transport.[337] As indicated in Fig. 25-18, the hydrolysis to glyoxylate, NH_4^+, and CO_2 follows a different pathway than in animals. See also Chapter 24, Section C.

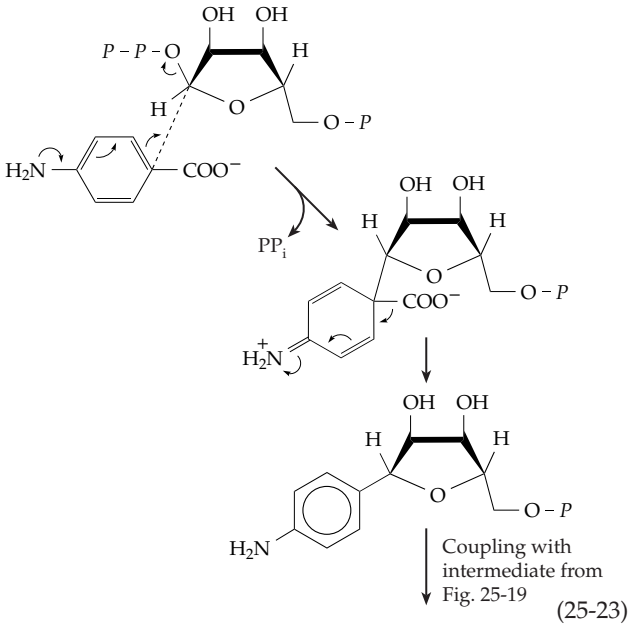
F. Pterins, Flavins, Dimethylbenzimidazole, Thiamin, and Vitamin B_6

Tracer studies have established that both folic acid and riboflavin originate from guanosine phosphates (Fig. 25-19).[338–341] All of the atoms of the purine ring are conserved in the products except for C-8 of the five-membered ring. The first step is opening of the 5-membered ring and the hydrolytic removal of formate (Fig. 25-19, step *a*). This is followed by an Amadori rearrangement (step *b*) and a simple ring closure between the resulting carbonyl and adjacent amino group (step c). The product is **7,8-dihydroneopterin triphosphate**. The single enzyme **GTP cyclohydrolase** catalyzes all three steps *a*–*c*.[342,342a–c] Dihydroneopterin is a central intermediate in pterin metabolism, being converted by one route into folate and methanopterin coenzymes and by another into biopterin, drosopterin, and others.[343]

An aldol cleavage (step *d*) followed by a series of reactions, shown at the left in in Fig. 25-19, leads to folate.[344–347] The reactions include the ATP-dependent conversion of 6-hydroxy-7,8-dihydropterin to its pyrophosphate ester.[347a] This is followed by coupling to *p*-aminobenzoate with elimination of PP_i to form dihydropteroic acid.[347b] ATP-dependent joining of glutamate and reduction yields **tetrahydrofolate**. Additional molecules of γ-linked glutamate are added to form the functional polyglutamate forms (p. 803).[347c,d] Gamma-glutamyl hydrolases provide essential turnover.[347e] Formation of biopterin is initiated by tautomerization of dihydroneopterin triphosphate (step *e*), a proton from the solvent binding to C-6 of the tetrahydropterin ring.[348] The single Zn^{2+}-dependent enzyme **6-pyruvoyltetrahydropterin synthase**[349–352] catalyzes the two consecutive tautomerization steps shown in Fig. 25-19 as well as elimination of the triphosphate group tripolyphosphate (step *f*), a reaction facilitated by the carbonyl group introduced in the preceding step.[349] The same enzyme promotes a final tautomerization to form 6-pyruvoyltetrahydropterin, a compound that is reduced by NADPH to tetrahydrobiopterin. The reaction is catalyzed by **sepiapterin reductase**.[353,354] Notice that biopterin, like the folates, is synthesized at the oxidation level of a tetrahydrobiopterin.

The *Drosophila* eye pigments **sepiapterin** and **drosopterins** (Figs. 15-17 and 25-19) arise from 6-pyruvoyltetrahydropterin.[355–357] Reduced glutathione appears to be the reducing agent needed to convert the 6-pyruvoyltetrahydropterin into the more reduced pyrimidinodiazepine (step *h*) with its 7-membered ring (Fig. 25-19).[357,358] Tetrahydrobiopterin can arise in mammalian cells, not only by the *de novo* pathway of Fig. 25-19 but also from salvage of sepiapterin.

Methanopterin (Fig. 15-17) is formed via a branch in the folate pathway. The 5-carbon chain that replaces the carboxyl group of *p*-aminobenzoate in the folates is derived by reaction of PRPP with *p*-aminobenzoate (Eq. 25-23).[359,360] White has proposed a



(25-23)

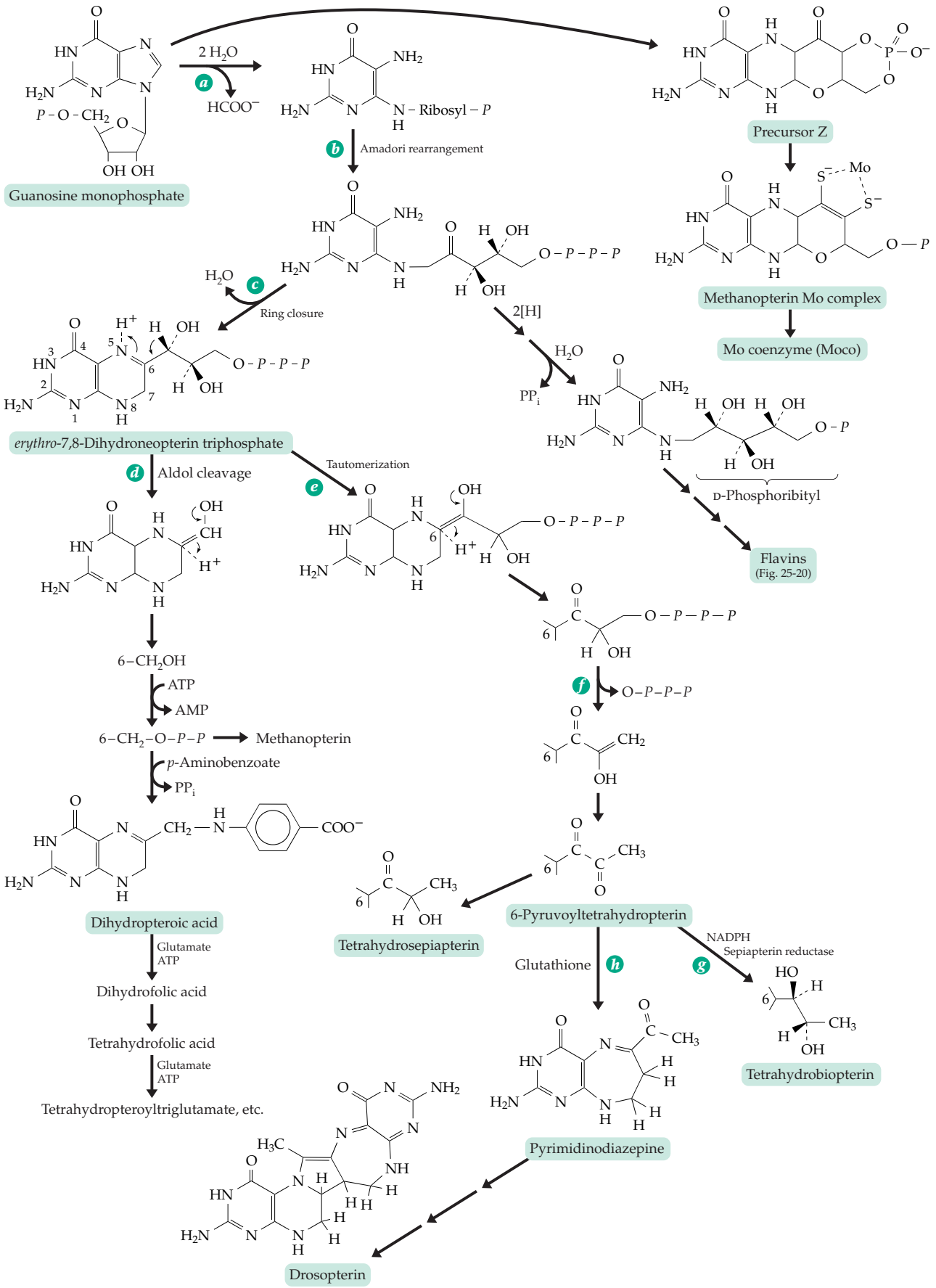


Figure 25-19 The biosynthesis of folic acid and other pterins.

detailed pathway for completion of the methanopterin synthesis.[359] After coupling with 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (shown in abbreviated form in Fig. 25-19) with loss of PP_i, the ribose ring in the product of Eq. 25-23 is reductively opened to give the ribityl side chain of methanopterin (see Fig. 15-17). Hydrolytic removal of the phosphate is then followed by transfer of the α-linked ribose phosphate. The *S*-hydroxyglutarate (derived by reduction of 2-oxoglutarate) is added in an ATP-dependent reaction. The extra methyl groups at positions 7 and 9 are transferred from *S*-adenosylmethione and the compound is reduced to the tetrahydropterin state.

Biosynthesis of **molybdopterin** (Fig. 15-17) also begins with a guanosine derivative, quite possibly GMP as is shown in Fig. 25-19. However, the C-8 carbon atom is not lost as formate, as in step *a* of Fig. 25-19, but is incorporated into the side chain in the molybdopterin precursor Z. A pathway was proposed by Wuebbens and Rajagopalan.[361] The first stage in the reaction sequence,[361a–d] which is identical in most organisms, is formation of the metastable compound Z, a proposed structure of which is shown in Fig. 25-19. The conversion of this compound into molybdopterin requires opening of the cyclic phosphate ring and incorporation of two atoms of sulfur. These may both be released from cysteine as S^0 and carried by a sulfurtransferase as an enzyme-bound persulfide group (see Chapter 24, Section G,3).[361b] A thiocarboxylate group generated from a C-terminal carboxylate of a molybdopterin synthase subunit, as in the ThiS protein (p. 1463), may be the direct sulfur donor. It is probably formed in an ATP-dependent process.[361b] Incorporation of molybdenum, perhaps from MoO_4^{2-}, completes the synthesis of the molybdenum cofactor Moco.[361c,e] In some molybdoenzymes, e.g., xanthine oxidase, an additional sulfur atom is bound to the Mo atom (Fig. 16-32). This is also obtained from cysteine using a PLP-dependent sulfurtransferase similar to the NifS protein.[361f,g] In many bacteria molybdopterin is joined to GMP, AMP, IMP, or CMP to form a dinucleotide.[361h,i]

Both the fungus *Eremothecium* (Box 15-B) and mutants of *Saccharomyces* have been used to deduce the pathways of **riboflavin** synthesis outlined in Figure 25-20. The first reaction (step *a*) is identical to step *a* of Fig. 25-19 but is catalyzed by a different GTP cyclohydrolase.[362] Instead of an Amadori rearrangement it catalyzes the hydrolytic deamination and dephosphorylation (step *b*) to give the flavin precursor 4-ribitylamino-5-amino-2,6-dihydroxypyrimidine. Additional carbon atoms to build the benzene ring of riboflavin are supplied in two stages from ribulose 5-phosphate. Isotopic labeling showed that carbon atoms 1, 2, 3, and 5 of this compound are utilized as is marked in Fig. 25-20, while C-4 is eliminated as formic acid in a rearrangement (step *c*, Fig. 25-20). A plausible mechanism has been suggested.[363,363a,b] The product L-3,4-dihydroxy-2-butanone 4-phosphate condenses (step *d*) with the product formed from GMP, possibly via the Schiff base shown. Elimination of H_2O and inorganic phosphate followed by tautomerization gives **6,7-dimethyl-8-ribityllumazine**.[363a,364–364c] Completion of the flavin ring requires an additional four carbon atoms, which are supplied by a second molecule of 6,7-dimethyl-8-ribityllumazine, as indicated in Fig. 25-20.[364d] This disproportionation reaction appears remarkable but is less so when one considers that the bimolecular reaction to form riboflavin occurs spontaneously under mild conditions. The precursor 4-ribitylamino-5-amino-2,6-dihydroxypyrimidine is regenerated in this process (Fig. 25-20).

The enzyme complex that catalyses steps *d* to *f* of Fig. 25-20 has an unusual composition. An α_3 trimer of 23.5-kDa subunits is contained within an icosahedral shell of 60 16-kDa β subunits, similar to the protein coats of the icosahedral viruses (Chapter 7). The β subunits catalyze the formation of dimethylribityllumazine (steps *d*, *e*), while the α_3 trimer catalyzes the dismutation reaction of step *f*, the final step in riboflavin formation.[365] A separate bifunctional bacterial ATP-dependent synthetase phosphorylates riboflavin and adds the adenylyl group to form FAD.[366] Two separate mammalian enzymes are required.[367] Synthesis of **deazaflavins** of methanogens (Fig. 15-22) follows pathways similar to those of riboflavin. However, the phenolic ring of the deazaflavin originates from the shikimate pathway.[368]

Dimethylbenzimidazole, a constituent of vitamin B_{12} (Box 16-B), also arises from 6,7-dimethyl-8-ribityllumazine in a process resembling that of riboflavin synthesis, but in which the riboflavin formed is hydrolytically degraded to remove the pyrimidine ring and to form the imidazole ring.[369] Conversion to an α-ribazole and linkage to the aminopropyl group of the corrin ring is described by Thompson *et al*.[370] Various related cobamides are also formed by bacteria.[371] Synthesis of the corrin ring is described briefly in Chapter 24, and the chemistry of the ligands to cobalt at the "top" of the vitamin B_{12} molecules is considered in Chapter 16, Section B.

Thiamin. Investigation of the biosynthesis of thiamin has been difficult because only minute amounts are formed by microorganisms such as *E. coli* or yeast. Furthermore, significant differences in the routes of synthesis in different organisms have caused confusion.[372,372a] The pathways outlined in Fig. 25-21 are incomplete.

The pyrimidine portion of thiamin (Fig. 25-21) is distinct in structure from the pyrimidines of nucleic acids. In bacteria it originates from the purine precursor 5-aminoimidazole ribotide, which is converted into a hydroxymethylpyrimidine (Fig. 25-21)[373] which is

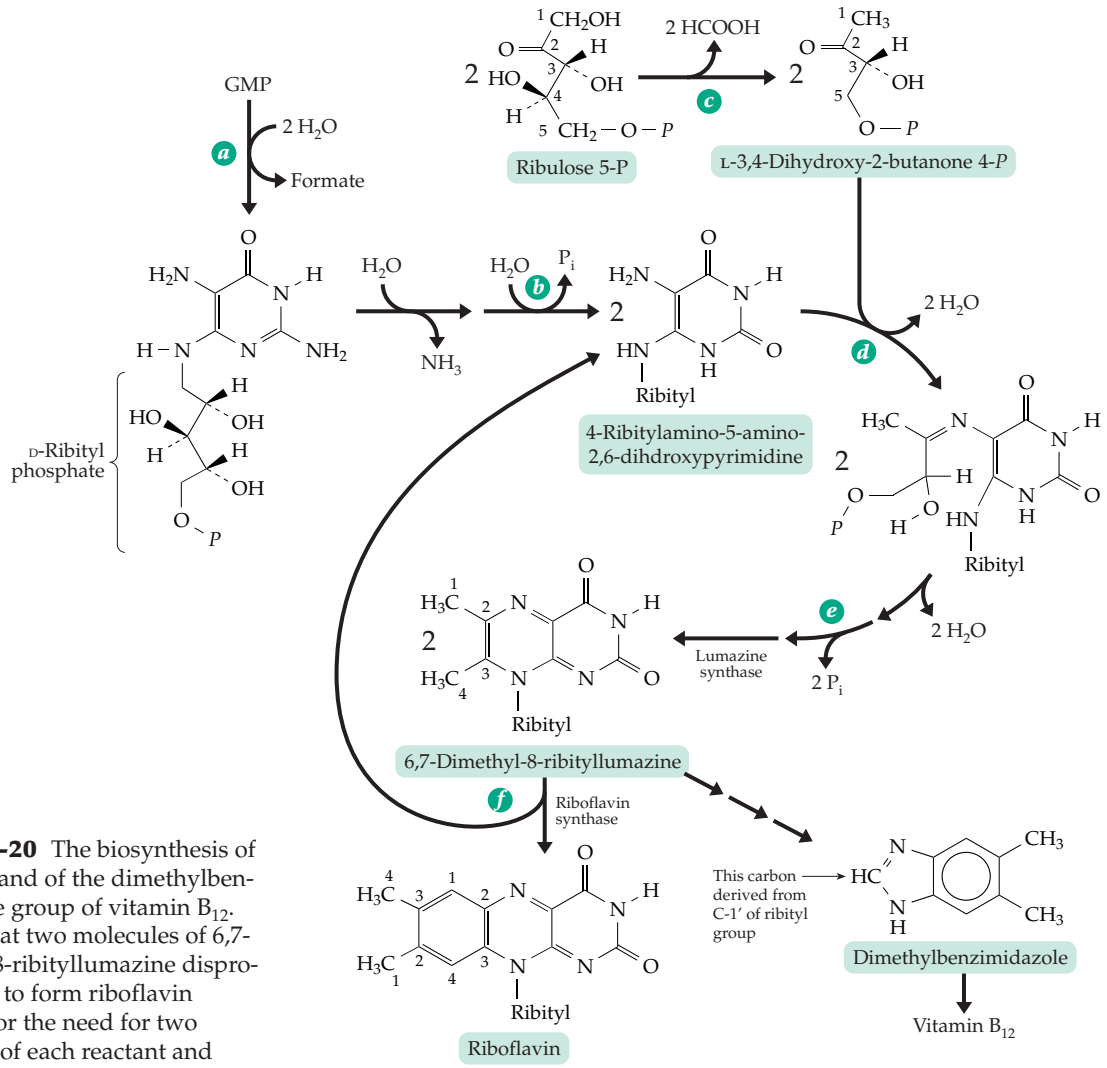
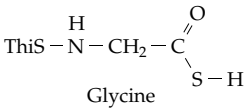


Figure 25-20 The biosynthesis of riboflavin and of the dimethylbenzimidazole group of vitamin B_{12}. The fact that two molecules of 6,7-dimethyl-8-ribityllumazine disproportionate to form riboflavin accounts for the need for two molecules of each reactant and product in many steps.

coupled with the thiazolium ring to form the vitamin. All of the carbon atoms of the substituted pyrimidine can be derived from the 5-aminoimidazole ribotide, but the pathway is uncertain. Both glycine and formate enter the pyrimidine, but labeling patterns are different in *E. coli* and in yeast.[372]

The thiazole ring is assembled on the 5-carbon backbone of **1-deoxyxylulose 5-phosphate**, which is also an intermediate in the alternative biosynthetic pathway for terpenes (Fig. 22-2) and in synthesis of vitamin B_6 (Fig. 25-21). In *E. coli* the sulfur atom of the thiazole comes from cysteine and the nitrogen from tyrosine.[374] The same is true for chloroplasts,[375] whereas in yeast glycine appears to donate the nitrogen.[372] The thiamin biosynthetic operon of *E. coli* contains six genes,[372a,376] one of which (***ThiS***) encodes a protein that serves as a sulfur carrier from cysteine into the thiazole.[374] The C-terminal glycine is converted into a thiocarboxylate:



The *ThiI* gene, which encodes aother sulfurtransferase protein, is also needed.[374a] The enzymology of the insertion of this sulfur into the thiazole is uncertain but may resemble that involved in synthesis of biotin, lipoic acid, and molybdopterin.[374] Linkage of the two parts of the thiamin molecule (step *d*, Fig. 25-21) is catalyzed by thiamin phosphate synthase, evidently via an S_N2 type reaction.[377–377b]

Pyridoxol (vitamin B_6). Again 1-deoxyxylulose 5-*P* serves as a precursor.[378] In *E. coli* only two genes have been implicated in the condensation of this compound with 4-(phosphohydroxy)-L-threonine (Fig. 25-21, step *f*).[378a] One is an NAD^+-dependent dehydrogenase that acts on the second substrate prior to the condensation. Significant differences from the path-

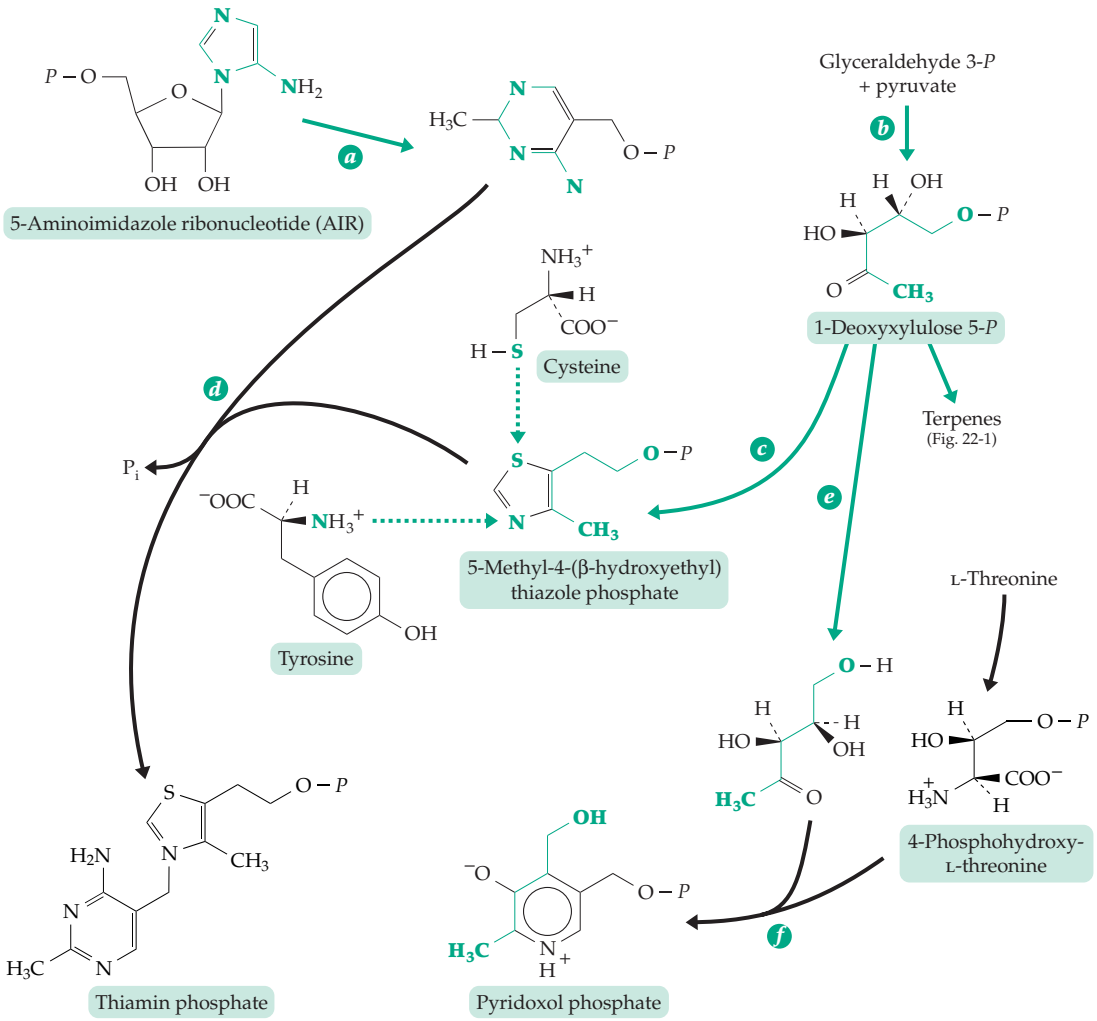
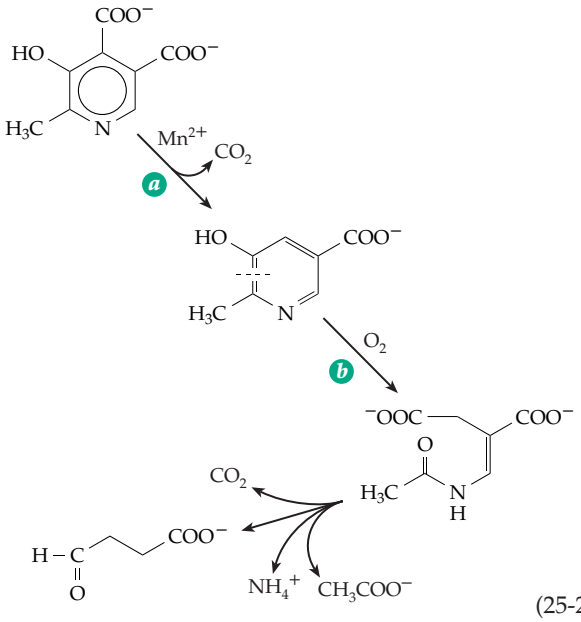


Figure 25-21 Proposed pathways for biosynthesis of thiamin phosphate and pyridoxol phosphate.

way shown in Fig. 25-21 exist in yeast and other fungi.[378b,c] Interconversion of pyridoxal phosphate and other forms of vitamin B_6 is discussed in Chapter 14.

The degradation of pyridoxol by bacteria has been investigated in detail.[379–381] In one pathway the hydroxymethyl group in the 5 position and the substituent in the 4 position are both oxidized in the early steps to carboxylate groups. Then, as indicated in Eq. 25-24, a decarboxylation is followed by the action of an unusual dioxygenase.

Isolated from a strain of *Pseudomonas*, this enzyme contains bound FAD, which must be reduced by external NADH. Like a typical dioxygenase the enzyme introduces two atoms of oxygen into the product. However, it also uses the reduced FAD to reduce the double bond system (either before or after the attack by oxygen).[381] Another enzyme of the same bacterium is remarkable in hydrolyzing the product of the oxygenation reaction to four different products without the accumulation of intermediates.



(25-24)

References

1. Floss, H. G. (1986) in *Recent Advances in Phytochemistry*, Vol. 20 (Conn, E. E., ed), pp. 13–55, Plenum, New York
2. Amrhein, N. (1986) in *Recent Advances in Phytochemistry*, Vol. 20 (Conn, E. E., ed), pp. 83–117, Plenum, New York
3. Marzabadi, M. R., Gruys, K. J., Pansegrau, P. D., Walker, M. C., Yuen, H. K., and Sikorski, J. A. (1996) *Biochemistry* **35**, 4199–4210
4. Davis, B. D. (1950) *Experientia* **6**, 4–50
5. Davis, B. D. (1950) *J. Biol. Chem.* **191**, 315–325
6. Weiss, U., and Edwards, J. M. (1980) *The Biosynthesis of Aromatic Compounds*, Wiley, New York
7. Pittard, A. J. (1987) in *Escherichia coli and Salmonella typhimurium*, Vol. I (Neidhardt, F. C., ed), pp. 368–394, Am. Soc. for Microbiology, Washington, DC
8. Gollub, E., Zalkin, H., and Sprinson, D. B. (1967) *J. Biol. Chem.* **242**, 5323–5328
9. Shultz, J., Hermodson, M. A., Garner, C. C., and Herrmann, K. M. (1984) *J. Biol. Chem.* **259**, 9655–9661
10. DeLeo, A. B., Dayan, J., and Sprinson, D. B. (1973) *J. Biol. Chem.* **248**, 2344–2353

10a. Yu, T.-W., Müller, R., Müller, M., Zhang, X., Draeger, G., Kim, C.-G., Leistner, E., and Floss, H. G. (2001) *J. Biol. Chem.* **276**, 12546–12555

10b. Guo, J., and Frost, J. W. (2002) *J. Am. Chem. Soc.* **124**, 10642–10643

10c. Eads, J. C., Beeby, M., Scapin, G., Yu, T.-W., and Floss, H. G. (1999) *Biochemistry* **38**, 9840–9849

10d. Arakawa, K., Müller, R., Mahmud, T., Yu, T-W., and Floss, H. G. (2002) *J. Am. Chem. Soc.* **124**, 10644 – 10645

11. Shultz, J., Hermodson, M. A., Garner, C. C., and Harrmann, K. M. (1984) *J. Biol. Chem.* **259**, 9655–9661
12. Dyer, W. E., Weaver, L. M., Zhao, J., Kuhn, D. N., Weller, S. C., and Harrmann, K. M. (1990) *J. Biol. Chem.* **265**, 1608–1614

12a. Wagner, T., Shumilin, I. A., Bauerle, R., and Kretsinger, R. H. (2000) *J. Mol. Biol.* **301**, 389–399

13. Herrmann, K. M., Schultz, J., and Hermodson, M. A. (1980) *J. Biol. Chem.* **255**, 7079–7081
14. Montchamp, J.-L., and Frost, J. W. (1997) *J. Am. Chem. Soc.* **119**, 7645–7653
15. Carpenter, E. P., Hawkins, A. R., Frost, J. W., and Brown, K. A. (1998) *Nature (London)* **394**, 299–302
16. Maitra, U. S., and Sprinson, D. B. (1978) *J. Biol. Chem.* **253**, 5426–5430
17. Widlanski, T., Bender, S. L., and Knowles, J. R. (1989) *Biochemistry* **28**, 7572–7582
18. Ranjeva, R., Refeno, G., Boudet, A. M., and Marmé, D. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5222–5224
19. Chaudhuri, S., and Coggins, J. R. (1985) *Biochem. J.* **226**, 217–223
20. Huang, L., Montoya, A. L., and Nester, E. W. (1975) *J. Biol. Chem.* **250**, 7675–7681
21. Krell, T., Coggins, J. R., and Lapthorn, A. J. (1998) *J. Mol. Biol.* **278**, 983–997
22. Ramilo, C., Appleyard, R. J., Wanke, C., Krekel, F., Amrhein, N., and Evans, J. N. S. (1994) *Biochemistry* **33**, 15071–15079
23. Kim, D. H., Tucker-Kellogg, G. W., Lees, W. J., and Walsh, C. T. (1996) *Biochemistry* **35**, 5435–5440
24. Lee, J. J., Asano, Y., Shiek, T.-L., Spreafico, F., Lee, K., and Floss, H. G. (1984) *J. Am. Chem. Soc.* **106**, 3367–3368
25. Jakeman, D. L., Mitchell, D. J., Shuttleworth, W. A., and Evans, J. N. S. (1998) *Biochemistry* **37**, 12012–12019
26. 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., Radgette, S. R., and Kishore, G. M. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5046–5050
27. Christensen, A. M., and Schaefer, J. (1993) *Biochemistry* **32**, 2868–2873

27a. Alibhai, M. F., and Stallings, W. C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2944–2946

28. Stalker, D. M., Hiah, W. R., and Comai, L. (1985) *J. Biol. Chem.* **260**, 4724–4728
29. Shuttleworth, W. A., Pohl, M. E., Helms, G. L., Jakeman, D. L., and Evans, J. N. S. (1999) *Biochemistry* **38**, 296–302
30. Morrell, H., Clark, M. J., Knowles, P. F., and Sprinson, D. B. (1967) *J. Biol. Chem.* **242**, 82–90
31. Balasubramanian, S., Coggins, J. R., and Abell, C. (1995) *Biochemistry* **34**, 341–348
32. Macheroux, P., Petersen, J., Bornemann, S., Lowe, D. J., and Thorneley, R. N. F. (1996) *Biochemistry* **35**, 1643–1652
33. Lauhon, C. T., and Bartlett, P. A. (1994) *Biochemistry* **33**, 14100–14108

33a. Osborne, A., Thorneley, R. N. F., Abell, C., and Bornemann, S. (2000) *J. Biol. Chem.* **275**, 35825–35830

33b. Kitzing, K., Macheroux, P., and Amrhein, N. (2001) *J. Biol. Chem.* **276**, 42658–42666

34. Hudson, G. S., Wong, V., and Davidson, B. E. (1984) *Biochemistry* **23**, 6240–6249
35. Hermes, J. D., Tipton, P. A., Fisher, M. A., O'Leary, M. H., Morrison, J. F., and Cleland, W. W. (1984) *Biochemistry* **23**, 6263–6275
36. Christendat, D., Saridakis, V. C., and Turnbull, J. L. (1998) *Biochemistry* **37**, 15703–15712
37. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Biochemistry* **37**, 10062–10073

37a. Guo, H., Cui, Q., Lipscomb, W. N., and Karplus, M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9032–9037

38. Sogo, S. G., Widlanski, T. S., Hoare, J. H., Grimshaw, C. E., Berchtold, G. A., and Knowles, J. R. (1984) *J. Am. Chem. Soc.* **106**, 2701–2703

38a. Pohnert, G., Zhang, S., Husain, A., Wilson, D. B., and Ganem, B. (1999) *Biochemistry* **38**, 12212–12217

39. Christendat, D., and Turnbull, J. L. (1999) *Biochemistry* **38**, 4782–4793
40. Christendat, D., and Turnbull, J. (1996) *Biochemistry* **35**, 4468–4479
41. Turnbull, J., Cleland, W. W., and Morrison, J. F. (1990) *Biochemistry* **29**, 10245–10254
42. Turnbull, J., Morrison, J. F., and Cleland, W. W. (1991) *Biochemistry* **30**, 7783–7788
43. Zamir, L. O., Jensen, R. A., Arison, B. H., Douglas, A. W., Albers-Schönberg, G., and Bowen, J. R. (1980) *J. Am. Chem. Soc.* **102**, 4499–4504
44. Zamir, L. O., Tiberio, R., Fiske, M., Berry, A., and Jensen, R. A. (1985) *Biochemistry* **24**, 1607–1612
45. Xia, T., and Jensen, R. A. (1990) *J. Biol. Chem.* **265**, 20033–20036
46. Jung, E., Zamir, L. O., and Jensen, R. A. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7231–7235
47. Jensen, R. A. (1986) in *Recent Advances in Phytochemistry*, Vol. 20 (Conn, E. E., ed), pp. 57–81, Plenum, New York
48. Teng, C.-Y. P., and Ganem, B. (1984) *J. Am. Chem. Soc.* **106**, 2463–2464
49. Morollo, A. A., Finn, M. G., and Bauerle, R. (1993) *J. Am. Chem. Soc.* **115**, 816–817
50. Green, J. M., and Nichols, B. P. (1991) *J. Biol. Chem.* **266**, 12971–12975
51. Paluh, J. L., Zalkin, H., Betsch, D., and Weith, H. L. (1985) *J. Biol. Chem.* **260**, 1889–1894
52. Walsh, C. T., Erion, M. D., Walts, A. E., Delany III, J. J., and Berchtold, G. A. (1987) *Biochemistry* **26**, 4734–4745
53. Roux, B., and Walsh, C. T. (1992) *Biochemistry* **31**, 6904–6910

53a. Knöchel, T., Ivens, A., Hester, G., Gonzalez, A., Bauerle, R., Wilmanns, M., Kirschner, K., and Jansonius, J. N. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9479–9484

54. Liu, D. J., and Day, L. A. (1994) *Science* **265**, 671–674
55. Liu, J., Quinn, N., Berchtold, G. A., and Walsh, C. T. (1990) *Biochemistry* **29**, 1417–1425
56. Kozlowski, M. C., and Bartlett, P. A. (1991) *J. Am. Chem. Soc.* **113**, 5897–5898

56a. Wildermuth, M. C., Dewdney, J., Wu, G., and Ausubel, F. M. (2001) *Nature (London)* **414**, 562–565

57. Sakaitani, M., Rusnak, F., Quinn, N. R., Tu, C., Frigo, T. B., Berchtold, G. A., and Walsh, C. T. (1990) *Biochemistry* **29**, 6789–6798
58. Rusnak, F., Liu, J., Quinn, N., Berchtold, G. A., and Walsh, C. T. (1990) *Biochemistry* **29**, 1425–1435
59. Gehring, A. M., Mori, I., and Walsh, C. T. (1998) *Biochemistry* **37**, 2648–2659

59a. May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001) *J. Biol. Chem.* **276**, 7209–7217

60. Schubert, K. R., Switzer, R. L., and Shelton, E. (1975) *J. Biol. Chem.* **250**, 7492–7500
61. Nosal, J. M., Switzer, R. L., and Becker, M. A. (1993) *J. Biol. Chem.* **268**, 10168–10175
62. Zalkin, H., Paluh, J. L., van Cleemput, M., Moye, W. S., and Yanofsky, C. (1984) *J. Biol. Chem.* **259**, 3985–3992
63. Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U., and Kirschner, K. (1995) *Biochemistry* **34**, 5419–5428
64. Knöchel, T. R., Hennig, M., Merz, A., Darimont, B., Kirschner, K., and Jansonius, J. N. (1996) *J. Mol. Biol.* **262**, 502–515
65. Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and Davies, D. R. (1988) *J. Biol. Chem.* **263**, 17857–17871
66. Schneider, T. R., Gerhardt, E., Lee, M., Liang, P.-H., Anderson, K. S., and Schlichting, I. (1998) *Biochemistry* **37**, 5394–5406
67. Rhee, S., Miles, E. W., and Davies, D. R. (1998) *J. Biol. Chem.* **273**, 8553–8555
68. Hyde, C. C., and Miles, E. W. (1990) *Bio/Technology* **8**, 27–32

68a. Miles, E. W. (2001) *Chem. Record* **1**, 140–151

68b. Weber-Ban, E., Hur, O., Bagwell, C., Banik, U., Yang, L.-H., Miles, E. W., and Dunn, M. F. (2001) *Biochemistry* **40**, 3497–3511

68c. Yamagata, Y., Ogasahara, K., Hioki, Y., Lee, S. J., Nakagawa, A., Nakamura, H., Ishida, M., Kuramitsu, S., and Yutani, K. (2001) *J. Biol. Chem.* **276**, 11062–11071

69. Sterner, R., Dahm, A., Darimont, B., Ivens, A., Liebl, W., and Kirschner, K. (1995) *EMBO J.* **14**, 4395–4402
70. Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K., and Jansonius, J. N. (1991) *Biochemistry* **30**, 9161–9169
71. Zhao, J., and Last, R. L. (1995) *J. Biol. Chem.* **270**, 6081–6087
72. Bentley, R., and Campbell, I. M. (1974) in *The Chemistry of the Quinonoid Compounds* (Patai, S., ed), pp. 683–736, Wiley, New York (Part 2)
73. Sippel, C. J., Goewert, R. R., Slachman, F. N., and Olson, R. E. (1983) *J. Biol. Chem.* **258**, 1057–1061

73a. Jonassen, T., and Clarke, C. F. (2000) *J. Biol. Chem.* **275**, 12381–12387

73b. Stenmark, P., Grünler, J., Mattsson, J., Sindelar, P. J., Nordlund, P., and Berthold, D. A. (2001) *J. Biol. Chem.* **276**, 33297–33300

References

73c. Miyadera, H., Amino, H., Hiraishi, A., Taka, H., Murayama, K., Miyoshi, H., Sakamoto, K., Ishii, N., Hekimi, S., and Kita, K. (2001) *J. Biol. Chem.* **276**, 7713–7716
73d. Hihi, A. K., Gao, Y., and Hekimi, S. (2002) *J. Biol. Chem.* **277**, 2202–2206
74. Marbois, B. N., and Clarke, C. F. (1996) *J. Biol. Chem.* **271**, 2995–3004
75. Hsu, A. Y., Poon, W. W., Shepherd, J. A., Myles, D. C., and Clarke, C. F. (1996) *Biochemistry* **35**, 9797–9806
76. Clarke, C. F., Williams, W., and Teruya, J. H. (1991) *J. Biol. Chem.* **266**, 16636 – 16644
77. Swiezewska, E., Dallner, G., Andersson, B., and Ernster, L. (1993) *J. Biol. Chem.* **268**, 1494–1499
78. Kalén, A., Appelkvist, E.-L., Chojnacki, T., and Dallner, G. (1990) *J. Biol. Chem.* **265**, 1158–1164
79. Folkers, K., Langsjoen, P., Willis, R., Richardson, P., Xia, L.-J., Ye, C.-Q., and Tamagawa, H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8931–8934
80. Threfall, D. R. (1971) *Vitamins and Hormones* **29**, 153–200
81. Goodwin, T. W., and Merces, E. I. (1986) *Introduction to Plant Biochemistry*, 2nd ed., Pergamon, Oxford
82. Fernández-Cañón, J. M., and Peñalva, M. A. (1995) *J. Biol. Chem.* **270**, 21199–21205
83. Shintani, D., and DellaPenna, D. (1998) *Science* **282**, 2098–2100
83a. Yamamoto, Y., Fujisawa, A., Hara, A., and Dunlap, W. C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13144–13148
84. Bentley, R., and Meganathan, R. (1987) in *Escherichia coli and Salmonella Typhimurium* (Nedihardt, F. C., ed), pp. 512–520, Am. Soc. for Microbiology, Washington, D.C.
85. Koike-Takeshita, A., Koyama, T., and Ogura, K. (1997) *J. Biol. Chem.* **272**, 12380–12383
86. Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. (1999) *Biochemistry* **38**, 4252–4258
86a. Stamps, S. L., Taylor, A. B., Wang, S. C., Hackert, M. L., and Whitman, C. P. (2000) *Biochemistry* **39**, 9671–9678
87. Daubner, S. C., and Fitzpatrick, P. F. (1999) *Biochemistry* **38**, 4448–4454
87a. Andersen, O. A., Flatmark, T., and Hough, E. (2001) *J. Mol. Biol.* **314**, 279 – 291
88. Nyhan, W. L. (1984) *Trends Biochem. Sci.* **9**, 71–72
89. Knox, W. E. (1972) in *The Metabolic Basis of Inherited Disease*, 3rd ed. (Standbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S., eds), pp. 266–295, McGraw-Hill, New York
90. Woo, S. L. C. (1989) *Biochemistry* **28**, 1–7
91. Kaufman, S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3160–3164
91a. Levy, H. L. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1811–1813
91b. Gámez, A., Pérez, B., Ugarte, M., and Desviat, L. R. (2000) *J. Biol. Chem.* **275**, 29737–29742
92. Ambrus, C. M., Ambrus, J. L., Horvath, C., Pederson, H., Skarma, S., Kant, C., Mirand, E., Guthrie, R., and Paul, T. (1978) *Science* **201**, 837–839
93. Scriver, C. R., Kaufman, S., Eisensmith, R. C., and Woo, S. L. C. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1015–1075, McGraw-Hill, New York
94. Onuffer, J. J., Ton, B. T., Klement, I., and Kirsch, J. F. (1995) *Protein Sci.* **4**, 1743–1749
95. Natt, E., Kida, K., Odievre, M., Di Rocco, M., and Scherer, G. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9297–9301
96. Hargrove, J. L., and Grammer, D. K. (1985) in *Transaminases* (Christen, P., and Metzler, D. E., eds), pp. 511–525, Wiley, New York
97. Wilcken, B., Hammond, J. W., Howard, N., Bohane, T., Hocart, C., and Halpern, B. (1981) *N. Engl. J. Med.* **305**, 865–869
98. La Du, B. N. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1371–1386, McGraw-Hill, New York
99. Fernández-Cañón, J. M., and Peñalva, M. A. (1998) *J. Biol. Chem.* **273**, 329–337
99a. Polekhina, G., Board, P. G., Blackburn, A. C., and Parker, M. W. (2001) *Biochemistry* **40**, 1567–1576
100. Kubo, S., Sun, M., Miyahara, M., Umeyama, K., Urakami, K.-i, Yamamoto, T., Jakobs, C., Matsuda, I., and Endo, F. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9552–9557
101. Frieden, E. (1981) *Trends Biochem. Sci.* **6**, 50–53
102. Tata, J. R. (1990) *Trends Biochem. Sci.* **15**, 282–284
103. Levy, O., De la Vieja, A., Ginter, C. S., Riedel, C., Dai, G., and Carrasco, N. (1998) *J. Biol. Chem.* **273**, 22657–22663
104. Sun, W., and Dunford, H. B. (1993) *Biochemistry* **32**, 1324–1331
105. Surks, M. I., and Sievert, R. (1995) *N. Engl. J. Med.* **333**, 1688–1693
106. Rawtich, A. B., Chernoff, S. B., Litwer, M. R., Rouse, J. B., and Hamilton, J. W. (1983) *J. Biol. Chem.* **258**, 2079–2082
107. Gentile, F., Ferranti, P., Mamone, G., Malorni, A., and Salvatore, G. (1997) *J. Biol. Chem.* **272**, 639–646
108. Gavaret, J.-M., Cahnmann, H. J., and Nunez, J. (1981) *J. Biol. Chem.* **256**, 9167–9173
108a. Ma, Y.-A., Sih, C. J., and Harms, A. (1999) *J. Am. Chem. Soc.* **121**, 8967–8968
109. Shupnik, M. A., Chin, W. W., Ross, D. S., Downing, M. F., Habener, J. F., and Ridgway, E. C. (1983) *J. Biol. Chem.* **258**, 15120–15124
110. Van Heuverswyn, B., Streydio, C., Brocas, H., Refetoff, S., Dumont, J., and Vassart, G. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5941–5945
111. Perlman, J. H., Colson, A.-O., Wang, W., Bence, K., Osman, R., and Gershengorn, M. C. (1997) *J. Biol. Chem.* **272**, 11937–11942
112. Kosugi, S., Okajima, F., Ban, T., Hidaka, A., Shenker, A., and Kohn, L. D. (1992) *J. Biol. Chem.* **267**, 24153–24156
113. Nettleton, E. J., Sunde, M., Lai, Z., Kelly, J. W., Dobson, C. M., and Robinson, C. V. (1998) *J. Mol. Biol.* **281**, 553–564
114. Sebastiao, M. P., Saraiva, M. J., and Damas, A. M. (1998) *J. Biol. Chem.* **273**, 24715–24722
115. Silva, J. E., and Larsen, P. R. (1983) *Nature (London)* **305**, 712–713
116. Pallud, S., Lennon, A.-M., Ramauge, M., Gavaret, J.-M., Croteau, W., Pierre, M., Courtin, F., and St. Germain, D. L. (1997) *J. Biol. Chem.* **272**, 18104–18110
116a. Leonard, J. L., Visser, T. J., and Leonard, D. M. (2001) *J. Biol. Chem.* **276**, 2600–2607
117. Maley, G. F., and Lardy, H. A. (1955) *J. Biol. Chem.* **215**, 377–388
118. Shi, Y.-B., and Brown, D. D. (1993) *J. Biol. Chem.* **268**, 20312–20317
119. Brown, D. D. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13011–13016
120. Kumara-Siri, M. H., Shapiro, L. E., and Surks, M. I. (1986) *J. Biol. Chem.* **261**, 2844–2852
121. Apriletti, J. W., David-Inouye, Y., Eberhardt, N. L., and Baxter, J. D. (1984) *J. Biol. Chem.* **259**, 10941–10948
122. Brent, G. A. (1994) *N. Engl. J. Med.* **331**, 847–853
123. Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D., and Fletterick, R. J. (1995) *Nature (London)* **378**, 690–697
123a. Wu, Y., Xu, B., and Koenig, R. J. (2001) *J. Biol. Chem.* **276**, 3929–3936
123b. Zhang, Y., Yin, L., and Hillgartner, F. B. (2001) *J. Biol. Chem.* **276**, 974–983
124. Hörlein, A. J., Näär, A. M., Heinzel, T., Torchia, J., Gloss, B., Kurokawa, R., Ryan, A., Kamei, Y., Söderström, M., Glass, C. K., and Rosenfeld, M. G. (1995) *Nature (London)* **377**, 397–404
125. Chen, Y., Chen, P.-L., Chen, C.-F., Sharp, Z. D., and Lee, W.-H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4443–4448
125a. Zhu, X.-G., Hanover, J. A., Hager, G. L., and Cheng, S. (1998) *J. Biol. Chem.* **273**, 27058–27063
125b. Tagami, T., Park, Y., and Jameson, J. L. (1999) *J. Biol. Chem.* **274**, 22345–22353
125c. Riedel, C., Dohán, O., De la Vieja, A., Ginter, C. S., and Carrasco, N. (2001) *Trends Biochem. Sci.* **26**, 490–496
126. Shaw-White, J. R., Bruno, M. D., and Whitsett, J. A. (1999) *J. Biol. Chem.* **274**, 2658–2664
127. Vassart, G., Dumont, J. E., and Refetoff, S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 2883–2928, McGraw-Hill, New York
128. Collingwood, T. N., and and 19 other authors. (1998) *EMBO J.* **17**, 4760–4770
129. Hetzel, B. S. (1994) *N. Engl. J. Med.* **331**, 1770–1771
130. Hearn, M. T. W. (1980) *Trends Biochem. Sci.* **5**, 75–79
131. Utiger, R. D. (1991) *N. Engl. J. Med.* **325**, 278–279
131a. Bertoldi, M., Frigeri, P., Paci, M., and Voltattorni, C. B. (1999) *J. Biol. Chem.* **274**, 5514 – 5521
132. Potterf, S. B., Muller, J., Bernardini, I., Tietze, F., Kobayashi, T., Hearing, V. J., and Gahl, W. A. (1996) *J. Biol. Chem.* **271**, 4002–4008
133. Jiménez-Cervantes, C., Solano, F., Kobayashi, T., Urabe, K., Hearing, V. J., Lozano, J. A., and García-Borrón, J. C. (1994) *J. Biol. Chem.* **269**, 17993–18001
134. Leonard, L. J., Townsend, D., and King, R. A. (1988) *Biochemistry* **27**, 6156–6159
135. Sugimoto, H., Taniguchi, M., Nakagawa, A., Tanaka, I., Suzuki, M., and Nishihira, J. (1999) *Biochemistry* **38**, 3268–3279
136. Rodríguez-López, J. N., Tudela, J., Varón, R., García-Carmona, F., and García-Cánovas, F. (1992) *J. Biol. Chem.* **267**, 3801–3810
137. Thomson, R. H. (1974) *Angew. Chem. Int. Ed. Engl.* **13**, 305–312
138. Deibel, R. M. B., and Chedekel, M. R. (1984) *J. Am. Chem. Soc.* **106**, 5884–5888
139. Pawelek, J., Körner, A., Bergstrom, A., and Bologna, J. (1980) *Nature (London)* **286**, 617–619
140. Chen, J. M., Xu, S. L., Wawrzak, Z., Basarab, G. S., and Jordan, D. B. (1998) *Biochemistry* **37**, 17735–17744
140a. Liao, D.-I., Thompson, J. E., Fahnestock, S., Valent, B., and Jordan, D. B. (2001) *Biochemistry* **40**, 8696–8704
141. Saul, S. J., and Sugumaran, M. (1990) *J. Biol. Chem.* **265**, 16992–16999
141a. Hall, M., Scott, T., Sugumaran, M., Söderhäll, K., and Law, J. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7764–7768
141b. Asano, T., and Ashida, M. (2001) *J. Biol. Chem.* **276**, 11100–11112
142. Sugumaran, M., Semensi, V., Kalyanaraman, B., Bruce, J. M., and Land, E. J. (1992) *J. Biol. Chem.* **267**, 10355–10361

References

143. McCormick, M. L., Gaut, J. P., Lin, T.-S., Britigan, B. E., Buettner, G. R., and Heinecke, J. W. (1998) *J. Biol. Chem.* **273**, 32030–32037
144. Malencik, D. A., and Anderson, S. R. (1994) *Biochemistry* **33**, 13363–13372
144a. Kanwar, R., and Balasubramanian, D. (2000) *Biochemistry* **39**, 14976–14983
145. Briza, P., Winkler, G., Kalchhauser, H., and Breitenbach, M. (1986) *J. Biol. Chem.* **261**, 4288–4294
146. Briza, P., Ellinger, A., Winkler, G., and Breitenbach, M. (1990) *J. Biol. Chem.* **265**, 15118–15123
147. Taylor, S. W., Waite, J. H., Ross, M. M., Shabanowitz, J., and Hunt, D. F. (1994) *J. Am. Chem. Soc.* **116**, 10803–10804
148. Waite, J. H., Jensen, R. A., and Morse, D. E. (1992) *Biochemistry* **31**, 5733–5738
149. Papov, V. V., Diamond, T. V., Biemann, K., and Waite, J. H. (1995) *J. Biol. Chem.* **270**, 20183–20192
149a. Yu, M., Hwang, J., and Deming, T. (1999) *J. Am. Chem. Soc.* **121**, 5825–5826
150. Dagley, S. (1978) *Naturwissenschaften* **65**, 85–95
151. Allewell, N. (1989) *Trends Biochem. Sci.* **14**, 473–474
152. Dagley, S., and Nicholson, D. E. (1970) *An Introduction to Metabolic Pathways*, Wiley, New York
153. Gibson, D. T., ed. (1984) *Microbial Degradation of Organic Compounds*, Dekker, New York
154. Nakai, C., Kagamiyama, H., Nozaki, M., Nakazawa, T., Inouye, S., Ebina, Y., and Nakazawa, A. (1983) *J. Biol. Chem.* **258**, 2923–2928
155. Correll, C. C., Batie, C. J., Ballou, D. P., and Ludwig, M. L. (1985) *J. Biol. Chem.* **260**, 14633–14635
155a. Palumbo, A., Poli, A., Cosmo, A. D., and d'Ischia, M. (2000) *J. Biol. Chem.* **275**, 16885–16890
156. Blasco, R., Wittich, R.-M., Mallavarapu, M., Timmis, K. N., and Pieper, D. H. (1995) *J. Biol. Chem.* **270**, 29229–29235
157. Doelle, H. W. (1975) *Bacterial Metabolism*, Academic Press, New York
158. Werlen, C., Kohler, H.-P. E., and van der Meer, J. R. (1996) *J. Biol. Chem.* **271**, 4009–4016
158a. Copley, S. D. (2000) *Trends Biochem. Sci.* **25**, 261–265
158b. Anandarajah, K., Kiefer, P. M., Donohoe, B. S., and Copley, S. D. (2000) *Biochemistry* **39**, 5303–5311
158c. Tame, J. R. H., Namba, K., Dodson, E. J., and Roper, D. I. (2002) *Biochemistry* **41**, 2982–2989
159. Seah, S. Y. K., Terracina, G., Bolin, J. T., Riebel, P., Snieckus, V., and Eltis, L. D. (1998) *J. Biol. Chem.* **273**, 22943–22949
160. Scholten, J. D., Chang, K.-H., Babbitt, P. C., Charest, H., Sylvestre, M., and Dunaway-Mariano, D. (1991) *Science* **253**, 182–185
160a. Zhang, W., Wei, Y., Luo, L., Taylor, K. L., Yang, G., and Dunaway-Mariano, D. (2001) *Biochemistry* **40**, 13474–13482
161. Harpel, M. R., and Lipscomb, J. D. (1990) *J. Biol. Chem.* **265**, 22187–22196
161a. Olivera, E. R., Minambres, B., Garcia, B., Muniz, C., Moreno, M. A., Ferrández, A., Díaz, E., García, J. L., and Luengo, J. M. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6419–6424
161b. Zaar, A., Eisenreich, W., Bacher, A., and Fuchs, G. (2001) *J. Biol. Chem.* **276**, 24997–25004
162. Yamaguchi, M., and Fujisawa, H. (1982) *J. Biol. Chem.* **257**, 12497–12502
163. Zylstra, G. J., and Gibson, D. T. (1989) *J. Biol. Chem.* **264**, 14940–14946
164. Díaz, E., and Timmis, K. N. (1995) *J. Biol. Chem.* **270**, 6403–6411
165. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. (1993) *Trends Biochem. Sci.* **18**, 372–376
166. Harwood, C. S., and Gibson, J. (1997) *J. Bacteriol.* **179**, 301–309
166a. Boll, M., Laempe, D., Eisenreich, W., Bacher, A., Mittelberger, T., Heinze, J., and Fuchs, G. (2000) *J. Biol. Chem.* **275**, 21889–21895
166b. Boll, M., Fuchs, G., and Lowe, D. J. (2001) *Biochemistry* **40**, 7612–7620
167. Härtel, U., Eckel, E., Koch, J., Fuchs, G., Linder, D., and Buckel, W. (1993) *Arch Microbiol* **159**, 174–181
168. Draths, K. M., Ward, T. L., and Frost, J. W. (1992) *J. Am. Chem. Soc.* **114**, 9725–9726
169. Richman, J. E., Chang, Y.-C., Kambourakis, S., Draths, K. M., Almy, E., Snell, K. D., Strasburg, G. M., and Frost, J. W. (1996) *J. Am. Chem. Soc.* **118**, 11587–11591
170. Draths, K. M., and Frost, J. W. (1991) *J. Am. Chem. Soc.* **113**, 9361–9363
171. Haslam, E. (1986) in *Recent Advances in Phytochemistry*, Vol. 20 (Conn, E. E., ed), pp. 163–200, Plenum, New York
172. Werner, I., Bacher, A., and Eisenreich, W. (1997) *J. Biol. Chem.* **272**, 25474–25482
173. Lauvergeat, V., Kennedy, K., Feuillet, C., McKie, J. H., Gorrichon, L., Baltas, M., Boudet, A. M., Grima-Pettenati, J., and Douglas, K. T. (1995) *Biochemistry* **34**, 12426–12434
174. Ellis, B. E., Kuroki, G. W., and Stafford, H. A., eds. (1994) *Genetic Engineering of Plant Secondary Metabolism*, Plenum, New York
174a. Crouteau, R., Kutchan, T. M., and Lewis, N. G. (2000) in *Biochemistry and Molecular Biology of Plants* (Buchanan, B., Gruissem, W., and Jones, R., eds), pp. 1250–1318, Am. Soc. Plant Physiologists, Rockville, Maryland
175. Schwede, T. F., Rétey, J., and Schulz, G. E. (1999) *Biochemistry* **38**, 5355–5361
175a. Stuible, H.-P., and Kombrink, E. (2001) *J. Biol. Chem.* **276**, 26893–26897
176. Becker-André, M., Schulze-Lefert, P., and Hahlbrock, K. (1991) *J. Biol. Chem.* **266**, 8551–8559
177. León, J., Shulaev, V., Yalpani, N., Lawton, M. A., and Raskin, I. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10413–10417
178. Lee, H.-i, León, J., and Raskin, I. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4076–4079
179. Delaney, T. P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., Gaffney, T., Gut-Rella, M., Kessmann, H., Ward, E., and Ryals, J. (1994) *Science* **266**, 1247–1250
180. Chen, Z., Malamy, J., Henning, J., Conrath, U., Sánchez-Casas, P., Silva, H., Ricigliano, J., and Klessig, D. F. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4134–4137
181. Durner, J., and Klessig, D. F. (1996) *J. Biol. Chem.* **271**, 28492–28501
182. Durner, J., and Klessig, D. F. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11312–11316
183. Jurivich, D. A., Pachetti, C., Qiu, L., and Welk, J. F. (1995) *J. Biol. Chem.* **270**, 24489–24495
184. Prince, R. C., and Gunson, D. E. (1994) *Trends Biochem. Sci.* **19**, 521
185. Li, K., and Frost, J. W. (1998) *J. Am. Chem. Soc.* **120**, 10545–10546
186. Rosen, M. A., Shapiro, L., and Patel, D. J. (1992) *Biochemistry* **31**, 4015–4026
187. Villegas, R. J. A., and Kojima, M. (1986) *J. Biol. Chem.* **261**, 8729–8733
188. von Röpenack, E., Parr, A., and Schulze-Lefert, P. (1998) *J. Biol. Chem.* **273**, 9013–9022
189. Balint, R., Cooper, G., Staebell, M., and Filner, P. (1987) *J. Biol. Chem.* **262**, 11026–11031
190. Crawford, R. L. (1981) *Lignin Biodegradation and Transformation*, Wiley, New York
190a. Humphreys, J. M., Hemm, M. R., and Chapple, C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10045–10050
190b. Li, L., Popko, J. L., Umezawa, T., and Chiang, V. L. (2000) *J. Biol. Chem.* **275**, 6537–6545
190c. Lim, E.-K., Li, Y., Parr, A., Jackson, R., Ashford, D. A., and Bowles, D. J. (2001) *J. Biol. Chem.* **276**, 4344–4349
191. Fujita, M., Gang, D. R., Davin, L. B., and Lewis, N. G. (1999) *J. Biol. Chem.* **274**, 618–627
191a. Xia, Z.-Q., Costa, M. A., Pélissier, H. C., Davin, L. B., and Lewis, N. G. (2001) *J. Biol. Chem.* **276**, 12614–12623
192. Lewis, N. G., and Davin, L. B. (1994) in *Isopentenoids and Other Natural Products* (Nes, W. D., ed), pp. 202–246, American Chemical Society, Washington D.C.
193. von Wettstein-Knowles, P. M. (1993) in *Lipid Metabolism in Plants* (Moore, T. S., Jr., ed), pp. 127–166, CRC Press, Boca Raton, Florida
194. Nimz, H. (1974) *Angew. Chem. Int. Ed. Engl.* **13**, 313–321
195. Ralph, J., Hatfield, R. D., Quideau, S., Helm, R. F., Grabber, J. H., and Jung, H.-J. G. (1994) *J. Am. Chem. Soc.* **116**, 9448–9456
196. Steelink, C. (1972) *Rec. Adv. Phytochem.* **4**, 239–271
197. Huang, K.-x, Fujii, I., Ebizuka, Y., Gomi, K., and Sankawa, U. (1995) *J. Biol. Chem.* **270**, 21495–21502
197a. Henriksen, A., Smith, A. T., and Gajhede, M. (1999) *J. Biol. Chem.* **274**, 35005–35011
198. Nakamura, N., Matsuzaki, R., Choi, Y.-H., Tanizawa, K., and Sanders-Loehr, J. (1996) *J. Biol. Chem.* **271**, 4718–4724
198a. Schwartz, B., Olgin, A. K., and Klinman, J. P. (2001) *Biochemistry* **40**, 2954–2963
198b. Klinman, J. P. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14766–14768
198c. Datta, S., Mori, Y., Takagi, K., Kawaguchi, K., Chen, Z.-W., Okajima, T., Kuroda, S., Ikeda, T., Kano, K., Tanizawa, K., and Mathews, F. S. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14268–14273
198d. Firbank, S. J., Rogers, M. S., Wilmot, C. M., Dooley, D. M., Halcrow, M. A., Knowles, P. F., McPherson, M. J., and Phillips, S. E. V. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12932–12937
198e. Xie, L., and van der Donk, W. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12863–12865
199. Roberts, M. F., and Wink, M., eds. (1998) *Alkaloids Biochemistry, Ecology, and Medicinal Applications*, Plenum, New York
200. Harborne, J. B. (1993) *Introduction to Ecological Biochemistry*, 4th ed., Academic Press, San Diego, California
201. Nakajima, K., Yamashita, A., Akama, H., Nakatsu, T., Kato, H., Hashimoto, T., Oda, J., and Yamada, Y. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4876–4881
202. Robinson, R. (1917) *J. Chem. Soc.* **111**, 876–899
203. Robinson, R. (1955) *The Structural Relations of Natural Products*, Oxford Univ. Press, London
204. Lenz, R., and Zenk, M. H. (1995) *J. Biol. Chem.* **270**, 31091–31096
204a. Grothe, T., Lenz, R., and Kutchan, T. M. (2001) *J. Biol. Chem.* **276**, 30717–30723
205. Kutchan, T. M., and Dittrich, H. (1995) *J. Biol. Chem.* **270**, 24475–24481
206. Hayaishi, O. (1993) *Protein Sci.* **2**, 472–475
207. Young, S. N., Oravec, M., and Sourkes, T. L. (1974) *J. Biol. Chem.* **249**, 3932–3936
208. Sono, M. (1989) *Biochemistry* **28**, 5400–5407
209. Taylor, M. W., and Feng, G. (1991) *FASEB J.* **5**, 2516–2522

References

209a. Phillips, R. S., Johnson, N., and Kamath, A. V. (2002) *Biochemistry* **41**, 4012–4019
210. Billker, O., Lindo, V., Panico, M., Etienne, A. E., Paxton, T., Dell, A., Rogers, M., Sinden, R. E., and Morris, H. R. (1998) *Nature (London)* **392**, 289–292
211. Garcia, G. E., Wirtz, R. A., Barr, J. R., Woolfitt, A., and Rosenberg, R. (1998) *J. Biol. Chem.* **273**, 12003–12005
212. Kayser, H. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, Vol. 10 (Kerkut, G. A., and Gilbert, L. I., eds), pp. 368–416, Pergamon, Oxford
213. Stone, T. W., ed. (1988) *Quinolinic Acid and the Kynurenines*, CRC Press, Boca Raton, Florida
213a. Aquilina, J. A., Carver, J. A., and Truscott, R. J. W. (1999) *Biochemistry* **38**, 11455–11464
213b. Wang, L., Erlandsen, H., Haavik, J., Knappskog, P. M., and Stevens, R. C. (2002) *Biochemistry* **41**, 12569 – 12574
214. Manthey, M. K., Pyne, S. G., and Truscott, R. J. W. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1954–1957
215. Schmidt, A. W., and Peroutka, S. J. (1989) *FASEB J.* **3**, 2242–2249
215a. McKinney, J., Teigen, K., Froystein, N. Å., Salaün, C., Knappskog, P. M., Haavik, J., and Martínez, A. (2001) *Biochemistry* **40**, 15591–15601
216. Hajduch, E., Rencurel, F., Balendran, A., Batty, I. H., Downes, C. P., and Hundal, H. S. (1999) *J. Biol. Chem.* **274**, 13563–13568
217. Klein, D. C., and Namboodiri, M. A. A. (1982) *Trends Biochem. Sci.* **7**, 98–102
218. Utiger, R. D. (1992) *N. Engl. J. Med.* **327**, 1377–1379
219. Ishida, I., Obinata, M., and Deguchi, T. (1987) *J. Biol. Chem.* **262**, 2895–2899
220. Rodriguez, I. R., Mazuruk, K., Schoen, T. J., and Chader, G. J. (1994) *J. Biol. Chem.* **269**, 31969–31977
221. Slominski, A., Baker, J., Rosano, T. G., Guisti, L. W., Ermak, G., Grande, M., and Gaudet, S. J. (1996) *J. Biol. Chem.* **271**, 12281–12286
222. Mayeno, A. N., Lin, F., Foote, C. S., Loegering, D. A., Ames, M. M., Hedberg, C. W., and Gleich, G. J. (1990) *Science* **250**, 1707–1708
223. Aldhous, P. (1991) *Nature (London)* **353**, 490
224. Moat, A. G., and Foster, J. W. (1987) in *Pyridine Nucleotide Coenzymes: Chemical, Biochemical and Medical Aspects*, Vol. B (Dolphin, D., Avramovic, O., and Poulson, R., eds), pp. 1–24, Wiley (Interscience), New York
225. Keys, L. D., III, and Hamilton, G. A. (1987) *J. Am. Chem. Soc.* **109**, 2156–2163
225a. Saridakis, V., Christendat, D., Kimber, M. S., Dharamsi, A., Edwards, A. M., and Pai, E. F. (2001) *J. Biol. Chem.* **276**, 7225–7232
226. Rizzi, M., Nessi, C., Mattevi, A., Coda, A., Bolognesi, M., and Galizzi, A. (1996) *EMBO J.* **15**, 5125–5134
227. Merkler, D. J., Glufke, U., Ritenour-Rodgers, K. J., Baumgart, L. E., DeBlassio, J. L., Merkler, K. A., and Vederas, J. C. (1999) *J. Am. Chem. Soc.* **121**, 4904–4905
228. Tedeschi, G., Zetta, L., Negri, A., Mortarino, M., Ceciliani, F., and Ronchi, S. (1997) *Biochemistry* **36**, 16221–16230
229. Nasu, S., Wicks, F. D., and Gholson, R. K. (1982) *J. Biol. Chem.* **257**, 626–632
230. White, H. B., III. (1982) in *The Pyridine Nucleotide Coenzymes* (Everse, J., Anderson, B., and You, K.-S., eds), pp. 225–242, Academic Press, New York
231. Tritz, G. J. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 557–563, Am. Soc. for Microbiology, Washington,DC
232. Malherbe, P., Köhler, C., Da Prada, M., Lang, G., Kiefer, V., Schwarcz, R., Lahm, H.-W., and Cesura, A. M. (1994) *J. Biol. Chem.* **269**, 13792–13797
233. Heyes, M. P., Saito, K., Lackner, A., Wiley, C. A., Achim, C. L., and Markey, S. P. (1998) *FASEB J.* **12**, 881–896
234. Emanuele, J. J., and Pitzpatrick, P. F. (1995) *Biochemistry* **34**, 3716–3723
234a. Gadda, G., Dangott, L. J., Johnson, W. H., Jr., Whitman, C. P., and Fitzpatrick, P. F. (1999) *Biochemistry* **38**, 5822–5828
234b. Zhao, Y., Christensen, S. K., Fankhauser, C., Cashman, J. R., Cohen, J. D., Weigel, D., and Chory, J. (2001) *Science* **291**, 306–309
235. Brown, H. M., and Purves, W. K. (1976) *J. Biol. Chem.* **251**, 907–913
236. Normanly, J., Cohen, J. D., and Fink, G. R. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10355–10359
236a. Mikkelsen, M. D., Hansen, C. H., Wittstock, U., and Halkier, B. A. (2000) *J. Biol. Chem.* **275**, 33712–33717
236b. Kato, Y., Nakamura, K., Sakiyama, H., Mayhew, S. G., and Asano, Y. (2000) *Biochemistry* **39**, 800–809
237. Kobayashi, M., Suzuki, T., Fujita, T., Masuda, M., and Shimizu, S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 714–718
238. Bartling, D., Seedorf, M., Schmidt, R. C., and Weiler, E. W. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6021–6025
238a. Kendrew, S. G. (2001) *Trends Biochem. Sci.* **26**, 218
238b. Jones, A. M. (1998) *Science* **282**, 2201 – 2202
239. Bartel, B., and Fink, G. R. (1995) *Science* **268**, 1745–1748
240. Szerszen, J. B., Szczyglowski, K., and Bandurski, R. S. (1994) *Science* **265**, 1699–1701
241. Roberto, F. F., Klee, H., White, F., Nordeen, R., and Kosuge, T. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5797–5801
241a. Crozier, A., Kamiya, Y., Bishop, G., and Yokota, T. (2000) in *Biochemistry and Molecular Biology of Plants* (Buchanan, B., Gruissem, W., and Jones, R., eds), pp. 850–929, Am. Soc. Plant Physiologists, Rockville, Maryland
242. Abel, S., Nguyen, M. D., and Theologis, A. (1995) *J. Mol. Biol.* **251**, 533–549
243. Abel, S., Nguyen, M. D., Chow, W., and Theologis, A. (1995) *J. Biol. Chem.* **270**, 19093–19099
244. Shibuya, M., Chou, H.-M., Fountoulakis, M., Hassam, S., Kim, S.-U., Kobayashi, K., Otsuka, H., Rogalska, E., Cassady, J. M., and Floss, H. G. (1990) *J. Am. Chem. Soc.* **112**, 297–304
245. Mizukami, H., Nordlöv, H., Lee, S.-L., and Scott, A. I. (1979) *Biochemistry* **18**, 3760–3763
245a. Stocking, E. M., Williams, R. M., and Sanz-Cervera, J. F. (2000) *J. Am. Chem. Soc.* **122**, 9089–9098
246. Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grün, S., Winklmair, A., Eisenreich, W., Bacher, A., Meeley, R. B., Briggs, S. P., Simcox, K., and Gierl, A. (1997) *Science* **277**, 696–699
246a. Beismann-Driemeyer, S., and Sterner, R. (2001) *J. Biol. Chem.* **276**, 20387–20396
247. Parker, A. R., Moore, J. A., Schwab, J. M., and Davisson, V. J. (1995) *J. Am. Chem. Soc.* **117**, 10605–10613
247a. Sivaraman, J., Li, Y., Larocque, R., Schrag, J. D., Cygler, M., and Mattte, A. (2001) *J. Mol. Biol.* **311**, 761–776
247b. Haruyama, K., Nakai, T., Miyahara, I., Hirotsu, K., Mizuguchi, H., Hayashi, H., and Kagamiyama, H. (2001) *Biochemistry* **40**, 4633–4644
248. Reizer, J., Michotey, V., Reizer, A., and Saier, M. H., Jr. (1994) *Protein Sci.* **3**, 440–450
249. Nagai, A., Ward, E., Beck, J., Tada, S., Chang, J.-Y., Scheidegger, A., and Ryals, J. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4133–4137
249a. Grubmeyer, C., and Teng, H. (1999) *Biochemistry* **38**, 7355–7362
250. Brenner, M., and Ames, B. N. (1971) in *Metabolic Pathways*, 3rd ed. (Vogel, H. J., ed), Academic Press, New York
251. Winkler, M. E. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 395–411, Am. Soc. for Microbiology, Washington,DC
252. Levy, H. L., Taylor, R. G., and McInnes, R. R. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1107–1123, McGraw-Hill, New York
253. Chen, J.-Y. C., and Bodley, J. W. (1988) *J. Biol. Chem.* **263**, 11692–11696
254. Jones, M. E. (1980) *Ann. Rev. Biochem.* **49**, 253–279
255. Gao, G., Nara, T., Nakajima-Shimada, J., and Aoki, T. (1999) *J. Mol. Biol.* **285**, 149–161
256. Serre, V., Guy, H., Liu, X., Penverne, B., Hervé, G., and Evans, D. (1998) *J. Mol. Biol.* **281**, 363–377
257. Neuhard, J., and Nygaard, P. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 446–473, Am. Soc. for Microbiology, Washington,DC
258. Switzer, R. L., and Quinn, C. L. (1993) in *Bacillus subtilis and Other Gram-Positive Bacteria* (Sonenshein, A. L., Hoch, J. A., and Losick, R., eds), pp. 343–358, American Society for Microbiology, Washington, D. C.
259. Guy, H. I., and Evans, D. R. (1996) *J. Biol. Chem.* **271**, 13762–13769
259a. Thoden, J. B., Raushel, F. M., Wesenberg, G., and Holden, H. M. (1999) *J. Biol. Chem.* **274**, 22502–22507
260. Braxton, B. L., Mullins, L. S., Rauschel, F. M., and Reinhart, G. D. (1999) *Biochemistry* **38**, 1394–1401
260a. Thoden, J. B., Phillips, G. N., Jr., Neal, T. M., Raushel, F. M., and Holden, H. M. (2001) *Biochemistry* **40**, 6989–6997
261. Williams, N. K., Manthey, M. K., Hambley, T. W., O'Donoghue, S. I., Keegan, M., Chapman, B. E., and Christopherson, R. I. (1995) *Biochemistry* **34**, 11344–11352
262. Zimmermann, B. H., Kemling, N. M., and Evans, D. R. (1995) *Biochemistry* **34**, 7038–7046
263. Hines, V., and Johnston, M. (1989) *Biochemistry* **28**, 1227–1234
264. Björnberg, O., Rowland, P., Larsen, S., and Jensen, K. F. (1997) *Biochemistry* **36**, 16197–16205
265. Rowland, P., Björnberg, O., Nielsen, F. S., Jensen, K. F., and Larsen, S. (1998) *Protein Sci.* **7**, 1269–1279
265a. Argyrou, A., Washabaugh, M. W., and Pickart, C. M. (2000) *Biochemistry* **39**, 10373–10384
266. Scapin, G., Ozturk, D. H., Grubmeyer, C., and Sacchettini, J. C. (1995) *Biochemistry* **34**, 10744–10754
267. Tao, W., Grubmeyer, C., and Blanchard, J. S. (1996) *Biochemistry* **35**, 14–21
268. Radzicka, A., and Wolfenden, R. (1995) *Science* **267**, 90–93
269. Beak, P., and Siegel, B. (1976) *J. Am. Chem. Soc.* **98**, 3601–3606
270. Smiley, J. A., and Jones, M. E. (1992) *Biochemistry* **31**, 12162–12168
270a. Appleby, T. C., Kinsland, C., Begley, T. P., and Ealick, S. E. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2005–2010

References

270b. Harris, P., Poulsen, J.-C. N., Jensen, K. F., and Larsen, S. (2000) *Biochemistry* **39**, 4217–4224
270c. Miller, B. G., Snider, M. J., Short, S. A., and Wolfenden, R. (2000) *Biochemistry* **39**, 8113–8118
270d. Phillips, L. M., and Lee, J. K. (2001) *J. Am. Chem. Soc.* **123**, 12067–12073
271. Girot, R., Hamet, M., Perignon, J.-L., Guesnu, M., Fox, R. M., Cartier, P., Durandy, A., and Griscelli, C. (1983) *N. Engl. J. Med.* **308**, 700–704
272. Lewis, D. A., and Villafrance, J. J. (1989) *Biochemistry* **28**, 8454–8459
273. Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York
274. Weiner, K. X. B., Ciesla, J., Jaffe, A. B., Ketring, R., Maley, F., and Maley, G. F. (1995) *J. Biol. Chem.* **270**, 18727–18729
274a. Kambampati, R., and Lauhon, C. T. (2000) *J. Biol. Chem.* **275**, 10727–10730
274b. Johansson, E., Mejlhede, N., Neuhard, J., and Larsen, S. (2002) *Biochemistry* **41**, 2563–2570
274c. Snider, M. J., Reinhardt, L., Wolfenden, R., and Cleland, W. W. (2002) *Biochemistry* **41**, 415–421
275. Kvalnes-Krick, K. L., and Traut, T. W. (1993) *J. Biol. Chem.* **268**, 5686–5693
275a. Dobritzsch, D., Schneider, G., Schnackerz, K. D., and Lindqvist, Y. (2001) *EMBO J.* **20**, 650–660
276. Kim, Y. S., and Bang, S. K. (1985) *J. Biol. Chem.* **260**, 5098–5104
277. Nygaard, P. (1993) in *Bacillus subtilis and Other Gram-Positive Bacteria* (Sonenshein, A. L., Hoch, J. A., and Losick, R., eds), pp. 359–378, American Society for Microbiology, Washington, D. C.
278. Iovannisci, D. M., Goebel, D., Allen, K., Kaur, K., and Ullman, B. (1984) *J. Biol. Chem.* **259**, 14617–14623
279. Schumacher, M. A., Carter, D., Scott, D. M., Roos, D. S., Ullman, B., and Brennan, R. G. (1998) *EMBO J.* **17**, 3219–3232
280. Muchmore, C. R. A., Krahn, J. M., Kim, J. H., Zalkin, H., and Smith, J. L. (1998) *Protein Sci.* **7**, 39–51
281. Smith, J. L., Zaluzec, E. J., Wery, J.-P., Niu, L., Switzer, R. L., Zalkin, H., and Satow, Y. (1994) *Science* **264**, 1427–1433
282. Wang, W., Kappock, T. J., Stubbe, J., and Ealick, S. E. (1998) *Biochemistry* **37**, 15647–15662
283. Benkovic, S. J. (1984) *Trends Biochem. Sci.* **9**, 320–322
284. Ononogbu, I. C. (1980) *Trends Biochem. Sci.* **5**, X
285. Shim, J. H., and Benkovic, S. J. (1998) *Biochemistry* **37**, 8776–8782
285a. Shim, J. H., and Benkovic, S. J. (1999) *Biochemistry* **38**, 10024–10031
285b. Gooljarsingh, L. T., Ramcharan, J., Gilroy, S., and Benkovic, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6565–6570
286. Kastanos, E. K., Woldman, Y. Y., and Appling, D. R. (1997) *Biochemistry* **36**, 14956–14964
287. Schrimsher, J. L., Schendel, F. J., Stubbe, J., and Smith, J. M. (1986) *Biochemistry* **25**, 4366–4371
287a. Mueller, E. J., Oh, S., Kavalerchik, E., Kappock, T. J., Meyer, E., Li, C., Ealick, S. E., and Stubbe, J. (1999) *Biochemistry* **38**, 9831–9839
288. Schrimsher, J. L., Schendel, F. J., and Stubbe, J. (1986) *Biochemistry* **25**, 4356–4365
289. Firestine, S. M., and Davisson, V. J. (1994) *Biochemistry* **33**, 11917–11926
290. Meyer, E., Kappock, T. J., Osuji, C., and Stubbe, J. (1999) *Biochemistry* **38**, 3012–3018
290a. Thoden, J. B., Kappock, T. J., Stubbe, J., and Holden, H. M. (1999) *Biochemistry* **38**, 15480–15492
291. Szabados, E., Hindmarsh, E. J., Phillips, L., Duggleby, R. G., and Christopherson, R. I. (1994) *Biochemistry* **33**, 14237–14245
292. Rayl, E. A., Moroson, B. A., and Beardsley, G. P. (1996) *J. Biol. Chem.* **271**, 2225–2233
292a. Wall, M., Shim, J. H., and Benkovic, S. J. (2000) *Biochemistry* **39**, 11303–11311
292b. Vergis, J. M., Bulock, K. G., Fleming, K. G., and Beardsley, G. P. (2001) *J. Biol. Chem.* **276**, 7727–7733
292c. Brosius, J. L., and Colman, R. F. (2000) *Biochemistry* **39**, 13336–13343
292d. Toth, E. A., Worby, C., Dixon, J. E., Goedken, E. R., Marqusee, S., and Yeates, T. O. (2000) *J. Mol. Biol.* **301**, 433–450
293. Silva, M. M., Poland, B. W., Hoffman, C. R., Fromm, H. J., and Honzatko, R. B. (1995) *J. Mol. Biol.* **254**, 431–446
293a. Poland, B. W., Lee, S.-F., Subramanian, M. V., Siehl, D. L., Anderson, R. J., Fromm, H. J., and Honzatko, R. B. (1996) *Biochemistry* **35**, 15753–15759
294. Lee, T. T., Worby, C., Bao, Z.-Q., Dixon, J. E., and Colman, R. F. (1999) *Biochemistry* **38**, 22–32
294a. Guicherit, O. M., Cooper, B. F., Rudolph, F. B., and Kellems, R. E. (1994) *J. Biol. Chem.* **269**, 4488–4496
294b. Iancu, C. V., Borza, T., Choe, J. Y., Fromm, H. J., and Honzatko, R. B. (2001) *J. Biol. Chem.* **276**, 42146–42152
295. Bass, M. B., Fromm, H. J., and Rudolph, F. B. (1984) *J. Biol. Chem.* **259**, 12330–12333
296. Webb, M. R., Reed, G. H., Cooper, B. F., and Rudolph, F. B. (1984) *J. Biol. Chem.* **259**, 3044–3046
297. Choe, J.-Y., Poland, B. W., Fromm, H. J., and Honzatko, R. B. (1999) *Biochemistry* **38**, 6953–6961
297a. Fleming, M. A., Chambers, S. P., Connelly, P. R., Nimmesgern, E., Fox, T., Bruzzese, F. J., Hoe, S. T., Fulghum, J. R., Livingston, D. J., Stuver, C. M., Sintchak, M. D., Wilson, K. P., and Thomson, J. A. (1996) *Biochemistry* **35**, 6990–6997
297b. Markham, G. D., Bock, C. L., and Schalk-Hihi, C. (1999) *Biochemistry* **38**, 4433–4440
297c. Barnes, B. J., Eakin, A. E., Izydore, R. A., and Hall, I. H. (2000) *Biochemistry* **39**, 13641–13650
297d. Kerr, K. M., Digits, J. A., Kuperwasser, N., and Hedstrom, L. (2000) *Biochemistry* **39**, 9804–9810
298. Zhang, R.-g, Evans, G., Rotella, F. J., Westbrook, E. M., Beno, D., Huberman, E., Joachimiak, A., and Collart, F. R. (1999) *Biochemistry* **38**, 4691–4700
299. Whitby, F. G., Luecke, H., Kuhn, P., Somoza, J. R., Huete-Perez, J. A., Phillips, J. D., Hill, C. P., Fletterick, R. J., and Wang, C. C. (1997) *Biochemistry* **36**, 10666–10674
300. Farazi, T., Leichman, J., Harris, T., Cahoon, M., and Hedstrom, L. (1997) *J. Biol. Chem.* **272**, 961–965
301. Lowenstein, J., and Tornheim, K. (1971) *Science* **171**, 397–400
301a. Bai, Y., Fox, D. T., Lacy, J. A., Van Lanen, S. G., and Iwata-Reuyl, D. (2000) *J. Biol. Chem.* **275**, 28731–28738
301b. Watanabe, M., Nameki, N., Matsuo-Takasaki, M., Nishimura, S., and Okada, N. (2001) *J. Biol. Chem.* **276**, 2387–2394
301c. Tercero, J. A., Espinosa, J. C., Lacalle, R. A., and Jiménez, A. (1996) *J. Biol. Chem.* **271**, 1579–1590
302. Coffee, C. J., and Kofke, W. A. (1975) *J. Biol. Chem.* **250**, 6653–6658
303. Schultz, V., and Lowenstein, J. M. (1978) *J. Biol. Chem.* **253**, 1938–1943
304. Merkler, D. J., and Schramm, V. L. (1993) *Biochemistry* **32**, 5792–5799
305. Sollitti, P., Merkler, D. J., Estupiñán, B., and Schramm, V. L. (1993) *J. Biol. Chem.* **268**, 4549–4555
306. Fishbein, W. N., Armbrustmacher, V. W., and Griffin, J. L. (1978) *Science* **200**, 545–548
306a. Craig, S. P., III, and Eakin, A. E. (2000) *J. Biol. Chem.* **275**, 20231–20234
306b. Phillips, C. L., Ullman, B., Brennan, R. G., and Hill, C. P. (1999) *EMBO J.* **18**, 3533–3545
307. Stiles, G. L. (1992) *J. Biol. Chem.* **267**, 6451–6454
308. Newby, A. C. (1984) *Trends Biochem. Sci.* **9**, 42–44
309. Linden, J. (1991) *FASEB J.* **5**, 2668–2676
310. Palmer, T. M., Benovic, J. L., and Stiles, G. L. (1995) *J. Biol. Chem.* **270**, 29607–29613
311. Erb, L., Garrad, R., Wang, Y., Quinn, T., Turner, J. T., and Weisman, G. A. (1995) *J. Biol. Chem.* **270**, 4185–4188
312. Osborne, W. R. A. (1981) *Trends Biochem. Sci.* **6**, 80–83
313. Anderson, W. F. (1995) *Sci. Am.* **273**(Sep), 124–128
314. Marshall, E. (1995) *Science* **269**, 1050–1055
315. Hershfield, M. S., and Mitchell, B. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1725–1768, McGraw-Hill, New York
316. Sideraki, V., Mohamedali, K. A., Wilson, D. K., Chang, Z., Kellems, R. E., Quiocho, F. A., and Rudolph, F. B. (1996) *Biochemistry* **35**, 7862–7872
317. Wang, Z., and Quiocho, F. A. (1998) *Biochemistry* **37**, 8314–8324
318. Mao.C, Cook, W. J., Zhou, M., Federov, A. A., Almo.SC, and Ealick, S. E. (1998) *Biochemistry* **37**, 7135–7146
319. Cook, W. J., Ealick, S. E., Krenitsky, T. A., Stoeckler, J. D., Helliwell, J. R., and Bugg, C. E. (1985) *J. Biol. Chem.* **260**, 12968–12969
320. Benveniste, P., and Cohen, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8373–8377
321. Deng, H., Kurz, L. C., Rudolph, F. B., and Callender, R. (1998) *Biochemistry* **37**, 4968–4976
321a. Fedorov, A., Shi, W., Kicska, G., Fedorov, E., Tyler, P. C., Furneaux, R. H., Hanson, J. C., Gainsford, G. J., Larese, J. Z., Schramm, V. L., and Almo, S. C. (2001) *Biochemistry* **40**, 853–860
321b. Tebbe, J., Bzowska, A., Wielgus-Kutrowska, B., Schröder, W., Kazimierczuk, Z., Shugar, D., Saenger, W., and Koellner, G. (1999) *J. Mol. Biol.* **294**, 1239–1255
322. McRoberts, J. A., and Martin, D. W., Jr. (1980) *J. Biol. Chem.* **255**, 5605–5615
322a. Xu, Y., and Grubmeyer, C. (1998) *Biochemistry* **37**, 4114–4124
322b. Canyuk, B., Focia, P. J., and Eakin, A. E. (2001) *Biochemistry* **40**, 2754–2765
323. Focia, P. J., Craig, S. P., III, Nieves-Alicea, R., Fletterick, R. J., and Eakin, A. E. (1998) *Biochemistry* **37**, 15066–15075
324. Wilson, J. M., and Kelley, W. N. (1984) *J. Biol. Chem.* **259**, 27–30
325. Rossiter, B. J. F., and Caskey, C. T. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 2 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1679–1706, McGraw-Hill, New York
326. Davidson, B. L., Pashmforoush, M., Kelley, W. N., and Palella, T. D. (1989) *J. Biol. Chem.* **264**, 520–525
327. Willis, R. C., Jolly, D. J., Miller, A. D., Plent, M. M., Esty, A. C., Anderson, P. J., Chang, H.-C., Jones, O. W., Seegmiller, J. E., and Friedmann, T. (1984) *J. Biol. Chem.* **259**, 7842–7849

