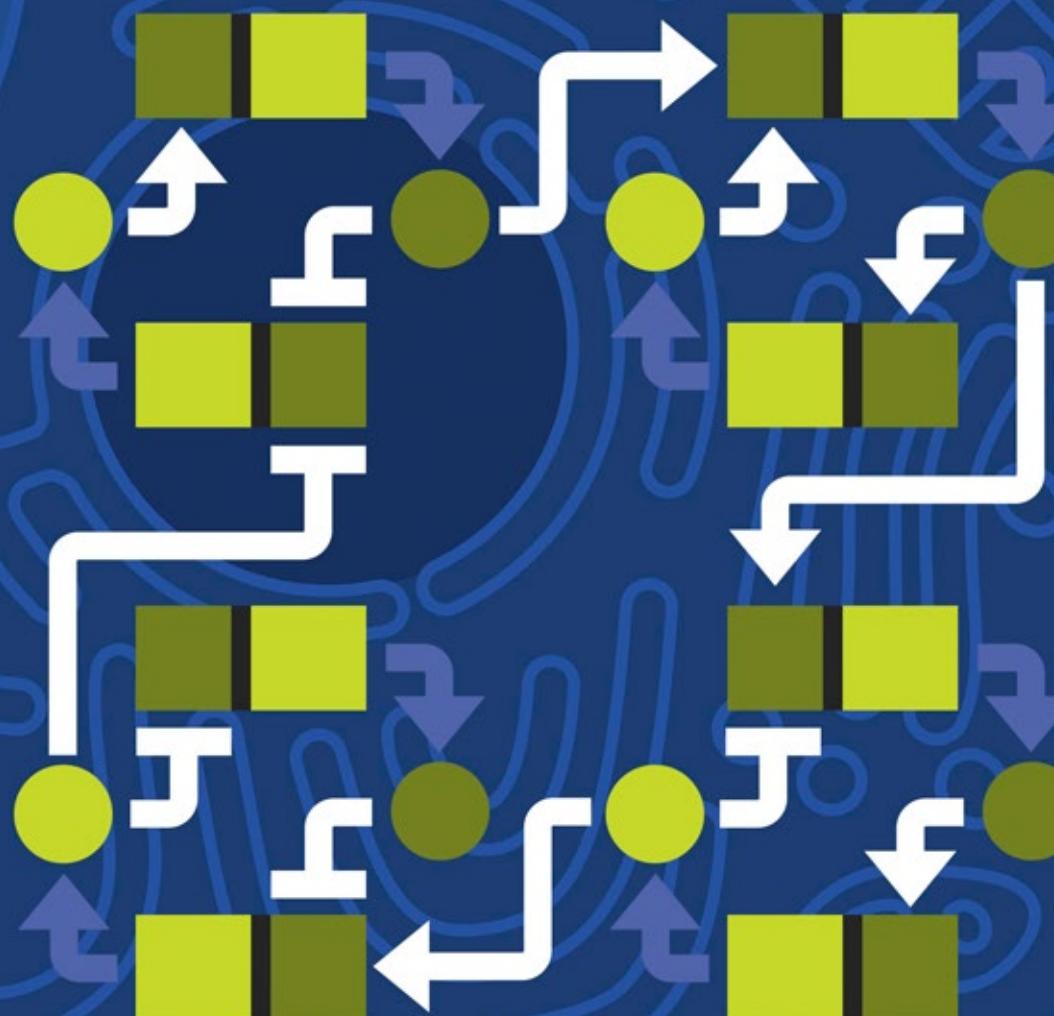


# Molecular Biology of **THE CELL**

Sixth Edition



ALBERTS

JOHNSON

LEWIS

MORGAN

RAFF

ROBERTS

WALTER

# WAYS OF WORKING WITH CELLS

## Analyzing Cells, Molecules, and Systems

### CHAPTER 8

Progress in science is often driven by advances in technology. The entire field of cell biology, for example, came into being when optical craftsmen learned to grind small lenses of sufficiently high quality to observe cells and their substructures. Innovations in lens grinding, rather than any conceptual or philosophical advance, allowed Hooke and van Leeuwenhoek to discover a previously unseen cellular world, where tiny creatures tumble and twirl in a small droplet of water ([Figure 8-1](#)).

The twenty-first century is a particularly exciting time for biology. New methods for analyzing cells, proteins, DNA, and RNA are fueling an information explosion and allowing scientists to study cells and their macromolecules in previously unimagined ways. We now have access to the sequences of many billions of nucleotides, providing the complete molecular blueprints for hundreds of organisms—from microbes and mustard weeds to worms, flies, mice, dogs, chimpanzees, and humans. And powerful new techniques are helping us to decipher that information, allowing us not only to compile huge, detailed catalogs of genes and proteins but also to begin to unravel how these components work together to form functional cells and organisms. The long-range goal is nothing short of obtaining a complete understanding of what takes place inside a cell as it responds to its environment and interacts with its neighbors.

In this chapter, we present some of the principal methods used to study cells and their molecular components. We consider how to separate cells of different types from tissues, how to grow cells outside the body, and how to disrupt cells and isolate their organelles and constituent macromolecules in pure form. We also present the techniques used to determine protein structure, function, and interactions, and we discuss the breakthroughs in DNA technology that continue to revolutionize our understanding of cell function. We end the chapter with an overview of some of the mathematical approaches that are helping us deal with the enormous complexity of cells. By considering cells as dynamic systems with many moving parts, mathematical approaches can reveal hidden insights into how the many components of cells work together to produce the special qualities of life.

#### IN THIS CHAPTER

[ISOLATING CELLS AND GROWING THEM IN CULTURE](#)

[PURIFYING PROTEINS](#)

[ANALYZING PROTEINS](#)

[ANALYZING AND MANIPULATING DNA](#)

[STUDYING GENE EXPRESSION AND FUNCTION](#)

[MATHEMATICAL ANALYSIS OF CELL FUNCTIONS](#)

## ISOLATING CELLS AND GROWING THEM IN CULTURE

Although the organelles and large molecules in a cell can be visualized with microscopes, understanding how these components function requires a detailed biochemical analysis. Most biochemical procedures require that large numbers of cells be physically disrupted to gain access to their components. If the sample is a piece of tissue, composed of different types of cells, heterogeneous cell populations will be mixed together. To obtain as much information as possible about the cells in a tissue, biologists have developed ways of dissociating cells from tissues and separating them according to type. These manipulations result in a relatively homogeneous population of cells that can then be analyzed—either directly or after their number has been greatly increased by allowing the cells to proliferate in culture.

### Cells Can Be Isolated from Tissues

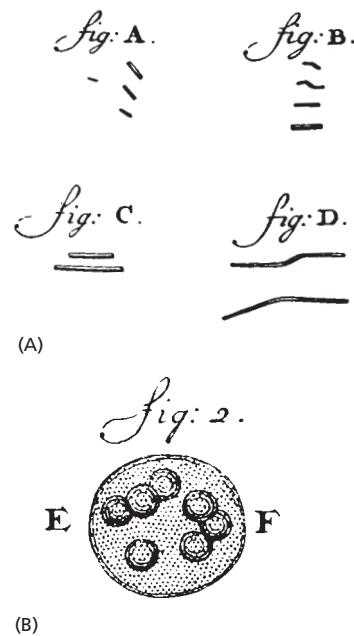
Intact tissues provide the most realistic source of material, as they represent the actual cells found within the body. The first step in isolating individual cells is to disrupt the extracellular matrix and cell-cell junctions that hold the cells together. For this purpose, a tissue sample is typically treated with proteolytic enzymes (such as trypsin and collagenase) to digest proteins in the extracellular matrix and with agents (such as ethylenediaminetetraacetic acid, or EDTA) that bind, or chelate, the  $\text{Ca}^{2+}$  on which cell-cell adhesion depends. The tissue can then be teased apart into single cells by gentle agitation.

For some biochemical preparations, the protein of interest can be obtained in sufficient quantity without having to separate the tissue or organ into cell types. Examples include the preparation of histones from calf thymus, actin from rabbit muscle, or tubulin from cow brain. In other cases, obtaining the desired protein requires enrichment for a specific cell type of interest. Several approaches are used to separate the different cell types from a mixed cell suspension. One of the most sophisticated cell-separation techniques uses an antibody coupled to a fluorescent dye to label specific cells. An antibody is chosen that specifically binds to the surface of only one cell type in the tissue. The labeled cells can then be separated from the unlabeled ones in a *fluorescence-activated cell sorter*. In this remarkable machine, individual cells traveling single file in a fine stream pass through a laser beam, and the fluorescence of each cell is rapidly measured. A vibrating nozzle generates tiny droplets, most containing either one cell or no cells. The droplets containing a single cell are automatically given a positive or a negative charge at the moment of formation, depending on whether the cell they contain is fluorescent; they are then deflected by a strong electric field into an appropriate container. Occasional clumps of cells, detected by their increased light scattering, are left uncharged and are discarded into a waste container. Such machines can accurately select 1 fluorescent cell from a pool of 1000 unlabeled cells and sort several thousand cells each second (**Figure 8–2**).

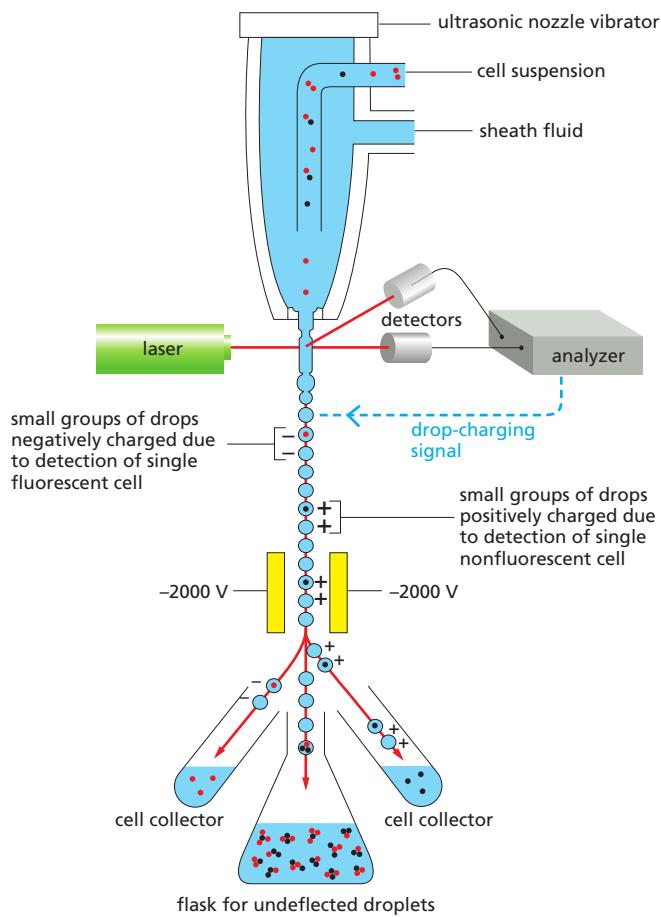
### Cells Can Be Grown in Culture

Although molecules can be extracted from whole tissues, this is often not the most convenient or useful source of material. The complexity of intact tissues and organs is an inherent disadvantage when trying to purify particular molecules. Cells grown in culture provide a more homogeneous population of cells from which to extract material, and they are also much more convenient to work with in the laboratory. Given appropriate surroundings, most plant and animal cells can live, multiply, and even express differentiated properties in a culture dish. The cells can be watched continuously under the microscope or analyzed biochemically, and the effects of adding or removing specific molecules, such as hormones or growth factors, can be systematically explored.

Experiments performed on cultured cells are sometimes said to be carried out *in vitro* (literally, “in glass”) to contrast them with experiments using intact organisms, which are said to be carried out *in vivo* (literally, “in the living organism”).



**Figure 8–1** Microscopic life. A sample of “diverse animalcules” seen by van Leeuwenhoek using his simple microscope. (A) Bacteria seen in material he excavated from between his teeth. Those in fig. B he described as “swimming first forward and then backwards” (1692). (B) The eukaryotic green alga *Volvox* (1700). (Courtesy of the John Innes Foundation.)



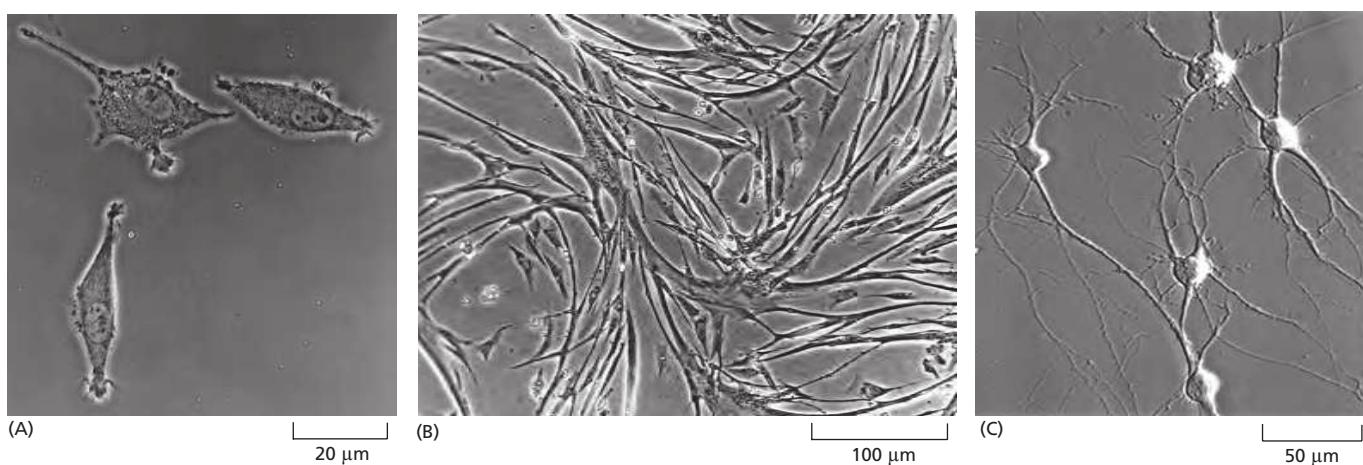
**Figure 8–2 A fluorescence-activated cell sorter.** A cell passing through the laser beam is monitored for fluorescence. Droplets containing single cells are given a negative or positive charge, depending on whether the cell is fluorescent or not. The droplets are then deflected by an electric field into collection tubes according to their charge. Note that the cell concentration must be adjusted so that most droplets contain no cells and flow to a waste container together with any cell clumps.

These terms can be confusing, however, because they are often used in a very different sense by biochemists. In the biochemistry lab, *in vitro* refers to reactions carried out in a test tube in the absence of living cells, whereas *in vivo* refers to any reaction taking place inside a living cell, even if that cell is growing in culture.

Tissue culture began in 1907 with an experiment designed to settle a controversy in neurobiology. The hypothesis under examination was known as the neuronal doctrine, which states that each nerve fiber is the outgrowth of a single nerve cell and not the product of the fusion of many cells. To test this contention, small pieces of spinal cord were placed on clotted tissue fluid in a warm, moist chamber and observed at regular intervals under the microscope. After a day or so, individual nerve cells could be seen extending long, thin filaments (axons) into the clot. Thus, the neuronal doctrine received strong support, and the foundation was laid for the cell-culture revolution.

These original experiments on nerve fibers used cultures of small tissue fragments called explants. Today, cultures are more commonly made from suspensions of cells dissociated from tissues. Unlike bacteria, most tissue cells are not adapted to living suspended in fluid and require a solid surface on which to grow and divide. For cell cultures, this support is usually provided by the surface of a plastic culture dish. Cells vary in their requirements, however, and many do not proliferate or differentiate unless the culture dish is coated with materials that cells like to adhere to, such as polylysine or extracellular matrix components.

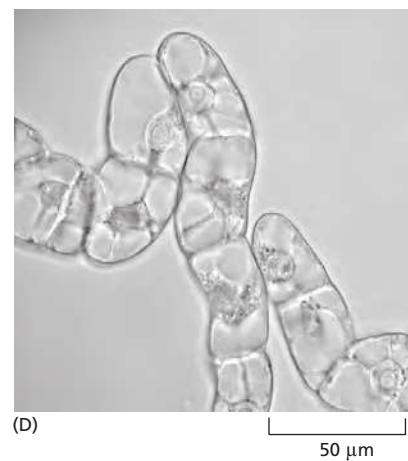
Cultures prepared directly from the tissues of an organism are called *primary cultures*. These can be made with or without an initial fractionation step to separate different cell types. In most cases, cells in primary cultures can be removed from the culture dish and recultured repeatedly in so-called secondary cultures; in this way, they can be repeatedly subcultured (*passaged*) for weeks or months. Such cells often display many of the differentiated properties appropriate to their



**Figure 8–3** Light micrographs of cells in culture. (A) Mouse fibroblasts. (B) Chick myoblasts fusing to form multinucleate muscle cells. (C) Purified rat retinal ganglion nerve cells. (D) Tobacco cells in liquid culture. (A, courtesy of Daniel Zicha; B, courtesy of Rosalind Zalin; C, from A. Meyer-Franke et al., *Neuron* 15:805–819, 1995. With permission from Elsevier; D, courtesy of Gethin Roberts.)

origin (Figure 8–3): fibroblasts continue to secrete collagen; cells derived from embryonic skeletal muscle fuse to form muscle fibers that contract spontaneously in the culture dish; nerve cells extend axons that are electrically excitable and make synapses with other nerve cells; and epithelial cells form extensive sheets with many of the properties of an intact epithelium. Because these properties are maintained in culture, they are accessible to study in ways that are often not possible in intact tissues.

Cell culture is not limited to animal cells. When a piece of plant tissue is cultured in a sterile medium containing nutrients and appropriate growth regulators, many of the cells are stimulated to proliferate indefinitely in a disorganized manner, producing a mass of relatively undifferentiated cells called a *callus*. If the nutrients and growth regulators are carefully manipulated, one can induce the formation of a shoot and then root apical meristems within the callus, and, in many species, regenerate a whole new plant. Similar to animal cells, callus cultures can be mechanically dissociated into single cells, which will grow and divide as a suspension culture (see Figure 8–3D).



### Eukaryotic Cell Lines Are a Widely Used Source of Homogeneous Cells

The cell cultures obtained by disrupting tissues tend to suffer from a problem—eventually the cells die. Most vertebrate cells stop dividing after a finite number of cell divisions in culture, a process called *replicative cell senescence* (discussed in Chapter 17). Normal human fibroblasts, for example, typically divide only 25–40 times in culture before they stop. In these cells, the limited proliferation capacity reflects a progressive shortening and uncapping of the cell's telomeres, the repetitive DNA sequences and associated proteins that cap the ends of each chromosome (discussed in Chapter 5). Human somatic cells in the body have turned off production of the enzyme, called *telomerase*, that normally maintains the telomeres, which is why their telomeres shorten with each cell division. Human fibroblasts can often be coaxed to proliferate indefinitely by providing them with the gene that encodes the catalytic subunit of telomerase; in this case, they can be propagated as an “immortalized” cell line.

Some human cells, however, cannot be immortalized by this trick. Although their telomeres remain long, they still stop dividing after a limited number of divisions because culture conditions cause excessive mitogenic stimulation, which

activates a poorly understood protective mechanism (discussed in Chapter 17) that stops cell division—a process sometimes called “culture shock.” To immortalize these cells, one has to do more than introduce telomerase. One must also inactivate the protective mechanisms, which can be done by introducing certain cancer-promoting oncogenes (discussed in Chapter 20). Unlike human cells, most rodent cells do not turn off production of telomerase and therefore their telomeres do not shorten with each cell division. Therefore, if culture shock can be avoided, some rodent cell types will divide indefinitely in culture. In addition, rodent cells often undergo spontaneous genetic changes in culture that inactivate their protective mechanisms, thereby producing immortalized cell lines.

Cell lines can often be most easily generated from cancer cells, but these cultures—referred to as *transformed cell lines*—differ from those prepared from normal cells in several ways. Transformed cell lines often grow without attaching to a surface, for example, and they can proliferate to a much higher density in a culture dish. Similar properties can be induced experimentally in normal cells by transforming them with a tumor-inducing virus or chemical. The resulting transformed cell lines can usually cause tumors if injected into a susceptible animal.

Transformed and nontransformed cell lines are extremely useful in cell research as sources of very large numbers of cells of a uniform type, especially since they can be stored in liquid nitrogen at -196°C for an indefinite period and retain their viability when thawed. It is important to keep in mind, however, that cell lines nearly always differ in important ways from their normal progenitors in the tissues from which they were derived.

Some widely used cell lines are listed in **Table 8–1**. Different lines have different advantages; for example, the PtK epithelial cell lines derived from the rat

**TABLE 8–1 Some Commonly Used Cell Lines**

Cell line*	Cell type and origin
3T3	Fibroblast (mouse)
BHK21	Fibroblast (Syrian hamster)
MDCK	Epithelial cell (dog)
HeLa	Epithelial cell (human)
PtK1	Epithelial cell (rat kangaroo)
L6	Myoblast (rat)
PC12	Chromaffin cell (rat)
SP2	Plasma cell (mouse)
COS	Kidney (monkey)
293	Kidney (human); transformed with adenovirus
CHO	Ovary (Chinese hamster)
DT40	Lymphoma cell for efficient targeted recombination (chick)
R1	Embryonic stem cell (mouse)
E14.1	Embryonic stem cell (mouse)
H1, H9	Embryonic stem cell (human)
S2	Macrophage-like cell ( <i>Drosophila</i> )
BY2	Undifferentiated meristematic cell (tobacco)

\*Many of these cell lines were derived from tumors. All of them are capable of indefinite replication in culture and express at least some of the special characteristics of their cells of origin.

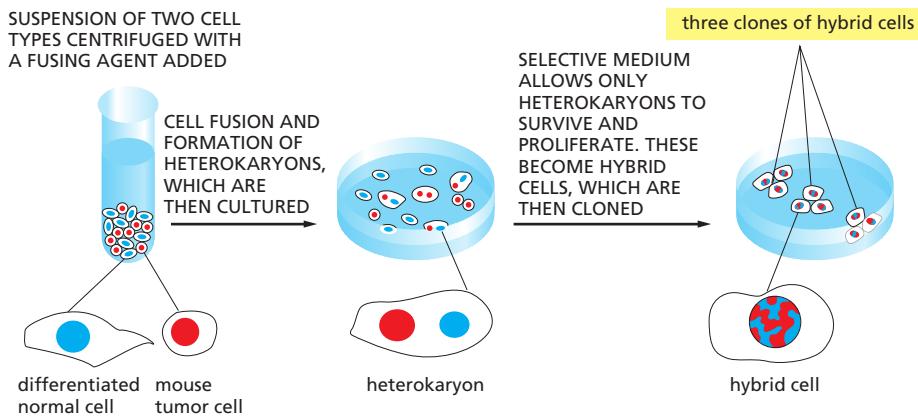
kangaroo, unlike many other cell types, remain flat during mitosis, allowing the mitotic apparatus to be readily observed in action.

### Hybridoma Cell Lines Are Factories That Produce Monoclonal Antibodies

As we see throughout this book, antibodies are particularly useful tools for cell biology. Their great specificity allows precise visualization of selected proteins among the many thousands that each cell typically produces. Antibodies are often produced by inoculating animals with the protein of interest and subsequently isolating the antibodies specific to that protein from the serum of the animal. However, only limited quantities of antibodies can be obtained from a single inoculated animal, and the antibodies produced will be a heterogeneous mixture of antibodies that recognize a variety of different antigenic sites on a macromolecule that differs from animal to animal. Moreover, antibodies specific for the antigen will constitute only a fraction of the antibodies found in the serum. An alternative technology, which allows the production of an unlimited quantity of identical antibodies and greatly increases the specificity and convenience of antibody-based methods, is the production of monoclonal antibodies by hybridoma cell lines.

This technology, developed in 1975, revolutionized the production of antibodies for use as tools in cell biology, as well as for the diagnosis and treatment of certain diseases, including rheumatoid arthritis and cancer. The procedure requires hybrid cell technology (Figure 8–4), and it involves propagating a clone of cells from a single antibody-secreting B lymphocyte to obtain a homogeneous preparation of antibodies in large quantities. B lymphocytes normally have a limited life-span in culture, but individual antibody-producing B lymphocytes from an immunized mouse, when fused with cells derived from a transformed B lymphocyte cell line, can give rise to hybrids that have both the ability to make a particular antibody and the ability to multiply indefinitely in culture. These **hybridomas** are propagated as individual clones, each of which provides a permanent and stable source of a single type of **monoclonal antibody**. Each type of monoclonal antibody recognizes a single type of antigenic site—for example, a particular cluster of five or six amino acid side chains on the surface of a protein. Their uniform specificity makes monoclonal antibodies much more useful than conventional antisera for many purposes.

An important advantage of the hybridoma technique is that monoclonal antibodies can be made against molecules that constitute only a minor component of a complex mixture. In an ordinary antiserum made against such a mixture, the proportion of antibody molecules that recognize the minor component would be too small to be useful. But if the B lymphocytes that produce the various components of this antiserum are made into hybridomas, it becomes possible to screen individual hybridoma clones from the large mixture to select one that produces the desired type of monoclonal antibody and to propagate the selected hybridoma



**Figure 8–4** The production of hybrid cells. It is possible to fuse one cell with another to form a *heterokaryon*, a combined cell with two separate nuclei. Typically, a suspension of cells is treated with certain inactivated viruses or with polyethylene glycol, each of which alters the plasma membranes of cells in a way that induces them to fuse. Eventually, a heterokaryon proceeds to mitosis and produces a hybrid cell in which the two separate nuclear envelopes have been disassembled, allowing all the chromosomes to be brought together in a single large nucleus. Such hybrid cells can give rise to immortal hybrid cell lines. If one of the parent cells was from a tumor cell line, the hybrid cell is called a hybridoma.

indefinitely so as to produce that antibody in unlimited quantities. In principle, therefore, a monoclonal antibody can be made against any protein in a biological sample. Once an antibody has been made, it can be used to localize the protein in cells and tissues, to follow its movement, and to purify the protein to study its structure and function.

### Summary

*Tissues can be dissociated into their component cells, from which individual cell types can be purified and used for biochemical analysis or for the establishment of cell cultures. Many animal and plant cells survive and proliferate in a culture dish if they are provided with a suitable culture medium containing nutrients and appropriate signal molecules. Although many animal cells stop dividing after a finite number of cell divisions, cells that have been immortalized through spontaneous mutations or genetic manipulation can be maintained indefinitely as cell lines. Hybridoma cells are widely employed to produce unlimited quantities of uniform monoclonal antibodies, which are used to detect and purify cell proteins, as well as to diagnose and treat diseases.*

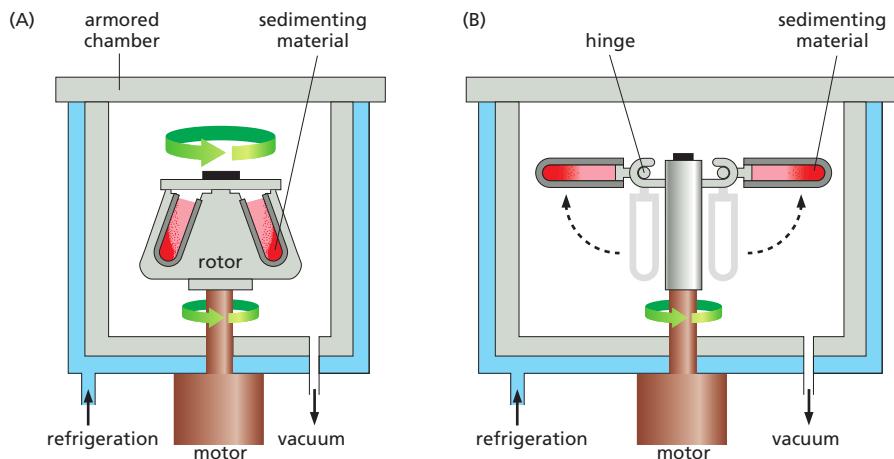
## PURIFYING PROTEINS

The challenge of isolating a single type of protein from the thousands of other proteins present in a cell is a formidable one, but must be overcome in order to study protein function *in vitro*. As we shall see later in this chapter, *recombinant DNA technology* can enormously simplify this task by “tricking” cells into producing large quantities of a given protein, thereby making its purification much easier. Whether the source of the protein is an engineered cell or a natural tissue, a purification procedure usually starts with subcellular fractionation to reduce the complexity of the material, and is then followed by purification steps of increasing specificity.

### Cells Can Be Separated into Their Component Fractions

To purify a protein, it must first be extracted from inside the cell. Cells can be broken up in various ways: they can be subjected to osmotic shock or ultrasonic vibration, forced through a small orifice, or ground up in a blender. These procedures break many of the membranes of the cell (including the plasma membrane and endoplasmic reticulum) into fragments that immediately reseal to form small closed vesicles. If carefully carried out, however, the disruption procedures leave organelles such as nuclei, mitochondria, the Golgi apparatus, lysosomes, and peroxisomes largely intact. The suspension of cells is thereby reduced to a thick slurry (called a *homogenate* or *extract*) that contains a variety of membrane-enclosed organelles, each with a distinctive size, charge, and density. Provided that the homogenization medium has been carefully chosen (by trial and error for each organelle), the various components—including the vesicles derived from the endoplasmic reticulum, called microsomes—retain most of their original biochemical properties.

The different components of the homogenate must then be separated. Such cell fractionations became possible only after the commercial development in the early 1940s of an instrument known as the *preparative ultracentrifuge*, which rotates extracts of broken cells at high speeds (**Figure 8–5**). This treatment separates cell components by size and density: in general, the largest objects experience the largest centrifugal force and move the most rapidly. At relatively low speed, large components such as nuclei sediment to form a pellet at the bottom of the centrifuge tube; at slightly higher speed, a pellet of mitochondria is deposited; and at even higher speeds and with longer periods of centrifugation, first the small closed vesicles and then the ribosomes can be collected (**Figure 8–6**). All of these fractions are impure, but many of the contaminants can be removed by resuspending the pellet and repeating the centrifugation procedure several times.

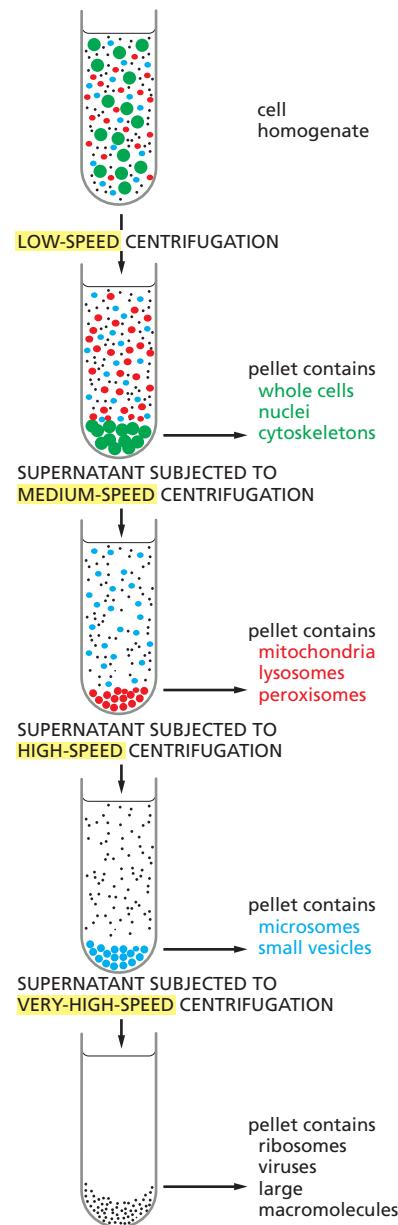


**Figure 8–5 The preparative ultracentrifuge.** (A) The sample is contained in tubes that are inserted into a ring of angled cylindrical holes in a metal rotor. Rapid rotation of the rotor generates enormous centrifugal forces, which cause particles in the sample to sediment against the bottom sides of the sample tubes, as shown here. The vacuum reduces friction, preventing heating of the rotor and allowing the refrigeration system to maintain the sample at 4°C. (B) Some fractionation methods require a different type of rotor called a *swinging-bucket rotor*. In this case, the sample tubes are placed in metal tubes on hinges that allow the tubes to swing outward when the rotor spins. Sample tubes are therefore horizontal during spinning, and samples are sedimented toward the bottom, not the sides, of the tube, providing better separation of differently sized components (see Figures 8–6 and 8–7).