References

327a. Héroux, A., White, E. L., Ross, L. J., and Borhani, D. W. (1999) *Biochemistry* **38**, 14485–14494
327b. Versées, W., Decanniere, K., Pellé, R., Depoorter, J., Brosens, E., Parkin, D. W., and Steyaert, J. (2001) *J. Mol. Biol.* **307**, 1363–1379
328. Yuan, G., Bin, J. C., McKay, D. J., and Snyder, F. F. (1999) *J. Biol. Chem.* **274**, 8175–8180
328a. Self, W. T., and Stadtman, T. C. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7208–7213
328b. Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H. (2000) *Nature (London)* **406**, 956–957
329. Rumsby, P. C., Kato, H., Waldren, C. A., and Patterson, D. (1982) *J. Biol. Chem.* **257**, 11364–11367
330. Onrot, J., Goldberg, M. R., Biaggioni, I., Hollister, A. S., Kincaid, D., and Robertson, D. (1985) *N. Engl. J. Med.* **313**, 549–554
331. Florkin, M. (1949) *Biochemical Evolution*, Academic Press, New York
332. Henderson, J. F., and Paterson, A. R. P. (1973) *Nucleotide Metabolism*, Academic Press, New York
333. Kahn, K., and Tipton, P. A. (1997) *Biochemistry* **36**, 4731–4738
334. Kahn, K., Serfozo, P., and Tipton, P. A. (1997) *J. Am. Chem. Soc.* **119**, 5435–5442
335. Kahn, K., and Tipton, P. A. (1998) *Biochemistry* **37**, 11651–11659
335a. Enomoto, A., and 18 other authors (2002) *Nature (London)* **417**, 447 – 452
335b. Hediger, M. A., (2002) *Nature (London)* **417**, 393, 395
336. Hayashi, S., Fujiwara, S., and Noguchi, T. (1989) *J. Biol. Chem.* **264**, 3211–3215
337. Reynolds, P. H. S., Boland, M. J., Blevins, D. G., Randall, D. D., and Schubert, K. R. (1982) *Trends Biochem. Sci.* **7**, 366–368
338. Burg, A. W., and Brown, G. M. (1968) *J. Biol. Chem.* **243**, 2349–2358
339. Shiota, T., Palumbo, M. P., and Tsai, L. (1967) *J. Biol. Chem.* **242**, 1961–1969
340. Brown, G. M., and Williamson, J. M. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), pp. 521–538, Am. Soc. for Microbiology, Washington,DC
341. Bracher, A., Eisenreich, W., Schramek, N., Ritz, H., Götze, E., Herrmann, A., Gütlich, M., and Bacher, A. (1998) *J. Biol. Chem.* **273**, 28132–28141
342. Nar, H., Huber, R., Auerbach, G., Fischer, M., Hösl, C., Ritz, H., Bracher, A., Meining, W., Eberhardt, S., and Bacher, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12120–12125
342a. Bracher, A., Fischer, M., Eisenreich, W., Ritz, H., Schramek, N., Boyle, P., Gentili, P., Huber, R., Nar, H., Auerbach, G., and Bacher, A. (1999) *J. Biol. Chem.* **274**, 16727–16735
342b. Bracher, A., Schramek, N., and Bacher, A. (2001) *Biochemistry* **40**, 7896–7902
342c. Maita, N., Okada, K., Hatakeyama, K., and Hakoshima, T. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1212–1217
343. Ahn, C., Byun, J., and Yim, J. (1997) *J. Biol. Chem.* **272**, 15323–15328
344. Haussmann, C., Rohdich, F., Schmidt, E., Bacher, A., and Richter, G. (1998) *J. Biol. Chem.* **273**, 17418–17424
345. Hampele, I. C., D'Arcy, A., Dale, G. E., Kostrewa, D., Nielsen, J., Oefner, C., Page, M. G. P., Schönfeld, H.-J., Stüber, D., and Then, R. L. (1997) *J. Mol. Biol.* **268**, 21–30
346. Sun, X., Bognar, A. L., Baker, E. N., and Smith, C. A. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6647–6652
347. Roy, K., Mitsugi, K., and Sirotnak, F. M. (1996) *J. Biol. Chem.* **271**, 23820–23827
347a. Bermingham, A., Bottomley, J. R., Primrose, W. U., and Derrick, J. P. (2000) *J. Biol. Chem.* **275**, 17962–17967
347b. Baca, A. M., Sirawaraporn, R., Turley, S., Sirawaraporn, W., and Hol, W. G. J. (2000) *J. Mol. Biol.* **302**, 1193–1212
347c. Sun, X., Cross, J. A., Bognar, A. L., Baker, E. N., and Smith, C. A. (2001) *J. Mol. Biol.* **310**, 1067–1078
347d. Ravanel, S., Cherest, H., Jabrin, S., Grunwald, D., Surdin-Kerjan, Y., Douce, R., and Rébeillé, F. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15360–15365
347e. Chave, K. J., Auger, I. E., Galivan, J., and Ryan, T. J. (2000) *J. Biol. Chem.* **275**, 40365–40370
348. Smith, G. K., Cichetti, J. A., Chandrasurin, P., and Nichol, C. A. (1985) *J. Biol. Chem.* **260**, 5221–5224
349. Switchenko, A. C., and Brown, G. M. (1985) *J. Biol. Chem.* **260**, 2945–2951
350. Oppliger, T., Thöny, B., Nar, H., Bürgisser, D., Huber, R., Heizmann, C. W., and Blau, N. (1995) *J. Biol. Chem.* **270**, 29498–29506
351. Bürgisser, D. M., Thöny, B., Redweik, U., Hess, D., Heizmann, C. W., Huber, R., and Nar, H. (1995) *J. Mol. Biol.* **253**, 358–369
352. Ploom, T., Thöny, B., Yim, J., Lee, S., Nar, H., Leimbacher, W., Richardson, J., Huber, R., and Auerbach, G. (1999) *J. Mol. Biol.* **286**, 851–860
353. Citron, B. A., Milstien, S., Gutierrez, J. C., Levine, R. A., Yanak, B. L., and Kaufman, S. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6436–6440
354. Auerbach, G., Herrmann, A., Gütlich, M., Fischer, M., Jacob, U., Bacher, A., and Huber, R. (1997) *EMBO J.* **16**, 7219–7230
355. Nichol, C. A., Smith, G. K., and Duch, D. S. (1985) *Ann. Rev. Biochem.* **54**, 729–764
356. Tanaka, K., Akino, M., Hagi, Y., Doi, M., and Shiota, T. (1981) *J. Biol. Chem.* **256**, 2963–2972
357. Wiederrecht, G. J., and Brown, G. M. (1984) *J. Biol. Chem.* **259**, 14121–14127
358. Jacobson, K. B., Dorsett, D., Pfleiderer, W., McCloskey, J. A., Sethi, S. K., Buchanan, M. V., and Rubin, I. B. (1982) *Biochemistry* **21**, 1238–1243
359. White, R. H. (1996) *Biochemistry* **35**, 3447–3456
360. Rasche, M. E., and White, R. H. (1998) *Biochemistry* **37**, 11343–11351
361. Wuebbens, M. M., and Rajagopalan, K. V. (1995) *J. Biol. Chem.* **270**, 1082–1087
361a. Unkles, S. E., Heck, I. S., Appleyard, M. V. C. L., and Kinghorn, J. R. (1999) *J. Biol. Chem.* **274**, 19286–19293
361b. Leimkühler, S., Wuebbens, M. M., and Rajagopalan, K. V. (2001) *J. Biol. Chem.* **276**, 34695–34701
361c. Leimkühler, S., and Rajagopalan, K. V. (2001) *J. Biol. Chem.* **276**, 1837–1844
361d. Schrag, J. D., Huang, W., Sivaraman, J., Smith, C., Plamondon, J., Larocque, R., Matte, A., and Cygler, M. (2001) *J. Mol. Biol.* **310**, 419–431
361e. Kuper, J., Palmer, T., Mendel, R. R., and Schwarz, G. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6475–6480
361f. Watanabe, T., Ihara, N., Itoh, T., Fujita, T., and Sugimoto, Y. (2000) *J. Biol. Chem.* **275**, 21789–21792
361g. Bittner, F., Oreb, M., and Mendel, R. R. (2001) *J. Biol. Chem.* **276**, 40381–40384
361h. Temple, C. A., and Rajagopalan, K. V. (2000) *J. Biol. Chem.* **275**, 40202–40210
361i. Lake, M. W., Temple, C. A., Rajagopalan, K. V., and Schindelin, H. (2000) *J. Biol. Chem.* **275**, 40211–40217
362. Ritz, H., Schramek, N., Bracher, A., Herz, S., Eisenreich, W., Richter, G., and Bacher, A. (2001) *J. Biol. Chem.* **276**, 22273–22277
363. Kis, K., Volk, R., and Bacher, A. (1995) *Biochemistry* **34**, 2883–2892
363a. Kelly, M. J. S., Ball, L. J., Krieger, C., Yu, Y., Fischer, M., Schiffmann, S., Schmieder, P., Kühne, R., Bermel, W., Bacher, A., Richter, G., and Oschkinat, H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13025–13030
363b. Illarionov, B., Eisenreich, W., and Bacher, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98** 7224 – 7229
364. García-Ramírez, J. J., Santos, M. A., and Revuelta, J. L. (1995) *J. Biol. Chem.* **270**, 23801–23807
364a. Jordan, D. B., Bacot, K. O., Carlson, T. J., Kessel, M., and Viitanen, P. V. (1999) *J. Biol. Chem.* **274**, 22114–22121
364b. Braden, B. C., Velikovsky, C. A., Cauerhff, A. A., Polikarpov, I., and Goldbaum, F. A. (2000) *J. Mol. Biol.* **297**, 1031–1036
364c. Meining, W., Mörtl, S., Fischer, M., Cushman, M., Bacher, A., and Ladenstein, R. (2000) *J. Mol. Biol.* **299**, 181–197
364d. Illarionov, B., Kemter, K., Eberhardt, S., Richter, G., Cushman, M., and Bacher, A. (2001) *J. Biol. Chem.* **276**, 11524–11530
365. Ritsert, K., Huber, R., Turk, D., Ladenstein, R., Schmidt-Bäse, K., and Bacher, A. (1995) *J. Mol. Biol.* **253**, 151–167
366. Efimov, I., Kuusk, V., Zhang, X., and McIntire, W. S. (1998) *Biochemistry* **37**, 9716–9723
367. Bowers-Komro, D. M., Yamada, Y., and McCormick, D. B. (1989) *Biochemistry* **28**, 8439–8446
368. Eisenreich, W., Schwarzkopf, B., and Bacher, A. (1991) *J. Biol. Chem.* **266**, 9622–9631
369. Hörig, J. A., Renz, P., and Heckmann, G. (1978) *J. Biol. Chem.* **253**, 7410–7414
370. Thompson, T. B., Thomas, M. G., Escalante-Semerena, J. C., and Rayment, I. (1998) *Biochemistry* **37**, 7686–7695
371. Trzebiatowski, J. R., and Escalante-Semerena, J. C. (1997) *J. Biol. Chem.* **272**, 17662–17667
372. Himmeldirk, K., Sayer, B. G., and Spenser, I. D. (1998) *J. Am. Chem. Soc.* **120**, 3581–3589
372a. Miranda-Rios, J., Navarro, M., and Soberón, M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9736–9741
373. Estramareix, B., and Thérisod, M. (1984) *J. Am. Chem. Soc.* **106**, 3857–3860
374. Taylor, S. V., Kelleher, N. L., Kinsland, C., Chiu, H.-J., Costello, C. A., Backstrom, A. D., McLafferty, F. W., and Begley, T. P. (1998) *J. Biol. Chem.* **273**, 16555–16560
374a. Mueller, E. G., Palenchar, P. M., and Buck, C. J. (2001) *J. Biol. Chem.* **276**, 33588–33595
375. Julliard, J.-H., and Douce, R. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2042–2045
376. Kelleher, N. L., Taylor, S. V., Grannis, D., Kinsland, C., Chiu, H.-J., Begley, T. P., and McLafferty, F. W. (1998) *Protein Sci.* **7**, 1796–1801
377. Chiu, H.-J., Reddick, J. J., Begley, T. P., and Ealick, S. E. (1999) *Biochemistry* **38**, 6460–6470
377a. Peapus, D. H., Chiu, H.-J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) *Biochemistry* **40**, 10103–10114
377b. Reddick, J. J., Nicewonger, R., and Begley, T. P. (2001) *Biochemistry* **40**, 10095–10102
378. Cane, D. E., Hsiung, Y., Cornish, J. A., Robinson, J. K., and Spenser, I. D. (1998) *J. Am. Chem. Soc.* **120**, 1936–1937
378a. Cane, D. E., Du, S., and Spenser, I. D. (2000) *J. Am. Chem. Soc.* **122**, 4213–4214
378b. Osmani, A. H., May, G. S., and Osmani, S. A. (1999) *J. Biol. Chem.* **274**, 23565–23569
378c. Gupta, R. N., Hemscheidt, T., Sayer, B. G., and Spenser, I. D. (2001) *J. Am. Chem. Soc.* **123**, 11353–11359

References

379. Huynh, M. S., and Snell, E. E. (1985) *J. Biol. Chem.* **260**, 2379–2383
380. Nelson, M. J. K., and Snell, E. E. (1986) *J. Biol. Chem.* **261**, 15115–15120
381. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) *Biochemistry* **36**, 13856–13864
382. Fleming, S. M., Robertson, T. A., Langley, G. J., and Bugg, T. D. H. (2000) *Biochemistry* **39**, 1522–1531
383. Krieger, C. J., Roseboom, W., Albracht, S. P. J., and Spormann, A. M. (2001) *J. Biol. Chem.* **276**, 12924–12927

Study Questions

1. The vitamin niacin (nicotinic acid) is converted to NAD through the intermediate, desamido NAD.

   a) Propose an enzymatic synthesis of desamido NAD beginning from niacin and ribose 5-phosphate
   b) Propose an enzymatic synthesis of NAD beginning from desamido NAD

2. AMP is formed from IMP in an ATP requiring reaction. The introduced nitrogen atom is derived directly from aspartate.

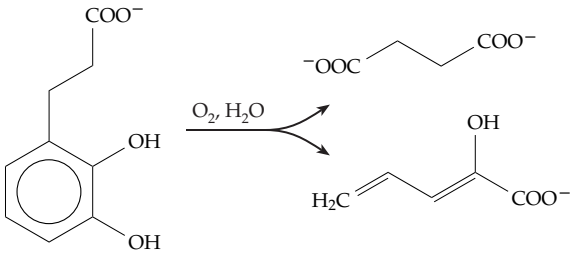
   a) Propose a reaction mechanism based on an analogous reaction sequence occurring during amino acid metabolism.
   b) Identify the analogous enzymes upon which you have based your answer by names of the enzymes or by the names of the substrates and products.

3. One of the mechanisms proposed for the decarboxylation of orotidine 5'-phosphate to UMP involves initial addition of an enzyme nucleophile to the pyrimidine ring. Describe and criticize this mechanism. Hint: The proposal has some similarity to the thymidylate synthase mechanism.

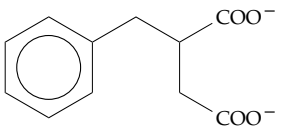
4. Name the enzymes and describe the chemical reaction that occurs for each of the four steps in the following pathway:

$$\text{UDP} \xrightarrow{1} \text{dUDP} \xrightarrow{2} \text{dUTP} \xrightarrow{3} \text{dUMP} \xrightarrow{4} \text{dTMP}$$

5. The following reaction is catalyzed by a dioxygenase from *E. coli*. The dioxygenase reaction opens the ring and the intermediate is cleaved hydrolytically. Propose a structure for the intermediate and a mechanism for the hydrolytic cleavage. See Fleming *et al*.[382]

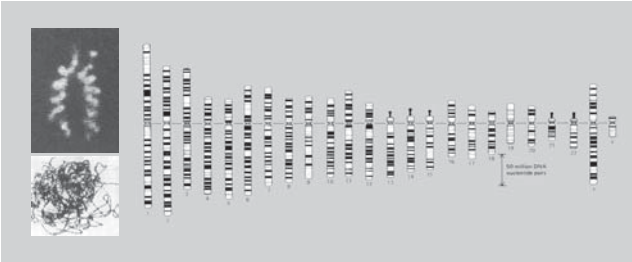


6. Anaerobic breakdown of toluene by denitrifying bacteria begins by the addition of toluene to fumarate to form benzylsuccinate. Benzylsuccinate synthase has an amino acid sequence homologous to that of pyruvate formate lyase (p. 799–801), contains a glycyl radical, and is activated in a manner similar to activation of pyruvate formate lyase. Propose a mechanism for formation of benzylsuccinate. See Krieger *et al.*[383]



Benzylsuccinate

7. Write out, using structural formulas, a step-by-step reaction sequence for the conversion of *O*-succinylbenzoate into isochorismate as indicated in Fig. 25-4.



Right: The human **karyotype**, a full set of chromosomes numbered according to size and content and showing characteristic banding patterns. The 22 autosomes plus either an X or Y chromosome make up the haploid set which, for the female, contains ~ 3,500 Mbp of DNA. Each diploid cell contains 46 chromosomes. From Alberts *et al.* "Molecular Biology of the Cell", Third edition, Garland, New York, 1994, p 356.

Upper Left: A pair of mitotic sister chromatids in a section stained with an antibody to topoisomerase II. Notice that the two chromatids are coiled with opposite helical handedness. Lower Left: Meiotic chromosomes of a lily at the pachytene stage in which sister chromatids are connected along their length in a **synaptonemal complex.** From Kleckner, N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8167–8174

Contents

1473 **A. Historical Sketch**
1473 1. DNA as the Genetic Material
1474 2. The Double Helix
1474 3. Ribonucleic Acids and Proteins
1474 *Messenger RNA*
1475 *Transfer RNA*
1475 4. Deciphering the Genetic Code
1476 **B. Genetic Methods**
1476 1. Mutations
1477 2. Mapping the Chromosome of a Bacteriophage
1478 3. The Cistron
1478 4. Nutritional Auxotrophs
1479 5. Establishing the Correspondence of a Genetic Map and an Amino Acid Sequence
1480 6. Conditionally Lethal Mutations
1480 7. The Nature of Suppressor Genes
1481 **C. Plasmids, Episomes, and Viruses**
1482 1. Bacterial Sex Factors
1483 2. Temperate Bacteriophage; Phage Lambda
1486 **D. Mapping of Chromosomes**
1486 1. The Chromosome Map of *E. coli*
1486 2. Restriction Endonucleases
1488 3. Restriction Mapping
1489 4. Electron Microscopy
1490 5. Optical Mapping
1490 **E. Cloning, Modifying, and Transferring Genes**
1490 1. Joining DNA Fragments
1491 *Some useful enzymes*
1491 *Forming homopolymeric tails*
1491 *Preparing material for cloning*
1492 2. Cloning Vehicles (Vectors)
1493 *Plasmids related to ColE1*
1494 *Typical cloning procedure with pBR322*
1494 *Selecting clones using β-galactosidase in pUC cloning vehicles*
1494 *Filamentous bacteriophages*
1496 *Lambda cloning vehicles*
1496 *Cosmid vehicles and in vitro packaging*
1497 3. Expression of Cloned Genes in Bacteria
1497 4. Cloning and Transferring Eukaryotic Genes
1497 *Plant genes*
1498 *Transferring genes with engineered animal viruses*
1498 *Physical methods of gene transfer*
1499 5. Genomic Libraries
1499 6. Probes
1500 7. Studies of Restriction Fragments
1500 *Locating mutations*
1500 *Positional cloning*
1500 8. Directed Mutation
1501 *Targeting and replacing genes*
1501 *Knockout mice*
1501 **F. The Genetics of Eukaryotic Organisms**
1501 1. Mitosis
1502 *Cyclin-dependent kinases*
1502 *Checking for completion of replication and for DNA damage*
1503 *Mitotic spindle formation and the spindle assembly checkpoint*
1503 *Anaphase*
1504 2. Meiosis
1506 3. Polytene Chromosomes
1507 4. Cytoplasmic Inheritance
1507 **G. The Human Genome Project**
1507 1. The Mammalian Genome and Human Health
1508 2. Understanding Gene Sequences
1509 *Human variation*
1510 *Other evolutionary relationships*
1511 *Metabolic studies of eukaryotic cells*
1513 3. Understanding Human Genetic Diseases
1516 *Triplet repeat diseases*
1517 *Cloned genes and diagnosis*
1517 *Vaccines, hormones, and other medicines*
1518 4. Gene Therapy
1518 5. Genetic Engineering of Bacteria, Plants, and Animals
1519 6. Ethical Problems

1520 **References**
1527 **Study Questions**

Boxes

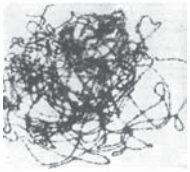
1513 Box 26-A Cystic Fibrosis

Tables

1485 Table 26-1 Some Genes of *E. coli*
1487 Table 26-2 Some Commonly Used Restriction Endonucleases, Their Sources, and Cleavage Sites
1492 Table 26-3 Some Enzymes Used in Molecular Cloning
1515 Table 26-4 Some Human Triple Repeat and Related Diseases

Biochemical Genetics

26



The most exciting biological discoveries of the 20th century include the unraveling of the genetic code (Chapter 5) and the understanding of the ways in which nucleic acids and proteins are synthesized. The biosynthesis of both nucleotides and amino acids has been considered in Chapters 24 and 25 and the basic chemistry of the polymerization processes in Chapters 12 and 17. Chapters 27–29 deal with the ways in which these polymerization reactions are controlled and by which the correct sequences of nucleotides or amino acids are obtained. The understanding of these matters is a development of genetics[1–2b] as well as of biochemistry; hence this introductory and historical chapter.

A. Historical Sketch

The discovery of deoxyribonucleic acid dates to 1869, when Miescher isolated a new chemical substance from white blood cells that he obtained from pus and later from sperm cells.[3] The material, which became known as nucleic acid, occurred in both plants and animals, thymus glands and yeast cells being among the best sources. Chemical studies indicated that the nucleic acids isolated from thymus glands and from yeast cells were different. As we now know, thymus nucleic acid was primarily DNA and yeast nucleic acid primarily RNA. For a while it was suspected that animals contained only DNA and plants only RNA, and it was not until the early 1940s that it was established that both substances were present in all organisms.[3–6]

1. DNA as the Genetic Material

In 1928, Griffith, using cells of *Diplococcus pneumoniae*, showed that genetic information that controls properties of the capsular polysaccharides could be transferred from one strain of bacteria to another. A material present in killed cells and in cell-free extracts permanently altered the capsular properties of cells exposed to the material.[7] This **transformation** of bacteria remained a mystery for many years. At the time of the experiments there was no hint of the genetic role of nucleic acids, which were generally regarded as strange materials. Furthermore, the covalent structure of nucleic acids was uncertain. A popular idea was that a tetranucleotide served as a repeating unit for some kind of regular polymer. Genes were most often thought to be protein in nature.

However, in 1944, Avery and associates showed that purified DNA extracted from pneumococci could carry out transformation.[8–12] The transforming principal appeared to contain little protein. It was not inactivated by proteolytic enzymes but was inactivated by deoxyribonuclease. Avery was 67 at the time of this discovery, refuting the popular contention that all important scientific discoveries are made by young people.

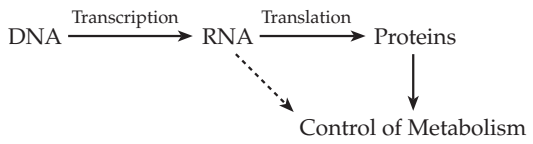
Other experiments also pointed to the conclusion that DNA was the genetic material. DNA was found localized in the nuclei of eukaryotic cells. The absolute amount per cell was constant for a given species. Studies of bacteriophage replication pointed strongly to DNA as the genetic material.[11] In 1952 Hershey and Chase showed that when a phage particle infects a cell the viral DNA enters the bacterium, but the protein "coat" remains outside.[13] This was demonstrated by

preparing two types of labeled bacteriophage T2. One contained ^{32}P, which had been incorporated into the DNA, and the other ^{35}S incorporated into the proteins. Cells of *E. coli* were infected with the labeled phage preparations and were then agitated violently in a blender to shear the phage particles off of the bacteria. Over 80% of the ^{35}S was removed from the bacteria by this treatment, but most of the ^{32}P entered the bacteria and could be recovered in the next generation of progeny bacteriophage.[14]

2. The Double Helix

Development of newer methods of investigation of the chemical composition of nucleic acids led Chargaff to an important discovery. Although the base composition of DNA was extremely variable from organism to organism, the molar ratio of adenine to thymine was nearly always 1:1, as was that of cytosine to guanine![15] This observation provided the basis for the concept of base pairing in the structure of DNA. The final important information was supplied by X-ray diffraction studies of stretched fibers of DNA, which showed that molecules of DNA were almost certainly helical structures containing more than one chain. The crucial experiments were done by Rosalind Franklin[16–17b] and M. H. F. Wilkins, whose data were used by Watson and Crick in 1952 in constructing their model of the double helical structure[12,18–20] (Fig. 5-6). Once established, the structure of DNA itself suggested both the nature of the genetic code and a replication mechanism. *The genetic code had to lie in the sequence of nucleotides. Base pairing provided a mechanism by which the two mutually complementary strands could be separated. Biosynthesis of a new complementary strand alongside each one would result in a precise replication of each gene and of the entire genome.* In a similar fashion, RNA could be synthesized alongside a DNA "template" and could then be used to direct protein synthesis.

The presence of RNA in the cytoplasm had been linked to protein synthesis by experiments done in the early 1940s. After the discovery of the double helix, the concept followed quickly that DNA was the master "blueprint" from which secondary blueprints or **transcripts** of RNA could be copied. The RNA copies, later identified as **messenger RNA** (mRNA), provided the genetic information for specifying protein sequence. The flow of information from DNA to RNA to proteins could be symbolized as in Eq. 26-1.



(26-1)

Proteins, in one way or another, control nearly all of metabolism. This includes the reactions that form the nucleotide precursors of the nucleic acids and that lead to polymerization of the amino acids and nucleotides, reactions catalyzed by **protein enzymes** and **ribozymes**. Thus, the flow of information from DNA to proteins is only part of a larger loop of metabolic processes (p. 973). Genetic information flows from DNA out into the cell, and copies of the master blueprint are passed nearly unchanged from generation to generation. The simple concepts implied by Eq. 26-1 quickly caught the imagination of the entire community of scientists and led to a rapid blossoming of the field of biochemical genetics.[20a,b]

3. Ribonucleic Acids and Proteins

By 1942 it was clear from ultraviolet cytophotometry developed by Caspersson[21] and from cytochemical work of Brachet[22,23] that RNA had something to do with protein synthesis. Use of radioautography with ^{3}H-containing uridine showed that RNA was synthesized in the nucleus of eukaryotic cells and was transported out into the cytoplasm.[24,25]

Ribosomes were discovered by electron microscopists examining the structure of the endoplasmic reticulum using ultrathin sectioning techniques. Their presence in cells was established by 1956, and the name ribosome was proposed in 1957. At first it was difficult to study protein synthesis *in vitro* using isolated ribosomes. No *net* synthesis could be detected until Hoagland and associates measured the rate of incorporation of ^{14}C-labeled amino acids into protein.[26] This sensitive method permitted measurement of very small amounts of protein synthesis in cell-free preparations from rat liver and paved the way toward studies with ribosomes themselves.

Immediately after the Watson–Crick proposals were made in 1953, it was generally thought that ribosomal RNA (rRNA), which constitutes up to 90% of the total RNA of some cells, carried the genetic message from the nucleus to the cytoplasm. By 1960 it seemed unlikely. For one thing the size and composition of rRNA was similar for different bacteria, despite differences in base composition of the DNA (Chapter 5).[27] It had been concluded that a relatively unstable, short-lived form of RNA must carry the message. Ribosomal RNA, however, was quite stable.[28]

Messenger RNA. In 1956, Volkin and Astrachan[29,30] detected a rapidly labeled and labile RNA in phage-infected bacterial cells. Studies of enzyme induction also suggested the existence of mRNA. Many bacteria, including *E. coli*, when grown on glucose as the sole source of energy and then suddenly switched to lactose, are unable to utilize the new sugar

immediately. However, transfer to lactose induces, within a period of two minutes, the synthesis of new proteins needed for the metabolism of this sugar. Among the new proteins is a **permease** for lactose and a **β-galactosidase** (Chapter 12)[31] that cleaves the disaccharide to glucose and galactose. When the lactose is exhausted, the level of the induced enzymes drops almost as quickly. These results suggested that the RNA that carries the genetic message for synthesis of the new enzymes must be unstable. It must be produced rapidly in response to the presence of the inducing sugar and must disappear rapidly in its absence.

In 1961, Jacob and Monod postulated messenger RNA (mRNA) as a short-lived polynucleotide.[30,32,33] An abundance of additional evidence supported the proposal. For example, RNA molecules produced after infection of *E. coli* by bacteriophage T4 underwent hybridization (Chapter 5) with denatured DNA of the bacteriophage. Furthermore, this virus-specific mRNA became associated with preexisting bacterial ribosomes and provided the template for synthesis of phage proteins.[34] The experiment provided direct evidence for transcription of mRNA from genes of the viral DNA.

Transfer RNA. Crick[35] suggested, in 1957, that special "adapter" molecules might be needed to align amino acids with their codons in the RNA transcript. He thought that the adapters might be polynucleotides. At the same time chemical studies of the RNA of cells revealed that a low-molecular-weight RNA made up 15% of the total RNA of *E. coli*. This RNA was recognized in the same year as constituting the needed adapters, when Hoagland demonstrated enzymatic "activation" of amino acids and their subsequent incorporation into protein. The name transfer RNA (tRNA; Figs. 5-30, 5-31) was proposed.[36]

4. Deciphering the Genetic Code

The genetic code consists of triplets of DNA base pairs (codons), each corresponding to a single amino acid. The triplets are consecutive, do not overlap, and are not separated by any "punctuation." Although this represented one of the simplest possibilities for a code, it required much effort more than a period of more than 10 years to prove it. Even after the triplet nature of the genetic code became evident, many questions remained. Were all of the 64 possible codons used by the living cell? If so, were they all used to code for amino acids or were some set aside for other purposes? How many codons were used for a single amino acid? Was the code universal or did different organisms use different codes? How could one decipher the code? Despite the complexity of these questions, they seem to have been almost completely answered.

In an important experiment[37] Nirenberg and Matthaei, in 1961, isolated ribosomes from *E. coli* and mixed them with crude extracts of soluble materials, also from *E. coli* cells. The extracts included tRNA molecules and aminocyl-tRNA synthases. The 20 amino acids, ATP, and an ATP-generating system (PEP + pyruvate kinase) were added. Nirenberg showed that under such conditions protein was synthesized by ribosomes in response to the presence of added RNA. For example, RNA from tobacco mosaic virus (Chapter 7) was very effective in stimulating protein synthesis. The crucial experiment, which was done originally simply as a "control," was one in which a synthetic polynucleotide consisting solely of uridylic acid units was substituted for mRNA. In effect, this was a synthetic mRNA containing only the codon UUU repeated over and over. The ribosomes read this code and synthesized a peptide containing only phenylalanine. Thus, poly(U) gave polyphenylalanine, and *UUU was identified as a codon specifying phenylalanine*. The first nucleotide triplet had been identified! In the same manner CCC was identified as a proline codon and AAA as a lysine codon. Study of mixed copolymers containing two different nucleotides in a random sequence suggested other codon assignments. A few years later, after Khorana had supplied the methods for synthesis of oligonucleotides and of regular alternating polymers of known sequence, the remaining codons were identified.

Another important technique was based on the observation that synthetic trinucleotides induced the binding to ribosomes of tRNA molecules that were "charged" with their specific amino acids.[38,39] For example, the trinucleotides UpUpU and ApApA stimulated the binding to ribosomes of ^{14}C-labeled phenylalanyl-tRNA and lysyl-tRNA, respectively. The corresponding dinucleotides had no effect, an observation that not only verified the two codons but also provided direct evidence for the triplet nature of the genetic code. Another powerful approach was the use of artificial RNA polymers, synthesized by combined chemical and enzymatic approaches.[40] For example, the polynucleotide CUCUCUCUCU · · · led to the synthesis by ribosomes of a regular alternating polypeptide of leucine and serine.

Table 5-5 shows the codon assignments, as we now know them, for each of the 20 amino acids. Table 5-6 shows the same 64 codons in a rectangular array. In addition to those codons assigned to specific amino acids, three are designated as **chain termination codons**: UAA, UAG, and UGA. These are frequently referred to as "nonsense" codons. The termination codons UAA and UAG are also known as *ochre* and *amber*, respectively, although these names have no scientific significance.[41] The codons AUG (methionine)

and much less often GUG (valine) serve as the **initiation codons** in protein synthesis. Consequently, the N-terminal amino acid in most newly synthesized eukaryotic proteins is methionine, and in bacterial proteins it is *N*-formylmethionine. As explained in Chapter 29, *N*-formylmethionyl-tRNA is specifically bound to initiation sites containing the AUG codon in bacterial mRNA-ribosome complexes.

A number of studies suggested that the genetic code as worked out for *E. coli* might be universal. For example, in the laboratories of Wittman and of Fraenkel-Conrat, RNA extracted from tobacco mosaic virus was treated with nitrous acid, a procedure known to deaminate many cytosine residues to uracil (Eq. 5-12). Such treatment could change the codon UCU (serine) to UUU (phenylalanine) and the codon CCC (proline) to CUC (leucine). When the nitrous acid-treated RNA was used to infect tobacco plants and virus particles were prepared in quantity from the resultant mutant strains, it was found that the amino acid sequence of the virus coat protein had been altered,[42] and that many of the alterations were exactly those that would be predicted from Table 5-6. Likewise, the amino acid substitutions in known defects of hemoglobin (Fig. 7-23) could be accounted for, in most cases, by single base alterations. Thus, hemoglobin S arose as a result of the following change in the sixth codon of the globin β chain gene: GAG (Glu) → GTG (Val).[43] Another argument favoring a universal code was based on the observation that mRNA coding for a globin chain could be translated by ribosomes and tRNA molecules from *E. coli.* The resulting protein was authentic mammalian globin.[44]

As often happens, a well-established conclusion may have to be modified. There are exceptions to the universal genetic code in mitochondrial DNA and in some protozoa (Chapter 5).[45]

B. Genetic Methods

Our present knowledge of molecular biology has depended greatly on the methods of genetics. The following introduction begins with a consideration of mutations.

1. Mutations

Changes in the structure of DNA occur only rarely. The average gene may be duplicated 10^6 times before some mistake results in a single detectable mutation.[46] Nevertheless, by using bacteria or bacterial viruses, it is possible to screen enormous numbers of individuals for the occurrence of mutations. If one million virus particles are spread on an agar plate under conditions where mutation in a certain gene can be recognized, on the average one mutant is found. The most common mutations are **base-pair switches** or **point mutations** that result from incorporation of the wrong base during replication or repair. In these mutations one base of a triplet codon is replaced by another to form a different codon, causing the substitution of one amino acid by another in the corresponding protein as was seen for hemoglobin S. Changes involving replacement of one pyrimidine by another (C → T or T → C) or of one purine by another are sometimes called **transition mutations**, whereas if a pyrimidine is replaced by a purine, or vice versa, the mutation is known as a **transversion**. An example is the previously mentioned A → T in hemoglobin S. Transition mutations are by far the most common, one possible cause being pairing with a minor tautomer of one of the bases (Chapter 5). For example, A could pair with a minor tautomer of C, causing a mutation from T to C. Note that substitution of an incorrect base in one strand will lead, in the next round of replication, to correct pairing again but with an AT pair replaced by GC, or vice versa, in one of the daughter DNA duplex strands. A base substitution does not always cause an amino acid replacement because of the "degeneracy" of the code, i.e., the fact that more than one codon specifies a given amino acid.

From the observed rate of appearance of point mutations (one mutation per 10^6 gene duplications), we can estimate that one mutation occurs per 10^9 replications at a single nucleotide site. Point mutants tend to "back mutate," often at almost the same rate as is observed for the forward mutation. That is, one in 10^9 times a mutation of the same nucleotide will take place to return the code to its original form. The phenomenon is easy to understand. For example, if T should be replaced by C because the latter formed a minor tautomer and paired with A, the mutation would appear in progeny duplexes as a GC pair. When this pair was replicated, there would be a finite probability that the C of the parental DNA strand would again assume the minor tautomeric structure and pair with A instead of G, leading to a back mutation.

Although the rates of spontaneous mutation are low, they can be greatly increased by mutagenic chemicals (Chapter 27) or by irradiation. It is perfectly practical to measure the rates of both forward and back mutation. When this was done, it was found that certain chemicals, e.g., acridine dyes, induce mutations that undergo reverse mutation at a very much lower frequency than normal. It was eventually shown that these mutations resulted either from **deletions** of one or more nucleotides from the chain or from **insertions** of extra nucleotides. Deletion and insertion mutations often result from errors during genetic recombination and repair at times when the DNA chain is broken.

Mutations involving deletion or insertion of one or a few nucleotides are called **frame-shift mutations**.

Messenger RNA is read by the protein synthesizing machinery from some starting point. As is illustrated in Chapter 5, Section E,1 the codons are read three bases at a time, and the proper amino acid corresponding to each codon is inserted. *When a deletion or insertion in the mRNA is met, all subsequent codons may be misread because the reading frame has shifted forward or backward by one or two nucleotides.* The protein synthesized bears little resemblance to that formed by the nonmutant organism and is usually completely nonfunctional. Mutations are considered further in Chapter 27.

2. Mapping the Chromosome of a Bacteriophage

Intensive work on the "T-even" phage T2, T4, and T6 (Box 7-C) was begun in 1938 by Delbrück and associates. The genetic information for these viruses is carried in a single linear DNA molecule, which in the case of T4 contains ~1.7×10^5 base pairs (170 kb), enough for about 200 genes. Before the sequences of the viral DNA were known, the positions of more than 60 of these genes were mapped in the following way. When a bacteriophage infects a cell of *E. coli*, it injects its DNA through the cell wall and into the cytoplasm. About 20 minutes later the cell bursts, and ~100 fully formed replicas of the original virus particle are released. This rapid rate of production of progeny is so fast that it is possible to carry out in a test tube in 20 min a genetic experiment that would require the entire population of the earth if done with humans. The approach is explained nicely by Seymour Benzer, the man who first mapped the fine structure of a gene.[47] Bacteriophage particles, like bacteria, can be "plated out" on agar plates, which must contain a uniform suspension of bacteria susceptible to the virus. Wherever a virus particle lies, a bacterium is infected. Soon the infection spreads to neighboring bacteria with production of a transparent "plaque" (Fig. 26-1). The number of active virus particles present in a suspension can be determined easily by plating and counting of the plaques.

Mutant bacteriophages can be identified in various ways. Some biochemical traits affect the appearance of the plaque. Other easily detected traits include alteration in the specificity toward strains of the host bacterium. A key discovery that made genetic mapping possible for bacteriophage was that *genetic recombination between two phage particles can take place within a host bacterium*. When large numbers of bacteriophages of two different mutant strains were grown and were mixed together in excess with many bacteria, a few of the progeny phage were found to contain both mutant traits in the same virus and an equal number were "wild type." Although recombinations between mutations that are located close together in the DNA are rare, their frequency still greatly exceeds that of new mutations. While this type of experiment gave no hint about the nature of the events involved, it showed that recombination had occurred.

Study of recombination frequencies between different strains of phage soon revealed that some sites of mutation are **closely linked**. Recombination between these sites occurs only rarely. Other sites are weakly linked, and recombination occurs often. This behavior was reminiscent of that established many years earlier for genes of the fruit fly *Drosophila*, maize, and other higher organisms. Recombination by "crossing over" in the chromosomes of *Drosophila* was established by Morgan and associates in 1911.[1,2] The basic idea behind chromosome mapping in any organism is the assumption that *recombination frequencies between two mutations are directly proportional to the distance between them on the genetic map*. For the T4 phage a recombination frequency of 1% is taken as one unit. The total T4 map is 700 units long. The fact that this is greater than 100% means that if genes are located at opposite ends of the chromosome multiple recombination events can occur between them. However, a maximum of 50% crossing over is observed for distant gene pairs, and the approximate linearity of map distance and recombination frequency holds only for distances of 10 units or less.[48]

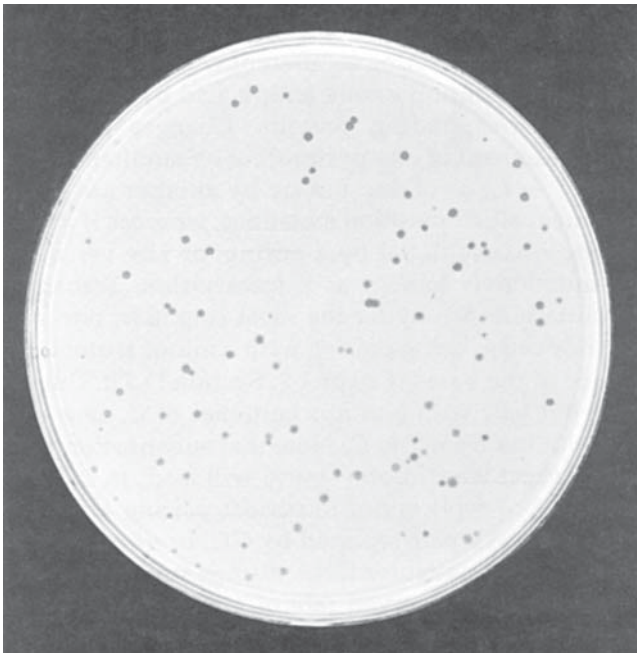


Figure 26-1 Plaques formed by bacteriophage ϕ11 growing on *Staphylococcus aureus*. Each transparent (dark) plaque is the result of lysis of bacteria by the progeny of a single bacteriophage particle. Courtesy of Peter Pattee.

How can recombinant bacteriophage be identified rapidly? Benzer used two strains of *E. coli*, the B strain and the K strain, as hosts. Mutants in gene *rII* form characteristic plaques on strain B but do not grow on strain K. To determine the recombination frequency between two different *rII* mutants, the viruses were added to liquid cultures of B cells (in which they replicate), and recombination was allowed to occur. Recombination permitted the emergence not only of a phage containing *both* mutations but also of a wild-type phage in which both mutations had been eliminated by the recombination process. Since only recombinants of the latter type grow in strain K, it was possible to detect a single recombinant among one billion progeny. Since the total DNA length in phage T4 is 166 kb, there are 237 base pairs for each of the 700 units of map length. Thus, a recombination frequency at 0.01% between two mutations meant that the two mutations were no more than three base pairs apart in the DNA. Benzer concluded that he had observed the expected recombination frequencies for mutations even of immediately adjacent bases in the DNA.

To make fine genetic mapping practical, a series of bacteriophage containing deletion mutations involving large segments of the *rII* gene were isolated. Using these, it was possible to establish in which segments of the gene a particular mutation lay. Then, recombination experiments with previously identified mutations in that same general region allowed the mutations to be pinpointed. In this way Benzer identified over 300 sites of mutations within the *rII* region. He concluded that the minimum distance between two mutable sites was compatible with the Watson–Crick structure of the gene.[47]

3. The Cistron

How can one tell whether two mutations are in the same gene or in nearby or adjacent genes? The answer can be supplied by a test of **complementation**. If two mutant bacteriophage are altered in different genes, they can often reproduce within a host if the bacterium is infected with both of the mutant phage. Since each one has a good gene for one of the two proteins involved, recombinant phage, in which all of the gene functions are fulfilled, will be formed. On the other hand, if both mutant phages are defective in the same gene (although at different locations), they usually cannot complement each other at high frequencies in a coinfection. The experiment is referred to as a ***cis–trans* comparison**. The coinfection with the two different mutants is the trans test. A control, the *cis* test, uses a recombinant containing both of the mutations in the same DNA and coinfection with a standard phage. Normal replication is expected in this instance.

When the complementation (or *cis–trans*) test was applied to various mutants in the *rII* region, it became clear that there are two genes, *rIIA* and *rIIB*. The name **cistron** was proposed by Benzer to represent that length of DNA identifiable in this fashion as a genetic unit. *For most purposes the terms gene and cistron are nearly synonymous.* When mapping of the *rII* region was done, there was no information about the functions of the proteins specified by these two cistrons. However, both the *rIIA* and *rIIB* proteins have since been shown to become incorporated in the membranes of phage-infected bacterial cells.[49,50] There they affect the ease of lysis of the infected cells and, in that manner, cause *rII* plaques to be larger and to have sharper edges than standard plaques.

4. Nutritional Auxotrophs

The beginning of biochemical genetics is often attributed to Beadle and Tatum, who, in 1940, discovered mutants of *Neurospora* with specific growth requirements. They X-rayed one parent strain to form mutants, then tested individual spores for their possible need of a specific nutrient for growth. The 299th spore tested required pyridoxine! Many other mutants requiring vitamins, amino acids, and nucleic acid bases were then discovered.[1,51] A few years later similar **nutritional auxotrophs**, as the mutants are called, were discovered for *E. coli*. Ordinary wild-type cells of *E. coli* can grow on a minimal medium containing a carbon compound as a source for energy together with inorganic nutrients. Irradiation with ultraviolet light or treatment with mutagenic chemicals produces many mutant cells that fail to grow on such a minimal medium. However, addition of one or more specific compounds, such as an amino acid or vitamin, usually permits growth. Selection of such nutritional auxotrophs can be accomplished by plating out large numbers of the irradiated or chemically treated cells on a solid, rich nutrient medium. Colonies (**clones**) are allowed to develop by multiplication of the individual bacteria. The auxotrophs are selected by **replica plating**.[2] In this procedure a sterile velveteen pad is pressed against a nutrient agar plate containing small colonies of bacteria and is used to "print" replica plates containing a minimal medium. The colonies on the initial and replica plates are compared and the colonies of auxotrophs (which do not grow on the minimal medium) are selected. In a second stage, the auxotrophs may be replica plated to minimal medium supplemented with various nutrients (amino acids, purines, pyrimidines, vitamins, etc.). Selection is made easier by pretreatment of the irradiated cells suspended in minimal medium with penicillin. Penicillin kills the growing cells, but the auxotrophs, which do not grow on the minimal medi-

um, survive. The penicillin is then destroyed by adding penicillinase (a β-lactamase; Box 20-G) leaving a suspension much enriched in the percentage of auxotrophic mutants.[2]

A nutritional auxotroph of a bacterium often has a defect in a gene specifying a protein needed for the biosynthesis of the required nutrient. Individual genes recognized in this way are named with a genetic symbol. For example, gene *trpA* specifies one of the two protein subunits of tryptophan synthase. Other kinds of mutations, e.g., those affecting motility or other properties of the cells, can also be detected and are given appropriate symbols. A few of these genetic symbols are indicated on the *E. coli* chromosome map in Fig. 26-4, and many others are used throughout this book. On the basis of such nutritional experiments Beadle, by 1945, had proposed his famous one-gene-one-enzyme hypothesis.[1]

5. Establishing the Correspondence of a Genetic Map and an Amino Acid Sequence

Although the studies of the *rII* region of the T4 chromosome established that genetic mapping could be carried to the level of individual nucleotides in the DNA, it was still necessary to prove a linear correspondence between the nucleotide sequence in the DNA and the amino acid sequence in proteins. This was accomplished by Yanofsky[52,53] and associates through study of the enzyme tryptophan synthase of *E. coli.* Tryptophan synthase (Fig. 25-3) consists of two subunits, α and β, the former containing only 268 amino acids and encoded by the *trpA* gene. To obtain a fine structure map of the A gene, a large series of tryptophan auxotrophs unable to grow in the absence of added tryptophan were isolated. Genetic crosses were carried out with the aid of a **transducing bacteriophage** Plkc. Transducing bacteriophage (Section E,3), while multiplying in susceptible bacteria, sometimes incorporate a portion of the bacterial chromosome into their own DNA. Then, when the virus infects other bacteria, some of the genetic information can be transferred through recombination into the chromosome of bacteria that survive infection. Use of a series of deletion mutants, as in the *rII* mapping, permitted division of the A gene into a series of segments, and observation of recombination frequencies permitted fine structure mapping.

The second part of the proof of colinearity of DNA and protein sequences was the determination of the complete amino acid sequence of tryptophan synthase and peptide mapping (Chapter 3) of fragments of the mutant enzymes. From the peptide maps it was possible to identify altered peptides and to establish the exact nature of the amino acid substitutions present in a variety of different tryptophan auxotrophs. When this was done, it was found that those mutations that mapped very close together had amino acid substitutions at adjacent or nearly adjacent sites in the peptide chain.

The same problem was approached by Sarabhai and associates[54] through the nonsense mutations (Section 6), which lead to premature chain termination during protein synthesis. During late stages of the infection of *E. coli* by phage T4, most of the protein synthesis is that of a single protein of the virus head. Synthesis of protein by infected cells was allowed to proceed in the presence of specific ^{14}C-labeled amino acids. Then cell extracts were digested with trypsin or chymotrypsin, the head-protein peptides were separated by electrophoresis, and autoradiograms were prepared. A series of T4 nonsense mutants that mapped within the head-protein gene were shown to give rise to incomplete head-protein chains. The peptide fragments were of varying lengths. By examining the radioautograms prepared from the enzymatically fragmented peptides, it was possible to arrange the mutants in a sequence based on the length of peptide formed and to show that this was the same as that deduced by genetic mapping. More recently the colinearity of codon and amino acid sequences has been verified repeatedly by comparison of experimentally determined nucleotide sequences in RNA and DNA molecules with the corresponding amino acid sequences for thousands of proteins.

Before the triplet nature of codons had been established, Crick and associates used frame-shift mutations in a clever way to demonstrate that the genetic code did consist of triplets of nucleotides.[7,55,55a] Consider what will happen if two strains of bacteria, each containing a frame-shift mutation (e.g., a −1 deletion), are mated. Genetic recombination can occur to yield mutants containing *both* of the frame-shift mutations. It would be difficult to recognize such recombinants because, according to almost any theory of coding, they would still produce completely defective proteins. However, Crick *et al*. introduced a third frame-shift mutation of the same type into the same gene and observed that the recombinants containing all three deletions (or insertions) were able to synthesize at least partially active proteins. Thus, while introduction of one or two single nucleotide deletions completely inactivates a gene, deletion of three nucleotides close together within a gene shortens the total message by just three nucleotides. The gene will contain only a short region in which the codons are scrambled. The reading frame for the remainder of the protein will not be changed. The protein specified will often be functional because it has a normal sequence except for a small region where some amino acid substitutions will be found and where one amino acid will be completely missing.

6. Conditionally Lethal Mutations

Studies of plaque morphology and of nutritional auxotrophs are directed narrowly at one gene or group of genes. It is desirable to have a general means of detecting mutations in the many other genes present within cells. However, most mutations are **lethal,** and this effect cannot be overcome by adding any nutrient. Lethal mutations are very common in higher organisms, but since eukaryotic cells have pairs of homologous chromosomes, they can be carried in one chromosome and the individual survives. With bacteria and viruses there is only one chromosome, and lethal mutants cannot survive.

Nutritional auxotrophs can be described as **conditionally lethal mutants**; they survive only if the medium is supplemented with the nutrient, whose synthesis depends upon the missing enzyme. Other kinds of conditional lethal mutations permit study of almost every gene in an organism. For example, **temperature-sensitive** (*ts*) mutants grow perfectly well at a low temperature, e.g., 25°C, but do not grow at a higher temperature, e.g., 42°C.[41,56] Many temperature-sensitive mutations involve an amino acid replacement that causes the affected protein to be less stable to heat than is the wild type protein. Others involve a loss in protein-synthesizing ability for reasons that may be obscure. Temperature-sensitive mutations occur spontaneously in nature, an example being the gene that controls hair pigment in Siamese cats.[41] The gene (or gene product) is inactivated at body temperatures but is active in the cooler parts of the body, such as the paws, tail, and nose, with the result that the cat's hair is highly pigmented only in those regions.

Screening for conditionally lethal temperature-sensitive mutants of bacteriophage T4 permitted isolation of hundreds of mutants involving sites at random over the entire viral chromosome. Complementation studies permitted assignment of these to individual genes, which at first were identified only by number (Fig. 26-2). Later specific functions were associated with the genes.[57,58] For example, the product of gene 42 was identified as an enzyme required in the synthesis of hydroxymethyl-dCMP (Chapter 5). Genes 20–24, among others, must code for head proteins because mutants produce normal tails but no heads. Gene 23 codes for the major head subunits, while gene 20 has something to do with "capping" the end of the head. These mutants produce cylindrical "polyheads" in place of the normal heads. Mutants of genes 25–29 have defective base plates and do not form tails, while mutants 34–38 lack tail fibers. The specific ways in which some of these gene products are assembled to form the base plates and tails of the phage are indicated in Fig. 7-29. The positions of the *rIIA* and *rIIB* genes are also shown in Fig. 26-2.

A second type of conditionally lethal mutation leads to alteration of an amino acid codon to one of the three **chain termination codons** UAG, UAA, and UGA (Table 5-3).[60,61] These are often called **nonsense mutations** in contrast to **missense mutations** in which one amino acid is replaced by another. A chain termination mutant synthesizes only part of the product of the defective gene because of the presence of the termination codon. A remarkable aspect of chain termination mutations is that they can be **suppressed** by other mutations in distant parts of the virus or bacterial chromosome. Many otherwise lethal mutations of bacteriophage T4 were discovered by their ability to grow in mutant strains of *E. coli*, which contained **suppressor genes**,[62] and their inability to grow in the normal B strain. Three different suppressor genes *supD*, *supE*, and *supF* were found to suppress mutations that formed UAG. These are commonly known as *amber* suppressor genes. A second group of *ochre* suppressor genes, including *supB* and *supC*, suppressed mutations that formed codon UAA. Suppressors for mutations that form codon UGA have also been found.[63] Like the temperature-sensitive mutants, *amber* and *ochre* mutants can be obtained in almost any genes of a bacterial virus. Chain termination mutants of unessential genes in bacteria can be recognized by transferring the genes by conjugation or by viral transduction into a strain (sup^+) that contains a desired suppressor gene.

Conditionally lethal mutants have been of great value in developing our understanding of the genetics of bacterial viruses. They have also provided a powerful technique for approaching complex problems of bacterial physiology. For example, we may ask how many genes are required for a bacterium to sense the presence of a food and to swim toward it (Chapter 19). Even though few clues as to the basic chemistry underlying these phenomena can be obtained in this way, the use of temperature-sensitive mutants and complementation tests permits us to establish the total number of genes involved in these complex processes and to map their positions on a bacterial or viral chromosome. This is often an important step toward a more complete understanding of a biological phenomenon.

7. The Nature of Suppressor Genes

How can one mutation be suppressed by a second mutation at a different point in the chromosome? Rarely, a mutation is suppressed by a second mutation within the *same* gene. Such **intragenic complementation** sometimes occurs when a mutation leads to an amino acid replacement that disrupts the structural stability or function of a protein. Sometimes a mutation at another site involving a residue, which interacts with the first amino acid replaced, will alter the inter-

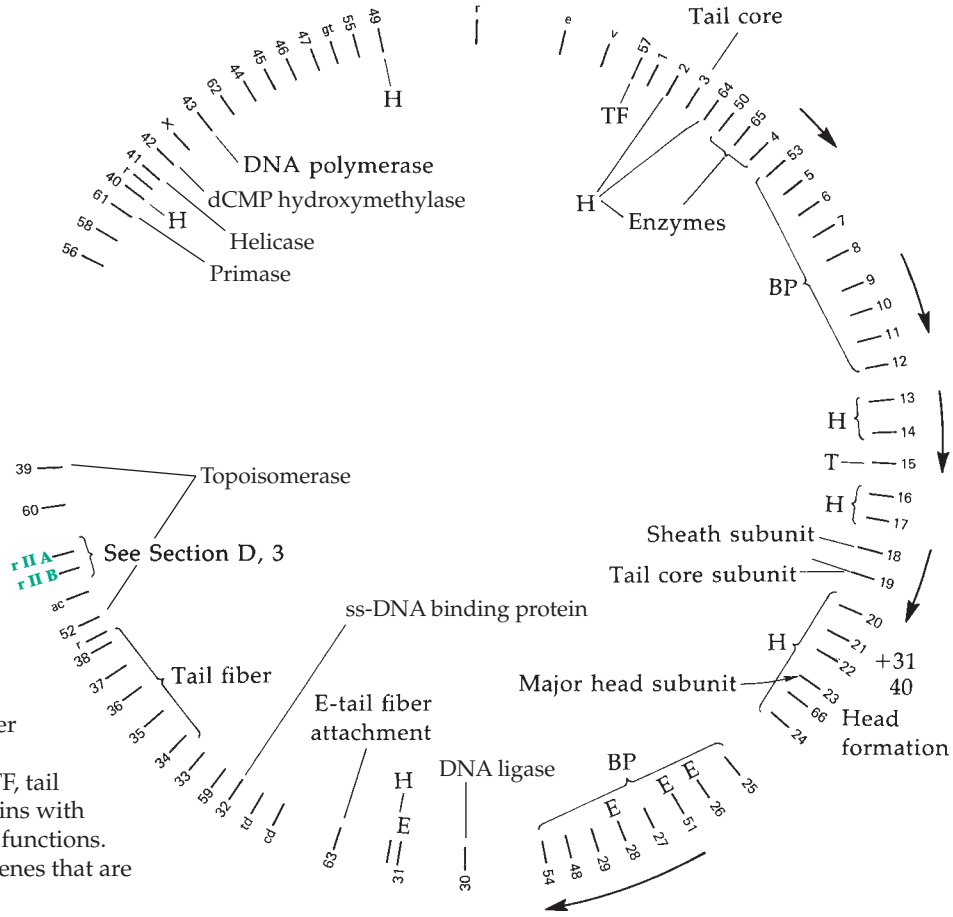


Figure 26-2 Linkage map for bacteriophage T4. See H. A. Sober and Eiserling and Dickson.[57,59] Abbreviations: H, head; T, tail; TF, tail fiber; BP, base plate; and E, proteins with enzymatic rather than structural functions. The bold arrows indicate some genes that are transcribed together (operons).

action between the two residues in a way that restores function to the protein. For example, if the first amino acid side chain is small and is replaced by mutation with a larger side chain, a second mutation leading to a decrease in the size of another side chain may permit the protein to fold and function properly. An example was found among mutants of tryptophan synthase.[64] Mutants in which Gly 211 of the α chain was replaced by Glu or Tyr 175 of the same chain by Cys both produced inactive enzymes. However, the double mutant with both replacements synthesized active tryptophan synthase. It is known now that these residues are adjacent to one another and form part of the binding site for the substrate indole-3-glycerol phosphate (Figs. 25-2; 25-3). Only the double mutant permits the substituted side chains to pack properly. In other cases, intragenic suppression involves changes in the subunit interactions in oligomeric proteins. These changes may affect the formation of correct quaternary structures of the proteins.

As discussed in the preceding section, the best known suppressor genes are those that suppress chain termination mutations (Section 6). These genes often encode mutant forms of tRNA molecules, which allow incorporation of an amino acid rather than chain termination to occur. They are discussed further in Chapter 29. Suppressor genes are not limited to bacteria. For example, the vermilion eye color mutation of *Drosophila* leads to a loss of brown eye pigments because of the inactivity of tryptophan 2,3-dioxygenase (Eq. 18-38). However, synthesis of the tryptophan dioxygenase from the *vermilion* mutant is inhibited by $tRNA_2^{Trp}$, one of the two tryptophanyl tRNAs. The suppressor mutation alters the tRNA in such a way that the inhibition is relieved.[61]

C. Plasmids, Episomes, and Viruses

The small pieces of DNA known as plasmids, which replicate independently of the chromosomes, have been discussed briefly in Chapter 5. Plasmids share a number of properties with viruses, and both are important to the techniques of contemporary molecular biology and genetic engineering. Bacterial plasmids may be present as one or several copies for each chromosome. Episomes are plasmids that are able to become integrated into the bacterial chromosome. Some extrachromosomal elements are episomes in one host and plasmids in another. Bacterial

plasmids may be infectious (transferable) or noninfectious. In the former case, they are able to transfer their DNA into another cell and are known as **sex factors** (F agents or fertility factors). A sex factor is able to integrate into a chromosome and later to come out and transfer other genes with it. In this property it resembles a transducing phage.

Plasmids and episomes vary in size. The F sex factor is a 100-kb circular molecule of supercoiled DNA. Colicinogenic factors,[65] which may also be present in *E. coli* in as many as 10–15 copies per bacterial chromosome, are often much smaller (6–7.5 kb). Some larger colicinogenic plasmids are also sex factors. They carry genes for toxic protein antibiotics known as **colicins** (Box 8-D) which attack other strains of *E. coli*, providing a selective advantage for the strain producing the colicin. They also carry a gene or genes conferring on the host bacterium resistance to antibiotics such as penicillin and chloramphenicol. Penicillin is inactivated because the plasmids carry a gene encoding a penicillinase that hydrolytically cleaves the β-lactam ring (Box 20-G). Chloramphenicol (Fig. 25-10) is inactivated by the action of chloramphenicol *O*-acyltransferase.

1. Bacterial Sex Factors

Bacteria usually reproduce by a simple cell division. The DNA in the chromosome is doubled in quantity and the cell divides, each daughter cell receiving an identical chromosome. However, in 1946 Lederberg and Tatum showed that sexual reproduction is also possible.[66,67] They studied nutritional auxotrophs of *E. coli* strain K-12, which lacked the ability to synthesize amino acids or vitamins. When cells of two different mutants were mixed together and allowed to grow for a few generations, a few individual bacteria regained the ability to grow on a minimal medium. Since each of the two strains had one defective gene, the creation of an individual with neither of the two defects required combining of genetic traits from both strains. The existence of bacterial conjugation was recognized. Later it was established that true **genetic recombination** had occurred, i.e., genes from the two mating cells had been integrated into a single molecule of bacterial DNA.

This transfer of DNA between bacterial cells requires the presence of a plasmid sex factor (F agent), whose presence confers a male character to the individual cell. The F agent is large enough to contain about 90 genes and has a length of ~30 μm, ~2.5% that of the *E. coli* chromosome. Among other things, the *E. coli* F agent contains the genes needed to direct the synthesis of the **F pili** (sex pili). These tiny appendages, 8.5 nm in diameter (see Fig. 7-9), grow out quickly during a period of 4–5 min to a length of about 1.1 μm. The end of an F pilus becomes attached to a female cell (a bacterium lacking the F agent) and may induce the transfer of DNA into the female cell. The mechanism of transfer has been uncertain. It may involve formation of a cytoplasmic bridge between cells in close contact. The pilus may be retracted into the membrane of the male cell, pulling the two cells close together. The DNA probably flows through the pilus into the female.[68,69]

On rare occasions an F agent becomes integrated into the chromosome of a bacterium. Both the F agent and the chromosome have been shown by electron microscopy to be circular. The integration process requires the enzymatic cleavage of the DNA of both the chromosome and the F agent and the rejoining of the ends in such a way that a continuous circle is formed (Fig. 26-3). The enzymes that catalyze these reactions are considered in Chapter 27. Different F agents can be incorporated into the chromosome at different points around the circle. A strain of bacteria containing an integrated F agent is known as an *Hfr* (high frequency of recombination) strain.

When an Hfr strain conjugates with an F^- (female), replication of the entire male chromosome commences at some point near the end of the integrated F agent, and genes of the bacterial chromosome followed by those of the F factor are transferred into the female. Only a single strand of DNA (customarily referred to as the **plus strand**) is transferred from the donor cell and into the recipient cell (Fig. 26-3). There the complementary **minus strand** is synthesized to form a complete double-stranded DNA molecule bearing the genes from the Hfr cell. Only rarely does a copy of the entire chromosome of the donor cell enter the female cell. More often the DNA strand, or perhaps the pilus itself, breaks and only part of the chromosome is transferred.

Partial chromosome transfer from a male cell transforms the F^- cell into a partial diploid (**merozygote**) containing double the usual number of some of the genes. Within this partial diploid genetic recombination between the two chromosomes takes place (Fig. 26-3) by the mechanisms discussed in Chapter 27. The end result of the recombination process is that the daughter cells formed by subsequent division contain only single chromosomes with the usual number of genes. However, some genes come from each of the two parental strains. Thus, an F^- mutant unable to grow on a medium deficient in a certain nutrient may receive a gene from the male and now be able to grow on a minimal medium. Even though the number of such recombinants is small, they are easily selected from the very large number of mutant bacteria that are mixed together initially.

One result of DNA transfer from Hfr into F^- bacteria is sometimes the introduction of a complete copy of the F agent into the female bacterium. Since this con-

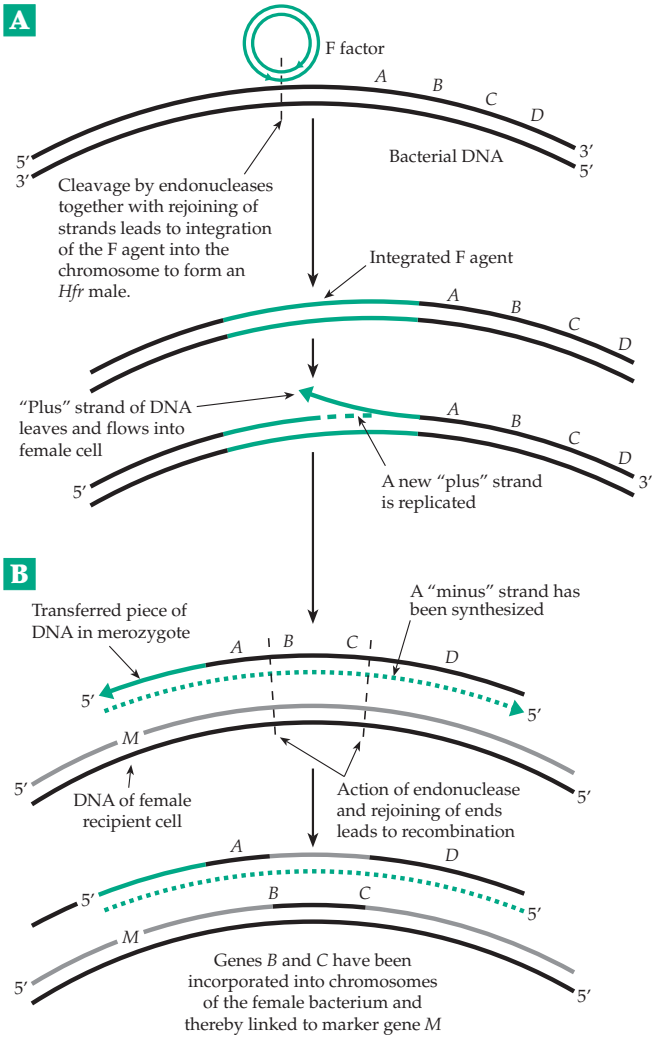


Figure 26-3 Integration of an F agent into a bacterial chromosome and transfer of some bacterial genes into another cell. (A) Incorporation of the F agent into *E. coli* genome and transfer of the "plus" strand of DNA out to a female recipient cell. (B) Genetic recombination between a piece of transferred DNA and the genome of the recipient cell.

verts the recipient into a male, Brinton referred to "bacterial sex as a virus disease." In fact, infectious plasmids and viruses display many similarities. For example, filamentous bacteriophages (Chapter 5; Fig. 7-7) adsorb to the F pili of male bacteria and the DNA, flowing in a direction opposite to that in bacterial conjugation, enters the cell.[70] The bacteriophage carry genes for the protein subunits of their coats (Figs. 26-2, 7-29), while F factors carry genes for synthesis of pilins. Pilins accumulate within the cell membrane and are extruded to generate F pili, just as viral subunits are extruded to form the virus coats. There is also a close similarity between episomes that can be integrated into bacterial chromosomes and the **temperate bacteriophages** considered in the next section.

2. Temperate Bacteriophage; Phage Lambda

When DNA from a typical bacteriophage enters a bacterial cell, it seizes control of the metabolic machinery of the cell almost immediately and directs it entirely toward the production of new virus particles. This leads within a period of about 20 min to the production of one or two hundred progeny viruses and to the lysis and death of the cell. However, the DNA from a temperate phage may become repressed and, like an F factor, be integrated with the bacterial genome (Fig. 26-3). In the resulting **prophage** or **lysogenic** state, the repressed phage DNA is replicated as part of the bacterial genome but does no harm to the host cells unless some factor "activates" the incorporated genetic material by release of the repression. Replication of the phage and lysis of the bacterium then ensues. Temperate phage may also exist as plasmids (e.g., plasmid P1).

The decision for lysis or lysogeny, which is very important for the survival of the bacteriophage, is governed primarily by the nutritional status of the host. For a bacterium growing in a relatively rich environment such as the colon, lysis will increase the chances of daughter phage encountering host bacteria. However, in soils, *E. coli* grows very slowly, and a bacteriophage capable of entering a lysogenic state has an increased chance of survival until the host bacterium finds a richer growth medium.

The best known temperate phage is **phage lambda** of *E. coli*.[70–72] A tailed virus resembling the T-even phages (Box 7-C), phage λ has a smaller (~48.5 kb) DNA genome.[73] Within the bacterial cell the ends of the λ DNA may be joined to form a circular replicative form of the virus. In ~30% of the infected cells the λ DNA becomes integrated into the *E. coli* chromosome at the special site, *att λ*, which is located at 17 min on the *E. coli* chromosome map (Fig. 26-4). The incorporated phage DNA now occupies a linear segment amounting to about 1.2% of the total length of the *E. coli* chromosome. It is replicated along with the rest of the chromosome and for, the most part, goes unnoticed.

The host, *E. coli* K12, contains useful *amber* suppressors that make it easy to detect mutations in the bacteriophage. The integrated prophage can undergo mutations of almost any type, including large deletion mutations, and can still be investigated through complementation studies with other strains of virus. Thus, a family of modified **defective λ phage** was developed. When the λ prophage is excised from the bacterial chromosome, adjacent bacterial genes are occasionally carried with it. This allowed development of **λ transducing phage,** which can carry genes and transfer them into bacteria lacking these genes. More recently an important series of cloning vehicles have been derived from phage λ.

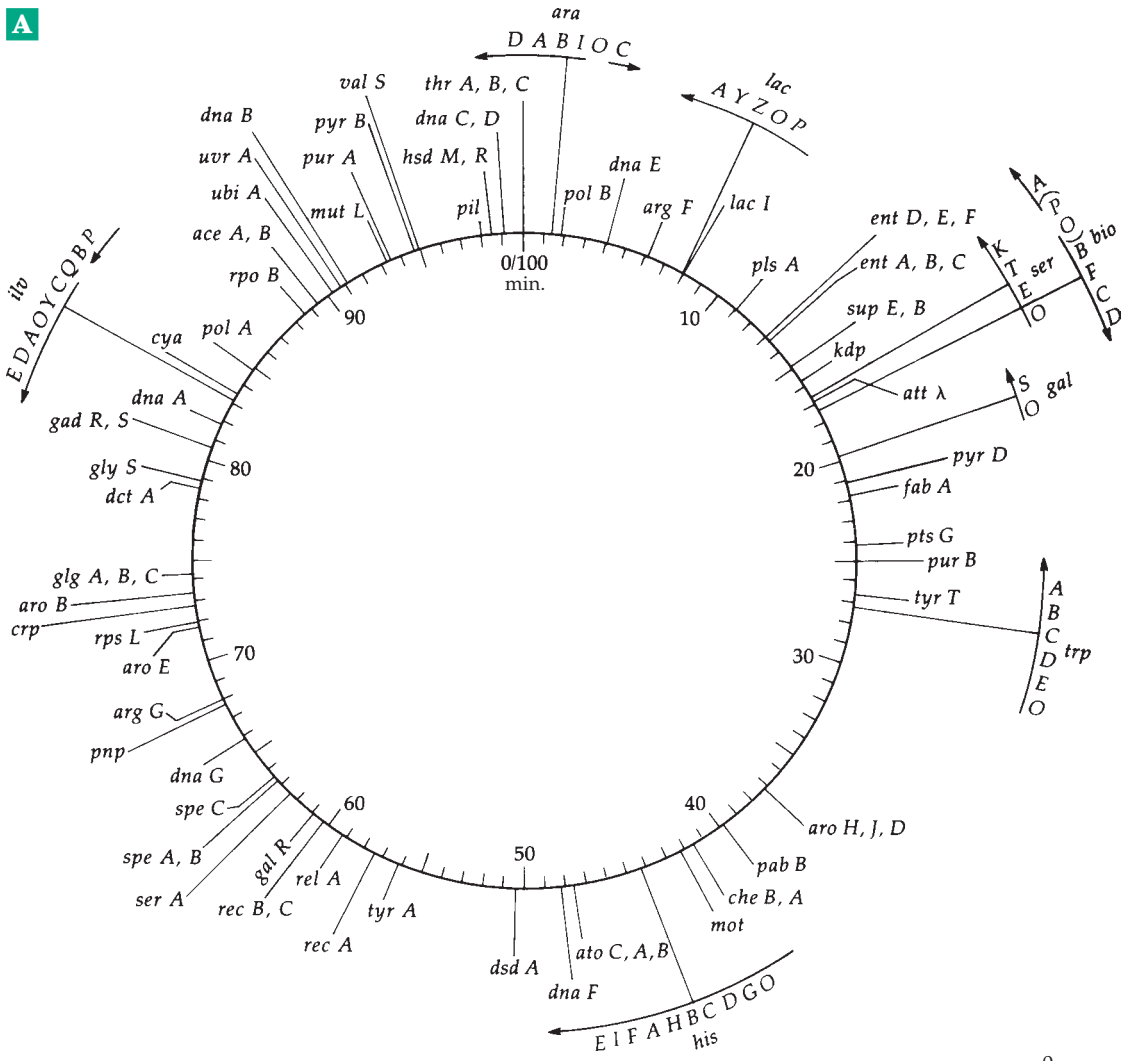


Figure 26-4 (A) Genetic map of *E. coli*, strain K-12 published in 1976. Data from Bachmann, Low, and Taylor.[75] The scale in minutes is based on the results of interrupted conjugation experiments. The *thr* locus is arbitrarily taken as zero. The genetic symbols are defined in Table 26-1. Only a single strand of the DNA molecule is normally transcribed into RNA over any given region. Because of the antiparallel nature of the DNA helix, transcription from one strand proceeds clockwise and transcription from the other counterclockwise (the *r* strand is conventionally taken as the one that is transcribed in the clockwise direction when the chromosome is oriented as in the drawing). The directions of transcription of certain operons are indicated by the arrows. The 1983 version of this map[76] contained 1027 loci and we now know the location of all of the more than 4300 genes. (B) Physical map of the *E. coli* chromosome obtained by study of large 20–100 kb fragments obtained by digestion with restriction endonuclease *Not*I and separated by pulsed field gel electrophoresis. The fragments A–V were ordered using genetic information and overlapping fragments from partial *Not*I digests (arrows). On the outside the genetic map has been superimposed after distortion of the scale in minutes to make the two maps coincident. The 100-kb F^+ plasmid and the 4.5-kb plasmid pBR322 are also drawn to show relative sizes. From Smith *et al.*[77] A finer restriction map, based on 3400 cloned fragments, was prepared by Kohara, Akiyama, and Isono in 1987.[78]

TABLE 26-1
Some Genes of *E. coli*[a]

Gene symbol	Mnemonic	Map position (min)[a]	Phenotypic trait affected
aceA	Acetate	89	Isocitrate lyase
aceB	Acetate	89	Malate synthetase A
araA	Arabinose	1	L-Arabinose isomerase
araB	Arabinose	1	L-Ribulokinase
araC	Arabinose	1	Regulatory gene
araD	Arabinose	1	L-Ribulose-5-phosphate-4-epimerase
araI	Arabinose	1	Initiator locus
araO	Arabinose	1	Operator locus
argF	Arginine	6	Ornithine carbamoyltransferase
argG	Arginine	68	Argininosuccinic acid synthetase
aroB	Aromatic	73	Dehydroquinate synthetase
aroD	Aromatic	37	Dehydroquinate dehydratase
aroE	Aromatic	71	Dehydroshikimate reductase
aroH	Aromatic	37	DAHP synthetase (tryptophan-repressible isoenzyme)
aroJ	Aromatic	37	Probable operator locus for *aro*H
atoA	Acetoacetate	48	Coenzyme A transferase
atoB	Acetoacetate	48	Thiolase II
atoC	Acetoacetate	48	Regulatory gene
*att*λ	Attachment	17	Integration site for prophage λ
bioA	Biotin	17	Group II; 7-oxo-8-aminopelargonic acid (7 KAP) → 7,8-diaminopelargonic acid (DAPA)
bioB	Biotin	17	Conversion of dethiobiotin to biotin
bioC	Biotin	17	Block prior to pimeloyl-CoA
bioD	Biotin	17	Dethiobiotin synthetase
bioF	Biotin	17	Pimeloyl-CoA → 7 KAP
bioO	Biotin	17	Operator for genes *bio*B through *bio*D
bioP	Biotin	17	Promoter site for genes *bio*B through *bio*D
cheA	Chemotaxis	42	Chemotactic motility
cheB	Chemotaxis	42	Chemotactic motility
crp		73	Cyclic adenosine monophosphate receptor protein
cya		83	Adenylate cyclase
dctA		79	Uptake of C_4-dicarboxylic acids
dnaA	DNA	82	DNA synthesis; initiation defective
dnaB	DNA	91	DNA synthesis
dnaC	DNA	99	*dnaD*; DNA synthesis; initiation defective
dnaE	DNA	4	*polC*, DNA polymerase III and mutator activity
dnaF	DNA	48	*nrdA*; ribonucleoside diphosphate reductase
dnaG	DNA	66	DNA synthesis
dsdA	D-Serine	50	D-Serine deaminase
entA	Enterochelin	13	2,3-Dihydro-2,3-dihydroxybenzoate dehydrogenase
entB	Enterochelin	13	2,3-Dihydro-2,3-dihydroxybenzoate synthetase
entC	Enterochelin	13	Isochorismate synthetase
entD,E,F	Enterochelin	13	Unknown steps in conversion of 2,3-dihydroxybenzoate to enterochelin
fabA		22	β-Hydroxydecanoylthioester dehydratase
gadR		81	Regulatory gene for *gadS*
gadS		81	Glutamic acid decarboxylase
galE	Galactose	17	Uridine diphosphogalactose 4-epimerase
galK	Galactose	17	Galactokinase
galO	Galactose	17	Operator locus
galT	Galactose	17	Galactose 1-phosphate uridyltransferase
galR	Galactose	61	Regulatory gene
glgA	Glycogen	74	Glycogen synthetase
glgB	Glycogen	74	α-1,4-Glucan: α-1,4-glucan 6-glucosyltransferase
glgC	Glycogen	74	Adenosine diphosphate glucose pyrophosphorylase
glyS	Glycine	79	Glycyl-transfer RNA synthetase
hisA	Histidine	44	Isomerase
hisB	Histidine	44	Imidazole glycerol phosphate dehydrase: histidinol phosphatase
hisC	Histidine	44	Imidazole acetol phosphate aminotransferase
hisD	Histidine	44	Histidinol dehydrogenase
hisE	Histidine	44	Phosphoribosyl-adenosine triphosphate-pyrophospho-hydrolase
hisF	Histidine	44	Cyclase
hisG	Histidine	44	Phosphoribosyl-adenosine triphosphate-pyro-phosphorylase
hisH	Histidine	44	Amidotransferase
hisI	Histidine	44	Phosphoribosyl-adenosine monophosphate-hydrolase
hisO	Histidine	44	Operator locus
dsdM	Host specificity	98	Host modification activity: DNA methylase M
hsdR	Host specificity	98	Host restriction activity: endonuclease R
ilvA	Isoleucine—valine	83	Threonine deaminase (dehydratase)
ilvB	Isoleucine—valine	83	Acetohydroxy acid synthetase I
ilvC	Isoleucine—valine	83	α-Hydroxy-β-oxo acid reductoisomerase
ilvD	Isoleucine—valine	83	Dehydrase
ilvE	Isoleucine—valine	83	Aminotransferase B
ilvO	Isoleucine—valine	83	Operator locus for genes *ilv*A,D,E
ilvP	Isoleucine—valine	83	Operator locus for gene *ilv*B
ilvQ	Isoleucine—valine	83	Induction recognition site for *ilv*C
ilvY	Isoleucine—valine	83	Positive control element for *ilv*C induction
kdp	K accumulation	16	Defect in potassium ion uptake
lacA	Lactose	8	Thiogalactoside transacetylase
lacI	Lactose	8	Regulator gene
lacO	Lactose	8	Operator locus
lacP	Lactose	8	Promoter locus
lacY	Lactose	8	Galactoside permease (M protein)
lacZ	Lactose	8	β-Galactosidase
mot	Motility	42	Flagellar paralysis
mutL	Mutator	93	Generalized high mutability (AT → GC)
pabB	*p*-Aminobenzoate	40	Requirement
pil	Pili	98	Presence or absence of pili (fimbriae)
plsA	Phospholipid	11	Glycerol-3-phosphate acyltransferase
pnp		68	Polynucleotide phosphorylase
polA	Polymerase	85	DNA polymerase I
polB	Polymerase	2	DNA polymerase II
ptsG	Phosphotranserase	24	Catabolite repression system
purA	Purine	93	Adenylosuccinic acid synthetase
purB	Purine	25	Adenylosuccinase
pyrB	Pyrimidine	95	Aspartate carbamoyltransferase
pyrD	Pyrimidine	21	Dihydroorotic acid dehydrogenase
recA	Recombination	58	Ultraviolet sensitivity and competence for genetic recombination
recB	Recombination	60	Ultraviolet sensitivity, genetic recombination; exonuclease V subunit
recC	Recombination	60	Ultraviolet sensitivity, genetic recombination; exonuclease V subunit
relA	Relaxed	59	Regulation of RNA synthesis
rpoB	RNA polymerase	89	RNA polymerase: β subunit (*rif* gene)
rpsL	Ribosomal protein, small	72	Ribosomal protein S12 (*strA* gene, streptomycin resistance)
serA	Serine	62	3-Phosphoglyceric acid dehydrogenase
serO	Serine	20	Operator locus
serS	Serine	20	Seryl transfer RNA synthetase
speA	Spermidine	63	Arginine decarboxylase
speB	Spermidine	63	Agmatine ureohydrolase
speC	Spermidine	63	Ornithine decarboxylase
supB	Suppressor	15	Suppressor of *ochre* mutations
supE	Suppressor	15	Suppressor of *amber* mutations (*su-2*)
thrA	Threonine	0	Aspartokinase I-homoserine dehydrogenase I complex
thrB	Threonine	0	Homoserine kinase
thrC	Threonine	0	Threonine synthetase
trpA	Tryptophan	27	Tryptophan synthetase, A protein
trpB	Tryptophan	27	Tryptophan synthetase, B protein
trpC	Tryptophan	27	*N*-(5-Phosphoribosyl) anthranilate
trpD	Tryptophan	27	Phosphoribosyl anthranilatetransferase
trpE	Tryptophan	27	Anthranilate synthetase
trpO	Tryptophan	27	Operator locus
tyrA	Tyrosine	56	Chorismate mutase T-prephenate dehydrogenase
tyrT	Tyrosine	27	Tyrosine transfer RNA_1 (*su-3* gene; amber suppressor)
ubiA	Ubiquinone	90	4-Hydroxybenzoate → 3-octaprenyl 4-hydroxybenzoate
uvrA	Ultraviolet	91	Repair of ultraviolet radiation damage to DNA, UV endo-nuclease
valS	Valine	95	Valyl-transfer RNA synthetase

[a] This list contains 126 of more than 1027 genes that had been mapped by 1983. (Bachmann, B. J. (1983) *Bacteriol. Rev.* **47**, 180—230). Their positions are shown diagrammatically in Fig. 26-4.

D. Mapping of Chromosomes

Let us now consider how knowledge of bacterial sex factors and of phage λ permitted the mapping of bacterial chromosomes. Together with the use of restriction endonucleases these techniques gave us the first precise physical maps of bacterial chromosomes and pointed the way toward the determination of complete genome sequences.

1. The Chromosome Map of *E. coli*

There are about 4,639,221 nucleotide pairs in the circular DNA molecule that is the chromosome of *E. coli* strain K-12. We now know the complete sequence, which includes all of the individual genes that are present.[74] However, our first knowledge of the location of these genes in the chromosome depended upon construction of a **linkage map** (Fig. 26-4A). Construction of this map, with 126 genes, began with the study of nutritional auxotrophs whose defective genes are located at many points on the chromosome. By 1983, 1027 genes had been mapped. In 1997, when the complete nucleotide sequence became known, 4288 protein coding genes could be recognized.[74] The map in Fig. 26-4 was established 30 years earlier, largely by use of interrupted bacterial mating.[79,80] In this procedure *Hfr* cells carrying specific mutations are mixed with wild-type F cells, and conjugation is allowed to proceed for a certain length of time. Then the cells are agitated violently, e.g., in a Waring blender. This breaks all of the conjugation bridges and interrupts the mating process. Mating is interrupted at different times, and the recipient bacteria are tested for the presence of genes transferred from the donor strain. Using this technique it was found that complete transfer of the chromosome takes ~100 min at 37°C, and that the approximate location of any gene on the chromosome can be determined by the length of time required for transfer of that gene into the recipient cell. It is a little more complex than this. Because complete chromosome transfer is rare, substrains of *E. coli* K-12 with an F agent integrated at different points were used. With certain F factors those genes lying clockwise around the circle in Fig. 26-4 immediately beyond the point of integration are transferred quickly and with high frequency.

The **time-of-entry map** in Fig. 26-4A is based not only on interrupted matings but also on the use of **transduction** by bacteriophage P1.[76,79] Transduction by phage permits the tranfer of a short fragment of DNA, about 2 min in length, on the *E. coli* map. Joint transduction, i.e., joint incorporation of two genes into the chromosome of the receptor, occurs with a frequency related to the map distance between these two genes. Thus, finer mapping was done within many segments of the *E. coli* chromosome. Meanings of the gene symbols used in the figure are given in Table 26-1. Similar maps were prepared for *Salmonella typhimurium* and *Bacillus subtilis*.

2. Restriction Endonucleases

Many of the procedures for cloning genes, synthesizing more copies of a DNA ("amplifying" the DNA), making genetic maps, and generating mutants, are dependent upon **restriction endonucleases**. The name comes from a property of bacteria, which often can digest and destroy DNA of invading viruses or DNA that has been injected during mating with a bacterium of an incompatible strain. Investigation of this phenomenon, known as **restriction**, revealed that the DNA of viruses that are able to replicate within a particular host is *marked* in some fashion at specific sites in the molecule. The marking often consists of the presence of methyl groups. Properly methylated DNA is not degraded, but unmethylated DNA is cleaved by a highly specific endonuclease at the same sites that are normally methylated. Each species of bacteria (and often an individual strain within a species) has its own restriction enzymes. Restriction enzymes are very specific and cut DNA chains at unique base sequences. Three types are recognized.[80a]

Type I restriction enzymes, such as those encoded in the chromosome of *E. coli*, are large 300- to 400-kDa proteins composed of at least three kinds of polypeptide chain. They bind at specific sites of a foreign DNA and apparently cleave the chain randomly nearby. They require ATP, Mg^{2+}, and *S*-adenosylmethionine and have the unusual property of promoting the hydrolysis of large amounts of ATP.[81,82] The significance of these properties is still unknown.

The type II restriction endonucleases, which are the ones most widely used in molecular biology, are relatively small 50- to 100-kDa monomeric or dimeric proteins. About 2400 different enzymes with 188 different specificities had been isolated by 1995.[83,84] The sites of attack, in most instances, are nucleotide sequences with a twofold axis of local symmetry.[85] For example, the following sites of cleavage have been identified for two restriction endonucleases encoded by the DNA of R-factor plasmids of *E. coli* and for a restriction enzyme from *Hemophilus influenzae*. In the diagrams ↓ are sites of cleavage, ∗ are sites of methylation, and • are local twofold axes (centers of palindromes); N can be any nucleotide with a proper base pairing partner.

Restriction enzymes often create breaks in each of the two strands in positions symmetrically arranged around the local twofold axis. This is what we might expect of a dimeric enzyme that binds in the major or minor groove of the double helix, each active site

E. coli R factor (*Eco*RI)

```
      ↓  *
5′–NNNGAA T TCNNN
3′–NNNCTT•AAGNNN
         *  ↑
```

E. coli R factor (*Eco*RII)

```
      ↓  *
5′–NNNNCCAGGNNNN–3′
3′–NNNNGGT•CCNNNN–5′
         *  ↑
```

H. influenzae (*Hin*dIII)

```
     *↓
5′–NNNAAGCTTNNN–3′
3′–NNNTTC•GAANNN–5′
         ↑*
```

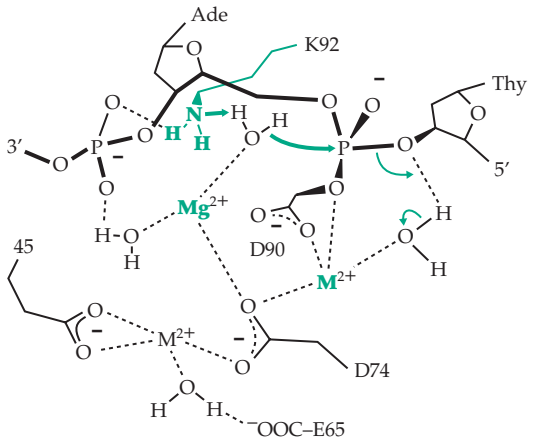
attacking one of the polynucleotide chains. In fact, the two 277-residue subunits of the *Eco*RI enzyme[86] bind primarily in the major grooves of the DNA, one active site on each strand. Each recognition unit makes 12 hydrogen bonds to the DNA. Each base pair forms two of these hydrogen bonds. Four arginines and two glutamates participate.[87] This provides a net charge of +2, which may be important for electrostatic interaction of the protein with backbone phosphate groups of the DNA. The binding of the protein affects the conformation of the DNA, widening the major groove from that in B DNA, and causing a torsional kink with some unwinding of the double helix.[88] It appears that the specificity for the GAATTC hexanucleotide is in part a result of direct complementary interactions between functional groups in the major groove (see Fig. 5-3), bound water molecules, and amino acid side chains from the enzyme (Fig. 26-5). Methylation of the 6-amino groups of the adenines in the center of the recognition sequence prevents cleavage by the *Eco*RI endonuclease, but appears to alter the interaction with the protein only slightly.[89]

Although they often share little sequence similarity and have quite different specificiities, many restriction enzymes have similar three-dimensional structures as well as mechanisms of action. This is true for the *Eco*RI, *Bam*HI (Fig. 26-5),[83,90] *Eco*RV,[91,91a] and *Cfr*10I enzymes,[84] and presumably many others. The specifically shaped and tightly packed active sites in the enzyme–substrate complexes ensure specificity. For example, the *Eco*RV endonuclease cleaves DNA at its recognition site at least a million times faster than at any other DNA sequence.[91] As mentioned in Chapter 12, restriction endonucleases require a metal ion, preferably Mg^{2+}, and probably act via a hydroxyl ion generated from Mg^{2+}–OH_2 at the active site. Three conserved active site residues, Asp 91, Glu 111, and Lys 113, in the *Eco*RI endonuclease interact with the DNA near the cleavage site. Lys 113 is replaced by Glu 113 in the *Bam*HI enzyme.[83,90]

The corresponding conserved residues in the smaller *Eco*RV enzyme are Asp 74, Asp 90, and Lys 92. They are shown in the following diagram that represents one of several possible metal-ion dependent mechanisms.[91–93] The metal-coordinated hydroxyl ion is generated by proton transfer to the $–NH_2$ group of Lys 92 and carries out an in-line attack on the backbone phospho group of thymidine at the cleavage point. At least two metal ions are needed, and three may be present, as shown in this diagram from Sam and Perona.[93]

TABLE 26-2
Some Commonly Used Restriction Endonucleases, Their Sources, and Cleavage Sites

Enzyme	Source	Cleavage Site
*Alu*I	*Arthrobacter luteus*	5' - - AG↓CT - - 3'
*Bam*HI	*Bacillus amyloliquefaciens H*	G↓GATCC
*Bcl*I	*Bacillus caldolyticus*	T↓GATCA
*Bgl*II	*Bacillus globigii*	A↓GATCT
*Cfr*10I	*Citrobacter freundii*	Pu↓CCGGPy
*Eco*RI	*Escherichia coli*	G↓AATTC
*Eco*RV	*Escherichia coli*	GAT↓ATC
*Hae*III	*Haemophilus aegypticus*	GG↓CC
*Hin*dIII	*Haemophilus influenzae*	A↓AGCTT
*Kpn*I	*Klebsiella pneumoniae*	GGTAC↓C
*Mbo*I	*Moraxella bovis*	↓GATC
*Pst*I	*Providencia stuartii*	CTGCA↓G
*Sal*I	*Streptomyces albus*	G↓TCGAC
*Sau*3AI	*Streptococcus aureus*	↓GATC
*Sfi*I	*Streptomyces fimbriatus*	GGCCNNNN↓NGGC



Restriction enzymes that cleave DNA at a large number of specific sequences are available commercially. A few are listed in Table 26-2. Another group of restriction enzymes have similar recognition sequences but cut the dsDNA that they recognize at a specific neighboring site rather than within the recognition sequence. An example is *Fok*I, which recognizes the nonpalindromic $\frac{\text{GGATG}}{\text{CCTAC}}$, but cuts the chains 9 and 13 base pairs to the right. This enzyme has been used by Szybalski and associates to devise a system for cutting ssDNA precisely at a desired point and converting it to ds fragments.[95]

3. Restriction Mapping

The calibration of the *E. coli* genetic map in minutes was a temporary expedient. It was followed by **physical maps** expressed directly as micrometers of DNA length (total length ~1.6 mm) or thousands of nucleotide units (kb). A physical map obtained by **restriction enzyme mapping** is shown in Fig. 26-4B. To obtain this map DNA fragments were prepared using specific restriction endonucleases (Section E, 1).

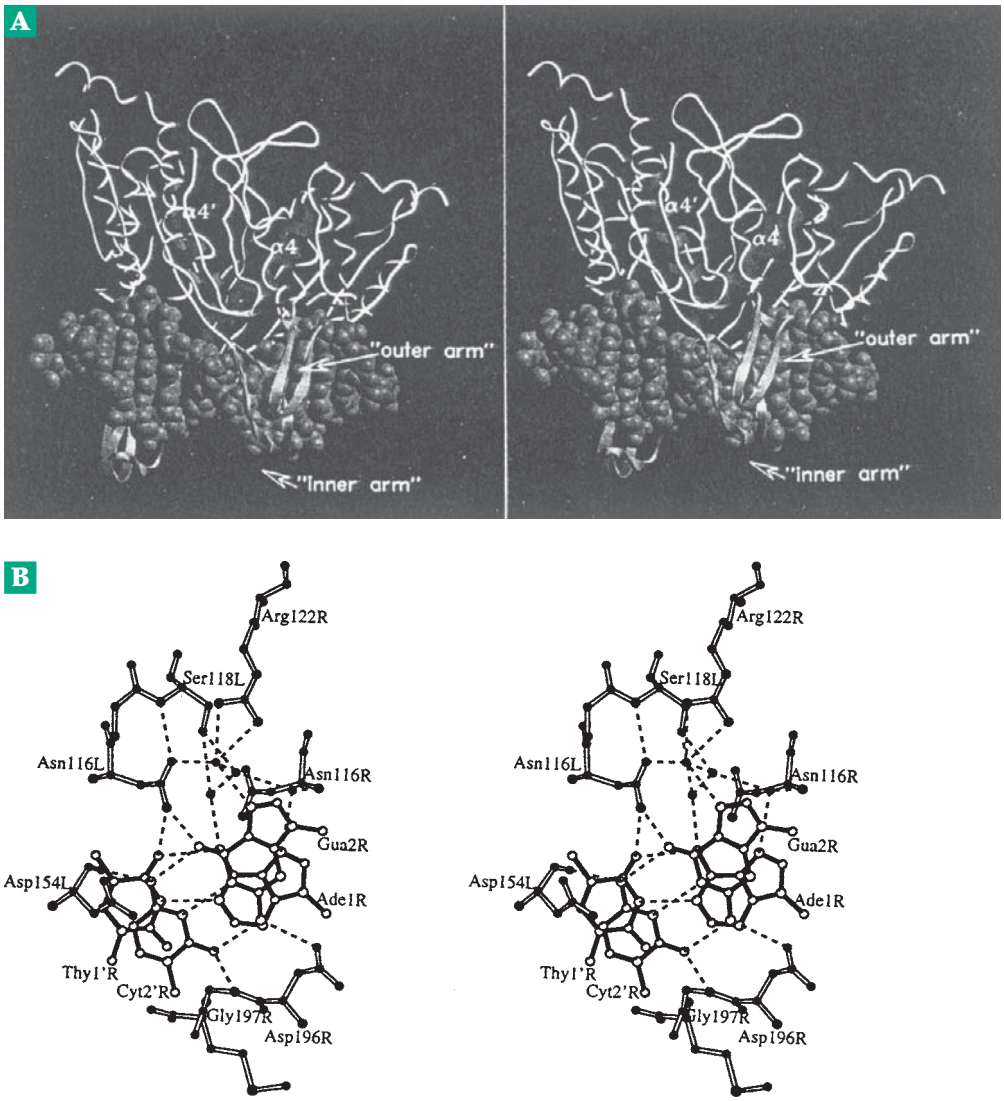


Figure 26-5 (A) Stereoscopic ribbon drawing of the dimeric *Eco*RI restriction endonuclease in a complex with DNA. The equivalent helices marked α4 and α4' point into the major groove of the DNA double helix while the inner and outer "arms" wrap around the DNA. From Bozic *et al.*[84] based on coordinates of Kim *et al.*[94] (B) Stereoscopic view of the two base pairs T•A and C•G of the right end of the recognition motif 5'-CGATCC-3' (Table 26-2) bound to the *Bam*HI restriction endonuclease. The third base pair C•G lies below the two that are shown. Notice the numerous hydrogen bonds, some of which bind atoms of the DNA directly to atoms of the protein and also hydrogen bonds to water molecules (filled circles). The tight packing of complementary charged and dipolar groups of protein, nucleic acid, water, and Mg^{2+} ions (not seen in this drawing) thoughout the complex accounts for the high specificity of these enzymes. From Newman *et al.*[83] Courtesy of Aneel Aggarwal.

The fragments were aligned using genetic markers, and their lengths were estimated by their electrophoretic mobilities.[77] The time-of-entry map has been added on a distorted scale in Fig. 26-4B. There are, on the average, 46.4 kb of DNA per minute, but this amount varies around the chromosome between 38 and 61 kb / min.

Mapping with restriction endonucleases was for many years an essential step in determination of the complete sequence of a piece of DNA. To make a restriction map the DNA, which may have been cut from a chromosome by a restriction endonuclease, is cloned. This permits isolation of a large amount of the DNA, which is then cut by other restriction endonucleases with differing specificities. Overlapping fragments resulting from the cleavages by single restriction enzymes are ordered to provide a map such as that of yeast mitochondrial DNA shown in Fig. 5-48. Any piece from the mapped DNA can now be cloned, and the exact sequence determined. The development of pulsed-field electrophoresis (Chapter 5), with its ability to separate DNA fragments 2000 kb or greater in length, allows restriction mapping with enzymes that cut at rare intervals to give very large fragments. For example, the *Not*I restriction endonuclease cuts the 4.7 Mb *E. coli* K12 genome into 22 fragments that were used to construct the complete restriction map of Fig. 26-4B.[77] Sequences of many viruses and plasmids, mitochondrial and plastid DNAs, and several bacterial genomes have been determined by use of restriction mapping and sequencing of the restriction fragments. Restriction fragment patterns have also been important to determine eukaryotic genome maps including the first genetic linkage map,[96] the first physical map of the human genome,[97] and the complete human genome sequences (Section G).

4. Electron Microscopy

Physical mapping by electron microscopy has been applied to bacteriophage, which can be obtained with large deletions in various parts of the genome.[98,99] The method can also be applied to cloned pieces of DNA. DNA might be isolated from two different phage strains, for example, from wild-type λ and from a mutant phage with a particular gene or genes deleted. The λ DNA can be denatured readily and separated into *r* strands and *l* strands by isopycnic centrifugation. If the isolated *l* strand of one strain is mixed with the *r* strand of another strain and annealed, a double-stranded DNA will be formed and, if there is a deletion in one strain, the homologous region in the normal λ DNA will form a single-stranded loop that can be visualized in the electron microscope. Figure 26-6 shows an example of a micrograph of such a **heteroduplex** molecule with a deletion loop and also a "bubble," where a segment of nonhomologous

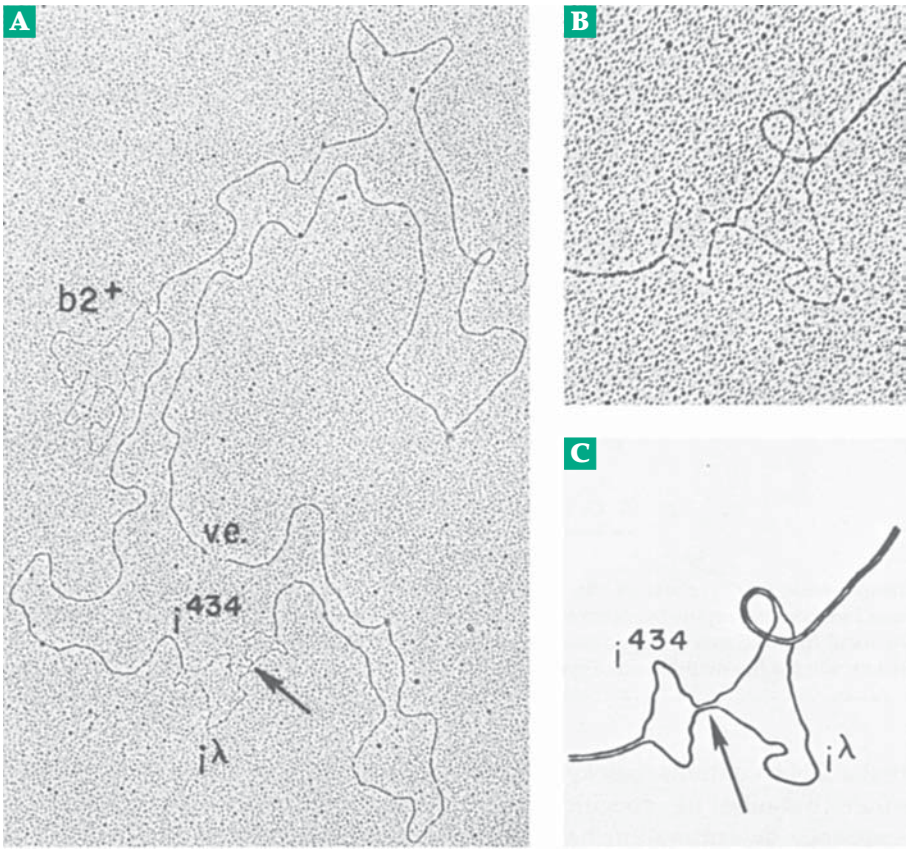


Figure 26-6 (A) Electron micrograph of a heteroduplex DNA molecule constructed from complementary strands of phages λb2 and λ*imm*434. In λb2 a segment of λ DNA has been deleted producing a deletion loop (labeled b2) and in λ*imm*434 a piece of DNA from phage 434 has been substituted for λ DNA resulting in a "nonhomology bubble" (labeled i^{434} / i^{λ}). The vegetative (cohesive) ends of the DNA are labeled v.e. (B) Enlargement of the nonhomology bubble. (C) Interpretative drawing of view in (B). Arrow marks a short (20–150 nucleotide) region of apparent homology. From Westmoreland, Szybalski, and Ris.[98]

DNA has been substituted in one strand.[98] Since distances can be measured accurately on the electron micrographs, rather precise (±50–100 bp) physical maps can be obtained. The chromosome map of phage λ was mapped in this way initially; now its complete nucleotide sequence is known. Another electron microscopic method is useful for location of AT-rich regions that denature readily. In a suitable concentration of formamide these regions melt to form visible single-stranded **denaturation loops** similar to the bubbles in Figure 26-6.

An important technique is to hybridize pieces of mRNA with DNA. If this is done with denatured (single-stranded) DNA and processed mRNA, which has been transcribed from genes with intervening sequences, the intervening sequences will form single-stranded loops in the DNA–RNA hybrid. A related technique depends upon the increased stability of DNA–RNA hybrids in high concentrations of formamide. Under these conditions an RNA segment will hybridize with its complementary strand of the DNA duplex displacing the other strand of DNA, which then appears as a visible **R-loop**. Intervening sequences appear as undisturbed DNA duplexes.

5. Optical Mapping

The long DNA molecules of bacterial or eukaryotic chromosomes are easily broken by vigorous stirring. However, if handled carefully very large fluorescently stained DNA molecules of 0.4–1.4 Mb lengths can be stretched out on a glass surface, and their lengths measured by optical microscopy. The technique depends upon binding of one of the ends of the DNA to the glass surface. If biotin is covalently attached to one end of the DNA, it will bind to a steptavidin-coated plate. However, unaltered DNA also binds to a vinyl-silane or a trichlorosilane coating under suitable conditions.[100,101] If DNA is incubated with a submerged silanized coverslip, which is then pulled out of the liquid mechanically at a constant speed, the DNA molecules are "dyamically combed" so that they are aligned for easy observation and measurement.[101] DNA can be transferred directly from agarose gels used for electrophoretic separation to the plates.[102] Used in combination with restriction enzymes these procedures allow rapid automated construction of physical maps. An example of whole-genome optical mapping is provided by the restriction map of the genome of the radiation-resistant bacterium *Deinococcus radiodurans*. The genome consists of two circular DNA molecules of 2.6 and 0.415 megabases and a smaller 176-kilobase DNA.[103] These were mapped without the laborious subcloning required by conventional restriction mapping. For example, the *E. coli* restriction map required analysis of 3400 phage clones.

Mapping of eukaryotic chromosomes has involved additional methods which are discussed in Sections E and G,1. These incluse **radiation hybrid mapping**,[104] use of **meiotic recombination**, identification of **restriction fragment length polymorphisms** (**RFLPs**; described in Section E,7), and use of **expressed sequence tags** (**ESTs**), short DNA sequences deduced from mRNA molecules transcribed from the DNA.[105,106]

E. Cloning, Modifying, and Transferring Genes

A true revolution in biology and in medicine is in progress as a result of our ability to clone, sequence, mutate, and manipulate genes at will. Methods of sequence determination are discussed in Chapter 5 as is the laboratory synthesis of oligonucleotides and of complete genes. Both of these techniques are essential to present-day genetic engineering as are the techniques of cloning,[99,107–113] which are considered in this section.

A diploid cell contains only two copies of many genes, and these two copies are often not identical. In a gram of any tissue, which may contain 10^9 cells, we will have only ~1 ng of a 1 kb (1 kilobase) gene. To isolate this gene we would have to fish it out from a huge excess of other genes. Present-day cloning techniques offer a way to locate the gene, increase its quantity by many orders of magnitude, learn its sequence, induce any desired mutations at any points, and transfer the gene into other organisms in such a form that it can be expressed, i.e., be transcribed and induce synthesis of proteins. The methods have taken years to develop and continue to be improved. Only some basic procedures and concepts are described here. Numerous manuals[99,108,109,111,114–118] as well as commercial "kits" are available.

1. Joining DNA Fragments

The cloning and manipulation of genes usually depends upon the precise cutting of DNA into discrete fragments by restriction endonucleases. Many restriction enzymes generate **cohesive ends** (sticky ends). Thus, *Eco*RI produces DNA fragments with the single-stranded "tails" shown here at the 5'-ends of the cut duplexes:

```
5′ ----- G 3′                    5′ pA A T T C ----- 3′
3′ ----- C T T A Ap 5′                      3′ G ----- 5′
```

These cohesive ends can be used to join together different restriction fragments. It is easy to see that the complementary single-stranded tails can form base pairs to regenerate the original hexanucleotide

sequence cleaved by *Eco*RI. There will still be nicks between G and A at the specific cleavage points, but these nicks can be closed enzymatically using DNA ligase. Thus, the original DNA cleaved by *Eco*RI can be reformed, or another piece of DNA that also has tails generated by *Eco*RI can be grafted onto an end.

Many of the other enzymes in Table 26-2 also form cohesive ends. Five of them (*Bam*HI, *Bcl*I, *Bgl*II, *Mbo*I, and *Sau*3A) have at the center of their recognition sites the same tetranucleotide: GATC. Enzymes *Sau*3A and *Mbo* I are called **isoschizomers** because they have just the same 4-base recognition sequence and also yield the same restriction patterns. Notice that they will both cut all of the *Bam*HI, *Bcl*I, or *Bgl*II sites, but *Bam*HI and *Bgl*II will not cut all *Mbo*I or *Sau*3A sites. However, cohesive ends made by any of these enzymes can be joined. The gaps left during the joining of certain of the fragments can be ligated enzymatically. *Sau*3A will cut at either methylated or unmethylated sites but *Mbo*I will not cut at methylated sites.

Two enzymes (*Kpn*I and *Pst*I in the list in Table 26-2) form 3'-cohesive ends rather than 5'-cohesive ends. In addition, there are three (*Alu*I, *Eco*RV, and *Hae*III) that cut at the local twofold axis; they form no cohesive ends but leave **blunt ends** (flush ends). Blunt end fragments are also much used in genetic engineering. "Linkers" that provide cohesive ends can be added.[119] The *Sfi*I endonuclease cuts between two 4-bp palindromes in a 13-bp recognition sequence (Table 26-2).[120]

Some useful enzymes. Several enzymes of use in cloning[121] are listed in Table 26-3. The detailed chemistry of most of these is discussed in Chapter 27. Among these are the **ligases** that allow DNA fragments to be joined. They act on DNA strands with adjacent 3'-OH and 5'-phosphate termini. The *E. coli* ligase seals single stranded nicks using NAD^+ as an energy source (Eq. 27-5). It is therefore able to ligate DNA fragments with cohesive ends. The T4 DNA ligase, which is obtained from *E. coli* infected with phage T4, not only can seal nicks but can ligate pieces of DNA with blunt ends. Its activity is linked to cleavage of ATP. If two DNA strands are joined, but with gaps in one or both strands, the gaps can be filled efficiently by the 109-kDa **DNA polymerase I** from *E. coli*. Most often the 76-kDa **Klenow fragment**, which is lacking the 5' → 3' exonuclease activity, is used. **T4 DNA polymerase** has similar properties.

A problem with DNA fragments with cohesive ends is that they spontaneously form closed circles, a process that may compete with a desired joining to another piece of DNA. One solution to this problem is to hydrolyze off the 5'-phosphate groups with an **alkaline phosphatase** (Chapter 12). This prevents formation of covalently closed circles. However, pieces of DNA that retain their 5'-phosphate groups can be ligated to these dephosphorylated pieces. **T4 polynucleotide kinase**[122] can be used to put a phospho group back onto the 5' end of a chain. A ^{32}P end label can be added to such a polynucleotide using ^{32}P-labeled ATP.

Forming homopolymeric tails. Chromosomal DNA may be cleaved with restriction enzymes that leave blunt ends or it may be cleaved randomly by shearing. In either case the blunt ends can be treated first with **λ-exonuclease**, then with **terminal deoxynucleotidyl transferase**[123] isolated from calf thymus. The exonuclease treatment cuts off a few nucleotides from the 5' termini leaving short single-stranded 3'–OH termini. The terminal transferase, a polynucleotide polymerase that acts on ssDNA, is nonspecific and requires no template. Using an appropriate nucleotide triphosphate, it will add a single-stranded tail of either deoxyribonucleotides or ribonucleotides to the exposed 3' termini of a polynucleotide of three or more residues. If deoxyATP is used, a 3' poly(dA) tail will be added to each 5' terminus. Such a poly(dA)-tailed DNA fragment can be annealed and ligated to DNA carrying poly(dT) tails. This approach has been used widely to insert a piece of DNA into a cloning vehicle. For example, a circular plasmid (Fig. 26-7) can be opened by a single cleavage with *Eco*RI or other suitable restriction enzyme. The opened plasmid is treated with exonuclease, and poly(dA) tails are added. The piece of DNA to be cloned is tailed with poly(dT). After annealing and ligation recombinant plasmids carrying the **passenger DNA** will be formed.

If DNA is cleaved with *Pst*I or *Kpn*I (Table 26-2), the resulting 3' cohesive ends can be extended with a poly(dC) tail. If the cloning vehicle also has a site for *Pst*I or *Kpn*I, it can be opened and poly(dG) tails can be applied. A useful feature is that after annealing, filling in the gaps, and ligation the original *Pst*I or *Kpn*I sites are restored. This provides for easy recovery of the cloned fragments (Fig. 26-8).

Preparing material for cloning. DNA may be prepared for cloning by (1) random cleavage by shearing or by enzymatic attack, (2) cleavage by one or more restriction endonucleases, (3) preparation of cDNA from mRNA, or (4) nonenzymatic chemical synthesis of DNA segments. The use of random cleavage has largely been replaced by cleavage with restriction enzymes. A major problem is the separation of the very large number of different restriction fragments formed from a large piece of DNA or from an entire genome. The creation of "libraries" of such fragments is described in Section 5. Considerable simplification comes from separation of individual eukaryotic chromosomes before the library is prepared. Careful purification of DNA to be used in cloning is helpful. This may be done by electrophoresis in agarose or

polyacrylamide gels or using HPLC. One technique is to embed cells directly in a gel, to diffuse in proteases and restriction enzymes that lyse the cells and release the DNA, and cleave it, and then to conduct the electrophoresis (see Chapter 5, Section H,1). DNA fragments of very large size can be separated. In addition to isolation of a fragment to be cloned the cloning vehicle must be prepared. This often involves release of a plasmid by lysis of bacteria that carry it and isolation using a suitable column. Likewise, after the DNA has been cloned and the content of DNA has been increased by growing a large bacterial culture, the plasmids must be released and purified, and the cloned DNA excised with a restriction enzyme. Alternatively, the cloning vehicle may be a virus, which must be isolated and disrupted to release the DNA. Use of the polymerase chain reaction (PCR; Fig. 5-47) allows cloning and amplification of DNA fragments with a minimum of purification.

2. Cloning Vehicles (Vectors)

Many cloning vehicles, more commonly referred to as **vectors**, originated with naturally occurring, independently replicating plasmids or viruses (replicons). More recently artificial chromosomes have been developed as cloning vehicles. Plasmids and viruses have been extensively engineered to provide convenience and safety. A large number of specialized

TABLE 26-3
Some Enzymes Used in Molecular Cloning

Name	Source	Reaction
T4 DNA polymerase	*E. coli* infected with bacteriophage T4	5'→ 3' chain growth 3'→ 5' exonuclease
E. coli DNA polymerase I and Klenow fragment	*E. coli*	5'→ 3' chain growth 3'→ 5' exonuclease 5'→ 3' exonuclease (lacking in Klenow fragment)
Reverse transcriptase	RNA tumor viruses, e.g., avian myoblastosis virus	5'→ 3' DNA chain growth
Ribonuclease H	*E. coli*	Cuts RNA in DNA-RNA hybrid
Lambda and T7 exonucleases	Bacteriophages	
Bal 31 nuclease	*Alteromonas espejiano*, a marine bacterium	Degrades both 3' and 5' termini of dsDNA
T7 RNA polymerase	Bacteriophage T7	DNA-dependent RNA polymerase
Terminal deoxyribonucleotide transferase	Thymus gland, plants	Limited 5'→ 3' chain growth; template independent addition of tails to DNA fragments
T4 DNA ligase	*E. coli* carrying an engineered λ phage	Ligation of DNA, either blunt or cohesive ends; uses ATP
E. coli DNA ligase	*E. coli*	Ligation of DNA with cohesive ends; uses NAD^+ as energy source
RNA ligase	Bacteriophage T4	Ligation of RNA and DNA
T4 polynucleotide kinase	Bacteriophage T4	Phosphorylation of 5′-OH terminus of a polynucleotide (DNA or RNA)
*Eco*RI methylase	*E. coli*	Transfer CH_3 from *S*-adenosylmethionine to adenines in *Eco*RI sites

See Chapter 12 for general discussion of nucleases.

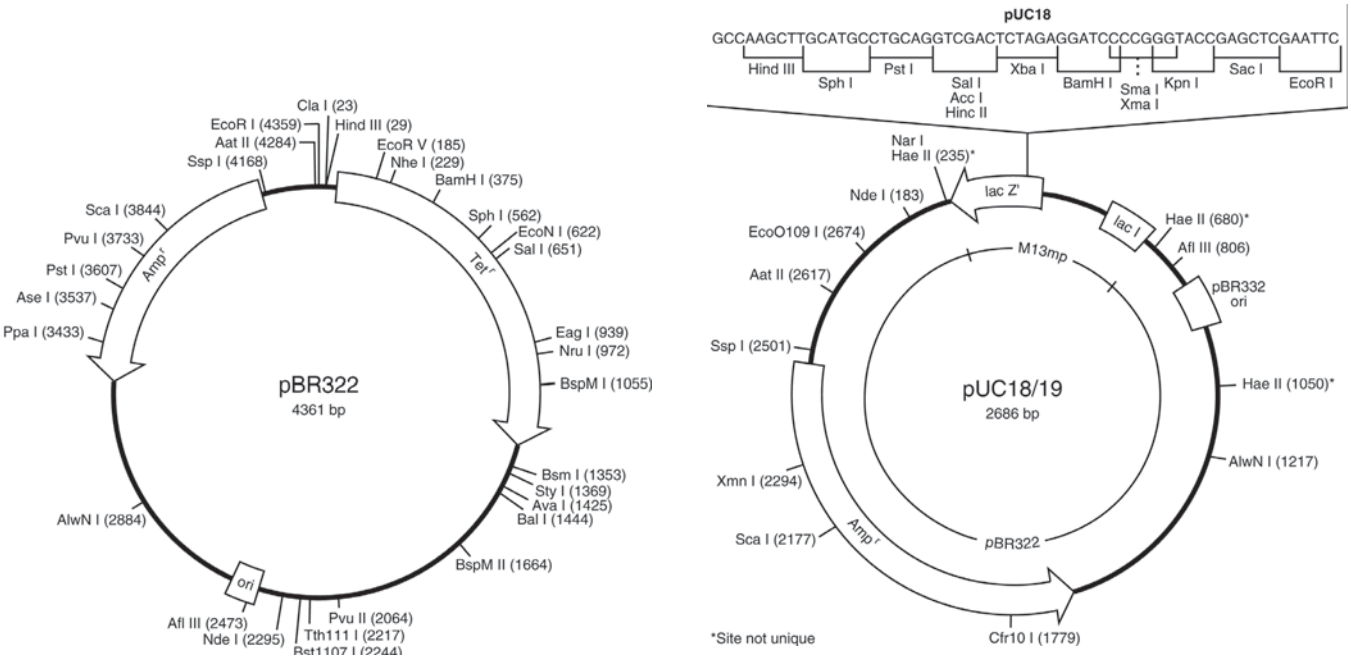


Figure 26-7 Genetic map of cloning plasmids pBR322 and pUC18. Abbreviations: ori, origin of replication; Amp^r, ampicillin resistance gene; Tet^r, tetracycline resistance gene. Other abbreviations are for sites cleaved by specific restriction endonucleases, a few of which are defined in Table 26-2. The nucleotide sequence numbers and directions of transcription are also indicated. Reproduced by permission of Amersham Pharmacia Biotech Inc.

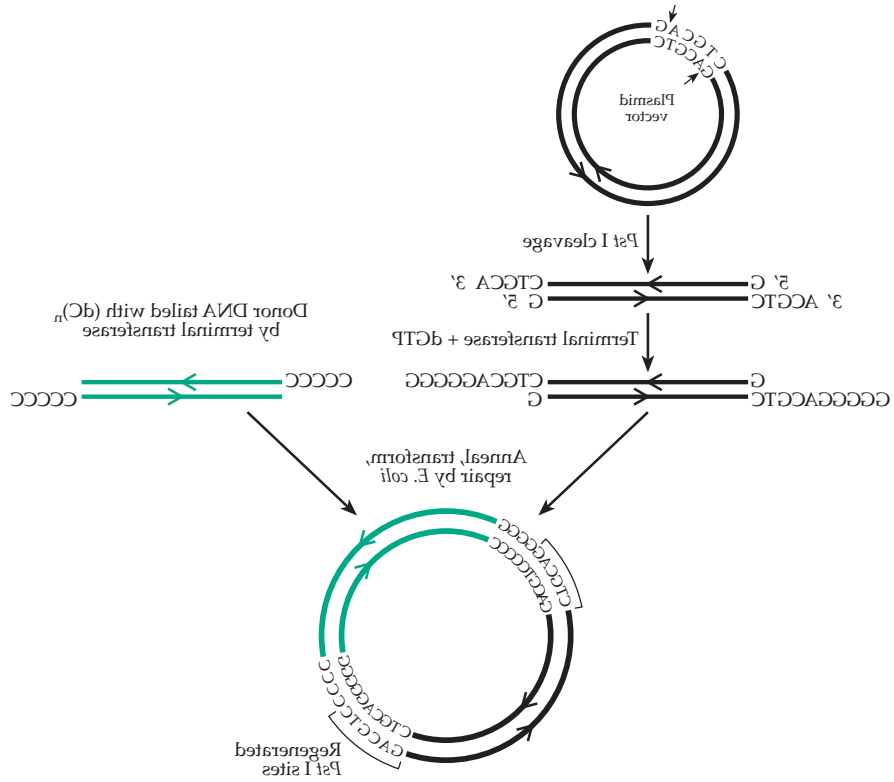


Figure 26-8 Regeneration of a *Pst* I site by dG:dC tailing with terminal transferase. See Glover.[99] Arrowheads indicate 5′→3′ directions.

vehicles have been devised. Only a few will be described briefly here. Suitable books and manuals must be consulted for details.[5,99,108–110,114,115,121,123a]

Plasmids related to ColE1.

The small colicinogenic plasmid, ColE1 of *E. coli*, is attractive for cloning because there are 20 copies per bacterial cell. If chloramphenicol is added during the logarithmic phase of growth, the *E. coli* cells will make 1000–3000 copies. Thus, a relatively small culture of the bacteria containing the cloned DNA in a ColE1 plasmid will yield a large amount of the desired DNA. In the past one of the most widely used cloning vehicles[124] has been plasmid pBR322, which was derived from a close relative of ColEl. The genome size of the original plasmid was reduced by deletion of genes unnecessary for its successful replication. Transposons

that might permit accidental transfer of DNA to other organisms were inactivated. Unneeded or undesirable restriction enzyme sites were eliminated, and useful restriction sites were introduced by point mutations. The resulting pBR322 has only one site of cleavage each for *Bam*HI, *Sal*I, *Pst*I, *Pvu*I, and *Eco*RI. These are at known positions in the 4363-nucleotide plasmid.[125]

Plasmid pBR322 contains two different antibiotic-resistance genes that were brought in from bacterial R-factors. These are used in selecting bacterial colonies that carry the desired recombinant plasmids. One of these is the β-lactamase gene, which confers resistance to ampicillin (Amp^r); the other provides resistance to tetracycline (Tet^r). Their positions are indicated in the pBR322 gene map (Fig. 26-7) as is the essential origin of replication (*ori*). The drug resistance genes are used as follows. If a unique restriction site such as that for *Bam*HI or *Sal*I that lies within the Tet^r gene is used to introduce the passenger DNA, the resistance to tetracycline is lost but that to ampicillin is retained. Thus, after incubation with the recombinant plasmids under conditions that favor their uptake by the host bacteria, the bacteria are plated onto an ampicillin-containing medium. Only those harboring the pBR322 plasmid with its Amp^r gene can grow. After these have produced small colonies, a replica plate is made on a tetracycline-containing medium. On this medium the desired recombinants do *not* grow because the Tet^r gene has been inactivated. This allows selection of colonies containing passenger DNA (Fig. 26-9). A further selection procedure is required to establish that the piece of DNA inserted into the recombinant plasmid is one that is desired.

Typical cloning procedure with pBR322. In much simplifed form the procedure might go as follows: (1) Purchase or isolate plasmid. (2) Cleave plasmid with *Bam*HI; heat at 70°C to inactivate the enzyme. (3) Treat with alkaline phosphatase to remove the 5'-phospho groups. (4) Mix with passenger DNA with cohesive ends generated by *Bam*HI, anneal, and join with DNA ligase. Although the resulting circular recombinant DNA contains a nick as a result of the missing 5'-phospho group, it will be taken up by bacteria and repaired. (5) Incubate joined DNA with cells of host *E. coli* that have been made permeable to DNA by treatment with Ca^{2+} ions. This type of transformation is called **transfection** and is widely used in cloning. (6) Plate transfected cells onto agar containing the first antibiotic, in this case ampicillin. (7) Make replica plate on medium containing second antibiotic, in this case tetracycline. (8) Screen selected colonies for desired DNA fragment. In one procedure a small sample from each of the selected colonies is placed onto spots on a nitrocellulose filter. Several colonies can be placed on one filter and the bacteria lysed, hybridized with a radioactive probe, and then viewed by autoradiography.

Selecting clones using β-galactosidase in pUC cloning vehicles. The newer **pUC** vehicles[126] contain the origin of replication and the ampicillin-resistance gene from pBR322. In addition, a segment of DNA from the *E. coli lac* operon (Fig. 28-2) has been grafted into an intergenic region (Fig. 26-7). It contains the lac control region as well as the coding sequence for the first 145 residues of β-galactosidase. Within bacteria containing the pUC DNA an N-terminal portion of β-galactosidase is synthesized. The specially designed host cell contains (in an episome) the gene for another defective β-galactosidase, one lacking the N-terminal portion. This defective enzyme, together with the N-terminal portions encoded in the pUC DNA, forms an active galactosidase. When the chromogenic substrate 5-chloro-4-bromo-3-indolyl-β-D-galactoside ("X-gal") and a suitable inducer, such as isopropylthio-β-galactoside (IPTG), are present, the unoccupied plasmid vehicles generate blue colonies. However, if passenger DNA is inserted within the galactosidase gene segment, formation of the enzyme will be disrupted and white colonies will appear. The restriction sites for cloning are placed in a **polylinker** near the 5' terminus of the galactosidase gene. The polylinker has been carefully engineered to maintain the correct reading frame and to avoid disruption of the galactosidase activity. It contains several different restriction sites for insertion of passenger DNA (Fig. 26-7). Such insertion does destroy the galactosidase activity allowing the user to detect the recombinant DNA from the white plaques.

Filamentous bacteriophages. An important series of cloning vehicles have been derived from the circular replicating forms of the filamentous bacteriophage M13 (Chapter 5; Fig. 7-7; Chapter 27).[127] Although the genome contains only a short intergenic region that can be deleted, up to 50 kb of passenger DNA can be inserted into these vehicles. Since long inserted sequences may be deleted spontaneously, M13 is most useful for cloning about 300- to 400-nucleotide chains. Many of the M13 vehicles also use the β-galactosidase blue-white screening technique. These modified viruses are highly infective, but the infected *E. coli* cells are not killed. Rather they produce large numbers of virus particles with single-stranded DNA representing one of the two DNA chains of the parental phage. These are widely used for sequencing by the chain-termination procedure of Sanger *et al.* (Chapter 5). The procedure requires a primer sequence. If M13 recombinants are sequenced, the primer consists of a synthetic oligonucleotide that can be annealed to the galactosidase gene fragment at its 3' end just in front of the DNA segment that is to be sequenced.

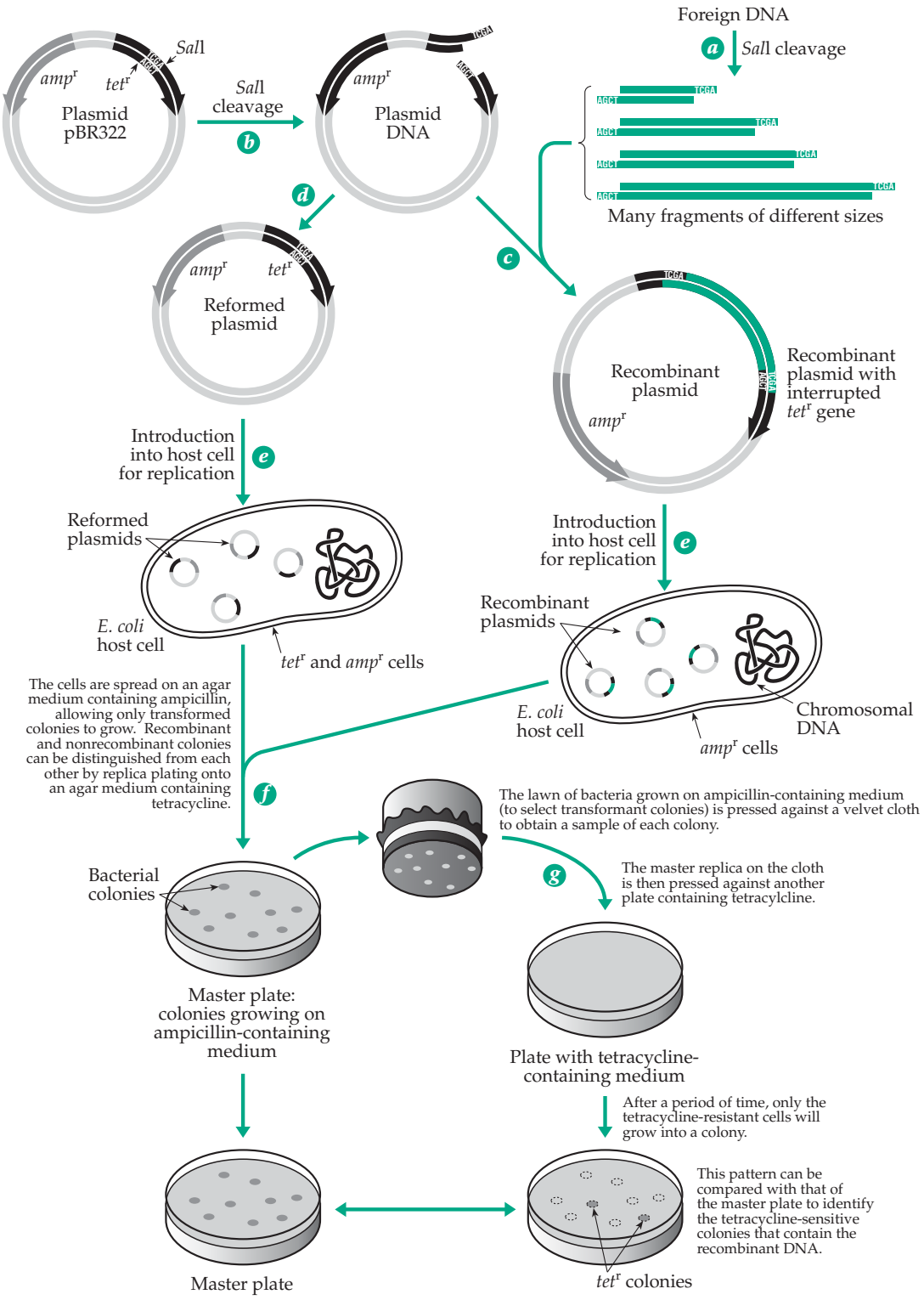


Figure 26-9 A classical scheme for cloning DNA in a pBR322 plasmid vehicle. A DNA sample is digested (step *a*) with one of the restriction endonucleases (e.g., *Sal*I), that cuts the cloning vehicle within the *Tet*^{r} gene (see Fig. 26-7). The plasmid is also cut with the *Sal*I restriction enzyme (step *b*). After mixing with the digested DNA sample and annealing, recombinant molecules are formed (step *c*). Some plasmids are reformed (step *d*). Both recombinant and reformed plasmids transform *E. coli* cells (step *e*). The transformed cells are plated on an agar medium containing ampicillin (step *f*). Only cells containing the ampicillin-resistance gene grow (step *g*). A replica plate is made and is pressed onto another plate containing tetracycline. The recombinant colonies do not grow on the medium because the tetracycline-resistance gene has been interrupted. By comparing the two plates recombinant colonies can be selected. These can be tested, using a suitable probe, to determine whether they carry a desired gene. After Atherly, Girton, and McDonald.[2]

Lambda cloning vehicles. Many cloning vehicles have been derived from the 48,502-bp[128] DNA from the temperate *E. coli* bacteriophage λ. The DNA from phage particles are taken up efficiently by *E. coli* cells, much more so than by transfection. The virus has a complex life cycle, which is discussed in Chapter 28. Within the phage head the λ DNA exists in a folded linear form with 12-base 5' cohesive ends (Fig. 28-11). After entrance into the bacterial cell the DNA cyclizes through its cohesive ends and is ligated by the *E. coli* ligase. Replication of the circular forms ensues. Later in the cycle, rolling circle replication (Eq. 27-7 and associated discussion) produces long concatamers with several phage genomes joined as a single chain. This DNA is "packaged" into new phage heads. As this is done, a nuclease cuts the concatamers at *cos* sites forming monomeric genomes with cohesive ends. This is the **lytic cycle** of the phage. In the alternative **lysogenic cycle** the DNA becomes integrated into the *E. coli* genome. Maintenance of the lysogenic state depends in part on gene *cI*, which encodes a repressor (see Chapter 28) that prevents expression of the genes required in the lytic pathway. Since only the lytic cycle is needed for cloning, it is convenient to place a cloning site within the *cI* gene. The screening of recombinant phage particles is done by examination of plaques. A phage without inserted DNA will be able to undergo both the lytic and lysogenic cycles and will form turbid plaques. However, if passenger DNA is inserted in the *cI* gene, the lysogenic cycle is prevented, and clear plaques are formed.

Of the ~50 genes present in native λ, only about half are necessary for replication in the lytic cycle. Thus, it is possible to delete about 1/3 of the genome to make room for more passenger DNA. However, to form mature phage particles the length of the DNA must be at least 75% of the native length. No more than 110% of the native amount may be present. The total DNA must fall between 38 and 53 kb in length. To accommodate these packaging requirements **replacement vectors** containing unnecessary "stuffer DNA" between two lambda "arms" are used. The unneeded stuffer piece has the same kind of restriction site or sites at each end so that it can easily be cut out and replaced by the passenger DNA. This permits cloning of DNA segments up to about 22 kb in length. Lambda vehicles have all been engineered to eliminate undesired restriction sites and to reduce the number of sites for *Eco*RI and other restriction enzymes commonly used for cloning. The widely used **Charon series**[129] have been further engineered so that they will grow only in strains of bacteria that cannot survive in the human intestinal tract. For example, amber mutations (Section B,6) are incorporated into genes needed for phage assembly, and the bacterial hosts must contain an amber suppressor gene. The bacteria are also nutritional auxotrophs with absolute requirements for thymidine and diaminopimelic acid in the medium. The latter compound is not found in the intestinal tract. The purpose of these alterations is to prevent the spread of recombinant DNA into the environment.

To sequence DNA carried in a lambda vehicle or to study it in other ways, it is often necessary to cut it with restriction enzymes, to prepare a restriction fragment map, and to subclone the fragments into a plasmid vehicle. Lambda vehicles, which will automatically transfer the passenger DNA into an M13 vehicle when propagated in a host carrying a special helper virus, have been devised.[130] The helper virus encodes proteins that recognize and cleave sequences that mark the initiation and termination of M13 DNA synthesis. These are used to mark the ends of the passenger DNA. As DNA synthesis occurs the displaced passenger DNA and M13 genes are excised, circularized, and converted into a replicating form of an M13 cloning vehicle.

Cosmid vehicles and in vitro packaging. Cosmids[131] are hybrids of a plasmid vehicle and phage λ. They contain the *cos* sites that are cleaved during packaging of λ DNA. A cosmid cleaved at a restriction site will form upon ligation a range of different sized DNA molecules that contain *cos* sites on both sides of a piece of passenger DNA, which may be up to 45 kb in length (Fig. 26-10). This can be cut at the *cos* sites and packed into heads using an *in vitro* packaging system. In this system the unassembled subunits of the phage particle are produced in special strains of bacteria and

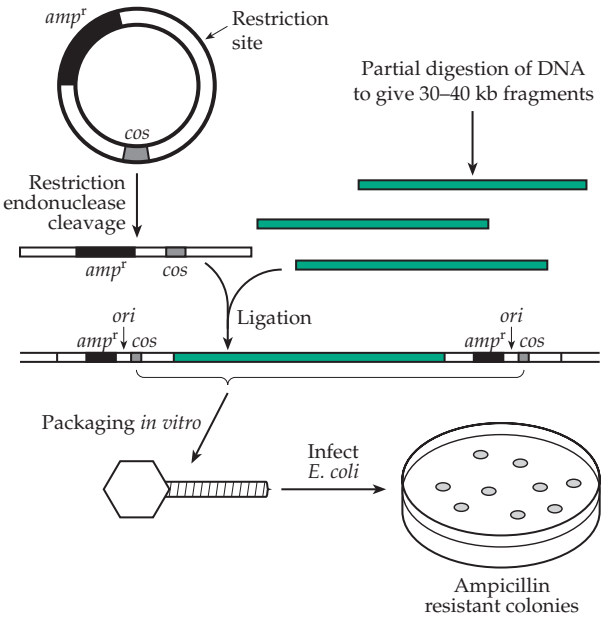


Figure 26-10 Cloning DNA in cosmids. See Glover.[99]

are allowed to assemble and package the cosmid DNA. Since cosmids contain relatively large pieces of DNA, they are useful for preparing sequentially overlapping clones that allow the investigator to "walk" along the DNA hunting for a target gene. Cosmid vehicles have been designed to allow both efficient genomic walking and restriction mapping.[132]

3. Expression of Cloned Genes in Bacteria

A major goal in recombinant DNA technology is the production of useful foreign proteins by bacteria, yeast, or other cultured cells. Protein synthesis depends upon both transcription and translation of the cloned genes and may also involve secretion of proteins from the host cells. The first step, transcription, is controlled to a major extent by the structures of promoters and other control elements in the DNA (Chapter 28). Since eukaryotic promoters often function poorly in bacteria, it is customary to put the cloned gene under the control of a strong bacterial or viral λ promoter. The latter include the λ promoter P_L (Fig. 28-8) and the *lac* (Fig. 28-2) and *trp* promoters of *E. coli*. These are all available in cloning vehicles.

It is often useful to create hybrid proteins fused to the *E. coli* β-galactosidase gene. If another gene is spliced in at either the N terminus or the C terminus of the galactosidase (*lac*Z) gene but is kept under control of the *lac* promoter, the resulting hybrid protein will have galactosidase activity, which can be used for screening. In addition, the hybrid protein will often react with antibodies directed against the protein whose gene is being cloned. Another kind of hybrid fuses the cloned gene to that of β-lactamase, for example at the *Pst I* site of plasmid pBR322 (Fig. 26-6). The β-lactamase activity will be gone, but the hybrid protein will be secreted because β-lactamase normally is secreted into the periplasmic space and its N-terminal signal sequence is now fused to the cloned protein. Engineering of a suitable site for cleavage by a protease can release the foreign protein in an active form. A variety of other **expression systems**, often using **reporter gene** products, have been developed.[133]

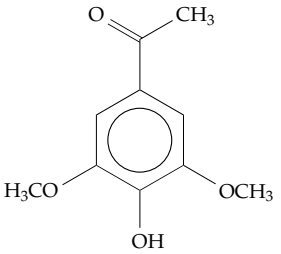
Bacteria often degrade foreign proteins by hydrolytic attack. One way in which such damage has been minimized is to clone multiple fused copies of the gene for the desired product. The resulting polyprotein may be resistant to degradation, and if the gene has been correctly engineered, may be cut apart by cyanogen bromide (Eq. 3-17) or by a specific protease.

For many years most cloning was done in *E. coli*, but cloning systems have now been developed for many other bacteria including *Bacillus* and other gram-positive bacteria, and also for yeast, insect cells, animals, and plants.[133–135]

4. Cloning and Transferring Eukaryotic Genes

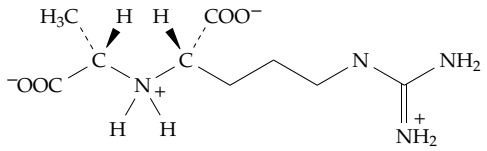
Eukaryotic genes cloned into bacterial plasmids are often poorly expressed. It is advantageous to clone such genes in eukaryotic cells, where the cutting and splicing of hnRNA to remove intervening sequences during formation of mRNA (Chapter 28) does occur. This permits expression of the cloned gene, something that is possible in bacteria only if cDNA that lacks the introns is cloned. The need for posttranslational modification of many proteins also interferes with expression in bacteria. Many methods of gene transfer and cloning have been developed.[136] The yeast *Saccharomyces cerevisiae* is often an ideal host for cloning. It grows rapidly in either its haploid or diploid stage (Chapter 1). Some strains carry a 2-μm 6.3-kb circular plasmid with 50–100 copies per cell.[109] This has been developed as a cloning vehicle. Recombinant plasmids can be used as **shuttle vehicles** for transferring genes cloned in *E. coli* into the yeast plasmid. Genes may also be cloned as minichromosomes, such as **yeast artificial chromosomes** (**YACs**). Artificial chromosomes contain origins of replication from yeast, human, or bacterial chromosomes as well as telomeres and centromeres[137] (see also Chapter 27). YACs have been widely used and became popular because they can accomodate 600 kbp or more of DNA.[137–139] However, their use in the human genome project resulted in serious problems of instability.[140] Bacterial artificial chromosomes (BACs), which accommodate only 200–300 kbp, **P1 artificial chromosomes** (**PACs**),[141,142] and human minochromosomes are more stable.[143] Another problem, which affects the use of yeast for production of eukaryotic proteins, is the tendency for poor removal of introns.

Plant genes. Much of the cloning in higher plant cells has made use of the **Ti plasmid** of *Agrobacterium tumefaciens*, a soil bacterium that enters wounds in dicotyledenous plants causing tumors known as **crown galls**.[144–146c] A related species *A. rhizogenes* harbors a similar plasmid that causes "hairy root" disease.[146c,147] The infecting bacteria respond to the synthesis of certain phenolic compounds such as **acetosyringone**, which are produced in plant wounds, by entering the plant cells.[148,149] Only bacteria carrying the Ti (tumor-inducing) plasmid cause tumors. The plasmid carries a 13-kb region

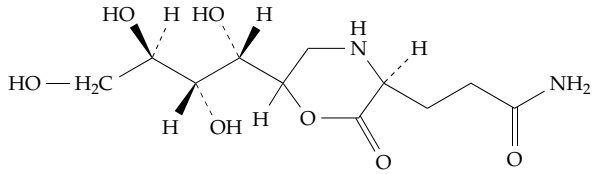


Acetosyringone, a compound that induces infection by *Agrobacterium*

(transferred region) known as T-DNA that encodes enzymes for production of the auxin indoleacetic acid (Fig. 25-12), the cytokinin isopentenyl-AMP, and the compounds known as **opines**. These are reduction products of Schiff bases of amino acids and 2-oxoacids or sugars.



Octopine: derived from pyruvate and arginine
Histopine: derived from pyruvate and histidine[150]
Lysopine: derived from pyruvate and lysine
Nopaline: derived from 2-oxoglutarate and arginine
Leucopine: derived from 2-oxoglutarate and leucine



Agropine: derived from D-mannose and L-glutamine

The auxin and cytokinin, whose production is normally controlled, are now formed in large amounts and cause uncontrolled tumor growth. The opines are used by *Agrobacterium* as a unique source of energy and of metabolites for biosynthesis. The host plant cells, however, cannot catabolize the opines.

Upon entrance into a plant cell the T-region is excised from the Ti plasmid and can become integrated into the DNA of the host plant much as occurs during bacterial conjugation (Fig. 26-3).[151–154a] By deleting the genes for synthesis of auxin, cytokinin, and opines, the Ti plasmid loses its tumorigenic property but is still able to transfer genes into the plant genome. Fusion of an *E. coli* plasmid vehicle into the modified Ti plasmid creates a useful plant shuttle vehicle. Also see below under "physical methods."

Transferring genes with engineered animal viruses. There is special interest in transferring cloned DNA into human cells to correct genetic defects.[155] Transfer of human genes into other animals is also important for a variety of reasons. For example, human proteins of therapeutic value could be produced in animal milk. Animals can be engineered to have defects that mimic those in humans and which can then be studied in animals.[156]

A number of different viruses have been used to transfer foreign genes into eukaryotic cells and also to create stable plasmids for cloning. One of the first was the 5380 bp SV40. The relatively small size of SV40 limits the amount of DNA that can be incorporated. However, many of the functions of this virus can be performed by simultaneous infection with a **helper virus**, often an adenovirus which is itself defective. With the helper virus making proteins essential for SV40 replication, all but 85 base pairs at the origin of replication of SV40 can be deleted and replaced by other DNA. More recently engineered **adenoviruses** and **retroviruses** have been widely used. In both cases genes required for replication of the viruses have been deleted.[155] Adenovirus vehicles can carry 7 kbp or more of passenger DNA. They are efficiently taken up into endosomes and into the nucleus by both replicating and nonreplicating cells. The DNA is not incorporated into the genome of the host cell.[155,157,158] Replication-deficient avian **retroviruses**, or **retrotransposons** (Chapter 28), can carry up to 9 kbp of passenger DNA and are incorporated into the host genome by the process of homologous recombination (Chapter 27).[155,159,160] Both adenovirus and retroviral vehicles suffer from serious problems. The adenovirus proteins can induce fatal inflammatory reactions,[160a] and DNA transfer from the retroviral vehicles is inefficient. A nonpathogenic human parvovirus, the **adeno-associated virus** (AAV), is also being developed as a gene therapy vector.[161,161a,b] Another possibility is to engineer a **lentivirus**, even HIV, to provide efficient integration (Chapter 28).[162,162a] Safety concerns have delayed human testing of this possibility.

A group of insect viruses, the **baculoviruses**, are being used in cultured insect cells for large-scale production of proteins. The cloned genes are placed under the control of the promoter region of the gene that encodes the major viral coat protein.[135,163] The **baculovirus** vehicles can carry over 15 kbp of passenger DNA and may also be useful in human gene therapy.[164,165] Filamentous bacteriophages have been reengineered for the same purpose.[166]

Nonviral methods for DNA incorporation into the human genome may utilize **transposons** (Chapter 27)[166a] or mobile **group II introns** (Chapter 28).[166b,c]

Physical methods of gene transfer. Genes can often be transferred without the use of a cloning vehicle. This is especially important for certain plant cells, such as those of cereal grains, for which transfer of genes via the Ti plasmid has been difficult.[167] If DNA, which may be in a plasmid, is coprecipitated with calcium phosphate, it can often be taken up directly either by animal cells or by plant protoplasts.[168,169] Polycations also facilitate DNA uptake; cationic **liposomes** seem to be especially effective.[170] In the widely used **electroporation** technique a short electrical pulse of a few hundred volts / cm is applied to create transient pores in the plasma membrane through which the DNA can enter a cell.[108,171–175] Chromosomes can be transferred by cell fusion and either

entire chromosomes, isolated DNA, or cloned DNA can be transferred into egg cells by microinjection.[176] Following transfer of DNA or of intact chromosomes, recombination sometimes leads to stable incorporation of some of the transferred genes into the host cell's genome.

A very important technique is the use of high-velocity microprojectiles shot from a particle gun. Spherical tungsten particles of ~0.5–1 μm diameter are coated with recombinant plasmids containing the genes to be transferred. The particles are then shot, using a gunpowder cartridge, into intact recipient cells. The particles penetrate cell membranes, mitochondria, and chloroplasts without serious damage if the number of particles is not too great.[177,178] The technique is very useful for genetic engineering of plants.[179]

5. Genomic Libraries

The human genome contains about 3×10^6 kb of DNA, about the average amount for a eukaryotic organism. If the entire genome is digested to completion with a restriction endonuclease, whose cleavage sites are distributed more or less randomly, the resulting restriction fragments constitute a "library" for that genome. If the average length of a fragment is 17 kb, about 1.8×10^5 unique fragments will be produced. To make a practical library these must be cloned into a suitable vehicle. Derivatives of phage λ or cosmids are most often used. The cloned fragments can then be packaged using the *in vitro* packaging system to form infectious phage particles, which can be propagated as plasmids in *E. coli* cells.

What is the probability that a given fragment among the total produced will be found in one of the recombinant phage in the library? From simple probability theory the number of clones that must be isolated and screened is given by Eq. 26-2.[180] Here N is the number of clones needed, p is the probability of

$$N = \ln(1-p)\,\ln(1-f) \qquad (26\text{-}2)$$

having the desired fragment in the library, and f is the fractional proportion of the genome represented by the fragment sought. For our example of a 17-kb fragment of the human genome this is 17 / 3000. From this equation one can calculate that to have a 99% chance of finding our fragment ($p = 0.99$) we need 8×10^5 clones.[99] What the equation does not show is that there will probably be some long fragments that cannot be cloned in the selected vehicle. These will be missing from the library.

Large numbers of clones obtained can be screened rapidly by colony hybridization using a labeled DNA probe. Thus, if it is desired to isolate a gene for a particular protein and some part of that protein has been sequenced, a synthetic DNA probe can be made. The phage containing the recombinant fragments can be plated, and after plaques form a nitrocellulose filter can be laid on the plate to form a replica. After release of DNA from the phage, denaturation by NaOH, and neutralization the single-stranded DNA fragments can be hybridized with the probe. Another screening method uses a probe carried in a plasmid that promotes homologous recombination between the probe sequence and restriction fragments with a similar sequence.[181] A problem that arises in preparing genomic libraries is that certain sequences, e.g., those involving highly polymorphic regions and inverted repetions, often cannot be propagated in most lambda cloning vehicles.[182]

In addition to genomic libraries, **cDNA libraries** can be prepared from mixed mRNAs. The total RNA of cells is isolated and passed through an affinity column containing oligo(dT) chains. These bind to the 3'-poly(A) tails of the mRNAs, allowing them to be isolated. The mixed mRNAs can then be cloned using a poly(AT)-tailed plasmid vehicle and a reverse transcriptase.[183,184]

6. Probes

The first step in screening the recombinant DNA in a library is use of some probe for detecting the desired DNA fragment. The most direct way is to synthesize a radioactive or fluorescent labeled oligonucleotide[185] complementary to a short known sequence in the protein. The number of codons for a single amino acid varies from 1 to 6 (Table 5-6). It is therefore desirable to prepare a probe complementary to segments of DNA containing a high proportion of codons for Trp and Met (1 codon each) and Asn, Asp, Cys, Glu, Gln, His, Lys, Phe, and Tyr (2 codons each). A popular procedure is to synthesize a mixture of probes containing all of the possible nucleotide sequences coding for the selected sequence of amino acids. The probe may be a mixture of more than 1024 different nucleotide sequences.[186] See also Chapter 5, Section H,4.

Antibodies are another popular type of probe. Antibodies to a specific protein may be utilized in isolating mRNA from ribosomes that are making that protein (Chapter 29). Thus one or more strongly binding antibodies may have already been obtained before the library clones are to be screened. To use this technique for screening recombinant DNA, the cloning must be done in a vehicle that causes expression of the gene, e.g., as proteins fused to *E. coli* galactosidase. One type of expression library is created by insertion of the cDNAs into copies of a bacteriophage gene that permits the expressed proteins to be displayed on surfaces of the phages[186a] (see Fig. 3-16).

7. Studies of Restriction Fragments

By cleavage with the correct restriction enzyme, cloned DNA fragments can be released from the vehicle in which they are carried in the library. What can be done with these fragments? The first obvious use is to sequence a gene that has been located with a probe. In many instances the gene will be longer than the cloned piece. However, the isolated restriction fragment, if labeled and denatured, becomes a highly specific probe for locating other restriction fragments that overlap it. For example, a fragment from an *Eco*RI library may bind to two or more fragments from a *Hin*dIII library. This permits "walking" along the chromosome to locate adjacent fragments. Cosmid vehicles that facilitate ^{32}P-labeling at the ends of the passenger DNA are useful.[187] A related approach called "jumping" depends upon converting very large DNA restriction fragments into circular molecules, digesting with restriction enzymes, and cloning the junction fragments of the circles. These fragments contain segments that may have been separated by as much as 100 kb in the genomic DNA, and enable the investigator to walk or jump again from a new location.[188]

Locating mutations. The study of restriction fragments provides a way of locating many mutations. For point mutations in genes for known proteins, sequencing reveals the exact defect. Some mutations, especially deletions, lead to changes in the lengths of restriction fragments. If the mutation causes loss of a restriction site, a longer piece of DNA will be present than in a digest of normal DNA. Such differences in length of restriction fragments are usually referred to as **restriction fragment length polymorphism** (RFLP, usually used in the plural as **RFLPs**). These polymorphisms are readily detectable by differences in mobility on gel electrophoresis.[96,189] They can often be mapped to particular chromosomes by hybridization *in situ* (Fig. 26-14),[190,191] by study of naturally occurring translocations of chromosome fragments, or other techniques.[192] An example is provided by the human hemoglobin abnormalities known as **thalassemias** (Chapter 32). Here, deletions remove certain restriction sites leading to observation of RFLP. Occasionally a point mutation is linked to RFLP at a nearby site. For example, in the United States most carriers of the sickle cell trait have a 13-kb *Hpa*I fragment that carries the globin S gene. However, noncarriers have their globin gene on a 7.6-kb *Hpa*I fragment.[99] Although the association is fortuitous, the linkage between the hemoglobin S gene and the mutated restriction site is broken only rarely by crossing-over during meiosis. RFLPs have been linked to many other human genetic defects and have also provided the basis for the first linkage maps of human chromosomes.[96,193]

Positional cloning. By using enzymes that cut at relatively rare sites, genomic DNA can be cut into very large restriction fragments. From studies of inheritance within families carrying specific genetic traits it is sometimes possible to find linkages between those traits and polymorphic restriction fragments.[194–196] This has been accomplished for a number of defective human genes including those responsible for sickle cell anemia,[197] cystic fibrosis (Box 26-A), Duchenne muscular dystrophy (Box 19-A), Huntington's disease,[198,199] X-linked chronic granulomatous disease[200] (p. 1072), neurofibromatosis (elephant man's disease),[201] the hereditary cancer retinoblastoma,[202] and others. By 1997 nearly 100 hereditary disease loci had been located by positional cloning.[203] These astonishing successes provided a major impetus for what became the Human Genome Project (Section G).[193]

Serious problems were met in actually locating these disease genes. Crossing-over is infrequent, occurring only at about 50 locations during each meiosis.[192] Therefore, linkage analysis does not tell us with any precision how close the linked gene is to a known DNA probe within a restriction fragment that may be up to 2000 kb in length. Finer restriction mapping or chromosome walking can be used to locate the precise piece of DNA that is defective.[204] This can still be a formidable problem. However, if the defective protein can be identified it can be sequenced. A specific oligonucleotide probe can be made for its gene and can be used to establish the exact chromosome location.

8. Directed Mutation

In addition to the developments of cloning and sequencing of DNA, a third technique is essential to the present revolution in molecular genetics. That is the ability to mutate any gene at any point in a specific way. Because of its precise nature the technique is called **directed mutation** in this book. However, the term **site-directed mutagenesis** is often used. Mutations can be introduced randomly in DNA in many ways including treatment with nitrous acid, bisulfite, formic acid, or hydrazine or by incorporation of nucleotide analogs.[205] Efficient procedures have been devised for isolating the mutants.[206] For many purposes **oligonucleotide-directed mutation** is the preferred technique.[207–210] An oligonucleotide of ~16–20 nucleotide length is synthesized with a sequence complementary to the coding strand containing the desired site of mutation. At that site the codon for the new amino acid is present. Despite this mismatch the oligonucleotide can be successfully hybridized with a single-stranded DNA such as that cloned in an M13 vehicle. Now the Klenow fragment of DNA polymerase I or a viral DNA polymerase (Chapter 27) is used to convert the single-stranded circular DNA into

a double-stranded replicating form. Many of the single-stranded progeny will contain the mutated DNA. They can be screened with a labeled probe made from the oligonucleotide used to induce the mutation. It will hybridize most tightly to the correctly mutated gene.[207] Use of the PCR reaction simplifies the procedure.[210–212] In another screening procedure the template strand is synthesized in an M13 phage vehicle using uracil rather than thymine. The circular heteroduplex obtained after synthesis with T4 DNA polymerase and T4 DNA ligase is taken up by *E. coli* cells, which select against the uracil-containing strand and, therefore, in favor of the mutated strand.[208]

A third approach is to completely synthesize a gene for the protein under study. The sequence does not have to be exactly the natural one but can be made with restriction enzyme cleavage sites that permit easy excision and readdition of particular fragments. Synthetic fragments containing various mutations can then be grafted in at will. Genes of this type have been made for rhodopsin and related proteins (Chapter 23) and for numerous other proteins.

What are the uses of directed mutation? As we have seen in previous chapters, the technique is being used in every area of biochemistry to bring new understanding of protein functions and of the chemical basis of disease. Together with complete synthesis of genes it provides the basis for genetic engineering of specific proteins of plants, animals, and microorganisms. Many protein products can probably be improved. For example, enzymes can be made more stable.[213] Specificities can be changed, but this is difficult.

Targeting and replacing genes. One goal of human genetic therapy is to replace a defective gene in body cells with a good gene. Is this really possible? It is essential to engineer the DNA that is to be transferred, so that it contains all of the components needed for efficient expression in the host following its incorporation as a **transgene**. A transcription-initiation region with suitable promoter, both 5' and 3' untranslated regions, start and stop codons, and polyadenylation site (Chapter 28) must all be present. It is hoped that a correctly constructed promoter region will allow the transgene to be picked up by the machinery of homologous recombination (discussed in Chapter 27) and be incorporated into the host's DNA and expressed in the appropriate tissues.[214,215] Only some cells will take up the new gene and discard the old. However, it may happen enough to benefit a patient. Targeted gene replacement has been very successful in mice.[216]

Knockout mice. If targeted DNA is injected into a fertilized mouse egg, there is a chance that the mouse will have the targeted gene replaced in one chromosome, and that it will be stably transmitted to some of its progeny. If the transgene is totally nonfunctional, the mouse will be a "knockout mouse," suffering from a hereditary defect that can be transmitted through carriers such as its mother. The standard knockout technique is to inactivate the gene of choice in cultured embryonic stem cells and to inject these into mouse embryos. Some progeny will carry the inactivated gene in their germ cells. A refinement of the technique utilizes the **Cre recombinase** or related enzymes discussed in Chapter 27 to selectively remove pieces of DNA from genes in specialized tissues of mice.[217,218] By 1996 several hundred different knockout mice had been created.[215] Nevertheless, interpretations of results of gene knockout are sometimes complex.[219]

F. The Genetics of Eukaryotic Organisms

Whereas DNA synthesis takes place almost continuously in a rapidly growing bacterium, replication of DNA occupies a more limited part of the **cell cycle** of eukaryotes (Fig. 11-15). In a mammalian cell mitosis proper (Fig. 26-11) may require about one hour. It is followed by the "gap" period, G_1, whose length is variable and depends upon the cell type, the nutritional state of the cell, and other factors. About 10 h is typical. During the S phase (~9 h) active DNA replication takes place. This is followed by a second gap (G_2) that occupies 4 h in the 24 h cell cycle shown in Fig. 11-15. The length of the different segments of the cell cycle varies widely among different organisms. Indeed, the concept of a cell cycle can be criticized.[219a] It is only in a rapidly growing culture that all, or most, cells can follow the same cycle. In the adult body most cells are inhibited from division (or are not stimulated to divide) most of the time.

1. Mitosis

The distribution of chromosomes to daughter cells of somatic cells undergoing division is accomplished by mitosis whose successive phases are referred to as **prophase**, **metaphase**, **anaphase**, and **telophase** (Fig. 26-11). As the chromosomes condense during prophase, it is seen that each one actually consists of two separate entities coiled together. These are the identical **chromatids,** which are formed from the two identical double-stranded DNA molecules formed by replication of the DNA of the chromosome during the S phase of the cell cycle. As the folding of the chromosomes occurs (during prophase), the nuclear envelope completely fragments or dissolves in many species.

An important event that *precedes* the main stages of mitosis is the formation of **poles** in the cell. In animal cells, the poles are formed by the **centrioles,**

which move apart and take up positions at opposite sides of the cell. Each of the centrioles is accompanied by a smaller "daughter" centriole lying at right angles to the larger parent. In plant cells, which lack centrioles, a more diffuse pole is formed. As the cell prepares for mitosis, fine microtubules (15 nm diameter) can be seen radiating from the poles. At the end of prophase the microtubules run from one pole to the other to form the **spindle**. Microtubules also become attached to the chromosomes at the **centromeres**.