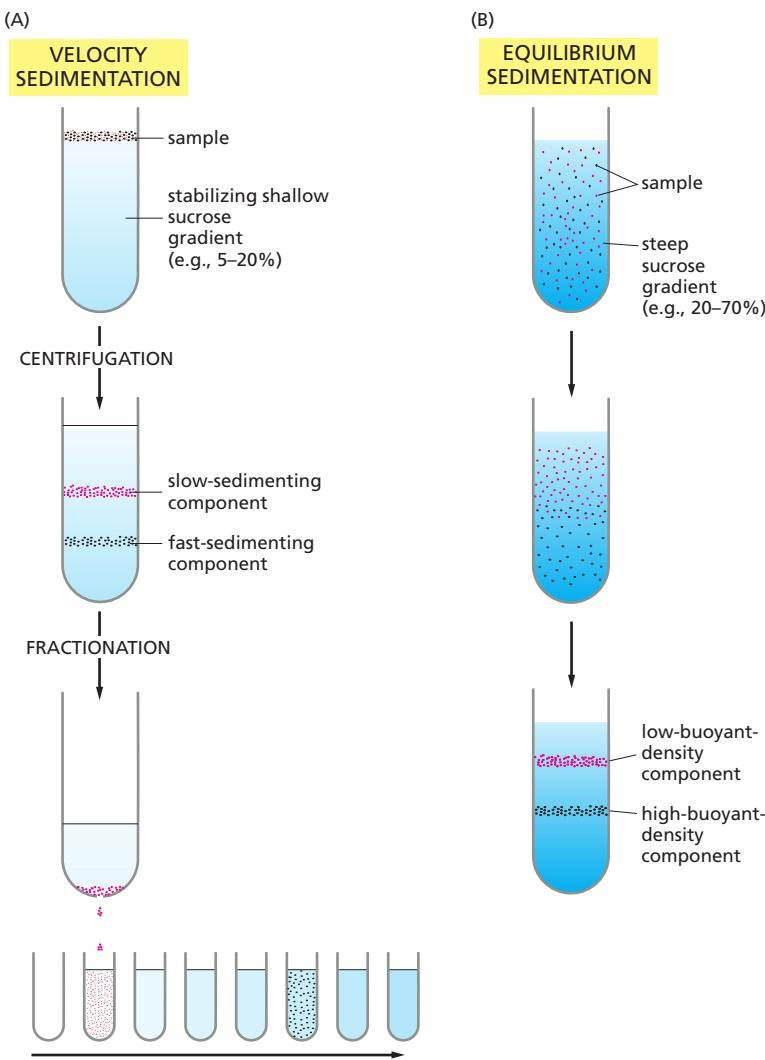
Centrifugation is the first step in most fractionations, but it separates only components that differ greatly in size. A finer degree of separation can be achieved by layering the homogenate in a thin band on top of a salt solution that fills a centrifuge tube. When centrifuged, the various components in the mixture move as a series of distinct bands through the solution, each at a different rate, in a process called *velocity sedimentation* (Figure 8–7A). For the procedure to work effectively, the bands must be protected from convective mixing, which would normally occur whenever a denser solution (for example, one containing organelles) finds itself on top of a lighter one (the salt solution). This is achieved by augmenting the solution in the tube with a shallow gradient of sucrose prepared by a special mixing device. The resulting density gradient—with the dense end at the bottom of the tube—keeps each region of the solution denser than any solution above it, and it thereby prevents convective mixing from distorting the separation.

When sedimented through sucrose gradients, different cell components separate into distinct bands that can be collected individually. The relative rate at which each component sediments depends primarily on its size and shape—normally being described in terms of its sedimentation coefficient, or S value. Present-day ultracentrifuges rotate at speeds of up to 80,000 rpm and produce forces as high as 500,000 times gravity. These enormous forces drive even small macromolecules, such as tRNA molecules and simple enzymes, to sediment at an appreciable rate and allow them to be separated from one another by size.

The ultracentrifuge is also used to separate cell components on the basis of their buoyant density, independently of their size and shape. In this case, the



**Figure 8–6 Cell fractionation by centrifugation.** Repeated centrifugation at progressively higher speeds will fractionate homogenates of cells into their components. In general, the smaller the subcellular component, the greater the centrifugal force required to sediment it. Typical values for the various centrifugation steps referred to in the figure are:  
*low speed:* 1000 times gravity for 10 minutes  
*medium speed:* 20,000 times gravity for 20 minutes  
*high speed:* 80,000 times gravity for 1 hour  
*very high speed:* 150,000 times gravity for 3 hours



**Figure 8–7** Comparison of velocity sedimentation and equilibrium sedimentation. (A) In velocity sedimentation, subcellular components sediment at different speeds according to their size and shape when layered over a solution containing sucrose. To stabilize the sedimenting bands against convective mixing caused by small differences in temperature or solute concentration, the tube contains a continuous shallow gradient of sucrose, which increases in concentration toward the bottom of the tube (typically from 5% to 20% sucrose). After centrifugation, the different components can be collected individually, most simply by puncturing the plastic centrifuge tube with a needle and collecting drops from the bottom, as illustrated here. (B) In equilibrium sedimentation, subcellular components move up or down when centrifuged in a gradient until they reach a position where their density matches their surroundings. Although a sucrose gradient is shown here, denser gradients, which are especially useful for protein and nucleic acid separation, can be formed from cesium chloride. The final bands, at equilibrium, can be collected as in (A).

sample is sedimented through a steep density gradient that contains a very high concentration of sucrose or cesium chloride. Each cell component begins to move down the gradient as in Figure 8–7A, but it eventually reaches a position where the density of the solution is equal to its own density. At this point, the component floats and can move no farther. A series of distinct bands is thereby produced in the centrifuge tube, with the bands closest to the bottom of the tube containing the components of highest buoyant density (Figure 8–7B). This method, called *equilibrium sedimentation*, is so sensitive that it can separate macromolecules that have incorporated heavy isotopes, such as  $^{13}\text{C}$  or  $^{15}\text{N}$ , from the same macromolecules that contain the lighter, common isotopes ( $^{12}\text{C}$  or  $^{14}\text{N}$ ). In fact, the cesium-chloride method was developed in 1957 to separate the labeled from the unlabeled DNA produced after exposure of a growing population of bacteria to nucleotide precursors containing  $^{15}\text{N}$ ; this classic experiment provided direct evidence for the semiconservative replication of DNA (see Figure 5–5).

### Cell Extracts Provide Accessible Systems to Study Cell Functions

Studies of organelles and other large subcellular components isolated in the ultracentrifuge have contributed enormously to our understanding of the functions of different cell components. Experiments on mitochondria and chloroplasts purified by centrifugation, for example, demonstrated the central function of these organelles in converting energy into forms that the cell can use. Similarly, resealed vesicles formed from fragments of rough and smooth endoplasmic reticulum

(microsomes) have been separated from each other and analyzed as functional models of these compartments of the intact cell.

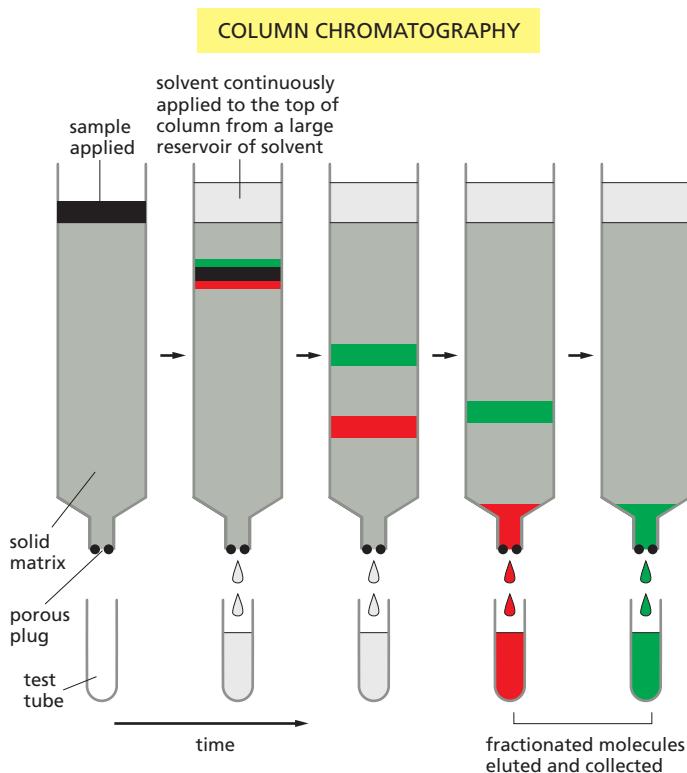
Similarly, highly concentrated cell extracts, especially extracts of *Xenopus laevis* (African clawed frog) oocytes, have played a critical role in the study of such complex and highly organized processes as the cell-division cycle, the separation of chromosomes on the mitotic spindle, and the vesicular-transport steps involved in the movement of proteins from the endoplasmic reticulum through the Golgi apparatus to the plasma membrane.

Cell extracts also provide, in principle, the starting material for the complete separation of all of the individual macromolecular components of the cell. We now consider how this separation is achieved, focusing on proteins.

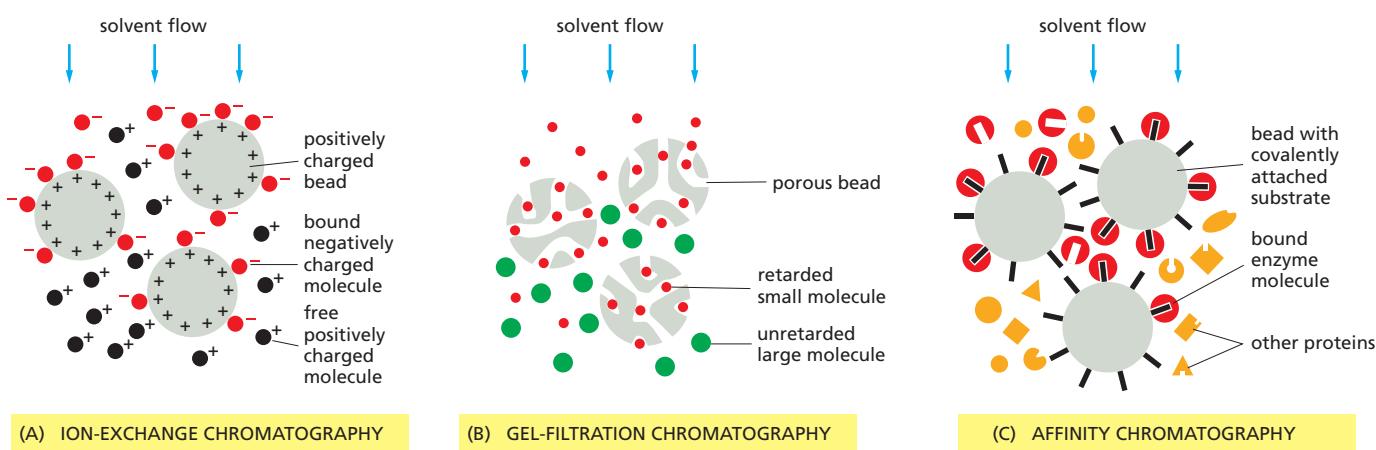
### Proteins Can Be Separated by Chromatography

Proteins are most often fractionated by **column chromatography**, in which a mixture of proteins in solution is passed through a column containing a porous solid matrix. Different proteins are retarded to different extents by their interaction with the matrix, and they can be collected separately as they flow out of the bottom of the column (Figure 8–8). Depending on the choice of matrix, proteins can be separated according to their charge (*ion-exchange chromatography*), their hydrophobicity (*hydrophobic chromatography*), their size (*gel-filtration chromatography*), or their ability to bind to particular small molecules or to other macromolecules (*affinity chromatography*).

Many types of matrices are available. Ion-exchange columns (Figure 8–9A) are packed with small beads that carry either a positive or a negative charge, so that proteins are fractionated according to the arrangement of charges on their surface. Hydrophobic columns are packed with beads from which hydrophobic side chains protrude, selectively retarding proteins with exposed hydrophobic regions. Gel-filtration columns (Figure 8–9B), which separate proteins according to their size, are packed with tiny porous beads: molecules that are small enough to enter the pores linger inside successive beads as they pass down the column, while larger molecules remain in the solution flowing between the beads and therefore move more rapidly, emerging from the column first. Besides providing



**Figure 8–8** The separation of molecules by column chromatography. The sample, a solution containing a mixture of different molecules, is applied to the top of a cylindrical glass or plastic column filled with a permeable solid matrix, such as cellulose. A large amount of solvent is then passed slowly through the column and collected in separate tubes as it emerges from the bottom. Because various components of the sample travel at different rates through the column, they are fractionated into different tubes.



a means of separating molecules, gel-filtration chromatography is a convenient way to estimate their size.

Affinity chromatography (Figure 8–9C) takes advantage of the biologically important binding interactions that occur on protein surfaces. If a substrate molecule is covalently coupled to an inert matrix such as a polysaccharide bead, the enzyme that operates on that substrate will often be specifically retained by the matrix and can then be eluted (washed out) in nearly pure form. Likewise, short DNA oligonucleotides of a specifically designed sequence can be immobilized in this way and used to purify DNA-binding proteins that normally recognize this sequence of nucleotides in chromosomes. Alternatively, specific antibodies can be coupled to a matrix to purify protein molecules recognized by the antibodies. Because of the great specificity of all such affinity columns, 1000- to 10,000-fold purifications can sometimes be achieved in a single pass.

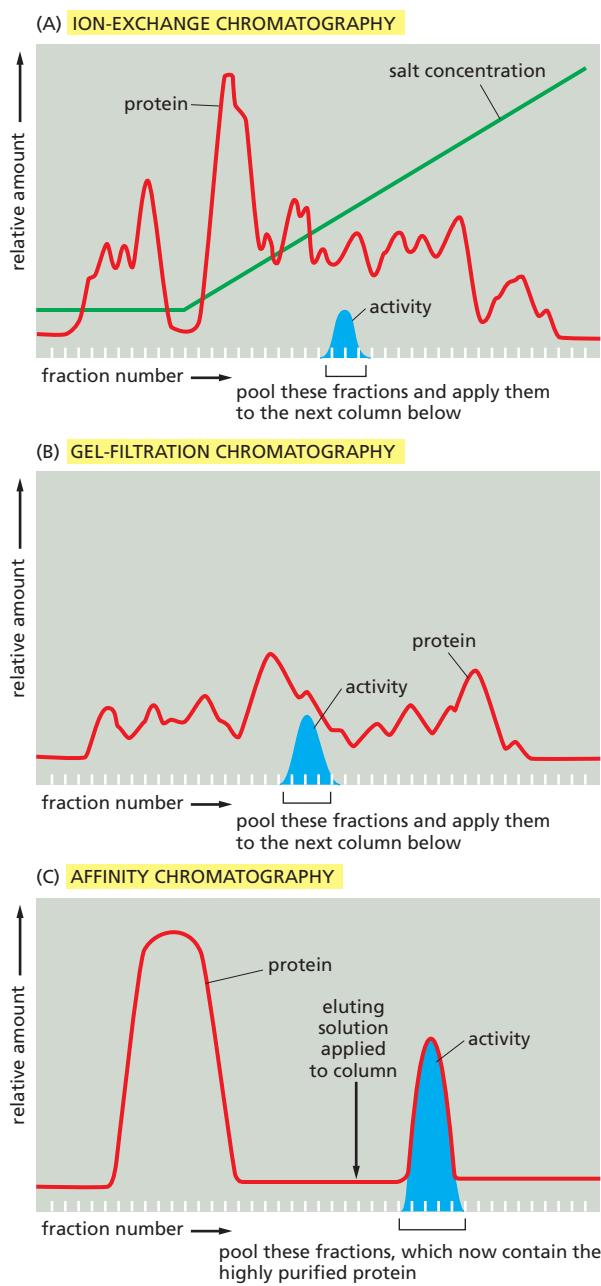
If one starts with a complex mixture of proteins, a single passage through an ion-exchange or a gel-filtration column does not produce very highly purified fractions, since these methods individually increase the proportion of a given protein in the mixture no more than twentyfold. Because most individual proteins represent less than 1/1000 of the total cell protein, it is usually necessary to use several different types of columns in succession to attain sufficient purity, with affinity chromatography being the most efficient (Figure 8–10).

Inhomogeneities in the matrices (such as cellulose), which cause an uneven flow of solvent through the column, limit the resolution of conventional column chromatography. Special chromatography resins (usually silica-based) composed of tiny spheres (3–10  $\mu\text{m}$  in diameter) can be packed with a special apparatus to form a uniform column bed. Such **high-performance liquid chromatography (HPLC)** columns attain a high degree of resolution. In HPLC, the solutes equilibrate very rapidly with the interior of the tiny spheres, and so solutes with different affinities for the matrix are efficiently separated from one another even at very fast flow rates. HPLC is therefore the method of choice for separating many proteins and small molecules.

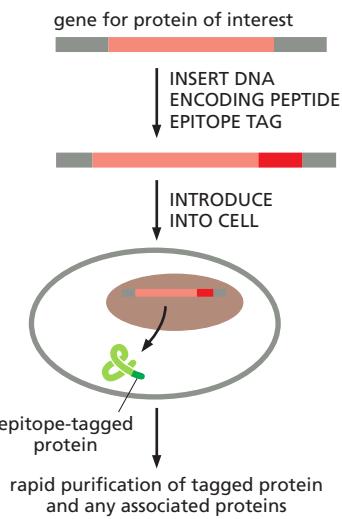
### Immunoprecipitation Is a Rapid Affinity Purification Method

Immunoprecipitation is a useful variation on the theme of affinity chromatography. Specific antibodies that recognize the protein to be purified are attached to small agarose beads. Rather than being packed into a column, as in affinity chromatography, a small quantity of the antibody-coated beads is simply added to a protein extract in a test tube and mixed in suspension for a short period of time—thereby allowing the antibodies to bind the desired protein. The beads are then collected by low-speed centrifugation, and the unbound proteins in the supernatant are discarded. This method is commonly used to purify small amounts of enzymes from cell extracts for analysis of enzymatic activity or for studies of associated proteins.

**Figure 8–9** Three types of matrices used for chromatography. (A) In ion-exchange chromatography, the insoluble matrix carries ionic charges that retard the movement of molecules of opposite charge. Matrices used for separating proteins include diethylaminoethylcellulose (DEAE-cellulose), which is positively charged, and carboxymethylcellulose (CM-cellulose) and phosphocellulose, which are negatively charged. Analogous matrices based on agarose or other polymers are also frequently used. The strength of the association between the dissolved molecules and the ion-exchange matrix depends on both the ionic strength and the pH of the solution that is passing down the column, which may therefore be varied systematically (as in Figure 8–10) to achieve an effective separation. (B) In gel-filtration chromatography, the small beads that form the matrix are inert but porous. Molecules that are small enough to penetrate into the matrix beads are thereby delayed and travel more slowly through the column than larger molecules that cannot penetrate. Beads of cross-linked polysaccharide (dextran, agarose, or acrylamide) are available commercially in a wide range of pore sizes, making them suitable for the fractionation of molecules of various molecular mass, from less than 500 daltons to more than  $5 \times 10^6$  daltons. (C) Affinity chromatography uses an insoluble matrix that is covalently linked to a specific ligand, such as an antibody molecule or an enzyme substrate, that will bind a specific protein. Enzyme molecules that bind to immobilized substrates on such columns can be eluted with a concentrated solution of the free form of the substrate molecule, while molecules that bind to immobilized antibodies can be eluted by dissociating the antibody–antigen complex with concentrated salt solutions or solutions of high or low pH. High degrees of purification can be achieved in a single pass through an affinity column.



**Figure 8-10 Protein purification by chromatography.** Typical results obtained when three different chromatographic steps are used in succession to purify a protein. In this example, a homogenate of cells was first fractionated by allowing it to percolate through an ion-exchange resin packed into a column (A). The column was washed to remove all unbound contaminants, and the bound proteins were then eluted by pouring a solution containing a gradually increasing concentration of salt onto the top of the column. Proteins with the lowest affinity for the ion-exchange resin passed directly through the column and were collected in the earliest fractions eluted from the bottom of the column. The remaining proteins were eluted in sequence according to their affinity for the resin—those proteins binding most tightly to the resin requiring the highest concentration of salt to remove them. The protein of interest was eluted in several fractions and was detected by its enzymatic activity. The fractions with activity were pooled and then applied to a gel-filtration column (B). The elution position of the still-impure protein was again determined by its enzymatic activity, and the active fractions were pooled and purified to homogeneity on an affinity column (C) that contained an immobilized substrate of the enzyme.



### Genetically Engineered Tags Provide an Easy Way to Purify Proteins

Using the recombinant DNA methods discussed in subsequent sections, any gene can be modified to produce its protein with a special recognition tag attached to it, so as to make subsequent purification of the protein simple and rapid. Often the recognition tag is itself an antigenic determinant, or *epitope*, which can be recognized by a highly specific antibody. The antibody can then be used to purify the protein by affinity chromatography or immunoprecipitation (Figure 8-11). Other types of tags are specifically designed for protein purification. For example, a repeated sequence of the amino acid histidine binds to certain metal ions, including nickel and copper. If genetic engineering techniques are used to attach a short string of histidines to one end of a protein, the slightly modified protein can be retained selectively on an affinity column containing immobilized nickel ions. Metal affinity chromatography can thereby be used to purify the modified protein from a complex molecular mixture.

**Figure 8-11 Epitope tagging for the purification of proteins.** Using standard genetic engineering techniques, a short peptide tag can be added to a protein of interest. If the tag is itself an antigenic determinant, or *epitope*, it can be targeted by an appropriate antibody, which can be used to purify the protein by immunoprecipitation or affinity chromatography.

In other cases, an entire protein is used as the recognition tag. When cells are engineered to synthesize the small enzyme glutathione S-transferase (GST) attached to a protein of interest, the resulting **fusion protein** can be purified from the other contents of the cell with an affinity column containing glutathione, a substrate molecule that binds specifically and tightly to GST.

As a further refinement of purification methods using recognition tags, an amino acid sequence that forms a cleavage site for a highly specific proteolytic enzyme can be engineered between the protein of choice and the recognition tag. Because the amino acid sequences at the cleavage site are very rarely found by chance in proteins, the tag can later be cleaved off without destroying the purified protein.

This type of specific cleavage is used in an especially powerful purification methodology known as *tandem affinity purification tagging* (*TAP-tagging*). Here, one end of a protein is engineered to contain two recognition tags that are separated by a protease cleavage site. The tag on the very end of the construct is chosen to bind irreversibly to an affinity column, allowing the column to be washed extensively to remove all contaminating proteins. Protease cleavage then releases the protein, which is then further purified using the second tag. Because this two-step strategy provides an especially high degree of protein purification with relatively little effort, it is used extensively in cell biology. Thus, for example, a set of approximately 6000 yeast strains, each with a different gene fused to DNA that encodes a TAP-tag, has been constructed to allow any yeast protein to be rapidly purified.

### Purified Cell-free Systems Are Required for the Precise Dissection of Molecular Functions

**Purified cell-free systems** provide a means of studying biological processes free from all of the complex side reactions that occur in a living cell. To make this possible, cell homogenates are fractionated with the aim of purifying each of the individual macromolecules that are needed to catalyze a biological process of interest. For example, the experiments to decipher the mechanisms of protein synthesis began with a cell homogenate that could translate RNA molecules to produce proteins. Fractionation of this homogenate, step by step, produced in turn the ribosomes, tRNAs, and various enzymes that together constitute the protein-synthetic machinery. Once individual pure components were available, each could be added or withheld separately to define its exact role in the overall process.

A major goal for cell biologists is the reconstitution of every biological process in a purified cell-free system. Only in this way can we define all of the components needed for the process and control their concentrations, which is required to work out their precise mechanism of action. Although much remains to be done, a great deal of what we know today about the molecular biology of the cell has been discovered by studies in such cell-free systems. They have been used, for example, to decipher the molecular details of DNA replication and DNA transcription, RNA splicing, protein translation, muscle contraction, and particle transport along microtubules, and many other processes that occur in cells.

### Summary

*Populations of cells can be analyzed biochemically by disrupting them and fractionating their contents, allowing functional cell-free systems to be developed. Highly purified cell-free systems are needed for determining the molecular details of complex cell processes, and the development of such systems requires extensive purification of all the proteins and other components involved. The proteins in soluble cell extracts can be purified by column chromatography; depending on the type of column matrix, biologically active proteins can be separated on the basis of their molecular weight, hydrophobicity, charge characteristics, or affinity for other molecules. In a typical purification, the sample is passed through several different columns in turn, with the enriched fractions obtained from one column being applied to the next. Recombinant DNA techniques (described later) allow special recognition tags to be attached to proteins, thereby greatly simplifying their purification.*

## ANALYZING PROTEINS

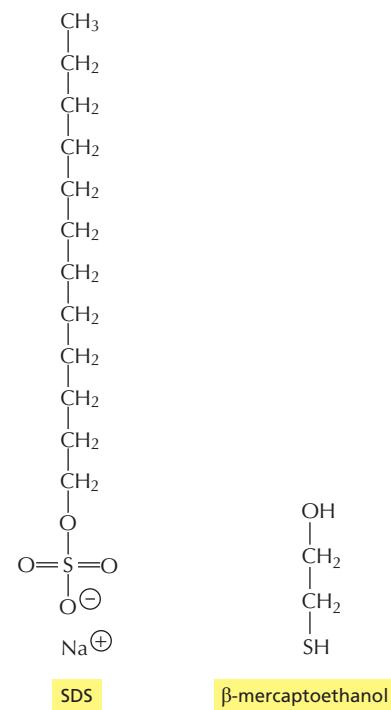
Proteins perform most cellular processes: they catalyze metabolic reactions, use nucleotide hydrolysis to do mechanical work, and serve as the major structural elements of the cell. The great variety of protein structures and functions has stimulated the development of a multitude of techniques to study them.

### Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis

Proteins usually possess a net positive or negative charge, depending on the mixture of charged amino acids they contain. An electric field applied to a solution containing a protein molecule causes the protein to migrate at a rate that depends on its net charge and on its size and shape. The most popular application of this property is **SDS polyacrylamide-gel electrophoresis (SDS-PAGE)**. It uses a highly cross-linked gel of polyacrylamide as the inert matrix through which the proteins migrate. The gel is prepared by polymerization of monomers; the pore size of the gel can be adjusted so that it is small enough to retard the migration of the protein molecules of interest. The proteins are dissolved in a solution that includes a powerful negatively charged detergent, sodium dodecyl sulfate, or SDS (Figure 8–12). Because this detergent binds to hydrophobic regions of the protein molecules, causing them to unfold into extended polypeptide chains, the individual protein molecules are released from their associations with other proteins or lipid molecules and rendered freely soluble in the detergent solution. In addition, a reducing agent such as  $\beta$ -mercaptoethanol (see Figure 8–12) is usually added to break any S-S linkages in the proteins, so that all of the constituent polypeptides in multisubunit proteins can be analyzed separately.

What happens when a mixture of SDS-solubilized proteins is run through a slab of polyacrylamide gel? Each protein molecule binds large numbers of the negatively charged detergent molecules, which mask the protein's intrinsic charge and cause it to migrate toward the positive electrode when a voltage is applied. Proteins of the same size tend to move through the gel with similar speeds because (1) their native structure is completely unfolded by the SDS, so that their shapes are the same, and (2) they bind the same amount of SDS and therefore have the same amount of negative charge. Larger proteins, with more charge, are subjected to larger electrical forces but also to a larger drag. In free solution, the two effects would cancel out, but, in the mesh of the polyacrylamide gel, which acts as a molecular sieve, large proteins are retarded much more than small ones. As a result, a complex mixture of proteins is fractionated into a series of discrete protein bands arranged in order of molecular weight (Figure 8–13). The major proteins are readily detected by staining the proteins in the gel with a dye such as Coomassie blue. Even minor proteins are seen in gels treated with a silver stain, so that as little as 10 ng of protein can be detected in a band. For some purposes, specific proteins can also be labeled with a radioactive isotope tag; exposure of the gel to film results in an *autoradiograph* on which the labeled proteins are visible (see Figure 8–16).

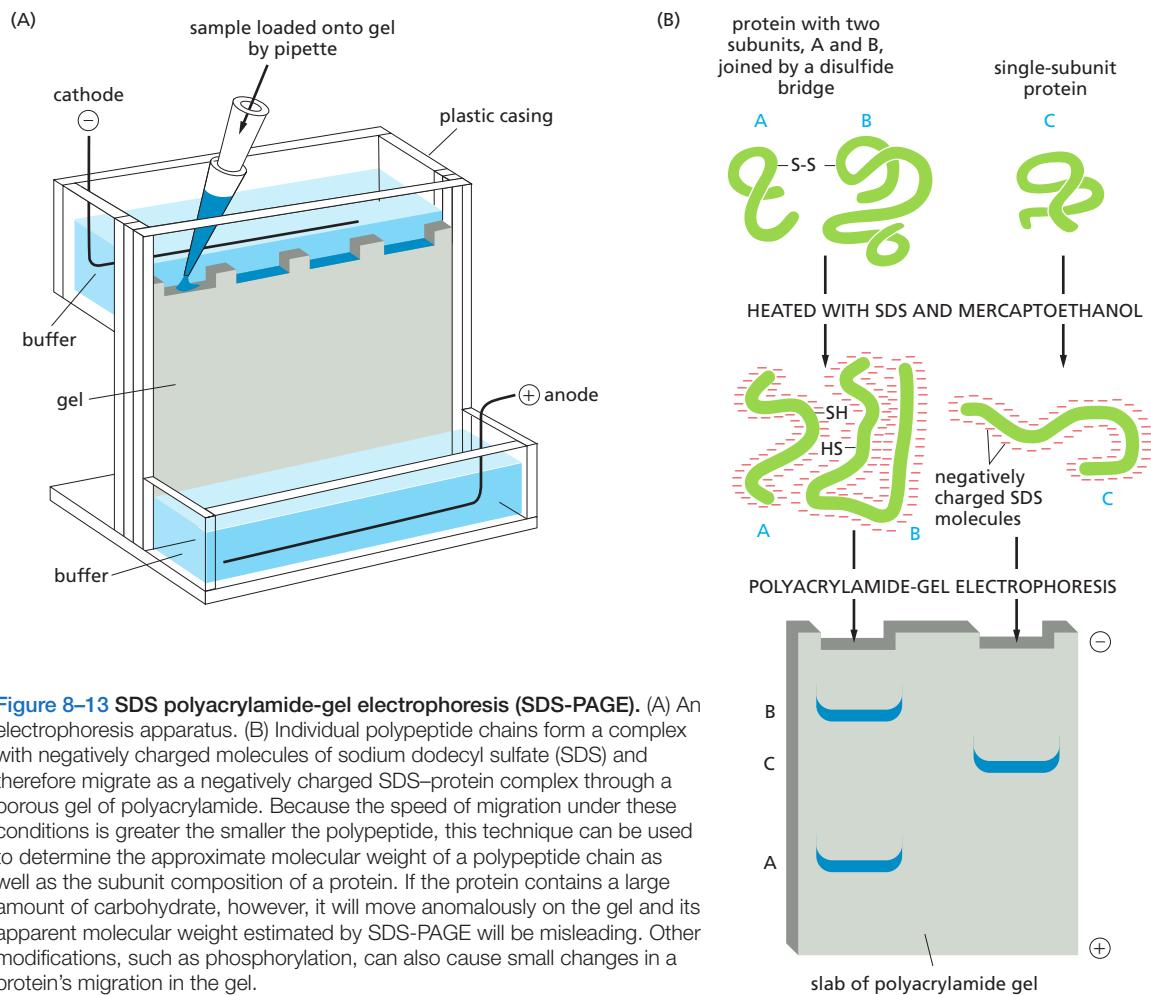
SDS-PAGE is widely used because it can separate all types of proteins, including those that are normally insoluble in water—such as the many proteins in membranes. And because the method separates polypeptides by size, it provides information about the molecular weight and the subunit composition of proteins. Figure 8–14 presents a photograph of a gel that has been used to analyze each of the successive stages in the purification of a protein.



**Figure 8–12** The detergent sodium dodecyl sulfate (SDS) and the reducing agent  $\beta$ -mercaptoethanol. These two chemicals are used to solubilize proteins for SDS polyacrylamide-gel electrophoresis. The SDS is shown here in its ionized form.

### Two-Dimensional Gel Electrophoresis Provides Greater Protein Separation

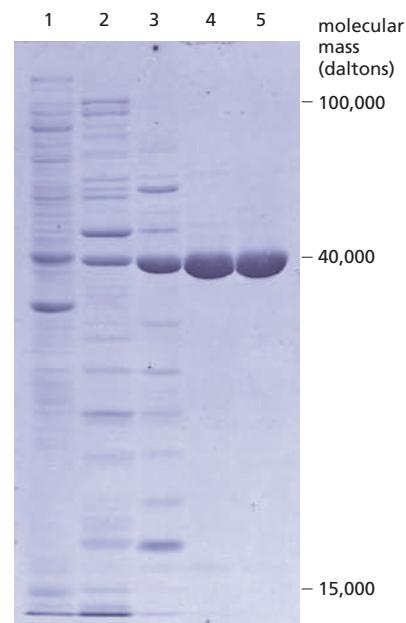
Because different proteins can have similar sizes, shapes, masses, and overall charges, most separation techniques such as SDS polyacrylamide-gel electrophoresis or ion-exchange chromatography cannot typically separate all the proteins



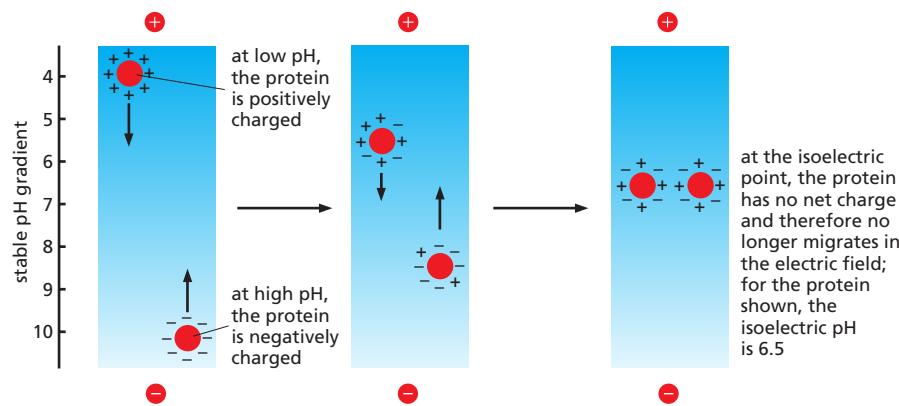
**Figure 8-13 SDS polyacrylamide-gel electrophoresis (SDS-PAGE).** (A) An electrophoresis apparatus. (B) Individual polypeptide chains form a complex with negatively charged molecules of sodium dodecyl sulfate (SDS) and therefore migrate as a negatively charged SDS-protein complex through a porous gel of polyacrylamide. Because the speed of migration under these conditions is greater the smaller the polypeptide, this technique can be used to determine the approximate molecular weight of a polypeptide chain as well as the subunit composition of a protein. If the protein contains a large amount of carbohydrate, however, it will move anomalously on the gel and its apparent molecular weight estimated by SDS-PAGE will be misleading. Other modifications, such as phosphorylation, can also cause small changes in a protein's migration in the gel.

in a cell or even in an organelle. In contrast, **two-dimensional gel electrophoresis**, which combines two different separation procedures, can resolve up to 2000 proteins in the form of a two-dimensional protein map.

In the first step, the proteins are separated by their intrinsic charges. The sample is dissolved in a small volume of a solution containing a nonionic (uncharged) detergent, together with  $\beta$ -mercaptoethanol and the denaturing reagent urea. This solution solubilizes, denatures, and dissociates all the polypeptide chains but leaves their intrinsic charge unchanged. The polypeptide chains are then separated in a pH gradient by a procedure called *isoelectric focusing*, which takes advantage of the variation in the net charge on a protein molecule with the pH of its surrounding solution. Every protein has a characteristic isoelectric point, the pH at which the protein has no net charge and therefore does not migrate in an electric field. In isoelectric focusing, proteins are separated electrophoretically in a narrow tube of polyacrylamide gel in which a gradient of pH is established by a mixture of special buffers. Each protein moves to a position in the gradient that



**Figure 8-14 Analysis of protein samples by SDS polyacrylamide-gel electrophoresis.** The photograph shows a Coomassie-stained gel that has been used to detect the proteins present at successive stages in the purification of an enzyme. The leftmost lane (lane 1) contains the complex mixture of proteins in the starting cell extract, and each succeeding lane analyzes the proteins obtained after a chromatographic fractionation of the protein sample analyzed in the previous lane (see Figure 8-10). The same total amount of protein (10  $\mu$ g) was loaded onto the gel at the top of each lane. Individual proteins normally appear as sharp, dye-stained bands; a band broadens, however, when it contains a large amount of protein. (From T. Formosa and B.M. Alberts, *J. Biol. Chem.* 261:6107–6118, 1986.)



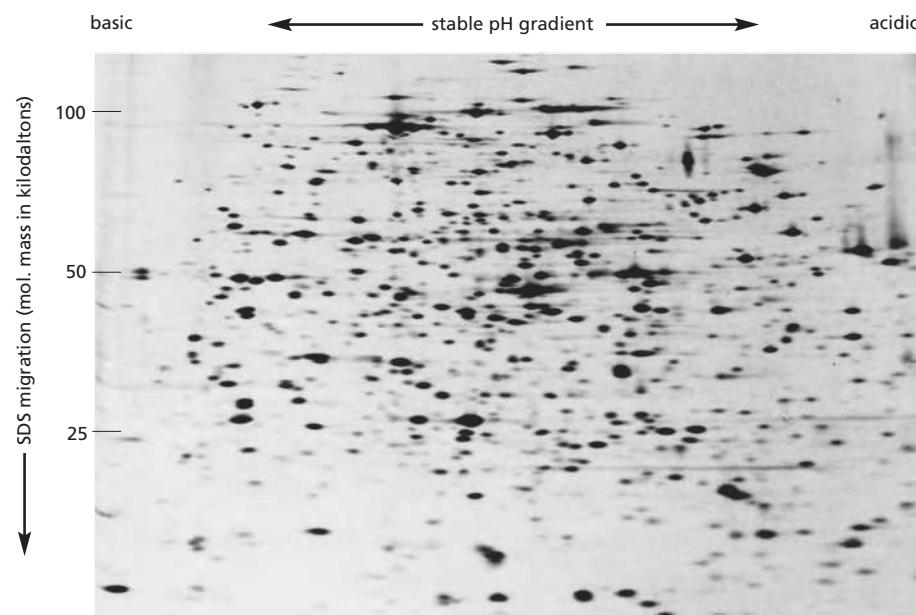
corresponds to its isoelectric point and remains there (Figure 8–15). This is the first dimension of two-dimensional polyacrylamide-gel electrophoresis.

In the second step, the narrow tube gel containing the separated proteins is again subjected to electrophoresis but in a direction that is at a right angle to the direction used in the first step. This time SDS is added, and the proteins separate according to their size, as in one-dimensional SDS-PAGE: the original tube gel is soaked in SDS and then placed along the top edge of an SDS polyacrylamide-gel slab, through which each polypeptide chain migrates to form a discrete spot. This is the second dimension of two-dimensional polyacrylamide-gel electrophoresis. The only proteins left unresolved are those that have both identical sizes and identical isoelectric points, a relatively rare situation. Even trace amounts of each polypeptide chain can be detected on the gel by various staining procedures—or by autoradiography if the protein sample was initially labeled with a radioisotope (Figure 8–16). The technique has such great resolving power that it can distinguish between two proteins that differ in only a single charged amino acid, or a single negatively charged phosphorylation site.

### Specific Proteins Can Be Detected by Blotting with Antibodies

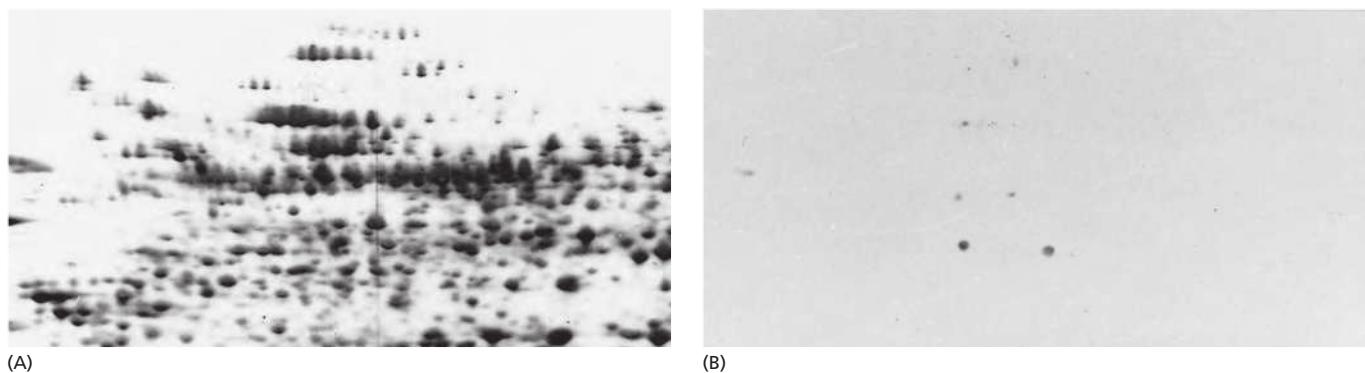
A specific protein can be identified after its fractionation on a polyacrylamide gel by exposing all the proteins present on the gel to a specific antibody that has been labeled with a radioactive isotope or a fluorescent dye. This procedure is normally carried out after transferring all of the separated proteins present in the

**Figure 8–15 Separation of protein molecules by isoelectric focusing.** At low pH (high H<sup>+</sup> concentration), the carboxylic acid groups of proteins tend to be uncharged ( $-COOH$ ) and their nitrogen-containing basic groups fully charged (for example,  $-NH_3^+$ ), giving most proteins a net positive charge. At high pH, the carboxylic acid groups are negatively charged ( $-COO^-$ ) and the basic groups tend to be uncharged (for example,  $-NH_2$ ), giving most proteins a net negative charge. At its *isoelectric pH*, a protein has no net charge since the positive and negative charges balance. Thus, when a tube containing a fixed pH gradient is subjected to a strong electric field in the appropriate direction, each protein species migrates until it forms a sharp band at its isoelectric pH, as shown.



**Figure 8–16 Two-dimensional polyacrylamide-gel electrophoresis.**

All the proteins in an *E. coli* bacterial cell are separated in this gel, in which each spot corresponds to a different polypeptide chain. The proteins were first separated on the basis of their isoelectric points by isoelectric focusing in the horizontal dimension. They were then further fractionated according to their molecular mass by electrophoresis from top to bottom in the presence of SDS. Note that different proteins are present in very different amounts. The bacteria were fed with a mixture of radioisotope-labeled amino acids so that all of their proteins were radioactive and could be detected by autoradiography. (Courtesy of Patrick O'Farrell.)



gel onto a sheet of nitrocellulose paper or nylon membrane. Placing the membrane over the gel and driving the proteins out of the gel with a strong electric current transfers the protein onto the membrane. The membrane is then soaked in a solution of labeled antibody to reveal the protein of interest. This method of detecting proteins is called **Western blotting**, or **immunoblotting** (Figure 8–17). Sensitive Western-blotting methods can detect very small amounts of a specific protein (1 nanogram or less) in a total cell extract or some other heterogeneous protein mixture. The method can be very useful when assessing the amounts of a specific protein in the cell or when measuring changes in those amounts under various conditions.

### Hydrodynamic Measurements Reveal the Size and Shape of a Protein Complex

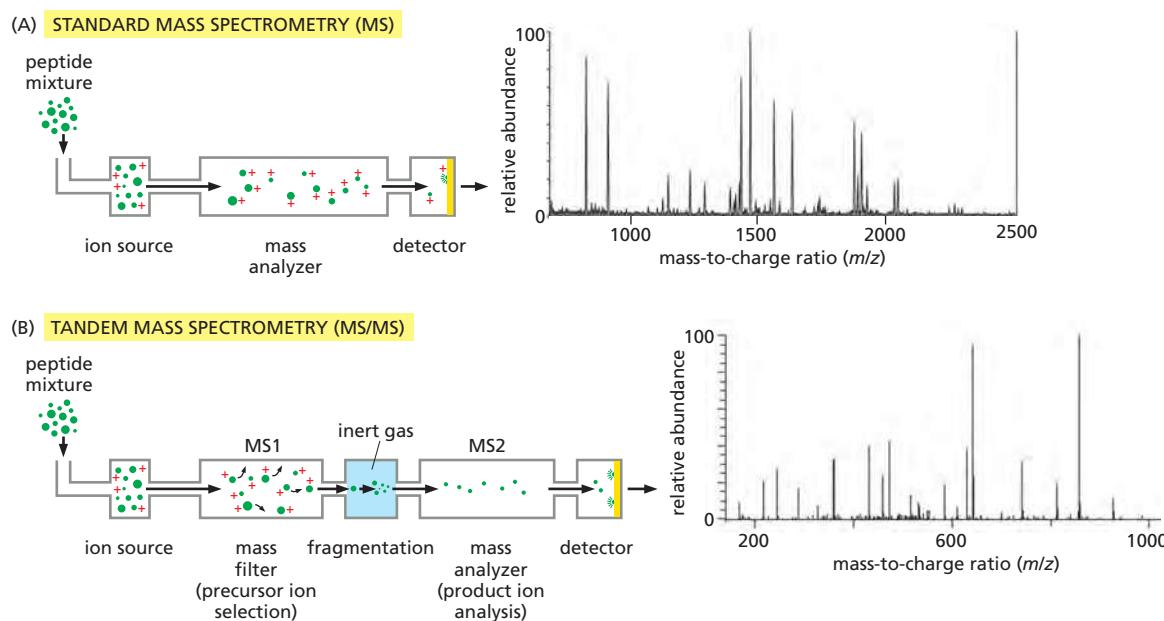
Most proteins in a cell act as part of larger complexes, and knowledge of the size and shape of these complexes often leads to insights regarding their function. This information can be obtained in several important ways. Sometimes, a complex can be directly visualized using electron microscopy, as described in Chapter 9. A complementary approach relies on the hydrodynamic properties of a complex; that is, its behavior as it moves through a liquid medium. Usually, two separate measurements are made. One measure is the velocity of a complex as it moves under the influence of a centrifugal field produced by an ultracentrifuge (see Figure 8–7A). The sedimentation coefficient (or S value) obtained depends on both the size and the shape of the complex and does not, by itself, convey especially useful information. However, once a second hydrodynamic measurement is performed—by charting the migration of a complex through a gel-filtration chromatography column (see Figure 8–9B)—both the approximate shape of a complex and its molecular weight can be calculated.

Molecular weight can also be determined more directly by using an analytical ultracentrifuge, a complex device that allows protein absorbance measurements to be made on a sample while it is subjected to centrifugal forces. In this approach, the sample is centrifuged until it reaches equilibrium, where the centrifugal force on a protein complex exactly balances its tendency to diffuse away. Because this balancing point is dependent on a complex's molecular weight but not on its particular shape, the molecular weight can be directly calculated.

### Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins

A frequent problem in cell biology and biochemistry is the identification of a protein or collection of proteins that has been obtained by one of the purification procedures discussed in the preceding pages. Because the genome sequences of most experimental organisms are now known, catalogs of all the proteins produced in those organisms are available. The task of identifying an unknown protein (or collection of unknown proteins) thus reduces to matching some of the amino acid

**Figure 8–17 Western blotting.** All the proteins from dividing tobacco cells in culture were first separated by two-dimensional polyacrylamide-gel electrophoresis. In (A), the positions of the proteins are revealed by a sensitive protein stain. In (B), the separated proteins on an identical gel were then transferred to a sheet of nitrocellulose and exposed to an antibody that recognizes only those proteins that are phosphorylated on threonine residues during mitosis. The positions of the few proteins that are recognized by this antibody are revealed by an enzyme-linked second antibody. (From J.A. Traas et al., *Plant J.* 2:723–732, 1992. With permission from Blackwell Publishing.)



**Figure 8–18 The mass spectrometer.** (A) Mass spectrometers used in biology contain an ion source that generates gaseous peptides or other molecules under conditions that render most molecules positively charged. The two major types of ion source are MALDI and electrospray, as described in the text. Ions are accelerated into a mass analyzer, which separates the ions on the basis of their mass and charge by one of three major methods: 1. *Time-of-flight (TOF)* analyzers determine the mass-to-charge ratio of each ion in the mixture from the rate at which it travels from the ion source to the detector. 2. *Quadrupole* mass filters contain a long chamber lined by four electrodes that produce oscillating electric fields that govern the trajectory of ions; by varying the properties of the electric field over a wide range, a spectrum of ions with specific mass-to-charge ratios is allowed to pass through the chamber to the detector, while other ions are discarded. 3. *Ion traps* contain doughnut-shaped electrodes producing a three-dimensional electric field that traps all ions in a circular chamber; the properties of the electric field can be varied over a wide range to eject a spectrum of specific ions to a detector. (B) Tandem mass spectrometry typically involves two mass analyzers separated by a collision chamber containing an inert, high-energy gas. The electric field of the first mass analyzer is adjusted to select a specific peptide ion, called a *precursor ion*, which is then directed to the collision chamber. Collision of the peptide with gas molecules causes random peptide fragmentation, primarily at peptide bonds, resulting in a highly complex mixture of fragments containing one or more amino acids from throughout the original peptide. The second mass analyzer is then used to measure the masses of the fragments (called *product* or *daughter ions*). With computer assistance, the pattern of fragments can be used to deduce the amino acid sequence of the original peptide.

sequences present in the unknown sample with known cataloged genes. This task is now performed almost exclusively by using mass spectrometry in conjunction with computer searches of databases.

Charged particles have very precise dynamics when subjected to electrical and magnetic fields in a vacuum. Mass spectrometry exploits this principle to separate ions according to their mass-to-charge ( $m/z$ ) ratio. It is an enormously sensitive technique. It requires very little material and is capable of determining the precise mass of intact proteins and of peptides derived from them by enzymatic or chemical cleavage. Masses can be obtained with great accuracy, often with an error of less than one part in a million.

Mass spectrometry is performed using complex instruments with three major components (Figure 8–18A). The first is the *ion source*, which transforms tiny amounts of a peptide sample into a gas containing individual charged peptide molecules. These ions are accelerated by an electric field into the second component, the *mass analyzer*, where electric or magnetic fields are used to separate the ions on the basis of their mass-to-charge ratios. Finally, the separated ions collide with a *detector*, which generates a mass spectrum containing a series of peaks representing the masses of the molecules in the sample.

There are many different types of mass spectrometer, varying mainly in the nature of their ion sources and mass analyzers. One of the most common ion sources depends on a technique called *matrix-assisted laser desorption ionization (MALDI)*. In this approach, the proteins in the sample are first cleaved into short peptides by a protease such as trypsin. These peptides are mixed with an organic

acid and then dried onto a metal or ceramic slide. A brief laser burst is directed toward the sample, producing a gaseous puff of ionized peptides, each carrying one or more positive charges. In many cases, the MALDI ion source is coupled to a mass analyzer called a *time-of-flight (TOF)* analyzer, which is a long chamber through which the ionized peptides are accelerated by an electric field toward a detector. Their mass and charge determine the time it takes them to reach the detector: large peptides move more slowly, and more highly charged molecules move more quickly. By analyzing those ionized peptides that bear a single charge, the precise masses of peptides present in the original sample can be determined. This information is then used to search genomic databases, in which the masses of all proteins and of all their predicted peptide fragments have been tabulated from the genomic sequences of the organism. An unambiguous match to a particular open reading frame can often be made by knowing the mass of only a few peptides derived from a given protein.

By employing two mass analyzers in tandem (an arrangement known as MS/MS; Figure 8–18B), it is possible to directly determine the amino acid sequences of individual peptides in a complex mixture. The MALDI-TOF instrument described above is not ideal for this method. Instead, MS/MS typically involves an *electrospray* ion source, which produces a continuous thin stream of peptides that are ionized and accelerated into the first mass analyzer. The mass analyzer is typically either a *quadropole* or *ion trap*, which employs large electrodes to produce oscillating electric fields inside the chamber containing the ions. These instruments act as *mass filters*: the electric field is adjusted over a broad range to select a single peptide ion and discard all the others in the peptide mixture. In tandem mass spectrometry, this single ion is then exposed to an inert, high-energy gas, which collides with the peptide, resulting in fragmentation, primarily at peptide bonds. The second mass analyzer then determines the masses of the peptide fragments, which can be used by computational methods to determine the amino acid sequence of the original peptide and thereby identify the protein from which it came.

Tandem mass spectrometry is also useful for detecting and precisely mapping post-translational modifications of proteins, such as phosphorylations or acetylations. Because these modifications impart a characteristic mass increase to an amino acid, they are easily detected during the analysis of peptide fragments in the second mass analyzer, and the precise site of the modification can often be deduced from the spectrum of peptide fragments.

A powerful, “two-dimensional” mass spectrometry technique can be used to determine all of the proteins present in an organelle or another complex mixture of proteins. First, the mixture of proteins present is digested with trypsin to produce short peptides. Next, these peptides are separated by automated high-performance liquid chromatography (LC). Every peptide fraction from the chromatographic column is injected directly into an electrospray ion source on a tandem mass spectrometer (MS/MS), providing the amino acid sequence and post-translational modifications for every peptide in the mixture. This method, often called LC-MS/MS, is used to identify hundreds or thousands of proteins in complex protein mixtures from specific organelles or from whole cells. It can also be used to map all of the phosphorylation sites in the cell, or all of the proteins targeted by other post-translational modifications such as acetylation or ubiquitylation.

### Sets of Interacting Proteins Can Be Identified by Biochemical Methods

Because most proteins in the cell function as part of complexes with other proteins, an important way to begin to characterize the biological role of an unknown protein is to identify all of the other proteins to which it specifically binds.

A key method for identifying proteins that bind to one another tightly is *co-immunoprecipitation*. A specific target protein is immunoprecipitated from a cell lysate using specific antibodies coupled to beads, as described earlier. If the target protein is associated tightly enough with another protein when it is captured by

the antibody, the partner precipitates as well and can be identified by mass spectrometry. This method is useful for identifying proteins that are part of a complex inside cells, including those that interact only transiently—for example, when extracellular signal molecules stimulate cells (discussed in Chapter 15).

In addition to capturing protein complexes on columns or in test tubes, researchers are developing high-density protein arrays to investigate protein interactions. These arrays, which contain thousands of different proteins or antibodies spotted onto glass slides or immobilized in tiny wells, allow one to examine the biochemical activities and binding profiles of a large number of proteins at once. For example, if one incubates a fluorescently labeled protein with arrays containing thousands of immobilized proteins, the spots that remain fluorescent after extensive washing each contain a protein that specifically binds the labeled protein.

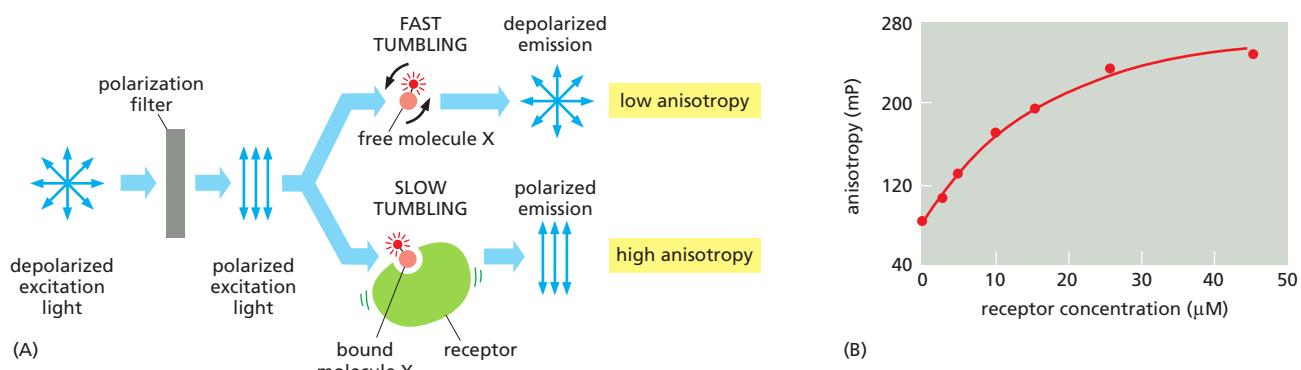
### Optical Methods Can Monitor Protein Interactions

Once two proteins—or a protein and a small molecule—are known to associate, it becomes important to characterize their interaction in more detail. Proteins can associate with each other more or less permanently (like the subunits of RNA polymerase or the proteasome), or engage in transient encounters that may last only a few milliseconds (like a protein kinase and its substrate). To understand how a protein functions inside a cell, we need to determine how tightly it binds to other proteins, how rapidly it dissociates from them, and how covalent modifications, small molecules, or other proteins influence these interactions.

As we discussed in Chapter 3 (see Figure 3–44), the extent to which two proteins interact is determined by the rates at which they associate and dissociate. These rates depend, respectively, on the association rate constant ( $k_{on}$ ) and dissociation rate constant ( $k_{off}$ ). The kinetic rate constant  $k_{off}$  is a particularly useful number because it provides valuable information about how long two proteins remain bound to one another. The ratio of the two kinetic constants ( $k_{on}/k_{off}$ ) yields another very useful number called the equilibrium constant ( $K$ , also known as  $K_{eq}$  or  $K_a$ ), the inverse of which is the more commonly used dissociation constant  $K_d$ . The equilibrium constant is useful as a general indicator of the affinity of the interaction, and it can be used to estimate the amount of bound complex at different concentrations of the two protein partners—thereby providing insights into the importance of the interaction at the protein concentrations found inside the cell.

A wide range of methods can be used to determine binding constants for a two-protein complex. In a simple *equilibrium binding experiment*, two proteins are mixed at a range of concentrations, allowed to reach equilibrium, and the amount of bound complex is measured; half of the protein complex will be bound at a concentration that is equal to  $K_d$ . Equilibrium experiments often involve the use of radioactive or fluorescent tags on one of the protein partners, coupled with biochemical or optical methods for measuring the amount of bound protein. In a more complex *kinetic binding experiment*, the kinetic rate constants are determined using rapid methods that allow real-time measurement of the formation of a bound complex over time (to determine  $k_{on}$ ) or the dissociation of a bound complex over time (to determine  $k_{off}$ ).

Optical techniques provide particularly rapid, convenient, and accurate binding measurements, and in some cases the proteins do not even need to be labeled. Certain amino acids (tryptophan, for example) exhibit weak fluorescence that can be detected with sensitive fluorimeters. In many cases, the fluorescence intensity, or the emission spectrum of fluorescent amino acids located in a protein–protein interface, will change when two proteins associate. When this change can be detected by fluorimetry, it provides a simple and sensitive measure of protein binding that is useful in both equilibrium and kinetic binding experiments. A related but more widely useful optical binding technique is based on *fluorescence anisotropy*, a change in the polarized light that is emitted by a fluorescently tagged protein in the bound and free states (Figure 8–19).



**Figure 8–19 Measurement of binding with fluorescence anisotropy.** This method depends on a fluorescently tagged protein that is illuminated with polarized light at the appropriate wavelength for excitation; a fluorimeter is used to measure the intensity and polarization of the emitted light. If the fluorescent protein is fixed in position and therefore does not rotate during the brief period between excitation and emission, then the emitted light will be polarized at the same angle as the excitation light. This directional effect is called *fluorescence anisotropy*. Protein molecules in solution rotate or tumble rapidly, however, so that there is a decrease in the amount of anisotropic fluorescence. Larger molecules tumble at a slower rate and therefore have higher fluorescence anisotropy. (A) To measure the binding between a small molecule and a large receptor protein, the smaller molecule is labeled with a fluorophore. In the absence of its binding partner, the molecule tumbles rapidly, resulting in low fluorescence anisotropy (top). When the small molecule binds to its larger partner, however, it tumbles less rapidly, resulting in an increase in fluorescence anisotropy (bottom). (B) In the equilibrium binding experiment shown here, a small, fluorescent peptide ligand was present at a low concentration, and the amount of fluorescence anisotropy (in millipolarization units, mP) was measured after incubation with various concentrations of a larger protein receptor for the ligand. From the hyperbolic curve that fits the data, it can be seen that 50% binding occurred at about  $10 \mu\text{M}$ , which is equal to the dissociation constant  $K_d$  for the binding interaction.

Another optical method for probing protein interactions uses *green fluorescent protein* (discussed in detail below) and its derivatives of different colors. In this application, two proteins of interest are each labeled with a different fluorescent protein, such that the emission spectrum of one fluorescent protein overlaps the absorption spectrum of the second. If the two proteins come very close to each other (within about 1–5 nm), the energy of the absorbed light is transferred from one fluorescent protein to the other. The energy transfer, called *fluorescence resonance energy transfer (FRET)*, is determined by illuminating the first fluorescent protein and measuring emission from the second (see Figure 9–26). When combined with fluorescence microscopy, this method can be used to characterize protein–protein interactions at specific locations inside living cells (discussed in Chapter 9).

### Protein Function Can Be Selectively Disrupted With Small Molecules

Small chemical inhibitors of specific proteins have contributed a great deal to the development of cell biology. For example, the microtubule inhibitor colchicine is routinely used to test whether microtubules are required for a given biological process; it also led to the first purification of tubulin several decades ago. In the past, these small molecules were usually natural products; that is, they were synthesized by living creatures. Although natural products have been extraordinarily useful in science and medicine (see, for example, Table 6–4, p. 352), they act on a limited number of biological processes. However, the recent development of methods to synthesize hundreds of thousands of small molecules and to carry out large-scale automated screens holds the promise of identifying chemical inhibitors for virtually any biological process. In such approaches, large collections of small chemical compounds are simultaneously tested, either on living cells or in cell-free assays. Once an inhibitor is identified, it can be used as a probe to identify, through affinity chromatography or other means, the protein to which the inhibitor binds. This general strategy, sometimes called **chemical biology**, has successfully identified inhibitors of many proteins that carry out key processes in cell biology. An inhibitor of a kinesin protein that functions in mitosis, for

**Figure 8–20 Small-molecule inhibitors for manipulating living cells.**  
 (A) Chemical structure of monastrol, a kinesin inhibitor identified in a large-scale screen for small molecules that disrupt mitosis. (B) Normal mitotic spindle seen in an untreated cell. The microtubules are stained green and chromosomes blue. (C) Monopolar spindle that forms in cells treated with monastrol, which inhibits a kinesin protein required for separation of the spindle poles in early mitosis. (B and C, from T.U. Mayer et al., *Science* 286:971–974, 1999. With permission from AAAS.)

example, was identified by this method (Figure 8–20). Chemical inhibitors give the cell biologist great control over the timing of inhibition, as drugs can be rapidly added to or removed from cells, allowing protein function to be switched on or off quickly.

### Protein Structure Can Be Determined Using X-Ray Diffraction

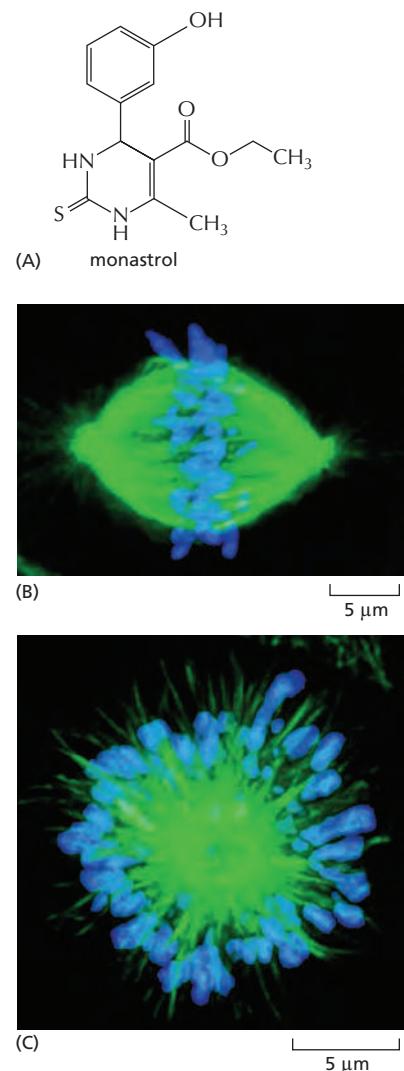
The main technique that has been used to discover the three-dimensional structure of molecules, including proteins, at atomic resolution is **x-ray crystallography**. X-rays, like light, are a form of electromagnetic radiation, but they have a much shorter wavelength, typically around 0.1 nm (the diameter of a hydrogen atom). If a narrow beam of parallel x-rays is directed at a sample of a pure protein, most of the x-rays pass straight through it. A small fraction, however, are scattered by the atoms in the sample. If the sample is a well-ordered crystal, the scattered waves reinforce one another at certain points and appear as diffraction spots when recorded by a suitable detector (Figure 8–21).

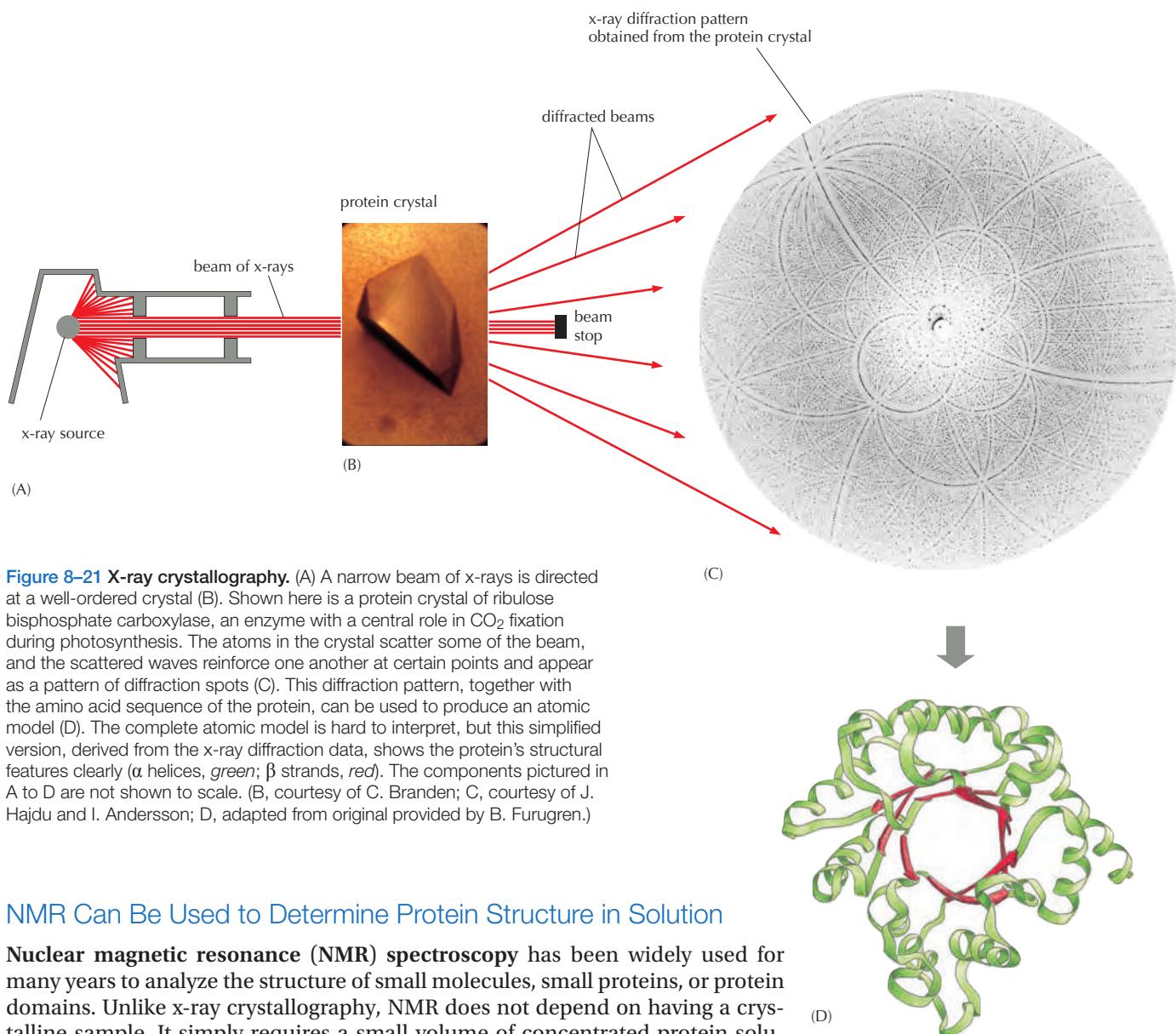
The position and intensity of each spot in the x-ray diffraction pattern contain information about the locations of the atoms in the crystal that gave rise to it. Deducing the three-dimensional structure of a large molecule from the diffraction pattern of its crystal is a complex task and was not achieved for a protein molecule until 1960. But in recent years x-ray diffraction analysis has become increasingly automated, and now the slowest step is likely to be the generation of suitable protein crystals. This step requires large amounts of very pure protein and often involves years of trial and error to discover the proper crystallization conditions; the pace has greatly accelerated with the use of recombinant DNA techniques to produce pure proteins and robotic techniques to test large numbers of crystallization conditions.