At metaphase the chromosomes are precisely lined up in the center of the cell to form the **metaphase plate**. Now each centromere divides, permitting the sister chromatids to be completely separated. A protein complex **cohesin**, which holds the sister chromatids together, undergoes proteolysis by a **separase** at this stage.[220,220a,b] During anaphase the separated chromatids, now referred to as **daughter chromosomes**, move to opposite poles as if pulled by contraction of the spindle fibers. Telophase is the final stage in which new nuclear envelopes are formed around each set of daughter chromosomes. In humans and many other species the cell pinches in two. In plants

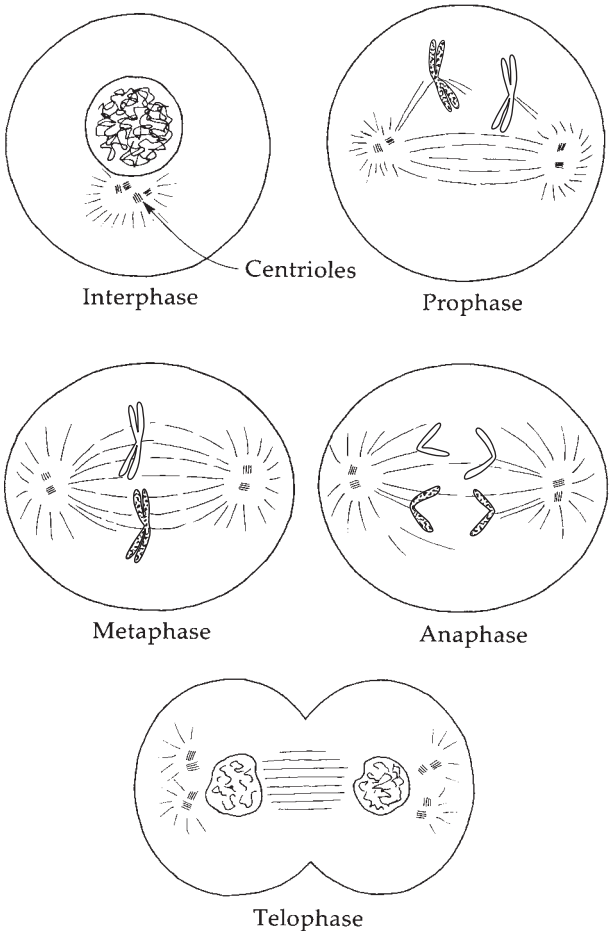


Figure 26-11 Mitosis. Illustrated for a cell with one homologous pair of chromosomes. After Mazia.[221]

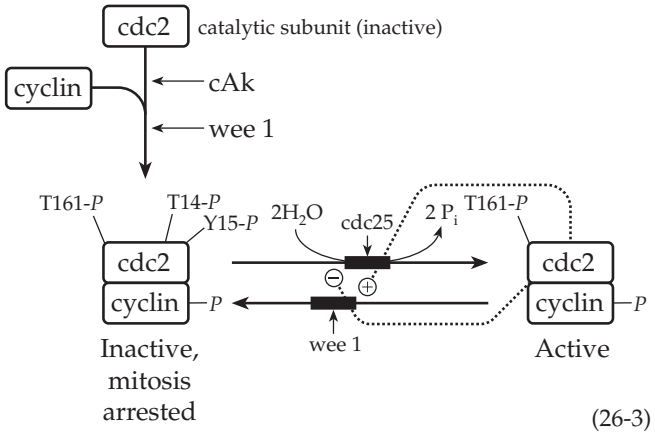
and fungi new plasma membranes and cell walls are constructed through the center of the cell. A partitioning of mitochondria and of Golgi components between cells must also occur.[220c]

The foregoing description overlooks the extreme complexity of mitosis, each stage of which must occur with precision and in the correct sequence.[222–225a] The replication of DNA, which takes place in the S phase of the cell cycle (and is discussed in Chapter 27) must be completed before mitosis begins. This is followed by condensation of the DNA into chromosomes (Chapter 27), breakdown of the nuclear membrane,[226,226a] assembly of the **kinetochores** by which the chromosomes attach to the spindle,[222] assembly of the spindle, attachment of chromosomes to the spindle, segregation of the chromosomes to opposite poles in anaphase, and finally the cleavage of the cell.

Cyclin-dependent kinases. As is shown in Fig. 11-15, the cell cycle is controlled by a series of complexes of the 30- to 45-kDa proteins called **cyclins** with **cyclin-dependent protein kinases** (**CDKs**). These kinases contain ~300-residue catalytic cores that resemble protein kinase A (Fig. 12-32). Like that kinase they transfer phospho groups from ATP to serine and threonine side chains of target proteins.[227–228] The kinases are inactive until a complex with the appropriate cyclin is formed and is activated by phosphorylation.[229] One or more additional kinases are required for this activation. Each stage in the cell cycle is controlled by one or more different cyclin–CDK complexes (Fig. 11-15). One of the best known CDKs is human **CDK2**, which functions in a complex with cyclin A during the S phase of the cycle.[228,230–232] Binding of the cyclin greatly alters the conformation of CDK2, opening the catalytic cleft and exposing threonine 160. Its hydroxyl group can be phosphorylated by the action of the **CDK-activating kinase** (**CAK**) with a 100-fold increase in catalytic activity.[228] Control of the CDKs also depends upon proteins that act as specific inhibitors and upon precise elimination of both cyclins and inhibitors via the ubiquitin system.[228a,b,c]

Checking for completion of replication and for DNA damage. The first checkpoint in the cell cycle is the **start** or **G_1 DNA damage checkpoint** (Fig. 11-15). Replication of DNA does not begin until the cell has had time to repair as much damage to DNA as is possible.[228d] As mentioned in Box 11-D, the cancer suppressor protein p53 is an essential component of the checking process.[224] A second checkpoint, the **G_2 checkpoint**, at the end of the G_2-phase (Fig. 11-15) requires verification that all DNA of all chromosomes has been replicated, checked for damage, and repaired if necessary. Control of these processes is accomplished by a mechanism first identified in

fission yeasts, such as *Schizosaccharomyces pombe*, and which has been conserved in metazoa.[223,233–236d] The CDK known as **cdc2** (fission yeasts), CDC28 in budding yeast, or CDC2 or CDK1 in mammals, functions from G_2 through the DNA damage checkpoint to mitosis (Fig. 11-15) and is sometimes described as the master cell-cycle switch.[236a] The catalytic subunit of cdc2, a serine / threonine protein kinase, is often called $p34^{cdc2}$. It is inactive unless complexed with a suitably phosphorylated cyclin (Eq. 26-3) and phosphorylated on Thr 161 by the action of kinase CAK (Eq. 26-3). However, when cdc2 becomes phosphorylated on Tyr 15 (in *Schizosaccharomyces pombe*) by the action of protein kinase wee 1 (Eq. 26-3) and on Thr 14 by another kinase, cdc2 is inhibited and mitosis is arrested at the G_2 checkpoint. This allows time to verify completion of replication as well as for repair before the replicated DNA strands are separated in mitosis. Hydrolytic removal of the phospho groups from Tyr 15 and Thr 14 of cdc2 by the action of phosphatase cdc25 then allows mitosis to begin.[223,237,237a] This phosphatase is also controlled by a phosphorylation-dephosphorylation cycle. The checkpoint kinase **Chk1** phosphorylates and inhibits cdc25.[238,238a]



(26-3)

When damaged DNA is present, protein p53 accumulates, just as at the G_1 checkpoint, and activates transcription of the seven phosphoserine-binding proteins of the 14-3-3 family.[224] These bind to the phosphorylated cdc25 phosphatase preventing activation of cdc2 by dephosphorylation.[224,238–240] In fission yeast one of the 14-3-3 proteins, **Rad 24**, apparently binds to the phosphorylated cdc25 and induces its export from the nucleus preventing mitosis.[241] Dephosphorylation of cdc25 and return to the nucleus allows mitosis to occur. A somewhat different regulatory mechanism is used by the budding yeast *Saccharomyces cerevisiae*.[242] In animal cells there are three cdc25 isoforms.[240,240a,b] In each case we probably have only a glimpse of a very complex process of checking for DNA damage and repair (see also Chapter 27).

Mitotic spindle formation and the spindle assembly checkpoint. Separation of the two copies of each replicated chromosome depends upon the spindle fibers. Their formation is preceded by replication of centrioles, if present, and formation of the two poles of the cell.[243–244b] Both γ-tubulin (Fig. 7-34) and acidic Ca^{2+}-binding proteins called **centrins** are involved.[244c–e] The microtubules appear to grow outward from the poles with their minus ends at the poles and their plus ends (Fig. 7-33) available for binding to the kinetochores, specialized protein complexes that assemble around the centromeric DNA (Chapter 27).[245–247] Each chromosome has two kinetochores, one for each daughter chromatid. These must be attached to spindle fibers coming from opposite poles. It has usually been assumed that random encounters of microtubule plus ends with kinetochores leads to correct linkage.[245,246] However, neither centrosomes or kinetochores are always essential to spindle assembly, and self-assembly occurs by motor-driven sorting according to polarity of the microtubules.[246,247] In addition to microtubules formation of spindles may require specialized matrix proteins.[247a]

The spindle isn't formed until replication is complete. A small G protein called **Ran**, a relative of Ras (Fig. 11-7), regulates spindle formation. Ran, in turn, depends upon a nucleotide exchange factor called **RCC1** (**regulator of chromatin condensation**).[248–250] RCC1 may signal that the replicated DNA is folding into chromosomes indicating that replication is complete. At the **spindle assembly checkpoint** (Fig. 11-15) the cell verifies that the metaphase spindle has been assembled correctly.[250a] All of the microtubules that pull the sister chromatids toward one pole or the other must be correctly attached to a kinetochore.[220,250b] In addition, there are interdigitated microtubules coming from both poles. Specific motor molecules (Chapter 19) then push the two poles apart. During the assembly and complex movements of the spindle both cytosolic dynein and four different types of kinesin-like motor molecules are required.[251,252] One of these is Kar3 (see Fig. 19-17). After assembly of the spindle is complete, a signal must be sent to the mitotic apparatus to move into anaphase. A clue to the nature of the signal has come from the observation that a single kinetochore lacking a spindle fiber connection causes arrest of mitosis.[253] Apparently unattached kinetochores send a "wait" signal, perhaps via the cytoskeleton.[253–255]

Anaphase. After the spindle has been checked, a sudden loss of cohesion between the sister chromatid pairs allows them to move toward the opposite poles. This process is catalyzed by the **anaphase-promoting complex** (**APC**, or cyclosome) and its activator protein **Cdc20**, a large multiprotein complex.[225] The APC also promotes proteolytic breakdown of cyclins and other

proteins by a ubiquitin- and proteosome-dependent mechanism.[225,256–259c] Its E_3-ubiquitin ligase (Box 10-C) targets the mitotic cyclins and other proteins for destruction.[260] A specific E_2 ubiquitin-conjugating enzyme (Box 10-C) is required for degradation of cyclin B and exit from mitosis.[261] The centrosome also plays an active role in cytokinesis, the final step in cell division.[261a]

The cell cycle encompasses so many different processes that it is clearly impossible to describe it by the single diagram of Fig. 11-15 or by the text written here. The cycle is influenced by a host of growth factors and external stimuli, many of which act on transcription of cyclins and other essential proteins. Transcription factors such as those of the Fos / Jun (AP-1) family in response to the MAP cascade (Fig. 11-13) are among those that control the transcription of cyclins.[262–263c] However, during mitosis most transcription of any genes is repressed.[264]

Among other factors influencing the cell cycle is the size of the cell and the availability of nutrients including purine and pyrimidine nucleotides.[263b,c,264a] Lack of cholesterol decreases the cdc2 kinase activity and causes apoptosis.[265] A cell cycle regulator in *S. pombe* known as **suc1** is essential for cell cycle progression. Although its three-dimensional structure is known, its function (like that of its human homolog CksHs2) is uncertain.[266,267]

2. Meiosis

The mechanism by which chromosomes are distributed during the formation of **gametes** (egg and sperm cells) is known as meiosis (Chapter 1; Fig. 26-12). Formation of gametes involves a halving of the chromosome content of a cell, each gamete receiving only one chromosome of each homologous pair. Genes found in the same chromosome are said to be **linked** because of their tendency to be passed together to the offspring. Genes present in different chromosomes are not linked, and their inheritance follows the pattern of **random segregation** established in Mendel's famous studies.

The simple fact that the genetic material is put up in several different packages (chromosomes) is sufficient to provide for considerable mixing of genetic information between different individuals in sexual reproduction. However, it doesn't provide a means for exchanging genes on the same chromosomes. Mixing of genetic information within chromosomes occurs by genetic recombination occurring during **crossing-over**, an aspect of meiosis with an essential biological role. In the S phase preceding meiosis, DNA is duplicated just as it is prior to mitosis. This provides sufficient genetic material to produce *four haploid cells*. These are formed during meiosis by two consecutive cell divisions (Fig. 26-12). Crosssing-over occurs prior to the first of these divisions, at the four-strand stage. The two homologous chromosomes of a pair come together to form what is called a **bivalent** or **tetrad** made up of four chromatids. For each chromosome, at least one chromatid is seen to come into intimate contact with a chromatid in the other homologous chromosome at points known as **chiasmata** (Fig. 26-13). During metaphase of the first meiotic cell division the homologous chromosomes, each still containing two chromatids, separate. Each chromatid now carries with it some genetic information that was previously found in the other member of the homologous pair and vice versa (Fig. 26-12). Now, without further replication of DNA in the second meiotic cell division, the chromatids separate to form haploid cells.

The process of crossing-over provides a means by which genes that are linked on the same chromosome can be separated, providing offspring with mixtures of genetic traits other than those predicted by simple Mendelian theory. The effects of crossing-over were first studied extensively by T. H. Morgan with the fruit fly *Drosophila melanogaster*. Morgan discovered his first mutant, a white-eyed fly,[267a] in 1910. The first genetic maps were made by assuming a direct relationship between the frequency of crossing-over and the linear distance between genes in a chromosome. Thus, the same approach to genetic mapping that was used later with *E. coli*, i.e., the measurement of recombination frequencies, was applied much earlier to crossing-over in the chromosomes of *Drosophila*. Extensive genetic maps involving many mutations were obtained for the four chromosomes of this organism, and similar techniques have been applied to many other organisms. The unit of distance in these chromosome maps is the **morgan** (named for T. H. Morgan[2,267a]). One centimorgan is the distance that allows recovery of 1% of recombinant progeny.[2] In the human chromosomes, this is ~1000 kb (1 Mb).

Biochemical and genetic studies of meiosis have been conducted in many organisms including fission[268,269] and budding[270–271a] yeasts, *Drosophila*,[272,273] starfish,[274] *Xenopus*,[275] and the mouse.[276] Meiosis can be viewed as a modification of mitosis but with the added initial step of crossing over and recombination. In addition, the S-phase of the cell cycle is absent in the second meiotic division. As was mentioned in Chapter 1, meiosis may occur at different stages of the life cycle of organisms. An important advantage in using fungi for genetic studies is that, like prokaryotes, they are haploid during much of their life cycle. Biochemical defects such as the inability to synthesize a particular nutrient can be recognized readily at this stage. At the same time genetic crosses can be made, and crossing-over frequencies can be measured and used for genetic mapping. The onset of meiosis may

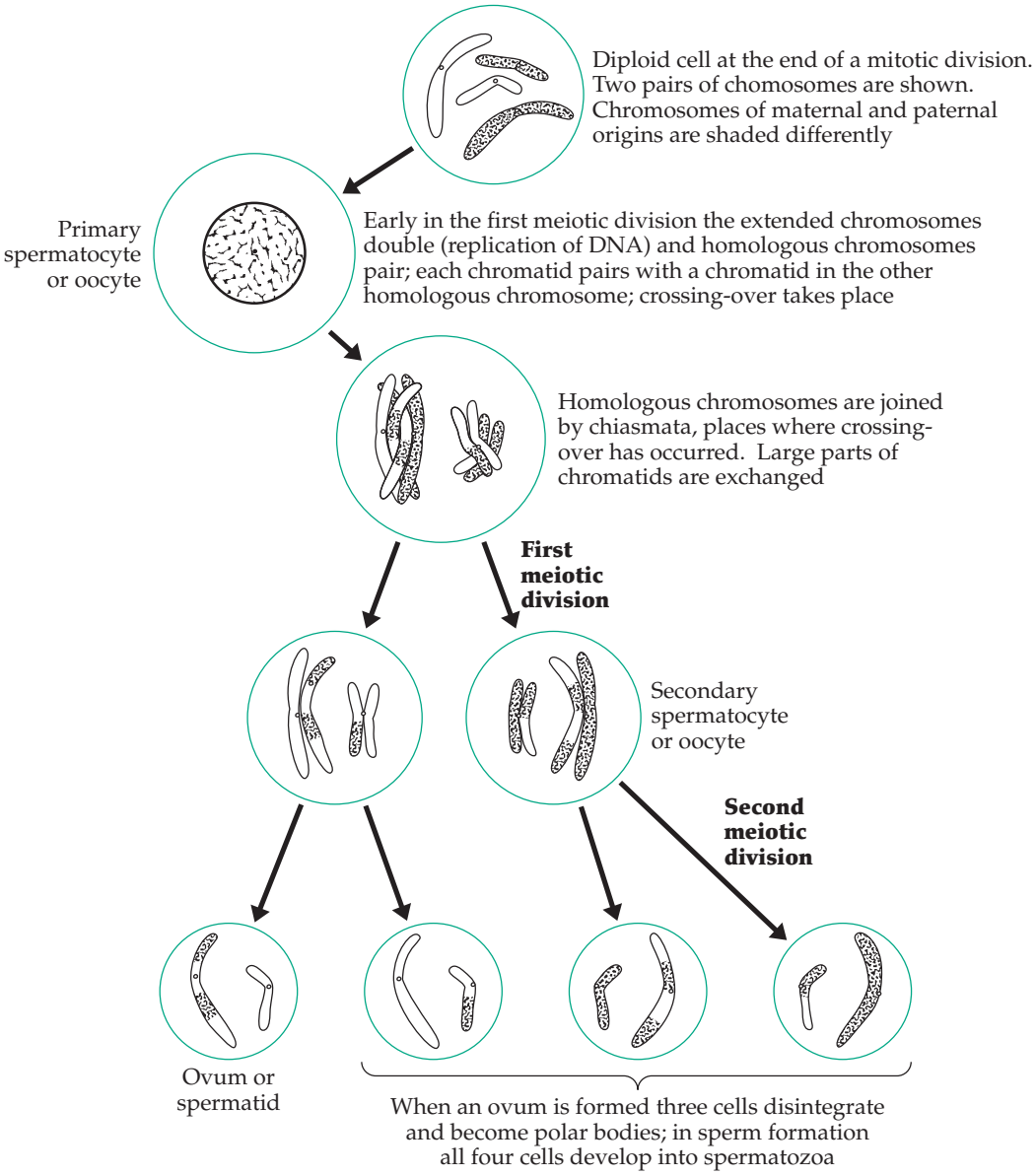


Figure 26-12 Meiosis. Cell division leading to formation of haploid gametes.

vary in time with diploid organisms. Meiosis in males may occur quickly. However, after chromatid pairing and recombination meiotic divisions of oocytes are arrested at the G_2 stage, in some species for many years.[274,275] The arrest is ended by hormonal stimulation, e.g., via progesterone acting through a cyclin B–cdc2 complex.[276] **Meiosis-activating sterols**, intermediates in lanosterol metabolism (Fig. 22-8) also accumulate and are thought to participate in control of meiosis.[276a] In haploid strains of fission yeasts sexual development is induced by starvation, especially of nitrogen. Cells of opposite mating types then fuse to form zygotes, which usually undergo meiosis immediately. Starvation is apparently signaled by the **cyclic AMP–protein kinase A cascade** (Fig. 11-13).[268]

During the prophase of the first meiotic division (meiosis I) two homologous pairs of partially "condensed" chromosomes must find each other and pair with appropriate orientation. A protein in the telomeres of the chromosomes seems to be involved.[269,277] The key structure in meiotic crossing-over is the ribbonlike **synaptonemal complex** formed by the pairs of homologous chromatids.[271,278–279b] This complex, in which a proteinaceous core or **axial element** separates the greatly extended chromatid pairs (Fig. 26-13), is fully formed in the **pachytene stage** of meiosis. Formation of the synaptonemal complex is preceded by development of a few double-stranded breaks in

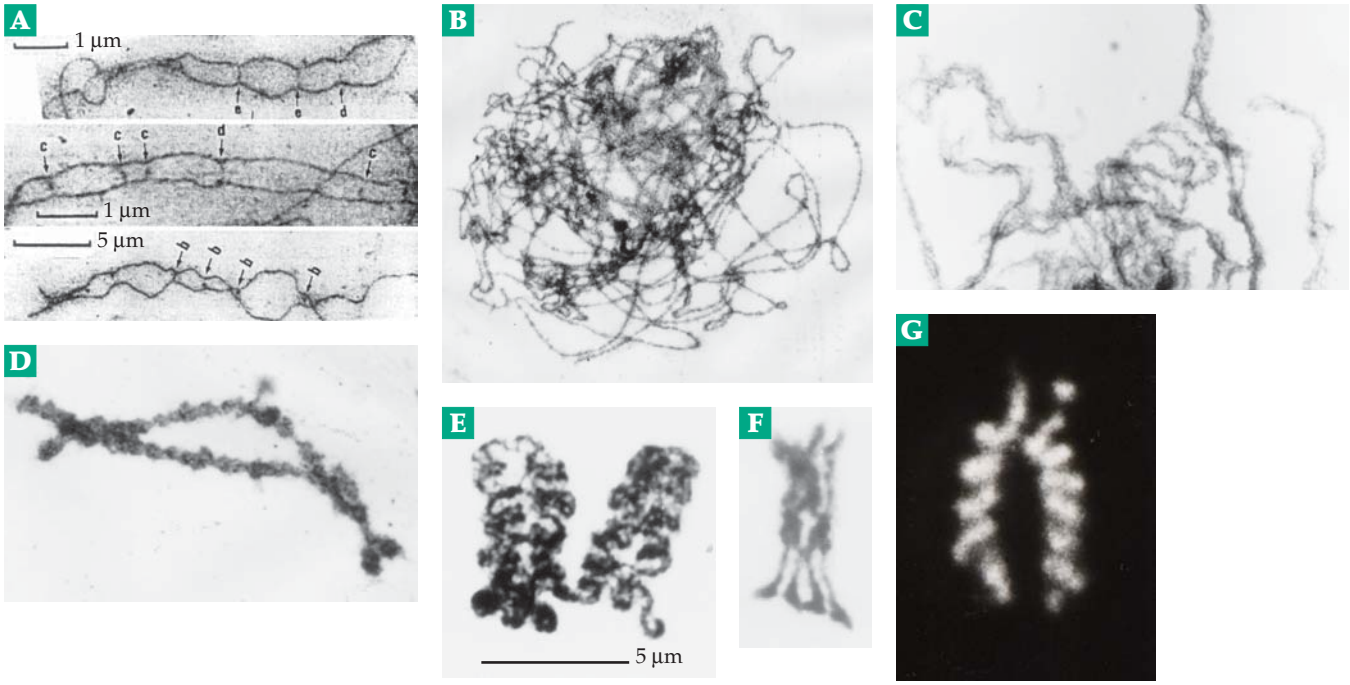


Figure 26-13 Synaptonemal complexes. (A) Aligned pairs of homologous chromatids lying ~0.4 µm apart in *Allium cepa*. Arrows indicate "recombination nodules" which may be involved in initiating formation of crossovers. Portions of meiotic chromosomes of lily are shown at successive stages: (B) Pachytene. (C) Portion of diplotene nucleus. (D) A bivalent at diplotene. (E) Two bivalents at diakinesis. Pairs of sister chromatids are coiled with appropriate handedness. (F) Sister chromatid cores are far apart in preparation for separation. A chiasma is present between the two central strands. (B) through (F) courtesy of Stephen Stack.[279,279d] (G) Pair of sister chromatids coiled with opposite handedness at metaphase. These are immunostained with anti-topoisomerase II antibodies. From Boy de la Tour and Laemmli.[280] Courtesy of U. K. Laemmli.

the DNA.[278,279c] The 3' ends of the DNA chains then invade the homologous chromatids to initiate the exchange via **Holliday junctions** (Fig. 5-28 and Fig. 27-11). The details of the recombination process are considered further in Chapter 27. The first meiotic division is a long process, for example, lasting ten days in mouse spermatocytes.[281] As cells pass to the metaphase stage of meiosis I, the chromosomes become much more compact, but the attachments between the homologous chromatids are still visible as chiasmata (Fig. 26-13). The chromatids then separate, the two homologs appearing as coils of opposite handedness. Within these coils the pairs of sister chromatids continue to be held together at their centromeres through metaphase and anaphase of meiosis I. One or more specific proteins, which must release their hold in meiosis II, are required.[270,277] A leading cause of embryonic and fetal death as well as severe mental and physical problems after birth is incorrect segregation of chromosomes during meiosis. If a gamete contains two copies of any chromosome the embryo will have three. Down syndrome, trisomy of chromosome 21, is the most frequent example. The spindle assembly checkpoint during meiosis I may sometimes be at fault.[281a,b]

3. Polytene Chromosomes

While most cells of higher organisms are normally diploid, the chromosome number may sometimes be doubled or increased even more. A cell with twice the diploid number of chromosomes is **tetraploid,** and with higher multiples of the haploid number it is **polyploid**. Plant breeders have succeeded in producing many tetraploid varieties of flowering plants often with increased size. One tetraploid mammal, the red viscacha rat, is known.[281] While most of our body cells are diploid, we, too, have polyploid cells. For example, some are always found in the liver. The most spectacular example of an increase in the normal DNA content of cells is provided by the giant **polytene chromosomes** of dipteran (fly) larvae. The DNA of cells in the salivary glands and some other parts of these organisms doubles as many as 11–14 times without cell division to give a several thousandfold (i.e.,

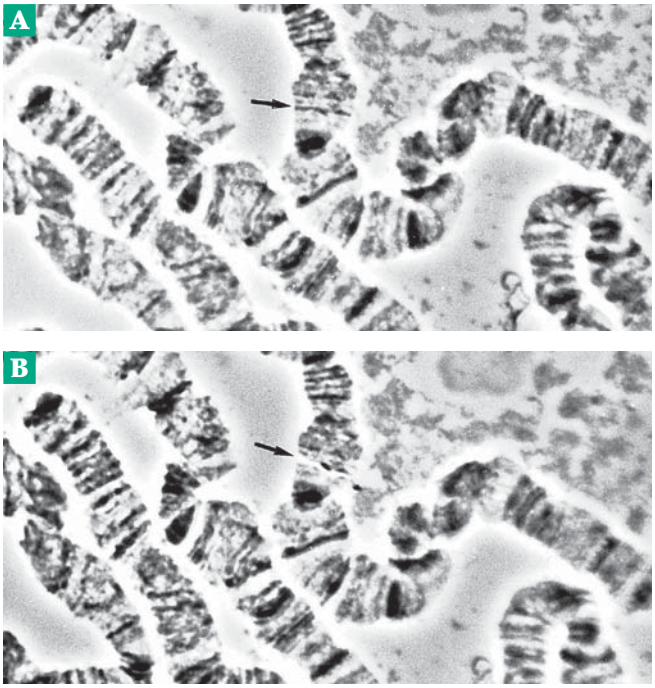


Figure 26-14 Microdissection of a *Drosophila* salivary gland chromosome. (A) Before the cut. (B) After cutting a medium size band. The arrow indicates the site of the cut. From Pirrotta.[282]

2^{14}-fold) increase. The supercoiled, duplicated DNA molecules all line up side by side in a much more extended form than in ordinary chromosomes. They can be seen readily by light microscopy. The total length of the four giant chromosomes of *Drosophila* is ~2 µm, compared to 7.5 cm in diploid cells. The giant chromosomes have a banded structure, ~5000 bands being visible along the length of the chromosome (Fig. 26-14). An average band contains ~36 kb of DNA in each of the strands. Since it has been possible to correlate visible changes in the appearance of these bands with particular mutations in the DNA, study of polytene chromosomes provided a second important method of mapping genes of the fruit fly. The maps produced by the two methods agree well.

Another use of polytene chromosomes is microdissection of DNA for cloning (Fig. 26-14). A piece of DNA containing 100–400 kb can be cut out of any desired spot, cleaved with restriction enzymes, and cloned.[282] Since it has been estimated that *Drosophila* may contain only ~9000–17,000 genes there may be 2–3 genes per band in these chromosomes.[283] The technique has been extended to human and other mammalian chromosomes.[284]

4. Cytoplasmic Inheritance

Not all hereditary traits follow the Mendelian patterns expected for chromosomal genes. Some are inherited directly from the maternal cell because their genes are carried in the cytoplasm rather than the nucleus. There are three known locations for cytoplasmic genes: the mitochondria, the chloroplasts, and certain other membrane-associated sites.[285,286] An example of the last is found in "killer" strains of yeast. Cells with the killer trait release a toxin that kills sensitive cells but are themselves immune. The genes are carried in **double-stranded RNA** rather than DNA, but are otherwise somewhat analogous to the colicin factors of enteric bacteria (Box 8-D). Similar particles (κ factors) are found in *Paramecium*.[287]

Mitochondrial and chloroplast genes are discussed in Chapters 18 and 23, respectively.

G. The Human Genome Project

The discovery of site-specific restriction endonucleases in 1970 and the development of efficient DNA sequencing methods in 1977 sparked a revolution in biology. On Oct 1, 1990, the 15-year project to sequence the complete human genome was officially begun.[288–290] The project is far ahead of schedule. "First drafts" of the genome were published in 2001.[290a–e] A more complete version will include **annotation**, a listing of predicted exons, mRNA transcripts,[290f] functions of recognizable genes, etc.[290e] Such documentation was provided first for chromosome 22[290e] and is now available for other human genes.[290g] It was a surprise to discover, at least initially, that only 30,000–40,000 protein-coding genes could be recognized.[290a,290h] Earlier 50,000–150,000 genes had been predicted (Table 3-1). A nearly complete version of the human genome is anticipated for spring of 2003.[290i]

Included in the genome project are completion of the sequences for *Caenorhabditis elegans*,[290j] *Drosophila melanogaster*, and the mouse (*Mus musculus*).[291,291a–e] However, substantial difficulties remain in filling hard-to-sequence gaps, and some new approaches may need to be developed.[292] More recently the project has been extended to include many additional species. Among them are the rat,[292a] the zebrafish,[292b] the pufferfish *Fugu rubripes*,[292c] the human malaria parasite *Plasmodium falciparum*,[292d,e] the rodent malaria *Plasmodium yoelii, yoelli*,[292f] and the mosquito *Anopheles gambia*, which carries the parasite.[292g,h] Plant genome sequences are also being determined (p. 1511).

1. The Mammalian Genome and Human Health

Human beings and other mammals all have about

3500 Mbp of DNA containing perhaps 40,000 protein-coding genes. However, ~ 97% of the genome is repetitive DNA and other DNA that, at present, has no recognizable function. The genome project will reveal all of these sequences and will doubtless provide us with many surprises.

A primary goal of the genome project is to understand the relationships between gene sequences and human diseases and health. Until recently little was known about the locations or structures of genes in human chromosomes. Human genetic diseases provided the first clues. Although systematic genetic experiments cannot be done with human beings, almost the entire population is under some kind of medical care. Genetic defects are being detected more and more often, and the inheritance within families is traced with increasing frequency. Consequently, a huge body of knowledge of mutations to the human genome is available.[293] By 1995 ~4500 human genetic disorders had been discovered.[294] Many more have been found since then.

The first mapping of human genetic defects came with the recognition that sex-linked traits are encoded on the X-chromosome. Some linkage analysis was also possible from studies of inheritance within families. For example, among individuals who have two X-linked traits, e.g., color blindness plus one other, naturally occurring crossing-over occasionally breaks the linkage in some individuals within the family. **Somatic cell fusion**[191] provided an additional approach. Human lymphocytes can be fused with rodent cells under the influence of inactivated Sendai virus, which causes the cells to adhere and then to fuse. From such human–mouse or human–hamster hybrid cells, strains in which the nuclei have also fused can be selected. Although such cell lines can be propagated for many generations, they tend to lose chromosomes, especially those of human origin. By observing loss of particular human enzymes or other proteins (separable from the hamster enzymes by electrophoresis), it was possible to assign genes to specific chromosomes. This also required identification of the chromosomes lost at each stage in the experiment. New staining techniques made it possible to identify each of the 22 pairs of human autosomes as well as the X and Y chromosomes (Fig. 26-15).[295] Using a variety of techniques, 500 loci in human chromosomes had been mapped by 1983.[296]

Separation of individual human chromosomes with a fluorescence-activated cell sorter (Box 3-B)[299] permitted the preparation of libraries of cloned DNA from individual chromosomes. Fragments, of an average length of 4 kbp, from digestion by *Eco*R I or *Hin*dIII were packaged into the Charon 21A cloning vehicles and "amplified" by culturing infectious phage particles in *E. coli*.[282] However, this approach was inadequate for sequencing the entire genome. By 1987 the use of yeast artificial chromosomes (YACs)[300] and later bacterial artificial chromosomes (BACs)[301,302] and radiation hybrid mapping[104,303] provided the means for sequencing the 34- to 260-Mbp human chromosomes.[288,304] Also essential was the development of high-speed automatic sequencing machines[305–305b] and of computers adequate to assist in compiling the sequences.[302,306,306a] Sets of overlapping and redundant clones define a continuous sequenced segment that is called a **contig**. As the contigs are enlarged the gaps are filled.

In 1987, the first global genetic linkage map of the human genome was published.[288,307] It was based on 403 polymorphic loci, of which 393 were RFLPs studied in 21 three-generation families. By 1996 a linkage map containing 5264 loci, in the form of polymorphism in **short tandem** repeats $(AC/TG)_n$ represented by **microsatellite DNA**, was available.[304] Also mapped by 1998 were 2227 **single nucleotide polymorphisms** (SNPs), where the two alternative bases occur with a frequency >1%.[308] This was increased to 1.4 million by 2001.[308a] As many as 5.3 million common SNPs may occur, each with a frequency of 10–50%. On the average there may be one SNP for every 600 base pairs.[308b] These SNPs account for a large fraction of the diversity in human DNA. By 1995 a physical map with >15,000 **sequence tagged sites** (STSs) with an average spacing of 199 kb had been constructed.[303,309] An STS is defined as a cloned fragment in a YAC (or BAC) library that has been amplified by PCR and tested to establish that it contains a known locus. DNA sequences known as **expressed sequence tags** have also been mapped (see p. 1490). Using STSs, RFLPs, ESTs, and the growing contigs, more than 16,000 human genes had been mapped by 1996.[310] By 1998 more than 1,060,000 ESTs had been reported.[311] In 1999, the nucleotide sequence of the smallest human chromosome, the 33.4-Mbp chromosome 22, was completed. It contains at least 545 genes and 134 pseudogenes.[312] In the entire genome >1000 seed contigs had been assembled, and completion of the first phase of the Human Genome Project was in sight.[313] In 2001 a whole-genome clone-based physical map was published.[313a]

2. Understanding Gene Sequences

The vast array of sequence data coming from the human genome and from genes of other species are deposited, as they are reported in the scientific literature, in the Human Genome Central. It can be found on the Web at the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov/genome/guide and European Bioinformatics Institute (EBI) at http://www.ensembl.org/genome/central/. The data have been doubling every 15 months. By

2000 almost 1.3 million human ESTs were included as were sequence data on more than 25,000 other species.[311,314] A problem is that there are many errors in the data with some seriously incorrect sequences.[315] It is hard to manage the mass of new data, but in time most of the errors will probably be corrected. Books[316–319] and computer programs[320,321–321c] are available to help understand genomes. The widely used programs **BLAST** and **FASTA**, available on the Web or in computational packages, routinely compare protein-coding genes with known genes in order to predict function.[322–325] An inexpensive high-powered desktop computer and an internet connection will enable a person to do complex biocomputing.[325a] See also Chapter 5, Section H,7.

Genes must be transcribed, and most transcripts must be spliced, modified, and translated by the ribosomal machinery. Genes cannot be fully understood without considering these processes, which are dealt with in Chapters 28 and 29.

Human variation. Between any two unrelated human beings there are on the average one base pair that is different out of every 500–1000 nucleotides. This amounts to $\sim 4 \times 10^6$ differences in the whole genome.[305,326] In addition to these single nucleotide polymorphisms (SNPs) there are many differences in the >100,000 nearly randomly dispersed short tandem repeats of microsatellite DNA.[189,327] The latter form the basis of DNA fingerprinting (Box 5-D), and together with the SNPs are helpful in tracking genetic diseases. One of the more difficult goals is to identify genes that confer increased susceptibility to cancer and to the complex syndromes of diabetes and mental problems. Searching for correlations with SNPs has proved difficult.[327a,b] A new approach termed **haplotype mapping** may permit correlations of disease susceptibility with larger blocks of conserved DNA sequence known as haplotypes.[327c,d] Some controversy surrounds possible uses of these maps, as also is the case for proposals to conduct extensive genetic testing of populations including newborn infants.[327e,f,g] Another planned project is the sequencing of 1000 individual human genomes.[327h]

Studies of DNA have also shed light on human evolution.[328] Mitochondrial DNA,[329] as well as a variety of nuclear genes,[330] is being studied in attempts to establish approximate dates of evolutionary divergence. If we assume a constant rate of incorporation of nucleotide substitutions during evolution, we can use sequence data as a **molecular clock** to construct phylogenetic trees such as that of Fig. 1-5, which is based on the gene for 16S ribosomal RNA. A problem is that mutation rates of genes are not all the same, and there are not yet enough data to draw firm conclusions. Estimates based upon the maternally transmitted mitochondrial DNA (see Chapter 18) suggest that we are all descended from Africans who lived 100,000 to 200,000 years ago. Limited data from the Y chromosome, which is transmitted through males, agree with the mitochondrial DNA data.[328,328a] However, study of the pyruvate dehydrogenase gene has been interpreted to indicate a more ancient divergence of African and other peoples.[331] Analysis of DNA from a fossil bone shows that the mtDNA of Neandertal people differed from modern mtDNAs in an average of 26 positions. From this figure the mitochondrial DNA molecular clock predicts that the extinct Neandertal line diverged from ours ~550,000 to 690,000 years ago. The most ancient mtDNA found is Australian.[331a,b] Unfortunately, DNA in fossils is very unstable and has not been recovered from older fossils, e.g., from dinosaurs. However, the search for ancient DNA goes on.[331c,d]

We diverged from our closest ape relative, the chimpanzee, about 4–6 million years ago.[196,332] Chimpanzee and bonobo DNA sequences are ~98.8% identical to those of humans.[332a–e] If differences in inserted and deleted segments of genes are included, however, the identity drops to 95%.[332f] One of the longest DNA sequences to be compared among humans and apes is a 7-kb length around the pseudogene in the β globin cluster (Fig. 27-10). In this sequence chimpanzees are closest to humans with gorillas being the next closest.[195] We may well ask in what way we differ from these apes? Some specific differences have been found. Notably, human beings do not hydroxylate the glycolyl groups of sialic acids (Chapter 4) to form *N*-glycolylneuraminic acid residues on glycoproteins.[328,333] Could this really be the most important difference between us and the apes? More genomic analysis may tell.

What does DNA analysis tell us about race? Most investigators conclude that there is only one human race with no detectable boundaries between the group commonly referred to as races.[334–336] As Pääbo put it, "in terms of the variation at most loci, we all seem to be Africans, either living on that continent or in recent exile."[328] The differences in skin color seem to reflect adaptation to the environment in which people live.[336a] Variations in the level of melanocyte-stimulating hormone receptor, one regulator of skin color (Box 25-A), are especially high in Africans.[328]

DNA analysis has also been useful in tracing human migration.[337–338e] For example, a genetic marker in the Y chromosome is carried by 85% of native Americans, suggesting that they are all descended from a man who lived ~20,000 years ago, probably an immigrant from Siberia.[339] Contrary to usual assumptions women, more often than men, seem to have spread their DNA to new locations in the world.[327b] Studies of cattle and of the wild ox reveal information about domestification of these animals about 10,000 years ago in Europe, Asia, and Africa.[340]

Other evolutionary relationships. Studies of chromosome banding, chromosome maps, restriction fragments, and detailed sequences provide many insights into relationships among species. For example, the chromosome banding pattern and also DNA sequences show close similarities between human beings and the mouse.[194,312] The latter is often regarded as the premier organism for the study of mammalian genetics and development.[291a–c,341] Dense genetic maps are available for both the mouse and rat as are moderate-resolution maps for livestock, companion animals, and other mammals.[327,342,343] Comparative gene maps are being constructed for more than 40 mammals[344] and other species of animals, plants, and fungi. Comparisons of these genomes reveal much of interest. For example, the pufferfish *Fugu rupripes* has a genome only 1/9 the size of the human genome. However, both species seem to contain about the same number of genes. Many of them can be directly correlated and some human-disease-causing mutations have been identified first in the Fugu.[344a]

Evolutionary history is being rewritten in molecular terms. Comparison of sequences of individual proteins allows evolutionary relationships of their genes to be traced. Many families of **homologous** genes can be identified.[344b] These include both **orthologs**, genes in different species that have evolved from a common ancestor, and **paralogs**, genes related by gene duplication within a genome. Orthologs have the same function in different organisms, but paralogs have different functions within a single species.[345]

Sophisticated molecular clock studies suggest that the evolution of metazoan organisms began earlier than had been supposed. Ancestral primates appeared at least 65 million years ago.[344] Gene sequence data for many species suggest that a great variety of mammals lived 100 million years ago in the age of dinosaurs,[346] a view also supported by new fossil evidence.[344,347] Metazoans appeared earlier than the "Cambrian explosion" generally thought to have occurred ~550 million years ago.[348] New geochemical data suggest that cyanobacteria diverged from other bacteria as early as 2.1×10^9 years ago.[349] Gram-negative bacteria diverged from gram-positive microbes ~3.2×10^9 years ago,[350] *Salmonella* from *Escherichia* only 0.1×10^9 years ago. However, DNA analysis shows that within this latter time period many genes from other microorganisms have been inserted into the *E. coli* chromosome and into other bacterial chromosomes.[351,352] Some of these transfers have occurred with the help of bacteriophages.[353] A puzzle is the fact that among eukaryotic cells the enzymes catalyzing the genetic **information transfer** via transcription and protein synthesis resemble those of archaea. However, **operational enzymes** that catalyze other basic metabolic processes tend to be more similar to those of eubacteria such as *E. coli*.[353,354] So much gene transfer between organisms has occurred that it is difficult to establish the earliest parts of a phylogenetic tree of the type shown in Fig. 1-5. Another factor that confuses our study of bacteria is that less than 1% of all living microorganisms have been grown in pure cultures.[355,356]

We are still dangerously ignorant of the complexity of the microbial world, which both threatens us with diseases and sustains our environment. We do have complete genome sequences for more than 60 different bacteria with hundreds more expected within a few years.[356,356a] They contain from 0.58 Mbp (*Mycoplasma genitalium*) to 8.7 Mbp for the antibiotic-producing *Streptomyces coelicolor*,[356b] a size shared by the legume symbiont *Bradyzhizobium japonicum*.[356c] The latter is one of many bacteria that have genomes split into two or more parts, often a major chromosome plus one or more plasmids. For example, the 6.7-Mbp genome of *Sinorhizobium melitoti*, an alfalfa symbiont, comprises a 3.65-Mbp chromosome and 1.35- and 1.68-Mbp megaplasmids.[356d] The 5.67-Mbp genome of *Agrobacterium tumefaciens*, much used in genetic engineering (Section E,4), consists of a circular chromosome, a linear chromosome, and two plasmids.[146a–c] The genome sequence revealed a close similarity to those of the above-mentioned legume symbionts.

Sequences of many pathogenic bacteria are known.[356e–o] These include the causative agents of cholera,[356c] typhoid fever,[356f] plague,[356g] brucellosis,[356h] leprosy,[356i] tuberculosis,[356i] and anthrax.[356j] Also included is a virulent strain of *Streptococcus pneumonia* (respiratory infections, ear aches, meningitis),[356k] *Pseudomonas aeruginosa* (a common "opportunistic" pathogen),[356l] *Listeria monocytogenes*, which causes a severe food-borne disease (Box 19-C),[356m] the pathogenic *E. coli* 0157:H7 (see Fig. 1-2),[356n] and a tiny mucosal pathogen *Ureaplasma urealyticum*.[256o] Many surprises were found in the genome sequences. For example, *E. coli* 0157:H7 is related to the laboratory strain *E. coli* K-12, and the two strains share a ~4.1 Mbp common "backbone" sequence. However, the pathogenic strain has hundreds of "islands" of additional DNA spread throughout the genome and amounting to ~1.54 Mbp. Many of these carry genes associated with virulence.[356n]

Specialized features appear in virtually every genome.[356a] For example, the mycoplasma *U. urealyticum* contains genes for enzymes that allow the bacterium to obtain almost all of its ATP from hydrolysis of urea.[356o] Both *Caulobacter crescentus*[356p] and *E. coli* develop specialized structures involved in motility (Figs. 19-3, 32-1). *C. crescentus* and also spore-forming bacteria have alternative developmental plans (Fig. 32-1). At the low end of the genome size range are species of *Buchnera*, which are endocellular symbionts of aphids.[356q] Their genome is only slightly larger than

that of *M. genitalium*. The extremely salt-tolerant *Halobacterium*[356r] and the heat-tolerant *Thermoplasma volcanium*[356s] are among the archaea for which complete genome sequences are known. Species of *Xanthomonas*, whose genomes have been sequenced, are economically important plant pathogens. *Xanthomonas campestris* is also grown commercially to produce xanthan gum (p. 179).[358a] Determination of the genomes of mycobacteria has been challenging.[356i,t] Among these slow-growing organisms is *Mycobacterium tuberculosis*, the causative agent of human tuberculosis. A large fraction of its ~4000 coding genes is devoted to metabolism of lipids and to the synthesis of unusual proteins and lipids of its cell wall.[356t]

The larger genomes code for many proteins of unknown function, but over 80% of the "ORFs" (presumed genes) of *Haemophilus influenzae* have been identified, as have their presumed functions.[357,358] The encoded proteins appear to catalyze 488 metabolic reactions on 343 different metabolites. Together these systems provide a **metabolic genotype**.[357] Results of such analyses are accumulating in **metabolic databases**.[359,359a] What is the minimum number of metabolic reactions needed for support of life? Transposon insertions can inactivate all but from 265 to 350 of the protein-encoding genes of *M. genitalium* without killing the bacteria.[360] From comparisons of a variety of bacterial chromosomes it seems likely that ~256 of these genes are truly essential.[350,361] This conclusion leads to an interesting question. Is it ethical to now try to generate such a minimal bacterium?[361a] Are there hazards, e.g., that its escape might endanger our health or the environment? On the other hand, genetic engineering of bacteria, which is already practiced, can provide useful improvements in bacteria used in foods and in industry.[362,362a]

Modeling with the aid of data available on the World Wide Web is leading to development of new mathematical descriptions of metabolic networks.[359a] An ambitious new project is to model the entire *E. coli* cell. Many experimental data will be required and it has been estimated that ten years will be needed. The effort involves investigators in many laboratories and will be at least ten times as complex as the determination of the human genome.[362b]

Metabolic studies of eukaryotic cells. The yeast, *Saccharomyces cerevisiae*, contains ~6200 genes of which, until recently, only 40% had been assigned a function. Now a variety of methods are being employed to understand this little fungus.[363–365] A useful approach is to systematically inactivate or "knock out" genes. Davis and associates[363] used a PCR-based strategy to delete one gene at a time of 2026 yeast genes. Of these genes 1620 were found not essential for growth in a rich medium. Ross-McDonald and coworkers engineered a transposon, a 274-bp derivative of the *E. coli* Tn3, and allowed it to be inserted into the genes of yeast cells by homologous recombination (Chapter 27, Section D). The transposon carried DNA for a short peptide tag in the form of a specific immunological epitope that could identify the transformed cells. More than 11,000 strains with disruptions in nearly 2000 genes were obtained. These and other deletion mutants are now available for study.[364a,365] A second yeast, *Schizosaccharomyces pombe*, has assumed major scientific importance in studies of the cell cycle and of metabolism.[365a] Its 13.8-Mbp genome is only a little smaller than that of *S. cerevisiae*, but it has ~1400 fewer recognized genes, a total of 4824.[365b] There are smaller eukaryotic genomes. That of the tiny marine chordate *Oikopleura dioica* may be only 51 Mbp. However, there may be a total of ~15,000 genes.[365c]

Transposon-induced mutations have also been created in nearly one-fourth of the ~12,000 genes of *Drosophila melanogaster*.[283] Studies of the expression of genes in both *Drosophila* and in the nematode *Caenorhabditis elegans* are directed toward understanding of development and differentiation. Of the nematode's predicted 19,293 genes, only 7% have been studied at the biochemical level. To understand what happens during development we need methods for studying simultaneously the expression of all of these genes. One approach is to look at messenger RNAs that are formed at different times during development. More than 9000 mRNAs have been identified in cells of *C. elegans*, and their patterns of expression have been observed using DNA microarrays.[366,367] A similar technique has been applied to yeast. DNA sequences of fragments of ~6400 yeast genes were amplified by PCR and printed onto a glass plate to form a "DNA chip." From mRNAs formed at different times during growth, fluorescent cDNA copies were made, and their amounts were checked by use of the DNA chip.[368] Another approach is to look directly at the proteins formed. Walhout and coworkers have devised a large-scale automated system for cloning all of the genes of *C. elegans*, expressing the protein products, and testing them in the yeast two-hybrid system (Box 29-F) for protein–protein interactions.[369] Another project is to use large-scale sequence comparisons between proteins of *C. elegans* and of other organisms to identify nematode genes that encode extracellular matrix proteins involved in cell adhesion and to trace their evolution.[370]

The genomes and the metabolism of the two insects *Drosophila melanogaster* and *Anopheles gambia*[370a] can now be compared. Many differences can be seen but almost half of the genes are orthologs. Many of these can be related also to those of pufferfish, mice,[370b] humans, and other species.

Among plant genomes that of *Arabidopsis thaliana* has been studied most. The sequences of its five chromosomes have been determined and analyzed.[371–375c]

About 25,498 genes encoding proteins from 11 different families have been found.[375a] More than 14,000 ESTs were established from cDNAs.[375b] Many of the genes represent new families, some of which may be peculiar to plants. However, many others are homologous to those of *C. elegans* and *H. sapiens*. For example, developmentally important homeotic genes, marked by a **homeo sequence** as in animal genes, are present as are thousands of cell surface receptors. However, only Ser / Thr and histidine kinases are present in *Arabidopsis*. No tyrosine kinases have been identified.[375]

The ~125-Mbp *Arabidopsis* genome is tiny compared to the 3000 Mbp of DNA present in the genomes of maize and of many other plants.[374] However, the rice genome, only ~15% as large as that of maize,[376] has been chosen for complete sequencing.[375d,e] Draft sequences for the genomes of two subspecies of rice (*Oryza sativa L.*), ssp. *indica*[375f] and ssp. *japonica*,[375g] were published in 2002. The much larger genomes of other cereal grains (2500, 4900, and 16,000 for maize, barley, and wheat, respectively) will probably all be sequenced within a few years. The genes of maize are being mapped and studied using transposon-tagging

Figure 26-15 (A) Numbering and staining patterns (G-banding) of human chromosomes. The horizontal line marks the centromeres. Chromosomes 13, 14, 15, 21, and 22 have nucleolar organizers located at the constrictions in their short arms. From Alberts *et al.*[297] Adapted from U. Franke.[298] (B) Map of genetic defects identified in X chromosomes by 1995. From McKusick and Amberger.[294]

methods.[377] The smaller, more compact genomes may have just as many genes as the longer ones but have less repetitive DNA.[378] Some ferns have 307,000 Mbp of DNA, nearly 100 times that of a human; bony fishes have ~307,000 Mbp and the ameba >200,000 Mbp. These organisms appear to lose unneeded repetitive DNA faster than those with smaller genomes.[378]

3. Understanding Human Genetic Diseases

Genetic diseases have always been with us, but it was not until 1949 that the first disease, sickle cell anemia (Box 7-B), was understood at the molecular level. A single base substitution in DNA and the resultant single amino acid substitution in hemoglobin causes this disastrous disease. It was soon recognized that defects in single proteins are causes of other inherited diseases. Many of the hundreds of other known genetic disorders[201,203,379] are discussed elsewhere in this book. Among them are muscular dystrophies and cardiomyopathies (Box 19-A),[380] lysosomal deficiency diseases (Chapter 22), problems with ion transporters[381–383] and channels (cystic fibrosis, Box 26-A), defective collagens (Box 8-E),[384] neurological disorders (Chapter 30), and defects in defense systems (X-linked granulomatous disease, Chapter 18, Section G). Many of these were first recognized by their frequent occurrence in boys. Some of these X-linked deficiencies are mapped in Fig. 26-15B.

One insight into molecular disease was the recognition that mutations that cause many diseases, e.g.,

BOX 26-A CYSTIC FIBROSIS

One of the commonest of genetic diseases, cystic fibrosis affects persons all over the world. The incidence is unusually high in persons of northern European descent, about one in 2500 children being born with the defect. The inheritance pattern showed that cystic fibrosis is recessive and is caused by a single-gene defect that is carried by almost 5% of white Americans. In the United States there are ~30,000 persons with the disease. Many die in early childhood and even with careful treatment only 50% live into their late twenties or beyond.[a,b] Through extensive linkage analysis the cystic fibrosis gene was mapped to chromosome 7, and its location was narrowed further to a 1600-kbp region between the oncogene *met* and another marker designated J3.11.[c] Random searching located closer markers, and "chromosome walking and jumping" led to identification and characterization of the gene in 1989.[c,d] The large 250-kbp gene contains 27 exons. The transcribed mRNA is 6129 bp in length, and the gene product is a 1480-residue amino acid protein,[e] which is known as the **cystic fibrosis transmembrane conductance regulator** (**CFTR**).[f]

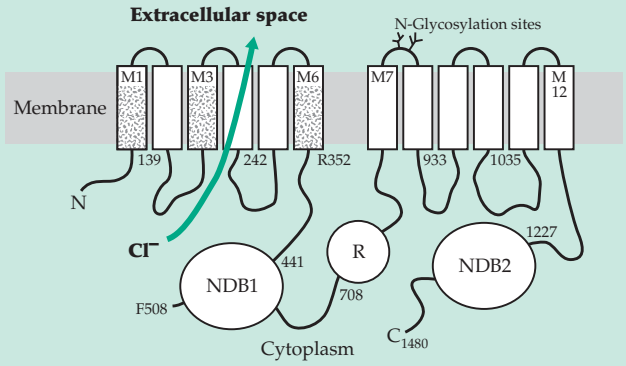
Children with cystic fibrosis lose excessive amounts of salt in perspiration and become dehydrated readily. A salty taste of the skin and an elevated chloride concentration of sweat are traditional diagnostic symptoms.[a] More serious problems arise from progressive respiratory failure and inadequate pancreatic secretion. Lung infections with *Pseudomonas aeruginosa* are the major cause of death.[g] The CFTR gene is expressed in many tissues, especially those of the mucous membranes. An alternatively spliced isoform may form chloride channels in heart.[h,i] As mentioned in Chapter 8, Section C,5 the CFTR protein is a member of the ATP-dependent ABC transporter family. However, it is atypical because it also contains a regulated chloride channel.[j] In secretory epithelia of intestines, pancreas, lungs, sweat glands, and kidneys Cl^- enters epithelial cells through their basolateral surfaces using an $Na^+ + K^+ + 2\,Cl^-$ cotransporter and exits the cells through their apical surfaces using the CFTR channel. Absorptive epithelia also contain both the cotransporter and the CFTR channel, but Cl^- flows into the cells from the exterior surface, and the distribution of the cotransporter and CFTR between basolateral and apical surfaces is opposite to that in secretory cells.[f]

From the amino acid sequence the CFTR protein is predicted to form two 6-helix transmembrane domains, two ~240-residue cytosolic ATP-binding domains, and a cytosolic regulatory (R) domain that contains at least five serine residues that may be phosphorylated by the cAMP-dependent **protein kinase A** (Chapter 11, Section C; Fig. 11-4).[k,l,m] See adjacent scheme. The two ATP-binding domains resemble those of myosin and other ATP-hydrolyzing proteins. The chloride channel, which is probably formed by helices M1, M3, and M6,[f] remains closed unless serine residues of the regulatory domain are phosphorylated. However, opening of the channels also requires binding of ATP in the nucleotide-binding domains. Binding of either vanadate or BeF_x stabilizes the open state of the channels.[k] There are four small 55- to 65-residue cytosolic loops, labeled 1–4 in the diagram. Transmembrane loop 3 also appears to function in regulation of the channel,[n] which involves a complex regulatory mechanism. Two sites of *N*-linked glycosylation are present in the short extracellular loop between helices M7 and M8.

cystic fibrosis, while affecting a single protein, occur at many places in the gene that encodes the protein. Not all mutations are caused by base substitutions; they are often a result of deletion or insertion of DNA. A whole group of diseases are caused by accumulation of repetitive DNA, often of nucleotide triplets, within genes or in control regions of genes. Another important insight is that understanding a newly discovered and very rare disease may help us to understand other related disorders. For example, after the Duchenne muscular dystrophy gene was located, the encoded protein dystrophin was found to have mutations resulting in other milder dystrophies (Box 19-A). Mutations are only rarely beneficial, but we know that many mutations alter the properties of proteins very little. We can anticipate that most genes may undergo mutations that cause some loss of good health and vitality without being diagnosed as causing disease.

We have also come to understand that many complex diseases such as diabetes, **polycystic ovary syndrome**,[385] **Crohn's disease** (inflammatory bowel disease),[386] and schizophrenia are in fact multiple diseases. Diabetes is a syndrome that can arise from causes such as defective insulin receptors or defective glucose transporters or from as yet unknown metabolic problems (Chapter 17).[387] Many cancers have a

BOX 26-A CYSTIC FIBROSIS (continued)



The CFTR protein can undergo endocytosis into clathrin-coated vesicles as part of its regulatory mechanism.[o] Since HCO_3^- is usually exchanged for Cl^- in epithelial ion transport, regulation of HCO_3^- uptake is also a significant aspect of CFTR function.[p]

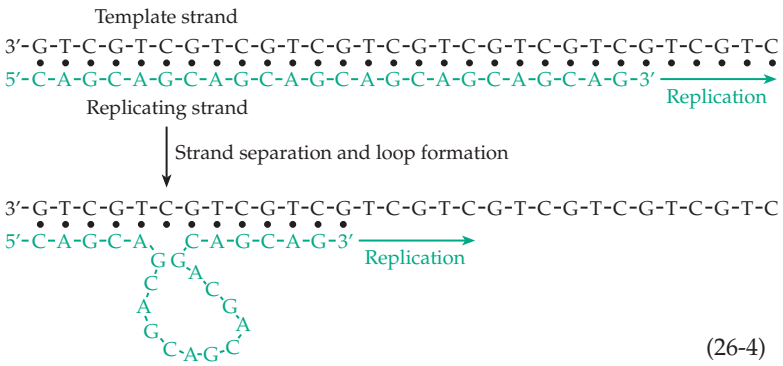
Mutations that cause cystic fibrosis are found at many locations in the gene. However, ~70% of the mutations are caused by the absence of phenylalanine 508, as a result of a three-nucleotide deletion, in the first nucleotide-binding domain.[e,k,p] This deletion causes misfolding of the CFTR to give an inactive protein.[p] Hundreds of other mutations, some in the regulatory domain and some in the cytosolic loops,[q,r] also cause the disease. Cystic fibrosis is one of the diseases for which targeted gene transfer may become an effective treatment.[a,s] Cystic fibrosis induced in mice by targeted disruption of the CFTR gene has been successfully treated by gene therapy.[t] However, the gene is large and efficiency of transfer into animals is low. Nevertheless, human gene therapy for cystic fibrosis is being pursued cautiously.[s]

The CFTR protein has additional medical significance. Its stimulation by bacterial toxins is responsible for **secretory diarrhea**, which kills 3 million young children annually.[f]

[a] Welsh, M. J., and Smith, A. E. (1995) *Sci. Am.* **273**(Dec), 52–59
[b] Koh, J., Sferra, T. J., and Collins, F. S. (1993) *J. Biol. Chem.* **268**, 15912–15921
[c] Marx, J. L. (1989) *Science* **245**, 923–925
[d] Riordan, J. R., Rommens, J. M., Kerem, B.-s, Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L.-C. (1989) *Science* **245**, 1066–1073
[e] Collins, F. S. (1992) *Science* **256**, 774–779
[f] Akabas, M. H. (2000) *J. Biol. Chem.* **275**, 3729–3732
[g] Ernst, R. K., Yi, E. C., Guo, L., Lim, K. B., Burns, J. L., Hackett, M., and Miller, S. I. (1999) *Science* **286**, 1561–1565
[h] Hart, P., Warth, J. D., Levesque, P. C., Collier, M. L., Geary, Y., Horowitz, B., and Hume, J. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6343–6348
[i] Howell, J. M., and Gawthorne, J. M., eds. (1987) *Copper in Animals and Man*, Vol. I and II, CRC Press, Boca, Raton,Florida
[j] Li, C., Ramjeesingh, M., Wang, W., Garami, E., Hewryk, M., Lee, D., Rommens, J. M., Galley, K., and Bear, C. E. (1996) *J. Biol. Chem.* **271**, 28463–28468
[k] Gadsby, D. C., and Nairn, A. C. (1994) *Trends Biochem. Sci.* **19**, 513–518
[l] Ikuma, M., and Welsh, M. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8675–8680
[m] Ostedgaard, L. S., Baldursson, O., and Welsh, M. J. (2001) *J. Biol. Chem.* **276**, 7689–7692
[n] Aleksandrov, L., Mengos, A., Chang, X.-b, Aleksandrov, A., and Riordan, J. R. (2001) *J. Biol. Chem.* **276**, 12918–12923
[o] Weixel, K. M., and Bradbury, N. A. (2000) *J. Biol. Chem.* **275**, 3655–3660
[p] Luo, X., Choi, J. Y., Ko, S. B. H., Pushkin, A., Kurtz, I., Ahn, W., Lee, M. G., and Muallem, S. (2001) *J. Biol. Chem.* **276**, 9808–9816
[q] Seibert, F. S., Jia, Y., Mathews, C. J., Hanrahan, J. W., Riordan, J. R., Loo, T. W., and Clarke, D. M. (1997) *Biochemistry* **36**, 11966–11974
[r] Cotten, J. F., Ostedgaard, L. S., Carson, M. R., and Welsh, M. J. (1996) *J. Biol. Chem.* **271**, 21279–21284
[s] Verma, I. M., and Somia, N. (1997) *Nature (London)* **389**, 239–242
[t] Hyde, S. C., Gill, D. R., Higgins, C. F., Trezise, A. E. O., MacVinish, L. J., Cuthbert, A. W., Ratcliff, R., Evans, M. J., and Colledge, W. H. (1993) *Nature (London)* **362**, 250–255

strong hereditary component, and many of these are being mapped to specific DNA locations.[388] Some specific proteins, such as **Ras** (Chapter 11) that are mutated in many cancers were first recognized as avian oncogenes. The tumor suppressors **Rb** (retinoblastoma protein, Box 11-D)[388,389] and **p53** (Fig. 11-15)[390,391] are major sites of mutation in cancer. See also Chapters 11 and 31. Since cancers contain multiple mutations, they are complex diseases. However, many specific susceptibility loci are being located, including some for breast cancer (Box 11-D),[392] prostate cancer,[393] and familial adenomatous polyposis, a hereditary disease leading to colon cancer.[388,394] Cancer has long been known to be associated with chromosome instability including deletion and insertion mutations at simple repeat sequences, frame-shift mutations,[395] DNA breakage, translocation,[396] and losses or gains of whole chromosomes.[397]

From a practical viewpoint the understanding that we are gaining will help us to provide better treatments of genetic diseases. At present almost every human genetic disease can be mimicked in a knockout mouse.[291a,398] A turnabout is that rare human hair-



(26-4)

TABLE 26-4
Some Human Triplet Repeat and Related Diseases[a]

Name	Repeat sequence (5' to 3')	Transmission[b]	Location
Fragile X syndrome (FRAXA)	$(CGG)_n$	XD	by 5' side of gene *FMR*-1 + C_pG island methylation
Type E (FRAXE)	$(CGG)_n$	XR	5' side of gene *FMR*-2
Synpolydactyly	$(GCG)_n$		HOXD13 gene (polyalanine)
Myotonic dystrophy (DM1)	$(CTG)_n$	AD	3' untranslated region of gene for cAMP-dependent protein kinase
DM2	$(CCTG)_n$	AD	Intron in a zinc finger protein gene
Friedreich ataxia	$(GAA)_n$	AR	Intron in frataxin, a mitochondrial protein
Huntington disease (HD)	$(CAG)_n$	AD	Huntingtin (polyglutamine)
Spinocerebellar ataxia, SCA-1	$(CAG)_n$	AD	Ataxin-1 (polyglutamine)
SCA-2	$(CAG)_n$	AD	Ataxin-2 (polyglutamine)
Machado–Joseph disease (SCA-3)	$(CAG)_n$	AD	Ataxin-3
SCA-6	$(CAG)_n$	AD	Calcium channel
Spinobulbar muscular atrophy (Kennedy disease; SBMA)	$(CAG)_n$	XR	Atrophin (androgen receptor)
Dentato-rubro-pallido-luysian atrophy (DRPLA)	$(CAG)_n$	AD	
Progressive myoclonus epilepsy	(G+C)-rich oligonucleotide repeat		Cystatin B

[a] See Mandel, J.-L. (1997) *Nature (London)* **386**, 767–769 and Richards, R. I., and Sutherland, G. R. (1997) *Trends Biochem. Sci.* **22**, 432–436
[b] XD, X dominant; XR, X recessive; AD, autosomal dominant; AR, autosomal recessive

less people have a mutation in a gene homologous to the well-known *hairless* gene of mice.[399] Genes homologous to those of many human genetic defects have also been identified in yeast.[400] Studies of both mutant mice and of mutant yeast cells can also help in understanding diseases and in devising therapies.

Triplet repeat diseases. With the exception of Down disease (extra chromosome 21, affecting 1 in 600 children) the most prevalent cause of mental retardation is the **fragile X syndrome,** which affects ~1 of 2000 newborn males. A fragile site,[401,402] where the X chromosome breaks easily, marks the location of the defect. Identification of nearby RFLPs led to cloning of the gene in 1991. The defect was found to be a repeated trinucleotide sequence 5'-$(CGG)_n$ in the DNA. The value of n varies, in most healthy individuals averaging ~30. However, for some normal individuals n may be 200–300. These persons may transmit the fragile X disease to their offspring in whom n increases from one generation to the next with increasing severity of the disease. There may be 2000 or more CGG triplets.[403] The defect lies at the 5' end in an untranslated part of the gene for the fragile X mental retardation protein (FMRP). A cytoplasmic RNA-binding protein, FMRP, may enter the nucleus and have an as yet uncertain function. In rare cases the fragile X syndrome arises from deletions or missense mutations in the *FMRP* gene.[404] There are actually two similar genes *FMRP*1 and *FMRP*2. Mutations in the latter are associated with a milder form of fragile X disease. FMRP genes with expanded $(CGG)_n$ tracts are not expressed, evidently because the mutation induces methylation of an adjacent "CpG island" (see Chapter 27) as well as of sites within the FMRP gene. Both FMRP proteins are apparently needed for normal brain function. The FMRP1 defect is genetically dominant, and female heterozygotes also suffer from the fragile X syndrome. However, FMRP2, also encoded on the X chromosome, is recessive.[405] FMRP appears to function in neuronal dendrites (Chapter 30) where it binds to polysomal aggregates and participates in regulation of translation of mRNA.[405a–c] Like DNA, mRNA may contain **G quartets** (Fig. 5-8, p. 227).[405d]

Twelve or more additional triplet repeat diseases, many with neurological symptoms, have been identified (Table 26-4).[405–407a] These involve other trinucleotide repeats 5'-$(GCG)_n$, 5'-$(CTG)_n$, 5'-$(GAA)_n$, and 5'-$(CAG)_n$. In **synpolydactyly**, an inherited developmental defect causing malformation of hands and feet, an expansion of a GCG trinucleotide occurs within the gene *HoxD*13. This results in incorporation of a polyalanine tract near the N terminus of the protein.[408] **Myotonic dystrophy** DM1 (Box 19-A) results from expansion of CTG to 6 kbp or more within the untranslated 3' region of a gene for cAMP-dependent protein kinase.[405] The mRNA transcripts accumulate in the nucleus and may bind to a CUG-binding protein that is involved in splicing other mRNAs, thereby poisoning the cell.[409] DM2 is caused by expansion of a CCTG quartet in an intron of a zinc-finger protein.[408a] Expansion of the GAA triplet is associated with the neurological disease **Friedreich ataxia**, which has a prevalence of ~1 in 50,000. The defect lies within an intron found in the gene for the 210-residue mitochondrial protein **frataxin**. The function of the protein is unknown.[409a] However, the defect in Freidrich ataxia leads to a deficit in mitochondrial ATP synthesis.[409b] Studies of a corresponding protein in yeast suggests that frataxin is an iron storage protein.[409c–e] Apparently the DNA defect in the intron interferes with splicing of the mRNA transcript.[410–413] The polyGAA strand in the triplet repeat region of the DNA is able to form various alternative structures including a parallel (GAA)•(TTC) duplex.[410a] Such structures in mRNA may interfere with proper splicing.

A defect in an α-tocopherol-transfer protein causes a similar set of neurological symptoms. Oxidative damage may therefore be a component of this disease.[410b]

Expansion of $(CAG)_n$ sequences causes a series of neurodegenerative diseases (Table 26-4),[413–416c] the commonest of which is **Huntington disease** (HD).[414–414c] In Huntington's disease the $(CAG)_n$ tract is found in the **huntington** gene. The protein is enormous, with over 3140 residues. It is essential for nerve development, but the function remains uncertain. The first exon at the HD gene encodes a polyglutamine tract of 6–35 residues, which when expanded to 36–100 or more causes the disease. A similar situation holds for the **spinocerebellar ataxias** (SCA) and other $(CAG)_n$ diseases (Table 26-4).[414,417–417b] The encoded proteins have a variety of functions. The SCA-1 protein ataxin-1 functions in the nuclear matrix.[418] The X-linked **spinobulbar muscular atrophy** (SBMA) gene encodes an androgen receptor protein.[414,417] The polyglutamine sequences inserted in those proteins seem to be toxic, but the mechanism of toxicity is uncertain. It may result from aggregation of the proteins within cells.[414] It has also been suggested that cleavage of these chains by **caspases** (cysteine proteases; Chapter 12) may produce truncated proteins, which induce apoptosis.[415]

A form of epilepsy (Table 27-4) appears to be a result of repeats of a (G + C)-rich sequence that may be a dodecamer.[405] Dinucleotide repeats and other "minisatellite" DNA sequences are also associated with instability of DNA and may undergo expansion.[419–421] A pentanucleotide repeat $(CCTTT)_n$ is associated with increased expression of the nitric oxide synthase gene NOS2A. Persons with $n = 14$ were found to have enhanced resistance to development of diabetic retinopathy. This seems to be a case of a beneficial "gain of function" mutation.[422]

How do repeat sequences expand from one generation to the next? There is probably more than one mechanism. One is **strand slippage** during DNA replication. If single-strand loops are present at any time one strand could be displaced (slipped) relative to the other. Replication could then either expand the repetitive sequence or cause a deletion (Eq. 26-4).[405,423–425] Expansion could also occur by gene conversion during homologous recombination (Chapter 27, Section D)[401,426] or during DNA repair.[421,425] Repeat sequences may also prevent proper formation of nucleosomes.[402]

Cloned genes and diagnosis. The first areas in which cloned DNA has affected medicine are in the diagnosis of genetic diseases and in the production of medicinally useful proteins in bacteria, yeast, or cultured cells. One of the first applications was for the diagnosis of the sickle-cell trait by use of PCR on DNA isolated from blood.[43] This was followed quickly by methods for recognition of other inherited diseases and by automated procedures. Intense efforts are now being made to develop DNA "chips" (usually small glass plates with an array of DNA fragments bound to the surface) that can recognize a great variety of defects. For example, DNA polymorphisms in the 16.6-kbp human mitochondrial genome can be recognized by a plate containing 135,000 oligonucleotide probes assembled in a regular grid by photolithography and solid-state synthesis (Chapter 3).[427] Cancers may be classified quickly.[428] Many systems are under development for binding and recognizing genomic DNA, cDNA, or mRNA using microchip arrays of up to 400,000 or more oligonucleotides on a 2 cm x 2 cm plate.[429–430b] One chip combines PCR with use of "zip-code primers" that direct the PCR products to specific zip-code addresses on the chip to give a universal array able to detect low abundance of mutations in any gene of interest.[431] Mass spectrometry is also being harnessed to identify oligonucleotides bound at any address on a chip.[432] Commercial chips are expensive, $100–2000 apiece, and good for only one use. The price will fall. To build a machine to make your own chips go to http//cmgm.stanford.edu/pbrown/mguide/.[433,434]

An alternative to DNA chips is to miniaturize DNA sequencing and analysis machinery. Using nanoliter droplets of fluid passing through microchannels built by photolithographic techniques of computer chip construction, these nanolaboratories may be the size of a credit card but able to cleave DNA and conduct PCR reactions, gel electrophoresis, and sequence determinations.[435,436]

Vaccines, hormones, and other medicines. A myriad of products of recombinant DNA technology are already in use in medical diagnostics.[437] In the past problems have arisen because vaccines and hormones can cause allergic reactions and may harbor viruses. A small percentage of diabetics are allergic to animal insulins, but human insulin produced in bacteria is now available. A number of children receiving human growth hormone isolated from cadavers contracted the fatal Creutzfeld–Jakob disease,[438] a neurological disorder caused by a prion (Chapter 29). One way in which AIDS has been spread is through contamination of the blood-clotting factors VIII or IX needed by hemophiliacs. These sources of contamination are being eliminated by the use of bacterially produced proteins.[439]

Recombinant DNA techniques can be used in two ways in the production of vaccines. The first is to find a protein in the virus or other infective agent that is a good inducer of antibody formation, i.e., a good antigen. This protein, or even a fragment of it, can then be produced from its cloned gene or can be made synthetically. Since the cloned protein can be purified highly, it may make a superior vaccine to those made from killed cells, inactivated virus particles, or mixtures of proteins.[440] The first commercial vaccine of this type was against viral hepatitis B, a major cause of liver cancer.[441] Particles consisting of viral envelope proteins can be produced in yeast and be used for vaccination. A DNA encoding these proteins has been transferred, using the Ti plasmid (Section E, 4), into lettuce. Human volunteers produced anti-hepatitis antibodies after eating the lettuce.[442] This suggests that vaccination through ingestion of antigenic proteins in food crops may be feasible. Injection of a small piece of DNA carrying the gene may also lead to antibody formation.[443,444] Using either purified antigens, proteins, or DNA, it may be possible to develop effective vaccines against Rocky Mountain spotted fever,[445] a rickettsial disease, and against malaria, for which there are no satisfactory vaccines.

The vaccinia virus, formerly used to vaccinate against smallpox, has been engineered for use against other diseases.[446] Much of its 187-kb DNA can be excised and replaced with passenger DNA. The virus particles are stable and highly infectious. The vaccinia virus is unusual in carrying genes for both RNA and DNA polymerases and other proteins that permit it to undergo replication and transcription of its own genes in the cytoplasm of the infected cells. One application of a recombinant vaccinia virus is oral vaccination of wild foxes and racoons against rabies.[447]

A major advance based on cloned genes is the production of new medicines previously unavailable or available in only small amounts. Among these are the **interferons**[287] (Chapter 31) and many hormones such as the **interleukins** produced by lymphocytes[448] and the **atrial natriuretic hormones** (Chapter 23). Another candidate is the **α1-protease inhibitor** (Chapter 12). Perhaps better inhibitors than the natural one can be devised and produced in bacteria.[449]

4. Gene Therapy

A few years ago it seemed like fantasy, but there is little doubt that we will soon be able to routinely treat some genetic illnesses by introducing new copies of genes into the body. A current goal is to insert cloned genes into body cells (somatic cells) to correct specific hereditary defects. For example, juvenile diabetics would benefit from introduction of genes for insulin production into cells that could replace their atrophied pancreatic beta cells. At present we don't know how to do this. However, genes have been transferred into human beings lacking adenosine deaminase and showing severe **combined immunodeficiency** (Chapter 31) and those lacking hypoxanthine guanine phosphoribosyltransferase and displaying the **Lesch–Nyhan syndrome** (Chapter 25, Section E,2). Corrections of **Gaucher disease** and other deficiencies of lysosomal enzymes is also an early goal (Chapter 20, Section G,2).

By 1986 more than 5000 children had received bone marrow transplants from close relatives to correct severe combined immunodeficiency caused by a defective adenosine deaminase gene.[450] However, the patients must receive chemotherapy or irradiation to suppress their immune system before the transplantation. Hospitalization may last for 30–60 days. Using genetic therapy some bone marrow cells can be removed from the patient with a needle. The cells can be treated to introduce the corrected genes using a suitable retrovirus. Clones of corrected bone marrow stem cells, which will give rise to lymphocytes in the body, can be selected, cutured and reintroduced into the patient. As discussed on p. 1498, retrovirus vehicles that locate and become integrated at appropriate sites in the genome have been developed.[216,451] Use of homologous recombination (Chapter 27) introduces the cloned gene into its normal chromosomal location.[452] The first children were treated by transfer of the adenosine deaminase gene in 1991. The procedure has been partially successful.[453,454]

By 1999 more than 400 clinical gene therapy trials were planned or in progress.[455] Nevertheless, development of suitable vehicles for gene delivery has been slow.[456,456a] Uncertainty about the safety of adenovirus vectors is one problem.[160a,454,457,457a] Poor efficiency of gene transfer is another. A glycogen storage disease of knockout mice has been cured by transfer of human α-glucosidase (Box 20-D) using an adenovirus vector.[458] Mice have also been used in developing gene therapy for hemophilia,[456a,457a] sickle-cell disease,[458a] and aspartylglycosaminuria.[458b] However, gene targeting in animals other than the mouse has been difficult.[459] Genetic therapy may be most effective when the gene is transferred into stem cells, which can then take up residence within the body (Chapter 32). Young stem cells, which can be obtained from unbilical cord blood at birth, can be used.[460,461] For many diseases therapy will probably be best soon after birth, or even prior to birth during the second trimester of pregnancy.[455]

An alternative approach is to synthesize highly specific hydrolysis-resistant DNA analogs that can form triple helical structures with DNA (Chapter 5). If these can be made specific enough they might bind to a targeted DNA site, such as the sickle cell anemia locus and induce a "back mutation" from the faulty A•T base pair (see Fig. 7-23) to a T•A pair in at least some of the hematopoietic cells that give rise to hemoglobin.[462] Another possibility is to make oligonucleotide analogs that serve as mimics of **antisense RNA** (discussed in Chapter 28). This could impede translation of bad mRNAs, such as that giving rise to polyglutamine chains in the triple repeat diseases.

At this time no effort is being made to alter the DNA in human germ cells. It seems undesirable to experiment with such changes.[160a,462a] However, as methods are developed for genetic therapy of somatic cells, we will rear more and more healthy carriers of serious genetic defects. Eventually we may need to develop therapy for germ cells.

5. Genetic Engineering of Bacteria, Plants, and Animals

Many improvements in bacteria used in industrial fermentations have been made.[463] The number of copies of a useful gene may be increased, and repressed genes may be made more active by deletion mutations in regulatory genes (Chapter 28). New genes are being transferred between bacteria and into plants and animals. For example, *Bacillus thuringensis* produces crystalline protein toxins, which are harmless to mammals but active against many insects. The *Bt* gene for this toxin has already been transferred into many different plants and is having a major effect on agriculture in allowing for decreased use of insecticides.[464,465] Genetic engineering may allow the toxin to be made more specific for particular insect species.

Genetic engineering of plant genes[466] may improve the quality of storage proteins in cereal grains[467] and the flavors of fruits,[468] provide more drought-resistant plants,[469] and offer increased resistance of crop plants to particular herbicides[464] (Chapter 25, Section A,1) and to viruses.[470] Many things can be done to enhance the nutritional qualities of foods.[464] For example, rice has been engineered to produce and store β-carotene in the grain, a development that may benefit 400 million people in the world deficient in vitamin A and often suffering from infections and blindness.[471] (See also p. 1240.) Specialty products such as poly-β-hydroxybutyrate and many others can also be produced in plants.[472] Genetic engineering of animals has produced not only knockout mice but the

possibility for nutritional improvement[156] and for production of useful products such as vaccines in milk.

6. Ethical Problems

With a host of new medicines and agricultural products coming and with the ability to alter genomes at will, we face new ethical problems. For example, who owns the human genome?[473–475] Should patenting of human genes be allowed?[476,477] How private are genetic data?[478] Who has access to human DNA data?[479] Insurance companies?[480,481] Should routine screening with new genetic tests be used (as is now the case for phenylketonuria, diabetes, and others), even if no effective treatment is available?[482,483]

Ethical questions arise in application of experimental forms of genetic therapy to patients dying of inherited disorders. What if the therapy simply provides a longer life of suffering to patients? Some biologists foresee a future in which human beings learn to control their own genes and prevent genetic deterioration resulting from the accumulation of harmful mutations. They find it exciting to think that we can, in an intelligent way, elect to continue our evolution in a desired direction. Others caution that our present knowledge is such that attempts to eliminate all "bad" genes from the population might be disastrous.[484] They point to the hemoglobin S gene (Box 7-B) and the role it once had in preserving life in a malaria-infested environment and urge that at this stage we allow maximum heterogeneity of genetic types. There are other dangers in allowing genetics to control human lives. In the past "eugenic" doctrines have been used to justify racist laws and (in Nazi Germany) genocide.

Problems have resulted from the availability of hormones and other products of genetic engineering. For example, should parents be allowed to request human growth hormone for short children who do not have a pituitary deficiency? Should it be available to athletes and to aging people? Use of stem cells in genetic therapy raises another set of questions. Present guidelines of the National Institutes of Health (NIH), which supports much of the gene therapy research, allows use of stem cells in therapy but doesn't allow their use in reproductive cloning or the combining of human stem cells with animal embryos.[485] Dangers from retroviruses may lurk in attempts to transplant animal organs (e.g., pig hearts) into people.[486–488]

In attempting to improve crop plants, proteins causing allergic reactions in some individuals may be tranferrred.[489] This is a compelling reason that mandatory monitoring and labeling of genetically modified foods is recommended.[490] The danger is emphasized by the highly popularized contamination of foods with "starlink" maize.

Should we rush commercial developments of genetic engineering to increase food supply at a time when there is an excess worldwide? Should we engineer increased herbicide-resistance into plants when drinking water supplies are being contaminated by herbicides already in use? On the other hand, less toxic herbicides effective at low levels can be used on plants engineered to resist them.[465] Will we generate superweeds resistant to herbicides by gene transfer? With so many genes being engineered, are there other dangers in their release into the environment?[465] A **terminator** technology that produces plants with sterile seeds has been rejected because it would force poor farmers to purchase seeds year after year from multinational corporations. Yet, it could eventually have beneficial consequences in preventing cross-fertilization with genetically modified plants.[491] Since plants with the *Bt* gene kill insects, will their widespread use decimate populations of butterflies or of various beneficial insects or adversely affect soil organisms in the **rhizosphere**?[492,493]

The darkest prospect may be the possibility that military organizations will develop new biological warfare weapons using recombinant DNA methods.[494] An international treaty bans such weapons. However, governments often respond with the argument "We must do it because we know that 'they' are doing it." Couldn't such an attitude bring on unparalleled disaster? Could new types of viruses spread throughout the world and literally tear the human genome to bits? Does any kind of imagined danger justify preparation to attack populations with new viruses or new toxins? What about the smallpox virus? Should remaining stocks in the United States and Russia be destroyed?[495,496]

References

1. Horowitz, N. H. (1995) *Protein Sci.* **4**, 1017–1019
2. Atherly, A. G., Girton, J. R., and McDonald, J. F. (1999) *The Science of Genetics*, Saunders, Philadelphia, Pennsylvania
2a. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002) *Molecular Biology of the Cell*, 4th ed., Garland Science, New York
2b. Lander, E. S., and Weinberg, R. A. (2000) *Science* **287**, 1777 – 1782
3. Mirsky, A. E. (1968) *Sci. Am.* **218**(Jun), 78–88
4. Portugal, F. H., and Cohen, J. S. (1977) *A Century of DNA: A History of the Discovery of the Structure and Function of the Genetic Substance*, MIT Press, Cambridge, Massachusetts
5. Adams, R. L. P., Knowler, J. T., and Leader, D. P. (1992) *The Biochemistry of the Nucleic Acids*, 11th ed., Chapman & Hall, London
6. Olby, R. (1975) *The Path to the Double Helix*, Univ. of Washington Press, Seattle, Washington
7. Stent, G. S., and Calendar, R. (1978) *Molecular Genetics*, 2nd ed., Freeman, San Francisco, California
8. Chargaff, E. (1971) *Science* **172**, 637–642
9. Olby, R. (1974) *Nature (London)* **248**, 782–785
10. McCarty, M. (1985) *The Transforming Principle*, Norton, New York
11. Cohen, S. S. (1984) *Trends Biochem. Sci.* **9**, 334–336
12. Chambers, D. A., Reid, K. B. M., and Cohen, R. L. (1994) *FASEB J.* **8**, 1219–1226
13. Hershey, A. D., and Chase, M. (1952) *J. Gen. Physiol.* **36**, 39–56
14. Hayes, W. (1968) *The Genetics of Bacteria and Their Viruses; Studies in Basic Genetics and Molecular Biology*, 2nd ed., Wiley, New York
15. Chargaff, E. (1950) *Experientia* **6**, 201–209
16. Brinton, C. C. (1971) *Crit. Rev. Microbiol* **1**, 105–160
17. Sayre, A. (1975) *Rosalind Franklin and DNA*, Norton, New York
17a. Maddox, B. (2002) *Rosalind Franklin: The Dark Lady of DNA*, HarperCollins, New York
17b. Blow, D. (2002) *Nature (London)* **418**, 725–726
18. Watson, J. D., and Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738
19. Crick, F. (1974) *Nature (London)* **248**, 766–769
20. Watson, J. D. (1968) *The Double Helix*, Atheneum, New York
20a. Morange, M. (2000) *A History of Molecular Biology*, Harvard Univ. Press, Cambridge,MA (translated by Matthew Cobb)
20b. Echols, H. (2001) *Operators and Promoters: The Story of Molecular Biology and Its Creators*, Univ. of California Press, Berkeley, California (edited by Carol A. Gross)
21. Caspersson, T. (1941) *Naturwissenschaften* **29**, 33–43
22. Brachet, J. (1942) *Arch. Biol.* **53**, 207–257
23. Brachet, J. (1987) *Trends Biochem. Sci.* **12**, 244–246
24. Prescott, D. M. (1964) *Prog. Nucleic Acid Res. Mol. Biol.* **3**, 33–57
25. Sirlin, J. L. (1972) *Biology of RNA*, Academic Press, New York
26. Hoagland, M. B., Keller, E. B., and Zamecnik, P. (1956) *J. Biol. Chem.* **218**, 345–358
27. Belozersky, A. N., and Spirin, A. S. (1960) in *The Nucleic Acids*, Vol. 3 (Chargaff, E., and Davidson, J. N., eds), pp. 147–185, Academic Press, New York
28. Davern, C. I., and Meselson, M. (1960) *J. Mol. Biol.* **2**, 153–160
29. Volkin, E., and Astrachan, L. (1956) *Virology* **2**, 149–161
30. Volkin, E. (1995) *Trends Biochem. Sci.* **20**, 206–209
31. Rouwenhorst, R. J., Pronk, J. T., and van Dijken, J. P. (1989) *Trends Biochem. Sci.* **14**, 416–418
32. Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356
33. Cohen, G. N. (1995) *FASEB J.* **9**, 981–982
34. Brenner, S., Jacob, F., and Meselson, M. (1961) *Nature (London)* **190**, 576–581
35. Crick, F. H. C. (1957) *Biochem. Soc. Symp.* **14**, 25–26
36. Hoagland, M. (1996) *Trends Biochem. Sci.* **21**, 77–80
37. Nirenberg, M. W., and Matthaei, J. H. (1961) *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1588–1602
38. Nirenberg, M., and Leder, P. (1964) *Science* **145**, 1399–1407
39. Matthaei, J. H., Voigt, H. P., Heller, G., Neth, R., Schöch, G., Kübler, H., Amelunxen, F., Sander, G., and Parmeggiani, A. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 25–38
40. Khorana, H. G., Büchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kössel, H., Morgan, R., Narang, S. A., Ohtsuka, E., and Wells, R. D. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 39–49
41. Edgar, R. S., and Epstein, R. H. (1965) *Sci. Am.* **212**(Feb), 70–78
42. Fraenkel-Conrat, H. (1964) *Sci. Am.* **211**(Oct), 47–54
43. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2757–2760
44. Lewin, B. (1974) *Gene Expression*, Vol. 2, Wiley, New York (p. 258)
45. Knight, R. D., Freeland, S. J., and Landweber, L. F. (1999) *Trends Biochem. Sci.* **24**, 241–247
46. Seid-Akhavan, M., Winter, W. P., Abramson, R. K., and Rucknagel, D. L. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 882–886
47. Benzer, S. (1962) *Sci. Am.* **206**(Jan), 70–84
48. Hayes, W. (1968) *The Genetics of Bacteria and Their Viruses*, 2nd ed., Wiley, New York (pp. 52–54)
49. Weintraub, S. B., and Frankel, F. R. (1972) *J. Mol. Biol.* **70**, 589–615
50. Ennis, H. L., and Kievitt, K. D. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1468–1472
51. Beadle, G. W., and Tatum, E. L. (1941) *Proc. Natl. Acad. Sci. U.S.A.* **27**, 499–
52. Yanofsky, C. (1967) *Sci. Am.* **216**(May), 80–94
53. Yanofsky, C. (1976) in *Reflections on Biochemistry* (Kornberg, A., ed), pp. 263–269, Pergamon, New York
54. Sarabhai, A. S., Stretton, A. O. W., Brenner, S., and Bolle, A. (1964) *Nature (London)* **210**, 13–17
55. Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961) *Nature (London)* **192**, 1227–1232
55a. Atkins, J. F., and Gesteland, R. F. (2001) *Nature (London)* **414**, 693
56. Watson, J. O. (1987) *Molecular Biology of the Gene*, 4th ed., Benjamin, New York
57. Eiserling, F. A., and Dickson, R. C. (1972) *Ann. Rev. Biochem.* **41**, 467–502
58. Wood, W. B., and Edgar, R. S. (1967) *Sci. Am.* **217**(Jul), 60–74
59. Sober, H. A., ed. (1968) *CRC Handbook of Biochemistry*, 1st ed., Chem. Rubber Publ. Co., Cleveland, Ohio (p. I25)
60. Garen, A. (1968) *Science* **160**, 149–159
61. Littauer, U. Z., and Inouye, H. (1973) *Ann. Rev. Biochem.* **42**, 439–470
62. Hartman, P. E., and Roth, J. R. (1973) *Adv. Genet.* **17**, 1–105
63. Lewin, B. (1974) *Gene Expression*, Vol. 1, Wiley, New York (p. 213)
64. Yanofsky, C., and Crawford, I. P. (1972) in *The Enzymes*, 3rd ed., Vol. 7 (Boyer, P. D., ed), Academic Press, New York (pp. 1–31)
65. Sherratt, D. (1975) *Nature (London)* **254**, 559–560
66. Lederberg, J., and Tatum, E. L. (1946) *Nature (London)* **158**, 558
67. Kornberg, A. (1974) *DNA Synthesis*, Freeman, San Francisco, California (p. 242–248)
68. Harrington, L. C., and Rogerson, A. C. (1990) *J. Bacteriol.* **172**, 7263–7264
69. Eisenbrandt, R., Kalkum, M., Lai, E.-M., Lurz, R., Kado, C. I., and Lanka, E. (1999) *J. Biol. Chem.* **274**, 22548–22555
70. Hendrix, R. W., Roberts, J. W., Stahl, F. W., and Weisberg, R., eds. (1983) *Lambda II*, Cold spring Harbor Lab, Cold Spring Harbor, New York
71. Hershey, A. D., ed. (1971) *The Bacteriophage Lambda*, Cold Spring Harbor Lab, Cold Spring Harbor, New York
72. Echols, H. (1971) *Ann. Rev. Biochem.* **40**, 827–854
73. O'Brien, S. J., ed. (1993) *Genetic Maps*, 6th ed., Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York ((Books 1–6))
74. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1453–1462
75. Bachmann, B. J., Low, K. B., and Taylor, A. L. (1976) *Bacteriol. Rev.* **40**, 116–167
76. Bachmann, B. J. (1983) *Bacteriol. Rev.* **47**, 180–230
77. Smith, C. L., Econome, J. G., Schutt, A., Klco, S., and Cantor, C. R. (1987) *Science* **236**, 1448–1453
78. Kohara, Y., Akiyama, K., and Isono, K. (1987) *Cell* **50**, 495–508
79. Low, K. B. (1987) in *Escherichia coli and Salmonella typhimurium* (Neidhardt, F. C., ed), Am. Soc. for Microbiology, Washington, D. C.
80. Willetts, N., and Skurray, R. (1987) in *Escherichia coli and Salmonella typhimurium*, Vol. 2 (Neidhardt, F. C., ed), pp. 1110–1133, Am. Soc. for Microbiology, Washington, D. C.
80a. Bist, P., Sistla, S., Krishnamurthy, V., Acharya, A., Chandrakala, B., and Rao, D. N. (2001) *J. Mol. Biol.* **310**, 93 – 109
81. Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E., and Dryden, D. T. F. (1999) *J. Mol. Biol.* **290**, 565–579
82. Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., and Bickle, T. A. (1999) *EMBO J.* **18**, 2638–2647
83. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1995) *Science* **269**, 656–663
84. Bozic, D., Grazulis, S., Siksnys, V., and Huber, R. (1996) *J. Mol. Biol.* **255**, 176–186
85. Luria, S. E. (1970) *Sci. Am.* **222**(Jan), 88–102
86. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., and Rosenberg, J. M. (1986) *Science* **234**, 1526–1541
87. Hager, P. W., Reich, N. O., Day, J. P., Coche, T. G., Boyer, H. W., Rosenberg, J. M., and Greene, P. J. (1990) *J. Biol. Chem.* **265**, 21520–21526
88. Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W., and Rosenberg, J. M. (1984) *Nature (London)* **309**, 327–331
89. Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., and Rich, A. (1988) *J. Biol. Chem.* **263**, 17872–17879

References

90. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1994) *Nature (London)* **368**, 660–664
91. Kostrewa, D., and Winkler, F. K. (1995) *Biochemistry* **34**, 683–696
91a. Horton, N. C., Otey, C., Lusetti, S., Sam, M. D., Kohn, J., Martin, A. M., Ananthnarayan, V., and Perona, J. J. (2002) *Biochemistry* **41**, 10754–10763
92. Stanford, N. P., Halford, S. E., and Baldwin, G. S. (1999) *J. Mol. Biol.* **288**, 105–116
93. Sam, M. D., and Perona, J. J. (1999) *Biochemistry* **38**, 6576–6586
94. Kim, Y., Grable, J. C., Love, R., Greene, P. J., and Rosenberg, J. M. (1990) *Science* **249**, 1307–1309
95. Kim, S. C., Podhajska, A. J., and Szybalski, W. (1988) *Science* **240**, 504–506
96. Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331
97. Cohen, D., Chumakov, I., and Weissenbach, J. (1993) *Nature (London)* **366**, 698–701
98. Westmoreland, B. C., Szybalski, W., and Ris, H. (1969) *Science* **163**, 1343–1348
99. Glover, D. M., ed. (1987) *DNA Cloning: A Practical Approach*, Vol. III, IRL Press, Oxford
100. Allemand, J.-F., Bensimon, D., Jullien, L., Bensimon, A., and Croquette, V. (1997) *Biophys. J.* **73**, 2064–2070
101. Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J. S., and Bensimon, A. (1997) *Science* **277**, 1518–1523
102. Samad, A. H., Cai, W. W., Hu, X., Irvin, B., Jing, J., Reed, J., Meng, X., Huang, J., Huff, E., Porter, B., Shenkar, A., Anantharaman, T., Mishra, B., Clarke, V., Dimalanta, E., Edington, J., Hiort, C., Rabbah, R., Skiada, J., and Schwartz, D. C. (1995) *Nature (London)* **378**, 516–517
103. Lin, J., Qi, R., Aston, C., Jing, J., Anantharaman, T. S., Mishra, B., White, O., Daly, M. J., Minton, K. W., Venter, J. C., and Schwartz, D. C. (1999) *Science* **285**, 1558–1561
104. Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990) *Science* **250**, 245–250
105. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991) *Science* **252**, 1651–1656
106. Mann, M. (1996) *Trends Biochem. Sci.* **21**, 494–495
107. Watson, J. D., Gilman, M., Witkowski, J., and Zoller, M., eds. (1992) *Recombinant DNA*, 2nd ed., Scientific American Publ. Co., New York
108. Sambrook, J., and Russell, D. (2000) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Lab. Press, Plainview, New York
109. Berger, S. L., and Kimmel, A. R., eds. (1987) *Guide to Molecular Cloning Techniques*, Academic Press, San Diego
110. Perbal, B. (1988) *A Practical Guide to Molecular Cloning*, Wiley, New York
111. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry, Part II*, Freeman, San Francisco, California (pp. 409–433)
112. Greene, J. J., and Rao, V. B., eds. (1998) *Recombinant DNA Principles and Methodologies*, Dekker, New York
113. Miller, J. H. (1992) *A Short Course in Bacterial Genetics*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
114. Wu, R., ed. (1987) *Methods in Enzymology, "Recombinant DNA,"* Vols. 153–155, Academic Press, San Diego
115. Pouwels, P. H., Enger-Valk, B. E., and Brammar, W. J., eds. (1985) *Cloning Vectors: A Laboratory Manual*, Elsevier, Amsterdam
116. Adolph, K. W., ed. (1993) *Methods in Molecular Genetics: Gene and Chromosome Analysis*, Vols. 1 and 2, Academic Press, San Diego, California
117. Birren, B., Green, E. D., Klapholz, S., Myers, R. M., and Roskams, J., eds. (1997) *Analyzing DNA: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
118. Birren, B., Green, E. D., Klapholz, S., Myers, R. M., Riethman, H., and Roskams, J., eds. (1999) *Cloning Systems: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
119. Lathe, R., Kieny, M. P., Skory, S., and Lecocq, J. P. (1984) *DNA* **3**, 173–182
120. Nobbs, T. J., and Halford, S. E. (1995) *J. Mol. Biol.* **252**, 399–411
121. Winnacker, E. L. (1987) *From Genes to Clones*, VCH, New York
122. Midgley, C. A., and Murray, N. E. (1985) *EMBO J.* **4**, 2695–2703
123. Bollum, F. J. (1981) *Trends Biochem. Sci.* **6**, 41–43
123a. Shanks, O. C., Bissonette, L., and Ream, W. (2000) *Science* **289**, 413
124. Bolivar, F., Rodriguez, R. L., Coreene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. (1977) *Gene* **2**, 95–113
125. Peden, K. W. C. (1983) *Gene* **22**, 277–280
126. Vieira, J., and Messing, J. (1982) *Gene* **19**, 259–268
127. Messing, J. (1981) *Nucleic Acids Res.* **9**, 309–321
128. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M., and Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246
129. Blattner, F. R., Williams, B. G., Blechi, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L.-A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. (1977) *Science* **196**, 161–169
130. Short, J. M. (1988) *Nucleic Acids Res.* **15**
131. Collins, J., and Hohn, B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4242–4246
132. Wahl, G. M., Lewis, K. A., Ruiz, J. C., Rothenberg, B., Zhao, J., and Evans, G. A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2160–2164
133. Fernandez, J. M., and Hoeffler, J. P., eds. (1998) *Gene Expression Systems*, Academic Press, San Diego, California
134. Moss, B., Elroy-Stein, O., Mizukami, T., Alexander, W. A., and Fuerst, T. R. (1990) *Nature (London)* **348**, 91–92
135. O'Reilly, D. R., Miller, L. K., and Luckow, V. A. (1992) *Baculovirus Expression Vectors*, Freeman, New York
136. Kriegler, M. (1990) *Gene Transfer and Expression*, Stockton Press, New York
137. Burke, D. T., Carle, G. F., and Olson, M. V. (1987) *Science* **236**, 806–812
138. Murray, A. W., and Szostak, J. W. (1987) *Sci. Am.* **257**(Nov), 62–68
139. Forget, B. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7909–7911
140. Anderson, C. (1993) *Science* **259**, 1684–1687
141. Sternberg, N. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 103–107
142. Pierie, J. C., and Sternberg, N. L. (1992) *Methods Enzymol.* **216**, 549–
143. Heller, R., Brown, K. E., Burgtorf, C., and Brown, W. R. A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7125–7130
144. Chilton, M.-D. (1983) *Sci. Am.* **248**(Jun), 51–59
145. Weiler, E. W., and Schröder, J. (1987) *Trends Biochem. Sci.* **12**, 271–275
146. Zambryski, P. C. (1992) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 465–490
146a. Wood, D. W., and 50 other authors. (2001) *Science* **294**, 2317–2323
146b. Goodner, B., and 30 other authors. (2001) *Science* **294**, 2323–2328
146c. Moriguchi, K., Maeda, Y., Satou, M., Hardayani, N. S. N., Kataoka, M., Tanaka, N., and Yoshida, K. (2001) *J. Mol. Biol.* **307**, 771–784
147. Slightom, J. L., Durand-Tardif, M., Jouanin, L., and Tepfer, D. (1986) *J. Biol. Chem.* **261**, 108–121
148. Hess, K. M., Dudley, M. W., Lynn, D. G., Joerger, R. D., and Binns, A. N. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7854–7858
149. Stachel, S. E., Messens, E., Van Montagu, M., and Zambryski, P. (1985) *Nature (London)* **318**, 624–629
150. Bates, H. A., Kaushal, A., Deng, P.-N., and Sciaky, D. (1984) *Biochemistry* **23**, 3287–3290
151. Tinland, B., Schoumacher, F., Gloeckler, V., Bravo-Angel, A. M., and Hohn, B. (1995) *EMBO J.* **14**, 3585–3595
152. Scheiffele, P., Pansegrau, W., and Lanka, E. (1995) *J. Biol. Chem.* **270**, 1269–1276
153. Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. J. (1995) *EMBO J.* **14**, 3206–3214
154. Fullner, K. J., Lara, J. C., and Nester, E. W. (1996) *Science* **273**, 1107–1109
154a. Ward, D. V., Draper, O., Zupan, J. R., and Zambryski, P. C. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11493–11500
155. Crystal, R. G. (1995) *Science* **270**, 404–410
156. Westphal, H. (1989) *FASEB J.* **3**, 117–120
157. Cotton, M., Wagner, E., Zatloukal, K., Phillips, S., Curiel, D. T., and Birnstiel, M. L. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6094–6098
158. Kojima, H., Ohishi, N., and Yagi, K. (1998) *Biochem. Biophys. Res. Commun.* **246**, 868–872
159. Yu, S.-F., von Rüden, T., Kantoff, P. W., Garber, C., Seiberg, M., Rüther, U., Anderson, W. F., Wagner, E. F., and Gilboa, E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3194–3198
160. Chakraborty, A. K., Zink, M. A., Boman, B. M., and Hodgson, C. P. (1993) *FASEB J.* **7**, 971–977
160a. Ferber, D. (2001) *Science* **294**, 1638–1642
161. Daly, T. M., Vogler, C., Levy, B., Haskins, M. E., and Sands, M. S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2296–2300
161a. Yan, Z., Zhang, Y., Duan, D., and Engelhardt, J. F. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6716–6721
161b. Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M. S. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10405–10410
162. Amado, R. G., and Chen, I. S. Y. (1999) *Science* **285**, 674–676
162a. Lois, C., Hong, E. J., Pease, S., Brown, E. J., and Baltimore, D. (2002) *Science* **295**, 868–872
163. Jasny, B. R. (1987) *Science* **238**, 1653
164. Hofmann, C., Sandig, V., Jennings, G., Rudolph, M., Schlag, P., and Strauss, M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10099–10103
165. Yin, H. L., and Stull, J. T. (1999) *J. Biol. Chem.* **274**, 32529–32530
166. Poul, M.-A., and Marks, J. D. (1999) *J. Mol. Biol.* **288**, 203–211
166a. Kaminski, J. M., Huber, M. R., Summers, J. B., and Ward, M. B. (2002) *FASEB J.* **16**, 1242–1247
166b. Saldanha, R., Chen, B., Wank, H., Matsuura, M., Edwards, J., and Lambowitz, A. M. (1999) *Biochemistry* **38**, 9069–9083
166c. Guo, H., Karberg, M., Long, M., Jones, J. P., III, Sullenger, B., and Lambowitz, A. M. (2000) *Science* **289**, 452–457
167. Cocking, E. C., and Davey, M. R. (1987) *Science* **236**, 1259–1262

References

168. Smith, H. O., Danner, D. B., and Deich, R. A. (1981) *Ann. Rev. Biochem.* **50**, 41–68
169. Wahl, G. M., de Saint Vincent, B. R., and DeRose, M. L. (1984) *Nature (London)* **307**, 516–520
170. Felgner, P. L., and Ringold, G. M. (1989) *Nature (London)* **337**, 387–388
171. Miller, J. F., Dower, W. J., and Tompkins, L. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 856–860
172. Solioz, M., and Bienz, D. (1990) *Trends Biochem. Sci.* **15**, 175–177
173. Hengen, P. N. (1995) *Trends Biochem. Sci.* **20**, 248–249
174. Collombet, J.-M., Wheeler, V. C., Vogel, F., and Coutelle, C. (1997) *J. Biol. Chem.* **272**, 5342–5347
175. Chang, D. C., Hunt, J. R., Zheng, Q., and Gao, P. Q. (1992) in *Guide to Electroporation and Electrofusion* (Chang, D. C., ed), pp. 303–326, Academic Press, San Diego, California
176. Klobutcher, L. A., and Ruddle, F. H. (1981) *Ann. Rev. Biochem.* **50**, 533–554
177. Klein, T. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1987) *Nature (London)* **327**, 70–73
178. Butow, R. A., and Fox, T. D. (1990) *Trends Biochem. Sci.* **15**, 465–468
179. Kohli, A., Leech, M., Vain, P., Laurie, D. A., and Christou, P. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7203–7208
180. Hohn, B., and Murray, K. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3259–3263
181. Lutz, C. T., Hollifield, W. C., Seed, B., Davie, J. M., and Huang, H. V. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4379–4383
182. Wyman, A. R., Wolfe, L. B., and Botstein, D. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2880–2884
183. Okayama, H., and Berg, P. (1982) *Mol. Cell. Biochem.* **2**, 161–170
184. Gubler, U., and Hoffman, B. J. (1983) *Gene* **25**, 263–269
185. Mason, W. T., ed. (1999) *Fluorescent and Luminescent Probes*, 2nd ed., Academic Press, San Diego, California
186. Keller, G. H., and Manak, M. M. (1989) *DNA Probes*, Stockton Press, New York
186a. Santi, E., Capone, S. Mennuni, C., Lahm, A., Tramontano, A, Luzzago, A., and Nicosia, A. (2000) *J. Mol. Biol.* **296**, 497 – 508
187. Little, P. F. R., and Cross, S. H. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3159–3163
188. Richards, J. E., Gilliam, T. C., Cole, J. L., Drumm, M. L., Wasmuth, J. J., Gusella, J. F., and Collins, F. S. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6437–6441
189. Housman, D. (1995) *N. Engl. J. Med.* **332**, 318–320
190. Viegas-Pequignot, E., Dutrilleax, B., Magdelenat, H., and Coppey-Moisam, M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 582–586
191. Abbott, C., and Povey, S. (1995) *Somatic Cell Hybrids*, IRL Press, Oxford
192. Williamson, B. (1981) *Nature (London)* **293**, 10–11
193. White, R., and Lalouel, J.-M. (1988) *Sci. Am.* **258**(Feb), 40–48
194. Gusella, J. F. (1986) *Ann. Rev. Biochem.* **55**, 831–854
195. Davies, K. E., ed. (1986) *Human Genetic Diseases*, IRL Press, Washington, D.C.
196. Martin, J. B. (1987) *Science* **238**, 765–772
197. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350–1354
198. Gusella, J. F., Tanzi, R. E., Bader, P. I., Phelan, M. C., Stevenson, R., Hayden, M. R., Hofman, K. J., Faryniarz, A. G., and Gibbons, K. (1985) *Nature (London)* **318**, 75–78
199. Gilliam, T. C., Bucan, M., MacDonald, M. E., Zimmer, M., Haines, J. L., Cheng, S. V., Pohl, T. M., Meyers, R. H., Whaley, W. L., Allitto, B. A., Faryniarz, A., Wasmuth, J. J., Frischauf, A.-M., Conneally, P. M., Lehrach, H., and Gusella, J. F. (1987) *Science* **238**, 950–952
200. Royer-Pokora, B., Kunkel, L. M., Monaco, A. P., Goff, S. C., Newburger, P. E., Baehner, R. L., Cole, F. S., Curnutte, J. T., and Orkin, S. H. (1986) *Nature (London)* **322**, 32–38
201. Reich, N. O., and Mashboon, N. (1993) *J. Biol. Chem.* **268**, 9191–9193
202. Lee, W.-H., Shew, J.-Y., Hong, F. D., Sery, T. W., Donoso, L. A., Young, L.-J., Bookstein, R., and Lee, E. Y.-H. P. (1987) *Nature (London)* **329**, 642–645
203. Thomson, G., and Esposito, M. S. (1999) *Trends Biochem. Sci.* **24**, M17–M20
204. Little, P. (1986) *Nature (London)* **321**, 558–559
205. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) *J. Mol. Biol.* **255**, 589–603
206. Myers, R. M., Lerman, L. S., and Maniatis, T. (1985) *Science* **229**, 242–247
207. Botstein, D., and Shortle, D. (1985) *Science* **229**, 1193–1201
208. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
209. Zoller, M. J., and Smith, M. (1983) *Methods Enzymol.* **100**, 468–500
210. Higuchi, R., Krummel, B., and Saiki, R. (1988) *Nucleic Acids Res.* **16**, 7351–7367
211. Jones, D. H., and Howard, B. H. (1990) *BioTechniques* **8**, 178–183
212. Jones, D. H., Sakamoto, K., Vorce, R. L., and Howard, B. H. (1990) *Nature (London)* **344**, 793–794
213. Wells, J. A., and Powers, D. B. (1986) *J. Biol. Chem.* **261**, 6564–6570
214. Shuldiner, A. R. (1996) *N. Engl. J. Med.* **334**, 653–655
215. Jasin, M., Moynahan, M. E., and Richardson, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8804–8808
216. Capecchi, M. R. (1994) *Sci. Am.* **270**(Mar), 52–59
217. Travis, J. (1992) *Science* **256**, 1392–1394
218. Barinaga, M. (1994) *Science* **265**, 26–28
219. Weissmann, C., and Aguzzi, A. (1999) *Science* **286**, 914–915
219a. Smith, J. A., and Martin, L. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 1263–1267
220. Nasmyth, K. (2002) *Science* **297**, 559–565
220a. Hauf, S., Waizenegger, I. C., and Peters, J.-M. (2001) *Science* **293**, 1320–1323
220b. Campbell, J. L., and Cohen-Fix, O. (2002) *Trends Biochem. Sci.* **27**, 492–495
220c. Seemann, J., Pypaert, M., Taguchi, T., Malsam, J., and Warren, G. (2002) *Science* **295**, 848–851
221. Mazia, D. (1961) *Sci. Am.* **205**(Sep), 101–120
222. Earnshaw, W. C., and Mackay, A. M. (1994) *FASEB J.* **8**, 947–956
223. Russell, P. (1998) *Trends Biochem. Sci.* **23**, 399–402
224. Piwnica-Worms, H. (1999) *Nature (London)* **401**, 535–537
225. Nasmyth, K. (1999) *Trends Biochem. Sci.* **24**, 98–104
225a. Cortez, D., and Elledge, S. J. (2000) *Nature (London)* **406**, 354–356
226. Warren, G. (1985) *Trends Biochem. Sci.* **10**, 439–443
226a. Lippincott-Schwartz, J. (2002) *Nature (London)* **416**, 31–32
227. Morgan, D. O. (1995) *Nature (London)* **374**, 131–133
227a. Kong, M., Barnes, E. A., Ollendorff, V., and Donoghue, D. J. (2000) *EMBO J.* **19**, 1378–1388
227b. Healy, J. M. S., Menges, M., Doonan, J. H., and Murray, J. A. H. (2001) *J. Biol. Chem.* **276**, 7041–7047
228. Pavletich, N. P. (1999) *J. Mol. Biol.* **287**, 821–828
228a. Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F. B., Mendenhall, M. D., Sicheri, F., Pawson, T., and Tyers, M. (2001) *Nature (London)* **414**, 514–521
228b. Bartek, J., and Lukas, J. (2001) *Science* **294**, 66–67
228c. Schwab, M., and Tyers, M. (2001) *Nature (London)* **413**, 268–269
228d. Zhou, B.-B. S., and Elledge, S. J. (2000) *Nature (London)* **408**, 433–439
229. Fotedar, A., Cannella, D., Fitzgerald, P., Rousselle, T., Gupta, S., Dorée, M., and Fotedar, R. (1996) *J. Biol. Chem.* **271**, 31627–31637
230. De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., and Kim, S.-H. (1993) *Nature (London)* **363**, 595–602
231. Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., and Pavletich, N. P. (1995) *Nature (London)* **376**, 313–320
232. Shiffman, D., Brooks, E. E., Brooks, A. R., Chan, S. C., and Milner, P. G. (1996) *J. Biol. Chem.* **271**, 12199–12204
233. Nurse, P. (1990) *Nature (London)* **344**, 503–507
234. Pines, J. (1993) *Trends Biochem. Sci.* **18**, 195–197
235. Solomon, M. J. (1994) *Trends Biochem. Sci.* **19**, 496–500
236. Lew, J., and Wang, J. H. (1995) *Trends Biochem. Sci.* **20**, 33–37
236a. Crawford, D. F., and Piwnica-Worms, H. (2001) *J. Biol. Chem.* **276**, 37166–37177
236b. Karsenti, E., and Vernos, I. (2001) *Science* **294**, 543–547
236c. Kastan, M. B. (2001) *Nature (London)* **410**, 766–767
236d. Davenport, R. J. (2001) *Science* **292**, 2415–2417
237. Morris, M. C., and Divita, G. (1999) *J. Mol. Biol.* **286**, 475–487
237a. Jacobs, H. W., Keidel, E., and Lehner, C. F. (2001) *EMBO J.* **20**, 2376–2386
238. Lopez-Girona, A., Furnari, B., Mondesert, O., and Russell, P. (1999) *Nature (London)* **397**, 172–175
238a. Yang, Q., Manicone, A., Coursen, J. D., Linke, S. P., Nagashima, M., Forgues, M., and Wang, X. W. (2000) *J. Biol. Chem.* **275**, 36892–36898
239. Reynolds, R. A., Yem, A. W., Wolfe, C. L., Deibel, M. R., Jr., Chidester, C. G., and Watenpaugh, K. D. (1999) *J. Mol. Biol.* **293**, 559–568
240. Hagting, A., Karlsson, C., Clute, P., Jackman, M., and Pines, J. (1998) *EMBO J.* **17**, 4127–4138
240a. Qian, Y.-W., Erikson, E., and Maller, J. L. (1998) *Science* **282**, 1701–1704
240b. Körner, K., Jérôme, V., Schmidt, T., and Müller, R. (2001) *J. Biol. Chem.* **276**, 9662–9669
241. Pines, J. (1999) *Nature (London)* **397**, 104 – 105
242. Sanchez, Y., Bachant, J., Wang, H., Hu, F., Liu, D., Tetzlaff, M., and Elledge, S. J. (1999) *Science* **286**, 1166–171
243. Kellogg, D. R., Moritz, M., and Alberts, B. M. (1994) *Ann. Rev. Biochem.* **63**, 639–674
244. Geier, B. M., Wiech, H., and Schiebel, E. (1996) *J. Biol. Chem.* **271**, 28366–28374
244a. Murray, A. W. (2001) *Science* **291**, 1499–1502
244b. Hinchcliffe, E. H., Miller, F. J., Cham, M., Khodjakov, A., and Sluder, G. (2001) *Science* **291**, 1547–1550
244c. Middendorp, S., Paoletti, A., Schiebel, E., and Bornens, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9141–9146
244d. do Carmo Avides, M., and Glover, D. M. (1999) *Science* **283**, 1733–1735

References

244e. Flory, M. R., Moser, M. J., Monnat, R. J., Jr., and Davis, T. N. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5919–5923
245. Nicklas, R. B. (1997) *Science* **275**, 632–637
246. Heald, R., Tournebize, R., Blank, T., Sandaltzopoulos, R., Becker, P., Hyman, A., and Karsenti, E. (1996) *Nature (London)* **382**, 420–425
246a. Fraschini, R., Beretta, A., Sironi, L., Musacchio, A., Lucchini, G., and Piatti, S. (2001) *EMBO J.* **20**, 6648–6659
247. Hyams, J. (1996) *Nature (London)* **382**, 397–398
247a. Bloom, K. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4757–4759
248. Dasso, M. (1993) *Trends Biochem. Sci.* **18**, 96–101
249. Ohba, T., Nakamura, M., Nishitani, H., and Nishimoto, T. (1999) *Science* **284**, 1356–1358
250. Wilde, A., and Zheng, Y. (1999) *Science* **284**, 1359–1362
250a. Gruneberg, U., Campbell, K., Simpson, C., Grindlay, J., and Schiebel, E. (2000) *EMBO J.* **19**, 6475–6488
250b. Martin-Lluesma, S., Stucke, V. M., and Nigg, E. A. (2002) *Science* **297**, 2267–2270
251. Barton, N. R., and Goldstein, L. S. B. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1735–1742
252. Lockhart, A., and Cross, R. A. (1996) *Biochemistry* **35**, 2365–2373
253. Pennisi, E. (1998) *Science* **279**, 477–478
254. Elledge, S. J. (1998) *Science* **279**, 999–1000
255. Hwang, L. H., Lau, L. F., Smith, D. L., Mistrot, C. A., Hardwick, K. G., Hwang, E. S., Amon, A., and Murray, A. W. (1998) *Science* **279**, 1041–1044
256. Yamano, H., Gannon, J., and Hunt, T. (1996) *EMBO J.* **15**, 5268–5279
257. Bachant, J. B., and Elledge, S. J. (1999) *Nature (London)* **398**, 757–758
258. Funabiki, H., Kumada, K., and Yanagida, M. (1996) *EMBO J.* **15**, 6617–6628
259. Page, A. M., and Hieter, P. (1999) *Ann. Rev. Biochem.* **68**, 583–609
259a. Prinz, S., and Amon, A. (1999) *Nature (London)* **402**, 133–135
259b. Nakaseko, Y., and Yanagida, M. (2001) *Nature (London)* **412**, 291–292
259c. Morgan, D. O., and Roberts, J. M. (2002) *Nature (London)* **418**, 495–496
260. Grossberger, R., Gieffers, C., Zachariae, W., Podtelejnikov, A. V., Schleiffer, A., Nasmyth, K., Mann, M., and Peters, J.-M. (1999) *J. Biol. Chem.* **274**, 14500–14507
261. Jiang, F., and Basavappa, R. (1999) *Biochemistry* **38**, 6471–6478
261a. Piel, M., Nordberg, J., Euteneuer, U., and Bornens, M. (2001) *Science* **291**, 1550–1553
262. Wisdom, R., Johnson, R. S., and Moore, C. (1999) *EMBO J.* **18**, 188–197
263. Takenaka, K., Moriguchi, T., and Nishida, E. (1998) *Science* **280**, 599–602
263a. Wilkinson, M. G., and Millar, J. B. A. (2000) *FASEB J.* **14**, 2147–2157
263b. Graves, L. M. (2000) *Nature (London)* **403**, 328–332
263c. Whitmarsh, A. J., and Davis, R. J. (2000) *Nature (London)* **403**, 255–256
264. Gottesfeld, J. M., and Forbes, D. J. (1997) *Trends Biochem. Sci.* **22**, 197–202
264a. Goldstein, L. S. B. (2001) *Science* **291**, 2102–2103
265. Martínez-Botas, J., Suárez, Y., Ferruelo, A. J., Gómez-Coronado, D., and Lasunción, M. A. (1999) *FASEB J.* **13**, 1359–1370
266. Bourne, Y., Arvai, A. S., Bernstein, S. L., Watson, M. H., Reed, S. I., Endicott, J. E., Noble, M. E., Johnson, L. N., and Tainer, J. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10232–10236
267. Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I., and Tainer, J. A. (1993) *Science* **262**, 387–395
267a. Rubin, G. M., and Lewis, E. B. (2000) *Science* **287**, 2216–2218
268. Yamamoto, M. (1996) *Trends Biochem. Sci.* **21**, 18–22
269. de Lange, T. (1998) *Nature (London)* **392**, 753–754
270. Conrad, M. N., Dominguez, A. M., and Dresser, M. E. (1997) *Science* **276**, 1252–1255
271. McKim, K. S., Green-Marroquin, B. L., Sekelsky, J. J., Chin, G., Steinberg, C., Khodosh, R., and Hawley, R. S. (1998) *Science* **279**, 876–878
271a. Lindgren, A., Bungard, D., Pierce, M., Xie, J., Vershon, A., and Winter, E. (2000) *EMBO J.* **19**, 6489–6497
272. Orr-Weaver, T. L. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10443–10449
273. McKim, K. S., and Hawley, R. S. (1995) *Science* **270**, 1595–1601
274. Picard, A., Galas, S., Peaucellier, G., and Dorée, M. (1996) *EMBO J.* **15**, 3590–3598
275. Lenormand, J.-L., Dellinger, R. W., Knudsen, K. E., Subramani, S., and Donoghue, D. J. (1999) *EMBO J.* **18**, 1869–1877
276. Sette, C., Barchi, M., Bianchini, A., Conti, M., Rossi, P., and Geremia, R. (1999) *J. Biol. Chem.* **274**, 33571–33579
276a. Xu, R., Wilson, W. K., and Matsuda, S. P. T. (2002) *J. Am. Chem. Soc.* **124**, 918–919
277. Watanabe, Y., and Nurse, P. (1999) *Nature (London)* **400**, 461–464
278. Roeder, G. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10450–10456
278a. Albini, S. M., and Jones, G. H. (1987) *Chromosoma* **95**, 324–328
279. Kleckner, N. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8167–8174
279a. Nabeshima, K., Kakihara, Y., Hiraoka, Y., and Nojima, H. (2001) *EMBO J.* **20**, 3871–3881
279b. Yuan, L., Liu, J.-G., Hoja, M.-R., Wilbertz, J., Nordqvist, K., and Höög, C. (2002) *Science* **296**, 1115–1118
279c. Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O., and Petes, T. D. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11383–11390
279d. Stack, S. M. (1991) *Genome* **34**, 900–908
280. Boy de la Tour, E., and Laemmli, U. K. (1988) *Cell* **55**, 937–944
281. Gallardo, M. H., Bickham, J. W., Honeycutt, R. L., Ojeda, R. A., and Köhler, N. (1999) *Nature (London)* **401**, 341
281a. Shonn, M. A., McCarroll, R., and Murray, A. W. (2000) *Science* **289**, 300–303
281b. Sluder, G., and McCollum, D. (2000) *Science* **289**, 254–255
282. Pirrotta, V. (1984) *Trends Biochem. Sci.* **9**, 220–221
283. Burtis, K. C., and Hawley, R. S. (1999) *Nature (London)* **401**, 125–127
284. Kao, F.-t, and Yu, J.-w. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1844–1848
285. Sager, R. (1965) *Sci. Am.* **212**(Jan), 71–79
286. Goodenough, U. W., and Levine, R. P. (1970) *Sci. Am.* **223**(Nov), 22–29
287. Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsodi, J., Boll, W., Cantell, K., and Weissmann, C. (1980) *Nature (London)* **284**, 316–320
288. Olson, M. V. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4338–4344
289. Collins, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10821–10823
290. Guyer, M. S., and Collins, F. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10841–10848
290a. International Human Genome Sequencing Consortium. (2001) *Nature (London)* **409**, 860–921
290b. Pennisi, E. (2001) *Science* **291**, 1177–1180
290c. Venter, J. C., and many others. (2001) *Science* **291**, 1304–1351
290d. Olivier, M., and 54 other authors. (2001) *Science* **291**, 1298–1302
290e. Shoemaker, D. D., and 34 other authors. (2001) *Nature (London)* **409**, 922–927
290f. Caron, H., and 12 other authors. (2001) *Science* **291**, 1289–1292
290g. Deloukas, P., and 126 other authors. (2001) *Nature (London)* **414**, 865–871
290h. Claverie, J.-M. (2001) *Science* **291**, 1255–1257
290i. Pennisi, E. (2002) *Science* **296**, 1600–1601
290j. Kim, S. K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J. M., Eizinger, A., Wylie, B. N., and Davidson, G. S. (2001) *Science* **293**, 2087–2092
291. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998) *Science* **282**, 682–689
291a. Malakoff, D. (2000) *Science* **288**, 248–253
291b. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. (2001) *Nature (London)* **409**, 685–690
291c. The International Mouse Mutagenesis Consortium. (2001) *Science* **291**, 1251–1255
291d. Gregory, S. G., and 85 other authors (2002) *Nature (London)* **418**, 743–750
291e. Mural, R. J., and 176 other authors (2002) *Science* **296**, 1661–1671
292. Roach, J. C., Siegel, A. F., van den Engh, G., Trask, B., and Hood, L. (1999) *Nature (London)* **401**, 843–845
292a. Marshall, E. (2001) *Science* **291**, 1872
292b. Vogel, G. (2000) *Science* **290**, 1671
292c. Aparicio, S., and 40 other authors (2002) *Science* **297**, 1301–1310
292d. Gardner, M. J., and 44 other authors (2002) *Nature (London)* **419**, 498–511
292e. Doolittle, R. F. (2002) *Nature (London)* **419**, 493–494
292f. Carlton, J. M., and 43 other authors (2002) *Nature (London)* **419**, 512–519
292g. Holt, R. A., and 121 other authors (2002) *Science* **298**, 129–149
292h. Kaufman, T. C., Severson, D. W., and Robinson, G. E. (2002) *Science* **298**, 97–115
293. White, R., and Caskey, C. T. (1988) *Science* **240**, 1483–1488
294. McKusick, V. A., and Amberger, J. S. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 119–125, McGraw-Hill, New York
295. Sawyer, J. R., and Hozier, J. C. (1986) *Science* **232**, 1632–1635
296. Tunnacliffe, A., Benham, F., and Goodfellow, P. (1984) *Trends Biochem. Sci.* **9**, 5–7
297. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994) *Molecular Biology of the Cell*, 3rd ed., Garland, New York
298. Franke, U. (1981) *Cytogenet. Cell Genet.* **31**, 24–32
299. Gray, J. W., and Langlois, R. G. (1986) *Ann. Rev. Biophys. Biophys. Chem.* **15**, 195–235
300. Chumakov, I., and plus 35 other authors (1992) *Nature (London)* **359**, 380–387
301. Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8794–8797
302. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998) *Science* **280**, 1540–1542
303. Marx, J. (1995) *Science* **270**, 1919–1920

References

304. Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J., and Weissenbach, J. (1996) *Nature (London)* **380**, 152–154, iii–v + supplement
305. Rogers, J. (1999) *Science* **286**, 429–432
305a. Liu, S., Ren, H., Gao, Q., Roach, D. J., Loder, R. T., Jr., Armstrong, T. M., Mao Q, Blaga, I., Barker, D. L., and Jovanovich, S. B. (2000) *Proc. Natl. Acad. Sci., U.S.A.* **97**, 5369 – 5374
305b. Paegel, B. M., Emrich, C. A., Wedemayer, G. J., Scherer, J. R., and Mathies, R. A. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 574 – 579
306. Marshall, E. (1999) *Science* **284**, 1906–1909
306a. Green, P. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4143 – 4144
307. Donis-Keller, H., and 32 other authors (1987) *Cell* **51**, 319–337
308. Wang, D. G., and 26 other authors (1998) *Science* **280**, 1077–1082
308a. The International SNP Map Working Group. (2001) *Nature (London)* **409**, 928–933
308b. Patil, N., and 21 other authors (2001) *Science* **294**, 1719–1723
309. Hudson, T. J., and 50 other authors (1995) *Science* **270**, 1945–1954
310. Schuler, G. D., and 103 other authors (1996) *Science* **274**, 540–546
311. Wang, S. M., and Rowley, J. D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11909–11914
312. Dunham, I., Shimizu, N., Roe, B. A., Chossoe, S., and 117 other authors (1999) *Nature (London)* **402**, 489–495
313. Butler, D. (1999) *Nature (London)* **402**, 447–448
313a. The International Human Genome Mapping Consortium. (2001) *Nature (London)* **409**, 934–941
314. Pandey, A., and Lewitter, F. (1999) *Trends Biochem. Sci.* **24**, 276–280
315. Pennisi, E. (1999) *Science* **286**, 447–450
316. Baxevanis, A. D., and Ouellette, B. F. F., eds. (1998) *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley, New York
317. Bishop, M. J., ed. (1999) *Genetics Databases*, Academic Press, San Diego, California
318. Durbin, R., Eddy, S., Krogh, A., and Mitchinson, G., eds. (1998) *Biological Sequence Analysis*, Cambridge Univ. Press., Cambridge
319. Bishop, M. J., ed. (1998) *Human Genome Computing*, 2nd ed., Academic Press, San Diego, California
320. Calvet, J. P. (1998) *Science* **282**, 1057–1058
321. Kereil, D. P., and Etzold, T. (1999) *Trends Biochem. Sci.* **24**, 155–157
321a. Tamames, J., and Tramontano, A. (2000) *Trends Biochem. Sci.* **25**, 402–403
321b. Jones, D. T., and Swindells, M. B. (2002) *Trends Biochem. Sci.* **27**, 161–164
321c. Zafar, N., Mazumder, R., and Seto, D. (2001) *Trends Biochem. Sci.* **26**, 514–516
322. Boguski, M. S. (1995) *N. Engl. J. Med.* **333**, 645–647
323. Borodovsky, M., Koonin, E. V., and Rudd, K. E. (1994) *Trends Biochem. Sci.* **19**, 309–313
324. Brown, N. P., Whittaker, A. J., Newell, W. R., Rawlings, C. J., and Beck, S. (1995) *J. Mol. Biol.* **249**, 342–359
325. Pennisi, E. (1998) *Science* **280**, 1692–1693
325a. Matthiessen, M. W. (2002) *Trends Biochem. Sci.* **27**, 586 – 588
326. Little, P. (1999) *Nature (London)* **402**, 467–468
327. O'Brien, S. J., Menotti-Raymond, M., Murphy, W. J., Nash, W. G., Wienberg, J., Stanyon, R., Copeland, N. G., Jenkins, N. A., Womack, J. E., and Graves, J. A. M. (1999) *Science* **286**, 458–481
327a. Gura, T. (2001) *Science* **293**, 593–595
327b. Buetow, K. H., Edmonson, M., MacDonald, R., Clifford, R., Yip, P., Kelley, J., Little, D. P., Strausberg, R., Koester, H., Cantor, C. R., and Braun, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 581–584
327c. Couzin, J. (2002) *Science* **296**, 1391,1393
327d. Adam, D. (2001) *Nature (London)* **412**, 105
327e. Yan, H., Kinzler, K. W., and Vogelstein, B. (2000) *Science* **289**, 1890–1892
327f. Wright, A. F., and Van Heyningen, V. (2001) *Nature (London)* **414**, 705–706
327g. Marshall, E. (2001) *Science* **294**, 2272–2274
327h. Trager, R. S. (2002) *Science* **298**, 947
328. Pääbo, S. (1999) *Trends Biochem. Sci.* **24**, M13–M16
328a. Templeton, A. R. (2002) *Nature (London)* **416**, 45–51
329. Olsen, G. J., and Woese, C. R. (1993) *FASEB J.* **7**, 113–123
330. Shenk, M. A., and Steele, R. E. (1993) *Trends Biochem. Sci.* **18**, 459–463
331. Harris, E. E., and Hey, J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3320–3324
331a. Adcock, G. J., Dennis, E. S., Easteal, S., Huttley, G. A., Jermiin, L. S., Peacock, W. J., and Thorne, A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 537–542
331b. Relethford, J. H. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 390–391
331c. Jones, M. (2001) *The Molecule Hunt: Archaeology and the Search for Ancient DNA*, Allen Lane,
331d. Fish, S. A., Shepherd, T. J., McGenity, T. J., and Grant, W. D. (2002) *Nature (London)* **417**, 432–436
332. Kahn, P., and Gibbons, A. (1997) *Science* **277**, 176–178
332a. Gibbons, A. (2001) *Science* **292**, 627–629
332b. Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T. D., Itoh, T., Tsai, S.-F., and 11 other authors. (2002) *Science* **295**, 131–134
332c. Enard, W., Khaitovich, P., Klose, J., Zöllner, S., and 9 other authors. (2002) *Science* **296**, 340–343
332d. Cyranoski, D. (2002) *Nature (London)* **418**, 910–912
332e. Stone, A. C., Griffiths, R. C., Zegura, S. L., and Hammer, M. F. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 43–48
332f. Britten, R. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13633–13635
333. Gibbons, A. (1998) *Science* **281**, 1432–1434
334. Kennedy, E. P. (1992) *Ann. Rev. Biochem.* **61**, 1–28
335. Barbujani, G., Magagni, A., Minch, E., and Cavalli-Sforza, L. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4516–4519
336. Molnar, S. (1998) *Human Variation, Races, Types and Ethnic Groups*, 4th ed., Prentice Hall, Upper Saddle River, New Jersey (pp. 230–247)
336a. Jablonski, N. G., and Chaplin, G. (2002) *Sci. Am.* **287**(Oct), 74–81
337. Rannala, B., and Mountain, J. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9197–9201
338. Sugimoto, C., Kitamura, T., Guo, J., Al-Ahdal, M. N., Shchelkunov, S. N., Otova, B., Ondrejka, P., Chollet, J.-Y., El-Safi, S., Ettayebi, M., Grésenguet, G., Kocagöz, T., Chaiyarasamee, S., Thant, K. Z., Thein, S., Moe, K., Kobayashi, N., Taguchi, F., and Yogo, Y. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9191–9196
338a. Pääbo, S. (2001) *Science* **291**, 1219–1220
338b. Pennisi, E. (2001) *Science* **291**, 1733–1734
338c. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J., and Feldman, M. W. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360–7365
338d. Gibbons, A. (2000) *Science* **290**, 1080–1081
338e. Cann, R. L. (2001) *Science* **291**, 1742–1747
339. De Mendoza, D. H., and Braginski, R. (1999) *Science* **283**, 1439–1440
340. Bradley, D. G., MacHugh, D. E., Cunningham, P., and Loftus, R. T. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5131–5135
341. Makalowski, W., and Boguski, M. S. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9407–9412
342. Nadeau, J. H., Grant, P. L., Mankala, S., Reiner, A. H., Richardson, J. E., and Eppig, J. T. (1995) *Nature (London)* **373**, 363–365
343. Collins, F. S., and Jegalian, K. G. (1999) *Sci. Am.* **281**(Dec), 86–91
344. O'Brien, S. J., and Stanyon, R. (1999) *Nature (London)* **402**, 365–366
344a. Hedges, S. B., and Kumar, S. (2002) *Science* **297**, 1283 – 1285
344b. Knight, J. (2002) *Nature (London)* **417**, 374–376
345. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) *Science* **278**, 631–637
346. Gibbons, A. (1998) *Science* **280**, 675–676
347. Rowe, T. (1999) *Nature (London)* **398**, 283–284
348. Bromham, L., Rambaut, A., Fortey, R., Cooper, A., and Penny, D. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12386–12389
349. Knoll, A. H. (1999) *Science* **285**, 1025–1026
350. Maniloff, J. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10004–10006
351. Lawrence, J. G., and Ochman, H. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9413–9417
352. Doolittle, W. F. (1999) *Trends Biochem. Sci.* **24**, M5–M8
353. Lake, J. A., Jain, R., and Rivera, M. C. (1999) *Science* **283**, 2027–2028
354. Pennisi, E. (1998) *Science* **280**, 672–674
355. Stein, J. L., and Simon, M. I. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6228–6230
356. Davies, J. (1999) *Trends Biochem. Sci.* **24**, M2–M5
356a. Doolittle, R. F. (2002) *Nature (London)* **416**, 697–700
356b. Bentley, S. D. and 42 other authors (2002) *Nature (London)* **417**, 141–147
356c. Barnett, M. J., and 25 other authors (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9883–9888
356d. Galibert, F., and 55 other authors (2001) *Science* **293**, 668–672
356e. Heidelberg, J. F., and 31 other authors (2000) *Nature (London)* **406**, 477–483
356f. Parkhill, J., and 40 other authors (2001) *Nature (London)* **413**, 848–852
356g. Parkhill, J., and 34 other authors (2001) *Nature (London)* **413**, 523–527
356h. Moreno, E., and Moriyón, I. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1–3
356i. Cole, S. T., and 43 other authors (2001) *Nature (London)* **409**, 1007–1011
356j. Enserink, M. (2002) *Science* **295**, 1442–1443
356k. Tettelin, H., and 38 other authors (2001) *Science* **293**, 498–506
356l. Stover, C. K., and 30 other authors (2000) *Nature (London)* **406**, 959–964
356m. Glaser, P., and 54 other authors (2001) *Science* **294**, 849–853
356n. Perna, N. T., and 27 other authors (2001) *Nature (London)* **409**, 529–533
356o. Glass, J. I., Lefkowitz, E. J., Glass, J. S., Heiner, C. R., Chen, E. Y., and Cassell, G. H. (2000) *Nature (London)* **407**, 757–761
356p. Nierman, W. C., and 36 other authors (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4136–4141
356q. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000) *Nature (London)* **407**, 81–86
356r. Ng, W. V., Kennedy, S. P., Mahairas, G. G., and 40 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12176 – 12181

References

356s. Kawashima, T., Amano, N., Koike, H., Makino, S.-i, Higuchi, S., and 10 other authors. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14257– 14262
356t. Cole, S. T., and 41 other authors. (1998) *Nature (London)* **393**, 537 – 544
357. Edwards, J. S., and Palsson, B. O. (1999) *J. Biol. Chem.* **274**, 17410 – 17416
358. Strauss, E. J., and Falkow, S. (1997) *Science* **276**, 707 – 711
358a. da Silva, A. C. R., and 64 other authors. (2002) *Nature (London)* **417**, 459 – 463
359. Karp, P. D. (1998) *Trends Biochem. Sci.* **23**, 114 – 116
359a. Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., and Palsson, B. O. (2001) *Trends Biochem. Sci* **26**, 179 – 186
360. Hutchison, C. A., III, Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O., and Venter, J. C. (1999) *Science* **286**, 2165 – 2169
361. Mushegian, A. R., and Koonin, E. V. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10268 – 10273
361a. Szostak, J. W., Bartel, D. P., and Luisi, P. L., (2001) *Nature (London)* **409**, 387 – 390
362. Cho, M. K., Magnus, D., Caplan, A. L., McGee, D., and Ethics of Genomics Group. (1999) *Science* **286**, 2087 – 2090
362a. Stephanopoulos, G., and Kelleher, J. (2001) *Science* **292**, 2024 – 2025
362b. Holden, C. (2002) *Science* **297**, 1459 – 1460
363. Winzeler, E. A., and 51 other authors (1999) *Science* **285**, 901 – 906
364. Hieter, P. (1999) *Nature (London)* **402**, 362 – 363
364a. Ross-Macdonald, P., and 17 other authors (1999) *Nature (London)* **402**, 413 – 418
365. Glaever, G., and 72 other authors. (2002) *Nature (London)* **418**, 387 – 391
365a. Eisen, J. A. (2002) *Nature (London)* **415**, 845 – 848
365b. Wood, V., and 132 other authors (2002) *Nature (London)* **415**, 871– 880
365c. Seo, H.-C., Kube, M., Edvardsen, R. B., Jensen, M. F., Beck, A., Spriet, E., Gorsky, G., Thompson, E. M., Leharch, H., Reinhardt, R., and Chourrout, D. (2001) *Science* **294**, 2506
366. Fields, S., Kohara, Y., and Lockhart, D. J. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8825 – 8826
367. Tabara, H., Motohashi, T., and Kohara, Y. (1996) *Nucleic Acids Res.* **24**, 2119 – 2124
368. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) *Science* **278**, 680 – 686
369. Walhout, A. J. M., Sordella, R., Lu, X., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., and Vidal, M. (2000) *Science* **287**, 116 – 122
370. Hutter, H., Vogel, B. E., Plenefisch, J. D., Norris, C. R., Proenca, R. B., Spieth, J., Guo, C., Mastwal, S., Zhu, X., Scheel, J., and Hedgecock, E. M. (2000) *Science* **287**, 989 – 994
370a. Zdobnov, E. M., and 35 other authors. (2002) *Science* **298**, 149 – 159
370b. Miki, R., and 29 other authors. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2199 – 2204
371. Lin, X., and 36 other authors (1999) *Nature (London)* **402**, 761 – 768
372. European Union Arabidopsis Genome Sequencing Consortium, and Cold Spring Harbor, W. U. in St. L. and P. E. B. A. S. C. (1999) *Nature (London)* **402**, 769 – 777
373. Mathé, C., Peresetsky, A., Déhais, P., Van Montagu, M., and Rouzé, P. (1999) *J. Mol. Biol.* **285**, 1977 – 1991
374. Gura, T. (2000) *Science* **287**, 412 – 414
375. Meyerowitz, E. M. (1999) *Nature (London)* **402**, 731 – 732
375a. The Arabidopsis Genome Initiative. (2000) *Nature (London)* **408**, 796 – 815
375b. Seki, M., and 19 other authors (2002) *Science* **296**, 141– 145
375c. Pennisi, E. (2000) *Science* **290**, 32 – 35
375d. Normile, D., and Pennisi, E. (2002) *Science* **296**, 32 – 39
375e. Bennetzen, J. (2002) *Science* **296**, 60 – 63
375f. Yu, J., and 99 other authors (2002) *Science* **296**, 79 – 92
375g. Goff, S. A., and 54 other authors (2002) *Science* **296**, 92 – 100
376. Somerville, C., and Somerville, S. (1999) *Science* **285**, 380 – 383
377. Pennisi, E. (1998) *Science* **282**, 652 – 654
378. Capy, P. (2000) *Science* **287**, 985 – 986
379. Scriver, C. R., Kaufman, S., Eisensmith, R. C., and Woo, S. L. C. (1995) in *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed., Vol. 1 (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds), pp. 1015 – 1075, McGraw-Hill, New York
380. Barinaga, M. (1998) *Science* **281**, 32 – 34
381. Schultheis, P. J., Lorenz, J. N., Meneton, P., Nieman, M. L., Riddle, T. M., Flagella, M., Duffy, J. J., Doetschman, T., Miller, M. L., and Shull, G. E. (1998) *J. Biol. Chem.* **273**, 29150 – 29155
382. Chairoungdua, A., Segawa, H., Kim, J. Y., Miyamoto, K.-i, Haga, H., Fukui, Y., Mizoguchi, K., Ito, H., Takeda, E., Endou, H., and Kanai, Y. (1999) *J. Biol. Chem.* **274**, 28845 – 28848
383. Chambers, E. J., Bloomberg, G. B., Ring, S. M., and Tanner, M. J. A. (1999) *J. Mol. Biol.* **285**, 1289 – 1307
384. Annunen, S., Paassilta, P., Lohiniva, J., Perälä, M., Pihlajamaa, T., Karppinen, J., Tervonen, O., Kröger, H., Lähde, S., Vanharanta, H., Ryhänen, L., Göring, H. H. H., Ott, J., Prockop, D. J., and Ala-Kokko, L. (1999) *Science* **285**, 409 – 412
385. Ward, C. J., Turley, H., Ong, A. C. M., Comley, M., Biddolph, S., Chetty, R., Ratcliffe, P. J., Gatter, K., and Harris, P. C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1524 – 1528
386. Hugot, J.-P., 16 other authors, and Groupe d'Etude Thérapeutique des Affections Inflammatoires Digestives. (1996) *Nature (London)* **379**, 821 – 823
387. Ghosh, S., and 39 other authors (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2198 – 2203
388. Kinzler, K. W., and Vogelstein, B. (1993) *Nature (London)* **363**, 495
389. Brehm, A., and Kouzarides, T. (1999) *Trends Biochem. Sci.* **24**, 142 – 145
390. Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. (1994) *Science* **265**, 346 – 355
391. Oren, M. (1999) *J. Biol. Chem.* **274**, 36031 – 36034
392. Venkitaraman, A. R. (1999) *Science* **286**, 1100 – 1102
393. Smith, J. R., and 22 other authors (1996) *Science* **274**, 1371 – 1374
394. Novelli, M. R., Williamson, J. A., Tomlinson, I. P. M., Elia, G., Hodgson, S. V., Talbot, I. C., Bodmer, W. F., and Wright, N. A. (1996) *Science* **272**, 1187 – 1190
395. Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., and Perucho, M. (1997) *Science* **275**, 967 – 969
396. Rabbitts, T. H. (1994) *Nature (London)* **372**, 143 – 149
397. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) *Nature (London)* **396**, 643 – 649
398. Kappel, C. A., Bieberich, C. J., and Jay, G. (1994) *FASEB J.* **8**, 583 – 592
399. Ahmad, W., and 17 other authors (1998) *Science* **279**, 720 – 724
400. Bassett, D. E., Jr., Boguski, M. S., and Hieter, P. (1996) *Nature (London)* **379**, 589 – 590
401. Richards, R. I., and Sutherland, G. R. (1997) *Trends Biochem. Sci.* **22**, 432 – 436
402. Wang, Y.-H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) *J. Mol. Biol.* **263**, 511 – 516
403. Weisman-Shomer, P., Naot, Y., and Fry, M. (2000) *J. Biol. Chem.* **275**, 2231 – 2238
404. Fridell, R. A., Benson, R. E., Hua, J., Bogerd, H. P., and Cullen, B. R. (1996) *EMBO J.* **15**, 5408 – 5414
405. Mandel, J.-L. (1997) *Nature (London)* **386**, 767 – 769
405a. Moine, H., and Mandel, J.-L. (2001) *Science* **294**, 2487– 2488
405b. Greenough, W. T., Klintsova, A. Y., Irwin, S. A., Galvez, R., Bates, K. E., and Weiler, I. J. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7101– 7106
405c. Sohn, E. (2001) *Science* **294**, 1809
405d. Deng, J., Xiong, Y., and Sundaralingam, M. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13665 – 13670
406. Wells, R. D. (1996) *J. Biol. Chem.* **271**, 2875 – 2878
407. Timchenko, L. T., and Caskey, C. T. (1996) *FASEB J.* **10**, 1589 – 1597
407a. Sinden, R. R. (2001) *Nature (London)* **411**, 757– 758
408. Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996) *Science* **272**, 548 – 551
408a. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. W. (2001) *Science* **293**, 864 – 867
409. Singer, R. H. (1998) *Science* **280**, 696 – 697
409a. Dhe-Paganon, S., Shigeta, R., Chi, Y.-I., Ristow, M., and Shoelson, S. E. (2000) *J. Biol. Chem.* **275**, 30753 – 30756
409b. Lodi, R., Cooper, J. M., Bradley, J. L., Manners, D., Styles, P., Taylor, D. J., and Schapira, A. H. V. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11492 – 11495
409c. Chen, O. S., and Kaplan, J. (2000) *J. Biol. Chem.* **275**, 7626 – 7632
409d. Foury, F., and Talibi, D. (2001) *J. Biol. Chem.* **276**, 7762 – 7768
409e. Gakh, O., Adamec, J., Gacy, A. M., Twesten, R. D., Owen, W. G., and Isaya, G. (2002) *Biochemistry* **41**, 6798 – 6804
410. Campuzano, V., and 26 other authors (1996) *Science* **271**, 1423 – 1427
410a. LeProust, E. M., Pearso, C. E., Sinden, R. R., and Gao, X. (2000) *J. Mol. Biol.* **302**, 1063 – 1080
410b. Rosenberg, R. N. (1995) *N. Engl. J. Med.* **333**, 1351– 1353
411. Mariappan, S. V. S., Catasti, P., Silds, L. A., III, Bradbury, E. M., and Gupta, G. (1999) *J. Mol. Biol.* **285**, 2035 – 2052
412. Dürr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., Mandel, J.-L., Brice, A., and Koenig, M. (1996) *N. Engl. J. Med.* **335**, 1169 – 1175
413. Rosenberg, R. N. (1996) *N. Engl. J. Med.* **335**, 1222 – 1224
414. Perutz, M. F. (1999) *Trends Biochem. Sci.* **24**, 58 – 63
414a. Holbert, S., Denghien, I., Kiechle, T., Rosenblatt, A., Wellington, C., Hayden, M. R., Margolis, R. L., Ross, C. A., Dausset, J., Ferrante, R. J., and Néri, C. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 1811– 1816
414b. Peters, M. F., and Ross, C. A. (2001) *J. Biol. Chem.* **276**, 3188 – 3194
414c. Bates, G. P. (2001) *Nature (London)* **413**, 691– 694
415. Wellington, C. L., and 20 other authors (1998) *J. Biol. Chem.* **273**, 9158 – 9167

References

416. Karlin, S., and Burge, C. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1560–1565
416a. Kaytor, M. D., and Warren, S. T. (1999) *J. Biol. Chem.* **274**, 37507– 37510
416b. Perutz, M. F., and Windle, A. H. (2001) *Nature (London)* **412**, 143 – 144
416c. Chai, Y., Wu, L., Griffin, J. D., and Paulson, H. L. (2001) *J. Biol. Chem.* **276**, 44889 – 44897
417. Diamond, M. L., Robinson, M. R., and Yamamoto, K. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 657–661
417a. Fernandez-Funez, P., and 14 other authors (2000) *Nature (London)* **408**, 101– 106
417b. Mushegian, A. R., Vishnivetskiy, S. A., and Gurevich, V. V. (2000) *Biochemistry* **39**, 6809 – 6813
418. Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1997) *Nature (London)* **389**, 971–974
419. Collick, A., Norris, M. L., Allen, M. J., Bois, P., Barton, S. C., Surani, M. A., and Jeffreys, A. J. (1994) *EMBO J.* **13**, 5745–5753
420. Sutherland, G. R., and Richards, R. I. (1994) *N. Engl. J. Med.* **331**, 191–193
421. McMurray, C. T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1823–1825
422. Warpeha, K. M., Xu, W., Liu, L., Charles, I. G., Patterson, C. C., Ah-Fat, F., Harding, S., Hart, P. M., Chakravarthy, U., and Hughes, A. E. (1999) *FASEB J.* **13**, 1825–1832
423. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) *J. Biol. Chem.* **273**, 5204–5210
424. Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L., and Sinden, R. R. (1998) *Biochemistry* **37**, 2701–2708
425. Iyer, R. R., Pluciennik, A., Rosche, W. A., Sinden, R. R., and Wells, R. D. (2000) *J. Biol. Chem.* **275**, 2174–2184
426. Jakupciak, J. P., and Wells, R. D. (1999) *J. Biol. Chem.* **274**, 23468–23479
427. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. A. (1996) *Science* **274**, 610–614
428. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) *Science* **286**, 531–537
429. Service, R. F. (1998) *Science* **282**, 396–399
430. Gerhold, D., Rushmore, T., and Caskey, C. T. (1999) *Trends Biochem. Sci.* **24**, 168–173
430a. Pennisi, E. (2002) *Science* **297**, 1985, 1987
430b. Schena, M. (2002) *Microarray Analysis*, Wiley, New York
431. Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G., and Barany, F. (1999) *J. Mol. Biol.* **292**, 251–262
432. Tang, K., Fu, D.-J., Julien, D., Braun, A., Cantor, C. R., and Köster, H. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10016–10020
433. Marshall, E. (1999) *Science* **286**, 444–447
434. Dalton, R. (2000) *Nature (London)* **403**, 234
435. Burns, M. A., Johnson, B. N., Brahmasandra, S. N., Handique, K., Webster, J. R., Krishnan, M., Sammarco, T. S., Man, P. M., Jones, D., Heldsinger, D., Mastrangelo, C. H., and Burke, D. T. (1998) *Science* **282**, 484–487
436. Service, R. F. (1998) *Science* **282**, 399–401
437. Landegren, U., Kaiser, R., Caskey, C. T., and Hood, L. (1988) *Science* **242**, 229–237
438. Kolata, G. (1986) *Science* **234**, 22–24
439. Schnieke, A. E., Kind, A. J., Ritchie, W. A., Mycock, K., Scott, A. R., Ritchie, M., Wilmut, I., Colman, A., and Campbell, K. H. S. (1997) *Science* **278**, 2130–2133
440. Arnon, R. (1986) *Trends Biochem. Sci.* **11**, 521–524
441. McAleer, W. J., Buynak, E. B., Maigetter, R. Z., Wampler, D. E., Miller, W. J., and Hilleman, M. R. (1984) *Nature (London)* **307**, 178–180
442. Kapusta, J., Modelska, A., Figlerowicz, M., Pniewski, T., Letellier, M., Lisowa, O., Yusibov, V., Koprowski, H., Plucienniczak, A., and Legocki, A. B. (1999) *FASEB J.* **13**, 1796–1799
443. Weiner, D. B., and Kennedy, R. C. (1999) *Sci. Am.* **281**(Jul), 50–57
444. Taubes, G. (1997) *Science* **278**, 1711–1714
445. McDonald, G. A., Anacker, R. L., and Garjian, K. (1987) *Science* **235**, 83–85
446. Brown, F., Schild, G. C., and Ada, G. L. (1986) *Nature (London)* **319**, 549–550
447. Rupprecht, C. E., Wiktor, T. J., Johnston, D. H., Hamir, A. N., Dietzschold, B., Wunner, W. H., Glickman, L. T., and Koprowski, H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7947–7950
448. Waldmann, T. A. (1986) *Science* **232**, 727–732
449. Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R. A., and Barr, P. J. (1985) *J. Biol. Chem.* **260**, 4384–4389
450. Parkman, R. (1986) *Science* **232**, 1373–1378
451. Sandig, V., Youil, R., Bett, A. J., Franlin, L. L., Oshima, M., Maione, D., Wang, F., Metzker, M. L., Savino, R., and Caskey, C. T. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1002–1007
452. Verma, I. M., and Somia, N. (1997) *Nature (London)* **389**, 239–242
453. Anderson, W. F. (1995) *Sci. Am.* **273**(Sep), 124–128
454. Marshall, E. (1995) *Science* **269**, 1050–1055
455. Zanjani, E. D., and Anderson, W. F. (1999) *Science* **285**, 2084–2088
456. Finkel, T., and Epstein, S. E. (1995) *FASEB J.* **9**, 843–851
456a. Miller, D. G., and Stamatoyannopoulos, G. (2001) *N. Engl. J. Med.* **344**, 1782 – 1783
457. Miller, H. I. (2000) *Science* **287**, 591–592
457a. Gura, T. (2001) *Science* **291**, 1692 – 1697
458. Amalfitano, A., McVie-Wylie, A. J., Hu, H., Dawson, T. L., Raben, N., Plotz, P., and Chen, Y. T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8861–8866
458a. Pawliuk, R., Westerman, K. A., Fabry, M. E., Payen, E., Tighe, R., Bouhassira, E. E., Acharya, S. A., Ellis, J., London, I. M., Eaves, C. J., Humphries, R. K., Beuzard, Y., Nagel, R. L., and Leboulch, P. (2001) *Science* **294**, 2368 – 2371
458b. Dunder, U., Kaartinen, V., Valtonen, P., Väänänen, E., Kosma, V.-M., Heisterkamp, N., Groffen, J., and Mononen, I. (2000) *FASEB J.* **14**, 361– 367
459. Capecchi, M. R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 956–957
460. Marshall, E. (1996) *Science* **271**, 586–588
461. Kohn, D. B., and Parkman, R. (1997) *FASEB J.* **11**, 635–639
462. Broitman, S., Amosova, O., Dolinnaya, N. G., and Fresco, J. R. (1999) *J. Biol. Chem.* **274**, 21763–21768
462a. Knight, J. (2001) *Nature (London)* **413**, 12 – 15
463. Alper, J. (1999) *Science* **283**, 1625–1626
464. Kishore, G. M., and Shewmaker, C. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5968–5972
465. Trewavas, A. (1999) *Nature (London)* **402**, 231–232
466. Goodman, R. M., Hauptli, H., Crossway, A., and Knauf, V. C. (1987) *Science* **236**, 48–54
467. Jones, M. G. K. (1985) *Nature (London)* **317**, 579–580
468. Roberts, L. (1988) *Science* **241**, 1290
469. Frommer, W. B., Ludewig, U., and Rentsch, D. (1999) *Science* **285**, 1222–1223
470. Kling, J. (1996) *Science* **274**, 180–181
471. Gura, T. (1999) *Science* **285**, 994–995
472. Simon Moffat, A. (1992) *Science* **256**, 770–771
473. Thomas, S. M., Davies, A. R. W., Birtwistle, N. J., Crowther, S. M., and Burke, J. F. (1996) *Nature (London)* **380**, 387–388
474. Marshall, E. (1996) *Science* **273**, 1788–1789
475. Enserink, M. (1998) *Science* **281**, 890–891
476. Dickson, D. (1993) *Nature (London)* **366**, 391
477. Poste, G. (1995) *Nature (London)* **378**, 534–536
478. Fuller, B. P., Ellis Kahn, M. J., Barr, P. A., Biesecker, L., Crowley, E., Garber, J., Mansoura, M. K., Murphy, P., Murray, J., Phillips, J., Rothenberg, K., Rothstein, M., Stopfer, J., Swergold, G., Weber, B., Collins, F. S., and Hudson, K. L. (1999) *Science* **285**, 1359–1361
479. Butler, D. (1998) *Nature (London)* **391**, 727
480. Hudson, K. L., Rothenberg, K. H., Andrews, L. B., Kahn, M. J. E., and Collins, F. S. (1995) *Science* **270**, 391–393
481. Lapham, E. V., Kozma, C., and Weiss, J. O. (1996) *Science* **274**, 621–624
482. Fost, N. (1992) *FASEB J.* **6**, 2813–2817
483. Holtzman, N. A. (1999) *Science* **286**, 409
484. Wills, C. (1970) *Sci. Am.* **222**(Mar), 98–107
485. Vogel, G. (1999) *Science* **286**, 2050–2051
486. Stoye, J. P. (1997) *Nature (London)* **386**, 126–127
487. Weiss, R. A. (1999) *Science* **285**, 1221–1222
488. Paradis, K., Langford, G., Long, Z., Heneine, W., Sandstrom, P., Switzer, W. M., Chapman, L. E., Lockey, C., Onions, D., XEN 111 Study Group, and Otto, E. (1999) *Science* **285**, 1236–1241
489. Enserink, M. (1999) *Science* **286**, 1662–1668
490. Haslberger, A. G. (2000) *Science* **287**, 431–432
491. Jackson, A., and Inglehearn, C. (1999) *Nature (London)* **402**, 457
492. Crawley, M. J. (1999) *Nature (London)* **400**, 501–502
493. Saxena, D., Flores, S., and Stotzky, G. (1999) *Nature (London)* **402**, 480
494. Smith, R. J. (1984) *Science* **224**, 1215–1216
495. Many, B. W. J., Almond, J. W., Berns, K. I., Chanock, R. M., Lvov, D. K., Pettersson, R. F., Schatzmayr, H. G., and Fenner, F. (1993) *Science* **262**, 1223–1224
496. Joklik, W. K., Moss, B., Fields, B. N., Bishop, D. H. L., and Sandakhchiev, L. S. (1993) *Science* **262**, 1225–1226
497. Jesson, L. K., and Barrett, S. C. H. (2002) *Nature (London)* **417**, 707
498. Human Genome Organization Ethics Committee. (2000) *Science* **290**, 49
499. Williamson, R., and Duncan, R. (2002) *Nature (London)* **418**, 585–586
500. Adam, D. (2002) *Nature (London)* **417**, 370
501. Aldhous, P. (2002) *Nature (London)* **418**, 355–356
502. Rothenberg, K. H., and Terry, S. F. (2002) *Science* **297**, 196–197
503. Nowlan, W. (2002) *Science* **297**, 195–196
504. McDowell, N. (2002) *Nature (London)* **416**, 571
505. Stokstad, E. (2002) *Science* **297**, 1797, 1799
506. Kaiser, J. (2001) *Science* **292**, 34–36
507. Enserink, M. (2002) *Science* **297**, 30–31
508. Clarke, T. (2002) *Nature (London)* **419**, 429–430
509. Rieger, M. A., Lamond, M., Preston, C., Powles, S. B., and Roush, R. T. (2002) *Science* **296**, 2386–2388
510. Spurgeon, D. (2001) *Nature (London)* **409**, 749

Study Questions

1. The two chromatids in Fig. 26-13G are said to be coiled with opposite handedness. Can you draw this conclusion from Fig. 26 alone? What are the biological implications for mitosis?[280] Does the DNA have a differing chirality at the molecular level? Compare this observation with the existence of snail shells or flowers[497] with both right and left handedness within the same species.

2. Will the sequencing of the human genome ever be complete?

3. Should patenting of human genes be allowed? Under what circumstances should patents be allowed on genetically engineered genes?

4. Should the human genome be regarded as a **common heritage** such that there is a guarantee that the medical and other benefits arising from genetic research are available to all persons on earth? See Human Genome Organization Ethics Committee.[498]

5. DNA testing (Box 5-D) is widely used by police throughout the world. It has been estimated that if ten loci in the DNA are tested the chance of a random match between two people is one in a billion. In the United Kingdom it is planned to hold DNA profiles on record for one of every 15 people. Is this wise? Or should DNA profiles be recorded for all people?[499] If DNA profiles are on record how can we be sure that they are not used dishonestly? Should police have access to DNA data bases? See Adam.[500]

6. Should "race" be used as a variable in biomedical studies? See Aldhous.[501]

7. Should insurance companies be allowed to have access to genetic information about insured people? Companies usually obtain other medical information. See Rothenberg and Terry,[502] Adam,[500] and Nowlan.[503]

8. Should genetically engineered fish be allowed in "farms" that are set up in ocean waters? See McDowell[504] and Stokstad.[505]

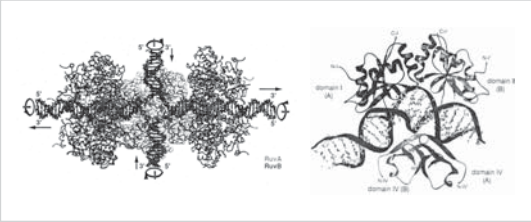
9. Do transgenic trees pose a threat to natural ecosystems? See Kaiser.[506]

10. Should we attempt to replace wild populations of mosquitoes with genetically engineered mosquitoes that can not transmit malaria? See Enserick[507] and Clarke.[508]

11. Could dispersal of pollen from genetically modified plants lead to undesirable "genetic pollution" of the environment? See Rieger *et al*.[509]

12. To what extent is genetic modification of plants and animals equivalent to changes made by conventional breeding? See Spurgeon.[510]

13. Choose one of the ethical questions that can be raised about application of our new knowledge about the genome (e.g., see pp. 1518, 1519, Chapter 32, and study questions on this page). Study literature available to you and prepare a recommendation to the public, Congress, or to local regulatory agencies. Follow a scientific approach. Try to find true facts that can be verified. Consider all viewpoints. State some of the uncertainties in your recommendation. Present your proposal to a class or to a friend for criticism. Then publish your view in a newspaper if you wish to.



Many proteins interact with DNA. These include polymerases that replicate DNA, helicases that unwind double helices, topoisomerases that cut and reseal DNA strands to avoid entangling, and enzymes that repair damage to DNA. At left is a motor complex consisting of the tetrameric binding protein RuvA (light shading) and the hexameric helicase RuvB (two copies in darker shading). See also Fig. 27-26B. From Putnam *et al.* (2001) *J. Mol. Biol.*, **311**, 297–310. Right bacterial protein (from *Thermus aquaticus*) MutS recognizes mispaired bases in DNA and initiates their removal. The DNA is bound in a bent conformation. In this duplex one strand contains an unpaired thymine (at top of bend), which would cause a mutation if not removed. From Obmolova *et al.* (2000) *Nature*, **407**, 703–7.

Contents

1529 **A. The Topology and Environment of DNA**
1529 1. DNA in Viruses
1530 2. Bacterial Chromosomes and Plasmids
1531 3. Protamines, Histones, and Nucleosomes
1531 *Nucleosomes*
1533 *The linker histones*
1534 *Folding of nucleosome chains; chromosomes*
1535 4. The Cell Nucleus
1535 *The nuclear matrix*
1535 *Other nonhistone nuclear proteins*
1536 *Nuclear membrane, pores, and lamina*
1537 *The nucleolus*
1537 **B. Organization of DNA**
1537 1. Repetitive DNA
1537 *Centromeres*
1538 *Telomeres*
1538 *Short interspersed sequences (SINES)*
1539 *Long interspersed sequences (LINES)*
1539 2. Genes for Ribosomal RNA and Small RNA Molecules
1539 3. Other Gene Clusters and Pseudogenes
1540 4. Introns, Exons, and Overlapping Genes
1540 5. DNA of Organelles
1541 6. Methylation of DNA
1541 *CG doublets*
1542 *Imprinting*
1542 **C. Replication**
1542 1. Early Studies
1543 *Autoradiography*
1543 *The chemistry of DNA polymerization*
1544 *Discontinuous replication and RNA primers*
1544 2. DNA Polymerases
1544 *Exonuclease activities, proofreading, and editing*
1547 *Other Class A polymerases*
1547 *Polymerases of Class B*
1548 *Other DNA polymerases*
1548 *DNA polymerase III*
1549 3. Other Replication Proteins
1549 *DNA ligases*
1549 *Single-strand binding proteins*
1550 *Helicases and topoisomerases*
1552 *Primases, initiator proteins, and ribonucleases*
1553 4. Replication of Bacterial DNA
1554 *Directions of replication*
1554 *Origins of replication*
1555 *Priming and initiation of DNA synthesis*
1556 *Elongation of DNA chains*
1557 *Termination of replication*
1557 5. The Replication of Viral DNA
1559 6. Packaging of Viral Genomes
1559 7. Plasmids
1559 8. Chromosome Ends
1561 9. Mitochondrial and Chloroplast DNA
1561 10. Replication of Eukaryotic Nuclear and Viral DNA
1562 *Eukaryotic viruses*
1562 *Artificial chromosomes*
1562 *Replication of nuclear DNA*
1564 **D. Integration, Excision, and Recombination of DNA**
1564 1. Recombination Mechanisms
1565 *The Holliday recombination intermediate*
1566 *RecA and other strand-exchange proteins*
1567 *Processing the Holliday junction*
1568 2. Nonreciprocal Recombination and Unequal Crossing-Over
1570 3. Site-Specific Recombination and the Integration and Excision of DNA
1570 *The integrase (tyrosine recombinase) family*
1572 *The resolvase/invertase family and invertible DNA sequences*
1573 4. Transposons and Insertion Sequences
1575 *Cut-and-paste (nonreplicative) transposons*
1575 *Replicative transposons*
1576 *The temperate bacteriophage Mu*
1576 *Some other transposons*
1577 5. Other Causes of Genetic Recombination
1578 **E. Damage and Repair of DNA**
1578 1. Causes of Mutations
1579 2. Fidelity of Replication
1579 *The role of polymerases*
1580 *Editing (proofreading)*
1580 3. Repair of Damaged DNA
1580 *Methyl-directed and other mismatch repair*
1581 *Excision repair*
1581 *Base excision repair (BER)*
1583 *Reactions of alkylated bases*
1583 *Repair of double-strand breaks*
1583 *The SOS response and translesion repair*
1584 *Poly (ADP-ribose)*
1584 **F. Mutagens in the Environment**

1590 **References**
1601 **Study Questions**

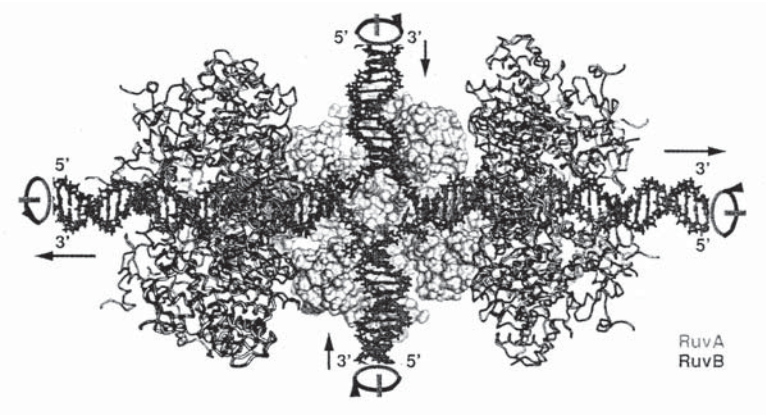
Boxes
1585 Box 27-A Deficiencies in Human DNA Repair

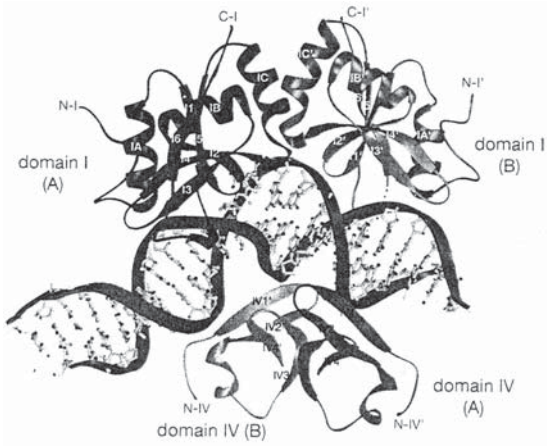
Tables
1545 Table 27-1 Families of DNA Polymerases

1551 Table 27-2 Some Proteins of DNA Replication in *E. coli*

Organization, Replication, Transposition, and Repair of DNA

27





Both the replication and transcription of DNA are complex processes. Although the basic chemistry is relatively simple many enzymes and other proteins are required. In part this reflects organizational and topological problems[1] associated with the huge amount of DNA present as a single molecule within a chromosome.