Analysis of the resulting diffraction pattern produces a complex three-dimensional electron-density map. Interpreting this map—translating its contours into a three-dimensional structure—is a complicated procedure that requires knowledge of the amino acid sequence of the protein. Largely by trial and error, the sequence and the electron-density map are correlated by computer to give the best possible fit. The reliability of the final atomic model depends on the resolution of the original crystallographic data: 0.5 nm resolution might produce a low-resolution map of the polypeptide backbone, whereas a resolution of 0.15 nm allows all of the non-hydrogen atoms in the molecule to be reliably positioned.

A complete atomic model is often too complex to appreciate directly, but simplified versions that show a protein's essential structural features can be readily derived from it (see Panel 3–2, pp. 142–143). The three-dimensional structures of tens of thousands of different proteins have been determined by x-ray crystallography or by NMR spectroscopy (see page 461)—enough to allow the grouping of common structures into families (Movie 8.1). These structures or protein folds often seem to be more conserved in evolution than are the amino acid sequences that form them (see Figure 3–13).

X-ray crystallographic techniques can also be applied to the study of macromolecular complexes. The method was used, for example, to determine the structure of the ribosome, a large and complex machine made of several RNAs and more than 50 proteins (see Figure 6–62). The determination required the use of a synchrotron, a radiation source that generates x-rays with the intensity needed to analyze the crystals of such large macromolecular complexes.



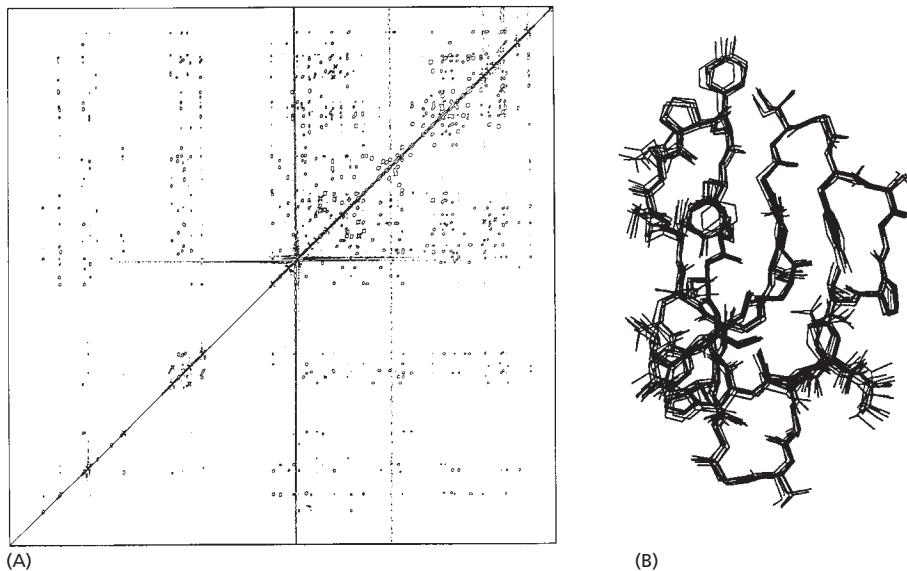


**Figure 8–21 X-ray crystallography.** (A) A narrow beam of x-rays is directed at a well-ordered crystal (B). Shown here is a protein crystal of ribulose bisphosphate carboxylase, an enzyme with a central role in  $\text{CO}_2$  fixation during photosynthesis. The atoms in the crystal scatter some of the beam, and the scattered waves reinforce one another at certain points and appear as a pattern of diffraction spots (C). This diffraction pattern, together with the amino acid sequence of the protein, can be used to produce an atomic model (D). The complete atomic model is hard to interpret, but this simplified version, derived from the x-ray diffraction data, shows the protein's structural features clearly ( $\alpha$  helices, green;  $\beta$  strands, red). The components pictured in A to D are not shown to scale. (B, courtesy of C. Branden; C, courtesy of J. Hajdu and I. Andersson; D, adapted from original provided by B. Furugren.)

### NMR Can Be Used to Determine Protein Structure in Solution

**Nuclear magnetic resonance (NMR) spectroscopy** has been widely used for many years to analyze the structure of small molecules, small proteins, or protein domains. Unlike x-ray crystallography, NMR does not depend on having a crystalline sample. It simply requires a small volume of concentrated protein solution that is placed in a strong magnetic field; indeed, it is the main technique that yields detailed evidence about the three-dimensional structure of molecules in solution.

Certain atomic nuclei, particularly hydrogen nuclei, have a magnetic moment or spin: that is, they have an intrinsic magnetization, like a bar magnet. The spin aligns along the strong magnetic field, but it can be changed to a misaligned, excited state in response to applied radiofrequency (RF) pulses of electromagnetic radiation. When the excited hydrogen nuclei return to their aligned state, they emit RF radiation, which can be measured and displayed as a spectrum. The nature of the emitted radiation depends on the environment of each hydrogen nucleus, and if one nucleus is excited, it influences the absorption and emission of radiation by other nuclei that lie close to it. It is consequently possible, by an ingenious elaboration of the basic NMR technique known as two-dimensional NMR, to distinguish the signals from hydrogen nuclei in different amino acid residues, and to identify and measure the small shifts in these signals that occur when these hydrogen nuclei lie close enough together to interact. Because the size of such a shift reveals the distance between the interacting pair of hydrogen atoms, NMR can provide information about the distances between the parts of the protein molecule. By combining this information with a knowledge of the amino acid

**Figure 8-22** NMR spectroscopy.

(A) An example of the data from an NMR machine. This two-dimensional NMR spectrum is derived from the C-terminal domain of the enzyme cellulase. The spots represent interactions between hydrogen atoms that are near neighbors in the protein and hence reflect the distance that separates them. Complex computing methods, in conjunction with the known amino acid sequence, enable possible compatible structures to be derived.

(B) Ten structures of the enzyme, which all satisfy the distance constraints equally well, are shown superimposed on one another, giving a good indication of the probable three-dimensional structure. (Courtesy of P. Kraulis.)

sequence, it is possible in principle to compute the three-dimensional structure of the protein (**Figure 8-22**).

For technical reasons, the structure of small proteins of about 20,000 daltons or less can be most readily determined by NMR spectroscopy. Resolution decreases as the size of a macromolecule increases. But recent technical advances have now pushed the limit to about 100,000 daltons, thereby making the majority of proteins accessible for structural analysis by NMR.

Because NMR studies are performed in solution, this method also offers a convenient means of monitoring changes in protein structure, for example during protein folding or when the protein binds to another molecule. NMR is also used widely to investigate molecules other than proteins and is valuable, for example, as a method to determine the three-dimensional structures of RNA molecules and the complex carbohydrate side chains of glycoproteins.

A third major method for the determination of protein structure, and particularly the structure of large protein complexes, is single-particle analysis by electron microscopy. We discuss this approach in Chapter 9.

### Protein Sequence and Structure Provide Clues About Protein Function

Having discussed methods for purifying and analyzing proteins, we now turn to a common situation in cell and molecular biology: an investigator has identified a gene important for a biological process but has no direct knowledge of the biochemical properties of its protein product.

Thanks to the proliferation of protein and nucleic acid sequences that are cataloged in genome databases, the function of a gene—and its encoded protein—can often be predicted by simply comparing its sequence with those of previously characterized genes. Because amino acid sequence determines protein structure, and structure dictates biochemical function, proteins that share a similar amino acid sequence usually have the same structure and usually perform similar biochemical functions, even when they are found in distantly related organisms. In modern cell biology, the study of a newly discovered protein usually begins with a search for previously characterized proteins that are similar in their amino acid sequences.

Searching a collection of known sequences for similar genes or proteins is typically done over the Internet, and it simply involves selecting a database and entering the desired sequence. A sequence-alignment program—the most popular is BLAST—scans the database for similar sequences by sliding the submitted

```

Score = 399 bits (1025), Expect = e-111
Identities = 198/290 (68%), Positives = 241/290 (82%), Gaps = 1/290

Query: 57 MENFQKVEKIGEGTYGVVYKA[RNK]T[G]EVVALKKIRLD[T]EGVPSTAIREISLLKELNH 116
        ME ++KVEKIGEGTYGVVYKA +K T E +ALKKIRL+ E EGVPSTAIREISLLKE+NH
Sbjct: 1  MEQYEKVEKIGEGTYGVVYKA[D]KAT[N]ETIALKKIRLE[Q]EGVPSTAIREISLLKEMNH 60

Query: 117 PNIVKL[L]DVIHTE[N]KLYLVF[EFL]HDLKKFMDSALTG[IPL]LIKSYLFQLLOGLAFCHS 176
        NIV+I DV+H+E ++YLVFE+L DLKKFMD+ LIKSYL+Q+I G+A+CHS
Sbjct: 61  GNIVRL[H]DVVHSEK[R]IYLVF[EY]L[D]DLKKFMDS[C]PEFAKNP[L]I[S]YLYQIL[H]GVAYCHS 120

Query: 177 HRVLHRDLKPQNLLINTE[GAIKLA]DFGLARAFG[P]VVRTYTHEVVT[LWYRAPEILLGCKY 235
        HRVLHRDLKPQNLLI+ A+KLADFG[LA]RAGF+PVRT+THEVVT[LWYRAPEILLG +
Sbjct: 121 HRVLHRDLKPQNLLID[RRTN]ALKLADFG[LA]RAGF[G]IPVVRTFTHEVVT[LWYRAPEILLGARO 180

Query: 236 YSTAVDIWSLGCIFAEMVTRRALFP[G]DSEIDQLFRIFR[L]GTPDE[V]WP[G]VT[SMPDYKPS 295
        YST VD+WS+GCFIAEMV ++ LPGDSEID+L+F+IFR LGTP+E WPGV+ +PD+K +
Sbjct: 181 YST[P]DVWSVGCIFAEMV[NQKPL]FP[G]DSEIDELFKIFR[L]GTPNEQS[WPGVSCLPDFKTA 240

Query: 296 FPKWARQDFSK[V]VPPPLDEDGRSLLSQMLHYDPNKRISAKA[A]LAH[P]FFQDV 345
        FP+W QD + VVP LD G LLS+ML Y+P+KRI+A+ AL H +F+D+
Sbjct: 241 FPRWQAQDLAT[V]VP[NLDPA]GLDLLS[KM]RYEPSKRITARQALEHEYFKDL 290

```

sequence along the archived sequences until a cluster of residues falls into full or partial alignment (Figure 8–23). Such comparisons can predict the functions of individual proteins, families of proteins, or even most of the protein complement of a newly sequenced organism.

As was explained in Chapter 3, many proteins that adopt the same conformation and have related functions are too distantly related to be identified from a comparison of their amino acid sequences alone (see Figure 3–13). Thus, an ability to reliably predict the three-dimensional structure of a protein from its amino acid sequence would improve our ability to infer protein function from the sequence information in genomic databases. In recent years, major progress has been made in predicting the precise structure of a protein. These predictions are based, in part, on our knowledge of the thousands of protein structures that have already been determined by x-ray crystallography and NMR spectroscopy and, in part, on computations using our knowledge of the physical forces acting on the atoms. However, it remains a substantial and important challenge to predict the structures of proteins that are large or have multiple domains, or to predict structures at the very high levels of resolution needed to assist in computer-based drug discovery.

While finding related sequences and structures for a new protein will provide many clues about its function, it is usually necessary to test these insights through direct experimentation. However, the clues generated from sequence comparisons typically point the investigator in the correct experimental direction, and their use has therefore become one of the most important strategies in modern cell biology.

## Summary

*Many methods exist for identifying proteins and analyzing their biochemical properties, structures, and interactions with other proteins. Small-molecule inhibitors allow the functions of proteins they act upon to be studied in living cells. Because proteins with similar structures often have similar functions, the biochemical activity of a protein can often be predicted by searching databases for previously characterized proteins that are similar in their amino acid sequences.*

## ANALYZING AND MANIPULATING DNA

Until the early 1970s, DNA was the most difficult biological molecule for the biochemist to analyze. Enormously long and chemically monotonous, the string of nucleotides that forms the genetic material of an organism could be examined only indirectly, by protein sequencing or by genetic analysis. Today, the situation has changed entirely. From being the most difficult macromolecule of the cell to

**Figure 8–23 Results of a BLAST search.** Sequence databases can be searched to find similar amino acid or nucleic acid sequences. Here, a search for proteins similar to the human cell-cycle regulatory protein Cdc2 (Query) locates maize Cdc2 (Sbjct), which is 68% identical to human Cdc2 in its amino acid sequence. The alignment begins at residue 57 of the Query protein, suggesting that the human protein has an N-terminal region that is absent from the maize protein. The green blocks indicate differences in sequence, and the yellow bar summarizes the similarities: when the two amino acid sequences are identical, the residue is shown; similar amino acid substitutions are indicated by a plus sign (+). Only one small gap has been introduced—indicated by the red arrow at position 194 in the Query sequence—to align the two sequences maximally. The alignment score (Score), which is expressed in two different types of units, takes into account penalties for substitutions and gaps; the higher the alignment score, the better the match. The significance of the alignment is reflected in the *Expectation* (E) value, which specifies how often a match this good would be expected to occur by chance. The lower the E value, the more significant the match; the extremely low value here ( $e^{-111}$ ) indicates certain significance. E values much higher than 0.1 are unlikely to reflect true relatedness. For example, an E value of 0.1 means there is a 1 in 10 likelihood that such a match would arise solely by chance.

analyze, DNA has become the easiest. It is now possible to determine the entire nucleotide sequence of a bacterial or fungal genome in a matter of hours, and the sequence of an individual human genome in less than a day. Once the nucleotide sequence of a genome is known, any individual gene can be easily isolated, and large quantities of the gene product (be it RNA or protein) can be made either by introducing the gene into bacteria or animal cells and coaxing these cells to over-express the foreign gene or by synthesizing the gene product *in vitro*. In this way, proteins and RNA molecules that might be present in only tiny amounts in living cells can be produced in large quantities for biochemical and structural analyses. And this approach can also be used to produce large quantities of human proteins (such as insulin, or interferon, or blood-clotting proteins) for use as human pharmaceuticals. As we will see later in this chapter, it is also possible for scientists to alter an isolated gene and transfer it back into the germ line of an animal or plant, so as to become a functional and heritable part of the organism's genome. In this way, the biological roles of any gene can be assessed by observing—in the whole organism—the results of modifying it.

The ability to manipulate DNA with precision in a test tube or an organism, known as **recombinant DNA technology** has had a dramatic impact on all aspects of cell and molecular biology, allowing us to routinely study cells and their macromolecules in ways that were unimaginable even twenty years ago. Central to the technology are the following manipulations:

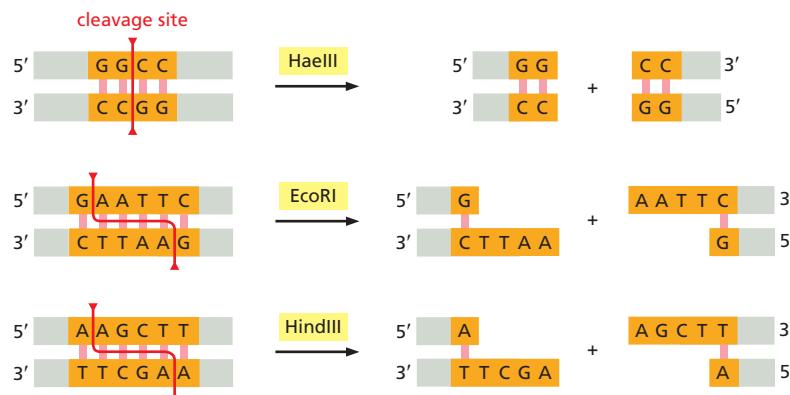
1. Cleavage of DNA at specific sites by restriction nucleases, which greatly facilitates the isolation and manipulation of individual pieces of a genome.
2. DNA ligation, which makes it possible to seamlessly join together DNA molecules from widely different sources.
3. DNA cloning (through the use of either cloning vectors or the polymerase chain reaction) in which a portion of a genome (often an individual gene) is “purified” away from the remainder of the genome by repeatedly copying it to generate many billions of identical molecules.
4. Nucleic acid hybridization, which makes it possible to identify any specific sequence of DNA or RNA with great accuracy and sensitivity based on its ability to selectively bind a complementary nucleic acid sequence.
5. DNA synthesis, which makes it possible to chemically synthesize DNA molecules with any sequence of nucleotides, whether or not the sequence occurs in nature.
6. Rapid determination of the sequence of nucleotides of any DNA or RNA molecule.

In the following sections, we describe each of these basic techniques which, together, have revolutionized the study of cell and molecular biology.

### Restriction Nucleases Cut Large DNA Molecules into Specific Fragments

Unlike a protein, a gene does not exist as a discrete entity in cells, but rather as a small region of a much longer DNA molecule. Although the DNA molecules in a cell can be randomly broken into small pieces by mechanical force, a fragment containing a single gene in a mammalian genome would still be only one among a hundred thousand or more DNA fragments, indistinguishable in their average size. How could such a gene be separated from all the others? Because all DNA molecules consist of an approximately equal mixture of the same four nucleotides, they cannot be readily separated, as proteins can, on the basis of their different charges and biochemical properties. The solution to this problem began to emerge with the discovery of **restriction nucleases**. These enzymes, which are purified from bacteria, cut the DNA double helix at specific sites defined by the local nucleotide sequence, thereby cleaving a long, double-stranded DNA molecule into fragments of strictly defined sizes.

Like many of the tools of recombinant DNA technology, restriction nucleases were discovered by researchers trying to understand an intriguing biological



phenomenon. It had been observed that certain bacteria always degraded “foreign” DNA that was introduced into them experimentally. A search for the mechanism responsible revealed a then unanticipated class of bacterial nucleases that cleave DNA at specific nucleotide sequences. The bacterium’s own DNA is protected from cleavage by methylation of these same sequences, thereby protecting a bacterium’s own genome from being overrun by foreign DNA. Because these enzymes restrict the transfer of DNA into bacteria, they were called restriction nucleases. The pursuit of this seemingly arcane biological puzzle set off the development of technologies that have forever changed the way cell and molecular biologists study living things.

Different bacterial species produce different restriction nucleases, each cutting at a different, specific nucleotide sequence (Figure 8–24). Because these target sequences are short—generally four to eight nucleotide pairs—many sites of cleavage will occur, purely by chance, in any long DNA molecule. The reason restriction nucleases are so useful in the laboratory is that each enzyme will always cut a particular DNA molecule at the same sites. Thus for a given sample of DNA (which contains many identical molecules), a particular restriction nuclease will reliably generate the same set of DNA fragments.

The size of the resulting fragments depends on the length of the target sequences of the restriction nucleases. As shown in Figure 8–24, the enzyme HaeIII cuts at a sequence of four nucleotide pairs; a sequence this long would be expected to occur purely by chance approximately once every 256 nucleotide pairs (1 in  $4^4$ ). In comparison, a restriction nuclease with a target sequence that is eight nucleotides long would be expected to cleave DNA on average once every 65,536 nucleotide pairs (1 in  $4^8$ ). This difference in sequence selectivity makes it possible to cleave a long DNA molecule into the fragment sizes that are most suitable for a given application.

### Gel Electrophoresis Separates DNA Molecules of Different Sizes

The same types of gel-electrophoresis methods that have proved so useful in the analysis of proteins (see Figure 8–13) can be applied to DNA molecules. The procedure is actually simpler than for proteins: because each nucleotide in a nucleic acid molecule already carries a single negative charge (on the phosphate group), there is no need to add the negatively charged detergent SDS that is required to make protein molecules move uniformly toward the positive electrode. Larger DNA fragments will migrate more slowly because their progress is impeded to a greater extent by the gel matrix. Over several hours, the DNA fragments become spread out across the gel according to size, forming a ladder of discrete bands, each composed of a collection of DNA molecules of identical length (Figure 8–25A and B). To separate DNA molecules longer than 500 nucleotide pairs, the gel is made of a diluted solution of agarose (a polysaccharide isolated from seaweed). For DNA fragments less than 500 nucleotides long, specially designed polyacrylamide gels allow the separation of molecules that differ in length by as little as a single nucleotide (see Figure 8–25C).

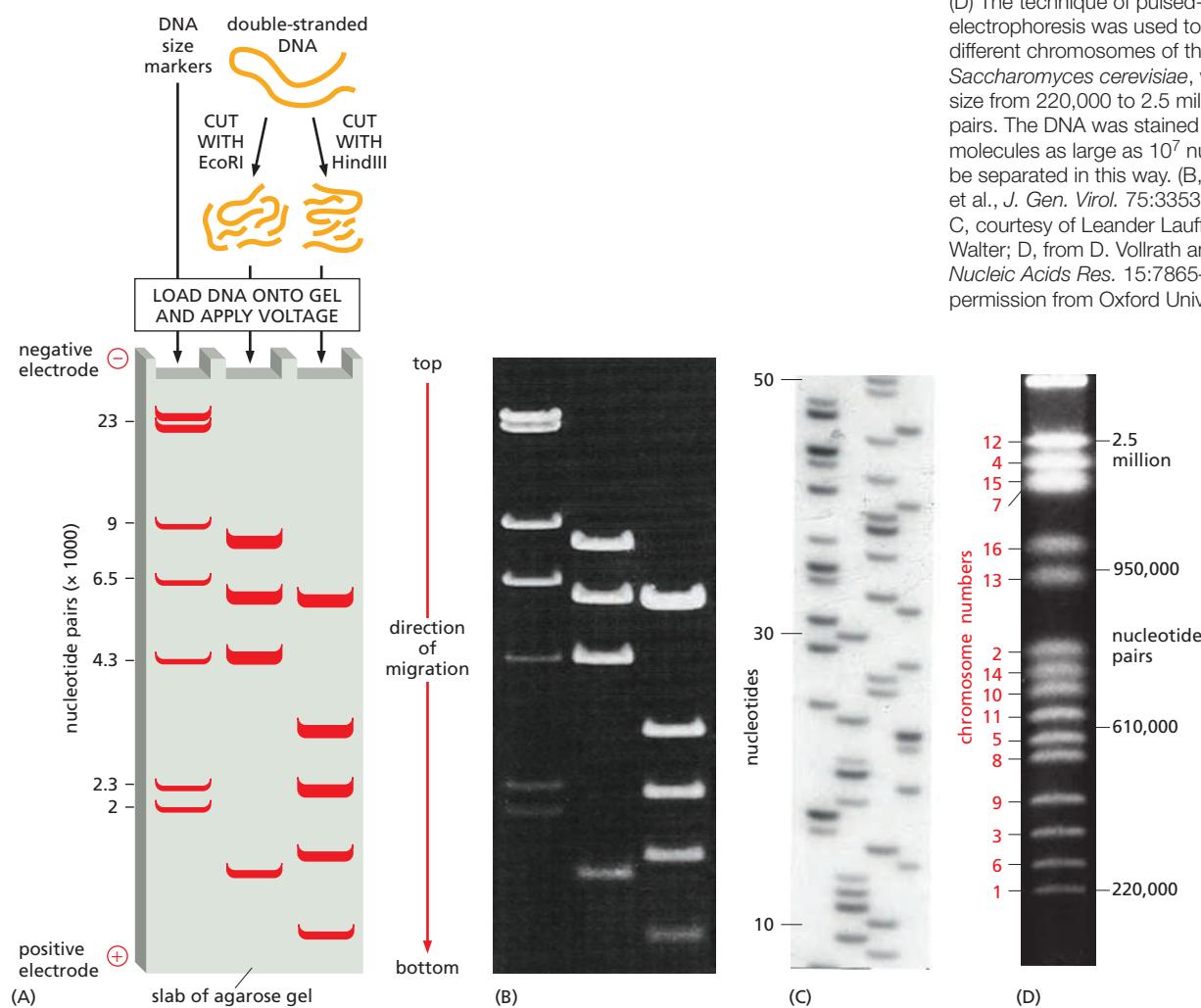
**Figure 8–24** Restriction nucleases cleave DNA at specific nucleotide sequences.

Like the sequence-specific DNA-binding proteins we encountered in Chapter 7 (see Figure 7–8), restriction enzymes often work as dimers, and the DNA sequence they recognize and cleave is often symmetrical around a central point. Here, both strands of the DNA double helix are cut at specific points within the target sequence (orange). Some enzymes, such as HaeIII, cut straight across the double helix and leave two blunt-ended DNA molecules; with others, such as EcoRI and HindIII, the cuts on each strand are staggered. These staggered cuts generate “sticky ends”—short, single-stranded overhangs that help the cut DNA molecules join back together through complementary base-pairing. This rejoining of DNA molecules becomes important for DNA cloning, as we discuss below. Restriction nucleases are usually obtained from bacteria, and their names reflect their origins: for example, the enzyme EcoRI comes from *Escherichia coli*. There are currently hundreds of different restriction enzymes available; they can be ordered from companies that commercially produce them.

A variation of agarose-gel electrophoresis, called *pulsed-field gel electrophoresis*, makes it possible to separate extremely long DNA molecules, even those found in whole chromosomes. Ordinary gel electrophoresis fails to separate very large DNA molecules because the steady electric field stretches them out so that they travel end-first through the gel in snakelike configurations at a rate that is independent of their length. In pulsed-field gel electrophoresis, by contrast, the direction of the electric field changes periodically, which forces the molecules to re-orient before continuing to move snakelike through the gel. This re-orientation takes much more time for larger molecules, so that longer molecules move more slowly than shorter ones. As a consequence, entire bacterial or yeast chromosomes separate into discrete bands in pulsed-field gels and so can be sorted and identified on the basis of their size (Figure 8–25D). Although a typical mammalian chromosome of  $10^8$  nucleotide pairs is still too long to be sorted even in this way, large segments of these chromosomes are readily separated and identified if the chromosomal DNA is first cut with a restriction nuclease selected to recognize sequences that occur only rarely.

The DNA bands on agarose or polyacrylamide gels are invisible unless the DNA is labeled or stained in some way. A particularly sensitive method of staining DNA is to soak the gel in the dye *ethidium bromide*, which fluoresces under ultraviolet light when it is bound to DNA (see Figure 8–25B and D). Even more sensitive detection methods incorporate a radioisotope or chemical marker into the DNA molecules before electrophoresis, as we next describe.

**Figure 8–25** DNA molecules can be separated by size using gel electrophoresis. (A) Schematic illustration comparing the results of cutting the same DNA molecule (in this case, the genome of a virus that infects wasps) with two different restriction nucleases, EcoRI (middle) and HindIII (right). The fragments are then separated by gel electrophoresis using a gel matrix of agarose. Because larger fragments migrate more slowly than smaller ones, the lowermost bands on the gel contain the smallest DNA fragments. The sizes of the fragments can be estimated by comparing them to a set of DNA fragments of known sizes (left). (B) Photograph of an actual agarose gel showing DNA “bands” that have been stained with ethidium bromide. (C) A polyacrylamide gel with small pores was used to separate short DNA molecules that differ by only a single nucleotide. Shown here are the results of a dideoxy sequencing reaction, explained later in this chapter. From left to right, the bands in the four lanes were produced by adding G, A, T, and C chain-terminating nucleotides (see Panel 8–1). The DNA molecules were labeled with  $^{32}\text{P}$ , and the image shown was produced by laying a piece of photographic film over the gel and allowing the  $^{32}\text{P}$  to expose the film, producing the dark bands observed when the film was developed. (D) The technique of pulsed-field agarose-gel electrophoresis was used to separate the 16 different chromosomes of the yeast species *Saccharomyces cerevisiae*, which range in size from 220,000 to 2.5 million nucleotide pairs. The DNA was stained as in (B). DNA molecules as large as  $10^7$  nucleotide pairs can be separated in this way. (B, from U. Albrecht et al., *J. Gen. Virol.* 75:3353–3363, 1994; C, courtesy of Leander Lauffer and Peter Walter; D, from D. Vollrath and R.W. Davis, *Nucleic Acids Res.* 15:7865–7876, 1987. With permission from Oxford University Press.)

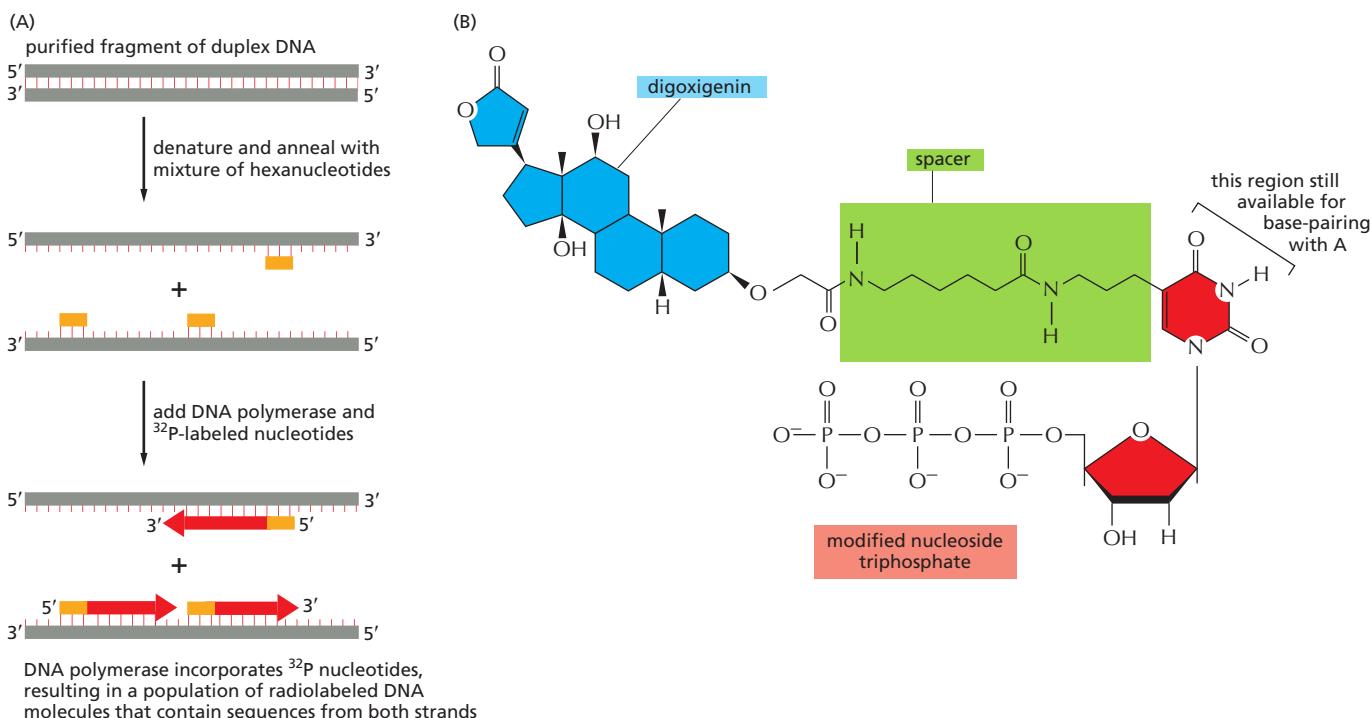


## Purified DNA Molecules Can Be Specifically Labeled with Radioisotopes or Chemical Markers *in vitro*

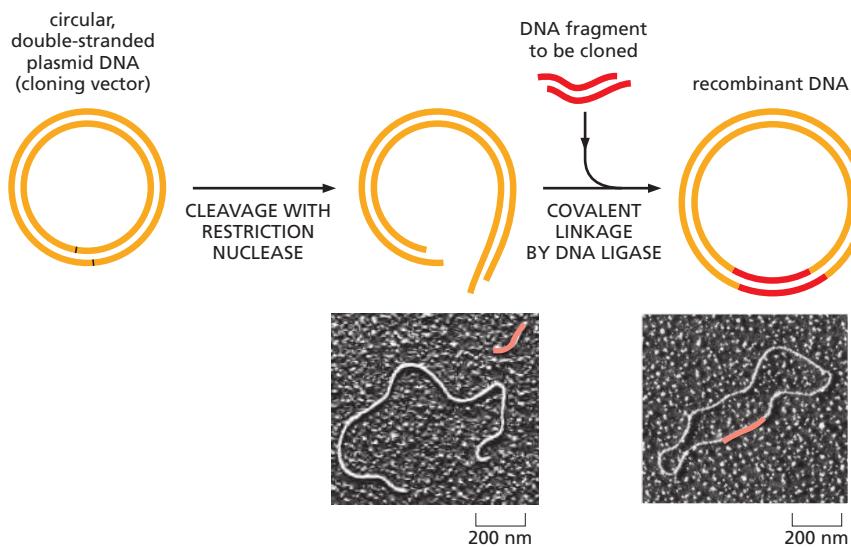
The DNA polymerases that synthesize and repair DNA (discussed in Chapter 5) have become important tools in experimentally manipulating DNA. Because they synthesize sequences complementary to an existing DNA molecule, they are often used in the test tube to create exact copies of existing DNA molecules. The copies can include specially modified nucleotides (Figure 8–26). To synthesize DNA in this way, the DNA polymerase is presented with a template and a pool of nucleotide precursors that contain the modification. As long as the polymerase can use these precursors, it automatically makes new, modified molecules that match the sequence of the template. Modified DNA molecules have many uses. DNA labeled with the radioisotope  $^{32}\text{P}$  can be detected following gel electrophoresis by placing the gel next to a piece of photographic film (see Figure 8–25C). The  $^{32}\text{P}$  atoms emit  $\beta$  particles which expose the film, producing a visible record of every band on the gel. Alternatively, the gel can be scanned by a detector that measures the  $\beta$  emissions directly. Other types of modified DNA, such as that labeled by digoxigenin (see Figure 8–26B), are useful for visualizing DNA molecules in whole cells, a topic we discuss later in this chapter.