A. The Topology and Environment of DNA

Although we can isolate DNA in the form of simple double helical fragments, the topology of natural DNA is always more complex. Covalently closed circular DNA such as that in plasmids, mitochondria, and bacterial chromosomes is supercoiled (Chapter 5) and bound to proteins. The DNA of chromosomes and bacteriophage particles is folded further into more compact forms. For example, the chromosome of *E. coli* contains DNA about 1.5 mm in length folded within a cell that is only 2 µm long. The diploid length of DNA in a 20 µm cell of a human is about 1.5 meters. At the time of cell division human DNA must all be replicated and packaged into chromosomes, 23 pairs in each cell. The density of the compacted DNA varies. A bacterial nucleoid may contain 10–30 mg / ml of DNA.[2] In chromatin of a eukaryotic nucleus there may be 200 mg / ml of DNA and in nucleosomes 330–400 mg / ml.[2a] The tightly compacted head of the T4 bacteriophage (Box 7-C) contains 520 mg / ml.[3]

1. DNA in Viruses

In the simplest filamentous DNA viruses such as M13 the DNA is coated by a helical protein sheath (Fig. 7-7), as it is extruded from a cell. The sheath is peeled off as the virus enters another cell. However, in the large tailed phage (Box 7-C), which contains ~160 kb of polynucleotide chains, the DNA is closely packed within the heads. In a model that seems to accommodate most experimental results, the DNA rod bends sharply into a series of folds, which are laid down around the long axis of the head in spirally arranged shells (Fig. 27-1).[4,4a] The end of the DNA that enters the phage head first appears to be located in the center with successive shells of DNA around it.[4] In the large bacteriophage G the DNA appears to be folded to form 12 icosahedrally arranged pear-shaped rings in the corners of the capsid.[5] The 2.0 nm diameter double helical segments of DNA lie roughly parallel and are separated by only 0.5–1.0 nm of solvent,[6] which contains cations such as the polycation of spermidine. Another possibility is that the DNA may be wound as on a spool of thread.[7] The DNA chains have an external diameter of ~2 nm with 0.6 nm additional for a hydration layer. In a phage head the adjacent parallel chains are 2.6–2.7 nm apart. Thus, the packing is very tight. Even so, capsids tend to be only about half filled with DNA. An exception is provided by the tobacco mosaic virus (Fig. 7-8) in which the RNA genome is held precisely by protein subunits, which dissociate to release the RNA during infection.[3]

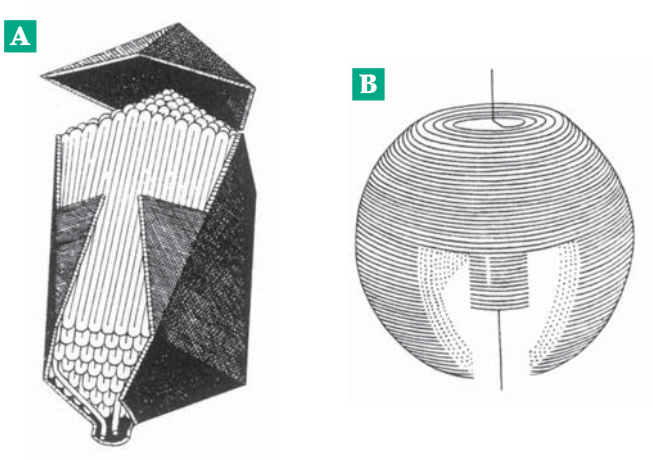


Figure 27-1 Possible ways of packing DNA into the heads of bacteriophage particles. (A) Spiral-fold. (B) Concentric shell model From Black *et al.*[4]

2. Bacterial Chromosomes and Plasmids

Most DNA in living organisms, whether bacteria or eukaryotes, is underwound. That is, the superhelix density (Chapter 5) is about −0.05 or one supercoil per 200 base pairs. In eukaryotes this negative supercoiling can be accounted for by the winding of DNA around the histones within the nucleosomes (Figs. 5-21, 27-4). The situation in bacteria is not as clear. There are many bacterial DNA-binding proteins,[8,9] but one of them known as HU is particularly abundant.[8–12] In *E. coli* it exists as mixed dimers and tetramers of 9.5 (α) and 9.2 (β) kDa subunits. Each HU tetramer can bind ~60 bp of DNA. There are about 60,000 HU monomers per cell, enough to coat ~20% of the genome. Possible modes of interaction with DNA have been proposed on the basis of the X-ray structure of HU.[8,13–14a] Binding to HU causes the DNA to be more tightly wound, introducing additional negative supercoils. The resulting unreleased torsional stress may be important in the functioning of the DNA.[15–17a] Binding is strongest to four-way junctions,[17a] and to DNA with nicks and gaps or to structures induced by supercoiling.[17b] Other basic histonelike proteins may also bind to the DNA. However, there are no structures that resemble eukaryotic nucleosomes.

If bacterial cells are lysed under certain conditions, e.g., in 1 M NaCl or in the presence of a "physiological" 5 mM spermidine, the entire bacterial chromosome can be isolated.[10] The DNA in these isolated chromosomes retains some torsional tension that, however, can be relaxed by nicking with nucleases or by γ-irradiation. However, a single nick relaxes the DNA very little. The explanation appears to be that the DNA is held by proteins of the nucleoid matrix in a series of loops (Fig. 27-2). A single nick relaxes just one loop. On this basis there are 43 ± 10 loops per genome with ~100 kb of DNA per loop.[18] A 136-residue protein **H-NS** is involved in condensation of bacterial DNA.[18a] It may act as a scaffolding protein, but it also functions in controlling transcription.[18b]

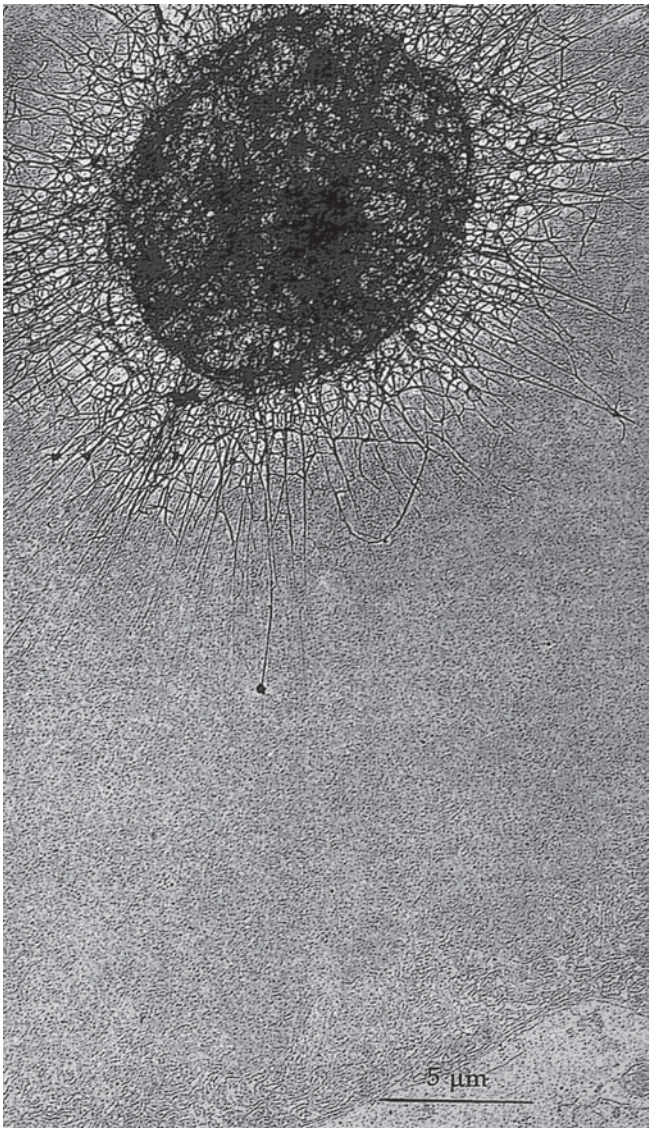


Figure 27-2 Electron micrograph of a bacterial nucleoid. The DNA is usually contained within the 'cage', but has been spread, using Kleinschmidt's procedure, to yield a surrounding skirt. The cage contains a protein network which includes elements of the cytoskeleton, enclosing the residual nuclear substructures. The denser fibrils radiating from this cage disappear to nuclease digestion and are probably aggregates of DNA fibers which merge with individual DNA strands at the extremities of the skirt. These strands are highly supercoiled, indicative of intact DNA. From Jackson and Patel, provided by Dr. S. J. McCready.[19]

3. Protamines, Histones, and Nucleosomes

Within bacterial cells the negatively charged phosphate groups of the DNA are neutralized to a large extent by the positively charged polyamines,[19a] by cations such as K^+ and Mg^{2+}, and by basic proteins such as HU. Within the mature heads of sperm cells of fish, the tightly packaged DNA is neutralized by the **protamines**, small ~5-kDa proteins rich in arginine.[4a,20–21a] Similar basic proteins are found in mammalian sperm.[22,23] However, within most eukaryotic cells, the charges on DNA are balanced principally by a group of basic proteins, first isolated and named **histone** by Kossel in 1884.[24]

There are five classes of histones, which range in molecular mass from ~11 to 21.5 kDa:[25,26]

H1 (including $H1^0$, H5; lysine-rich "linker" histone)
H2A, H2B (moderately lysine-rich) } "core" histones
H3, H4 (arginine-rich)

All of the core histones share a conserved 65-residue **histone fold.**[27,28] The arginine-rich histones have a strongly conserved amino acid sequence, histone H4 from pea seedlings differing from that of the bovine thymus by only two amino acids. On the other hand, the lysine-rich H1 is almost species-specific in its sequence. Differentiated tissues contain at least seven variant forms of histone H1 including proteins designated $H1^0$, H1t, and H5.[29–31]

The N-terminal 25–40 amino acids of the core histones are positively charged and highly conserved.[32] The 135-residue histone H3 of calf thymus carries a net charge of +18 within the first 53 residues. This is probably the portion that binds to DNA. On the other hand, the carboxyl terminal end is hydrophobic and only slightly basic.[33] Histones undergo substantial amounts of **micromodification** including phosphorylation, acetylation, and methylation.[33a,b] Mono-, di-, and tri-methyllysine residues may be present.[33c,d] The core histones all undergo acetylation on specific lysyl side chains. Nuclear histone acetyltransferases[34–38b] transfer acetyl groups from acetylCoA and hydrolytic deacetylases may remove them.[39–40b] The amount of acetylation varies during different stages of the cell cycle, suggesting a regulatory role.[41] Acetylation sites in H3 and H4 are highly conserved in all eukaryotes.[42]

A small fraction of histone H2A undergoes phosphorylation and dephosphorylation continuously,[43] but H1 and H3 are phosphorylated and dephosphorylated at specific stages of the cell cycle. Phosphorylation of H1 has been thought essential for "condensation" of chromatin,[44,45] the folding into the tightly packed chromosome structures. However, more recent experiments point to the N terminus of histone H2B as the required site of phosphorylation for chromosome condensation.[45a] In addition, histone H1 may interact with membrane lipids.[46] Histone H2A exhibits the greatest heterogeneity and appears to function in regulation of transcription, of gene silencing, and of repair of double-strand breaks in DNA.[46a] Each of the histones appears to be regulated separately. In animal chromatin ~10% of histone H2A and small fractions of H2B and of tissue-specific histones are covalently linked to ubiquitin (Box 10-C). However, this monoubiquitination may not be related to proteolytic degradation.[47] Archea contain histones that dimerize and bind DNA to form nucleosomes.[47a,b]

Nucleosomes. An early idea of the function of histones was that they serve as gene repressors. To some extent this view is still valid. However, the large quantity of histone and uniform distribution over the DNA suggested some other role. This was clarified when electron micrographs showed that chromatin fibers form **nucleosomes**,[48–50] regular repeating structures resembling beads on a string. The same structure is seen in the "minichromosomes" formed from virus SV40 (Fig. 27-3).[51–53] Two molecules each of histones H2A, H2B, H3, and H4 form the core of the nucleosome around which ~146 bp of dsDNA is coiled into approximately two negative, left-handed toroidal superhelical turns (Figs. 5-21, 27-4).

Digestion of chromatin by nucleases causes rapid cleavage into ~200-bp fragments and slower cleavage to 146 ± 20-bp fragments. This suggested that ~200-bp segments of DNA are folded around a histone octomer, contracting the 68 nm extended length of relaxed B-DNA into a 10-nm nucleosome. A short linker region of variable length, up to 80 bp, lies between the nucleosomes.[54] The fifth histone, H1 (or H5 in some species), may bind to this linker DNA. A nucleosome with bound H1 is sometimes called a **chromatosome**.[55]

The superhelix density of ~0.05 observed for DNA extracted from eukaryotic cells is just equal to one negative superhelix turn per nucleosome. For example, the number of nucleosomes seen in the minichromosome of Fig. 27-3 matches the numbers of supercoils in the SV40 DNA (Fig. 5-20). If there are two negative supercoils per nucleosome, as shown in Fig. 27-4, the DNA in the nucleosome must be wound more tightly than in relaxed DNA (10.0 bp per turn instead of the 10.6 of relaxed DNA).[56] NMR data suggest that within the nucleosome the regular base pairing in the DNA may be partially disrupted and that some parts of the histone have a high degree of mobility.[57]

Although nucleosomes are distributed rather evenly along the DNA of a cell, there are some DNA sequences that favor nucleosome formation. The resulting **positioned nucleosomes** are often found in the vicinity of gene promoters, enhancers, and other

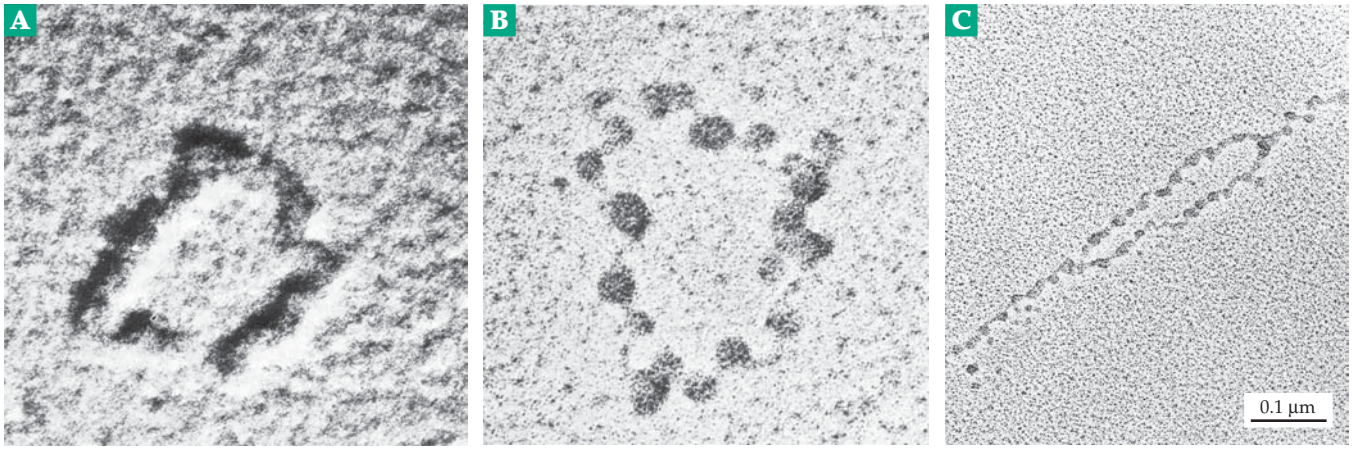


Figure 27-3 (A) Electron micrograph of "minichromosome" formed from virus SV40 growing in monkey cells in culture.[51] In this native form the nucleoprotein fiber is ~11 nm in diameter and ~210 nm in length. (B) Beaded form of minichromosome observed when the ionic strength was lowered. The 20 beads have diameters of ~11 nm and are joined by bridges roughly 2 nm in diameter and 13 nm long. Deproteinization and relaxation of the DNA revealed that the overall length of the DNA present is seven times the length of the native minichromosome. (C) Electron micrograph of chromatid of a blastoderm-stage embryo of *Drosophila melanogaster* in the process of replication. Nucleosomal particles are visible immediately adjacent to the replication forks. Courtesy of Steven L. McKnight and Oscar L. Miller, Jr.

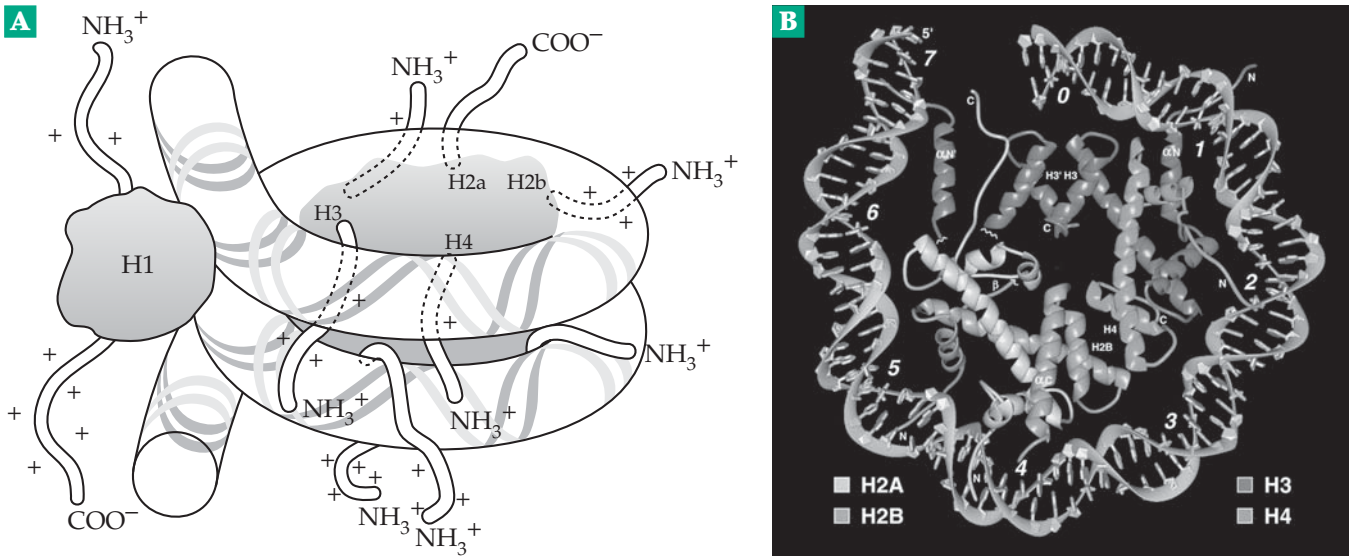


Figure 27-4 (A) A nucleosome is formed when dsDNA, shown schematically as a tube, wraps roughly twice around a histone octomer. This complex, or nucleosomal core particle, includes at its center two copies of histone H3, two of histone H4, and two H2A–H2B dimeric pairs, one of which is not visible. Ends of each histone molecule are thought to protrude like tails from the core, ready to interact with other molecules. In many organisms, histone H1, portrayed at left in one possible position, helps to anchor DNA to the core and promotes further compaction of the DNA into a 30-nanometer fiber. See Grunstein.[49] (B) Nucleosome core particle: 73-bp half. The view is down the superhelix axis with the pseudo dyad axis aligned vertically. The central base pair through which the dyad passes is above the superhelix axis location labeled 0. Each additional numerical label 1–7 represents one further DNA double helix turn. The complete histone proteins (except for the tail regions) that are primarily associated with the 73-bp superhelix half are shown. The two copies of each histone pair are distinguished as unprimed and primed, where the histone fold of the unprimed copy is primarily associated with the 73-bp DNA half and the primed copy with the 72-bp half. Four-helix bundles are labeled as H3', H3, and H2B, H4; histone-fold extensions of H3 and H2B are labeled as αN′, αN, and αC, respectively; the interface between the H2A docking domain and the H4 C terminus as β; and N- and C-terminal tail regions as N or C. From Luger *et al.*[50]

control sequences.[58–61a] It seems likely that positioning of nucleosomes is related to control of transcription or of other activities. The same basic structure for chromatin has been found in animals, fungi, and green plants.

The linker histones. A nucleosome is pictured in Fig. 27-4A as if held in a compact configuration by the binding of histone H1 at a position that marks a pseudo twofold axis that lies in the plane of the nucleosome. However, this is only one of several possible locations for linker histones of the H1, H5 family.[62–67] The structure of the linker histones is somewhat different from that of core histones. They have an 80-residue globular domain with long N-terminal *and* C-terminal chains, both of which are rich in basic residues and evidently available for binding to DNA.[68] Perhaps they bind both to the DNA entering the nucleosome and to that leaving the nucleosome, reducing electrostatic repulsion of those two parts of the DNA superhelix.[69]

Another possible location for histone H1 or H5 is above the histone surface as shown in Fig. 27-4 and *inside* the DNA loop.[65] A third suggested location for the globular linker core is *between* the two turns of the DNA strand.[66] While one function of the linker histones may be to stabilize mononucleosomes, they may also play a role in compaction of the DNA into the 30-nm fibers universally seen in nuclei of cells.[62,64–64b]

Histone H1 can also be regarded as a general repressor, holding chromatin tightly folded and preventing transcription. The possible roles of acetylation, phosphorylation, methylation, ubiquitination, and other modifications of histones in controlling transcription, replication, and DNA repair are receiving increasing attention.[70–73] **Active chromatin**, where transcription is occurring, has an altered nucleosome structure and increased susceptibility to nuclease action. It appears to be less tightly packed than inactive chromatin and to contain regions called **hypersensitive sites** that are accessible to nucleases or chemical modification reagents.[74–76] There seems to be a direct link between increased acetylation of histones and enhanced initiation of transcription by RNA polymerase II.[35,38,77–82] Conversely, deacetylation is associated with repression of transcription. Both histone acetylase and deacetylase activities have been found in transcriptional regulators.[80–82]

The observation that H1 becomes phosphorylated during the initiation step of mitosis suggests another control mechanism for its repressor functions.[83] Several multiprotein complexes that "remodel" chromatin have been identified.[73,84] These complexes contain

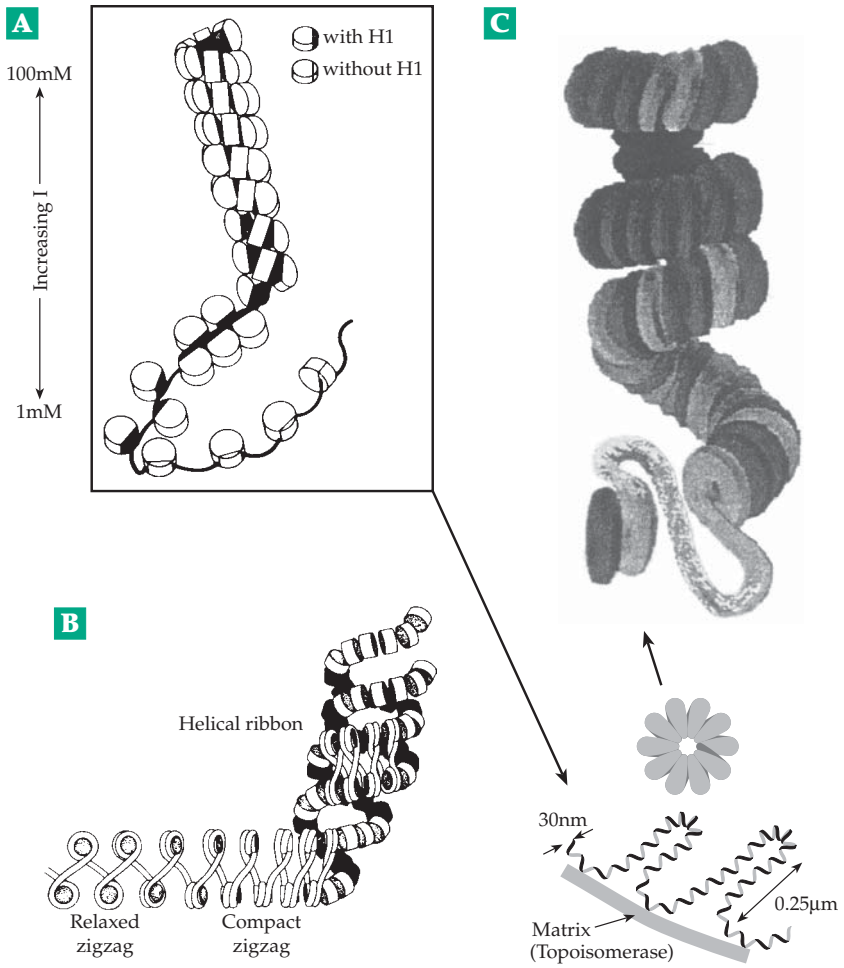


Figure 27-5 (A, B) Two possible models of the 30-nm chromatin fiber.[55] (A) Thoma *et al.*[85] (B) Woodcock *et al.*[64,87] The fully compacted structure is seen at the top of each figure. The bottom parts of the figures illustrate proposed intermediate steps in the ionic strength-induced compaction. (C) Possible organization of the DNA within a metaphase chromosome. Six nucleosomes form each turn of a solenoid in the 30-nm filament as in (A). The 30-nm filament forms ~30 kb-loop domains of DNA and some of these attach at the base to the nuclear matrix that contains topoisomerase II. About ten of the loops form a helical radial array of 250-nm diameter around the core of the chromosome. Further winding of this helix into a tight coil ~700 nm in diameter, as at the top in (C), forms a metaphase chromatid. From Manuelidis[91].

ATP-dependent **helicase** activities that open up DNA for transcription (see Chapter 28, Section B).

Folding of nucleosome chains; chromosomes. Electron microscopy shows that chromatin is packed into the nucleus largely as fibers of ~30- to 36-nm diameter.[55,84a] Thoma *et al.*[85] proposed that the fibers could be formed by winding the string of nucleosomes into a simple one-start helix with six nucleosomes per turn and a pitch of ~11 nm (Fig. 27-5A). The H1 molecules would be close together in the center. An alternative model, one of several suggested,[55,86] is shown in Fig. 27-5B. Its zigzag pattern of adjacent nucleosomes would generate a two-start helix.[64,87,88] Another alternative model envisions larger solenoids with interdigitated nucleosomes.[89]

A somewhat similar structure appears to be present in the specialized eukaryotic **lampbrush chromosomes** (Fig. 27-6), which are observed during the meiotic prophase of oocytes. They have been studied intensively in amphibians such as *Xenopus*. A lampbrush chromosome is actually a homologous pair of chromosomes, each one in turn consisting of two closely associated chromatids. The chromosomes are highly expanded, and about 5% of the DNA is extended in the form of ~4000 perfectly paired loops visible with an electron microscope. Each loop consists of ~50 µm or ~150 kb of extended DNA. No evidence of any breaks in the DNA is seen, a fact that supports the belief that a single DNA molecule extends from one end of the chromosome to the other through all of the loops.

Like the puffs of polytene chromosomes (Chapter 28), which may have a similar structure, lampbrush chromosomes appear to be actively engaged in transcription. Approximately 3% of the DNA may be functional in producing mRNA that is accumulated within the oocyte and is used as a template for protein synthesis during early embryonic development.[90]

A different arrangement is present in metaphase chromosomes, which appear as two dense parallel sister chromatids of ~700-nm diameter.[91] The DNA must be highly folded. In the model shown in Fig. 27-5C the 30-nm fiber is folded into ~30-kb loops, each one formed from ~25 turns of the 30-nm helix. The loops then form a helical array 250 nm in diameter with ~ ten loops (300 kb) per turn. This helix is further wound into a tight helix of ~700-nm diameter. A single turn of this helix may contain as much as 9 Mb of DNA.[91] About four hundred coils (an average of 18 coils per human haploid chromosome) could accommodate the entire genome. A group of five proteins, some of which are designated **SMC** (structural maintenance of chromosomes) proteins, form a complex called **cohesin**. SMC proteins, large multidomain proteins found in all eukaryotes, are also present in bacteria.[91a,b]

Interphase chromatin must be much less tightly

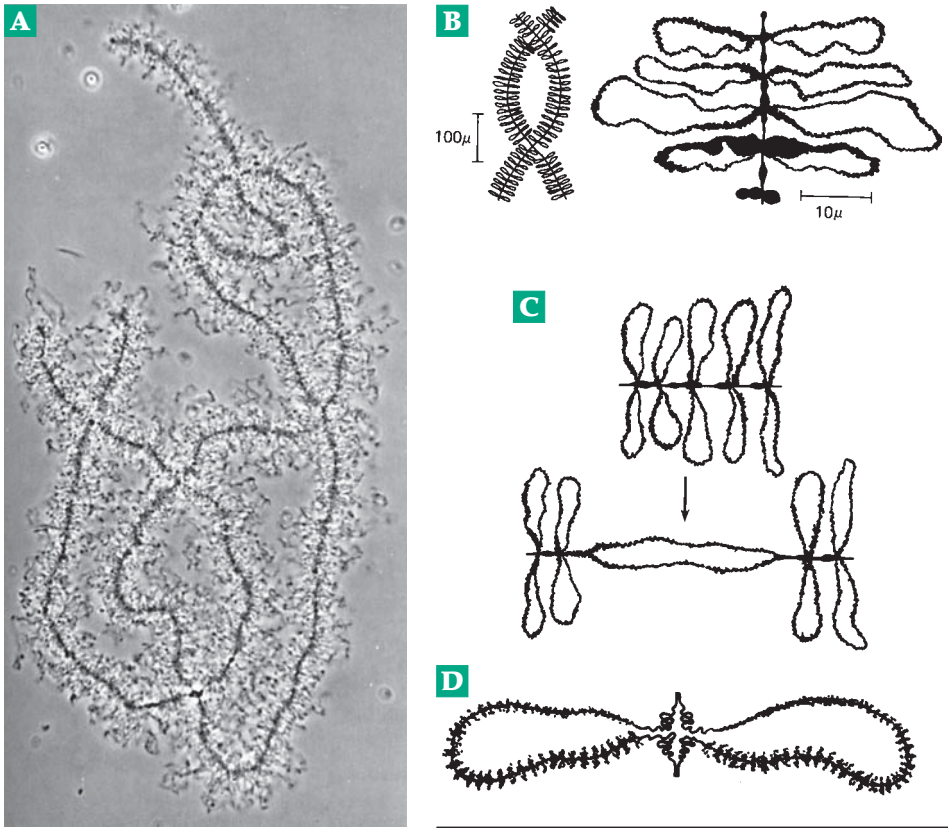


Figure 27-6 (A) Photomicrograph of a lampbrush chromosome from the nucleus of an oocyte of the newt *Triturus*. From L. M. Mays, *Genetics, A Molecular Approach*, Macmillan, New York, 1981, p. 227. (B–D) Diagrammatic views of lampbrush chromosomes. (B) The two homologous chromosomes (left) are held together by two chiasmata. A portion of the central chromosome axis (right) shows that two loops with identical morphology emerge at a given point, evidence that each chromosome has already split into two chromatids. (C) Accidental stretching of a chromosome reveals the continuity of the loop axis with the central axis. (D) A single loop pair, showing the single DNA molecules on which RNA chains (indicated by fuzzy shading) are being transcribed. From J. Gall, *Brookhaven Symp. Biol.* **8**, 17 (1955).

packed and contains regions in which large loops, e.g., of 20–120-kb size, are uncoiled enough to allow transcription factors and other proteins to locate their target sequences.[91c] Many models of interphase chromatin have been proposed.[92,93]

4. The Cell Nucleus

It has been clear for many years that the nucleus has a well-organized structure. However, techniques such as fluorescent labeling have only recently been used to provide important details. These studies show that individual interphase chromosomes occupy discrete territories within the nucleus.[94–96] Parts of the chromosomes may be unfolded and active in transcription, while others are more tightly coiled. Some parts, known as **heterochromatin**,[97–98a] are very tightly coiled and metabolically almost inert. These regions include the highly repetitive DNA of telomeres and centromeres as well as other regions and complete inactivated female X chromosomes. Replication, transcription, and RNA splicing complexes are found at distinct locations within cells.[99–102a] They are apparently fixed, perhaps attached to the inner nuclear membrane, while the DNA passes through the complexes.

The nuclear matrix. The lipid bilayers, the histones and other soluble proteins, and the DNA can all be removed from nuclei by extraction and enzymatic digestion. An insoluble residue, the **nuclear matrix,** is left.[103–107] Largely protein in nature, this matrix is spread throughout the nucleus. Remnants of the membranes remain in the form of proteins that were in or along the bilayer. The nucleolus is clearly defined. The DNA appears to be bound to the nuclear matrix proteins. A specific 320-kb piece of a *Drosophila* chromosome has been mapped and used to locate nontranscribed scaffold (or matrix) attachment regions of DNA bound to matrix proteins. These were found at intervals of 26–112 kb, the intervening loops containing up to five or more genes.[105,106,108,109] A 120-kDa protein together with topoisomerase II (Section C,2)[102,104] may be components of a **nuclear scaffold** that constrains the loops of DNA. The scaffold may also provide locations for the complexes of proteins involved in replication and other processes.[103] The **matrix attachment regions** (MARs) may also act as **insulators** that shield promoters for transcription within certain loops from control elements such as **enhancers** that may be present in adjacent loops.[91b,110] At least one nuclear matrix component becomes phosphorylated and moves to the nuclear poles during mitosis.[111]

Other nonhistone nuclear proteins. Polyacrylamide gel electrophoresis revealed more than 450 components in HeLa cell nuclei. Most are present in small amounts of <10,000 molecules per cell and are not detectable in cytoplasm.[112] Among the more acidic proteins are many enzymes including RNA polymerases. There are also gene repressors, hormone-binding proteins, protein kinases, and topoisomerases.[113] Among the six most abundant nonhistone nuclear proteins in the rat are the cytoskeletal proteins myosin, actin, tubulin, and tropomyosin.[114]

A group of small (<30 kDa) proteins, the **high mobility group (HMG)** proteins,[112,115] can be extracted from chromatin with 0.35 M NaCl. Two pairs, HMG-1 + HMG-2 R (renamed **HMGB**)[115a,b] and HMG-14 + HMG-17 (renamed **HMGN**),[115b,c] are present in nuclei of all mammals and birds. HMG-14

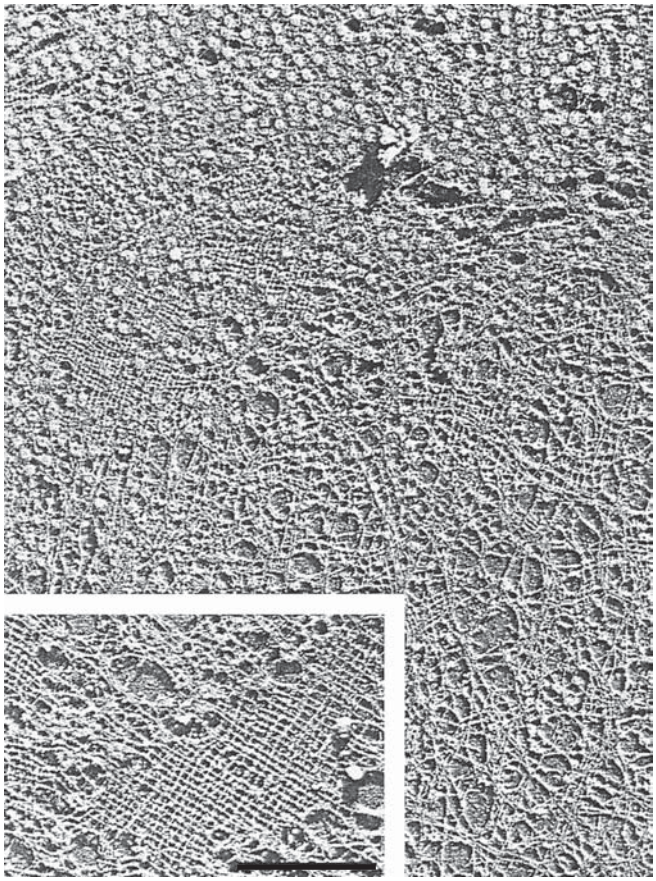


Figure 27-7 Native nuclear lamina of *Xenopus* oocytes. Freeze-dried metal-shadowed nuclear envelope extracted with Triton X-100, revealing the nuclear lamina meshwork partially covered with arrays of nuclear pore complexes. Inset, relatively well-preserved area of the meshwork of nearly orthogonal filaments from which pore complexes have been mechanically removed. Bar, 1 µm. From Aebi *et al.*[121]

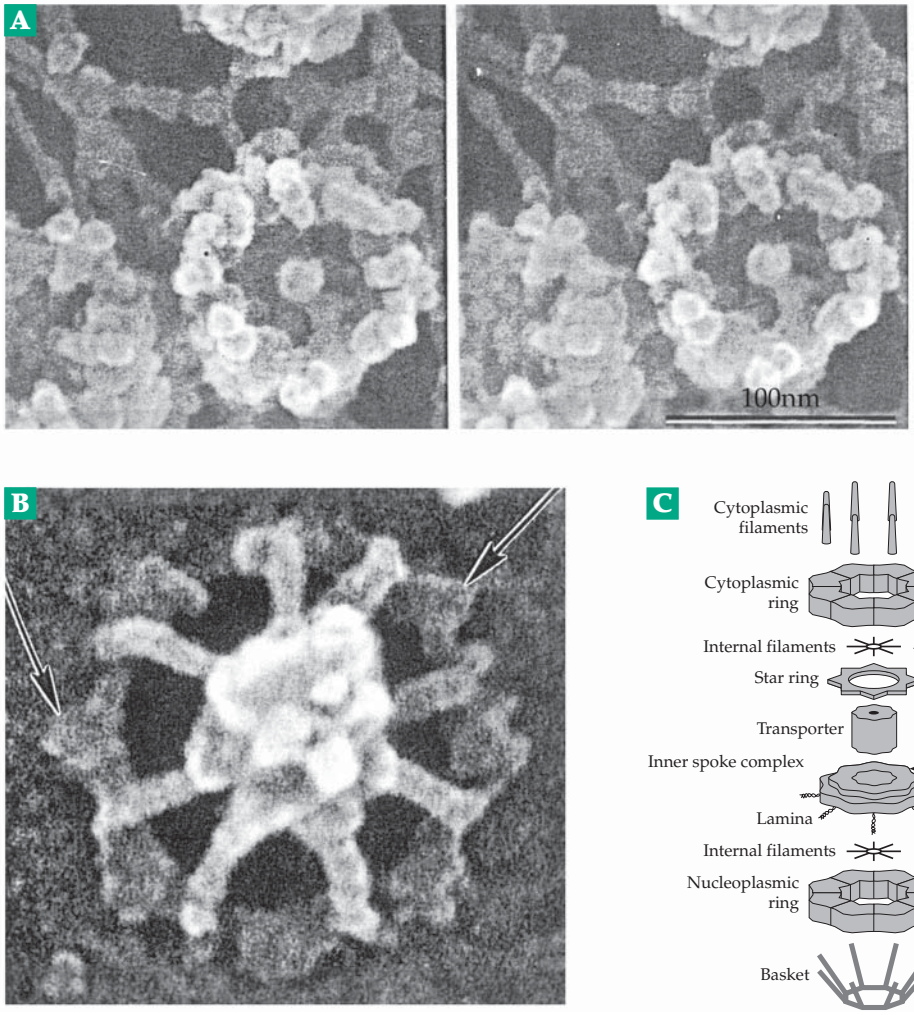


Figure 27-8 Electron micrographs of nuclear pores from a *Xenopus* oocyte. (A) Cytoplasmic face of a detergent-extracted membrane. Stereo view showing structure of a single pore. Lamin fibers are seen in the background. (B) Nucleoplasmic face of a trypsin-treated sample. The trypsin-resistant basket is clearly visible. Arrows point to the triangular pieces of the nucleoplasmic ring. (C) Diagram illustrating various structural features of the pores and the way in which they may fit together. From Goldberg and Allen.[120] Courtesy of Martin W. Goldberg.

and HMG-17 appear to be concentrated where active transcription is occurring. HMG-17 undergoes a variety of modifications including acetylation, methylation, phosphorylation, ubiquitination, glycosylation, and ADP-ribosylation, suggesting that it may assist in regulation of transcription. Other members of the HMG family bind more specifically to certain DNA sequences or structures.[115d,e]

Nuclear membrane, pores, and lamina. The nuclear membrane consists of two bilayer membranes separated by a 40- to 60-nm perinuclear space.[116,117] Both of the membranes and the intervening space are penetrated by large proteinaceous **nuclear pores**.[118–120b] The two membranes are fused together around the pores. On the inside of the innermost nuclear membrane is a relatively insoluble meshwork of intermediate filaments, the **nuclear lamina** (Fig. 27-7). It acts as a scaffold for the pores and may also interact with chromatin.[113,121–123a] Three proteins, the **lamins** A, B, and C, are the major components of the nuclear lamina. The A and C lamins, which differ only at their C termini, are homologous to cytoplasmic intermediate filament proteins and may exist as a network of coiled-coil polymers.[124] Lamins are phosphorylated, and the lamina is disassembled during the prophase of mitosis, apparently under the influence of the cdc2–cyclin B complex (Eq. 26-3),[125] and reappears in telophase at the end of mitosis.

Nuclear pores consist of octameric rings of protein subunits with a complex structure and an outside diameter of 120 nm and an inside diameter of ~80 nm (Fig. 27-8). A pore may consist of as many as 80 to 100 different proteins and have a mass of ~125 MDa.[119,126,126a,b] The pores are sometimes seen to be blocked by ~35 nm granules, perhaps pre-ribosomes. Transport through nuclear pores occurs in both directions. Numerous proteins enter the nucleus where many bind to RNA and are then exported as ribosomal subunits. Transfer RNAs and messenger RNAs must also be exported. A family of nuclear transporters known as **importins** and **exportins** mediate the movement of protein and RNA through the pores.[119,126b,127–130] They depend upon the G protein **Ran**, which also functions in spindle formation (p. 1503), and the hydrolysis of GTP.[131–132b] Large

conformational changes may be linked to the transport cycle[133,134] and also have been observed in response to changes in the Ca^{2+} level.[135,136]

The nucleolus. This organelle is a dynamic structure, which breaks down and reforms in each mitotic cycle. It is organized around clusters of genes for the 28S, 18S, and 5.8S ribosomal RNA subunits. Both chromatin and ribonucleoproteins that transport the rRNA out to the cytoplasm are present.[95,137] The major nucleolar protein **nucleolin** is characterized by its glycine-rich C terminus, which contains seven repeats of GGRGG and also contains N^{γ},N^{γ}-dimethylarginine.[138,139] Nucleolin may control transcription of the DNA that carries the rRNA genes. It appears to be required for ribosome synthesis (Chapter 29) and for attachment to the nuclear matrix. A variety of other proteins in smaller amounts are also needed in nucleoli. In addition to the true nucleoli other regions described as **speckles** and **coiled bodies** may develop around RNA splicing centers.[95,139a]

B. Organization of DNA

A vast amount of new information on DNA sequences and structures is available for yeast, *Arabidopsis*,[140] *Drosophila*, the nematode *Caenorhabditis elegans*,[141] human beings, and other species (see Chapter 26). Nevertheless, it may be worthwhile to consider older discoveries, some dating back 20 years or more.

When DNA is cut into ~10-kb fragments by shearing and the fragments are denatured by heat, the renaturation of the resulting single-stranded fragments upon cooling takes place in two or more steps. Some material reforms double helices rapidly, whereas other material is slow to renature (Fig. 5-46). At least four kinetically distinct fractions have been recognized: (1) About 70% of mammalian DNA appears to exist largely as single copies, i.e., with unique sequences. (2) About 20% of the total DNA is moderately repetitious, containing sequences that may be present 10^3–10^5 times. (3) About 5% reassociates very rapidly and is identified as highly repetitive **satellite DNA** of which there may be ~10^6 copies. (4) A smaller fraction contains long palindromic sequences. As a consequence, the single-stranded pieces can fold back almost instantaneously to form hairpin structures.[142,143] These differences also show up in density gradient centrifugation in CsCl, whether on large pieces of DNA[144] in the presence of ligands such as Ag^+ or on smaller restriction fragments which have often been studied by polyacrylamide gel electrophoresis.

If there are ~35,000 human genes, protein or RNA coding sequences must occupy only 2–3% of the genome.[145] The pufferfish *Fuga rubripes* probably has almost as many genes as we but only ~13% as much DNA. In further contrast the newt *Triturus cristatus* has six times as much DNA as a human.[146] The compact genome of the green plant *Arabidopsis thaliana* occupies only 120 Mb. In contrast are the 415-, 2500-, and 5300-Mb genomes of rice, maize, and barley, respectively.[147] What is the function of all of the apparently noncoding DNA present in some organisms? It is often viewed as "junk," whose only function may be to facilitate evolutionary changes in the genome. However, there is doubtless important undiscovered information in these regions.[145] It has been very hard to determine sequences, in part because of the large amount of highly repetitive DNA. Some regions have been "unclonable" in prokaryotic systems because of the presence of transposon-like sequences or "kinkable" elements (TG•CA steps), palindromes, etc.[145a,b]

1. Repetitive DNA

Rapidly renaturing DNA fragments often have a different base composition than the bulk of the DNA and, consequently, often separate as small satellite bands upon centrifugation in a CsCl gradient. Satellite DNA is usually associated with regions of the chromosome that do not unravel in telophase as does the bulk of the DNA. Satellite DNA usually consists of short highly repetitive sequences,[148] which occur in large clusters of up to 100 Mb of DNA, often near centromeres or telomeres or on a Y chromosome.[146] The DNA of a satellite band from the kangaroo rat contains the sequence 5'-GGACACAGCG-3' repeated so often that it accounts for 11% of the entire DNA of the cell. Longer repeating sequences ~170 bp are also often present as are **microsatellites** of 2- to 5-base-pair repeats. At least 30,000 microsatellite loci are present in the human genome.[149]

Centromeres. The attachment of spindle fibers to chromosomes depends upon the segments of DNA known as centromeres to organize the attached kinetochores. In the yeast *Saccharomyces cerevisiae* a 120-bp region containing three short conserved sequences is present in the centrosomes of all ten chromosomes.[150,151] This may fold into a distinctive looped structure.[152] Human centrosomes are large and complex,[153] but the DNA is highly repetitive, giving rise to **α-satellite DNA**.[154–157b] Sequences such as $(TGGAA)_n$ are repeated many times. Such sequences can form self-complementary looped structures containing some unpaired guanines that intercalate and stack between sheared G•A pairs.[158,158a] Complex regional centromeres involving kilobases or megabases of DNA have also been identified in fission yeasts, *Drosophila*, and green plants.[156,157] The great variability indicates that centromeric sequences undergo rapid

evolution. This may be related to the fact that of the four cells formed by female meiosis only one becomes an egg.[158b] A series of unique **centromeric proteins** (CENP-A to CENP-G) bind to the DNA sequences of centromeres[159–162a] and direct the formation of the kinetochores. Even for the simpler centromere of budding yeasts, kinetochores have a complex structure.[150] The CENP proteins were first identified as autoantigens in sera of patients with the autoimmune disorder **scleroderma** (Chapter 31).[160,162]

Telomeres. The DNA sequences at the chromosome ends have a TG-rich strand, such as the $(TTGGGG)_{50–70}$ (*Tetrahymena*)[163] and the $(TTAGGG)_n$ of both human and trypanosome chromosomes.[164–166] The complementary DNA strand is CA-rich. The *S. cerevisiae* telomers have ~350 base pairs containing the sequences ($TG_{1–3}$ / $C_{1–3}A$) as well as one or more copies of a 6.7-kb nonrepetitive sequence and other elements.[166,167] In many species the repetitive telomeric sequences have 3' poly(G) tails at the ends of the DNA molecules. These tails are able to form G quartet structures (Fig. 5-26 and Chapter 5, Section C,4). A variety of telomere-binding proteins have been isolated.[167a–d] Some of these bind to G quartet structures[168–169a] and some, such as **RAP1** of yeast,[170–171a] to double-stranded telomere repeat regions. Special problems associated with replication of telomeres are discussed in Section C,8.

Short interspersed sequences (SINES). Much of the reiterated DNA is present in repeated segments 100–500 bp in length that lie between 1- to 2-kb segments of unique DNA.[172–174] The best known example is the human ***Alu*** family,[175–177b] so-called because it contains a site for cleavage by restriction enzyme *Alu*I. Sequences of the *Alu* family also exists in other primates and in rodents. The ~300-bp *Alu* sequence is reiterated over 500,000 times in the human genome with various sequence alterations, but an 80–90% homology. This sequence (Fig. 27-9) consists of two similar ~130-bp segments called the "left monomer" and "right monomer." The right monomer contains a 31-bp insertion and the left end carries a poly(dA) sequence. In addition there is a short 7- to 20-bp sequence, which is variable between different *Alu* sequences but is directly repeated at each end of a given *Alu* sequence.

The *Alu* sequence has strong homology with the 7S RNA (Table 5-4) that is part of the signal recognition particle involved in transport of newly synthesized peptide chains across the membranes of the ER (Chapter 10). *Alu* sequences are transcribed into hnRNA, the precursors to mRNA. Some *Alu* sequences are present in intervening sequences (introns) within genes and others are in noncoding sequences between genes. Sharp[176] suggested that specific proteins in the nucleus may bind to the *Alu* sequences, preventing hnRNA from leaving the nucleus before it has been processed to remove introns and other sequences absent in mRNA. However, the presence of the poly(dA) regions and the direct terminal repeats suggests that the whole *Alu* sequence is **pseudogene** derived from a **retroposon**, a type of transposon that originated from an RNA molecule. *Alu* and other SINES contain an RNA polymerase III promoter and are transcribed. Active retroposons form reverse transcripts (cDNAs) that can be integrated at various points in the genome (see Section D,4 and Chapter 28). One theory is that retroposons have no biological function but have invaded the genome at random locations.

Many SINES and other families of repetitive sequences have been characterized by the presence of a restriction enzyme cleavage site in each copy of the sequence. *Eco*RI cuts the previously mentioned α-satellite DNA.[177] A 319-bp reiterated sequence in the human genome surrounds a *Hin*f site,[178] and *Sau*3A cleaves a 849-bp sequence with ~1000 copies per haploid genome.[179] In the hermit crab 30% of the genome consists of repeated sequences, one 156-bp unit occuring ~7 million times. Many identical 14-bp GC-rich inverted repeats are present.[180] *Neurospora* and yeast both contain many copies of the GC-rich palindromic sequence 5'-CCCTGCAGTACTGCAGGG-3', which contains the two underlined *Pst*I sites.[181] Some repetitive sequences such as the **CAT** family and the **homeodomain sequence** (Chapter 28, Section C,6) are found in the control regions preceding the 5' ends of groups of genes that are regulated coordinately.[182] Some repetitive sequences seem to be unstable in the genome and may be excised and lost or may increase in number during aging.[183] Repeated DNA

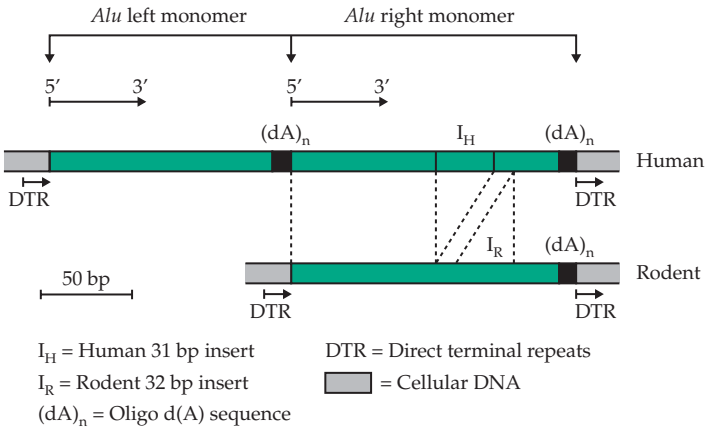


Figure 27-9 Structure of human and rodent *Alu* sequences in DNA. From Ullu.[175]

sequences have apparently originated not only with 7S RNA but also from mRNA or tRNA molecules.[172] A characteristic SINE in rodent DNA is known as the ID (identifier) sequence because it was once thought to be a marker for genes transcribed in neural tissues. Rat DNA contains ~130,000 copies of the ID sequence.[172a] Other **clustered repeats** are frequently found between genes that are present in large numbers. These include the genes for ribosomal RNA, tRNA, small nuclear RNAs, and histones. Triplet repeats are considered in Chapter 26, Section G,3.

Long interspersed repeat sequences (LINES). These moderately repetitive sequences may be several kb in length. Just one type seems to be abundant in each mammalian species.[122] The human L1 sequence is over 6 kb in length. Like the *Alu* sequence it has a poly(dA) sequence at the 3'-end, but it does not contain short terminal direct repeats. These interspersed repeat sequences usually contain genes which may be functional, but many of the copies contain pseudogenes or genes that are randomly truncated, having lost a segment from the 5' end. These sequences, too, seem to have been dispersed throughout the genome by retroposons.[122,184,185] Human chromosome 22, sequenced in 1999, contains within its 33.4 Mb of euchromatin at least 545 genes, 134 pseudogenes, and 8043 L1 sequences (9.7% of the DNA), as well as 20,188 *Alu* sequences (16.8% of the DNA) and many other interspersed repeats.[186] Repetitive DNA sequences are less common in prokaryotes, but they do exist. For example, many dispersed and clustered repeating units are present in *Halobacterium* DNA.[187]

2. Genes for Ribosomal RNA and Small RNA Molecules

Most genes are present as one copy each per haploid genome. However, there are many copies of the genes for ribosomal RNA and tRNA. In *Xenopus* DNA there are ~450 repeats of the 28 S and 18 S rRNA genes on one chromosome and ~24,000 copies of the 5 S RNA genes at the ends of the long arms of most of the chromosomes.[188] Nontranscribed spacer regions lie between the repeats of the 28S and 18S gene pairs, as can be seen with the electron microscope in Fig. 28-17. The gene for 5 S RNA has a high GC content. A denaturation map of the DNA shows easily melted regions separated by shorter 120-bp sequences, apparently of high GC content and presumably coding of the 5 S RNA. The easily denatured AT-rich spacers are ~630 bases long. Using restriction enzymes much of this DNA can be cut into segments that contain repeats within repeats. One 15-unit polynucleotide contains the sequence $A_4CUCA_3CU_3G$ repeated about 30 times.[189]

About 200 copies of rRNA genes per haploid genome are located at the constrictions in the short arms of human chromosomes 13, 14, 15, 21, and 22 (Chapter 26, banner). As in *Xenopus*, clusters of tandemly repeated 5S RNA genes are found at the ends of the long arms of most chromosomes. A similar organization of rRNA genes is found in the rat.[190] The ten different chromosome ends carrying rRNA genes in the human diploid nucleus come together to form the nucleolus, the site of synthesis of ribosomal subunits.[191] At first there are ten small nucleoli, but these fuse to form the single highly structured but membraneless nucleolus. The rRNA genes of the macronucleus of *Tetrahymena* have been "amplified" and are found on linear 21-kb palindromic molecules, about 10^4 copies being present per cell. Initiation of transcription begins near the center and proceeds outward in both directions.[192] These short chromosomes contain typical telomeric ends.

Genes for the tRNAs are spread throughout the genome of bacteria, mitochondria (Fig. 18-3), chloroplasts,[193] and eukaryotic nuclei. They sometimes occur in clusters but more often are far apart. In *Drosophila* there are probably at least 600 tRNA genes, many occurring in pairs of opposite polarity, i.e., as inverted repeat sequences. The genes for small nuclear RNAs U1–U6 (Chapter 28) are organized in a variety of ways. The ~100 human U1 genes occur on a single chromosome, perhaps organized in a tandem array and interspersed with as many as 10,000 defective pseudogenes.[194]

3. Other Gene Clusters and Pseudogenes

Closely related structural genes often occur in clusters.[195] Among these are clusters of **immunoglobulin** genes (Fig. 31-17) and clustered genes for the α and β **globins** (Fig. 27-10), which encode the protein sequences for the hemoglobins. The human α globin gene cluster occupies about 30 kb on chromosome 16 and the β globin genes 60 kb on chromosome 11.[196–197d] The α1 and α2 genes (Fig. 27-10) encode identical peptides, while the related ζ gene encodes the corresponding subunit of embryonic hemoglobin. The β cluster includes, in addition to the adult gene, a pair of fetal globin genes (γ^G and γ^A) differing by only one amino acid (Gly vs Ala) at position 136, the embryonic ε chain, and the minor adult δ chain.

Besides the functional genes the globin cluster contains **pseudogenes**, which are given the prefix ψ in Fig. 27-10. These are nonfunctional genes, which appear to encode peptides homologous to the known globins. However, they contain mutations that prevent expression. For example, deletion of a single nucleotide near the beginning of the pseudogene will scramble the genetic message by changing the reading

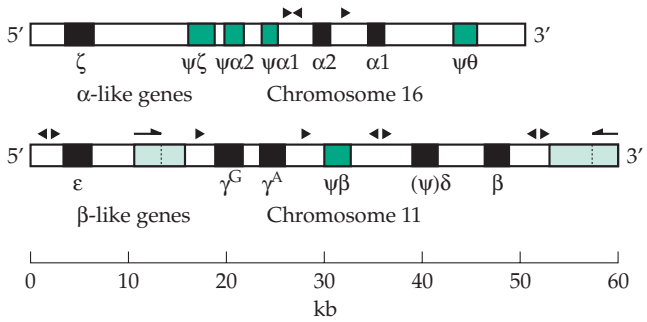


Figure 27-10 Organization of the globin genes on human chromosomes 11 and 16. The composition of the various embryonic, fetal, and adult hemoglobins is also indicated. Closed boxes indicate active genes and open boxes pseudogenes. The triangles (▶) indicate *Alu* repetitive sequences and their orientation. The shaded boxes indicate Kpn repeat sequences and the half-arrows their respective orientation. The Kpn sequence between the ε and γ^G genes in fact consists of two tandemly linked Kpn repeats. From Karlsson and Nienhuis[196] and Proudfoot.[197]

frame. Globin pseudogene ζψ has a nonsense mutation at codon 6, ψα1 contains a whole set of mutations that prevent transcription and translation, while ψα2 is mutated almost to the point of being unrecognizable as a globin relative.[198]

Do pseudogenes have a physiological function? The presence of pseudogenes in close proximity to functional genes suggests that there may be an as yet unrecognized controlling function.[198] On the other hand, they may be relics of evolution. If a gene family arises by duplication and mutation of duplicated copies, we may expect to find gene copies that have not been selected as useful but are still present in the genome.

There are regions of the genome that seem to be largely single copy DNA with few reiterated sequences or gene clusters.[144] While some gene families exist as clusters others are dispersed throughout the genome.

4. Introns, Exons, and Overlapping Genes

A major difference between prokaryotes and eukaryotes is the presence of intervening sequences (introns) between the coding sequences (exons) in eukaryotes. Introns are especially numerous in higher organisms. For example, the gene for the myosin heavy chain present in rat embryonic skeletal muscle[199] encodes a 1939-residue peptide but occupies a length of 22 kb of DNA. The gene is split into 41 exons, whose transcribed RNA must be cut and spliced at 40 places to form the mRNA. By comparison the corresponding gene from the nematode *Caenorhabdatis elegans* is far less fragmented. It is not clear why some genes have so many introns and others so few. The very compact chromosomes of viruses and most bacteria do not contain introns. However, they are sometimes present in mitochondrial and chloroplast genes.

A surprise, which was first recognized in viral RNA and DNA, is that genes sometimes overlap. For example, two proteins, one long and one shorter, are synthesized starting at the same point in the RNA genome of phage Qβ.[200] In the DNA of phage φX174 the third nucleotide of the stop signals for some genes are also the first nucleotides for the start signal for translation of the next gene. Pairs of genes have been found in which one of the genes of the pair is found completely inside the other gene but is translated in a different reading frame.

A four-base overlap between dihydrofolate reductase and thymidylate synthase has been found[201] in the DNA of phage T4. A transposable DNA insertion sequence (see Section D,5) in *E. coli* encodes two genes, one of which is contained within the other and which is transcribed from the opposite strand of DNA.[202] The double-stranded RNA of a reovirus produces two peptides from the same sequence using two different AUG initiation codons in different reading frames.[203]

Overlapping obviously limits severely the mutational alterations that are allowable. Perhaps it is for this reason that eukaryotic genes seem to be dispersed in widely separated locations and overlap is rare. However, the human type IV collagen genes for the α1 and α2 chains are encoded on opposite strands with their 5' flanking regions overlapping.[204] Some introns contain genes.[205,206] Chlorarachniophyte algae contain a multimembraned chloroplast thought to be a vestigial remnant of an endosymbiont. Its 380-kb genome consists of three short chromosomes that encode overlapping genes and contain the shortest introns (18–20 base pairs) known.[207]

5. DNA of Organelles

Mitochondrial DNA (mtDNA), discussed in Chapter 18, varies in size from <6 kb to 367 kb.[208] The small circular mtDNA of animals is extremely efficiently packed with genes for tRNAs, rRNA, and a small number of protein subunits (Fig. 18-3). However, the 78-kb yeast mtDNA contains many long AT-rich spacers as well as long introns, some of which contain genes for splicing enzymes. Otherwise, it is similar to animal mitochondrial DNA. Mitochondrial genomes of higher plants are much larger; that of maize is a 570-kb circle, which contains both direct and inverted repeat sequences. Recombination between these sequences is apparently the origin of smaller incom-

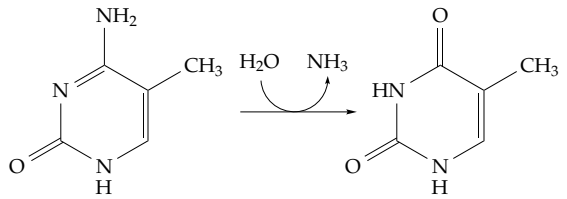
plete circular genomes that are also present. Plant mtDNA appears to contain the genes present in animal mitochondria plus additional genes. For example, one encodes a 5S RNA. The most unusual mtDNA is found in hemoflagellates such as *Trypanosoma* and *Leishmania*. The single large mitochondrion or **kinetoplast** located near the base of the flagellum contains a network of catenated circular DNA molecules (Fig. 5-16).[209,210] The genetic makeup is similar to that of other mtDNA, but the rRNA genes are unusually short. Chloroplast DNA, which varies in size from 120 to 160 kb, is discussed in Chapter 23.

6. Methylation of DNA

As mentioned in Chapter 5, a significant fraction of the pyrimidine and purine bases in DNA is methylated. One function of such methylation in bacteria, discussed in Chapter 26, is to protect against the action of restriction endonucleases.[211] For example, the gene for the well-known *Eco*RI endonuclease is carried in *E. coli* cells by an R factor. This plasmid also carries (just 29 base pairs away) the gene for a 326-residue **N^6-adenine methylase**.[212–213a] This enzyme uses *S*-adenosylmethionine (AdoMet) to methylate the two adenines (marked by asterisks) in the six-base-pair recognition sequence 5'-G A A* T T C / 3'-C T T A* AG converting them to N^6-methyladenines (m^6A). Other DNA methyltransferases place methyl groups on N-4 of cytosine or on C-5 of cytosine.[214–215a] The latter utilizes a mechanism illustrated in Eq. 12-4.[216,217] Such enzymes are components of both type II and type I restriction–methylation systems.[218] However, most of the m^6A in the *E. coli* chromosome arises from action of a different methylase, one that recognizes the palindromic sequence 5'-GATC and methylates adenines in both chains.[219] This **DNA adenine methylase**, a product of the *dam* gene, plays an important role in mismatch repair, transposition, regulation of transcription, and initiation of DNA replication.[220] The same methylase regulates at least 20 genes induced during infection by *Salmonella typhimurium*. Some of these genes are essential to virulence.[221] A similar methyltransferase appears to control differentiation of the stalked *Caulobacteria* cells.[220]

CG doublets. The only modified base commonly found in eukaryotes is 5-methylcytosine,[222,223] which upon deamination becomes thymine (Eq. 27-1). Most methylation occurs when C is followed by G. Usually 60–90% of all 5'-CG sequences (CpG sequences) in eukaryotic DNA are methylated. However, the fraction of methylated cytosine varies from almost zero for *Drosophilia, Caenorhabditis*, and *Saccharomyces* to as much as 30% in higher plants.[224]

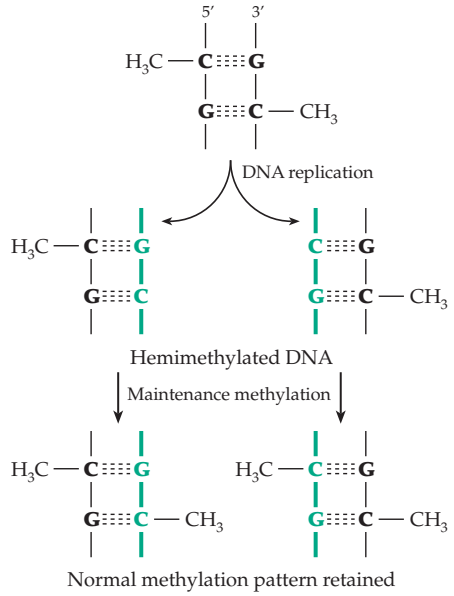
When CG pairs are methylated on both cytosines,



5-Methylcytosine (m^5C) Thymine (T) (27-1)

an interesting result arises upon replication of the DNA. The methyl groups don't prevent replication, but only one of the DNA strands in each daughter duplex is methylated (Eq. 27-2). However, additional methylation of DNA occurs within three hours of DNA replication by a **maintenance methyltransferase** that recognizes methylated CpG sequences in the old DNA strand and methylates the cytosine in the 3′-GpC of the newly synthesized strand.[220,225,226] Mammalian cells contain at least three different DNA (cytosine-5) methyltransferases.[227–229b] Enzymatic **demethylation**, which converts m^5C residues back to cytosine, may also occur.[230–233] It has been difficult to demonstrate[233a] but enzymatic demethylation of 5-methylcytosine has been reported.[233b,c] The observed loss of methylation during embryonic development (Chapter 32) may be a result of loss of the maintenance methylase. Most CpG doublets are found in large "islands" of several hundred bases to about 2 kb in length, which are unusually rich in G + C.[234,235] These islands lie near the 5' ends of ~60% of all human genes[236] and near origins of replication.[237] They are also found in regions of the DNA that are compacted into heterochromatin. The genes in methylated regions tend to be "silent," i.e., they are not actively transcribed.[236] Demethylation may permit transcription.

The degree of methylation of the CG doublets is variable both in position within a chromosome and



(27-2)

with stage of development. Cytosine methylation is essential to embryonic development,[238] and mice lacking the maintenance DNA methytransferase are developmentally retarded and die at mid-gestation.[239] Not all CG sequences are methylated and various patterns of DNA methylation are generated at different stages of development by rounds of methylation and by the action of demethylases.[232] The maintenance methylase (Eq. 27-2) ensures that a stable methylation pattern persists until altered by new rounds of methylase or demethylase action. A parallel is found in the prokaryotic *Caulobacter* in which three chromosomal sites are fully methylated in swarmer cells, become hemimethylated in stalked cells, and are fully methylated again just prior to cell division.[220] See Fig. 32-1.

DNA methylation affects transcription, either directly by preventing the binding of transcription factors or indirectly via a series of binding proteins specific for methylated CG doublets. In early stages of embryonic development there is very little methylation, but some genes are quickly silenced as methylation takes place. Heterochromatic regions including inactivated X chromosomes are heavily methylated. However, additional alteration in chromatin is required for complete silencing of genes.[236,240] Recent studies indicate that an abundant mammalian protein binds to the methylated DNA along with a **histone deacetylase**.[239,241–244] The latter acts on acetylated histones to free lysine side chains, which may interact in an inhibitory manner with the DNA.

In female mammalian cells most of the genes on one of the two X-chromosomes are completely inactivated. DNA methylation plays a major role in this process.[244,245] A perfect correlation has been observed between 5'-methylation of cytosines in CpG islands and inactivation of X-chromosome genes.[246] Methylation may also play a role in recombination and repair.[247] Methylation of DNA decreases with increasing age.[248] It increases as a result of oncogenic transformation of cells.[249] Some other modifications of DNA largely limited to bacteriophages are discussed on p. 234.[247,250]

Imprinting. With the exception of X-linked genes each person has two copies of each gene, one of maternal and one of paternal origin. Both copies of most of these genes are expressed. However, a few of the genes receive from one parent or the other an **imprint**, a mark that distinguishes the parental origin.[251–253] Such imprints are maintained in cells through embryonic development but are erased in embryonic gonads to allow for a new imprint in the germ cells. Imprinting depends upon DNA methylation, and all imprinted genes show the presence of differentially methylated regions.[253] See also Chapter 32, Section A,1.

C. Replication

Following the discovery of the double helix and the enthusiasm that it engendered many people thought that the synthesis of DNA was simple. The nucleotide precursors would align themselves along separated DNA template strands and perhaps spontaneously snap onto the growing chains. In fact, replication is a complex process that requires the cooperative action of many different gene products and perhaps an association with membrane sites. The matter is made more complex by the fact that some of the enzymes involved in replication are also required in the processes of genetic recombination, in repair of damaged DNA molecules, and in defensive systems of cells.

1. Early Studies

That the DNA content doubles prior to cell division was established by microspectrophotometry. It was clear that both daughter cells must receive one or more identical molecules of DNA. However, it was not known whether the original double-stranded DNA molecule was copied in such a way that an entirely new double-stranded DNA was formed or whether, as we now know to be the case, the two chains of the original molecule separated. The latter is called **semiconservative** replication, each of the separated strands having a new complementary strand synthesized along it to form the two identical double-stranded molecules.

The first definitive evidence for semiconservative replication was reported by Meselson and Stahl[254] in 1958. Cells of *E. coli* were grown on a medium containing isotopically pure $^{15}NH_4^+$ ions as the sole source of nitrogen. After a few generations of growth in this medium the DNA contained exclusively ^{15}N. Then the cells were transferred abruptly to a medium containing $^{14}NH_4^+$ and were allowed to grow and to double and quadruple in number. At various stages DNA was isolated and subjected to ultracentrifugation in a CsCl gradient. Small but easily detectable differences in density led to separation of dsDNA molecules containing only ^{15}N from those containing partly ^{15}N and from those containing only ^{14}N. At the beginning of the experiment only DNA containing entirely ^{15}N was present. However, after one generation of growth in the ^{14}N-containing medium, the density of *all* the DNA was such as to indicate a content of one-half ^{14}N and one-half ^{14}N. After a second generation of growth half of the DNA still contained both nitrogen isotopes in equal quantity, whereas half contained only ^{14}N, exactly the result expected for semiconservative replication. A similar experiment using 5-bromodeoxyuridine, a thymidine analog, is shown in Fig. 27-11.

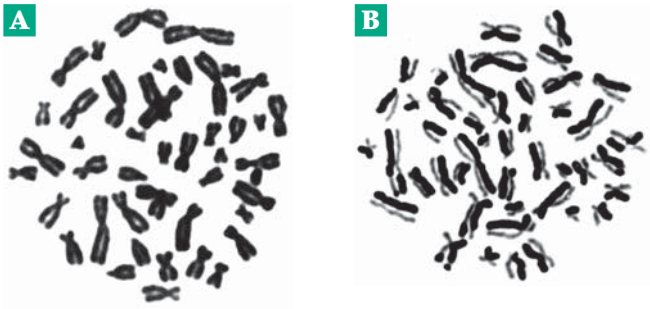
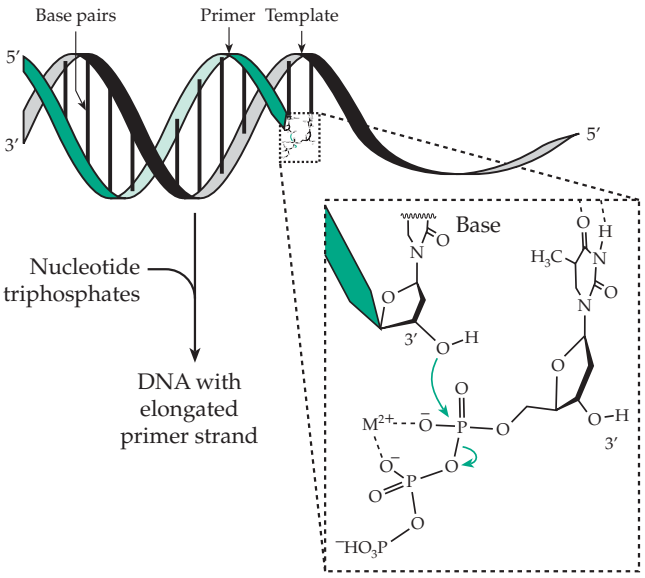


Figure 27-11 (A) Human chromosomes after one replication in the presence of 5-bromodeoxyuridine (BrdU). Both chromatids of each chromosome contain BrdU in one strand of the DNA duplex and normal thymidine in the other. (B) After two replications in the presence of BrdU one chromatid of each chromosome contains BrdU in both strands of the duplex and stains strongly with a special differential staining procedure. The other chromatid contains only normal thymidine in one strand of the duplex and is not stained. Courtesy of Carolina Biologicals.

Autoradiography. Later a technique of direct autoradiography of DNA using ^{3}H-labeled thymidine[255] was applied by Cairns.[256,257] Cells of *E. coli* were grown on a medium containing the radioactive thymidine for various times but typically for 1 h (~2 generations). The cells were then ruptured, the DNA was spread on a thin membrane filter, and autoradiograms were prepared. When the DNA contained [^{3}H]thymidine the exposed trace in the autoradiogram could be followed around the entire 1.1–1.4 mm circumference of the spread DNA molecule. Molecules partially labeled and in the process of replication could also be identified. After two hours of growth in the presence of ^{3}H-labeled thymidine about half of the bacterial DNA was fully labeled, but half contained regions that were labeled only half as heavily. They presumably contained ^{3}H label in a single strand and, therefore, represented unreplicated regions. All of the molecules had undergone one round of replication with tritiated thymidine to yield lightly labeled molecules; parts of the molecules had not completed the second round. The more heavily labeled regions were interpreted as fully replicated. The shapes of the "replication figures" suggested that DNA is synthesized in a continuous manner starting from one point and continuing around the circular molecule at a constant rate. Although subsequent experiments (considered in Section 4) show that replication is usually **bidirectional**, the experiments of Cairns were important because they introduced a technique for direct visualization of replication *in vivo*.

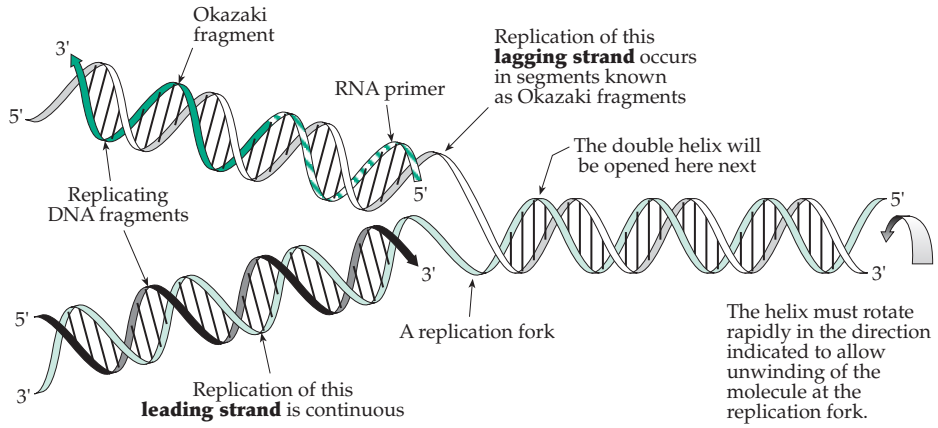


(27-3)

The chemistry of DNA polymerization. What are precursors of DNA? Early experiments showed that the nucleoside [^{3}H]thymidine was efficiently incorporated into DNA, but for energetic reasons it seemed unlikely that thymidine was an immediate precursor. Evidence favoring the nucleoside triphosphates was provided in 1958 when Arthur Kornberg identified a DNA polymerase from *E. coli*. Kornberg's enzyme, now known as **DNA polymerase I**, was isolated in the amount of 600 mg from 90 kg of bacterial cells[258,259] (over 400 molecules of enzyme per cell). The 928-residue enzyme displayed many of the properties expected of a DNA-synthesizing enzyme. It requires a **template strand** of DNA as well as a shorter **primer strand**. As indicated in Eq. 27-3, the enzyme recognizes the 3' end of the primer strand and binds the proper nucleoside triphosphate to pair with the next base in the template strand. Then it catalyzes the displacement of pyrophosphate, at the same time linking the new nucleotide unit onto the 3' end of the primer strand. Continuing in this way, the enzyme is able to turn a single-stranded template DNA into a double-stranded DNA in which the newly synthesized strand contains, at each point, the base complementary to the one in the template strand.[259a]

Although the action of the DNA polymerase I, according to Eq. 27-3, provided a straightforward way to form a complementary strand of DNA, it did not explain how double-stranded DNA could be copied. One problem is that the two strands must be separated and unwound. If unwinding and replication occured at a single **replication fork** in the DNA, as indicated by Cairns' experiment, the entire molecule would have to spin at a speed of 300 revolutions per second to permit replication of the *E. coli* chromosome in 20 min. It also required that some kind of a **swivel**, or at least a

nick in one chain, be present in the chromosome as indicated in the following drawing.



Another problem was posed by the fact that the two chains in DNA have opposite orientations. Thus, at the replication fork one of the new chains might be expected to grow by addition of a new nucleotide at the 3' end, while the other chain would grow at the 5' end. *If so, there should be two DNA polymerases*, one specific for polymerization at each of the two kinds of chain end. Nevertheless, despite intensive search the only DNA polymerases found added new units only at the 3' end.

Discontinuous replication and RNA primers. In 1968, Okazaki reported that during the time that replication of DNA is taking place bacterial cells contain short fragments of DNA. These are now called **Okazaki fragments** or **replication fragments**.[260] A second development was the discovery of the enzyme **DNA ligase**,[261,262] which is able to join two pieces of DNA to form a continuous chain. These two discoveries provided an explanation for the lack of a second kind of DNA polymerase. One strand, the **leading strand**, of the replicating DNA could be synthesized continuously in the 5' to 3' direction while the other strand, the **lagging strand**, would have to be synthesized in segments (the Okazaki fragments), which could then be joined by the DNA ligase. In 1971 Brutlag *et al*. reported that initiation of synthesis of the DNA of phage M13 in *E. coli* required formation of a short segment of RNA as a primer.[263] It was subsequently shown that *with few exceptions priming by RNA synthesis is always required for replication*.

Newer studies have revealed great complexity in the mechanisms of replication. As for synthesis of any polymer there are distinct steps of *initiation*, *elongation*, and *termination*. Topological problems associated with the unwinding and rewinding of the double helices and with disconnection of catenated circles and untying of knots are solved with the aid of special enzymes, the **helicases** and **topoisomerases**. Replication requires both DNA and RNA polymerases, a ligase, and ancillary proteins, some of whose functions aren't yet clear. Many of these proteins associate to form large multiprotein complexes, which are given names such as **primosome** (for priming) and **replisome**. Many bacteriophages and plasmids also replicate within cells of *E. coli* utilizing bacterial proteins as well as proteins encoded in the viral or plasmid genome.

2. DNA Polymerases

Immediately after its identification DNA polymerase I was generally assumed to be the principal chain elongation enzyme. However, an *amber* mutant of *E. coli* deficient in DNA polymerase I (gene *pol*A; Fig. 26-4A) synthesized DNA normally. This finding stimulated an intensive search for new polymerases. Two were found: DNA polymerases II (gene *pol B*)[264] and III. Both are present in amounts less than 25% of that of DNA polymerase I.[265,266] Both have properties similar to those of polymerase I, but there are important differences. By now DNA polymerases have been isolated from many organisms, many genes have been cloned and many sequences, both of bacterial and eukaryotic polymerases are known. Comparisons of both sequences and three-dimensional structures,[266a,b] a few of which are shown in Fig. 27-12, suggest that the polymerases belong to at least six families (Table 27-1). These include the RNA-dependent DNA polymerases known as **reverse transcriptases** as well as some **RNA polymerases**.[267–268b]

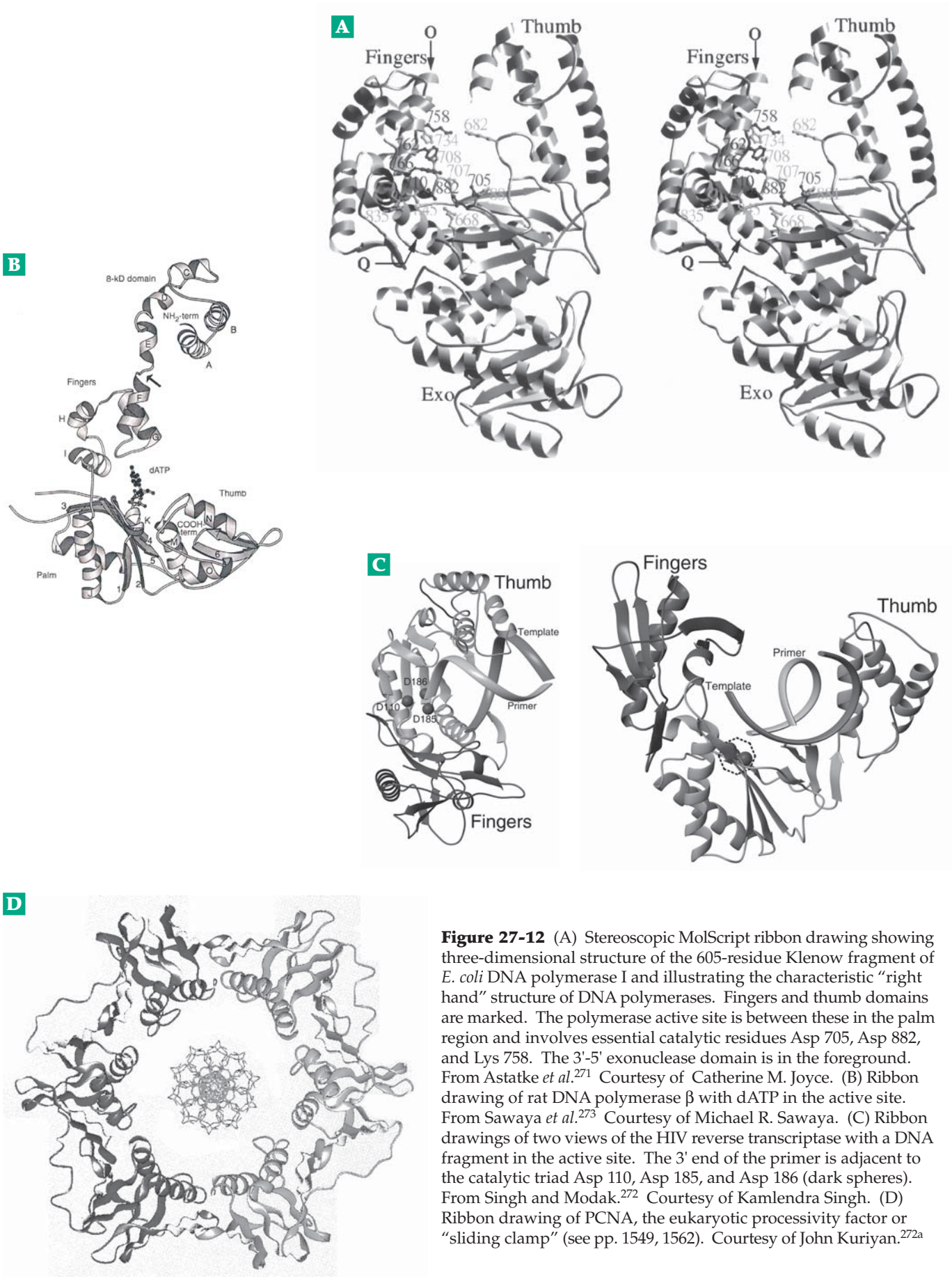
Some of the polymerases exist as single polypeptide chains, while others function only as large complexes. In every case a two-metal ion catalytic mechanism with in-line nucleotidyl transfer,[269] illustrated in Fig. 27-13, appears to be used by the enzymes.[267,270] Two-metal ion catalysis is also observed for phosphatases and ribozymes (Chapter 12).

Exonuclease activities, proofreading, and editing. DNA polymerase I not only catalyzes the growth of DNA chains at the 3' end of a primer strand but also, at about a 10-fold slower rate, the hydrolytic removal of nucleotides from the 3' end (**3'- 5' exonuclease activity**). The same enzyme also catalyzes hydrolytic removal of nucleotides from the 5' end of DNA chains. This latter **5'- 3' exonuclease activity**, the DNA polymerase activity, and the 3'-5' exonuclease activity all arise from separate active sites in the protein. DNA polymerases II and III do not catalyze

TABLE 27-1
Families of DNA Polymerases[a,]

Class	Name	Function	Molecular mass (kDa)
A	*E. coli* polymerase I (Pol I)	DNA excision repair	103
	Klenow fragment		68
	Bacillus subtilis Pol I[b]	DNA excision repair	
	Thermus aquaticus DNA polymerase (Taq)[c,d]	DNA excision repair	
	T7 DNA polymerase[e]	Virus replication	80
	T7 RNA polymerase[f,g]		99
	Eukaryotic Pol γ (gamma)	Mitochondrial replication	
	Eukaryotic Pol θ (theta)	DNA repair	
B	Eukaryotic Pol α (alpha)	DNA replication	180 (core)
	Eukaryotic Pol δ (delta)	DNA replication	
	Eukaryotic Pol ε (epsilon)	DNA replication	
	Eukaryotic Pol ζ (zeta)	Bypass synthesis	
	Bacteriophage T4 DNA Pol + accessory proteins[h,i]	DNA replication	43
	E. coli Pol II[j]		90
C	Bacterial DNA Pol III[p] + accessory proteins	DNA replication	~900 (holoenzyme)
D	Euryarchaeotic Pol II		
X	Eukaryotic DNA Pol β[n,o]	DNA repair	39 x 2
	Eukaryotic Pol λ (lambda)	Base excision repair	
	Eukaryotic Pol μ (mu)	Non-homologous end-joining	
	Eukaryotic Pol σ (sigma)	Sister chromatid cohesion	
Y	*E. coli* UmuC protein	SOS response	
	Eukaryotic Pol η (eta, XP-V, RAD 30)	Bypass synthesis	
	Eukaryotic Pol ι (iota)	Bypass synthesis	
	Eukarytoic Pol κ (kappa)	Bypass synthesis	
Reverse transcriptase family			
	HIV reverse transcriptase[k,l]		
	Telomerase[m]		
	RNA-dependent RNA polymerases		

[a] Based on reviews by Burgers, P. M. J. *et al.* (2001) *J. Biol. Chem.* **276**, 43487–43490 and Steitz, T. A. (1999) *J. Biol. Chem.* **274**, 17395–17398
[b] Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature (London)* **391**, 304–307
[c] Eom, S. H., Wang, J., and Steitz, T. A. (1996) *Nature (London)* **382**, 278–281
[d] Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17**, 7514–7525
[e] Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature (London)* **391**, 251–258
[f] Sousa, R., Chung, Y. J., Rose, J. P., and Wang, B.-C. (1993) *Nature (London)* **364**, 593–599
[g] Sastry, S., and Ross, B. M. (1999) *Biochemistry* **38**, 4972–4981
[h] Wang, J., Yu, P., Lin, T. C., Konigsberg, W. H., and Steitz, T. A. (1996) *Biochemistry* **35**, 8110–8119
[i] Jing, D. H., Dong, F., Latham, G. J., and von Hippel, P. H. (1999) *J. Biol. Chem.* **274**, 27287–27298
[j] Cai, H., Yu, H., McEntee, K., Kunkel, T. A., and Goodman, M. F. (1995) *J. Biol. Chem.* **270**, 15327–15335
[k] Morris, M. C., Berducou, C., Mery, J., Heitz, F., and Divita, G. (1999) *Biochemistry* **38**, 15097–15103
[l] Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C., and Harrison, S. C. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1222–1226
[m] Lundblad, V. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8415–8416
[n] Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science* **264**, 1930–1935
[o] Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) *Biochemistry* **36**, 11205–11215
[p] Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York


A
O
Thumb
Fingers
758
682
734
708
707
705
882
668
Q
Exo
B
8-kD domain
NH2-term
B
A
Fingers
H
I
dATP
Thumb
COOH term
Palm
C
Thumb
Template
D186
D110
D185
Primer
Fingers
Fingers
Thumb
Primer
Template
D
Figure 27-12 (A) Stereoscopic MolScript ribbon drawing showing three-dimensional structure of the 605-residue Klenow fragment of E. coli DNA polymerase I and illustrating the characteristic "right hand" structure of DNA polymerases. Fingers and thumb domains are marked. The polymerase active site is between these in the palm region and involves essential catalytic residues Asp 705, Asp 882, and Lys 758. The 3'-5' exonuclease domain is in the foreground. From Astatke et al.271 Courtesy of Catherine M. Joyce. (B) Ribbon drawing of rat DNA polymerase β with dATP in the active site. From Sawaya et al.273 Courtesy of Michael R. Sawaya. (C) Ribbon drawings of two views of the HIV reverse transcriptase with a DNA fragment in the active site. The 3' end of the primer is adjacent to the catalytic triad Asp 110, Asp 185, and Asp 186 (dark spheres). From Singh and Modak.272 Courtesy of Kamlendra Singh. (D) Ribbon drawing of PCNA, the eukaryotic processivity factor or "sliding clamp" (see pp. 1549, 1562). Courtesy of John Kuriyan.272a

hydrolysis from the 5' end.

Treatment with proteolytic enzymes cuts a 323-residue piece containing the 5'-3' exonuclease from DNA polymerase I leaving a larger C-terminal piece known as the **Klenow fragment**. This fragment retains the polymerase activity as well as the 3'-5' exonuclease activity and is widely used in genetic engineering. Its three-dimensional structure[271,275] is shown in Fig. 27-12A. In the Klenow fragment the large C-terminal domain contains the polymerase. The N-terminal domain contains the 3', 5'-exonuclease activity, which is thought to fulfill a **proofreading** and **editing** function.[275a] The polymerase acts at the 3' end of the growing DNA chain. Before moving on to the next position, the enzyme verifies that the correct base pair has been formed in the preceding polymerization event. If it has not, the exonuclease action removes the incorrect nucleotide and allows the polymerase to add the correct one. Thus, each base pair is checked twice, first before polymerization and then after polymerization.

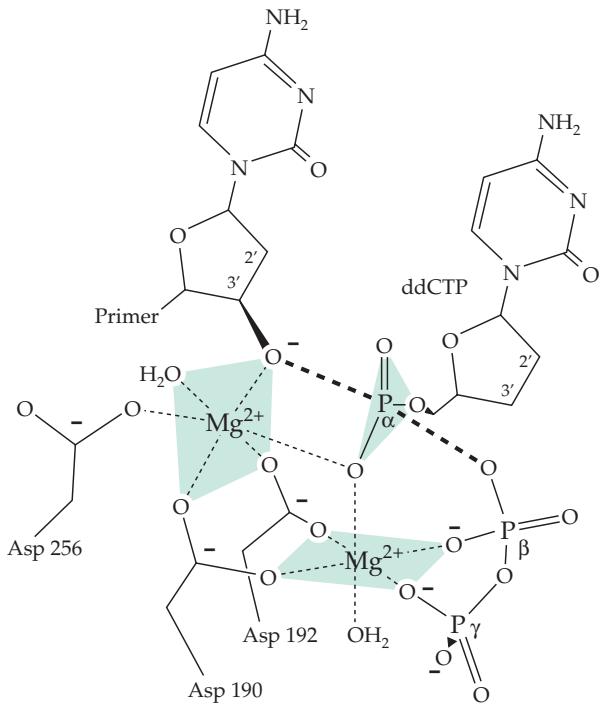


Figure 27-13 Proposed mechanism and transition state structure for the synthetic nucleotidyltransfer activity of DNA polymerase β (and other DNA polymerases). The chain-terminating inhibitor dideoxy CTP is reacting with the 3'-OH group of a growing polynucleotide primer chain. This –OH group (as $-O^-$) makes an in-line nucleophilic attack on P_α of the dideoxy-CTP. Notice the two metal ions, which interact with the phospho groups and which are held by three aspartate side chains. Two of the latter, Asp 190 and Asp 256, are present in similar positions in all of the polymerases. The active centers for the hydrolytic 3'-5' and 5'-3' exonuclease activities of some of the polymerases also appear to involve two-metal catalysis and in-line displacement. See Sawaya *et al.*[274]

A puzzle was the fact that structural studies indicate that the editing center is over 3 nm away from the catalytic center in Pol I.[267] The proposed solution to the puzzle is illustrated in Fig. 27-14. When the catalytic center "identifies" a nucleotide triphosphate as able to form a proper Watson–Crick nucleotide pair, it catalyzes the formation of the new nucleotide linkage. Then it releases the 3'-chain end, which sometimes "melts" and over a 10- to 100-ms time interval is able to reach over into the exonuclease site where the newly added nucleotide may be hydrolytically removed. However, if the newly formed nucleotide is properly paired, it will be less likely to melt, and the new nucleotide will be more likely to be retained.[259a,265,267] DNA polymerase I and other related polymerases utilize **processive mechanisms**, moving from one site to the next without diffusing away from the DNA. The schematic picture shown in Fig. 27-14 also indicates how the 5'-3' exonuclease activity can come into play, when the polymerase reaches the end of a single-stranded gap.

Other Class A polymerases. The *Thermus aquaticus* (***Taq***) polymerase is best known for its widespread use in the polymerase chain reaction (PCR; Fig. 5-47). Like *E. coli* I the enzyme is a large multidomain protein. The structure of the catalytic domains of the two enzymes are nearly identical, but the *Taq* polymerase has poor 3'-5' editing activity.[276] The enzyme has been carefully engineered to improve its characteristics for use in the PCR reaction.[277]

Some bacteriophage encode their own DNA polymerases. However, they usually rely on the host cell to provide accessory proteins. The sequence of the DNA polymerase from phage T7 is closely homologous to that of the Klenow fragment and the 3D structures are similar. The 80-kDa T7 polymerase requires the 12-kDa thioredoxin from the host cell as an additional subunit. It has been genetically engineered to improve its usefulness in DNA sequencing.[278]

About 45% of the sequence of the **RNA polymerase** encoded by phage T7, which transcribes RNA from the phage DNA, is also similar to that of the Klenow fragment. Sequences of these DNA polymerases are distantly related to those of reverse transcriptases.[279,280] The 136-kDa polymerase γ functions in mitochondria but is encoded in a nuclear gene. It is the only DNA polymerase that is inhibited by antiviral nucleotide analogs such as AZT (Box 28-C).[280a,b]

Polymerases of Class B. Although *E. coli* polymerase II is a member of this family, relatively little is known about its function. It may participate in DNA repair in the "SOS" response (Section E).[265]

The catalytic subunits of the major *eukaryotic* DNA

polymerases as well as of archaeal DNA polymerases are members of the B family.[267,281–282a] Eukaryotic cells contain at least 13 DNA polymerases which are designated by Greek letters (Table 27-1). Polymerases α, δ, and ε are essential for nuclear DNA synthesis.[283–286b] They function together with accessory proteins in primase and replisome complexes considered in Section 8. Others participate in DNA repair (pp. 1583, 1584).

Phage T4 encodes a DNA polymerase much used in the laboratory because of its ability to polymerize using a long single-stranded template. It also depends upon accessory factors provided by the bacterial host (Section 5).[287,287a]

Other DNA polymerases. **Reverse transcriptases** synthesize DNA using an RNA template strand. They are best known for their function in retroviruses (Chapter 28). The HIV reverse transcriptase is a heterodimer of 51- and 66-kDa subunits. The larger subunit contains a **ribonuclease H** domain.[288–289a] The enzyme is a prime target for drugs such as AZT and others.[290,291] A different reverse transcriptase is found in all eukaryotic cells in **telomerase**, an enzyme essential for replication of chromosome ends. Reverse transcriptases have also been found in rare L1 sequences that are functioning retrotransposons (Section D).[292]

The ~335-residue catalytic subunit of **eukaryotic polymerase β** , which has a DNA repair function, is the simplest known DNA polymerase. The active enzyme (Fig. 27-12B) is as small as 38 kDa. It lacks proofreading and is less accurate than the eukaryotic replicative polymerases listed in Table 27-1.[293–293c] Structural analysis revealed a folding pattern (Class X) related to that of a **nucleotidyltransferase superfamily** that includes enzymes such as terminal deoxynucleotidyltransferase (Chapter 12) and the glutamine synthase adenylytransferase (Fig. 24-7).[294] However, its active site (Fig. 27-13) is similar to that of other DNA polymerases.

DNA polymerase III. This Class C enzyme is the major bacterial polymerase for DNA replication. In its complete holoenzyme form it can synthesize new DNA strands at rates as high as ~1 kilobase s^{-1} without dissociation from a template.[294a] A genetic approach has provided important information about DNA replication.[265,266] A series of temperature-sensitive mutants of *E. coli* unable to carry out DNA synthesis were obtained. From these mutations, genes *dnaA*, *F*, *G*, *I*, *J*, *K*, *L*, *P*, *I*, *X*, and *Z* were identified and located at various points on the chromosome map. Genes *C* and *D* map close together at 89 min, and it now appears that they are one gene. Gene *F* encodes a ribonucleotide reductase (Eq. 16-22). The functions of genes *A*–*E*, *G*, *X*, and *Z* are discussed in the following sections. None of these genes code for DNA polymerase I, but gene *dnaE* was identified as that of DNA polymerase III, which is now known to be the major polymerase in bacterial DNA replication.[265,295] To obtain rapid error-

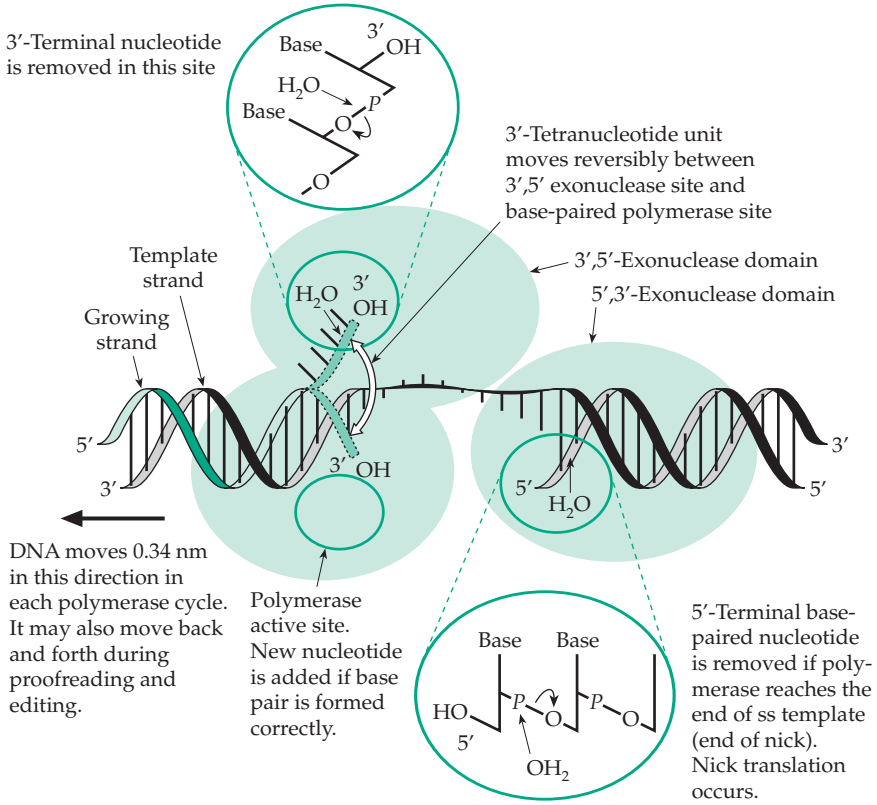


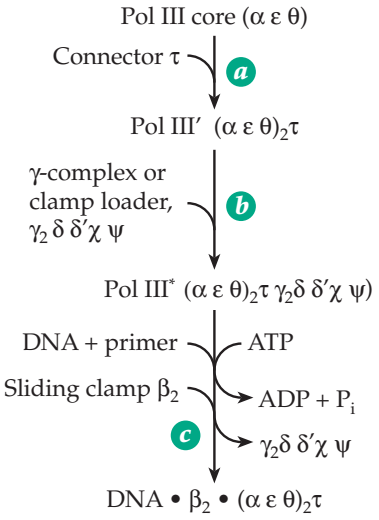
Figure 27-14 Schematic representation of DNA polymerase action on a nicked strand of DNA in which the nick has been enlarged. At the catalytic center new nucleotide units are added at the 3' end of a growing strand. At the 3'-5' exonuclease site the 3' terminal nucleotide may be removed hydrolytically. This will happen to the greatest extent if the nucleotide is poorly paired in the duplex. At the 5'-3' exonuclease site nucleotides are hydrolytically removed from the 5' end of a strand in the chain.[265,267]

free synthesis it must be combined with a number of other subunits to form **polymerase III holoenzyme** (or replisome). Ten different subunits, some as two or more copies, form the holoenzyme.[294a,296–299] Subunits are listed in Table 27-2. The polymerase catalytic center is in the 132-kDa α subunit. The 27.5-kDa ε subunit contains the 3'-5' exonuclease editing activity.[298,300,300a] Mutation in its gene (*dnaQ*) leads to a high spontaneous mutation rate in bacteria. Together with the θ subunit, α and ε form the polymerase III core. This complex has polymerase activity and improved proofreading ability but is still unable to act rapidly, accurately, and processively. Full catalytic activity requires at least the additional β, γ, δ, and τ subunits (Table 27-2). The presence of the τ subunit causes the core complex to dimerize to form Pol III′ (Eq. 27-4, step *a*). Pol III′ can add only about ten nucleotides to a growing DNA chain before it dissociates.[301] The presence of the β_2 dimer, known as the **processivity factor** or **sliding clamp**,[301a,b] is essential for highly processive polymerization. The β protein forms a ring around the duplex DNA and interacts with the polymerase clamping it tightly to the DNA. Putting the β_2 clamp onto the DNA is an ATP-dependent process that first involves binding of the **γ-complex** or **clamp loader** (Eq. 27-4, step *b*), ATP-dependent opening of the β_2 ring, and insertion of the DNA duplex (step *c*).[301a,c] This complex may form a replisome structure that acts simultaneously on the leading and lagging strands (see Fig. 27-19).[302,303] The *Bacillus subtilis* replisome appears to contain two different catalytic (α) subunits, perhaps one for each strand.[303a]

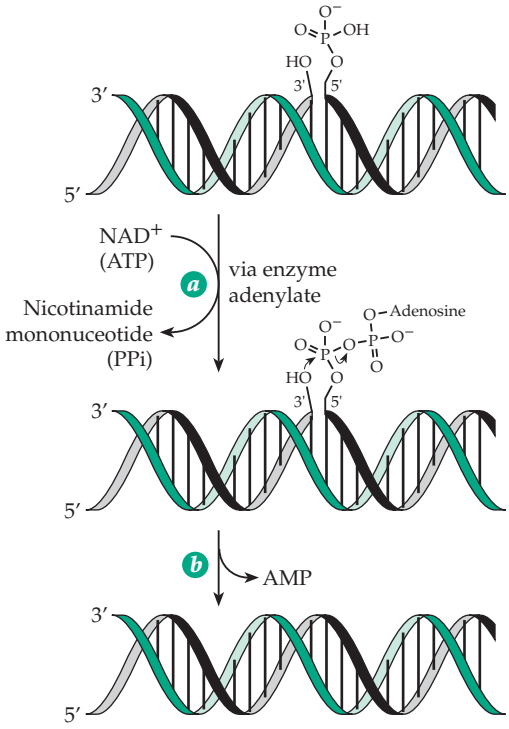
3. Other Replication Proteins

DNA ligases. These enzymes, which are essential to replication, have a specific function of repairing "nicked" DNA.[261,304–307a] Such DNA, as indicated in Eq. 27-5, has a break in one strand and contains a 3'-hydroxyl group and a 5'-phosphate group, which must be rejoined. The ligase from *E. coli* activates the phosphate group in an unusual way by transfer of an adenylyl group from NAD^+, with displacement of nicotinamide mononucleotide (Eq. 27-5, step *a*). The reaction is completed by displacement of AMP as indicated in Eq. 27-5, step *b*. Cells infected by bacteriophage T4 synthesize a virally encoded ligase, which utilizes ATP rather than NAD^+ as the activating reagent. The ~190-kDa mammalian DNA ligase I has been found deficient in some patients with the Bloom syndrome, a condition associated with poor DNA repair and a high incidence of cancer (see also p. 1550).[304]

Single-strand binding proteins. Genetic analysis of replication of the DNA of phage T4 within cells of *E. coli* revealed that at least five genes of the virus are required. One of these, gene 43 specifies the T4 DNA polymerase, while gene 32 codes for a single-strand binding protein, also known as the DNA unwinding, melting, or helix-destabilizing protein. It has a greater affinity for ssDNA than for dsDNA and binds to a length of ssDNA causing unwinding of the double helix and exposure of the purine and pyrimidine bases of the template strand.[308] The protein is required for replication, genetic recombination, and repair of DNA. Similar proteins are encoded in the genomes of many viruses.[309] The 87-residue single-strand binding protein encoded by gene 5 of phage M13 forms a dimer, which completely coats newly synthesized ssDNA preventing the DNA polymerase system of the host



(27-4)



(27-5)

bacteria from converting it into dsDNA. The polynucleotide chain binds into a groove in the protein with one tyrosine intercalated between the DNA bases.[310,311]

The *E. coli* single-strand binding protein, another helix-destabilizing protein that is usually called simply **SSB**, is a tetramer of 18.5-kDa subunits.[265,312,313] It is essential to DNA replication. About 35 nucleotides may bind to each tetramer.[314] The situation is not as clear in eukaryotes where DNA is largely coiled around histones in the nucleosomes. Several single-strand binding proteins have been identified,[315] but the need for SSB proteins in eukaryotic nuclear replication is uncertain.[316] A human mitochondrial SSB resembles that of *E. coli*.[317]

Helicases and topoisomerases. Cells of *E. coli* contain at least 12 DNA-dependent ATPases that cause unwinding of DNA at the expense of hydrolysis of ATP.[318] The activity of these **helicases** is essential to replication, repair, and recombination of DNA in all organisms.[319–320b] The primary replicative helicase of *E. coli*, which unwinds DNA ahead of the replication fork, is encoded by gene *dnaB*. The active form is a hexamer, which exists in at least two conformational states.[321,321b] A segment of ~20 nucleotide units of DNA binds to one hexamer.[322] Helicases have ATPase activity, and the dnaB hexamer contains six ATP-binding sites. However, only three of them may be occupied.[323,324] Many helicases have a hexameric ringlike structure; that of *Bacillus subtilis* is seen clearly in electron micrographs.[325] Three-dimensional structures are known for some.[321a,325a] Although these enzymes may bind to duplex DNA, they also bind to and move along single-stranded DNA in either the 5'→3' or 3'→5' direction. The directionality of a helicase can be determined by annealing two small pieces of ssDNA to the 5' and 3' ends of a longer strand of ssDNA. A 5'→3' helicase will translocate along the long ssDNA and displace the oligonucleotide annealed at the 3' end of the strand, while a 3'→5' helicase will displace the oligonucleotide annealed at the 5' end.[265]

The dnaB protein is a 5'→3' helicase. However, the first helicase identified, the product of the *E. coli rep* gene, is a 3'→5' helicase. It is one of the host proteins needed for the propagation of phages such as ϕX174 and M13. It catalyzes the unwinding of the double-stranded replication forms of these viruses. It binds to a stretch of ~20 nucleotides in a single-stranded region of nicked DNA. The hydrolysis of ATP moves the enzyme along the bound strand in a 3' to 5' direction opening up the DNA at a replication fork. Another *E. coli* 3'→5' helicase is protein *priA* (also called n′), a component of some primasome structures involved in replication of viruses.[318]

The bacteriophage T4 gene 41 protein, a 5'→3' helicase, functions together with the gene 61 primase in replication of that virus.[326–326b] The phage T7 gene 4 protein and virus SV40 large T antigen are also hexameric ringlike helicases. The *E. coli* protein **RecQ** is required for various aspects of recombination and is the prototype of a large group of helicases present in both prokaryotes and eukaryotes.[326c] The bacterial ruvB protein is a hexameric helicase that propels branch migration in Holliday junctions (Fig. 5-28) during genetic recombination,[327–327b] while helicase rho is required for termination of RNA synthesis.[328] Numerous eukaryotic helicases have been identified and purified.[329–331] Helicase DNA2 is needed for DNA replication in nuclear extracts from yeast.[332] Human 3'→5' DNA helicases, members of the RecQ family, are defective in some patients with **Bloom syndrome** and **Werner Syndrome**[332a] (Box 27-A). Bloom syndrome causes growth retardation, immunodeficiency, sensitivity to sunlight, and a predisposition to skin cancers and leukemias.[331a,b] The yeast (*S. cerevisiae*) genome contains genes for 134 different proteins that are probably helicases.[332b] RNA helicases are also known.[333]

A characteristic of all helicases is their ATPase activity, which apparently provides energy for "melting" the DNA. The mechanisms are not clear, but rapid separation of the stacked and hydrogen-bonded base pairs may be impossible without some assistance from an ATP-dependent process. In the case of the rep protein hydrolysis of two molecules of ATP seems to be required to melt one base pair.[323] It isn't clear whether one strand of DNA passes through the center of the oligomeric ring, as shown in Fig. 27-15, or whether both strands pass through. Helicases vary in their amino acid sequences, but they all possess several characteristic **signature sequences** including the Walker A and B motifs found in other ATPases, in

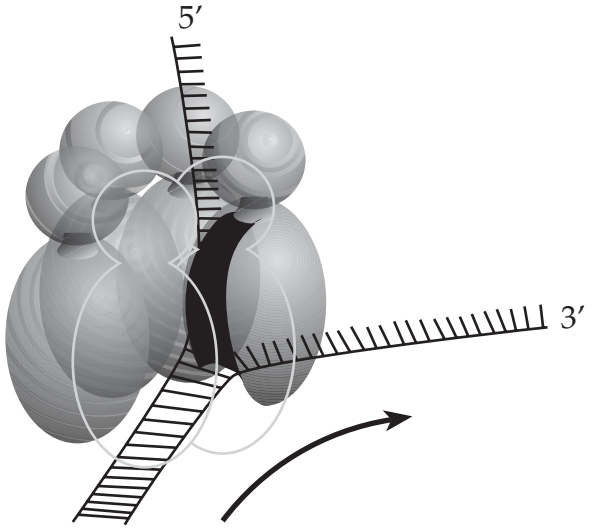
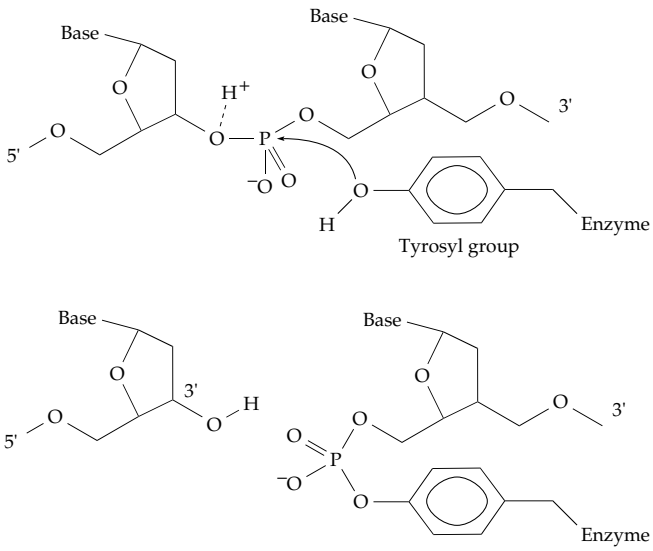


Figure 27-15 Model of DnaB hexamer proposed by Jezewska *et al*.[322] The arrow indicates the direction of movement of the DNA relative to the position of the helicase.

sequences related to the *E. coli* **recA** proteins (Section D) and in synthases (Chapter 18).[328,334]

While helicases can cause the unwinding of linear DNA duplexes, they cannot alter the linking number of covalently closed circular double-stranded DNA. The latter is the function of **topoisomerases**, which have been found in organisms from bacteria to higher plants and animals and may also be encoded by viruses.[265,335–337] There are two basic types of DNA topoisomerase. Those of type I change the linking number in steps of 1. One way that this might be accomplished is for an enzyme to nick one strand in the DNA allowing one of the cut ends to swivel around the unbroken strand, then to reseal the chain. However, it was found that topoisomerases can also cause catenation or decatenation of circular duplex DNA as long as at least one of the reacting DNA molecules is nicked. This observation suggested that a topoisomerase binds to a single-stranded region at a nick and cuts the chain but does not release the ends. This permits either a single strand or a duplex to pass through the broken strand, which is then resealed.

Topoisomerases of type I usually act most rapidly on negatively supercoiled DNA. They relax it by decreasing the number of negatively supercoiled turns one at a time. Negative supercoiling presumably facilitates binding of the enzyme to a single-stranded region by unwinding of the duplex. No ATP or other obvious source of energy is needed by type I topoisomerases. The chain cleavage involves a simple nucleophilic displacement by an –OH group of a tyrosine side chain (Tyr 319 of *E. coli* topoisomerase I),[338] which attacks a phosphorus atom in the DNA chain (Eq. 27-6). The result is covalent attachment of the enzyme to the 5'-end of the cut strand in type IA topoisomerases and at the 3'-end in type IB topoisomerases.[336,337,339] After the passage of the other



(27-6)

TABLE 27-2
Some Proteins of DNA Replication in *E. coli*[a]

Name	Gene	Mass of monomer (kDa)	Map location (min) Fig. 26-4
Polymerase I (also has 3'-5' and 5'-3' exonuclease, and RNase activities)	*polA*	103	87
Polymerase III			
Core			
α subunit (polymerase)	*dnaE*	130	4
ε subunit (3'-5' exonuclease)	*dnaQ*	27.5	5
θ subunit	*hol E*	10	
Sliding clamp, β_2	*dnaN*	37 × 2	83
Connector τ (ATPase)	*dnaX*	71	
Clamp loader (γ complex)			
γ subunit	*dnaX*	47.5 × 2	11
σ subunit		35	
σ′ subunit		33	
χ subunit		15	
ψ subunit		12	
DNA binding proteins			
Single-strand, SSB	*ssb*	18.5 × 4	92
Double-strand, HU			
α subunit		9.5 × 2	
β subunit		9.5 × 2	
Helicases (ATP-dependent)			
Primary replicative	*dnaB*	52 × 6	92
Dna C protein	*dnaC*		99
PriA (n′), primosome	*priA*	76	88
Rep	*rep*	76.4	85
Initiation and priming proteins			
Dna A protein	*dnaA*	52	83
Primase (an RNA polymerase)	*dnaG*	60	67
PriB (n) primosome	*priB*		96
PriC (n″) primosome	*priC*		
DnaT (primosome assembly)	*dnaT*		99
Ribonuclease HI	*rnhA*		
DNA ligase	*lig*	75	52
Topoisomerases			
Type I	*topA*		28
Type II, DNA gyrase ($\alpha_2\beta_2$)			
Subunit α	*gyrA*		97
Subunit β	*gyrB*		90

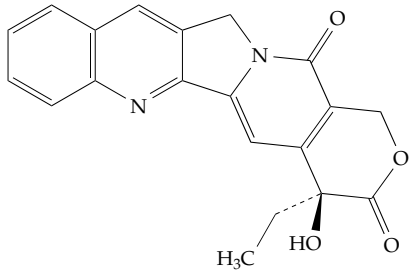
[a] In large part from Kornberg, A., and Baker, T. A. (1992) *DNA Replication*, 2nd ed., Freeman, New York

strand through the gap in the cut strand of a type IB topoisomerase (Fig. 27-16) the free 3'-OH oxygen atom (or 5'-OH) attacks the phosphorus atom in the phosphotyrosine diester to reform the chain and release the enzyme. X-ray diffraction studies show that both the Type IA *E. coli* topoisomerase I and the human type IB topoisomerase[339–341a] are large proteins with holes of appropriate diameter for a DNA double helix. As illustrated in Fig. 27-16B, the protein may open to allow a double helix to enter and occupy a suitable position for cleavage of one chain by the active-site Tyr 319. Topoisomerases are metalloenzymes, usually functioning best with Mg^{2+}. *E. coli* topoisomerase I also contains 3–4 tightly bound Zn^{2+} ions.

Topoisomerases of type II change the linking number by 2 in either the positive or negative direction and hydrolyze ATP in the process. The best known of these is the *E. coli* **DNA gyrase**, an $\alpha_2\beta_2$ dimer of 97-kDa (α) and 90-kDa (β) subunits.[343,344] The enzyme catalyzes the ATP-dependent introduction of negative supercoils into DNA. It also relaxes negatively supercoiled DNA slowly in the absence of ATP. Type II topoisomerases are found in all organisms.[335,345] They are encoded by some bacteriophage such as T4[346] and by plasmids.[347] However, most differ from bacterial DNA gyrase in not coupling DNA supercoiling to ATP hydrolysis. They require ATP but like topoisomerase I cause a relaxation of the supercoiling. Strands of one segment of DNA (called the "gate" or **G-segment**) are cleaved by the enzyme with staggered cuts four base pairs apart. Another segment of DNA (the "transport" or **T-segment**) is then passed through the gate and is thought to be released from a second gate in the complex (Fig. 27-16C).[342,348] The enzyme subunits bind through phosphotyrosine linkages as in Eq. 27-5 to the 5' phospho groups of the two cleaved chains, while the subunits bind ATP and may like tiny muscles twist the DNA.[349] Topoisomerases II are large dimeric proteins. The subunits of yeast topoisomerase II (Fig. 27-16C) are 1200-residue multidomain proteins.[342,350] Mechanisms of DNA cleavage by types I and II topoisomerases appear to be related.[351] However, the ATP-dependent conformational changes involved in a two-gate mechanism are unique to topoisomerases II.[348,348a] The bacterium *Sulfolobus* contains a type I topoisomerase that is called **reverse gyrase** because it utilizes ATP to introduce *positive* supercoils into DNA.[352] This is in contrast to gyrase, a type II topoisomerase that introduces negative supercoils.

Type II topoisomerases are essential and function in replication, DNA repair, transcription, and chromosome segregation at mitosis.[345,349] Yeast with a *top2* mutation dies during mitosis with hopelessly entangled daughter chromosomes.[353] A fluorescent antibody to eukaryotic topoisomerase II binds to chromosomes, probably at the bases of the radial loops present during mitosis,[353] and also to centrosomes[353a] Higher organisms contain more than one topoisomerase II.[354] Their specific functions are uncertain, but one appears to be to unknot entangled chromosomal DNA. In the crowded conditions of a cell nucleus topoisomerase can also cause inadvertent *formation* of knots.[355]

The functional role of topoisomerases of type I is less clear. Staining with fluorescent antibodies to the enzyme has revealed its presence in the transcriptionally active "puffs" of polytene chromosomes (p. 1635)[356] and in centromeres of mitotic cells.[357] A current hypothesis is that in *E. coli* class I topoisomerases act to relax negatively supercoiled strands of DNA behind transcription complexes, while gyrase acts to generate superhelical twists, which favor opening of the duplex ahead of transcription complexes.[354,358] Transcription of a supercoiled rRNA gene *in vitro* is diminished by the selective topoisomerase I inhibitor **camptothecin**, one of a group of antitumor drugs directed against topoisomerase of both types I and II.[349,354,359]



Camptothecin, a topoisomerase I inhibitor

In the autoimmune diseases scleroderma and systematic lupus erythematosus antigens to nuclear proteins or nucleic acids are present in the blood. Many patients with severe scleroderma have an antibody against topoisomerase I.[360]

Primases, initiator proteins, and ribonucleases. The priming segment needed for initiation of DNA replication is either a short segment of RNA or an oligonucleotide containing a mixture of ribonucleotide and deoxyribonucleotide units. The enzyme forming the primer is an RNA polymerase called **primase**. In *E. coli* it is encoded by gene *dna*G.[361–362a] Under some circumstances other RNA polymerases can act as the primase. Bacteriophages and plasmids may also encode primases (Table 27-1). For example, gene 61 of phage T4 (Fig. 26-2) encodes a primase, which together with the T4 helicase forms the priming particle or **primosome**.[363] The phage T7 gene 4 encodes a 63-kDa mutlifunctional protein that is both a primase and a helicase.[364–366b] The primase active site is on the outside of the hexameric ring. Additional proteins repre-

senting products of genes *dna*A and C are also required for initiation of replication in *E. coli*. Several molecules of the dnaA **initiator protein** bind to a specific DNA **origin** sequence and participate in assembling a primosome that also contains the hexameric dnaB helicase and, transiently, protein dnaC.[367–370] Replication of some single-stranded phages, such as ϕX174, in *E. coli* also require the *E. coli* priA, priB, and priC proteins (Table 27-2).[265,371,372] For successful completion of replication chaperonins the products of genes *dna*J and *grp* E are needed[373] as is a ribonuclease that digests the primer segments after they have been used as replication initiators.[374–376]

4. Replication of Bacterial DNA

The basic mechanisms of replication implied in Eq. 27-3 seems to be universal, but several questions had to be asked. "Is replication initiated at a fixed point or points in a chromosome?" and "Does replication occur in one direction only or do two forks form at the point

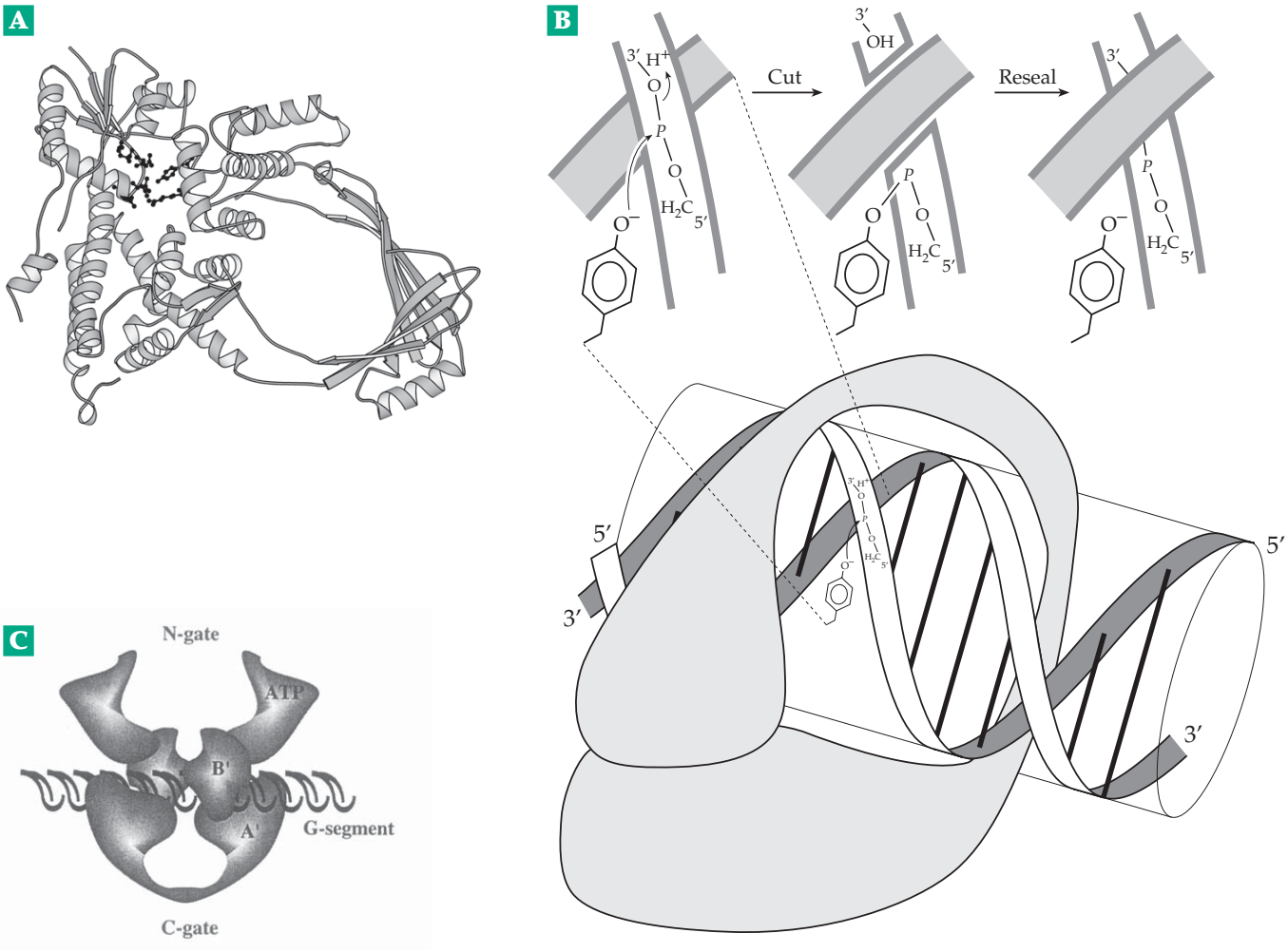


Figure 27-16 (A) Ribbon drawing of a large 67-kDa fragment of the 97-kDa (864-residue) *E. coli* topoisomerase I showing the position of the active-site tyrosine 319 and an adjacent arginine. From Lima *et al.*[341] Courtesy of Alfonso Mondragon. (B) Schematic diagram indicating a way in which topoisomerases of type 1 may pass one strand of DNA through another. The protein is shown binding to a single strand of a DNA duplex. This binding is facilitated by negative supercoiling. The enzyme then cuts the same strand by means of a nucleophilic displacement on a phosphorus atom using a tyrosinate side chain. The other cut end is held noncovalently by the enzyme, while the second strand passes through the gap. Then the gap is resealed by a reversal of the cleavage reaction. From Lima *et al.*[341] (C) Schematic model of a type II topoisomerase bound to a G-segment of DNA. This double helix is cut and another double helical strand, the T-strand, enters the N-gate. The gate then closes, and the central gate opens to allow the T-strand to pass through and exit through the C-gate. The shapes of the three domains are based on crystallographic data for the yeast enzyme. The ATPase, B′, and A′ domains consist of residues 1 to 409, 410 to 660, and 660 to 1200, approximately. From Olland and Wang.[342] Courtesy of James C. Wang. See also Champoux.[340c]

of origin and travel in opposite directions?" To answer these questions both genetic methods and electron microscopy have been employed.

Directions of replication. One technique for establishing the direction of replication in *E. coli* was to insert a λ prophage at the *att* site (Fig. 26-4, 17 min) and DNA from phage Mu-1 at a variety of other sites around the chromosome.[377] Phage Mu-1 was especially useful because it can be integrated at many different locations within the well-mapped genome. Integration within a gene inactivates that gene (an addition mutation) and allows the localization of the Mu prophage. Bacteria were prepared containing both λ and Mu-1 prophage, the latter at various sites. The bacteria were also auxotrophic for certain amino acids. Because of this, replication could be stopped by amino acid starvation. The bacteria usually completed any replication cycle in progress and then stopped. When the missing amino acids were added, replication began, again starting from the replication origin. Bromouracil, which enters DNA in place of thymine, was added at the same time. Consequently, the newly synthesized DNA strands were denser than the parent strands (see Fig. 27-11). After various times of replication the newly formed strands were separated by centrifugation in a CsCl gradient and were tested for hybridization with both λ and Mu-1 DNA. Since the cells did not all begin replication at the same time after addition of amino acids, a variety of lengths of newly replicated DNA were present. Nevertheless, from the observed ratios of Mu-1 DNA to λ DNA for the various strains it was possible to map the progress of replication beginning at an **origin** *ori*C near gene *ilv* at 74 min (Fig. 26-4). Replication was found to progress bidirectionally around the chromosome and to terminate between genes *trp* and *his* at ~25 min.

The use of autoradiographic methods confirmed bidirectional replication. Strains of amino acid auxotrophs with small nucleoside triphosphate pools were used. The addition of amino acids after starvation led to initiation of replication with only a 6-min lag. The cells were labeled with [^{3}H]thymidine, and after the replication forks had moved a short distance from the origin of replication the cells were given a pulse of "super-hot" [^{3}H]thymidine. Using autoradiography it was possible to observe the clearly bidirectional replication forks[378] (Fig. 27-17). Replication in other bacteria is also bidirectional.

Origins of replication. Replication of the *E. coli* chromosome begins and proceeds bidirectionally from its defined origin *ori*C. Replication of linear phage T7 is also bidirectional and begins at a point 17% of the way from one end.[265] In mammalian mitochondrial DNA the origin of replication for the H-strand is in the D-loop but that for the L-strand is 2 / 3 of the way around the circular chromosome within a cluster of tRNA genes (Fig. 18-3).[381,382] The single-stranded circular DNA genomes of Ff, ϕX174, and G4 phage also have distinct origins for initiation of replication to give RF circles.[265]

Most origins have quasi-palindromic nucleotide sequences, perhaps so that DNA can be looped out from the main duplex as is shown in Fig. 27-18A and B. The lengths of *ori* sequences vary, as does the complexity of their possible folding patterns. Plasmids have been constructed which not only contain the *E. coli* origin, but are dependent upon that origin for their own replication.[382a] Study of those plasmids indicate

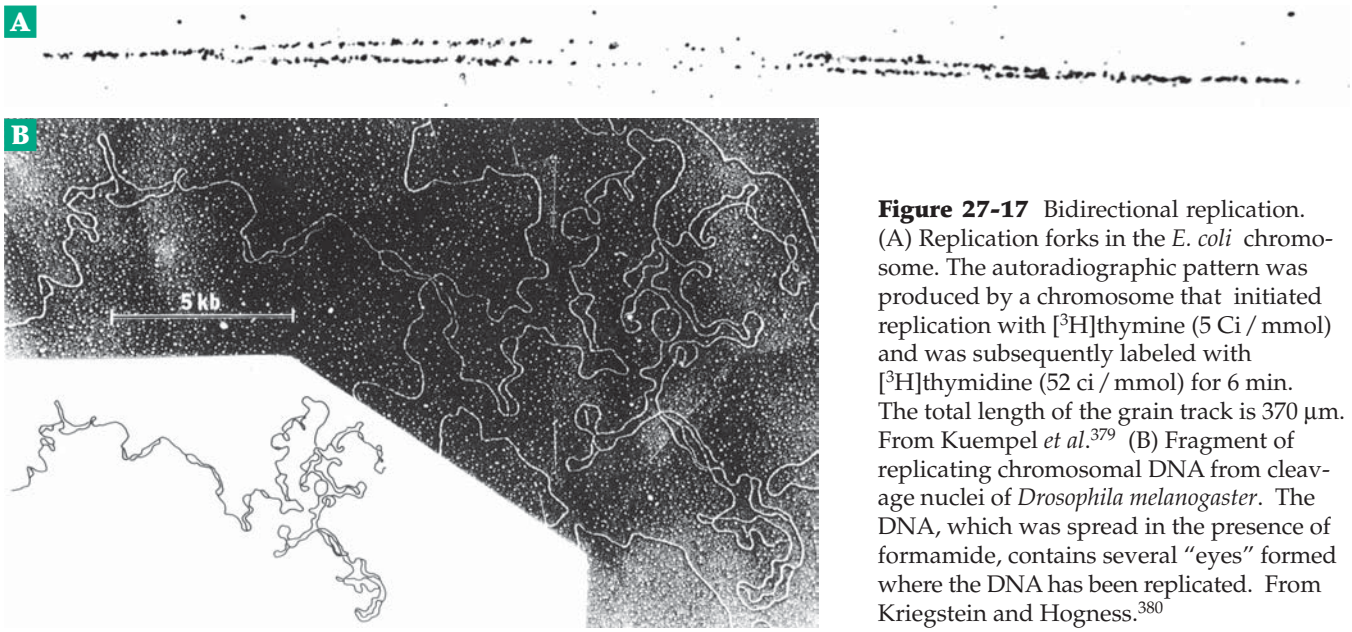


Figure 27-17 Bidirectional replication. (A) Replication forks in the *E. coli* chromosome. The autoradiographic pattern was produced by a chromosome that initiated replication with [^{3}H]thymine (5 Ci / mmol) and was subsequently labeled with [^{3}H]thymidine (52 ci / mmol) for 6 min. The total length of the grain track is 370 μm. From Kuempel *et al*.[379] (B) Fragment of replicating chromosomal DNA from cleavage nuclei of *Drosophila melanogaster*. The DNA, which was spread in the presence of formamide, contains several "eyes" formed where the DNA has been replicated. From Kriegstein and Hogness.[380]

that a 245-bp *ori*C sequence is essential.[265,383,383a] This sequence, which is shown in Fig. 27-18, contains several repeated oligonucleotides including 11 GATC sequences, which are sites of adenine N^6-methylation (see Section B, 6), and four "9-mers" (commonly known as dnaA "boxes") with the consensus sequence

$$\mathrm{TTAT}\begin{smallmatrix}\mathrm{A}\\\mathrm{C}\end{smallmatrix}\mathrm{CA}\begin{smallmatrix}\mathrm{A}\\\mathrm{C}\end{smallmatrix}\mathrm{C}$$

These appear in both 5' to 3' and 3' to 5' orientations, allowing them to form two base paired "stems." In addition, there are three direct repeats of a 13-residue consensus sequence 5'-GATCTNTTNTTTT (shaded in green in Fig. 27-18), which form an AT rich duplex. Other bacterial replication origins often follow a pattern similar to that in Fig. 27-18.[384] However, the origin for *Mycoplasma genitalium* has been hard to detect.[385] Replication origins of archaea have characteristics similar to those of bacteria and of organelles.[385a]

Priming and initiation of DNA synthesis. The first step in initiating a new round of replication and a new cell cycle in *E. coli* appears to be the binding of the initiator protein, dnaA, to the 9-mers in the negatively supercoiled origin.[295,383,386–388] The resulting complex is visible in an electron microscope and may consist of a core of up to 20–40 molecules of dnaA protein with the DNA wrapped around them (Fig. 27-19). ATP is also required and is bound to the protein where it hydrolyzes slowly. Similar initiator proteins are found

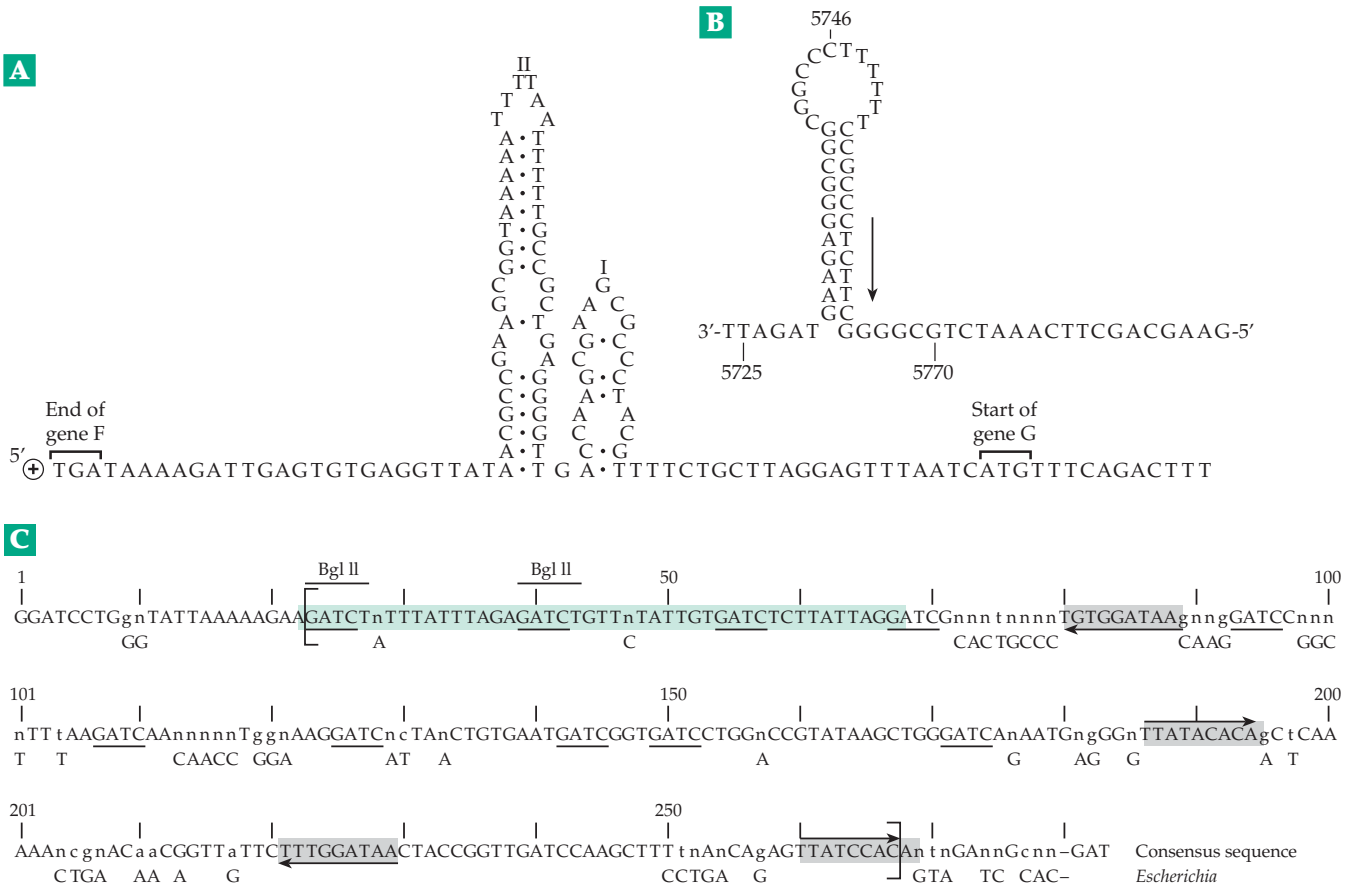


Figure 27-18 Nucleotide sequences and proposed secondary structures for some origins of replication. (A) Bacteriophage ϕX174. The ends of the neighboring genes F and G are shown. From Kornberg and Baker.[265] (B) Human mitochondrial DNA; origin of replication of the L strand. The sequence shown is for the H strand. The arrow indicates the direction of replication. From Hixson *et al*.[381] (C) Bacterial *ori* region. The consensus sequence of origins of replication that will function in *E. coli*. Derived from sequences from *E. coli, Salmonella typhimurium, Enterobacter aerogenes, Klebsiella pneumoniae,* and *Erwinia carotovora.* In the consensus sequence a large capital letter means that the same nucleotide is present in all six origins; a lowercase letter means that the same nucleotide is present in three of the five origins; n means that any of the four nucleotides may be present; – represents a deletion, G being present here in the *Enterobacter* origin. Underlined are *dam* methylation sites. Green shaded sequences, three 13-mers in a region that undergoes easy opening of the duplex for insertion of replication proteins; gray boxes with arrows, four 9-mers that are specific sites for binding of replication protein A. Brackets enclose a 245-bp minimum origin. From Kornberg and Baker.[265] Data from Zyskind *et al*.[390]

in many bacteria and are encoded by some viruses and plasmids. In archaea, as well as in humans, the initiator proteins seem to combine functions of the *E. coli* dnaA protein with those of SSB.[391,391a]

Following the binding of the dnaA protein the hexameric helicase dnaB (Fig. 27-15) is loaded onto the adjacent DNA in the region of the 13-residue repeated sequences, which are labeled 1, 2, and 3 in Fig. 27-19. As shown in this figure introduction of the helicase is assisted by protein dnaC, which forms a complex with the helicase. Additional binding of protein HU and a temperature of >30°C are essential for tight binding of the proteins. Although the dnaC protein is needed for formation of the prepriming complex, it dissociates after the dnaB protein becomes firmly bound. One dnaB hexamer binds to each single strand of the DNA duplex with opposite orientations (Fig. 27-19).[383a] With the dnaB helicase in place on each strand, this ATP-driven enzyme processes along the DNA, unwinding the duplex in both di-rections. (Perhaps it may be more accurate to say that the *DNA moves through the helicase*.) The resulting "bubble" is held open by binding of SSB tetramers. The primase (dnaG protein) then binds adjacent to the helicase and synthesizes the RNA primer along each strand of DNA. As the helicase processes to the right in the fork shown in Fig. 27-19, the complex with the dnaA protein dissociates, permitting primer synthesis into the origin region. Alternatively, RNA polymerase (Chapter 28) can prime replication by initiating transcription on both strands of the DNA. Suitable promoters are present and oriented in opposite directions on the two strands.[392]

It has long been postulated that the bacterial chromosome is attached to the plasma membrane. At least one such attachment site may be at or near the origin of replication.[393] Furthermore, the exchange of ADP for ATP in the dnaA protein is catalyzed specifically by cardiolipin and phosphatidylglycerol containing the unsaturated oleic acid.[386,393] Inositol polyphosphates may also play a role.[394]

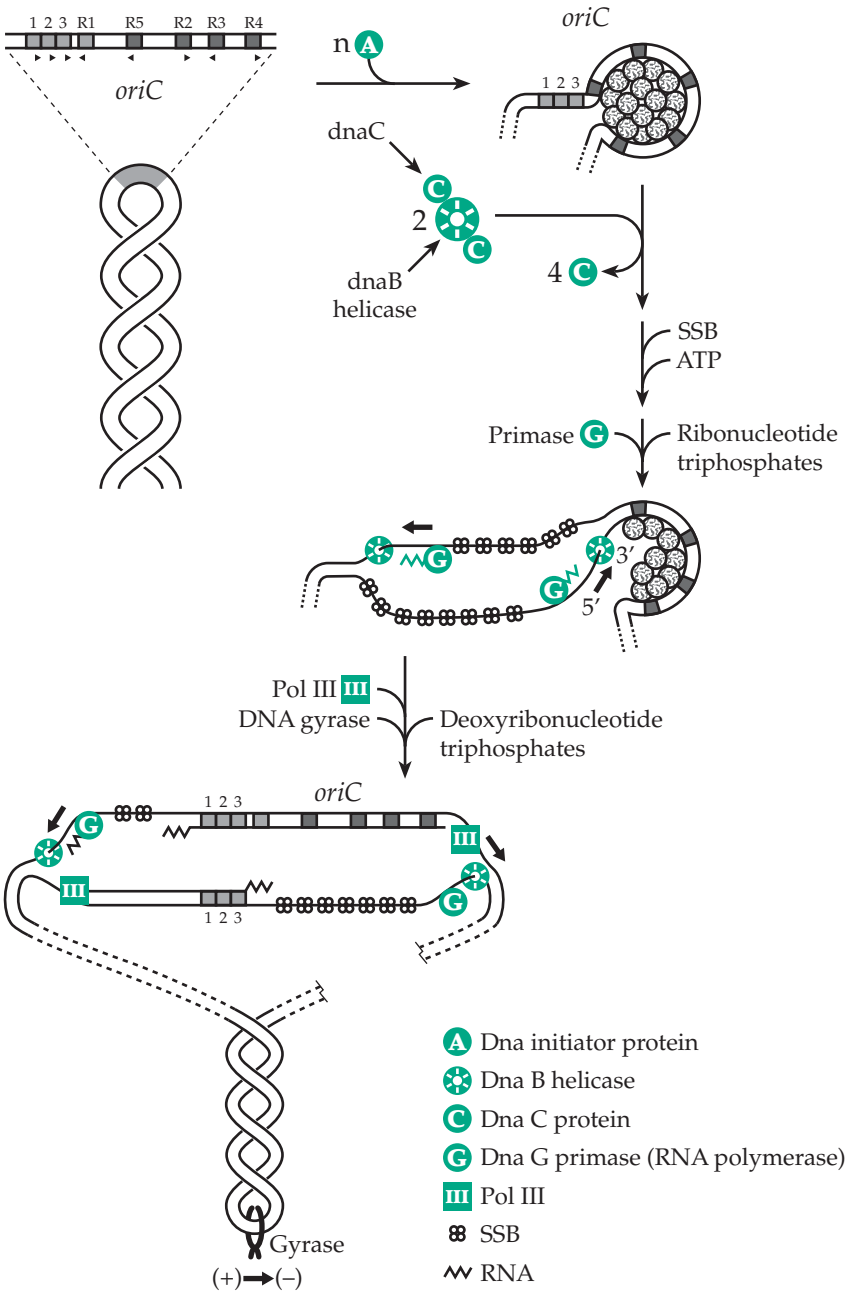


Figure 27-19 Hypothetical scheme for initiation of bidirectional replication in *E. coli*. The closed boxes R1 through R5 represent the 9-residue recognition sequences for the *E. coli* dnaA protein. The open boxes 1, 2, and 3 represent the three 13-residue repeats, possible sites for binding of the dnaB-dnaC protein complex. From McMacken *et al.*[383] Redrawn in simplified form.

Elongation of DNA chains.

DNA polymerase III in its holoenzyme form is the major polymerase for DNA replication. It elongates the primer chains rapidly and processively leaving only very small gaps at the ends of single-stranded regions. The rate of elongation, which is ~3 nucleotides / s for 8 kb

*ori*C plasmids, may be determined by the rate of action of the dnaB helicase. A completely unwound *ori*C plasmid, bound to SSB, undergoes primer elongation ~ 10 times faster.[383] However, the rate for the intact *E. coli* replisome is nearly 1000 nucleotide s^{-1} with a rate of misincorporation of only one in 10^9 nucleotides.[394a]

Small *ori*C plasmids need to be primed at only one location, but the large bacterial chromosome must undergo priming at many sites on the lagging strand to permit DNA polymerase III to act on that strand with formation of the Okazaki fragments. The DNA polymerase complex may be a dimer that works on both strands at once. The lagging strand may be looped out to allow it to lie parallel with the leading strand (Fig. 27-20). The appearance of the electron micrograph in Fig. 27-20B supports this suggestion. However, the manner in which the lagging strand can be shifted to bring the next primed initiation site to the replication complex is not clear. DNA polymerase III holoenzyme itself may be organized as an asymmetric oligomer[395] that operates on both strands in a complex such as is shown in Fig. 27-20. The primase, either the dnaG protein[265,383] or the primosome used by phage ϕX174 (Section 5),[396] may synthesize the RNA primers on the lagging strand.

Termination of replication. As each Okazaki fragment is completed along the template for the lagging strand, the RNA primer piece is digested out, replaced by DNA, and the nick sealed by action of DNA ligase. Ribonuclease H, which is found in both bacteria and eukaryotic cells, specifically degrades the RNA component in these RNA–DNA hybrid regions.[396a] In bacteria another mechanism for primer removal is available. The 5'- 3' exonuclease activity of DNA polymerase I will cut out the RNA segment, while the 5'- 3' polymerase activity of the same enzyme will fill the gap.

Replication of *ori*C plasmids may occur by simply allowing replication of the leading strands at both replication forks to continue all the way around the circle.[397] However, in *E. coli* bidirectional replication continues only until the two replication forks converge. This can occur anywhere between two **terminators**, T1 and T2, located at 28.1 min and 35.6 min. The terminators slow replication in the counterclockwise and clockwise directions, respectively. A gene (*tus*) near T1 encodes a **terminator utilization substance**, a DNA-binding protein that associates with T1 and T2 and causes termination.[398–401] Another problem may be the separating of catenated DNA circles by action of a topoisomerase. Finally, it is essential to **partition** the original chromosome and its replica, one to each daughter cell. This requires at least three other gene products including one large 170-kDa protein.[402]

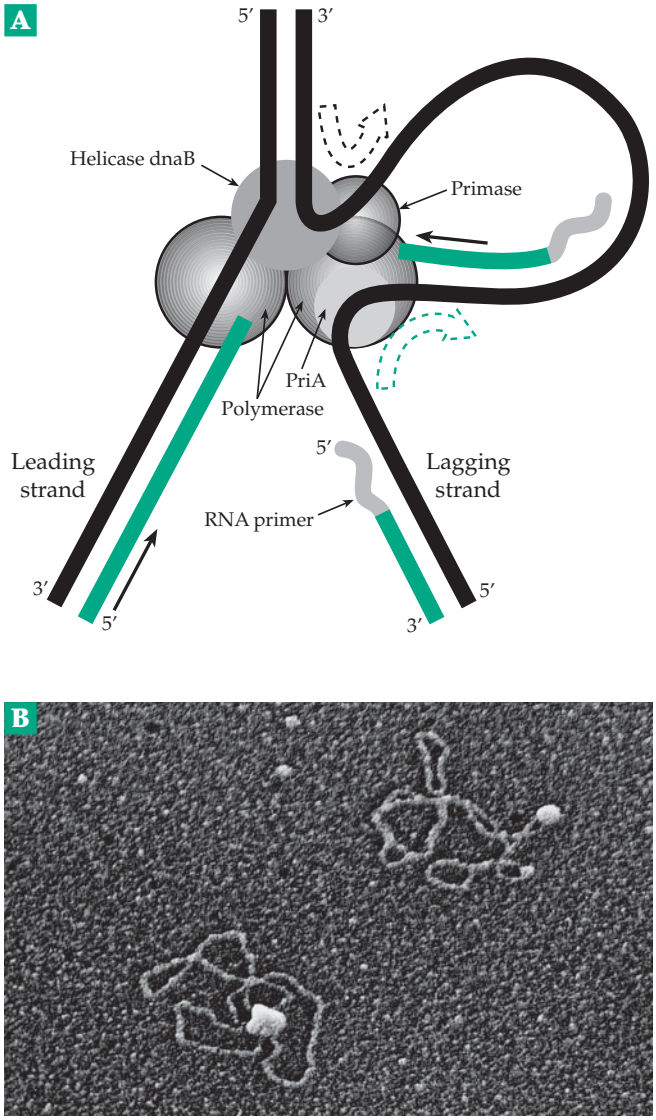


Figure 27-20 (A) Hypothetical replisome for concurrent replication of leading and lagging strands by a dimeric polymerase associated with helicase dnaB and a primosome. Open arrows indicate directions of movement of DNA, which is forming a loop as the polymerase fills a gap to complete an Okazaki fragment. The primase will then form a new primer and a new loop. From Kornberg and Baker.[265] (B) Electron micrograph of the primosome bound to covalently closed ϕX174 duplex replicative form. These enzymatically synthesized duplexes invariably contain a single primosome with one or two associated small DNA loops. From A. Kornberg in Hubscher and Spadari,[266] pp. 9,10.

5. The Replication of Viral DNA

The replication of viral DNA usually depends upon the genes of both the host organism and the virus. For example, *ts* mutations in the *E. coli* genes *dnaB*, *D*, *E*, *F*, and *G* lead to a loss in ability to support

growth of phage λ as well as loss of ability to reproduce under conditions where the *ts* gene products are inactivated. However, the phage can replicate in *E. coli* with mutated genes *dnaA* and *C* because phage λ encodes its own initiator proteins by genes *O* and *P* (marked on the gene map in Fig. 26-4). In addition to these two proteins, seven *E. coli* proteins are required to initiate replication at the lambda origin *ori* λ and to complete replication.[403–407] The *E. coli* dnaB helicase and the dnaC protein are needed, as in Fig. 27-19. The chaperonins dnaJ, dnaK, and GrpE are also necessary for replication of phage λ and other viruses.[408,409] As we have seen (Section 2) many viruses contain genes specifying their own DNA polymerases and primases, which function in cooperation with host proteins. For many dsDNA viruses the origins of replication, priming reactions, and chain elongation processes closely resemble those of *E. coli*.[363,410]

In contrast, the first step in replication of the filamentous Ff viruses (f1, fd, and M13) or the small icosahedral ϕX174 or G4 is conversion of the single-stranded closed circular DNA molecules of the infecting virus particles into circular double-stranded **replicative forms** (RF).[265,411,412] This occurs as the DNA enters the bacterial cell and is accomplished entirely by the enzymes of the host cell. Phage G4 DNA, whose replication has the simplest known requirements, contains a tight hairpin region at its origin. The rest of the DNA must be coated by SSB for replication to occur. The hairpin resists melting and serves as a binding site for *E. coli* primase. This is the only known case in which no other priming proteins are needed.[265] Primase synthesizes up to a 29-ribonucleotide primer after which DNA polymerase III holoenzyme copies the rest of the chain.

Replication of the closely related ϕX174 is more complex. It requires assembly of a **primosome** made up of at least seven host proteins[265,371,413]: dnaB, dnaC, primase (dnaG), and proteins priB, A, C (n, n′, n″), and dnaT (i), Table 27-2. The 76-kDa helicase priA (n′) may locate the **primosome assembly sequences**, which are ~70 nucleotides in length,[414] and displace SSB from them. These sequences can adopt secondary structures with a pair of hairpin loops.[265] Kornberg and associates suggested that the same kind of primosome formed at these sites in the ϕX DNA may participate in replication of the lagging strand of the chromosomal DNA. If so, helicase priA presumably functions in the replisome on one strand and dnaB helicase on the other as depicted in Fig. 27-20.

In the second and third stages of replication ϕX174 RF molecules are themselves replicated and are then used for synthesis of new viral (+) strands. At both stages a virally encoded **gene A protein**, which has endonuclease activity, nicks the duplex. Cutting the (+) strand it leaves a free 3'-OH on DNA residue 4305, while the 5'-phospho group of residue 4306 becomes covalently attached to a tyrosyl residue in the A protein.[411,415] The free 3'-OH serves as the primer for a **rolling-circle synthesis** (Eq. 27-7).[412,416,416a] As a new viral strand is synthesized along the complementary (-) strand as a template, the original viral DNA (+) strand is displaced (Eq. 27-7) as a single-stranded tail.

(27-7)

A strand complementary to the single-stranded tail is then formed in segments. A complete turn of the circle produces a viral strand twice the normal length. Cleavage by the endonuclease activity of the gene A protein and closure of the circle completes the replication. The displaced (+) strand can be cut off and either incorporated into a progeny phage or converted into another RF circle. The A protein, attached to the 5'-terminus of the (+) strand, is involved in either case. It can participate in repeated sequences of initiation and termination of viral (+) strand synthesis.[265,411] Once double-stranded circles are formed, they undergo several replications to give additional RF circles, which serve as templates for the synthesis of many single strands of viral (+) DNA, which are incorporated into the mature viruses. This synthesis of additional RF circles requires transcription of some viral RF genes.

In the final stage of replication the single-stranded (+) chains formed by the rolling-circle mechanism are packaged into phage particles. The gene 5 single-strand binding protein of M13 coats the DNA chains as they are formed, evidently preventing their conversion to RF circles. In the case of ϕX174 the new single-stranded DNA circles are packaged as they are synthesized (Fig. 7-28) to form complete icosahedral virus particles. See Kornberg and Baker[265] for details about these and many other virus replication systems.

Replication of the larger tailed viruses, which have many genes, is complex and varied. The lytic phage λ resembles the smaller viruses in using the host replication enzymes.[265,417,418] In the final stages a rolling-circle mechanism is utilized to form **concatemers** consisting of linear DNA duplexes with numerous successive copies of the viral DNA. The ssDNA that is formed in the rolling circle is converted to dsDNA as it is formed. Finally a **terminase** cleaves the DNA at specific *cos* sites, using staggered cuts, to form cohesive ends.[417,419–420a] However, there are uncertainties.[421] The linear dsDNA enters an empty preformed procapsid, apparently pumped in an ATP-dependent fashion, perhaps by a rotating portal ring.[421a]

6. Packaging of Viral Genomes

The construction of intact virus particles from the genomic DNA and protein subunits is often a complex process. It is simplest for the small filamentous ssDNA viruses (Fig. 7-7). The subunits are synthesized as soluble proteins, which enter the cell membrane, then lose their leader sequences. As the viral DNA coated by the viral gene 5 ssDNA-binding protein enters the membrane, the binding protein is replaced by the coat subunits.[422–424]

The process is somewhat more complex for the icosahedral viruses. In the ϕX, G4, α3 family the icosahedral procapsid is constructed with the aid of both internal and external **scaffolding proteins**[425,426] as is illustrated in Fig. 7-28. In phage that replicate via concatameric dsDNA the terminase that cleaves the DNA also interacts in a precise way with the packaging apparatus of the prohead.[427–429] For the tailed phage the ring-shaped oligomeric head–tail connector (Fig. 7-29), together with an ATPase, may function as a rotatory pump to feed the DNA into the prohead.[430–430c] This has been demonstrated for phage ϕ29.[430a–c] In some cases the terminase produces new phage DNA of unit size, but in other cases, e.g., with phage T4, the DNA may be cut more randomly when the head is full or when another piece is needed to fill it.[431] After the DNA is packaged the virus capsids usually expand and become stronger.[432,432a]

7. Plasmids

Most bacteria contain plasmids which are self-replicating but stably maintained at well-defined numbers of copies per cell.[265,433] They are usually not essential to the cell but may carry traits such as antibiotic resistance or toxin formation that benefit the bacterial host. A plasmid always carries in its DNA an orign (*ori*) of replication and a gene, usually designated *rep*, for an initiator protein. It usually encodes other proteins as well but may depend largely on host proteins for replication. Plasmids may use the *oriC* copied from the bacterial host's DNA, the origin from phage λ, yeast autonomously replicating sequences (ARSs, Section 10), or other origins. Replication of the small 6.6-kb plasmid ColE1, which is present at ~20 copies per cell, depends entirely on the host-cell replication machinery.[265] However, the control of copy number depends upon synthesis of **antisense RNA** and its reaction with the plasmid DNA (see also Chapter 28, p. 1615).[434,435] Similar copy number control is used by the larger ~100-kb resistance factor **R plasmids**.[436] Some plasmids use replication systems very similar to those of viruses such as ϕX174, often using rolling-circle replication.[437] However, the plasmids lack the proteins for virus coat formation and maturation.[438]

The F factor plasmids, discussed in Chapter 26, are large 100-kb circular DNA molecules containing ~60 genes, about 20 of which encode proteins involving transfer of DNA into another bacterial cell (Fig. 26-3).[265,439,440] F plasmids display strict copy number control with only 1–2 copies per host chromosome. The controls lie in a region known as the partition locus, which resembles regions of the host chromosomes that are involved in partition of the bacterial genome. They have repetitive sequences suggesting a similarity to centromeres of eukaryotic chromosomes.[441]

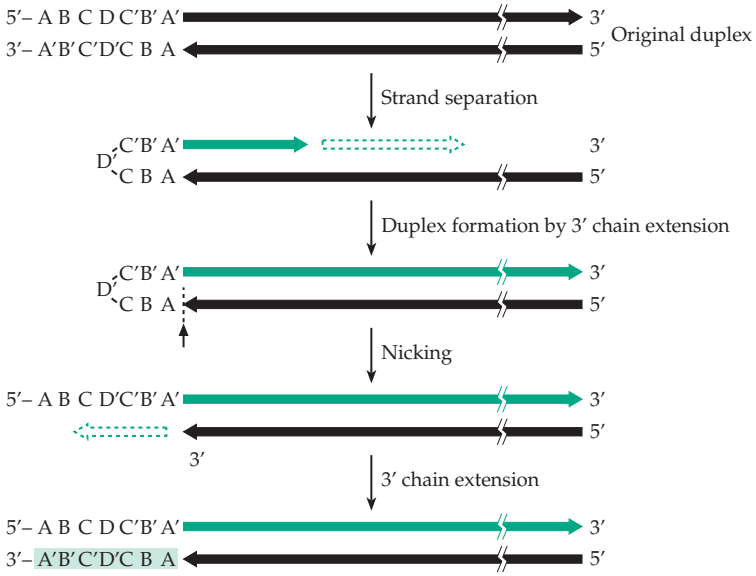
8. Chromosome Ends

The T-odd bacteriophages T1, T3, T5, and T7 are medium-sized phage with linear duplex DNA genomes. Replication of linear DNA in these and in many other genomes presents a problem. Even if the RNA primer segment is made at the very 3' end of the template strand, there will be a gap in the final replicated strand when the primer is digested out. Since there is no known enzyme that will add to the 3' end of a chain, this gap will remain unfilled. The problem is solved by **terminal redundancy**, the presence of a common 260-nucleotide

sequence at both ends. Several daughter DNA molecules with gaps at the 5' ends can be joined by their cohesive ends to form a long **concatamer**. DNA polymerase I fills any gaps, and the chains can then be ligated and cleaved at different points to generate complete 5' ends.[442]

Another mechanism, which is utilized by some single-stranded parvoviruses,[416,443] obviates the need for an RNA primer by use of a palindromic sequence to form a hairpin loop (Eq. 27-8).

Yet another solution to this problem is used by some viruses. **Phage ϕ29** of *Bacillus subtilis* primes the replication of its 19,285 bp dsDNA at both ends by a **terminal protein,** which is linked covalently through its Ser 232 –OH group to dAMP.[443a] The 3'–OH of the deoxyadenosyl group primes the DNA replication. In a similar fashion replication of the eukaryotic **adenoviruses**, whose genome is a 35- to 36-kb linear DNA duplex, starts at the ends and is primed by one residue of dCMP covalently attached through a 5'-phosphodiester linkage to a serine side chain in a 80-kDa **preterminal protein**. It substitutes for the RNA oligonucleotides that prime most DNA synthesis.[444–447] The dCMP pairs with guanine at the 3' terminus of the template strand and provides the initiating 3'-OH group. During the replication the



(27-8)

preterminal protein is cleaved to the 55-kDa **terminal protein,** which remains covalently attached, one molecule at the 5' end of each strand. The genome can be replicated *in vitro* by five proteins: the virally encoded preterminal protein, DNA polymerase, DNA-binding protein, and two cellular transcription factors that bind in the adenovirus origin region.

The chromosome end problem is solved in another way in eukaryotes. As discussed in Section B,1, **telomeres** (chromosome ends) contain repeated sequences of variable length. One DNA strand is always G-rich. For example, in human cells the sequence 5'-$(TTAGGG)_n$–3', where n may be ~20, oc- curs at the 3' ends of the G-rich strands. The other strand, whose 5' end is at the telomere end, is C-rich and has the complementary sequence 3'-$(AACTCCC)_n$–5'. The 3' end of the G-rich strand is always longer by 12–16 nucleotides than the end of the C-rich strand. This 3' extension may fold back to form non-Watson–Crick structures that apparently involve G-quartets (Chapter 5).[447a,b] The shorter 5' end is thought to result from the need for a short RNA primer during replication. As shown in Fig. 27-21, when replicated in the normal fashion the full G-rich leading strand will be formed, but the C-rich lagging strand will be 8–12 nucleotides short, when the RNA primer is digested away. A result of this is that human somatic cells gradually lose telomeric repeats. However, in tumor cells, germline cells, and unicellular organisms the enzyme **telomerase** prevents this telomere loss.[448–449a]

Telomerase is a reverse transcriptase that copies the DNA sequence of the telomeric repeats from a small **guide RNA** that is part of the enzyme. The first telomerase studied was the relatively abundant enzyme from *Tetrahymena*. It contains a 159-nucleotide RNA with the sequence 5'-CAACCCCAA-3' at positions 43–51. This sequence is complementary to the 5'-TTGGGG-3' repeat sequence of the *Tetrahymena* telomeres.[450,451] A 127-kDa human protein contains a similar guide RNA with the sequence 5'-CUAACCCUACC-3', which is complementary to the human telomere repeat sequence as is illustrated in Fig. 27-21.[452–454] Telomerases[455,456] evidently allow the cell to elongate the telomere 5'-ends using the

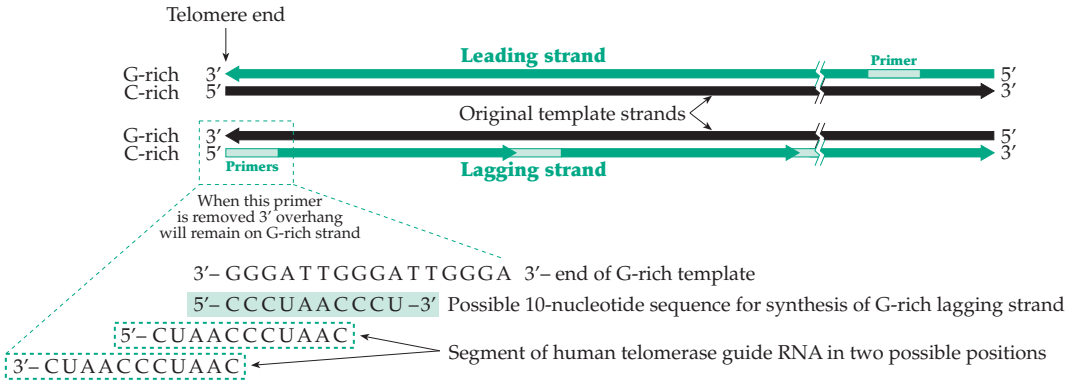


Figure 27-21 Aspects of telomere synthesis. The end of the chromosome and the 5' end of the C-rich strand is at the left. This is the template for replication of the leading G-rich strand (green). The primer lies far back in the chromosome. The C-rich strand is replicated in segments from several RNA oligonucleotide primers, one of which lies at the 5'-terminus. This first primer is removed by RNase activity leaving a 12–16 nucleotide 3'-overhang. The telomerase guide RNA can hybridize with the 3'-end of the G-rich strands providing a template that allows additional growth of the G-rich strand and extension of the C-rich strand also in the next replication.

guide sequence and the reverse transcriptase activity of the telomerase. Any number of additional repeats may be added to the 5' ends. The shortened 3' ends can also be lengthened in the next round of replication.