### Genes Can Be Cloned Using Bacteria

Any DNA fragment can be cloned. In molecular biology, the term **DNA cloning** is used in two senses. It literally refers to the act of making many identical copies (typically billions) of a DNA molecule—the amplification of a particular DNA sequence. However, the term also describes the isolation of a particular stretch of DNA (often a particular gene) from the rest of the cell's genome; the same term



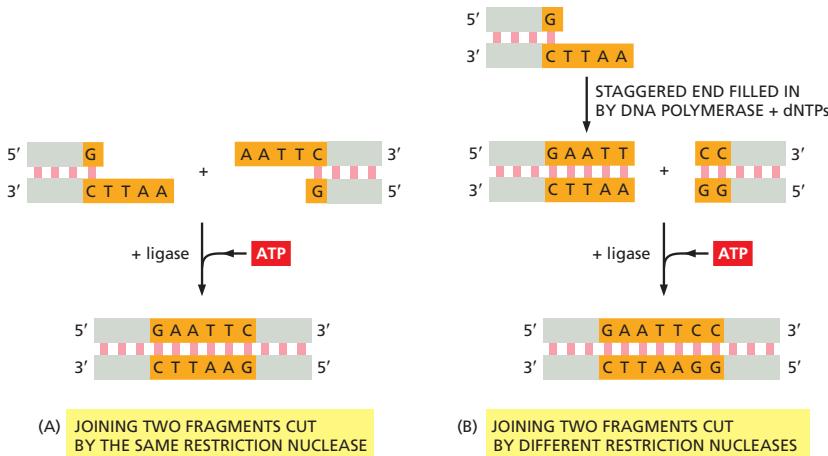
**Figure 8–26** Methods for labeling DNA molecules *in vitro*. (A) A purified DNA polymerase enzyme can incorporate radiolabeled nucleotides as it synthesizes new DNA molecules. In this way, radiolabeled versions of any DNA sequence can be prepared in the laboratory. (B) The method in (A) is also used to produce nonradioactive DNA molecules that carry a specific chemical marker that can be detected with an appropriate antibody. The base on the nucleoside triphosphate shown is an analog of thymine, in which the methyl group on T has been replaced by a spacer arm linked to the plant steroid digoxigenin. An anti-digoxigenin antibody coupled to a visible marker such as a fluorescent dye is then used to visualize the DNA. Other chemical labels, such as biotin, can be attached to nucleotides and used in the same way. The only requirements are that the modified nucleotides properly base-pair and appear “normal” to the DNA polymerase.



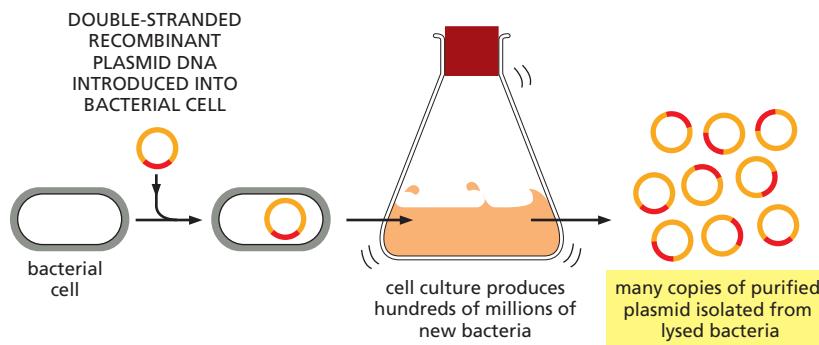
**Figure 8–27** The insertion of a DNA fragment into a bacterial plasmid with the enzyme DNA ligase. The plasmid is cut open with a restriction nuclease (in this case, one that produces staggered ends) and is mixed with the DNA fragment to be cloned (which has been prepared with the same restriction nuclease). DNA ligase and ATP are added. The staggered ends base-pair, and DNA ligase seals the nicks in the DNA backbone, producing a complete recombinant DNA molecule. In the accompanying micrographs, the inserted DNA is colored red. (Micrographs courtesy of Huntington Potter and David Dressler.)

is used because this isolation is usually accomplished by making many identical copies of only the DNA of interest. We note that elsewhere in the book, cloning, particularly when used in the context of developmental biology, can also refer to the generation of many genetically identical cells starting from a single cell or even to the generation of genetically identical organisms (see, for example, Figure 7–2). In all cases, cloning refers to the act of making many identical copies, and in this section, we use the term to refer to methods designed to generate many identical copies of a defined segment of nucleic acid.

DNA cloning can be accomplished in several ways. One of the simplest involves inserting a particular fragment of DNA into the purified DNA genome of a self-replicating genetic element—usually a plasmid. The **plasmid vectors** most widely used for gene cloning are small, circular molecules of double-stranded DNA derived from plasmids that occur naturally in bacterial cells. They generally account for only a minor fraction of the total host bacterial cell DNA, but owing to their small size, they can easily be separated from the much larger chromosomal DNA molecules, which precipitate as a pellet upon centrifugation. For use as cloning vectors, the purified plasmid DNA circles are first cut with a restriction nuclease to create linear DNA molecules. The DNA to be cloned is added to the cut plasmid and then covalently joined using the enzyme DNA ligase (Figure 8–27 and Figure 8–28). As discussed in Chapter 5, this enzyme is used by the cell to stitch together the Okazaki fragments produced during DNA replication. The recombinant DNA circle is introduced back into bacterial cells that have been made transiently permeable to DNA. As the cells grow and divide, doubling in number every 30 minutes, the recombinant plasmids also replicate to produce an



**Figure 8–28** DNA ligase can join together any two DNA fragments *in vitro* to produce recombinant DNA molecules. ATP provides the energy necessary to reseal the sugar-phosphate backbone of DNA (see Figure 5–12). (A) DNA ligase can readily join two DNA fragments produced by the same restriction nuclease, in this case EcoRI. Note that the staggered ends produced by this enzyme enable the ends of the two fragments to base-pair correctly with each other, greatly facilitating their rejoining. (B) DNA ligase can also be used to join DNA fragments produced by different restriction nucleases—for example, EcoRI and HaeIII. In this case, before the fragments undergo ligation, DNA polymerase plus a mixture of deoxyribonucleoside triphosphates (dNTPs) are used to fill in the staggered cut produced by EcoRI. Each DNA fragment shown in the figure is oriented so that its 5' ends are at the left end of the upper strand and the right end of the lower strand, as indicated.



**Figure 8–29** A DNA fragment can be replicated inside a bacterial cell. To clone a particular fragment of DNA, it is first inserted into a plasmid vector, as shown in Figure 8–27. The resulting recombinant plasmid is then introduced into a bacterium, where it is replicated many millions of times as the bacterium multiplies. For simplicity, the genome of the bacterial cell is not shown.

enormous number of copies of DNA circles containing the foreign DNA (Figure 8–29). Once the cells are lysed and the plasmid DNA isolated, the cloned DNA fragment can be readily recovered by cutting it out of the plasmid DNA with the same restriction nuclease that was used to insert it, and then separating it from the plasmid DNA by gel electrophoresis. Together, these steps allow the amplification and purification of any segment of DNA from the genome of any organism.

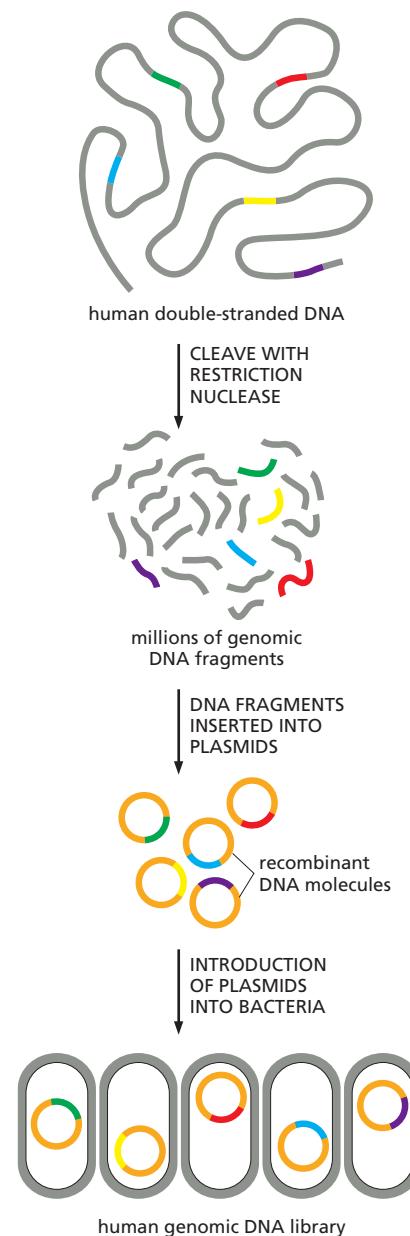
A particularly useful plasmid vector is based on the naturally occurring F plasmid of *E. coli*. Unlike smaller bacterial plasmids, the F plasmid—and its engineered derivative, the **bacterial artificial chromosome (BAC)**—is present in only one or two copies per *E. coli* cell. The fact that BACs are kept in such low numbers means that they can stably maintain very long DNA sequences, up to 1 million nucleotide pairs in length. With only a few BACs present per bacterium, it is less likely that the cloned DNA fragments will become scrambled by recombination with sequences carried on other copies of the plasmid. Because of their stability, ability to accept large DNA inserts, and ease of handling, BACs are now the preferred vector for handling large fragments of foreign DNA. As we will see below, BACs were instrumental in determining the complete nucleotide sequence of the human genome.

### An Entire Genome Can Be Represented in a DNA Library

Often it is useful to break up a genome into much smaller fragments and clone every fragment, separately, using a plasmid vector. This approach is useful because it allows scientists to work with easily managed, discrete pieces of a genome instead of whole, unwieldy chromosomes.

This strategy involves cleaving genomic DNA into small pieces using a restriction nuclease (or, in some cases, by mechanically shearing the DNA) and ligating the entire collection of DNA fragments into plasmid vectors, using conditions that favor the insertion of a single DNA fragment into each plasmid molecule. These recombinant plasmids are then introduced into *E. coli* at a concentration that ensures that no more than one plasmid molecule is taken up by each bacterium. The collection of cloned plasmid molecules is known as a **DNA library**. Because the DNA fragments were derived directly from the chromosomal DNA of the organism of interest, the resulting collection—called a **genomic library**—will represent the entire genome of that organism (Figure 8–30), spread out over tens of thousands of individual bacterial colonies.

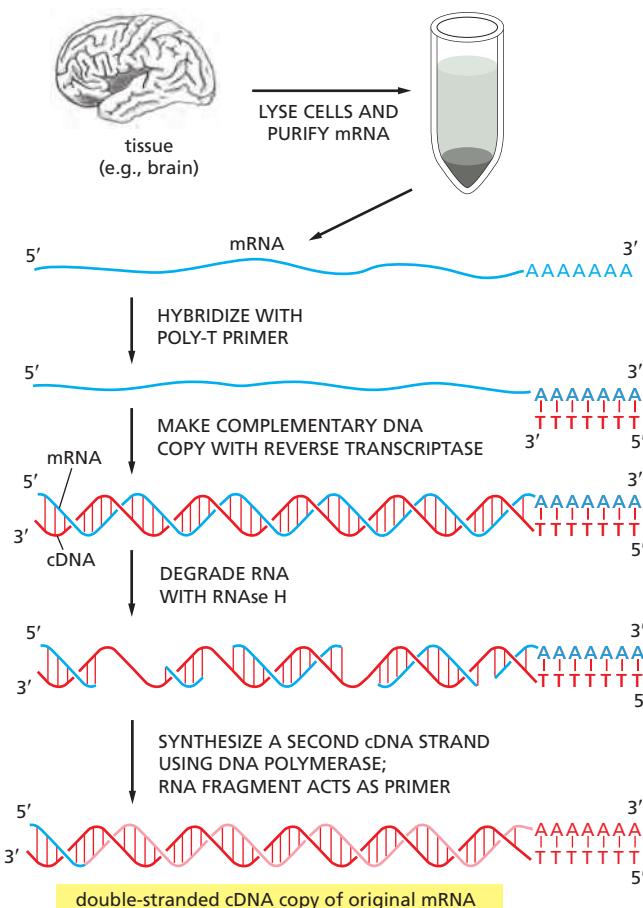
An alternative strategy, one that enriches for protein-coding genes, is to begin the cloning process by selecting only those DNA sequences that are transcribed into mRNA and thus correspond to protein-encoding genes. This is done by extracting the mRNA from cells and then making a DNA copy of each mRNA



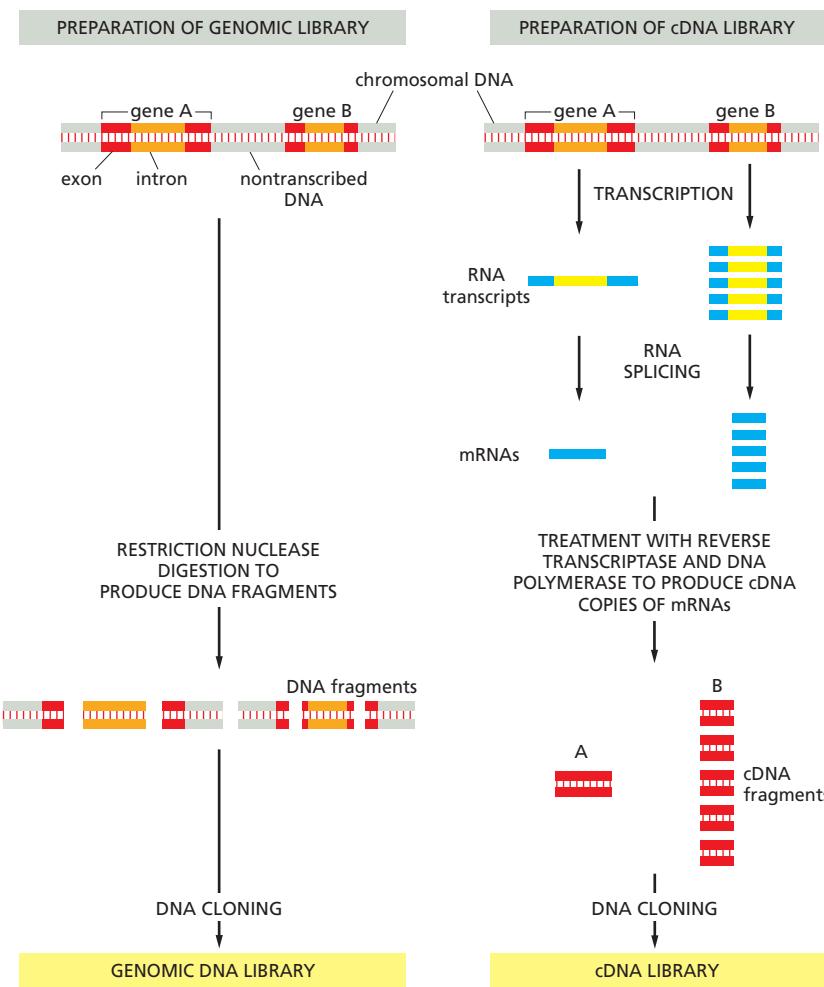
**Figure 8–30** Human genomic libraries containing DNA fragments that represent the whole human genome can be constructed using restriction nucleases and DNA ligase. Such a genomic library consists of a set of bacteria, each carrying a different fragment of human DNA. For simplicity, only the colored DNA fragments are shown in the library; in reality, all of the different gray fragments will also be represented.

molecule present—a so-called *complementary DNA*, or *cDNA*. The copying reaction is catalyzed by the reverse transcriptase enzyme of retroviruses, which synthesizes a complementary DNA chain on an RNA template. The single-stranded cDNA molecules synthesized by the reverse transcriptase are converted by DNA polymerase into double-stranded cDNA molecules, and these molecules are inserted into a plasmid or virus vector and cloned (Figure 8–31). Each clone obtained in this way is called a **cDNA clone**, and the entire collection of clones derived from one mRNA preparation constitutes a **cDNA library**.

**Figure 8–32** illustrates some important differences between genomic DNA clones and cDNA clones. Genomic clones represent a random sample of all of the DNA sequences in an organism—both coding and noncoding—and, with very rare exceptions, are the same regardless of the cell type used to prepare them. By contrast, cDNA clones contain only those regions of the genome that have been transcribed into mRNA. Because the cells of different tissues produce distinct sets of mRNA molecules, a distinct cDNA library is obtained for each type of cell used to prepare the library.



**Figure 8–31 The synthesis of cDNA.** Total mRNA is extracted from a particular tissue, and the enzyme reverse transcriptase (see Figure 5–62) is used to produce DNA copies (cDNA) of the mRNA molecules. For simplicity, the copying of just one of these mRNAs into cDNA is illustrated. A short oligonucleotide complementary to the poly-A tail at the 3' end of the mRNA (discussed in Chapter 6) is first hybridized to the RNA to act as a primer for the reverse transcriptase, which then copies the RNA into a complementary DNA chain, thereby forming a DNA–RNA hybrid helix. Treating the DNA–RNA hybrid with a specialized nuclease (RNase H) that attacks only the RNA produces nicks and gaps in the RNA strand. DNA polymerase then copies the remaining single-stranded cDNA into double-stranded cDNA. Because DNA polymerase can synthesize through the bound RNA molecules, the RNA fragment that is base-paired to the 3' end of the first DNA strand usually acts as the primer for the second strand synthesis, as shown. Any remaining RNA is eventually degraded during subsequent cloning steps. As a result, the nucleotide sequences at the extreme 5' ends of the original mRNA molecules are often absent from cDNA libraries.



**Figure 8–32** The differences between cDNA clones and genomic DNA clones derived from the same region of DNA. In this example, gene A is infrequently transcribed, whereas gene B is frequently transcribed, and both genes contain introns (orange). In the genomic DNA library, both the introns and the nontranscribed DNA (gray) are included in the clones, and most clones contain, at most, only part of the coding sequence of a gene (red). In the cDNA clones, the intron sequences (yellow) have been removed by RNA splicing during the formation of the mRNA (blue), and a continuous coding sequence is therefore present in each clone. Because gene B is transcribed more frequently than gene A in the cells from which the cDNA library was made, it is represented much more frequently than A in the cDNA library. In contrast, A and B are represented equally in the genomic DNA library.

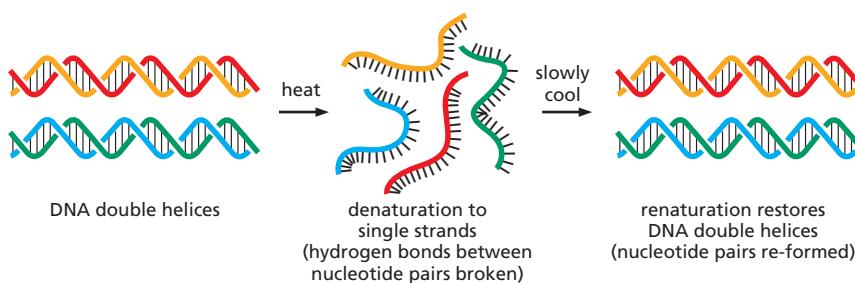
## Genomic and cDNA Libraries Have Different Advantages and Drawbacks

Genomic libraries are especially useful in determining the nucleotide sequences of a whole genome. For example, to determine the nucleotide sequence of the human genome, it was broken up into roughly 100,000-nucleotide-pair pieces, each of which was inserted into a BAC plasmid and amplified in *E. coli*. The resulting genomic library consisted of tens of thousands of bacterial colonies, each containing a different human DNA insert. The nucleotide sequence of each insert was determined separately and the sequence of the entire genome was stitched together from the pieces.

The most important advantage of cDNA clones, over genomic clones, is that they contain the uninterrupted coding sequence of a gene. When the aim of the cloning, for example, is to produce the protein in large quantities by expressing the cloned gene in a bacterial or yeast cell, it is more preferable to start with cDNA.

Genomic and cDNA libraries are inexhaustible resources, which are widely shared among investigators. Today, many such libraries are also available from commercial sources. Because the identity of each insert in a library is often known (through sequencing the insert), it is often possible to order a particular region of a chromosome (or, in the case of cDNA, a complete, intron-less protein-coding gene) and have it delivered by mail.

Cloning DNA by using bacteria revolutionized the study of genomes and is still in wide use today. However, there is an even simpler way to clone DNA, one that can be carried out entirely *in vitro*. We discuss this approach, called the *polymerase chain reaction*, below. However, first we need to review a fundamental, far-reaching property of DNA and RNA called *hybridization*.



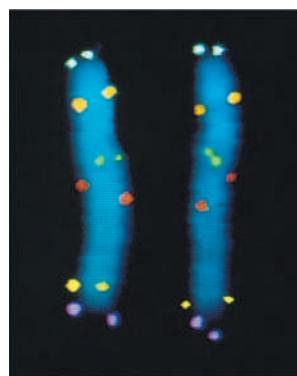
**Figure 8–33** A molecule of DNA can undergo denaturation and renaturation (hybridization). For two single-stranded molecules to hybridize, they must have complementary nucleotide sequences that allow base-pairing. In this example, the red and orange strands are complementary to each other, and the blue and green strands are complementary to each other. Although denaturation by heating is shown, DNA can also be renatured after being denatured by alkali treatment.

### Hybridization Provides a Powerful, But Simple Way to Detect Specific Nucleotide Sequences

Under normal conditions, the two strands of a DNA double helix are held together by hydrogen bonds between the complementary base pairs (see Figure 4–3). But these relatively weak, noncovalent bonds can be fairly easily broken. Such *DNA denaturation* will release the two strands from each other, but does not break the covalent bonds that link together the nucleotides within each strand. Perhaps the simplest way to achieve this separation involves heating the DNA to around 90°C. When the conditions are reversed—by slowly lowering the temperature—the complementary strands will readily come back together to re-form a double helix. This **hybridization**, or *DNA renaturation*, is driven by the re-formation of the hydrogen bonds between complementary base pairs (Figure 8–33). We saw in Chapter 5 that DNA hybridization underlies the crucial process of homologous recombination (see Figure 5–47).

This fundamental capacity of a single-stranded nucleic acid molecule, either DNA or RNA, to form a double helix with a single-stranded molecule of a complementary sequence provides a powerful and sensitive technique for detecting specific nucleotide sequences. Today, one simply designs a short, single-stranded DNA molecule (often called a *DNA probe*) that is complementary to the nucleotide sequence of interest. Because the nucleotide sequences of so many genomes are known—and are stored in publicly accessible databases—designing a probe to hybridize anywhere in a genome is straightforward. Probes are single-stranded, typically 30 nucleotides in length, and are usually synthesized chemically by a commercial service for pennies per nucleotide. A DNA sequence of 30 nucleotides will occur by chance only once every  $1 \times 10^{18}$  nucleotides ( $4^{30}$ ); so, even in the human genome of  $3 \times 10^9$  nucleotide pairs, a DNA probe designed to match a particular 30-nucleotide sequence will be highly unlikely to hybridize—by chance—anywhere else on the genome. This, of course, presumes that the sequence complementary to the probe does not occur multiple times in the genome, a condition that can be checked beforehand by scanning the genomic sequence *in silico* (using a computer) and designing probes that match only one spot. The hybridization conditions can be set so that even a single mismatch will prevent hybridization to “near-miss” sequences. The exquisite specificity of nucleic acid hybridization can be easily appreciated by the *in situ* (Latin for “in place”) **hybridization** experiment shown in Figure 8–34. As we will see throughout this chapter, nucleic acid

**Figure 8–34** *In situ* hybridization can be used to locate genes on isolated chromosomes. Here, six different DNA probes have been used to mark the locations of their complementary nucleotide sequences on human Chromosome 5, isolated from a mitotic cell in metaphase (see Figure 4–59 and Panel 17–1, pp. 980–981). The DNA probes have been labeled with different chemical groups (see Figure 8–26B) and are detected using fluorescent antibodies specific for those groups. The chromosomal DNA has been partially denatured to allow the probes to base-pair with their complementary sequences. Both the maternal and paternal copies of Chromosome 5 are shown, aligned side by side. Each probe produces two dots on each chromosome because chromosomes undergoing mitosis have already replicated their DNA; therefore, each chromosome contains two identical DNA helices. The technique employed here is nicknamed FISH, for fluorescence *in situ* hybridization. (Courtesy of David C. Ward.)



hybridization has many uses in modern cell and molecular biology; one of the most powerful is in the cloning of DNA by the polymerase chain reaction, as we next discuss.

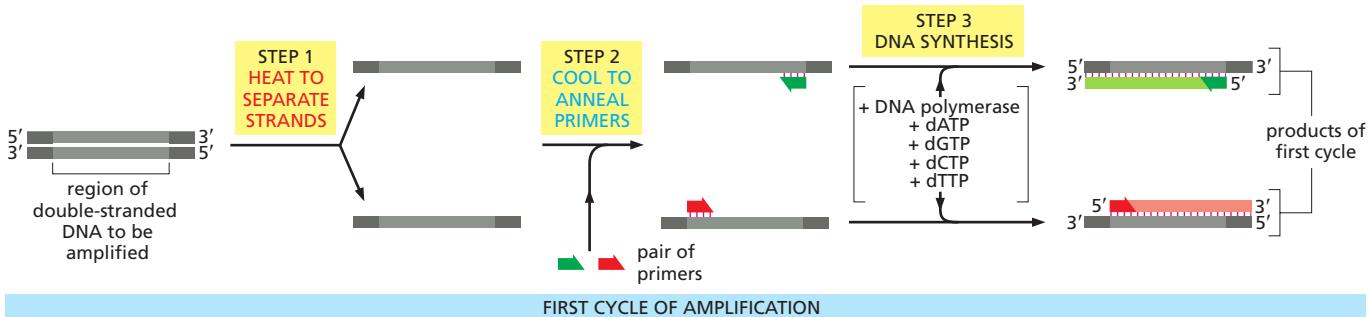
### Genes Can Be Cloned *in vitro* Using PCR

Genomic and cDNA libraries were once the only route to cloning genes and they are still used for cloning very large genes and for sequencing whole genomes. However, a powerful and versatile method for amplifying DNA, known as the **polymerase chain reaction (PCR)**, provides a more rapid and straightforward approach to DNA cloning, particularly in organisms whose complete genome sequence is known. Today, since genome sequences are abundant, most cloning is carried out by PCR.

Invented in the 1980s, PCR revolutionized the way that DNA and RNA are analyzed. The technique can amplify any nucleotide sequence selectively and is performed entirely in a test tube. Eliminating the need for bacteria makes PCR convenient and rapid—billions of copies of a nucleotide can be generated in a matter of hours. Starting with an entire genome, PCR allows DNA from a specified region—selected by the experimenter—to be greatly amplified, effectively “purifying” this DNA away from the remainder of the genome, which remains unamplified. Because of its power to greatly amplify nucleic acids, PCR is remarkably sensitive: the method can be used to detect the trace amounts of DNA in a drop of blood left at a crime scene or in a few copies of a viral genome in a patient’s blood sample.

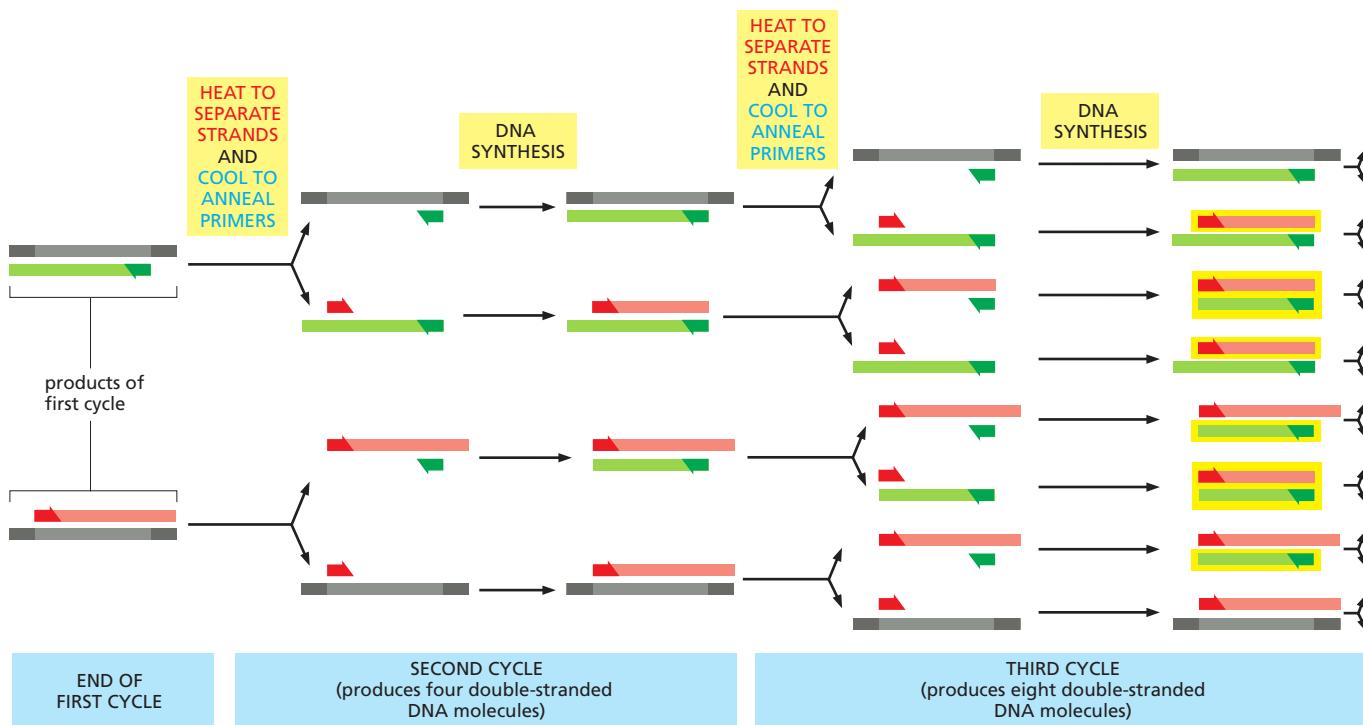
The success of PCR depends both on the selectivity of DNA hybridization and on the ability of DNA polymerase to copy a DNA template faithfully through repeated rounds of replication *in vitro*. As discussed in Chapter 5, this enzyme adds nucleotides to the 3' end of a growing strand of DNA (see Figure 5–4). To copy DNA, the polymerase requires a primer—a short nucleotide sequence that provides a 3' end from which synthesis can begin. For PCR, the primers are designed by the experimenter, synthesized chemically, and, by hybridizing to genomic DNA, “tell” the polymerase which part of the genome to copy. As discussed in the previous section, *DNA primers* (in essence, the same type of molecules as DNA probes but without a radioactive or fluorescent label) can be designed to uniquely locate any position on a genome.

PCR is an iterative process in which the cycle of amplification is repeated dozens of times. At the start of each cycle, the two strands of the double-stranded DNA template are separated and a different primer is annealed to each. These primers mark the right and left boundaries of the DNA to be amplified. DNA polymerase is then allowed to replicate each strand independently (**Figure 8–35**). In



**Figure 8–35** A pair of primers directs the synthesis of a desired segment of DNA in a test tube. Each cycle of PCR includes three steps: (1) The double-stranded DNA is heated briefly to separate the two strands. (2) The DNA is exposed to a large excess of a pair of specific primers—designed to bracket the region of DNA to be amplified—and the sample is cooled to allow the primers to hybridize to complementary sequences in the two DNA strands. (3) This mixture is incubated with DNA polymerase and the four deoxyribonucleoside triphosphates so that DNA can be synthesized, starting from the two primers. To amplify the DNA, the cycle is repeated many times by reheating the sample to separate the newly synthesized DNA strands (see Figure 8–36).

The technique depends on the use of a special DNA polymerase isolated from a thermophilic bacterium; this polymerase is stable at much higher temperatures than eukaryotic DNA polymerases, so it is not denatured by the heat treatment shown in step 1. The enzyme therefore does not have to be added again after each cycle.