The control of telomerase must be important. The enzyme is active in early embryonic cells and some stem cells. However, most normal cells have little or no telomerase activity and lose telomere length throughout their lifespan with eventual growth arrest.[448,454,457] On the other hand, excessive telomerase activity may induce cancer.[451,458,459] Certain mutations in the telomerase guide RNA can cause greatly increased telomerase activity.[459] The control of telomerase is still poorly understood but involves specific telomere-binding proteins.[448,454,460–460b]

9. Mitochondrial and Chloroplast DNA

Replication of the ~16-kb mammalian mtDNA begins with RNA priming within a small **displacement loop** or D-loop. One daughter strand, the heavier or H strand, starts to grow on the primer. As it does, the parental H strand is displaced and the D-loop is enlarged. The H-strand grows until ~70% of the parental H strand has been displaced and the L-strand origin (Fig. 18-3) is uncovered. Then a new light L strand is laid down to form the second daughter duplex. The rate of formation of the new L strand is only 10 nucleotides / s, an hour being required to complete the process. The DNA formed is initially relaxed, another 40 min being needed to introduce the 100 superhelical turns present in the finished chromosome.[461]

The kinetoplast DNA of trypanosomes (Fig. 5-16) consists of thousands of catenated circular DNA molecules. Among these the smaller minicircles always contain the sequence GGGGTTGGTGTA at their origins of replication. The minicircles are individually removed from the mass prior to replication. The two progeny circles are then both recatenated into the mass.[462]

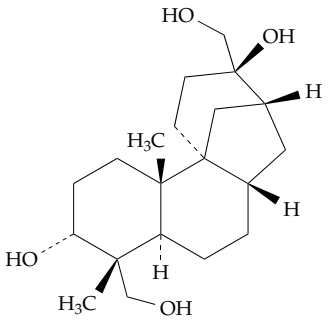
Chloroplasts contain large 120- to 169-kb circular genomes encoding about 100 proteins (Chapter 23). A characteristic feature of most chloroplast DNA is the presence of long inverted repeat sequences (10,058 bp in the liverwort, 25,339 bp in tobacco).[463,464] These are separated by 19,813 and 81,095 bp single copy regions in the liverwort and by similar sized regions in tobacco. Plastid DNA exists as a mixture of monomeric molecules with smaller amounts of dimers, trimers, and tetramers.[464]

Ethidium bromide inhibits the replication of chloroplast DNA and causes partial degradation of existing DNA in chloroplasts without interfering with replication of DNA in the nucleus. The effect is similar to that of the same drug on mitochondrial DNA. However, cells of *Chlamydomonas* treated with ethidium bromide are able later to regenerate their chloroplast DNA. This result has been interpreted to mean that there may be one or a few "master copies" of chloroplast DNA in specially protected locations. The result should also be considered in relationship to the following observation. Although nuclear and organelle DNA molecules replicate at different times in the cell cycle, constant proportions of the organelle and nuclear DNA tend to be maintained. Thus, there must be some kind of control mechanism leading to a coupling of DNA replication in nuclei, mitochondria, and chloroplasts.[465]

10. Replication of Eukaryotic Nuclear and Viral DNA

Replication in eukaryotes is similar in many ways to that in bacteria.[284,466–467a] However, the ~10^6 kb of DNA in a typical eukaryotic genome is divided into many **replicons**, segments of DNA 30–150 kb in length, each having its own origin of replication. In the relatively small ~14-Mb yeast genome there are ~400 replicons,[468,468a] but in mammalian DNA and also in plant DNA[464] there are probably thousands. DNA synthesis is initiated at different times during the S-phase of the cell cycle at the various origins in an ordered pattern.[469,469a,b] An important unanswered question is how the cell is able to replicate all segments of all of the chromosomes just once before entering mitosis.

Many of the proteins of eukaryotic replication are closely related in sequences and functions to those of bacteria. There are initiator proteins analogous to *E. coli* dnaA (Fig. 27-19). The DNA polymerases have been discussed in Section C, 2 (see Table 27-1). Eukaryotic polymerases α and δ and possibly ε are essential for replication.[470,471] Polymerase α, which is inhibited specifically by the fungal metabolite **aphidocolin**,[472] is a complex of a ~170-kDa DNA polymerase core, an RNA-synthesizing primase consisting



Aphidocolin

of 58- and 49-kDa subunits, and a 70-kDa subunit of uncertain function.[473–474a] The complex makes an RNA–DNA primer consisting of ~10 nucleotides of RNA and ~30 of DNA.[475] This pol α/primer is replaced early in replication by the highly processive polymerase δ and, perhaps, under some conditions by polymerase ε. The ringlike processivity factor or "clamp" that is provided in *E. coli* by protein dnaB is called the **proliferating cell nuclease antigen** (**PCNA**) in eukaryotes.[476] It is loaded onto the DNA by a **clamp loader**, the 5-subunit replication factor C (RFC).[476a] As in *E. coli* (Fig. 27-19) an SSB type protein known as **replication protein A** (**RPA**) is also essential.[467,477] PCNA is not only essential to eukaryotic replication but is also required for recombination and repair.[476] The polα/primer primes leading strand synthesis initially but then switches to replication origins on the lagging strand where, together with other proteins, it primes the formation of the Okazaki fragments. Polα/primer is also a logical participant in the control of the initiation of the S-phase of the cell cycle.[478]

Eukaryotic viruses. Investigation of viruses provided the first insights into eukaryotic DNA replication. Most of the factors needed for replication of the DNA of adenoviruses, simian virus 40 (SV40), and polyomavirus[447] within animal cells are supplied by the host. Replication of the 5-kb SV40 DNA, whose DNA forms typical nucleosomes (Fig. 27-3), appears to be an excellent model for eukaryotic replication in general.[479–482] The single SV40 origin of replication is a 64-bp sequence containing the 5-bp sequence GAGGC four times as pairs of inverted repeats. These are recognized by the 95-kDa virally encoded initiation protein which also has helicase activity and is known as the **T antigen**.[483,483a] A nearby 17-bp sequence containing only AT pairs is presumably the region of entry of the host cell's polymerase α/primer. Single-stranded DNA regions are coated with the replication protein A. After the primer is formed, the RFC complex loads the sliding clamp PCNA, and polymerase α is replaced by polymerase δ on both leading and lagging DNA strands permitting highly processive bidirectional chain elongation. Topoisomerase activity is required to decatenate the replicated chromosomes. Since SV40 DNA forms typical nucleosome (Fig. 27-3), its replication is thought to mimic chromosomal replication quite closely. The more complex herpes simplex virus HSV-1 has a 153-kb genome, a linear DNA duplex. It has ~75 genes and encodes its own DNA polymerase, origin-binding protein, SSB, and other proteins needed for replication within eukaryotic cells.[484,485]

Artificial chromosomes. Another approach to understanding eukaryotic replication, similar to the use of *ori*C plasmids in *E. coli*, is to study **autonomously replicating sequences** (**ARSs**)[469,486] and plasmids[468] and **artificial chromosomes**[487] made from them. ARS sequences were first found in the budding yeast *S. cerevisiae*. Plasmids containing an ARS, whose core consensus sequence is 5'-(A/T)TTTAT(A/G)TTT(A/T), replicate autonomously during S-phase. Such plasmids have been genetically engineered, providing them with telomeres and some kind of functional centromere, to form artificial chromosomes. **Yeast artificial chromosomes** (**YACs**) have become extremely important as cloning vehicles (Chapter 26), and they also serve as important tools for studying eukaryotic replication and its control. They can be cultured in yeast cells or can be transferred into animal cells, etc.

As mentioned in Section B, 1, human centromeres are rich in the repetitive α-satellite DNA. By joining α-satellite DNA-containing fragments of the X-chromosome to cloned telomeric DNA, human **minichromosomes** have been created.[488] These have been developed into **human artificial chromosomes**,[489] which may be practical vehicles for gene transfer in human therapy.

Replication of nuclear DNA. The budding yeast *Saccharomyces cerevisiae* has permitted the most detailed picture of DNA replication in a eukaryote. The complete genome sequences are known and the ARSs have been physically mapped.[468a] For example, in chromosome VI there are nine origins that differ in frequency of initiation and which replicate sequentially during the S-phase of the cell cycle.[490] The initiation (**replicator**) regions surround the 11-bp consensus sequence of the ARSs, each occupying at most ~150 bp. However, in metazoa and even in the fission yeast *Schizosaccharomyces pombe* the ARSs range from 500 to 1500 bp in length. These origin regions frequently overlap the **promoter** sequences, which control initiation of transcription (Chapter 28).[491] This association with transcription origins has also been observed in metazoan cells, where replication origins are often clustered.[492] However, there is no sequence homology between the ARSs of *S. cerevisiae* and replication origins in other species, even those of *S. pombe*.[490]

The study of replication in yeast ARSs and artificial chromosomes has revealed that initiation of replication requires not only an initiator protein but a complex of six proteins that form an **origin recognition complex** (**ORC**).[493–495a] This complex, which is essential to initiation of replication, may be joined by additional proteins in a **prereplication complex**. At least some of the ORC proteins have their homologous counterparts in metazoa, suggesting a highly conserved initiation machinery.[495a–c]

One or more of the proteins that bind to the ORC may constitute a **license** to replicate. The licensing

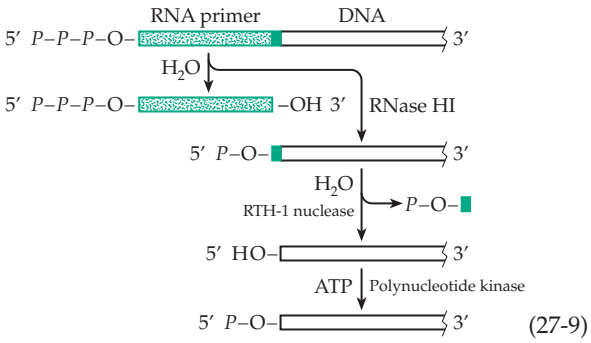
concept states that when replication occurs the license is destroyed and the origin involved cannot initiate replication again without a new license. The **replication licensing factor** (**LRF**) is postulated to be unable to pass through the nuclear membrane.[494,496] It can only reach the replication origin after the S-phase has concluded and mitosis has taken place. At this time the membrane has been disrupted. A second signal, the **S-phase promoting factor** (**SPF**), cannot act without an intact nucleus and a license in place.[469,495,497–500] This system ensures that DNA is replicated only once per cell cycle. Among the proteins involved in the licensing is a group of **minichromosome maintenance** (**MCM**) **proteins**, so-named because of their importance to replication of ARSs and artificial chromosomes.[495] Six of these proteins (MCM2–MCM7) can form a hexameric complex with one subunit of each type as well as other complexes, e.g., $(MCM4,6,7)_2$.[500a–c] The latter acts as an ATP-dependent helicase. A somewhat simpler MCM complex is found in archaea.[500d] Some of the cell cycle proteins (Chapter 26, Section F,1), including Cdc6 and the protein kinases Cdc7 and Cdc28 as well as other proteins, are also required for regulation of replication origins. Proteins homologous to those of yeast have been identified in humans and other eukaryotes. Licensing of replication involves association of the MCM helicases with each ORC during the G1 phase of the cell cycle. Binding of the initiation factor Cdc6/18 and of a recently discovered loading factor **Cdt-1** apparently completes the licensing. Once licensing has occurred both cdc6/18 and Cdt-1 can dissociate from the DNA.[496] Removal of Cdt is facilitated by its binding to another protein **Geminen**, found first in the frog *X. laevis*.[500e–g] The ORC complexes may remain at the origins. It has been estimated that a yeast cell contains ~400–600 molecules of the very stable ORC, about one ORC per replication origin. However, a large excess of MCM proteins may be present. Their concentrations may regulate the number of ORC molecules that associate with DNA. Replication of DNA is not the only aspect of cell growth. For example, as DNA is replicated histones must be synthesized and assembled. This synthesis occurs during the S-phase and is tightly coupled to replication.[500h]

Initiation of replication in metazoans is still confusing. Almost any piece of DNA will be replicated if introduced into a *Xenopus* egg, where initiation appears to occur just once at a random position.[501,501a,b] However, in differentiated tissue the origins of replication seem to be fewer in number and more specifically located. A possible explanation is that high concentrations of ORC and MCM proteins in the embryo may lead to many relatively nonspecific origins and a replicon size of ~7 kb. The lower concentrations of these auxiliary factors in somatic cells may lead to fewer but more specific origins with a replicon size of ~170 kb.[495] Three distinct mammalian origins have been studied in detail. One is in the β globin locus (Fig. 27-10).[500] A second is near the dihydrofolate reductase gene.[502] A third, which is activated early in S-phase, is at the 3' end of the lamin B2 gene.[503] The latter has been localized to a 500-bp region. These findings suggest that the replicon concept, as developed for yeast, may be generally applicable.

Replication of the intact genome of *Drosophila* has been studied in rapidly dividing nuclei by electron microscopy.[504] The replication rate in these nuclei is ~300,000 bases / s, but it has been estimated that replication forks in animal chromosomes move no faster than ~50 bases / s. Thus, we would anticipate at least 6000 forks, or one fork per 10 kK bases. Indeed, this number of forks has been observed.[505] They occur in pairs with many short regions containing single-stranded DNA as if one strand at the fork is replicated more rapidly than the other as in mitochondrial DNA. The arrangement of the ssDNA regions at the two forks in a pair suggests bidirectional replication. However, replication forks are rarely seen in higher eukaryotes, but extensive regions of single-stranded DNA are often visible. Benbow and associates suggested that in higher eukaryotes the strands of duplex DNA may be separated throughout a whole looped domain of DNA. Replication could then occur with initiation at many points along each strand.[505]

Replication reactions are similar in bacteria and eukaryotes, but some details differ. In eukaryotes at least two DNA polymerases, α and δ, are required. In budding yeast polymerase ε is also essential.[506] Both polymerases δ and ε may replicate separate strands at the replication fork.[506a] Processing of Okazaki fragments also differs from that in bacteria, where either RNase H or the 5' to 3' exonuclease activity of DNA pol I removes the RNA primer (Fig. 27-14). This exonuclease activity is lacking in eukaryotic polymerases. Replication primers are removed in a two-step process by **RNase HI**, which makes an endonucleolytic attack that removes all but one nucleotide residue of the primer in a single piece, leaving a 5'-phospho group on the remaining ribonucleotides. That residue is removed by a 5' to 3' exonuclease designated RTH-1 nuclease (Eq. 27-9).[467,507,507a,b] This is a homolog of the yeast RAD27 protein. A polynucleotide kinase may then phosphorylate the 5' end of the DNA fragment.[508]

Another difference between bacterial and eukaryotic replication is the presence of nucleosomes in eukaryotes. Some evidence suggests that nucleosomes may open and close to allow replication forks to pass through.[509] Studies of SV40 minichromosomes indicate that passage of the replication machinery does destabilize nucleosomes, which must be partially reconstructed about 260 nucleotides past the elongation point.[510] Another factor is the variable extent and location of modifications to histones, in particular to



(27-9)

the H3 and H4 histone tails (Section A,3). A code has been proposed according to which certain modifications would favor transcription or mitosis, while lack of modification would silence the genes.[72]

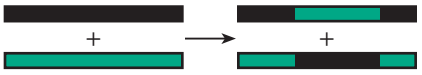
Much of the control of replication is at the initiation stage. Growth factors and other mitogenic stimuli acting at the plasma membrane can stimulate expression of such nuclear proteins as those encoded by the proto-oncogenes *c-myc*, *c-myb*, and *c-fos*. These may initiate a regulatory cascade (Fig. 11-13) and trigger mitosis.[511] As indicated in Chapter 26, the 34-kDa protein kinase encoded by fission yeast gene *cdc*2 (budding yeast CDC28) is essential for progression of the cell cycle through the G1 phase into mitosis (Eq. 26-3). A single oscillation in this kinase activity induced by a B-type cyclin can promote both replication and mitosis. However, in *S.cerevisiae* there are 14 different cyclinlike proteins, and their individual functions are not clear.[512] The signal that is sent to the ORCs is likewise unclear.[513] However, theoretical models involving Eq. 26-3 and many additional components have been proposed.[514] Multiple phosphorylations may occur, some on the RPA initiator protein.[515] Many proteins required for replication, including DNA polymerase and primase, are associated with the nuclear matrix.[516] The nuclear membrane may also be important in controlling replication.

D. Integration, Excision, and Recombination of DNA

The exchange of genetic information between chromosomes, plasmids, and viruses occurs in many ways, which are described collectively as recombination.[517–520] Mutants of *E. coli* deficient in recombinational ability often have defects in genes designated *recA, B, C,* etc. (for recombination), or *ruvA, B, C* . . . (for resistance to ultraviolet light). Some of these mutants are unusually sensitive to ultraviolet light because of their inability to repair damage to DNA. Several of the recombination enzymes are used for repair of ultraviolet damage and of double-strand breaks in DNA arising from other causes.[521,522] In eukaryotes recombination occurs during meiosis. Many viruses, including phage, also carry genes for their own general recombination systems. In addition, the DNA of some viruses, such as the temperate phage λ, undergoes recombination with the host DNA. This can happen during the processes of integration of the viral DNA into the host genome or excision of the viral DNA during the lysogenic cycle of replication (Chapter 28, Section B). Recombination occurs around specific sites in the chromosomes of both the virus and its host and is called **site-specific recombination**. Genetic recombination is essential to the development of genomic diversity, to the survival of a species, and to evolution.

1. Recombination Mechanisms

How can the homologous regions of two different DNA duplexes be brought together? As illustrated schematically in Eq. 27-10, the strand exchange must occur at exactly the same point in each duplex. An early attempt to explain this postulated a "copy choice" mechanism of replication. It was assumed that replication occurred along one DNA strand up to some random point at which the polymerase jumped and



(27-10)

began to copy from the second of a pair of homologous chromosomes. The newly formed DNA molecule would be complementary to different parts of both parental DNA duplexes. To test the idea, Meselson and Weigle infected *E. coli* with two strains of phage λ containing ^{13}C- and ^{15}N-labeled DNA, respectively.[523] Recombinant DNA was found to contain some ^{13}C and some ^{15}N, as judged by density gradient centrifugation. It was clear that DNA from both parents was incorporated into DNA of recombinant progeny, a finding that ruled out the copy choice hypothesis.

If recombination occurs instead by enzymatic cutting of two homologous duplex DNA molecules followed by rejoining, how is it possible to avoid inactivation of genes by addition or deletion of genetic material? Recombination cannot depend upon the random action of a nonspecific enzyme with random rejoining. Yet, general recombination can occur at any point and with a roughly constant frequency throughout the DNA chain. The explanation of these facts lies in the occurrence of base pairing between at least some short homologous regions of strands of the two different DNA duplexes.

The Holliday recombination intermediate. In 1964, Holliday suggested a recombination process that would give rise to characteristic H-shaped intermediates.[524] Recombination could be initiated at special points on the duplexes, recognizable by a recombination enzyme (Fig. 27-22). A short amount of unraveling would be followed by strand exchange with the two broken strands being rejoined by a ligase as indicated in Fig. 27-22. The crossover points would then migrate up or down the chains as the two helices turned about their own axes. Long regions of heteroduplex DNA could be generated in this way, and the process could be terminated at a random distance from the starting point, accounting for the observed uniformity of genetic recombination events. Chain cleavage and rejoining of two of the strands would terminate the process. If these were the same strands broken in the initiation event (cleavage at points *aa′* in Fig. 27-22), genes lying outside the heteroduplex region would not be recombined, but cleavage of the other chains (at points *bb′*) would lead to their recombination. Intermediates of the type predicted by the Holliday model were soon observed by electron microscopy (Fig. 27-23).[525] Three-dimensional structures have been determined by X-ray crystallography[525a] and have been studied by atomic force microscopy.[525b]

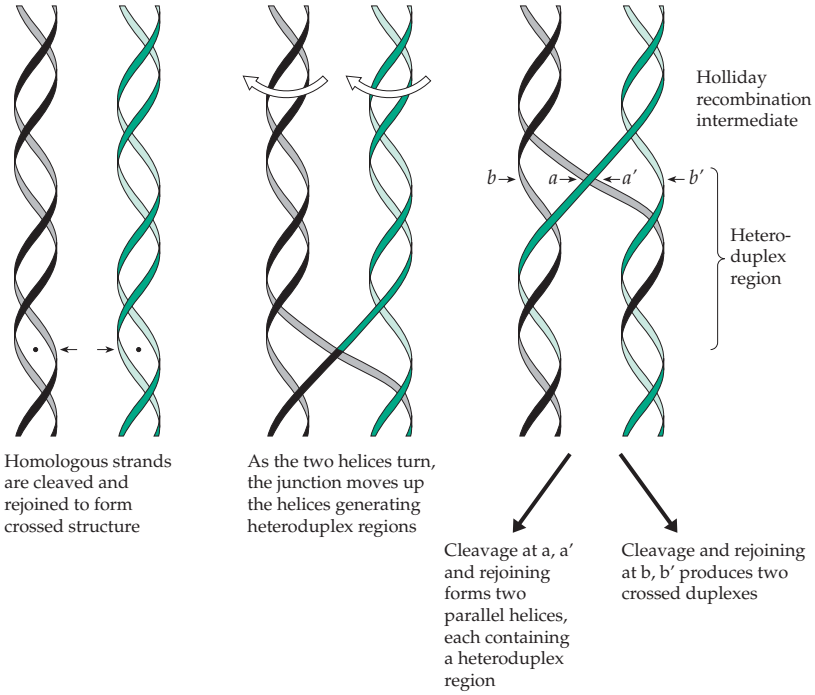
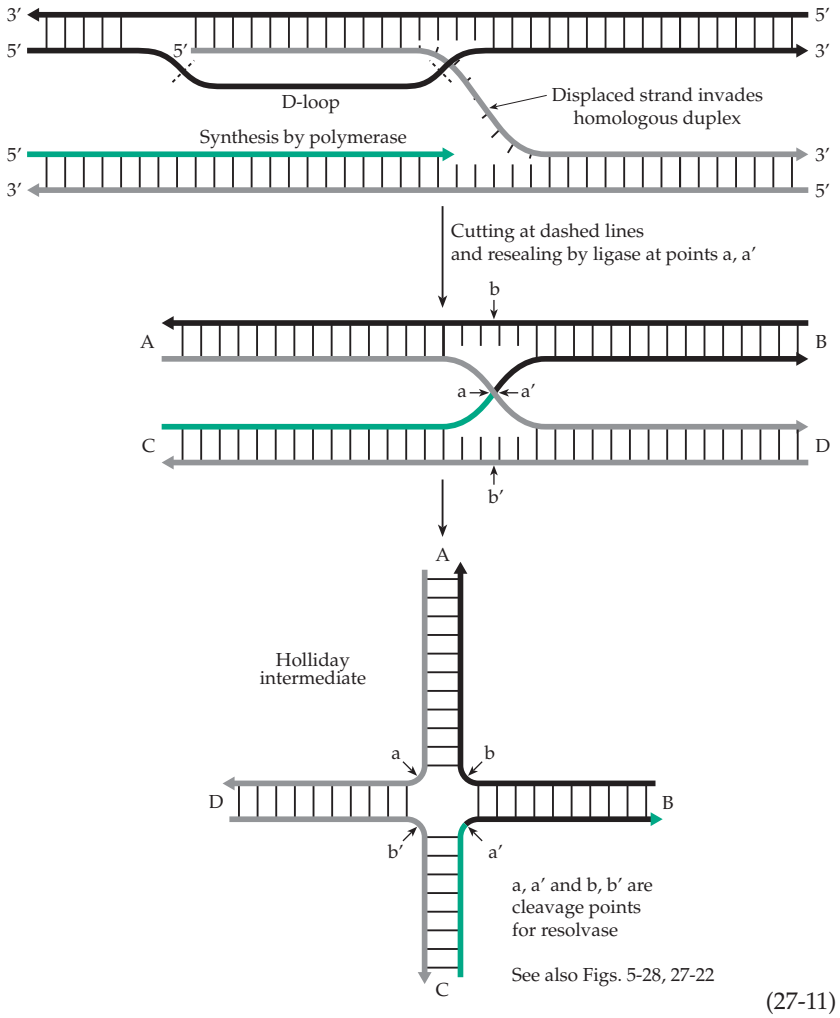


Figure 27-22 A recombination mechanism involving single-stranded exchanges. After Holliday.[524]

The cross-stranded structure shown in Fig. 27-22 can be formed with all base pairs in both duplexes intact.[526,527] All that is required is formation of a nick in each of the two polynucleotide chains and a rejoining of the backbones across the close gap between the duplexes. This model also accounts for the cutting of the two crossed strands at exactly equivalent points to terminate the process. Various mechanisms of recombination exist, and most make use of the key Holliday recombination intermediate (Fig. 27-22, Fig. 5-28, Eq. 27-11).[521,528–530a] Such four-way junctions can arise in several ways.[530a–d] For example, a 3' or 5' single-stranded tail in a piece of dsDNA can "invade"

Figure 27-23 A chi form of DNA from the colicin E1 plasmid. These forms are thought to be derived from recombination intermediates of the Holliday type, which appear as "figure eight"-shaped molecules twice the length of the colicin genome. This figure eight form was cut at a specific site that occurs only once in the genome (twice in the figure eight) by restriction enzyme *Eco*R1 to give the chi form. The pairs of long and short arms are believed to represent homologous duplexes. The single strands in the crossover have pulled apart revealing the strand connections clearly. Such a structure would be expected from the Holliday intermediate (upper right corner of Fig. 27-22), e.g., if one of the two vertical duplexes were rotated end over end. From Potter and Dressler.[525]



(27-11)

another dsDNA that has a homologous sequence as indicated in Eq. 27-11. In this drawing a 5' tail has been displaced during repair of a gap in one strand. The resulting D loop may be trimmed out and a new connection made to give the Holliday intermediate. The cleavage points a, a′, b, b′ marked in Eq. 27-11 correspond to those in Fig. 27-22. Holliday junctions may also be formed in stalled replication forks and must be removed to allow replication and transcription to continue.[530e] The existence of the Holliday intermediate has been supported not only by electron micrographs such as that of Fig. 27-23 but also by the identification of endonucleases that carry out the necessary cleavages of synthetic Holliday intermediates that have been made artificially (see Chapter 5.) Endonucleases with a high specificity for Holliday junctions have been found in bacteria, among proteins encoded by viruses, and in a wide variety of eukaryotic cells. Additional proteins including helicases, DNA-binding proteins, and specialized **strand exchange proteins** are also required to catalyze the individual steps in the recombination process.

The main **RecBCD pathway** of recombination in *E. coli* depends upon a dsDNA nuclease and an unwinding complex consisting of proteins RecB, C, and D.[531–533] The complex is a powerful *exonuclease*, which can digest the ends of a DNA duplex. It degrades the 3' ends most rapidly, leaving 5'-tails that can invade other homologous duplexes as in Eq. 27-11. The RecBCD complex is also an ATP-dependent helicase, which unwinds the DNA, preparing ssDNA for reaction with the strand-exchange protein **RecA**. The RecBCD complex also functions to completely degrade foreign dsDNA such as that from invading bacteriophages.[532] Why doesn't it also degrade the genomic DNA of the *E. coli* cell in which it functions? The answer lies in an eight-base DNA sequence, a recom-bination "hot spot" called **chi (χ)**: 5'–GCTGGTGG–3'. This χ sequence occurs 761 times in the leading strands for DNA replication in the *E. coli* genome.[533a–534a] When the RecBCD complex reaches a χ sequence, when approaching it from the 3' end, the enzyme stops its exonuclease action by inactivating the nucleolytic activity of the D subunit and promotes recombination about five- to ten-fold as fast as at other sites.[533,535]

RecA and other strand-exchange proteins. The 352-residue product of the *E. coli RecA* gene is a multifunctional **recombinase,** which is required both for recombination and also for DNA repair.[536–540a] In its repair function the RecA protein acts as a DNA-dependent protease that cleaves a number of repressors in response to damage to DNA. It has a quite different role in recombination where it (1) brings a piece of single-stranded DNA (an end or a gap) together with a duplex; (2) locates homologous sequences; and (3) forms a synaptic complex in which strand exchange can occur. Electron microscopic observations[536] show that the RecA protein binds to either single-stranded or duplex DNA in a cooperative manner to form long rodlike spiral filaments (Fig. 27-24). Measurement of the lengths of RecA protein-covered duplexes shows that the DNA is underwound and stretched by about 50%. It contains ~18 nucleotides per turn.[536,539,539a] Similar filaments are formed with single-stranded DNA, 3–4 nucleotides being bound per RecA protein monomer. Formation of this ssDNA complex, which may be regarded as an initiation

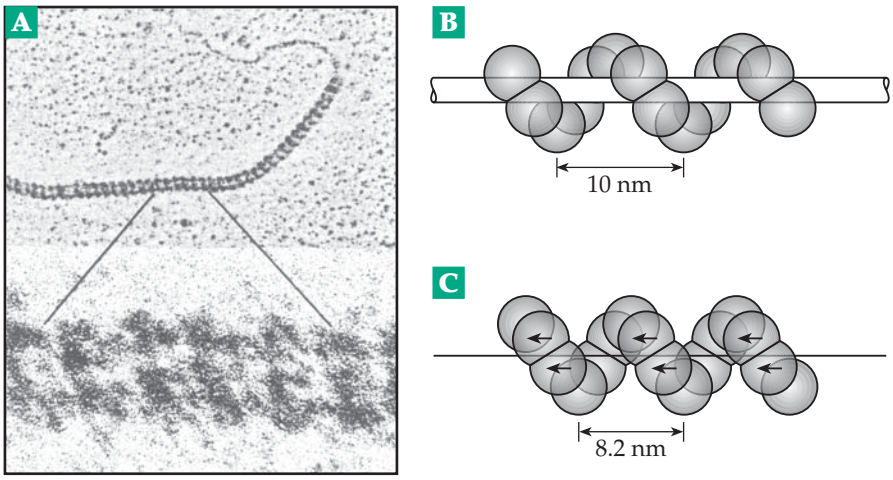
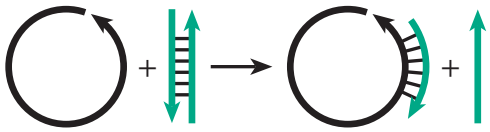


Figure 27-24 Structures of RecA protein spiral filaments. (A) RecA protein filament formed on circular duplex DNA in the presence of ATP(γ-S), shadowed with Pt and seen by electron microscopy. (B) Diagram of RecA bound to duplex DNA in the presence of ATP(γ-S), as determined by electron microscopy. RecA monomers are shown as spheres, but their exact shape is unknown. (C) Diagram of RecA spiral filament in crystals of RecA protein free of DNA, based on X-ray crystallographic data. Arrows indicated alignment of monomers. From Howard-Flanders, West, and Stasiak.[536]

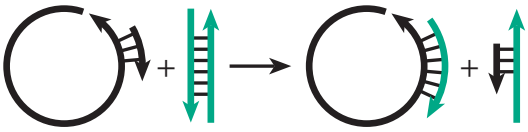
complex for recombination, requires MgATP and is faciliated by prior coating of the DNA with SSB protein. The RecA protein subunits are added in the 5'- 3' direction of the DNA, and SSB is displaced in the process.[541]

The initiation complex binds to duplex DNA rapidly and more slowly promotes strand exchange. In related reactions a single-stranded SSB-coated circular DNA will bind to RecA protein, then exchange strands with a linear duplex (Eq. 27-12). The strand exchange requires ATP and advances in the 5' to 3'



(27-12)

direction along the original single strand at the rate of a few bases / s. Strand exchange can also occur between two duplexes if there is a suitable gap in one strand, e.g., as is illustrated in Eq. 27-13.



(27-13)

A possible mechanism of strand exchange is illustrated in Figs. 27-25 and 27-26. The RecA protein has binding sites that can accommodate nucleotides from two DNA molecules, one single-stranded and the other a duplex. It may also accommodate two DNA duplexes.[542] As shown in Fig. 27-25, the RecA protein could test the hydrogen-bonding between many base pairs at once in a search of homologous regions. The two DNA chains would have to either slide past each other or repeatedly dissociate and reassociate[543] until a homologous region was found. Then strand exchange could occur. As is shown in Fig. 27-26, the single strand may be wound into the major groove of the duplex to form an interwound triplex. The matching of hydrogen-bonding atoms may be an attractive way of searching for homology, but the actual search seems to substitute speed for precision. Base substitutions are quite permissive. The need for precise hydrogen-bonding has not been demonstrated, and the exact recognition mechanisms in homologous recombination remain uncertain.[544–544b] Whole chromosomes must be aligned and checked rapidly in the homology search.[545]

Many proteins similar to the RecA protein and with similar functions have been found.[546,547] These include the products of gene *uvsX* of phage T4,[548,549] the β protein of phage lambda,[550] the yeast RAD51[551,551a] and human RAD51 proteins,[552,553] a meiosis-specific human homolog of the RecA protein,[554] and corresponding proteins from plastids of higher plants.[555] Both the UvsX protein of phage T4 and human RAD51 protein yield strands of coated DNA similar to those in Fig. 27-24.

Processing the Holliday junction. Completion of the recombination process requires "resolution" of the Holliday intermediate by endonuclease action followed by ligation and perhaps by gap repair. The major recombination pathway in *E. coli* employs a binding protein, a nuclease, and a helicase encoded by genes *RuvA*, *B*, and *C*.[528] **RuvA** is a DNA binding protein specific for symmetric Holliday junctions.[529] **RuvB** is a closely associated ATP-dependent helicase.[556–558] On the basis of genetic and X-ray crystallographic evidence it is now evident that some of the functions previously attributed to RecA are carried out by the RuvABC complex. As indicated in Fig. 27-26B, RuvA binds to the Holliday junctions, holding it in the symmetric square configuration in which branch

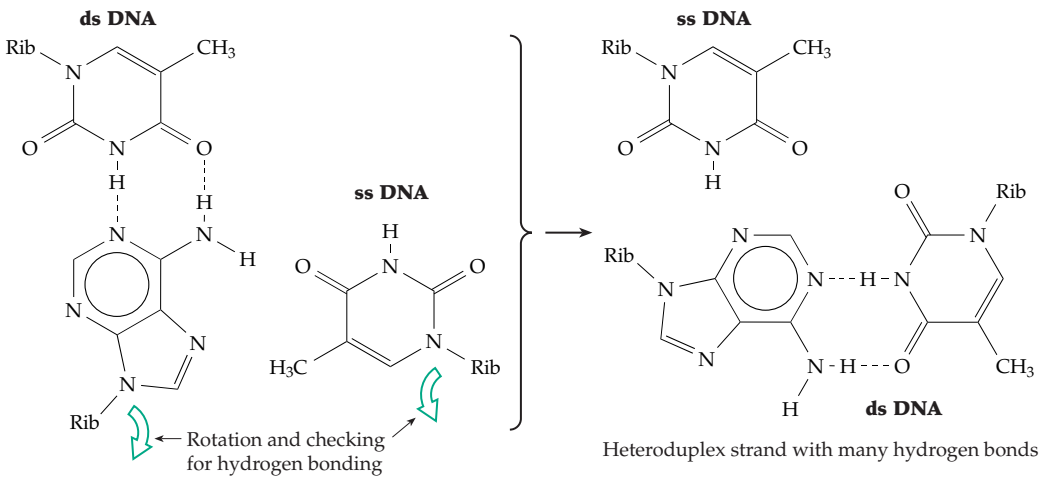


Figure 27-25 A possible mechanism for homologous pairing of an ssDNA with a duplex DNA and strand exchange. The ssDNA (right) binds together with a hydrogen-bonded duplex (left). The RecA protein rotates the bases into the heteroduplex configuration, where hydrogen bonds may be formed in many of the base pairs. After Howard-Flanders, West, and Stassiak.[536]

migration is possible. Two molecules of the oligomeric RuvB helicase apparently rotate the DNA, causing branch migration and movement of the DNA through the RuvAB complex.[559,559a,b] Under some circumstances a different helicase, encoded by *E. coli* gene **RecG,** moves Holliday junctions in the opposite direction.[560] In addition to RuvA a variety of other Holliday junction-binding proteins are known.[561] These include **p53** and the nuclear HMG proteins.[560,561]

RuvC is an endonuclease that is highly specific for Holliday junctions. It is a **resolvase** that cuts at either points a,a′ or b,b′ of Eq. 27-11 to form either "patched" or "spliced" recombinant DNA (Fig. 27-26C). Similar resolvases process bacteriophage DNA[562–564] and have also been found in yeasts and in mammals.[565,566] All are dimeric metal ion-dependent proteins.[567]

2. Nonreciprocal Recombination and Unequal Crossing-Over

The phenomenon of **gene conversion** or **nonreciprocal recombination**[568,569] was first recognized in genetic studies of fungi for which the four haploid miotic products can be examined individually (tetrad analysis; p. 20). Instead of the normal Mendelian ratio of 2:2 for the gene distribution in the progeny at a heterozygous locus a ratio of 3:1 is sometimes observed. One of the recombinant chromosomes appears to have been altered to a parental type. A reasonable mechanism by which this can occur arises from the fact that heteroduplex regions, which are present in recombination intermediates, contain defects in base pairing. One strand of the heteroduplex will have a base that does not properly pair with the base in the other strand, or will have an extra base that loops out from the heteroduplex. Since cells contain repair mechanisms that search for defects and carry out a repair process, there is a likelihood that one strand in the heteroduplex region will be altered to restore perfect base pairing, thus causing the observed gene conversion. "Flanking" genetic markers outside of the heteroduplex region are unaffected by gene conversion, and during meiosis crossing-over between these markers occurs in about 50% of gene conversion events as would be predicted from the model of Fig. 27-21. Data from yeast show that nonreciprocal recombination during meiosis may also result from double-strand breaks and gap formation followed by repair synthesis using both strands of the homologous chromosome as templates.[569]

Recombination is not limited to meiosis but can occur between homologous chromosomes during mitosis, during the G_1 period preceding mitosis, or even during the G_2 period.[570,571] Certain mutations in yeast abolish meiotic recombination but have much less effect on mitotic recombination.[572] Thus, the two processes are not identical. It has been suggested that mitotic recombination is utilized to maintain sequence homogeneity between repeated eukaryotic genes.[572,573]

Since DNA contains many repeated sequences, crossing-over sometimes occurs between locations that are not the same in the two duplexes. Such **unequal crossing-over** has the effect of lengthening one duplex and shortening the other. This may be very important in evolution. It may also, surprisingly, function to preserve homogeneity of chromosomes within a species.[574,575] For example, tandem arrays of ribosomal RNA genes (Section B,3) in yeast have 140 identical copies of their 9-kb repeat unit.[575] Unequal crossing-over between either sister chromatids or homologous chromosomes, when repeated often enough, can lead statistically to a highly homogeneous population.

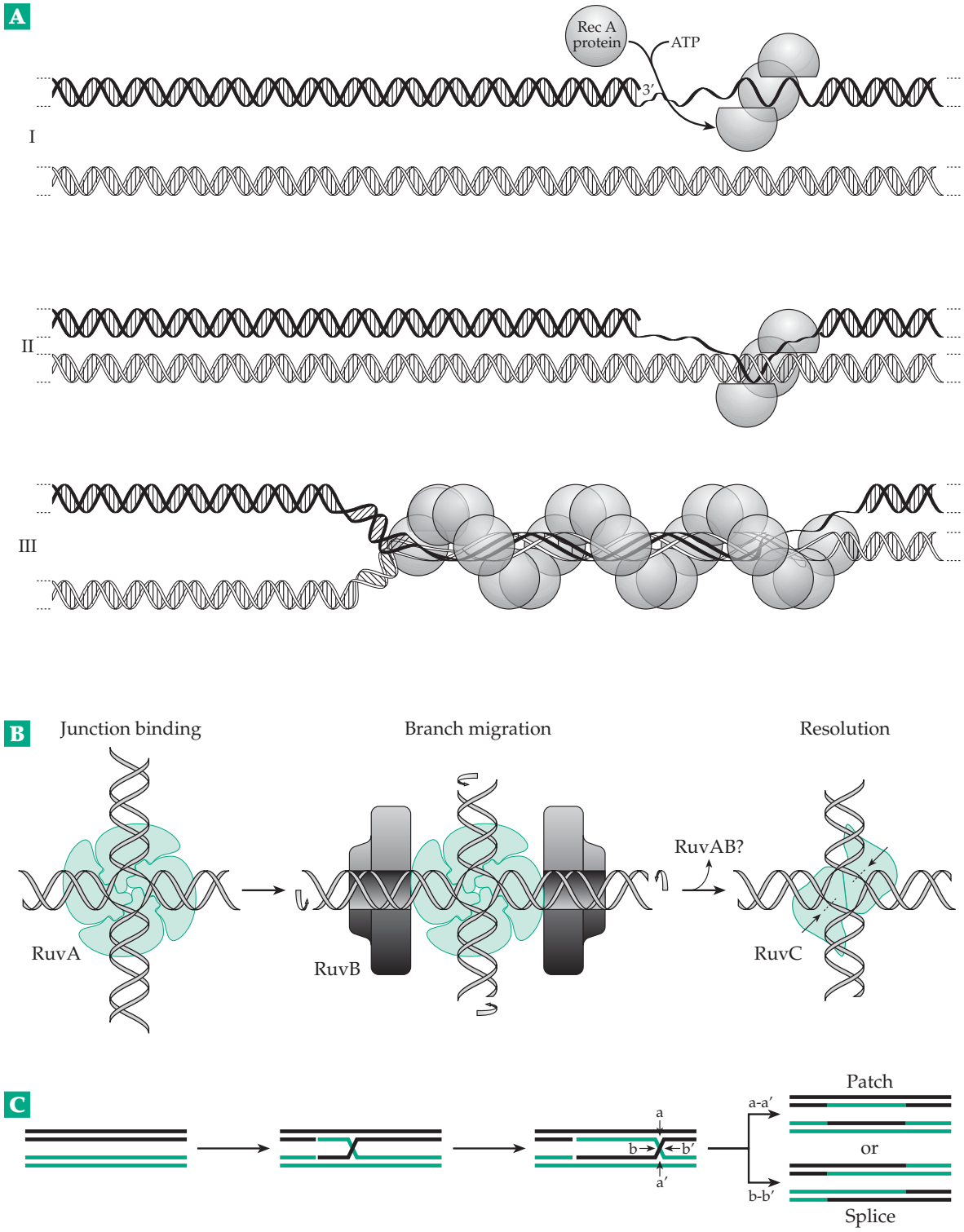


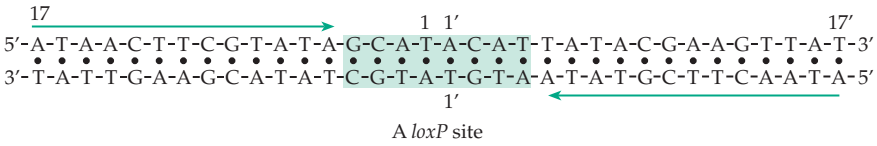
Figure 27-26 (A) Model for genetic recombination proposed by Howard-Flanders *et al.*[536] (I) RecA protein binds cooperatively to the single strand in a gapped duplex to form an initiation complex in preparation for pairing. (II) The initiation complex binds to the intact duplex, making transient contacts until a homologous site is reached. For clarity, the initiation complex is drawn with only a few protein monomers, but in reality it is likely to extend over hundreds or thousands of nucleotides. (III) When homologous contacts are made and the strands become paired locally, the initiation complex acts as nucleus for further cooperative binding, which extends the RecA spiral filament around all three or perhaps all four interacting strands. (B) Arrangement of the proteins and DNA during three stages of recombination catalyzed by the RuvABC system. The two RuvB hexameric rings are shown in cross section with the DNA passing through their centers. After Rafferty *et al.*[529] See also chapter banner, p. 1527. (C) Scheme of DNA rearrangement during homologous recombination in *E. coli*.

3. Site-Specific Recombination and the Integration and Excision of DNA

Recombination at specific sites in DNA is responsible for integration of DNA from viruses into the genome and for the cutting out of viral DNA and other pieces of DNA from the genome. The temperate bacteriophage λ and the F factors and R factors of bacteria can all be integrated into the genomic DNA of the host in this way. Genes encoded by the phage or plasmid are required. In the case of phage λ the viral genes *int* and *xis* are required for integration and excision, respectively.[576] These are not the same as the enzymes of the *rec* loci of the bacterium or the general recombination genes *exo* and *bet* of the phage. In addition, both integration and excision require an *E. coli* protein called integration host factor (**IHF**), a DNA bending protein resembling the DNA-binding HU.[18b,265,577,577a]

Integration of λ DNA (Fig. 27-27) occurs at the ~25-bp site *att B* in *E. coli* (Fig. 26-4) and the ~240-bp site *att P* in the λ chromosome (Fig. 28-11). These two sites contain identical 15-bp core sequences within which the recombination occurs. In a manner similar to that of the recA protein a homologous region is located by the complex of the Int (**integrase**) protein and IHF. Several molecules of Int protein bind and, together with the IHF protein, hold the phage DNA in a nucleosomelike structure (an **intasome**) in which the recombination occurs.[406,578–580b] Strand cleavage and rejoining occur within the short core sequence (Fig. 27-27). Parts of both *att P* and *att B* are recombined to give sites *att L* (left) and *att R* (right) in the DNA of the integrated prophage. In the integration complex the two core *att* sequences are aligned, and single-strand cuts are made at one of the points *a* or *b* that are indicated by the small arrows located on opposite strands and seven bp apart in the core sequence shown in Fig. 27-27. Rejoining of strands from the opposite duplex yields a Holliday intermediate. That this really occurs is shown by the fact that the Int protein cleaves synthetic Holliday intermediates derived from the *att* core and reseals the strands to give the expected products.[581] Cleavage of the Holliday intermediate at points a (Fig. 27-27) will lead to excision of the viral circle, but cutting at points b followed by resealing with opposite strands, as is observed, will yield integrated prophage. Although Int and IHF proteins are sufficient to promote integration, the **excisionase** encoded by phage λ gene *xis* is needed together with the Int protein for excision of the λ prophage.[576,582]

The integrase (tyrosine recombinase) family. The lambda integrase is the first recognized member of a family of a hundred or more closely related enzymes that are involved not only in integration and excision of phage DNA but also in converting multimeric forms of bacterial and plasmid chromosomes into monomers. One of the best known integrases is the 38.5-kDa **Cre recombinase**, which functions to keep the lysogenic phage P1 in a monomeric form by recombination between pairs of 34-bp core sequences designated *loxP*.



Since the reaction doesn't require accessory protein factors and can be performed *in vitro* with a variety of DNA substrates, the *Cre–loxP* system is much used in genetic engineering.[583,584] A pair of related integrase subunits known as **XerC** and **XerD** perform a similar function for the *E. coli* chromosome as well as for multicopy plasmids.[585,586] The XerC / XerD system is

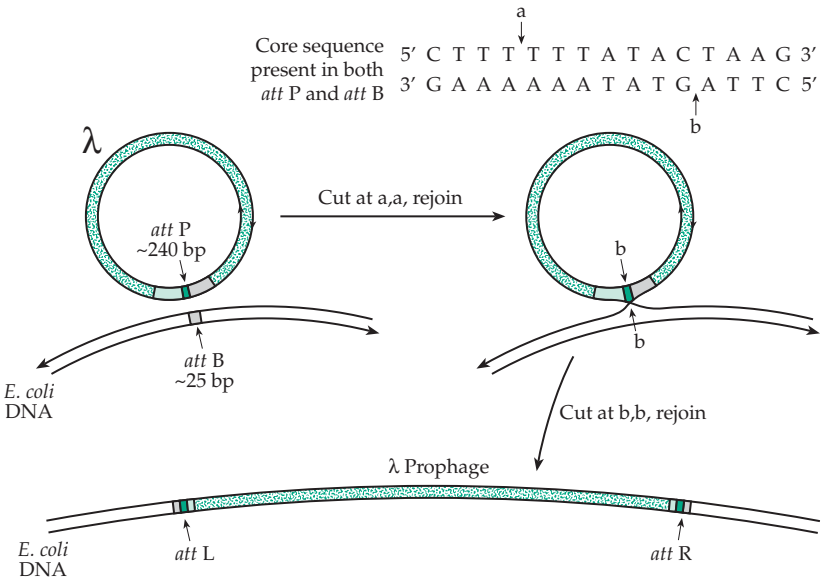


Figure 27-27 Integration of the phage λ genome into the *E. coli* chromosome at site *att*B. The same recognition sequences are present at *attB* and *attP*. These are cut at points *a* with rejoining to give the structure at center right. This is cleaved at points *b* with rejoining to give the integrated prophage.

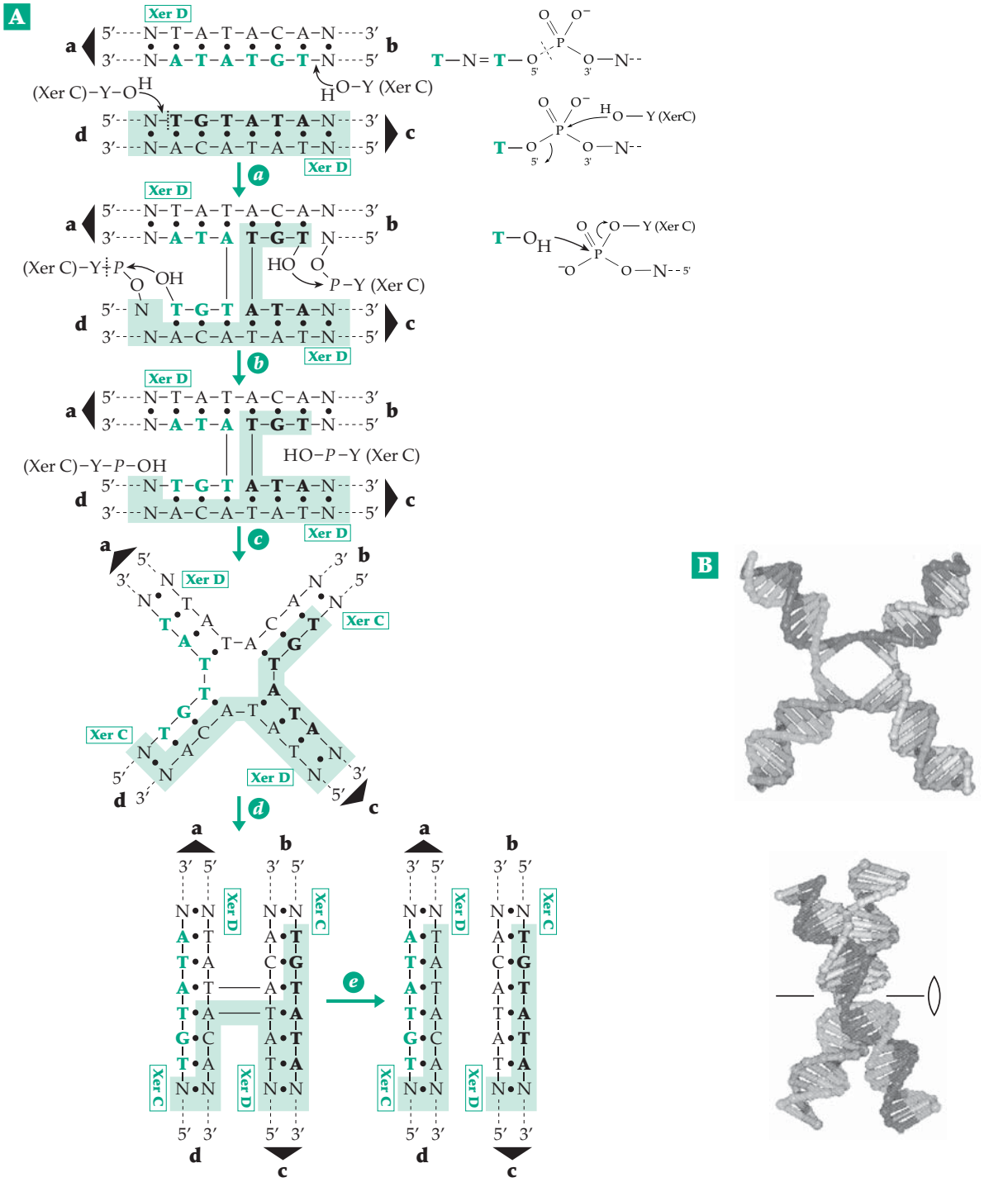


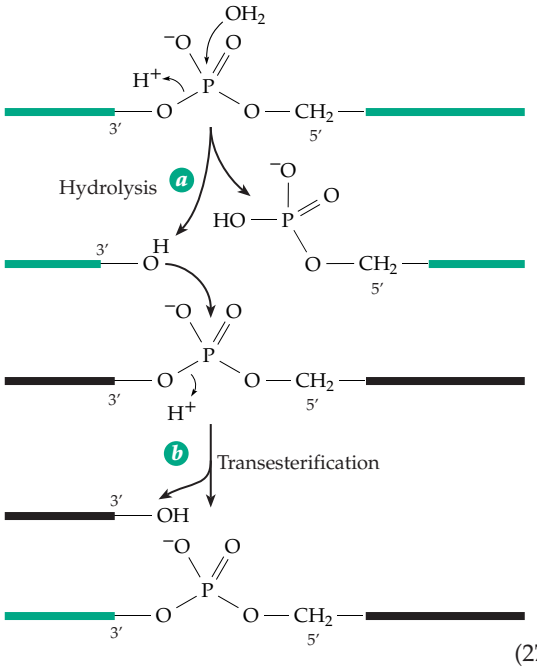
Figure 27-28 (A) Action of the integrase XerC / XerD on a pair of *E. coli dif* sites, containing the central six bp sequences TATACA/ATATGT, which are shown in an antiparallel orientation. In step *a* the active site tyrosine hydroxyl groups (Y-OH) of a pair of XerC subunits carry out transesterification reactions on the 5'-terminal thymidylate residues of the central hexanucleotide sequences to yield 3'-phosphodiester linkages to the XerC tyrosines. In step *b* the cut 5'-ends, containing free thymidylate 5'–OH groups, fold back to form new base pairs, and the strands are resealed in a second transesterification. This generates a Holliday junction, which *isomerizes* (steps *c* + *d*) to an isomeric species that is acted on by a second pair of transesterification steps (*e*) that are catalyzed by protein XerD, again with folding back of the central trinucleotides. After Arciszewska *et al.*[586] (B) The two isoforms of an antiparallel stacked X Holliday junction are shown. These can be reached from the symmetric square form shown in Fig. 5-28 and schematically in (A) by folding into X-conformations in which all base pairs are stacked either in pairs a,b and c,d or a,c and b,d. From Eichman *et al.*[525a]

atypical because it utilizes a pair of integrase subunits rather than just one. However, the basic chemistry (Fig. 27-28) is the same for the entire family.[587,587a] As with the λ integrase (Fig. 27-27) the XerC / XerD complex acts on a pair of identical core sequences that are aligned in an antiparallel fashion. Active sites in all of the integrases contain the conserved amino acid sequence Arg-His-Arg-Tyr. All of these residues are essential for catalysis.[585] Staggered cuts are made sequentially in the core sequences, e.g., at points a and b in Fig. 27-27 and adjacent to the 5'-terminal thymidylate residues in Fig. 27-28A. A transesterification reaction forms 3'-phosphotyrosyl linkages from the cut DNA to the integrase protein. The freed 5'-OH groups on the other cut end fold back and recombine with the bound 3' ends of the second duplex to generate a Holliday junction. In the mechanism proposed in this figure, three nucleotide units are involved in the folding back. The hydrogen bonds of their initial base pairs are broken, and new bonds are formed. This "base swapping" process is the equivalent of a short branch migration and verifies homology of the two recombination sites.

The Holliday junction can isomerize readily between two forms. In one a base-stacked double helix runs between ends a and b in Fig. 27-28B and another runs between ends c and d. The core hexanucleotide sequences lie between the marked XerC and XerD cleavage sites. In the other isomer (lower drawing) one helix has ends a and d and the other b and c. Following the isomerization (steps *c* and *d* in Fig. 27-28A) the XerD active sites act in two transesterification reactions (not shown but analogous to those in steps *a* and *b*) with base swapping of the trinucleotides at the cut ends. This generates the two separate recombinant duplexes. These might be two circular chromosomes or plasmids formed by recombination from a double length chromosome or plasmid. The previously mentioned λ amd Cre recombinases appear to act by closely similar mechanisms.

Tyrosine recombinases of the lambda family also function in eukaryotes. Best known is the **FLP (Flip) recombinase,** which is encoded by the 2-µm plasmid of *Saccharomyces cerevisiae* and is thought to function in amplifying the number of plasmid copies.[265] The 6.3-kbp plasmid contains a unique DNA sequence that lies between two 599-bp repeats in inverted orientation. Embedded in each repeat is an FLP recombination target (**FRT**) sequence, which is recognized by the plasmid recombinase. Each FRT segment includes inverted repeats 13 bp in length with an 8-bp spacer between them. As with other integrase systems the 8-bp spacer or **strand exchange region** is **asymmetric** and establishes the orientation of the recombination sites.[587–590] The role of the recombinase is to invert one of the 599-bp repeats with respect to the other (see Eq. 27-15). This switches replication of the plasmid to a rolling circle pattern.[265]

The resolvase/invertase family and invertible DNA sequences. A second large family of recombinases act by cleaving a target DNA sequence hydrolytically leaving a free 3'-OH end (Eq. 27-14, step *a*). This free end then attacks a phosphodiester linkage in a second strand of DNA, cleaving that strand with an in-line nucleophilic displacement (step *b*). Active sites usually contain a characteristic cluster of aspartate and

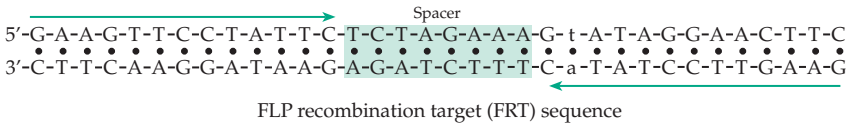


(27-14)

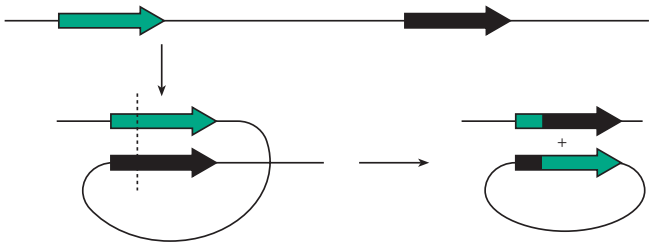
glutamate (DDE) side chains, which probably act together with a metal ion, perhaps as in Fig. 27-13.[591] Enzymes in this resolvase/invertase family act either to resolve cointegrates in transposon action (next section) or to invert DNA sequences.

If recombination occurs within a piece of DNA at two homologous sites such as the *attL* and *attR* sites at the boundaries of the λ prophage, the intervening DNA will be excised as a circular particle (Eq. 27-15). In this instance the two homologous regions must be repeated in the same direction, as is indicated by the arrow structures in Eq. 27-15. If the homologous sequences are oriented in opposite directions, i.e., they are inverted repeats, excision will not occur but the piece of DNA between the repeats will be inverted (Eq. 27-16).

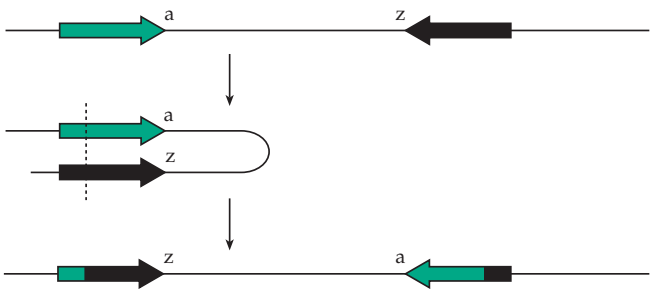
A number of such invertible



FLP recombination target (FRT) sequence



(27-15)



(27-16)

DNA recombination systems are known and are sometimes used to control specific genes.[592] For example, many strains of *Salmonella* have two types of flagella, which are composed of flagellin subunits encoded by genes H1 and H2 respectively.[592–594] On rare occasions an individual bacterium switches from one flagellar "phase" to the other. This occurs by recombination, as in Eq. 27-16, between two 26-recombination sites *hixL* and *hixR*, each of which contains a 14-bp inverted repeat. The 993-bp invertible segment encodes a recombinase gene called *hin* and a promoter, i.e., an mRNA initiation site, which has a specific orientation. In one orientation mRNA is transcribed from a short operon that includes the right inverted repeat IRR, the *H2* flagellin gene and gene *rh*1, which encodes a repressor for flagellin gene *H1*. Consequently, only gene *H2* is expressed. In the other orientation the RNA transcription is in the opposite direction so that neither *H2* nor *rh*1 is expressed. However, *H1*, which is located elsewhere, is expressed freely.

Two other proteins are required for efficient inversion by the Hin recombinase. A dimer of a 98-residue helix–turn–helix DNA binding protein called **Fis** (factor for inversion stimulation), a relative of protein HU,[18b] must bind to an enhancer, a 65-bp DNA segment. Binding of Fis to the enhancer helps to hold the supercoiled DNA and the recombinase in a correct orientation for reaction.[576,595,596] Protein HU is also needed. The same Fis protein binds to an enhancer for a 3-kbp invertible DNA sequence, which controls alternative host preferences for bacteriophage Mu (Fig. 27-29).[576,596] The chemistry of the inversion reaction is related to that of the replicative transposons discussed in Section 4.

Microorganisms sometimes control the synthesis of surface proteins using segments of invertible DNA. The pathogenic bacterium *Campylobacter fetus* utilizes DNA rearrangements to allow one of a large family of surface layer (S-layer) proteins to be formed.[597] The yeast FLP recombinase, mentioned in the preceding section, also inverts the sequence flanked by the 599-bp repeats.[589]

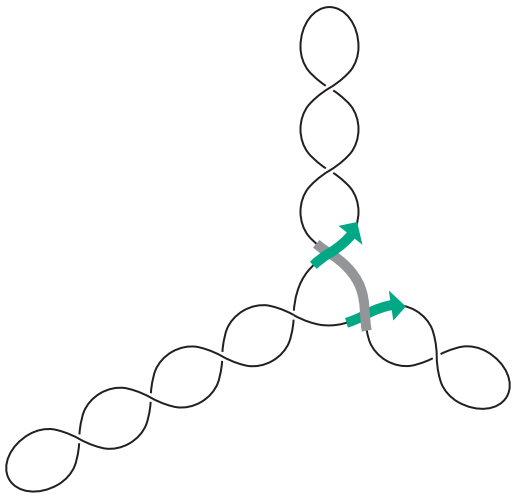
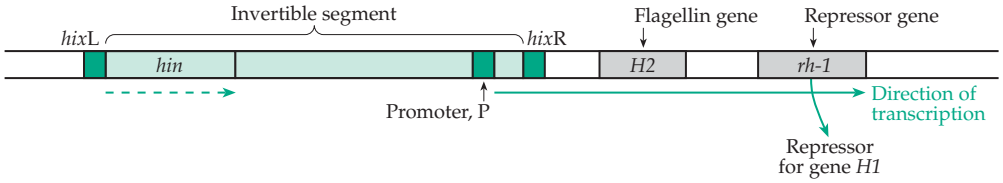


Figure 27-29 Formation of a synaptic complex of a supercoiled circular DNA containing the sites *gix* (green), which pass over and under the enhancer (gray). The recombinase Gin and the enhancer-binding Fis form a synaptic complex with DNA in this form as seen directly by electron microscopy. From Sadowski.[576]

4. Transposons and Insertion Sequences

The first evidence that some genes can move from one location to another within the genome came from studies of *Zea mays* by Barbara McClintock in the late 1940s.[598–602] She concluded that the variegated kernels found in some colored maize were a result of **controlling elements**, which could move from place to place turning on or inhibiting expression of various genes including some of those determining anthocyanin pigment formation. Two of these systems have been



studied especially intensively: the "**activator (Ac)–dissociation (Ds)**" system, discovered by McClintock and the "enhancer (En)–inhibitor (I)" system, discovered by Peterson[599,600] and independently by McClintock.[603] Each contains two genetic elements (segments of DNA) of which Ac and En are autonomous, i.e., they can move by themselves. Ds and I, however, cannot move unless the other element of the pair is also present. Both Ds and En have now been cloned and sequenced.

The general importance of transposable genetic elements was not appreciated by most molecular biologists until about 20 years after McClintock's discoveries. Then several moveable **insertion sequences** (Is elements) were found in enteric bacteria. Like the controlling elements of maize these small (0.8–1.4 kb) sequences can move and insert themselves at many points in the genome, often inactivating genes which they enter.[602–608] The *E. coli* genome contains eight copies of IS1, and five of IS2, as well as several others. Most species of *Shigella* contain more than 40 copies of IS1.[604] Mobile elements similar to those present in maize also exist in archaea.[604a]

Following the discovery of the IS elements it was found that transposable elements named **transposons** could transfer resistance to antibiotics between bacteria. All of these transposable elements have inverted repeat sequences at the ends. For example, IS1 contains the following sequence at both ends but with opposite orientation as if in a

$$\text{CAACTTA}\begin{smallmatrix}\text{C}\\\text{T}\end{smallmatrix}\text{TGAT}$$

palindrome.[605] Some complex transposons have an IS sequence at each end. For example, Tn10 contains IS10, a relative of IS1, at both ends.[608] These provide the characteristic inverted repeat termini. Figure 27-30 shows a schematic drawing of a bacterial drug resistance plasmid containing IS1, IS2, and IS10 as well as transposons Tn3, Tn10, and Tn55. The two IS1 elements surround the large resistance determinant, which can be transferred as a block.[602] To be autonomously mobile a transposon must contain a **transposase** that enables it to be transferred. It usually carries other genes as well and may also contain one or more signals for transcriptional regulation such as promoters.

The chemistry of transposition is more complex than that of simple site-specific recombination. Transposition can occur at many sites in a genome, and no homology with the transposon termini is required. Transposition is accompanied by *duplication of a short sequence of the recipient DNA* exactly at the ends of the transposon. Usually 5, 9, or 11 base pairs are duplicated as in the hypothetical example of Eq. 27-17. This happens because staggered cuts 5, 9, or 11 bp apart in the recipient DNA are made during recombination. These are indicated by the small arrows in Eq. 27-17. Transposons causing 5-bp duplications are Tn3, Tn7,[608a] γδ, phage Mu of *E. coli*, and Ty1 of yeast[609]; IS1, Tn10,[609a] and Tn5 of *E. coli* cause 9-bp duplications.[610–611b] When a transposon of one major group moves to a new location, the original copy remains. In this case transposition involves a combination of replication and site-specific (for the transposon) recombination. As a consequence, a circular DNA molecule containing the transposon will often react with a second circular DNA to form a large circle, a **cointegrate**, which contains two copies of the transposon. However, another group of transposons utilize a "cut-and-paste" mechanism that doesn't require extensive DNA duplication.

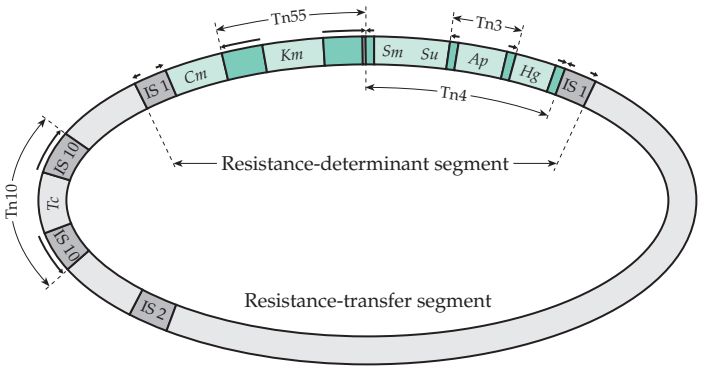
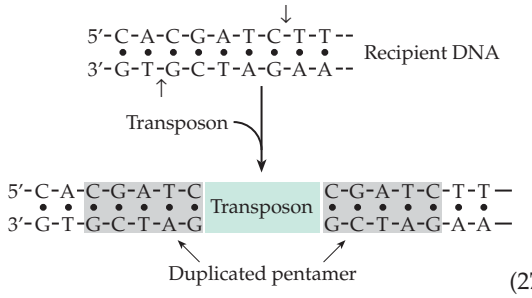


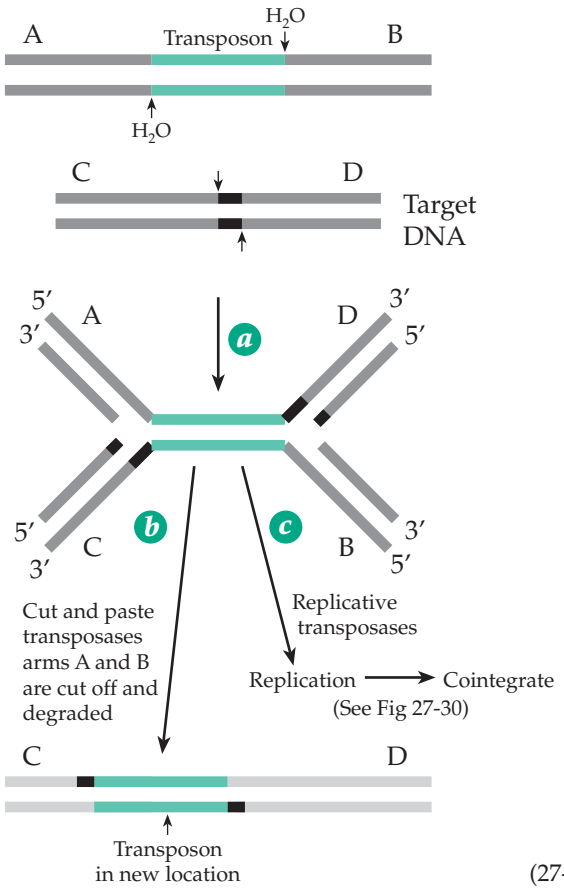
Figure 27-30 Transposons in an antibiotic-resistance plasmid. The plasmid appears to have been formed by the joining of a resistance-determinant segment and a resistance-transfer segment; there are insertion elements (*IS1*) at the junctions, where the two segments sometimes dissociate reversibly. Genes encoding resistance to the antibiotics chloramphenicol (*Cm*), kanamycin (*Km*), streptomycin (*Sm*), sulfonamide (*Su*), and ampicillin (*Ap*) and to mercury (*Hg*) are clustered on the resistance-determinant segment, which consists of multiple transposable elements; inverted-repeat termini are designated by arrows pointing outward from the element. A transposon encoding resistance to tetracycline (*Tc*) is on the resistance-transfer segment. Transposon Tn3 lies within Tn4. Each transposon can be transferred independently. From Cohen and Shapiro.[602]



(27-17)

Cut-and-paste (nonreplicative) transposons. The transposases of *E. coli* Tn5, Tn7, and Tn10 act by hydrolytically cutting both strands of duplex DNA at the transposon ends leaving the phospho groups attached to the 5' cut ends, as is depicted in detail in Eq. 27-14, step *a*. The two 3' ends then carry out transesterification reactions, as in Eq. 27-14, step *b*. These two steps are used to nick both strands of the DNA carrying the transposon and to join them to a target DNA sequence to give a branched intermediate (Eq. 27-18, step *a*). Nonreplicative transposons apparently cut off two arms, e.g., A and B, and heal up the small gaps by repair synthesis, leaving the transposon in a new location between C and D. The gap repair accounts for the duplication of the end sequences of the cut target DNA.[610–612a]

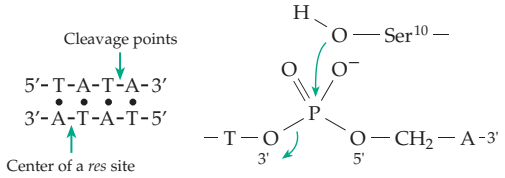
Replicative transposons. In 1979 Shapiro proposed the mechanism illustrated in Fig. 27-31 for replicative transposons. The two inversely repeated segments (green) at the ends of the transposon are aligned with the recipient DNA whose ends are labeled C and D. In fact, the recombining DNA molecules must be supercoiled.[1,613] Staggered cuts are made in the recipient DNA at points *a* and *b*, which are 5, 9, or 11 bp apart, depending upon the specific recombinase. Nicks are also made in the transposon ends. The 3' ends from the transposon are resealed with the 5' ends from the recipient DNA (step *a*) to give a structure that in effect has two replication forks. Replication (step *b*) yields the cointegrate, which contains two copies of the transposon as indicated. In a third step (step *c*) recombination between the two integrated transposons yields a copy of the original transposon-containing donor and the recipient DNA, which now also contains a copy of the transposon.



(27-18)

When a transposon reacts with another part of the same DNA circle there are two possibilities. The piece of DNA lying between the transposon and the recipient site may be excised as a circle containing a copy of the transposon. Alternatively, there will be inversion of that sequence as well as replication of the transposon (compare with Eqs. 27-15, 27-16).

The closely related transposons Tn3 and γδ are understood best. They contain not only a transposase gene but also a **resolvase** gene.[613–616] The transposase carries out the recombination reactions that yield the cointegrates, while the resolvase catalyzes site-specific recombination between the two transposons in the cointegrate to complete the transposition. Several subunits of the resolvases bind the two **resolution (res) sites** of the supercoiled DNA in a parallel orientation with the DNA supercoiled as in Fig. 27-29. However, no enhancer-binding protein is needed, and the two *res* sites must be supercoiled. Purified γδ resolvase uses hydroxyl groups of the serine-10 side chains as the nucleophiles to cleave the DNA by displacement at specific *res* sites to give transient enzyme-bound phosphodiester linkages.[615]



The synaptic complex contains 240 bp of DNA and at least two resolvase dimers. All four DNA chains are cut to give eight ends. Four of these are bound to the serine side chains in phosphodiester linkage. In the second step the freed 3'-OH groups react with the bound ends of the other duplex via a transesterification reaction to form the recombinant chains.

The resolvases act on supercoiled cointegrated DNA molecules that contain two directly repeated res sites to produce two singly linked circles (which are still supercoiled) each containing one res site as shown in Fig. 27-32. The two *res* (resolution) sites within the transposons are aligned, the open circle of DNA shown at the upper left being folded as shown in the lower part of the drawing. The DNA substrate is not knotted. However, after recombination it is catentated and will require action of a topoisomerase to separate

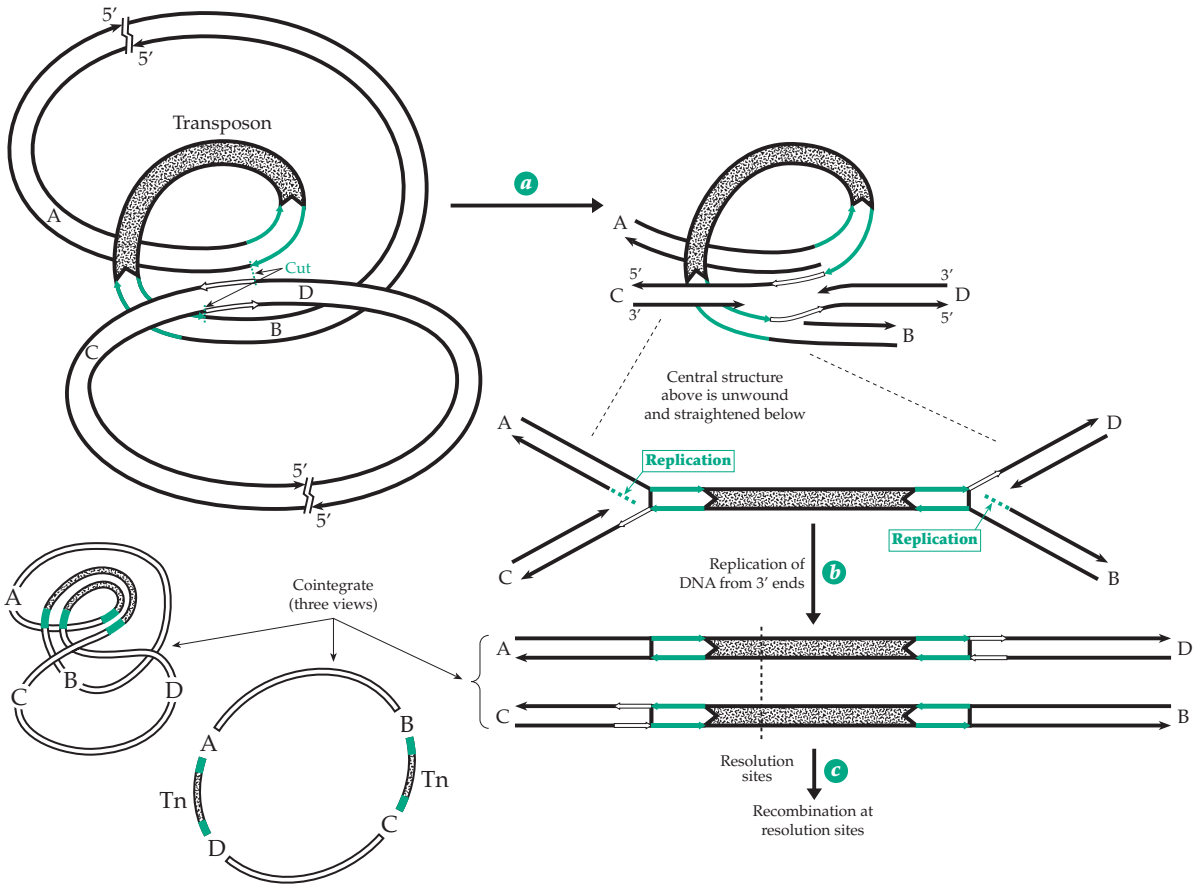


Figure 27-31 Scheme for integration of a transposon (stippled duplex) present in a piece of DNA with ends A and B into another piece of DNA with ends C and D and containing a suitable recognition sequence (open bars). Inversely repeated sequences in the transposon are shown as solid bars with a direction arrowhead. Arrowheads point toward 5′ strand ends. Cleavage and rejoining at points a,a and b,b yield an intermediate with two replication forks. Replication through the transposon yields one unchanged DNA segment with ends A and B and a transposon inserted into the other DNA segment. If A is continuous with B, a cointegrate structure is formed. See Cohen and Shapiro.[602]

the two products. Occasionally additional recombination events occur, perhaps processively along interwound double helices. This produces various knotted products.[617] An electron micrograph of one of these is shown in Fig. 5-17.

The temperate bacteriophage Mu. The most efficient transposon known is the 37-kb genome of the **mutator bacteriophage Mu**.[265,618–621] Once it becomes integrated into a bacterial chromosome, it replicates by repeated rounds of transposition within the host bacterium. During the lytic cycle some of the replicating DNA is excised as extrachromosomal circles of various sizes, which are packaged into virus particles by a headfull mechanism. The circles contain copies of some host DNA, but this is left behind when a virus particle infects a new cell.[265] The phage Mu DNA is integrated into host DNA by a cut-and-paste mechanism in a transposome that contains a tetramer of the virally encoded transposase (MuA protein) bound to a supercoiled DNA. The transposome resembles that in Fig. 27-29 and contains cleavage sites at the ends of the transposon and also an enhancer sequence.[620,621] Several steps involving conformational alterations occur in the transposome. One is an ATP-dependent action of a second Mu-encoded protein MuB.[622] During the lytic cycle replicative transposition predominates.

Some other transposons. Transposons have a variety of biological functions. For example, haploid cells of the yeast *S. cerevisiae* exist as one of two mating types *a* or α. The mating type is established by transposition of one of two "cassettes" of genes from two different "silent" locations to a location from which they can be expressed.[623,624] See Chapter 28.

One of the best known eukaryotic transposons is the P element of *Drosophila*, which transposes only within the germ line cells of developing embryos, somatic cells being unaffected.[265,625,626] It belongs to

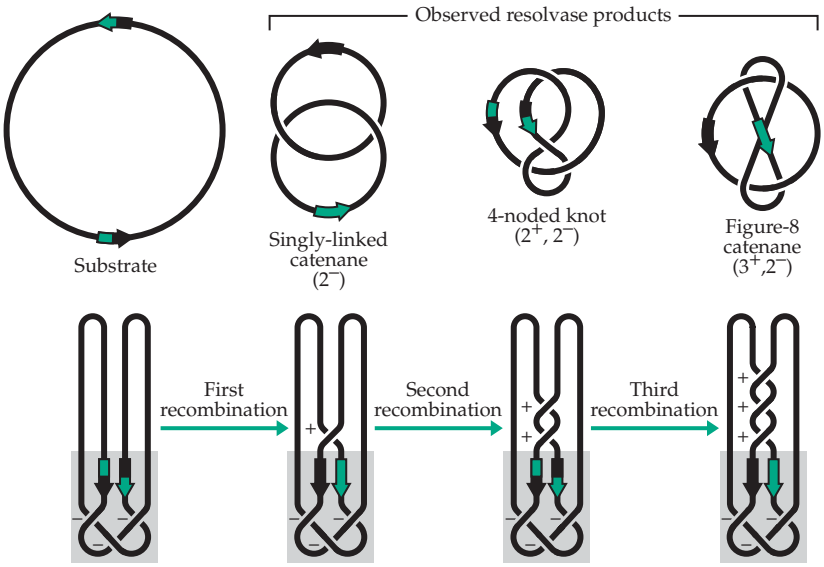


Figure 27-32 Scheme for resolution of an unknotted cointegrate molecule by a resolvase that cuts the transposons at resolution (*res*) sites and recombines them. The resolvase may act only once or repeatedly as shown. In the upper row, the duplex DNA substrate and products are represented in standard topological form as they might appear after nicking. In the lower row the DNAs are depicted as folded forms bound to the resolvase with the two directly repeated *res* sites (thick arrows) dividing the substrate into two domains (thick and thin regions). The substrate at synapsis has three (–) supercoils that entail crossing of the two domains. Successive rounds of recombination, each introducing a single (+) interdomainal node (see Fig. 5-17), are drawn in the lower row. Bound resolvase maintains the three synaptic supercoils. After dissociation from the resolvase at any stage, the product supercoil nodes either cancel with ones of opposite sign or are removed by subsequent nicking. The node composition is indicated in parentheses. From Wasserman *et al.*[617]

the same family as Barbara McClintock's Ac element of maize and the *Tc*1 family of nematodes.[627] P elements use a nonreplicative cut-and-paste method of transposition. The 87-kDa transposase protein requires GTP and Mg^{2+} for activity.[628] It was only in the past few decades that P elements have been found in *D. melanogaster*. They may have entered this fruit fly from another species, possibly transferred by a mite.[627,629]

A second *Drosophila* transposon called **mariner**[630] typifies the *mariner* / *Tc*1 transposon superfamily, which also contains members from nematodes,[631] other invertebrates, fishes,[632] amphibia,[633] and possibly human beings.[634] These transposons encode a transposase containing a D, D, D or D, D, E motif[630] but no other proteins. They contain short ~30-bp terminal inverted repeats and become inserted into host TA sequences.[631] Movement of some repetitive sequences of the LINE[635] and SINE[636] families within the human genome may be assisted by *mariner* transposons.[637]

The maize transposon *ac* is widely used as a means of inactivating genes and placing a "tag" that can be used to map the gene and to permit it to be cloned and sequenced.[598,638] Although initially of use only in maize the method has been extended to other plants,[639–641] and genetically engineered transposons have allowed it to be utilized in animals.[642]

A different kind of transposition controls self-sterility in maize. The cause of self-sterility in one strain has been traced to the presence of two linear episomes called S-1 (614 kb) and S-2 (514 kb) within mitochondria.[643] These have inverted terminal 208 bp repeats. On rare occasions they recombine with the circular mtDNA converting it to a linear form with the episomes covalently linked to one end. The change is accompanied by reversion to fertility.

As described in Section B,1, mammalian DNA contains many **retrotransposons** (retroposons) that lie within short direct repeats characteristic of transposons. However, they contain a poly(A) tail at the 3' end, an indication of their relationship to RNA transcripts, and are discussed in Chapter 28.

5. Other Causes of Genetic Recombination

Because their integration sites in host DNA do not depend upon homology with the transposon ends, transposition is sometimes called "**illegitimate recombination.**" Although no homology is required, there are prefered sites for integration. For example, Tn10 is transposed most readily into certain "hot spots" in the *Salmonella* DNA among which is the sequence 5'-GC^{m5}CAGGC.[608] Illegitimate recombination can also be induced by other processes that involve DNA chain cleavage, e.g., by topoisomerases.[644] Whenever a DNA chain breaks, it must be repaired, a process that often also involves recombination. Recombination is often observed to occur between direct repeat sequences, which are a major cause of instability in the genome.[645]

Under some circumstances selected segments of the genome are **amplified** by repeated replication of a gene or genes.[646,647] Amplication of specific genes occurs in viruses,[648,649] in bacteria where it may provide