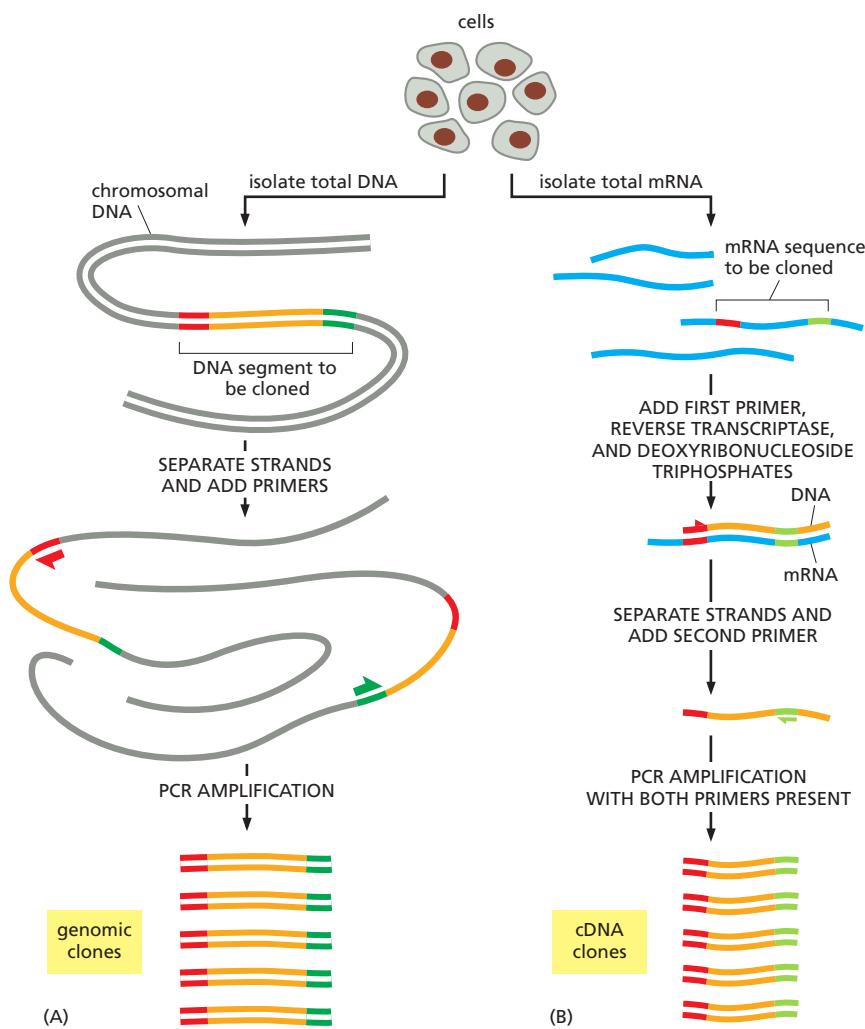
**Figure 8–36** PCR uses repeated rounds of strand separation, hybridization, and synthesis to amplify DNA. As the procedure outlined in Figure 8–35 is repeated, all the newly synthesized fragments serve as templates in their turn. Because the polymerase and the primers remain in the sample after the first cycle, PCR involves simply heating and then cooling the same sample, in the same test tube, again and again. Each cycle doubles the amount of DNA synthesized in the previous cycle, so that within a few cycles, the predominant DNA is identical to the sequence bracketed by and including the two primers in the original template. In the example illustrated here, three cycles of reaction produce 16 DNA chains, 8 of which (boxed in yellow) correspond exactly to one or the other strand of the original bracketed sequence. After four more cycles, 240 of the 256 DNA chains will correspond exactly to the original sequence, and after several more cycles, essentially all of the DNA strands will be this length. Typically, 20–30 cycles are carried out to effectively clone a region of DNA starting from genomic DNA; the rest of the genome remains unamplified, and its concentration is therefore negligible compared with that of the amplified region ([Movie 8.2](#)).

subsequent cycles, all the newly synthesized DNA molecules produced by the polymerase serve as templates for the next round of replication (Figure 8–36). Through this iterative amplification process, many copies of the original sequence can be made—billions after about 20 to 30 cycles.

PCR is now the method of choice for cloning relatively short DNA fragments (say, under 10,000 nucleotide pairs). Each cycle takes only about five minutes, and automation of the whole procedure enables cell-free cloning of a DNA fragment in a few hours. The original template for PCR can be either DNA or RNA, so this method can be used to obtain either a genomic clone (complete with introns and exons) or a cDNA copy of an mRNA (Figure 8–37).

### PCR Is Also Used for Diagnostic and Forensic Applications

The PCR method is extraordinarily sensitive; it can detect a single DNA molecule in a sample if at least part of the sequence of that molecule is known. Trace amounts of RNA can be analyzed in the same way by first transcribing them into DNA with reverse transcriptase. For these reasons, PCR is frequently employed for uses that go beyond simple cloning. For example, it can be used to detect invading pathogens at very early stages of infection. In this case, short sequences complementary to a segment of the infectious agent's genome are used as primers and following many cycles of amplification, even a few copies of an invading bacterial or viral genome in a patient's sample can be detected (Figure 8–38). For many infections, PCR has replaced the use of antibodies against microbial molecules to

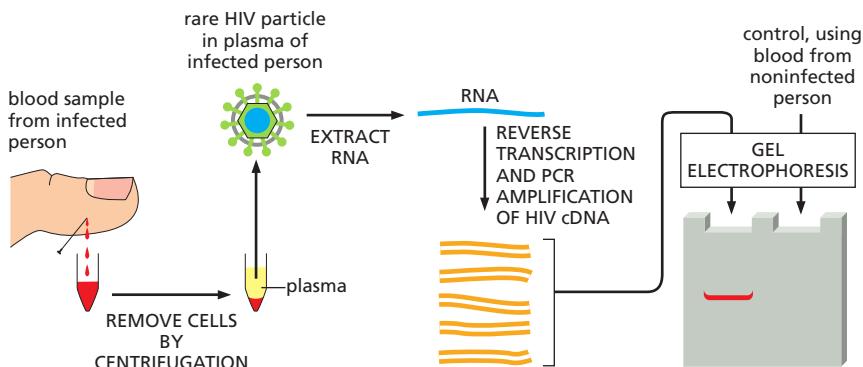


**Figure 8–37** PCR can be used to obtain either genomic or cDNA clones.

(A) To use PCR to clone a segment of chromosomal DNA, total genomic DNA is first purified from cells. PCR primers that flank the stretch of DNA to be cloned are added, and many cycles of PCR are completed (see Figure 8–36). Because only the DNA between (and including) the primers is amplified, PCR provides a way to obtain selectively any short stretch of chromosomal DNA in an effectively pure form. (B) To use PCR to obtain a cDNA clone of a gene, total mRNA is first purified from cells. The first primer is added to the population of mRNAs, and reverse transcriptase is used to make a DNA strand complementary to the specific RNA sequence of interest. The second primer is then added, and the DNA molecule is amplified through many cycles of PCR.

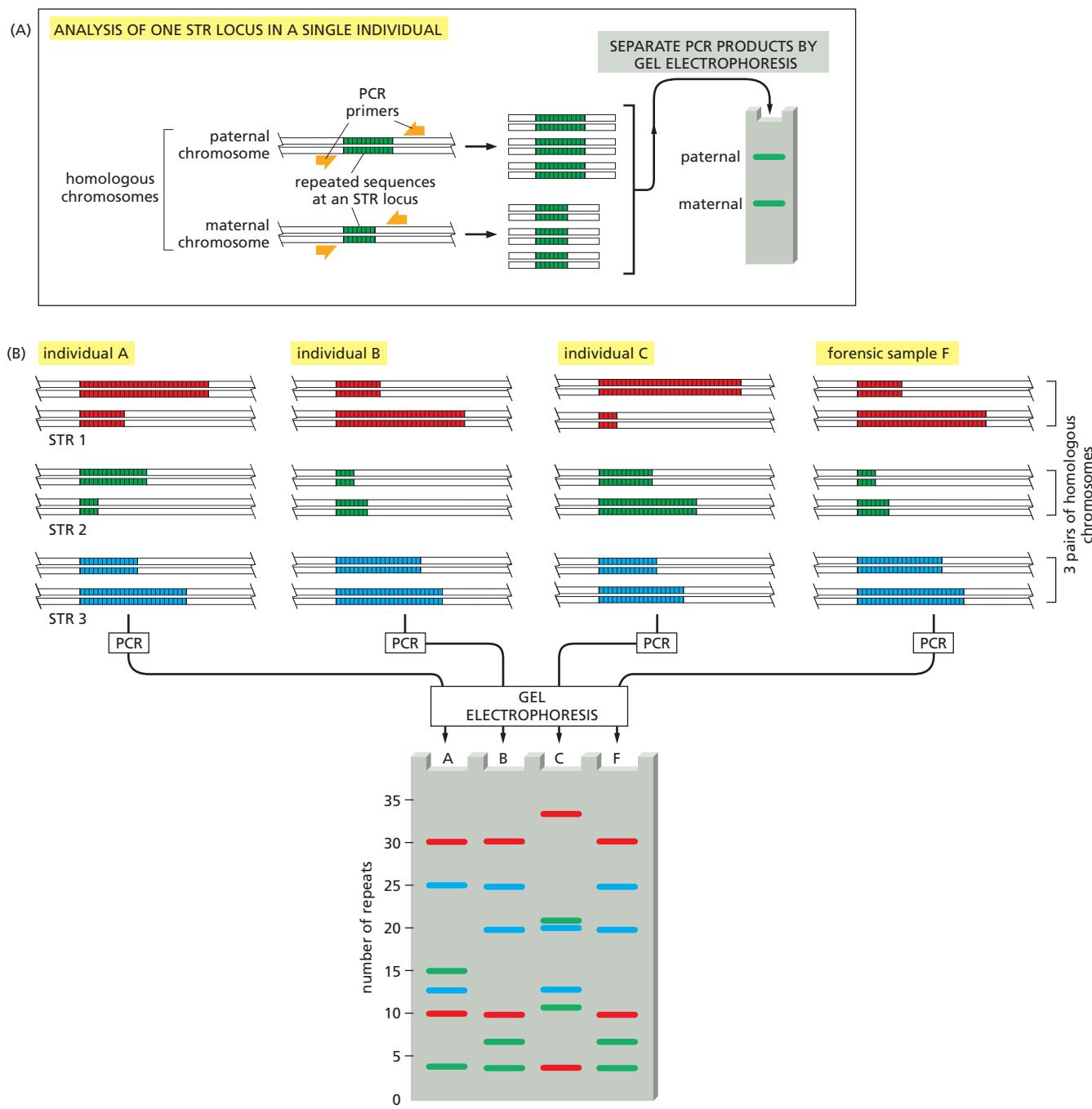
detect the presence of the invader. It is also used to verify the authenticity of a food source—for example, whether a sample of beef actually came from a cow.

Finally, PCR is now widely used in forensics. The method's extreme sensitivity allows forensic investigators to isolate DNA from minute traces of human blood or other tissue to obtain a *DNA fingerprint* of the person who left the sample behind.



**Figure 8–38** PCR can be used to detect the presence of a viral genome in a sample of blood.

Because of its ability to amplify enormously the signal from a single molecule of nucleic acid, PCR is an extraordinarily sensitive method for detecting trace amounts of virus in a sample of blood or tissue, without the need to purify the virus. For HIV, the virus that causes AIDS, the genome is a single-stranded molecule of RNA, as illustrated here. In addition to HIV, many other viruses that infect humans are now detected in this way.



**Figure 8–39** PCR is used in forensic science to distinguish one individual from another. The DNA sequences analyzed are short tandem repeats (STRs) composed of sequences such as CACACA... or GTGTGT... STRs are found in various positions (loci) in the human genome. The number of repeats in each STR locus is highly variable in the population, ranging from 4 to 40 in different individuals. Because of the variability in these sequences, individuals will usually inherit a different number of repeats at each STR locus from their mother and from their father; two unrelated individuals, therefore, rarely contain the same pair of sequences at a given STR locus. (A) PCR using primers that recognize unique sequences on either side of one particular STR locus produces a pair of bands of amplified DNA from each individual, one band representing the maternal STR variant and the other representing the paternal STR variant. The length of the amplified DNA, and thus its position after gel electrophoresis, will depend on the exact number of repeats at the locus. (B) In the schematic example shown here, the same three STR loci are analyzed in samples from three suspects (individuals A, B, and C), producing six bands for each individual. Although different people can have several bands in common, the overall pattern is quite distinctive for each person. The band pattern can therefore serve as a DNA fingerprint to identify an individual nearly uniquely. The fourth lane (F) contains the products of the same PCR amplifications carried out on a hypothetical forensic DNA sample, which could have been obtained from a single hair or a tiny spot of blood left at a crime scene.

The more loci that are examined, the more confident one can be about the results. When examining the variability at 5–10 different STR loci, the odds that two random individuals would share the same fingerprint by chance are approximately one in 10 billion. In the case shown here, individuals A and C can be eliminated from inquiries, while B is a clear suspect. A similar approach is used routinely in paternity testing.

With the possible exception of identical twins, the genome of each human differs in DNA sequence from that of every other person on Earth. Using primer pairs targeted at genome sequences that are known to be highly variable in the human population, PCR makes it possible to generate a distinctive DNA fingerprint for any individual ([Figure 8–39](#)). Such forensic analyses can be used not only to help identify those who have done wrong, but also—equally important—to exonerate those who have been wrongfully accused.

### Both DNA and RNA Can Be Rapidly Sequenced

Most current methods of manipulating DNA, RNA, and proteins rely on prior knowledge of the nucleotide sequence of the genome of interest. But how were these sequences determined in the first place? And how are new DNA and RNA molecules sequenced today? In the late 1970s, researchers developed several strategies for determining, simply and quickly, the nucleotide sequence of any purified DNA fragment. The one that became the most widely used is called **dideoxy sequencing** or **Sanger sequencing** ([Panel 8–1](#)). This method was used to determine the nucleotide sequence of many genomes, including those of *E. coli*, fruit flies, nematode worms, mice, and humans. Today, cheaper and faster methods are routinely used to sequence DNA, and even more efficient strategies are being developed (see [Panel 8–1](#)). The original “reference” sequence of the human genome, completed in 2003, cost over \$1 billion and required many scientists from around the world working together for 13 years. The enormous progress made in the past decade makes it possible for a single person to complete the sequence of an individual human genome in less than a day.

The methods summarized in [Panel 8–1](#) for rapidly sequencing DNA can also be applied to RNA. Although methods are being developed to sequence RNA directly, it is most commonly carried out by converting the RNA to complementary DNA (using reverse transcriptase) and using one of the methods described for DNA sequencing. It is important to keep in mind that although genomes remain the same from cell to cell and from tissue to tissue, the RNA produced from the genome can vary enormously. We will see later in this chapter that sequencing the entire repertoire of RNA from a cell or tissue (known as **deep RNA sequencing**, or **RNA-seq**) is a powerful way to understand how the information present in the genome is used by different cells under different circumstances. In the next section, we shall see how RNA-seq has also become a valuable tool for annotating genomes.

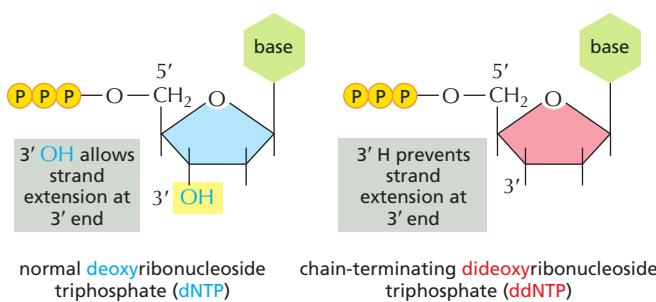
### To Be Useful, Genome Sequences Must Be Annotated

Long strings of nucleotides, at first glance, reveal nothing about how this genetic information directs the development of a living organism—or even what types of DNA, protein, and RNA molecules are produced by a genome. The process of **genome annotation** attempts to mark out all the genes (both protein-coding and noncoding) in a genome and ascribe a role to each. It also seeks to understand more subtle types of genome information, such as the *cis*-regulatory sequences that specify the time and place that a given gene is expressed and whether its mRNA undergoes alternative splicing to produce different protein isoforms. Clearly, this is a daunting task, and we are far short of completing it for any form of life, even the simplest bacterium. For many organisms, we know the approximate number of genes, and, for very simple organisms, we understand the functions of about half their genes.

In this section, we discuss broadly how genes are identified in genome sequences and what clues we can discern about their roles from simply inspecting their sequences. Later in the chapter, we turn to the more difficult problem of experimentally determining gene function.

How does one begin to make sense of a genome sequence? The first step is usually to translate *in silico* the entire genome into protein. There are six different reading frames for any piece of double-stranded DNA (three on each strand). We saw in Chapter 6 that a random sequence of nucleotides, read in frame, will

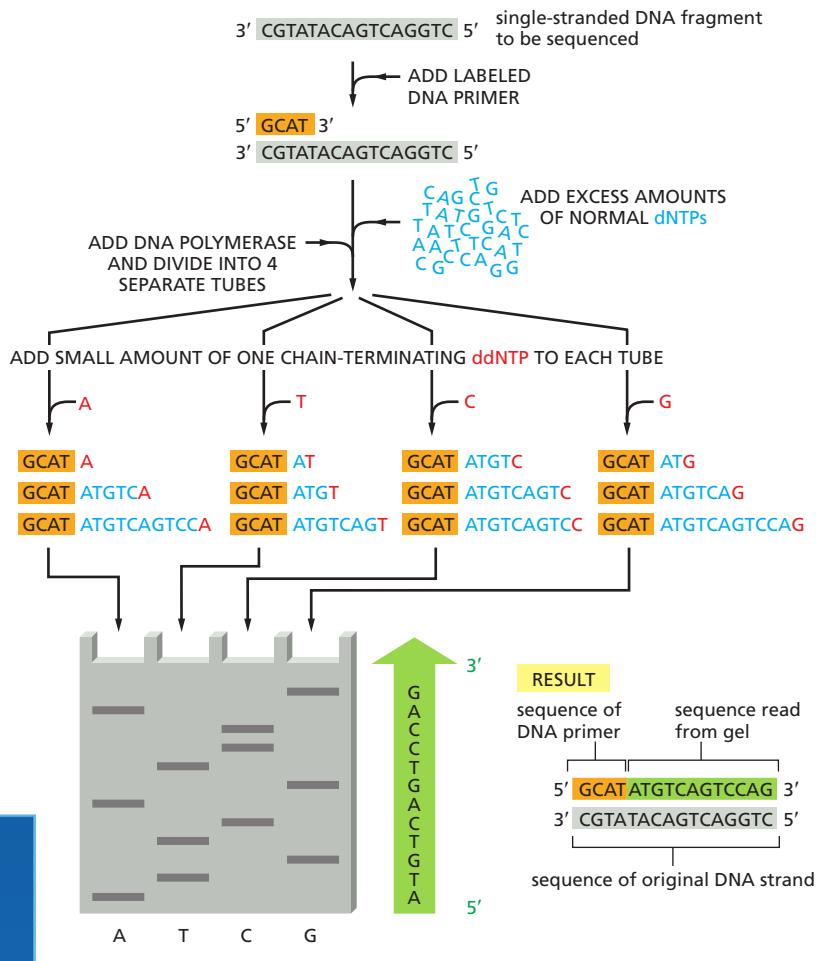
## DNA SEQUENCING



**Dideoxy sequencing, or Sanger sequencing** (named after the scientist who invented it), uses DNA polymerase, along with special chain-terminating nucleotides called dideoxyribonucleoside triphosphates (left), to make partial copies of the DNA fragment to be sequenced. These ddNTPs are derivatives of the normal deoxyribonucleoside triphosphates that lack the 3' hydroxyl group. When incorporated into a growing DNA strand, they block further elongation of that strand.

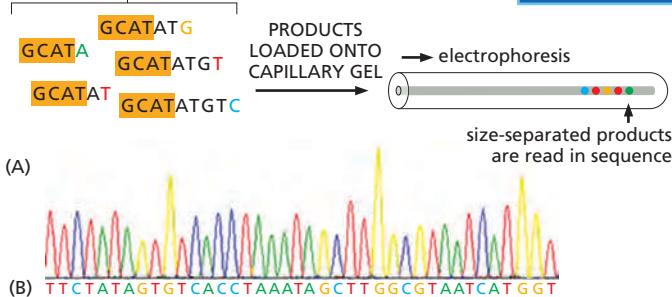
## MANUAL DIDEOXY SEQUENCING

To determine the complete sequence of a single-stranded fragment of DNA (gray), the DNA is first hybridized with a short DNA primer (orange) that is labeled with a fluorescent dye or radioisotope. DNA polymerase and an excess of all four normal deoxyribonucleoside triphosphates (blue A, C, G, or T) are added to the primed DNA, which is then divided into four reaction tubes. Each of these tubes receives a small amount of a single chain-terminating dideoxyribonucleoside triphosphate (red A, C, G, or T). Because these will be incorporated only occasionally, each reaction produces a set of DNA copies that terminate at different points in the sequence. The products of these four reactions are separated by electrophoresis in four parallel lanes of a polyacrylamide gel (labeled here A, T, C, and G). In each lane, the bands represent fragments that have terminated at a given nucleotide but at different positions in the DNA. By reading off the bands in order, starting at the bottom of the gel and reading across all lanes, the DNA sequence of the newly synthesized strand can be determined (see Figure 8-25C). The sequence, which is given in the green arrow to the right of the gel, is complementary to the sequence of the original gray single-stranded DNA.



## AUTOMATED DIDEOXY SEQUENCING

mixture of DNA products, each containing a chain-terminating ddNTP labeled with a different fluorescent marker



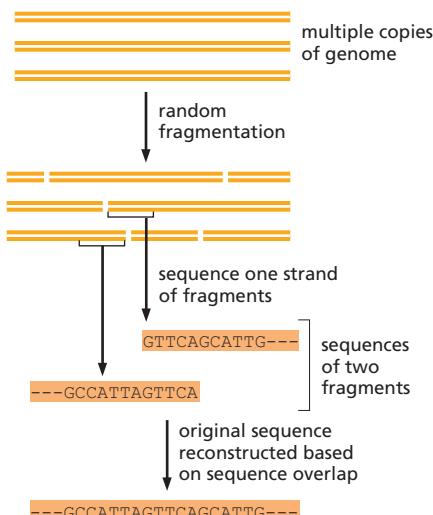
Fully automated machines can run dideoxy sequencing reactions. (A) The automated method uses an excess amount of normal dNTPs plus a mixture of four different chain-terminating ddNTPs, each of which is labeled with a fluorescent tag of a different color. The reaction products are loaded onto a long, thin capillary gel and separated by electrophoresis. A camera (not shown) reads the color of each band as it moves through the gel and feeds the data to a computer that assembles the sequence. (B) A tiny part of the data from such an automated sequencing run. Each colored peak represents a nucleotide in the DNA sequence.

## SEQUENCING WHOLE GENOMES

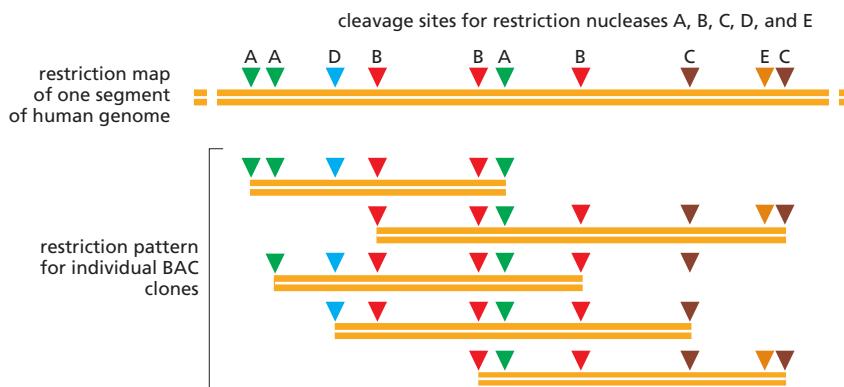
**Shotgun sequencing:** To determine the nucleotide sequence of a whole genome, the genomic DNA is first fragmented into small pieces and a genomic library is constructed, typically using plasmids and bacteria (see Figure 8–30). In *shotgun sequencing*, the nucleotide sequence of tens of thousands of individual clones is determined; the full genome sequence is then reconstructed by stitching together (*in silico*) the nucleotide sequence of each clone, using the overlaps between clones as a guide. The shotgun method works well for small genomes (such as those of viruses and bacteria) that lack repetitive DNA.

**BAC clones:** Most plant and animal genomes are large (often over  $10^9$  nucleotide pairs) and contain extensive amounts of repetitive DNA spread throughout the genome. Because a nucleotide sequence of a fragment of repetitive DNA will “overlap” every instance of the repeated DNA, it is difficult, if not impossible, to assemble the fragments into a unique order solely by the shotgun method.

To circumvent this problem, the human genome was first broken down into very large DNA fragments (each approximately 100,000 nucleotide pairs) and cloned into BACs (see p. 469). The order of the BACs along a chromosome was determined by comparing the pattern of restriction enzyme cleavage sites in a given BAC clone with that of the whole genome. In this way, a given BAC clone can be mapped, say, to the left arm of human Chromosome 3. Once a collection of BAC clones was



obtained that spanned the entire genome, each individual BAC was sequenced by the shotgun method. At the end, the sequences of all the BAC inserts were stitched together using the knowledge of the position of each BAC insert in the human genome. In all, approximately 30,000 BAC clones were sequenced to complete the human genome.



Thousands of genomes from individual humans have now been sequenced and it is not necessary to painstakingly reconstruct the order of DNA sequence “reads” each time; they are simply assembled using the order determined from the original human genome sequencing project. For this reason, *resequencing*, the term applied when the genome of a species is sequenced again (even though it may be from a different individual), is far easier than the original sequencing.

## SECOND-GENERATION SEQUENCING TECHNOLOGIES

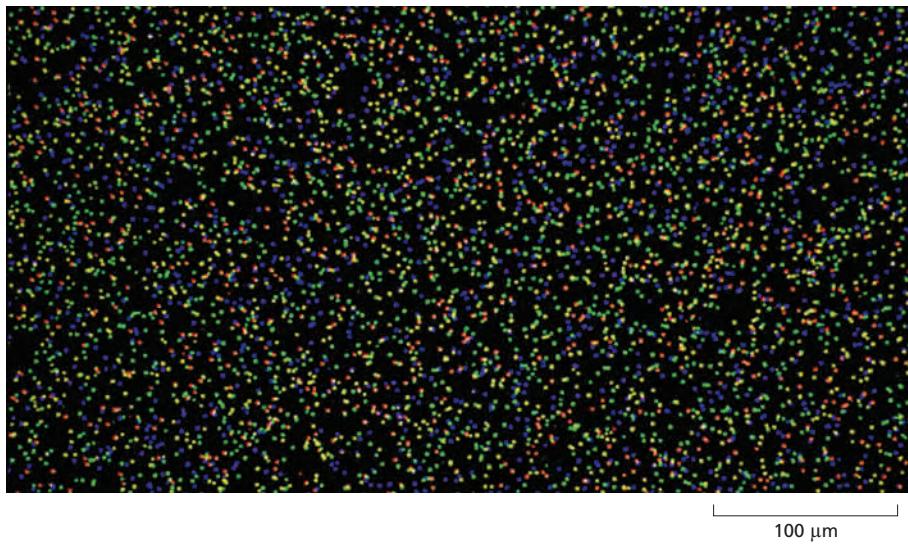
The dideoxy method made it possible to sequence the genomes of humans and most of the other organisms discussed in this book. But newer methods, developed since 2005, have made genome sequencing even more rapid—and very much cheaper. With these so-called second-generation sequencing methods, the cost of sequencing DNA has decreased dramatically. Not surprisingly, the number of genomes that have been sequenced has increased enormously. These rapid methods allow multiple genomes to be sequenced in parallel in a matter of weeks, enabling investigators to examine thousands of individual human genomes, catalog the variation in nucleotide sequences

from people around the world, and uncover the mutations that increase the risk of various diseases, from cancer to autism. These methods have also made it possible to determine the genome sequence of extinct species, including Neanderthal man and the wooly mammoth. By sequencing genomes from many closely related species, they have also made it possible to understand the molecular basis of key evolutionary events in the tree of life, such as the “inventions” of multicellularity, vision, and language. The ability to rapidly sequence DNA has had major impacts on all branches of biology and medicine; it is almost impossible to imagine where we would be without it.

## ILLUMINA® SEQUENCING

Several second-generation sequencing methods are now in wide use, and we will discuss two of the most common. Both rely on the construction of libraries of DNA fragments that represent—*in toto*—the DNA of the genome. Instead of using bacterial cells to generate these libraries, as we saw in Figure 8–30, they are made using PCR amplification of billions of DNA fragments, each attached to a solid support. The amplification is

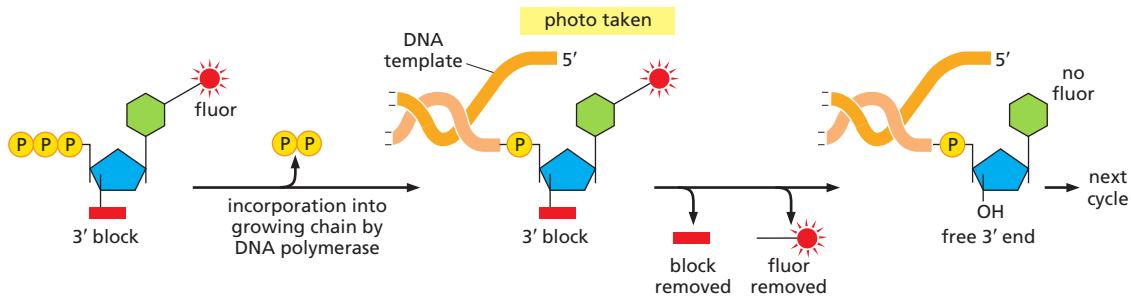
carried out so that the PCR-generated copies, instead of floating away in solution, remain bound in proximity to the original DNA fragment. This process generates clusters of DNA fragments, where each cluster contains about 1000 identical copies of a small bit of the genome. These clusters—a billion of which can fit in a single slide or plate—are then sequenced at the same time; that is, in parallel.



A slide showing individual clusters of PCR-generated DNA molecules. Each cluster carries about 1000 identical DNA molecules; the four colors are produced by incorporation of C, G, A, or T, each of which has a different color fluorophore. The image has been taken just after a fluorescent nucleotide has been incorporated into each growing DNA chain. (From Illumina Sequencing Overview, 2013.)

One method, known as *Illumina sequencing*, is based on the dideoxy method described above, but it incorporates several innovations. Here, each nucleotide is attached to a removable fluorescent molecule (a different color for each of the four bases) as well as a special chain-terminating chemical adduct: instead of a 3'-OH group, as in conventional dideoxy sequencing, the nucleotides carry a chemical group that blocks elongation by DNA polymerase but which can be removed chemically. Sequencing is then carried out as follows: the four fluorescently labeled nucleotides along with DNA polymerase are added to billions of DNA clusters immobilized on a slide. Only the appropriate nucleotide (that is complementary to the next nucleotide in the template) is covalently incorporated at each

cluster; the unincorporated nucleotides are washed away, and a high-resolution digital camera takes an image that registers which of the four nucleotides was added to the chain at each cluster. The fluorescent label and the 3'-OH blocking group are then removed enzymatically, washed away, and the process is repeated many times. In this way, billions of sequencing reactions are carried out simultaneously. By keeping track of the color changes occurring at each cluster, the DNA sequence represented by each spot can be read. Although each individual sequence read is relatively short (approximately 200 nucleotides), the billions that are carried out simultaneously can produce several human genomes worth of sequence in about a day.



**Principle behind Illumina sequencing.** This reaction is carried out stepwise, on billions of DNA clusters at once. The method relies on a color digital camera that rapidly scans all the DNA clusters after each round of modified nucleotide incorporation. The DNA sequence of each cluster is then determined by the sequence of color changes it undergoes as the elongation reaction proceeds stepwise. Each round of modified nucleotide incorporation,

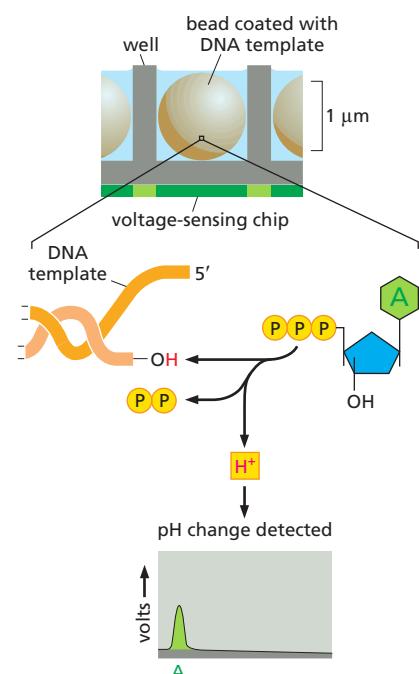
image acquisition, and removal of the 3' block and the fluorescent group takes less than an hour. Each cluster on the slide contains many copies of different, random bits of a genome; in preparing the clusters, a DNA sequence (specified by the experimenter) is joined to each copy in every cluster, and a primer complementary to this sequence is used to begin the elongation reaction by DNA polymerase.

## ION TORRENT™ SEQUENCING

Another widely used strategy for rapid DNA sequencing is called the *ion torrent* method. Here, a genome is fragmented, and the individual fragments are attached to microscopic beads. Using PCR, each DNA fragment is then amplified so that copies of it eventually coat the bead to which it was initially attached. This process produces a library of billions of individual beads, each covered with identical copies of a particular DNA fragment.

Like eggs in a carton, the beads are placed into individual wells on an array that can hold a billion beads in a square inch. Beginning with a primer, DNA synthesis is then initiated on each bead. A hydrogen ion ( $H^+$ ) is released (along with pyrophosphate) each time a nucleotide is incorporated into a growing DNA chain (see Figure 5–3), and the ion torrent

method is based on this simple fact. Each of the four nucleotides is washed in, one at a time, over the array of beads; when a nucleotide is incorporated in the DNA of a given bead, the release of an  $H^+$  ion changes the pH, which is registered by a semiconductor chip placed beneath the array of wells. In this way, the DNA sequence on a given bead can be read from the pattern of pH changes observed as nucleotides are washed over them. Like a high-resolution sensor in a digital camera, the ion torrent semiconductor chip can register enormous amounts of information and can thus keep track of billions of parallel sequencing reactions. Using this technology it is currently possible, using a single chip, to determine the nucleotide sequences of several human genomes in just a few hours.

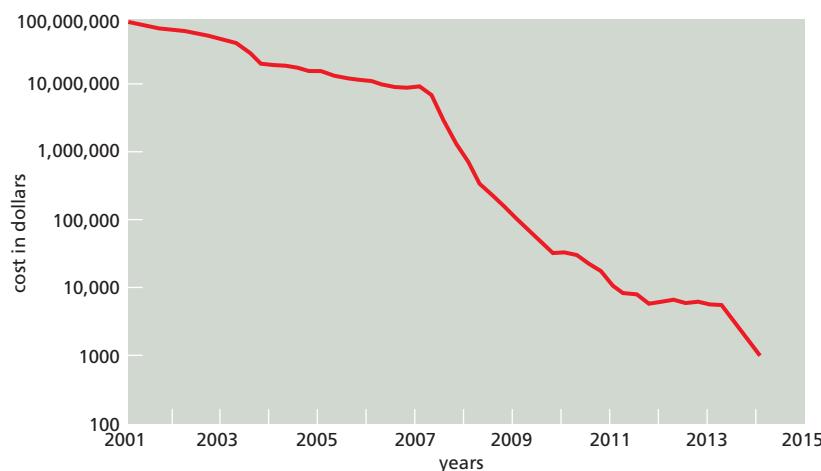


**DNA sequencing by the ion torrent method.** Beads, each coated with a DNA molecule that has been amplified many times, are placed in wells along with primers and DNA polymerase. As nucleotides are sequentially washed over the beads, those incorporated by the polymerase cause a pH change. In the example shown, an A is incorporated; thus, the template must have a T in this position. As the four nucleotides are sequentially washed over the beads, the sequence of the DNA on each bead can be “read” by the pattern of pH fluctuations. Billions of beads are monitored at once by a voltage-sensitive semiconductor chip placed below the array of beads.

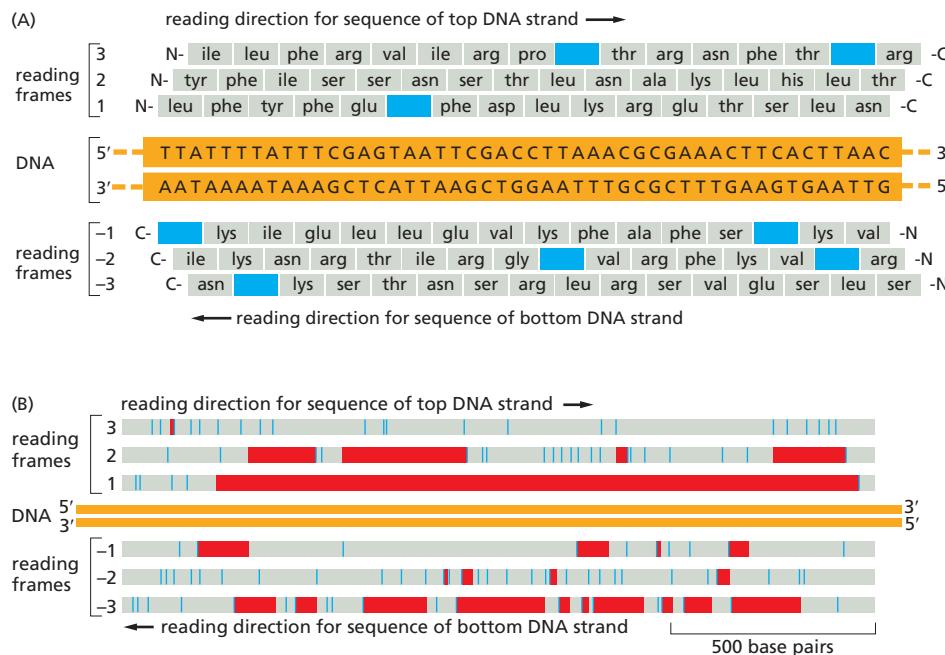
## THE FUTURE OF DNA SEQUENCING

Even newer, potentially faster, methods of sequencing DNA are being developed. Some of these “third-generation” technologies bypass the DNA amplification steps altogether and determine the sequence of single molecules of DNA. In one technique, a DNA molecule is pushed through a tiny channel, like a thread through the eye of a needle. As the DNA molecule moves through the pore, it generates electrical currents that depend on its sequence of nucleotides; the pattern of currents can then be used to deduce the nucleotide sequence. Other methods visualize single DNA molecules using electron or atomic force microscopy; the nucleotide sequence is read from the small differences in the “appearance” of the DNA as it is scanned. Finally, another method is based on immobilizing a single DNA polymerase molecule (with a template) and measuring the “dwell” time of each of the four nucleotides, which are labeled with different removable fluorescent dyes.

Nucleotides that reside longer on the polymerase (before their dye is removed) are those incorporated by the polymerase. Although the two methods we have described in detail (Illumina and ion torrent) are now used extensively, it is likely that faster and cheaper methods will continue to be developed.



Shown here are the costs of sequencing a human genome, which was \$100 million in 2001 and about a thousand dollars by the end of 2014. (Data from the National Human Genome Research Initiative.)



**Figure 8–40** Finding the regions in a DNA sequence that encode a protein. (A) Any region of the DNA sequence can, in principle, code for six different amino acid sequences, because any one of three different reading frames can be used to interpret the nucleotide sequence on each strand. Note that a nucleotide sequence is always read in the 5'-to-3' direction and encodes a polypeptide from the N-terminus to the C-terminus. For a random nucleotide sequence read in a particular frame, a stop signal for protein synthesis is encountered, on average, about once every 20 amino acids. In this sample sequence of 48 base pairs, each such signal (*stop codon*) is colored *blue*, and only reading frame 2 lacks a stop signal. (B) Search of a 1700-base-pair DNA sequence for a possible protein-encoding sequence. The information is displayed as in (A), with each stop signal for protein synthesis denoted by a *blue line*. In addition, all of the regions between possible start and stop signals for protein synthesis (see pp. 347–349) are displayed as *red bars*. Only reading frame 1 actually encodes a protein, which is 475 amino acid residues long.

contain a stop codon about every 20 amino acids; protein-coding regions will, in contrast, usually contain much longer stretches without stop codons (Figure 8–40). Known as **open reading frames (ORFs)**, these usually signify bona fide protein-coding genes. This assignment is often “double-checked” by comparing the ORF amino acid sequence to the many databases of documented proteins from other species. If a match is found, even as imperfect one, it is very likely that the ORF will code for a functional protein (see Figure 8–23).

This strategy works very well for compact genomes, where intron sequences are rare and ORFs often extend for many hundreds of amino acids. However, in many animals and plants, the average exon size is 150–200 nucleotide pairs (see Figure 6–31) and additional information is usually required to unambiguously locate all the exons of a gene. Although it is possible to search genomes for splicing signals and other features that help to identify exons (codon bias, for example), one of the most powerful methods is simply to sequence the total RNA produced from the genome in living cells. As can be seen in Figure 7–3, this RNA-seq information, when mapped onto the genome sequence, can be used to accurately locate all the introns and exons of even complex genes. By sequencing total RNA from different cell types, it is also possible to identify cases of alternative splicing (see Figure 6–26).

RNA-seq also identifies noncoding RNAs produced by a genome. Although the function of some of them can be readily recognized (tRNAs or snoRNAs, for example), many have unknown functions and still others probably have no function at all (discussed in Chapter 7, pp. 429–436). The existence of the many noncoding RNAs and our relative ignorance of their function is the main reason that we know only the approximate number of genes in the human genome.

But even for protein-coding genes that have been unambiguously identified, we still have much to learn. Thousands of genomes have been sequenced, and we know from *comparative genomics* that many organisms share the same basic set of proteins. However, the functions of a very large number of identified proteins remain unknown. Depending on the organism, approximately one-third of the proteins encoded by a sequenced genome do not clearly resemble any protein that has been studied biochemically. This observation underscores a limitation of the emerging field of genomics: although comparative analysis of genomes reveals a great deal of information about the relationships between genes and organisms, it often does not provide immediate information about how these genes function, or what roles they have in the physiology of an organism. Comparison of the full

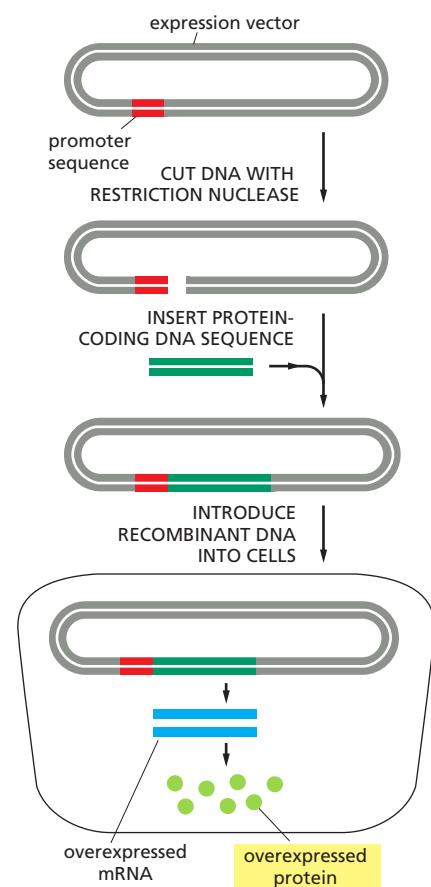
gene complement of several thermophilic bacteria, for example, does not reveal why these bacteria thrive at temperatures exceeding 70°C. And examination of the genome of the incredibly radioresistant bacterium *Deinococcus radiodurans* does not explain how this organism can survive a blast of radiation that can shatter glass. Further biochemical and genetic studies, like those described in the other sections of this chapter, are required to determine how genes, and the proteins they produce, function in the context of living organisms.

### DNA Cloning Allows Any Protein to be Produced in Large Amounts

In the last section, we saw how protein-coding genes can be identified in genome sequences. Using the genetic code (and provided the intron and exon boundaries are known), the amino acid sequence of any protein coded in a genome can be deduced. As was discussed earlier, this sequence can often provide an important clue to the protein's function if found to be similar to the amino acid sequence of a protein that has already been studied (see Figure 8–23). Although this strategy is often successful, it typically provides only the likely biochemical function of the protein; for example, whether the protein resembles a kinase or a protease. It usually remains for the experimenter to verify (or refute) this assignment and, most importantly, to discover the protein's biological function in the whole organism; that is, to what attributes of the organism does the kinase or the protease contribute and in what molecular pathways does it function? Nowadays, most new proteins are “discovered” through genome sequencing, and it often remains a great challenge to ascertain their functions.

An important approach in determining gene function is to alter the gene (or in some cases, its expression pattern), to put the altered copy back into the germ line of the organism, and to deduce the function of the normal gene by the changes caused by its alteration. Various techniques to implement this strategy are discussed in the next section of this chapter. But it is equally important to study the biochemical and structural properties of a gene product, as outlined in the first part of this chapter. One of the most important contributions of DNA cloning to cell and molecular biology is the ability to produce any protein, even the rare ones, in nearly unlimited amounts—as long as the gene coding for it is known. Such high-level production is usually carried out in living cells using expression vectors (Figure 8–41). These are generally plasmids that have been designed to produce a large amount of stable mRNA that can be efficiently translated into protein when the plasmid is introduced into bacterial, yeast, insect, or mammalian cells. To prevent the high level of the foreign protein from interfering with the cell's growth, the expression vector is often designed to delay the synthesis of the foreign mRNA and protein until shortly before the cells are harvested and lysed (Figure 8–42).

Because the desired protein made from an expression vector is produced inside a cell, it must be purified away from the host-cell proteins by chromatography following cell lysis; but because it is such a plentiful species in the cell (often 1–10% of the total cell protein), the purification is usually easy to accomplish in only a few steps. As we saw in the first part of this chapter, many expression



**Figure 8–41** Production of large amounts of a protein from a protein-coding DNA sequence cloned into an expression vector and introduced into cells. A plasmid vector has been engineered to contain a highly active promoter, which causes unusually large amounts of mRNA to be produced from an adjacent protein-coding gene inserted into the plasmid vector. Depending on the characteristics of the cloning vector, the plasmid is introduced into bacterial, yeast, insect, or mammalian cells, where the inserted gene is efficiently transcribed and translated into protein. If the gene to be overexpressed has no introns (typical for genes from bacteria, archaea, and simple eukaryotes), it can simply be cloned from genomic DNA by PCR. For cloned animal and plant genes, it is often more convenient to obtain the gene as cDNA, either from a cDNA library (see Figure 8–32) or cloned directly by PCR from RNA isolated from the organism (see Figure 8–37). Alternatively, the DNA coding for the protein can be made by chemical synthesis (see p. 472).

vectors have been designed to add a molecular tag—a cluster of histidine residues or a small marker protein—to the expressed protein to facilitate easy purification by affinity chromatography (see Figure 8–11). A variety of expression vectors is available, each engineered to function in the type of cell in which the protein is to be made.

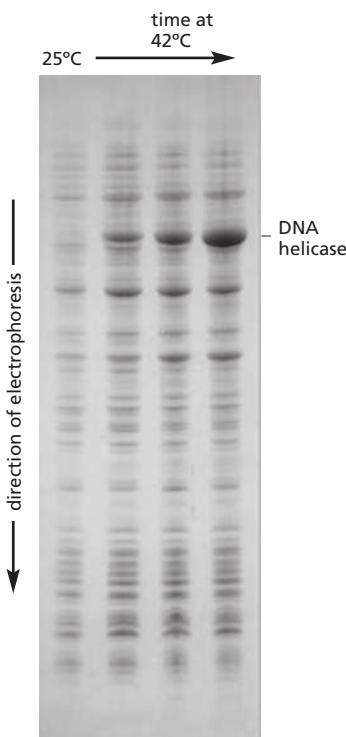
This technology is also used to make large amounts of many medically useful proteins, including hormones (such as insulin and growth factors) used as human pharmaceuticals, and viral coat proteins for use in vaccines. Expression vectors also allow scientists to produce many proteins of biological interest in large enough amounts for detailed structural studies. Nearly all three-dimensional protein structures depicted in this book are of proteins produced in this way. Recombinant DNA techniques thus allow scientists to move with ease from protein to gene, and vice versa, so that the functions of both can be explored on multiple fronts (Figure 8–43).

## Summary

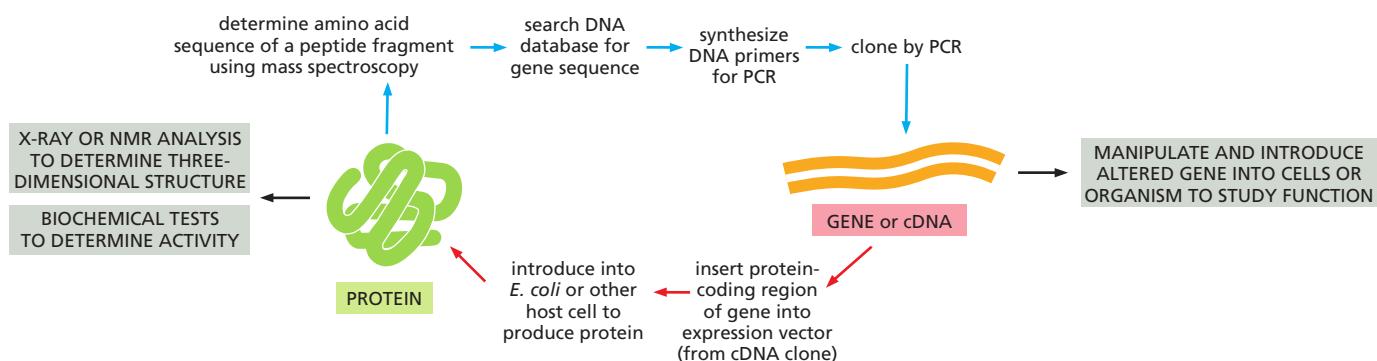
*DNA cloning allows a copy of any specific part of a DNA or RNA sequence to be selected from the millions of other sequences in a cell and produced in unlimited amounts in pure form. DNA sequences can be amplified after breaking up chromosomal DNA and inserting the resulting DNA fragments into the chromosome of a self-replicating genetic element such as a plasmid. The resulting “genomic DNA library” is housed in millions of bacterial cells, each carrying a different cloned DNA fragment. Individual cells from this library that are allowed to proliferate produce large amounts of a single cloned DNA fragment. Bypassing cloning vectors and bacterial cells altogether, the polymerase chain reaction (PCR) allows DNA cloning to be performed directly with a DNA polymerase and DNA primers—provided that the DNA sequence of interest is already known.*

The procedures used to obtain DNA clones that correspond in sequence to mRNA molecules are the same, except that a DNA copy of the mRNA sequence, called cDNA, is first made. Unlike genomic DNA clones, cDNA clones lack intron sequences, making them the clones of choice for analyzing the protein product of a gene.

Nucleic acid hybridization reactions provide a sensitive means of detecting any nucleotide sequence of interest. The enormous specificity of this hybridization reaction allows any single-stranded sequence of nucleotides to be labeled with a radioisotope or chemical and used as a probe to find a complementary partner strand, even in a cell or cell extract that contains millions of different DNA and RNA sequences. DNA hybridization also makes it possible to use PCR to amplify any section of any genome once its sequence is known.



**Figure 8–42** Production of large amounts of a protein by using a plasmid expression vector. In this example, an expression vector that overproduces a DNA helicase has been introduced into bacteria. In this expression vector, transcription from this coding sequence is under the control of a viral promoter that becomes active only at a temperature of 37°C or higher. The total cell protein, either from bacteria grown at 25°C (no helicase protein made) or after a shift of the same bacteria to 42°C for up to 2 hours (helicase protein has become the most abundant protein species in the lysate), has been analyzed by SDS polyacrylamide-gel electrophoresis. (Courtesy of Jack Barry.)



**Figure 8–43** Recombinant DNA techniques make it possible to move experimentally from gene to protein and from protein to gene. If a gene has been identified (right), its protein-coding sequence can be inserted into an expression vector to produce large quantities of the protein (see Figure 8–41), which can then be studied biochemically or structurally. If a protein has been purified based on its biochemical properties, mass spectrometry (see Figure 8–18) can be used to obtain a partial amino acid sequence, which is used to search a genome sequence for the corresponding nucleotide sequence. The complete gene can then be cloned by PCR from a sequenced genome (see Figure 8–37). The gene can also be manipulated and introduced into cells or organisms to study its function, a topic covered in the next section of this chapter.

The nucleotide sequence of any genome can be determined rapidly and simply by using highly automated techniques based on several different strategies. Comparison of the genome sequences of different organisms allows us to trace the evolutionary relationships among genes and organisms, and it has proved valuable for discovering new genes and predicting their functions.

Taken together, these techniques for analyzing and manipulating DNA have made it possible to sequence, identify, and isolate genes from any organism of interest. Related technologies allow scientists to produce the protein products of these genes in the large quantities needed for detailed analyses of their structure and function, as well as for medical purposes.

## STUDYING GENE EXPRESSION AND FUNCTION

Ultimately, one wishes to determine how genes—and the proteins they encode—function in the intact organism. Although it may seem counterintuitive, one of the most direct ways to find out what a gene does is to see what happens to the organism when that gene is missing. Studying mutant organisms that have acquired changes or deletions in their nucleotide sequences is a time-honored practice in biology and forms the basis of the important field of **genetics**. Because mutations can disrupt cell processes, mutants often hold the key to understanding gene function. In the classical genetic approach, one begins by isolating mutants that have an interesting or unusual appearance: fruit flies with white eyes or curly wings, for example. Working backward from the **phenotype**—the appearance or behavior of the individual—one then determines the organism's **genotype**, the form of the gene responsible for that characteristic (**Panel 8–2**).

Today, with numerous genome sequences available, the exploration of gene function often begins with a DNA sequence. Here, the challenge is to translate sequence into function. One approach, discussed earlier in the chapter, is to search databases for well-characterized proteins that have similar amino acid sequences to the protein encoded by a new gene. From there, the protein (or for noncoding genes, the RNA molecule) can be overexpressed and purified and the methods described in the first part of this chapter can be employed to study its three-dimensional structure and its biochemical properties. But to determine directly a gene's function in a cell or organism, the most effective approach involves studying mutants that either lack the gene or express an altered version of it. Determining which cell processes have been disrupted or compromised in such mutants will usually shed light on a gene's biological role.

In this section, we describe several approaches to determining a gene's function, starting either from an individual with an interesting phenotype or from a DNA sequence. We begin with the classical genetic approach, which starts with a *genetic screen* for isolating mutants of interest and then proceeds toward identification of the gene or genes responsible for the observed phenotype. We then describe the set of techniques that are collectively called *reverse genetics*, in which one begins with a gene or gene sequence and attempts to determine its function. This approach often involves some intelligent guesswork—searching for similar sequences in other organisms or determining when and where a gene is expressed—as well as generating mutant organisms and characterizing their phenotype.

### Classical Genetics Begins by Disrupting a Cell Process by Random Mutagenesis

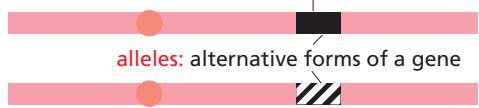
Before the advent of gene cloning technology, most genes were identified by the abnormalities produced when the gene was mutated. Indeed, the very concept of the gene was deduced from the heritability of such abnormalities. This classical genetic approach—identifying the genes responsible for mutant phenotypes—is most easily performed in organisms that reproduce rapidly and are amenable to genetic manipulation, such as bacteria, yeasts, nematode worms, and fruit flies. Although spontaneous mutants can sometimes be found by examining extremely

## GENES AND PHENOTYPES

**Gene:** a functional unit of inheritance, usually corresponding to the segment of DNA coding for a single protein.

**Genome:** all of an organism's DNA sequences.

**locus:** the site of the gene in the genome



A detailed illustration of a goldfish, showing its characteristic orange body, large eye, and flowing fins.

**Wild-type:** the normal, naturally occurring type

**Mutant:** differing from the wild-type because of a genetic change (a mutation)

**GENOTYPE:** the specific set of alleles forming the genome of an individual

**PHENOTYPE:** the visible character of the individual

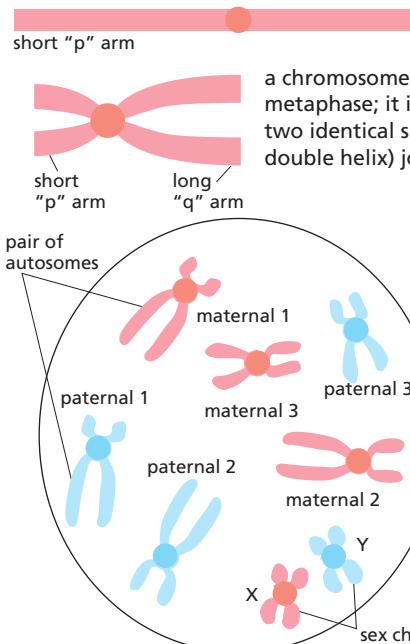


allele A is **dominant** (relative to a); allele a is **recessive** (relative to A)

In the example above, the phenotype of the heterozygote is the same as that of one of the homozygotes; in cases where it is different from both, the two alleles are said to be co-dominant.

## CHROMOSOMES

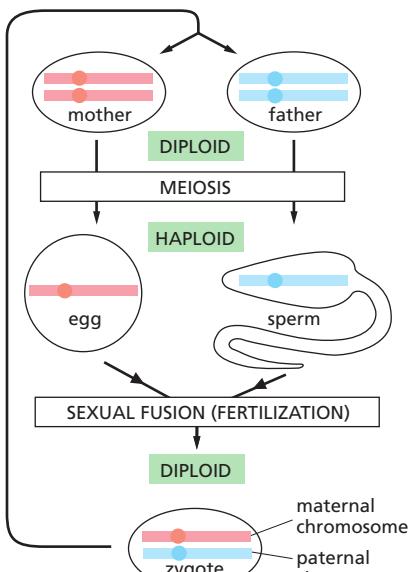
a chromosome at the beginning of the cell cycle, in G<sub>1</sub> phase; the single long bar represents one long double helix of DNA



a chromosome near the end of the cell cycle, in metaphase; it is duplicated and condensed, consisting of two identical sister chromatids (each containing one DNA double helix) joined at the centromere.

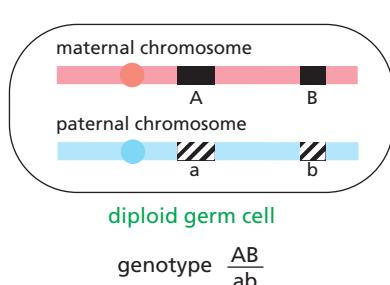
A normal diploid chromosome set, as seen in a metaphase spread, prepared by bursting open a cell at metaphase and staining the scattered chromosomes. In the example shown schematically here, there are three pairs of autosomes (chromosomes inherited symmetrically from both parents, regardless of sex) and two sex chromosomes—an X from the mother and a Y from the father. The numbers and types of sex chromosomes and their role in sex determination are variable from one class of organisms to another, as is the number of pairs of autosomes.

## THE HAPLOID–DIPLOID CYCLE OF SEXUAL REPRODUCTION



For simplicity, the cycle is shown for only one chromosome/chromosome pair.

## MEIOSIS AND GENETIC RECOMBINATION



**MEIOSIS AND RECOMBINATION**

genotype Ab

A      b

site of crossing-over

genotype aB

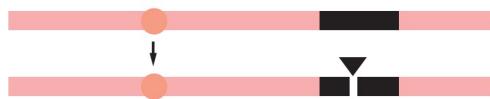
a      B

a      B

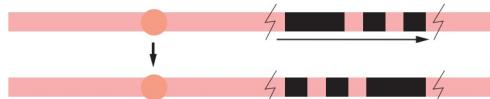
haploid gametes (eggs or sperm)

The greater the distance between two loci on a single chromosome, the greater is the chance that they will be separated by crossing over occurring at a site between them. If two genes are thus reassorted in  $x\%$  of gametes, they are said to be separated on a chromosome by a **genetic map distance** of  $x$  **map units** (or  $x$  **centimorgans**).

## TYPES OF MUTATIONS



**POINT MUTATION:** maps to a single site in the genome, corresponding to a single nucleotide pair or a very small part of a single gene



**INVERSION:** inverts a segment of a chromosome

**lethal mutation:** causes the developing organism to die prematurely.

**conditional mutation:** produces its phenotypic effect only under certain conditions, called the *restrictive* conditions. Under other conditions—the *permissive* conditions—the effect is not seen. For a *temperature-sensitive* mutation, the restrictive condition typically is high temperature, while the permissive condition is low temperature.

**loss-of-function mutation:** either reduces or abolishes the activity of the gene. These are the most common class of mutations. Loss-of-function mutations are usually *recessive*—the organism can usually function normally as long as it retains at least one normal copy of the affected gene.

**null mutation:** a loss-of-function mutation that completely abolishes the activity of the gene.



**DELETION:** deletes a segment of a chromosome



**TRANSLOCATION:** breaks off a segment from one chromosome and attaches it to another

**gain-of-function mutation:** increases the activity of the gene or makes it active in inappropriate circumstances; these mutations are usually *dominant*.

**dominant-negative mutation:** dominant-acting mutation that blocks gene activity, causing a loss-of-function phenotype even in the presence of a normal copy of the gene. This phenomenon occurs when the mutant gene product interferes with the function of the normal gene product.

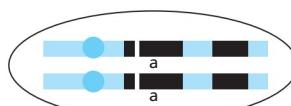
**suppressor mutation:** suppresses the phenotypic effect of another mutation, so that the double mutant seems normal. An *intragenic* suppressor mutation lies within the gene affected by the first mutation; an *extragenic* suppressor mutation lies in a second gene—often one whose product interacts directly with the product of the first.

## TWO GENES OR ONE?

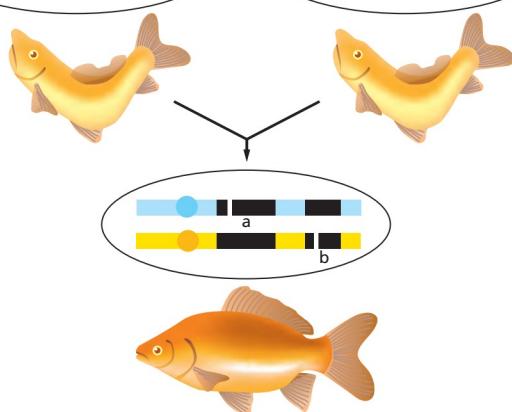
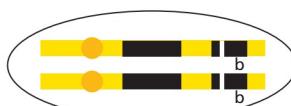
Given two mutations that produce the same phenotype, how can we tell whether they are mutations in the same gene? If the mutations are recessive (as they most often are), the answer can be found by a **complementation test**.

### COMPLEMENTATION: MUTATIONS IN TWO DIFFERENT GENES

homozygous mutant mother



homozygous mutant father

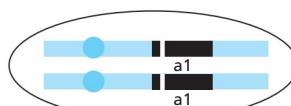


hybrid offspring shows normal phenotype:  
one normal copy of each gene is present

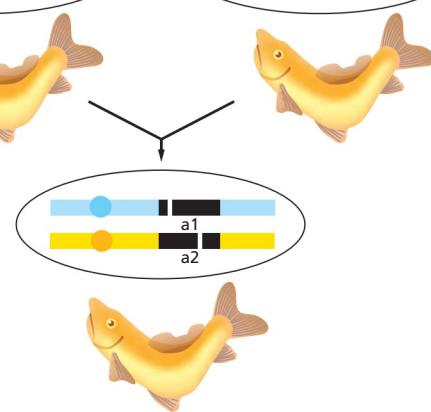
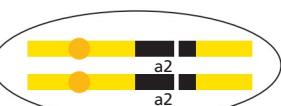
In the simplest type of complementation test, an individual who is homozygous for one mutation is mated with an individual who is homozygous for the other. The phenotype of the offspring gives the answer to the question.

### NONCOMPLEMENTATION: TWO INDEPENDENT MUTATIONS IN THE SAME GENE

homozygous mutant mother



homozygous mutant father



hybrid offspring shows mutant phenotype:  
no normal copies of the mutated gene are present

large populations—thousands or tens of thousands of individual organisms—isolating mutant individuals is much more efficient if one generates mutations with chemicals or radiation that damage DNA. By treating organisms with such mutagens, very large numbers of mutant individuals can be created quickly and then screened for a particular defect of interest, as we discuss shortly.

An alternative approach to chemical or radiation mutagenesis is called *insertional mutagenesis*. This method relies on the fact that exogenous DNA inserted randomly into the genome can produce mutations if the inserted fragment interrupts a gene or its regulatory sequences. The inserted DNA, whose sequence is known, then serves as a molecular tag that aids in the subsequent identification and cloning of the disrupted gene (Figure 8–44). In *Drosophila*, the use of the transposable P element to inactivate genes has revolutionized the study of gene function in the fly. Transposable elements (see Table 5–4, p. 288) have also been used to generate mutations in bacteria, yeast, mice, and the flowering plant *Arabidopsis*.

### Genetic Screens Identify Mutants with Specific Abnormalities

Once a collection of mutants in a model organism such as yeast or fly has been produced, one generally must examine thousands of individuals to find the altered phenotype of interest. Such a search is called a **genetic screen**, and the larger the genome, the less likely it is that any particular gene will be mutated. Therefore, the larger the genome of an organism, the bigger the screening task becomes. The phenotype being screened for can be simple or complex. Simple phenotypes are easiest to detect: one can screen many organisms rapidly, for example, for mutations that make it impossible for the organism to survive in the absence of a particular amino acid or nutrient.

More complex phenotypes, such as defects in learning or behavior, may require more elaborate screens (Figure 8–45). But even genetic screens that are used to dissect complex physiological systems can be simple in design, which permits the simultaneous examination of large numbers of mutants. As an example, one particularly elegant screen was designed to search for genes involved in visual processing in zebrafish. The basis of this screen, which monitors the fishes' response to motion, is a change in behavior. Wild-type fish tend to swim in the direction of a perceived motion, whereas mutants with defects in their visual processing systems swim in random directions—a behavior that is easily detected. One mutant discovered in this screen is called *lakritz*, which is missing 80% of the retinal ganglion cells that help to relay visual signals from the eye to the brain. As the cellular organization of the zebrafish retina is similar to that of all vertebrates, the study of such mutants should also provide insights into visual processing in humans.

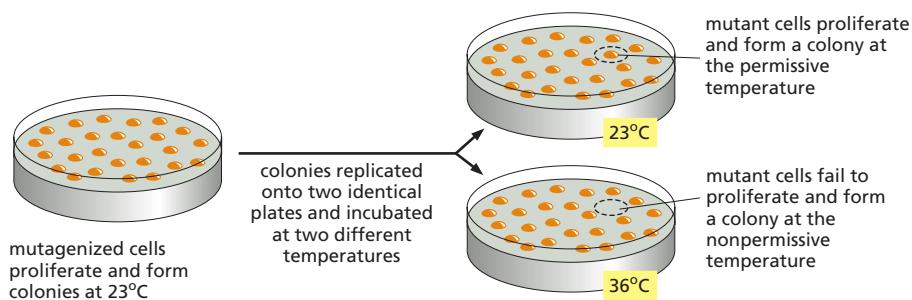
Because defects in genes that are required for fundamental cell processes—RNA synthesis and processing or cell-cycle control, for example—are usually lethal, the functions of these genes are often studied in individuals with



**Figure 8–44 Insertional mutant of the snapdragon, *Antirrhinum*.** A mutation in a single gene coding for a regulatory protein causes leafy shoots (left) to develop in place of flowers, which occur in the normal plant (right). The mutation causes cells to adopt a character that would be appropriate to a different part of the normal plant, so instead of a flower, the cells produce a leafy shoot. (Courtesy of Enrico Coen and Rosemary Carpenter.)



**Figure 8–45 A behavioral phenotype detected in a genetic screen.** (A) Wild-type *C. elegans* engage in social feeding. The worms migrate around until they encounter their neighbors and commence feeding on bacteria. (B) Mutant animals feed by themselves. (Courtesy of Cornelia Bargmann, *Cell* 94: cover, 1998. With permission from Elsevier.)



**conditional mutations.** The mutant individuals function normally as long as “permissive” conditions prevail, but demonstrate abnormal gene function when subjected to “nonpermissive” (restrictive) conditions. In organisms with *temperature-sensitive mutations*, for example, the abnormality can be switched on and off experimentally simply by changing the ambient temperature; thus, a cell containing a temperature-sensitive mutation in a gene essential for survival will die at a nonpermissive temperature but proliferate normally at the permissive temperature (Figure 8–46). The temperature-sensitive gene in such a mutant usually contains a point mutation that causes a subtle change in its protein product; for example, the mutant protein may function normally at low temperatures but unfold at higher temperatures.

Temperature-sensitive mutations were crucial to find the bacterial genes that encode the proteins required for DNA replication. The mutants were identified by screening populations of mutagen-treated bacteria for cells that stop making DNA when they are warmed from 30°C to 42°C. These mutants were later used to identify and characterize the corresponding DNA replication proteins (discussed in Chapter 5). Similarly, screens for temperature-sensitive mutations led to the identification of many proteins involved in regulating the cell cycle, as well as many proteins involved in moving proteins through the secretory pathway in yeast. Related screening approaches demonstrated the function of enzymes involved in the principal metabolic pathways of bacteria and yeast (discussed in Chapter 2) and identified many of the gene products responsible for the orderly development of the *Drosophila* embryo (discussed in Chapter 21).

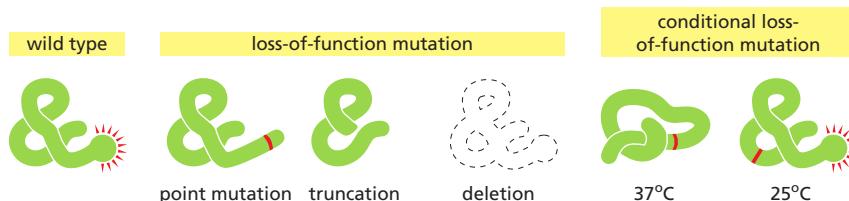
### Mutations Can Cause Loss or Gain of Protein Function

Gene mutations are generally classed as “loss of function” or “gain of function.” A loss-of-function mutation results in a gene product that either does not work or works too little; thus, it can reveal the normal function of the gene. A gain-of-function mutation results in a gene product that works too much, works at the wrong time or place, or works in a new way (Figure 8–47).

An important early step in the genetic analysis of any mutant cell or organism is to determine whether the mutation causes a loss or a gain of function. A standard test is to determine whether the mutation is *dominant* or *recessive*. A dominant mutation is one that still causes the mutant phenotype in the presence of a single copy of the wild-type gene. A recessive mutation is one that is no longer able to cause the mutant phenotype in the presence of a single wild-type copy of the gene. Although cases have been described in which a loss-of-function mutation is dominant or a gain-of-function mutation is recessive, in the vast majority of cases, recessive mutations are loss of function and dominant mutations are gain

**Figure 8–46** Screening for temperature-sensitive bacterial or yeast mutants.

Mutagenized cells are plated out at the permissive temperature. They divide and form colonies, which are transferred to two identical Petri dishes by replica plating. One of these plates is incubated at the permissive temperature, the other at the nonpermissive temperature. Cells containing a temperature-sensitive mutation in a gene essential for proliferation can divide at the normal, permissive temperature but fail to divide at the elevated, nonpermissive temperature. Temperature-sensitive mutations of this type were especially useful for identifying genes needed for DNA replication, an essential process.



**Figure 8–47** Gene mutations that affect their protein product in different ways.

In this example, the wild-type protein has a specific cell function denoted by the red rays. Mutations that eliminate this function or inactivate it at higher temperatures are shown. The conditional mutant protein carries an amino acid substitution (red) that prevents its proper folding at 37°C, but allows the protein to fold and function normally at 25°C. Such temperature-sensitive conditional mutations are especially useful for studying essential genes; the organism can be grown under the permissive condition and then be moved to the nonpermissive condition to study the consequences of losing the gene product.

of function. It is easy to determine if a mutation is dominant or recessive. One simply mates a mutant with a wild type to obtain diploid cells or organisms. The progeny from the mating will be heterozygous for the mutation. If the mutant phenotype is no longer observed, one can conclude that the mutation is recessive and is very likely to be a loss-of-function mutation (see Panel 8–2).

### Complementation Tests Reveal Whether Two Mutations Are in the Same Gene or Different Genes

A large-scale genetic screen can turn up many different mutations that show the same phenotype. These defects might lie in different genes that function in the same process, or they might represent different mutations in the same gene. Alternative forms of the same gene are known as **alleles**. The most common difference between alleles is a substitution of a single nucleotide pair, but different alleles can also bear deletions, substitutions, and duplications. How can we tell, then, whether two mutations that produce the same phenotype occur in the same gene or in different genes? If the mutations are recessive—if, for example, they represent a loss of function of a particular gene—a **complementation test** can be used to ascertain whether the mutations fall in the same gene or in different genes. To test complementation in a diploid organism, an individual that is homozygous for one mutation—that is, it possesses two identical alleles of the mutant gene in question—is mated with an individual that is homozygous for the other mutation. If the two mutations are in the same gene, the offspring show the mutant phenotype, because they still will have no normal copies of the gene in question (Figure 8–48). If, in contrast, the mutations fall in different genes, the resulting offspring show a normal phenotype, because they retain one normal copy (and one mutant copy) of each gene; the mutations thereby complement one another and restore a normal phenotype. Complementation testing of mutants identified during genetic screens has revealed, for example, that 5 different genes are required for yeast to digest the sugar galactose, 20 genes are needed for *E. coli* to build a functional flagellum, 48 genes are involved in assembling bacteriophage T4 viral particles, and hundreds of genes are involved in the development of an adult nematode worm from a fertilized egg.

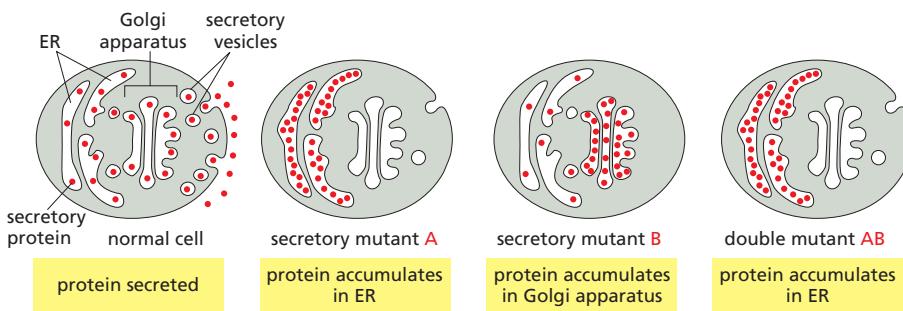
### Gene Products Can Be Ordered in Pathways by Epistasis Analysis

Once a set of genes involved in a particular biological process has been identified, the next step is often to determine in which order the genes function. Gene order is perhaps easiest to explain for metabolic pathways, where, for example, enzyme A is necessary to produce the substrate for enzyme B. In this case, we would say that the gene encoding enzyme A acts before (upstream of) the gene encoding enzyme B in the pathway. Similarly, where one protein regulates the activity of another protein, we would say that the former gene acts before the latter. Gene order can, in many cases, be determined purely by genetic analysis without any knowledge of the mechanism of action of the gene products involved.

Suppose we have a biosynthetic process consisting of a sequence of steps, such that performance of step B is conditional on completion of the preceding step A; and suppose gene A is required for step A, and gene B is required for step B. Then a null mutation (a mutation that abolishes function) in gene A will arrest the process at step A, regardless of whether gene B is functional or not, whereas a null mutation in gene B will cause arrest at step B only if gene A is still active. In such a case, gene A is said to be **epistatic** to gene B. By comparing the phenotypes of the different combinations of mutations, we can therefore discover the order in which the genes act. This type of analysis is called **epistasis analysis**. As an example, the pathway of protein secretion in yeast has been analyzed in this way. Different mutations in this pathway cause proteins to accumulate aberrantly in the endoplasmic reticulum (ER) or in the Golgi apparatus. When a yeast cell is engineered to carry both a mutation that blocks protein processing in the ER *and* a mutation that blocks processing in the Golgi apparatus, proteins accumulate in



**Figure 8–48** A complementation test can reveal that mutations in two different genes are responsible for the same abnormal phenotype. When an albino (white) bird from one strain is bred with an albino from a different strain, the resulting offspring (bottom) have normal coloration. This restoration of the wild-type plumage indicates that the two white breeds lack color because of recessive mutations in different genes. (From W. Bateson, Mendel's Principles of Heredity, 1st ed. Cambridge, UK: Cambridge University Press, 1913.)



**Figure 8–49** Using genetics to determine the order of function of genes. In normal cells, secretory proteins are loaded into vesicles, which fuse with the plasma membrane to secrete their contents into the extracellular medium. Two mutants, A and B, fail to secrete proteins. In mutant A, secretory proteins accumulate in the ER. In mutant B, secretory proteins accumulate in the Golgi. In the double mutant AB, proteins accumulate in the ER; this indicates that the gene defective in mutant A acts before the gene defective in mutant B in the secretory pathway.

the ER. This indicates that proteins must pass through the ER before being sent to the Golgi before secretion (Figure 8–49). Strictly speaking, an epistasis analysis can only provide information about gene order in a pathway when both mutations are null alleles. When the mutations retain partial function, their epistasis interactions can be difficult to interpret.

Sometimes, a double mutant will show a new or more severe phenotype than either single mutant alone. This type of genetic interaction is called a *synthetic phenotype*, and if the phenotype is death of the organism, it is called *synthetic lethality*. In most cases, a synthetic phenotype indicates that the two genes act in two different parallel pathways, either of which is capable of mediating the same cell process. Thus, when both pathways are disrupted in the double mutant, the process fails altogether, and the synthetic phenotype is observed.

### Mutations Responsible for a Phenotype Can Be Identified Through DNA Analysis

Once a collection of mutant organisms with interesting phenotypes has been obtained, the next task is to identify the gene or genes responsible for the altered phenotype. If the phenotype has been produced by insertional mutagenesis, locating the disrupted gene is fairly simple. DNA fragments containing the insertion (a transposon or a retrovirus, for example) are amplified by PCR, and the nucleotide sequence of the flanking DNA is determined. The gene affected by the insertion can then be identified by a computer-aided search of the complete genome sequence of the organism.

If a DNA-damaging chemical was used to generate the mutations, identifying the inactivated gene is often more laborious, but there are several powerful strategies available. If the genome size of the organism is small (for example, for bacteria or simple eukaryotes), it is possible to simply determine the genome sequence of the mutant organism and identify the affected gene by comparison with the wild-type sequence. Because of the continuous accumulation of neutral mutations, there will probably be differences between the two genome sequences in addition to the mutation responsible for the phenotype. One way of proving that a mutation is causative is to introduce the putative mutation back into a normal organism and determine whether or not it causes the mutant phenotype. We will discuss how this is accomplished later in the chapter.

### Rapid and Cheap DNA Sequencing Has Revolutionized Human Genetic Studies

Genetic screens in model experimental organisms have been spectacularly successful in identifying genes and relating them to various phenotypes, including many that are conserved between these organisms and humans. But how can we study humans directly? They do not reproduce rapidly, cannot be treated with mutagens, and, if they have a defect in an essential process such as DNA replication, would die long before birth.

Despite their limitations compared to model organisms, humans are becoming increasingly attractive subjects for genetic studies. Because the human

population is so large, spontaneous, nonlethal mutations have arisen in all human genes—many times over. A substantial proportion of these remain in the genomes of present-day humans. The most deleterious of these mutations are discovered when the mutant individuals call attention to themselves by seeking medical help.

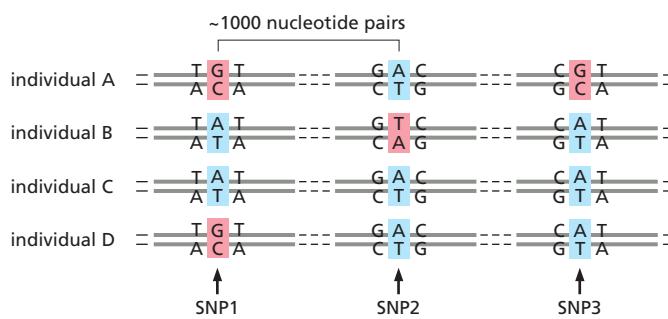
With the recent advances that have enabled the sequencing of entire human genomes cheaply and quickly, we can now identify such mutations and study their evolution and inheritance in ways that were impossible even a few years ago. By comparing the sequences of thousands of human genomes from all around the world, we can begin to identify directly the DNA differences that distinguish one individual from another. These differences hold clues to our evolutionary origins and can be used to explore the roots of disease.

### Linked Blocks of Polymorphisms Have Been Passed Down from Our Ancestors

When we compare the sequences of multiple human genomes, we find that any two individuals will differ in roughly 1 nucleotide pair in 1000. Most of these variations are common and relatively harmless. When two sequence variants coexist in the population and both are common, the variants are called **polymorphisms**. The majority of polymorphisms are due to the substitution of a single nucleotide, called **single-nucleotide polymorphisms** or SNPs (**Figure 8–50**). The rest are due largely to insertions or deletions—called *indels* when the change is small, or *copy number variations* (CNVs) when it is large. Although these common variants can be found throughout the genome, they are not scattered randomly—or even independently. Instead, they tend to travel in groups called **haplotype blocks**—combinations of polymorphisms that are inherited as a unit.

To understand why such haplotype blocks exist, we need to consider our evolutionary history. It is thought that modern humans expanded from a relatively small population—perhaps around 10,000 individuals—that existed in Africa about 60,000 years ago. Among that small group of our ancestors, some individuals will have carried one set of genetic variants, others a different set. The chromosomes of a present-day human represent a shuffled combination of chromosome segments from different members of this small ancestral group of people. Because only about two thousand generations separate us from them, large segments of these ancestral chromosomes have passed from parent to child, unbroken by the crossover events that occur during meiosis. As described in Chapter 5, only a few crossovers occur between each set of homologous chromosomes during each meiosis (see Figure 5–53).

As a result, certain sets of DNA sequences—and their associated polymorphisms—have been inherited in linked groups, with little genetic rearrangement across the generations. These are the haplotype blocks. Like genes that exist in different allelic forms, haplotype blocks also come in a limited number of variants that are common in the human population, each representing a combination of DNA polymorphisms passed down from a particular ancestor long ago.



**Figure 8–50** Single-nucleotide polymorphisms (SNPs) are sites in the genome where two or more alternative choices of a nucleotide are common in the population. Most such variations in the human genome occur at locations where they do not significantly affect a gene's function.

## Polymorphisms Can Aid the Search for Mutations Associated with Disease

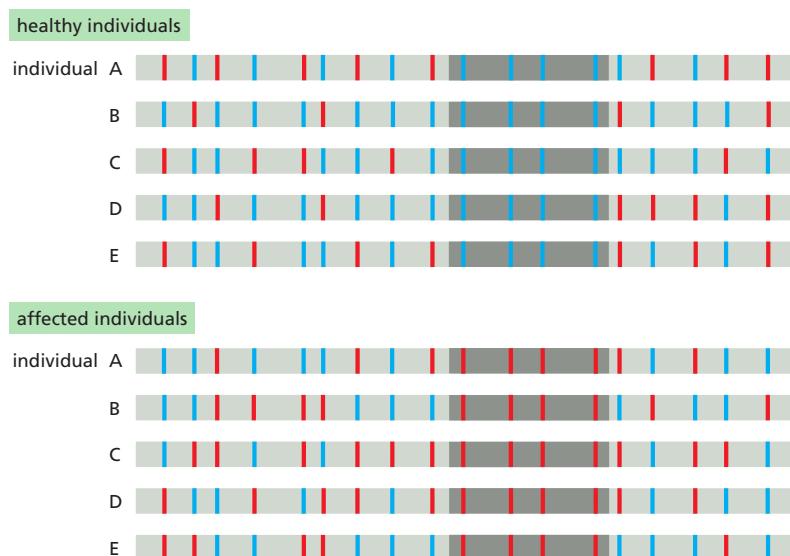
Mutations that give rise, in a reproducible way, to rare but clearly defined abnormalities, such as albinism, hemophilia, or congenital deafness, can often be identified by studies of affected families. Such single-gene, or monogenic, disorders are often referred to as *Mendelian* because their pattern of inheritance is easy to track. Moreover, individuals who inherit the causative mutation will exhibit the abnormality irrespective of environmental factors such as diet or exercise. But for many common diseases, the genetic roots are more complex. Instead of a single allele of a single gene, such disorders stem from a combination of contributions from multiple genes. And often, environmental factors have strong influences on the severity of the disorder. For these *multigenic* conditions, such as diabetes or arthritis, population studies are often helpful in tracking down the genes that increase the risk of getting the disease.

In population studies, investigators collect DNA samples from a large number of people who have the disease and compare them to samples from a group of people who do not have the disease. They look for variants—SNPs, for example—that are more common among the people who have the disease. Because DNA sequences that are close together on a chromosome tend to be inherited together, the presence of such SNPs could indicate that an allele that increases the risk of the disease might lie nearby (Figure 8–51). Although, in principle, the disease could be caused by the SNP itself, the culprit is much more likely to be a change that is merely linked to the SNP as part of a haplotype block.

Such *genome-wide association studies* have been used to search for genes that predispose individuals to common diseases, including diabetes, coronary artery disease, rheumatoid arthritis, and even depression. For many of these conditions, the DNA polymorphisms identified increase the risk of disease only slightly. Moreover, environmental factors (diet, exercise, for example) play an important role in the onset and severity of the disease. Nonetheless, the identification of genes affected by these polymorphisms is leading to a mechanistic understanding of some of our most common disorders.

## Genomics Is Accelerating the Discovery of Rare Mutations That Predispose Us to Serious Disease

The genetic variants that have thus far allowed us to identify some of the genes that increase our risk of disease are common ones. They arose long ago in our evolutionary past and are now present, in one form or another, in a substantial portion (1% or more) of the population. Such polymorphisms are thought to account



**Figure 8–51** Genes that affect the risk of developing a common disease can often be tracked down through linkage to SNPs. Here, the patterns of SNPs are compared between two sets of individuals—a set of healthy controls and a set affected by a particular common disease. A segment of a typical chromosome is shown. For most polymorphic sites in this segment, it is a random matter whether an individual has one SNP variant (red vertical bars) or another (blue vertical bars); this same randomness is seen both for the control group and for the affected individuals. However, in the part of the chromosome that is shaded in *darker gray*, a bias is seen: most normal individuals have the blue SNP variants, whereas most affected individuals have the red SNP variants. This suggests that this region contains, or is close to, a gene that is genetically linked to these red SNP variants and that predisposes individuals to the disease. Using carefully selected controls and thousands of affected individuals, this approach can help track down disease-related genes, even when they confer only a slight increase in the risk of developing the disease.

for about 90% of the differences between one person's genome and another's. But when we try to tie these common variants to differences in disease susceptibility or other heritable traits, such as height, we find that they do not have as much predictive power as we had anticipated: thus, for example, most confer relatively small increases—less than twofold—in the risk of developing a common disease.

In contrast to polymorphisms, rare DNA variants—those much less frequent in humans than SNPs—can have large effects on the risk of developing some common diseases. For example, a number of different loss-of-function mutations, each individually rare, have been found to increase greatly the predisposition to autism and schizophrenia. Many of these are *de novo* mutations, which arose spontaneously in the germ-line cells of one or the other parent. The fact that these mutations arise spontaneously with some frequency could help explain why these common disorders—each observed in about 1% of the population—remain with us, even though the affected individuals leave few or no descendants. These rare mutations may arise in any one of hundreds of different genes, which could explain much of the clinical variability of autism and schizophrenia. Because they are kept rare by natural selection, most such variants with a large effect on risk would be missed by genome-wide association studies.

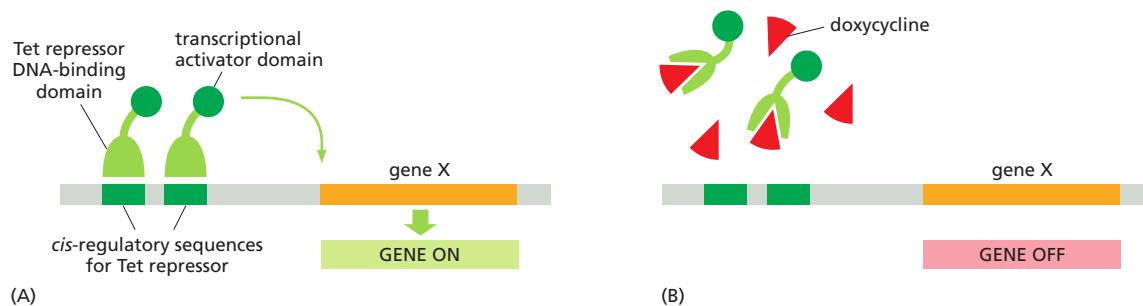
Now that DNA sequencing has become fast and inexpensive, the most efficient and cost-effective way to identify these rare, large-effect mutations is by sequencing the genomes of affected individuals, along with those of their parents and siblings as controls.

### Reverse Genetics Begins with a Known Gene and Determines Which Cell Processes Require Its Function

As we have seen, classical genetics starts with a mutant phenotype (or, in the case of humans, a range of characteristics) and identifies the mutations, and consequently the genes, responsible for it. Recombinant DNA technology has made possible a different type of genetic approach, one that is used widely in a variety of genetically tractable species. Instead of beginning with a mutant organism and using it to identify a gene and its protein, an investigator can start with a particular gene and proceed to make mutations in it, creating mutant cells or organisms so as to analyze the gene's function. Because this approach reverses the traditional direction of genetic discovery—proceeding from genes to mutations, rather than vice versa—it is commonly referred to as **reverse genetics**. And because the genome of the organism is deliberately altered in a particular way, this approach is also called *genome engineering* or *genome editing*. We shall see in this chapter that this approach can be scaled upward so that whole collections of organisms can be created, each of which has a different gene altered.

There are several ways a gene of interest can be altered. In the simplest, the gene can simply be deleted from the genome, although in a diploid organism, this requires that both copies—one on each chromosome homolog—be deleted. Although somewhat counterintuitive, one of the best ways to discover the function of a gene is by observing the effects of not having it. Such “gene knockouts” are especially useful if the gene is not essential. Through reverse genetics, the gene in question (even if it is essential) can also be replaced by one that is expressed in the wrong tissue or at the wrong time in development; this type of manipulation often provides important clues to the gene's normal function. For example, a gene of interest can be modified to be expressed at will by the experimenter (**Figure 8–52**). Finally, genes can also be engineered so that they are expressed normally in most cell types and tissues but deleted in certain cell types or tissues selected by the experimenter (see Figure 5–66). This approach is especially useful when a gene has different roles in different tissues.

It is also possible to make subtler changes to a gene. It is sometimes useful to make slight changes in a protein's structure so that one can begin to dissect which portions of a protein are important for its function. The activity of an enzyme, for example, can be studied by changing a single amino acid in its active site. It is also possible, through genome engineering, to create new types of proteins in an



**Figure 8–52** Engineered genes can be turned on and off with small molecules. Here, the DNA-binding portion of a bacterial protein (the tetracycline, Tet, repressor) has been fused to a portion of a mammalian transcriptional activator and expressed in cultured mammalian cells. The engineered gene *X*, present in place of the normal gene, has its usual gene control region replaced by *cis*-regulatory sequences recognized by the tetracycline repressor. In the absence of doxycycline (a particularly stable version of tetracycline), the engineered gene is expressed; in the presence of doxycycline, the gene is turned off because the tetracycline repressor causes the fusion protein to dissociate from the DNA. This strategy can also be used in mice by incorporating the engineered genes into the germ line. In many tissues, the gene can be turned on and off simply by adding or removing doxycycline from the animal's water. If the tetracycline repressor construct is placed under the control of a tissue-specific gene control region, the engineered gene will be turned on and off only in that tissue.

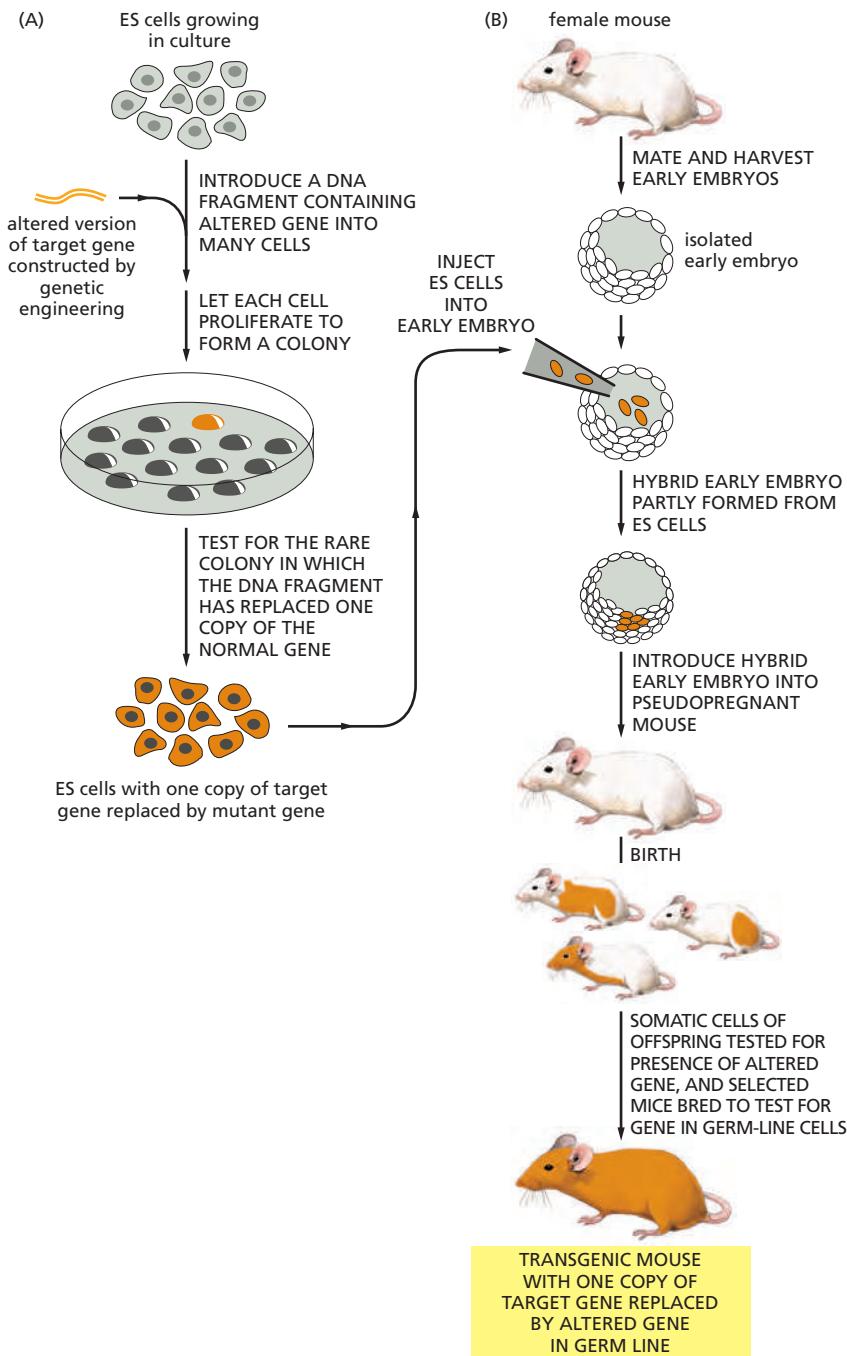
animal. For example, a gene can be fused to the gene for a fluorescent protein. When this altered gene is introduced into the genome, the protein can be tracked in the living organism by monitoring its fluorescence.

Altered genes can be created in several ways. Perhaps the simplest is to chemically synthesize the DNA that makes up the gene. In this way, the investigator can specify any type of variant of the normal gene. It is also possible to construct altered genes using recombinant DNA technology, as described earlier in this chapter. Once obtained, altered genes can be introduced into cells in a variety of ways. DNA can be microinjected into mammalian cells with a glass micropipette or introduced by a virus that has been engineered to carry foreign genes. In plant cells, genes are frequently introduced by a technique called particle bombardment: DNA samples are painted onto tiny gold beads and then literally shot through the cell wall with a specially modified gun. *Electroporation* is the method of choice for introducing DNA into bacteria and some other cells. In this technique, a brief electric shock renders the cell membrane temporarily permeable, allowing foreign DNA to enter the cytoplasm.

To be most useful to experimenters, the altered gene, once it is introduced into a cell, must recombine with the cell's genome so that the normal gene is replaced. In simple organisms such as bacteria and yeasts, this process occurs with high frequency using the cell's own homologous recombination machinery, as described in Chapter 5. In more complex organisms that have elaborate developmental programs, the procedure is more complicated because the altered gene must be introduced into the germ line, as we next describe.

### Animals and Plants Can Be Genetically Altered

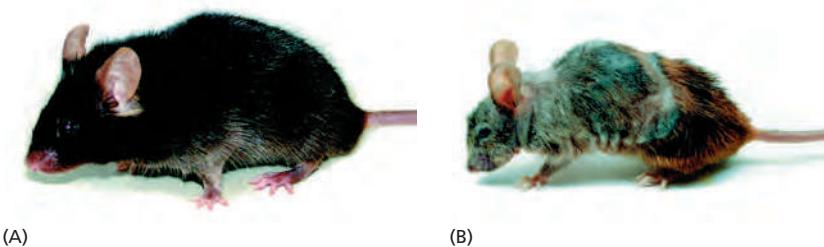
Animals and plants that have been genetically engineered by gene deletion or gene replacement are called **transgenic organisms**, and any foreign or modified genes that are added are called **transgenes**. We discuss transgenic plants later in this chapter and, for now, concentrate our discussion on transgenic mice, as enormous progress has been made in this area. If a DNA molecule carrying a mutated mouse gene is transferred into a mouse cell, it often inserts into the chromosomes at random, but methods have been developed to direct the mutant gene to replace the normal gene by homologous recombination. By exploiting these “gene targeting” events, any specific gene can be altered or inactivated in a mouse cell by a direct gene replacement. In the case in which both copies of the gene of interest are completely inactivated or deleted, the resulting animal is called a “*knockout*” mouse. The technique is summarized in [Figure 8–53](#).



**Figure 8–53** Summary of the procedures used for making gene replacements in mice. In the first step (A), an altered version of the gene is introduced into cultured ES (embryonic stem) cells. These cells are described in detail in Chapter 22. Only a few ES cells will have their corresponding normal genes replaced by the altered gene through a homologous recombination event. These cells can be identified by PCR and cultured to produce many descendants, each of which carries an altered gene in place of one of its two normal corresponding genes. In the next step of the procedure (B), these altered ES cells are injected into a very early mouse embryo; the cells are incorporated into the growing embryo, and a mouse produced by such an embryo will contain some somatic cells (indicated by orange) that carry the altered gene. Some of these mice will also contain germ-line cells that contain the altered gene; when bred with a normal mouse, some of the progeny of these mice will contain one copy of the altered gene in all of their cells.

The mice with the transgene in their germ line are then bred to produce both a male and a female animal, each heterozygous for the gene replacement (that is, they have one normal and one mutant copy of the gene). When these two mice are mated (not shown), one-fourth of their progeny will be homozygous for the altered gene.

The ability to prepare transgenic mice lacking a known normal gene has been a major advance, and the technique has been used to determine the functions of many mouse genes (Figure 8–54). If the gene functions in early development, a knockout mouse will usually die before it reaches adulthood. These lethal defects can be carefully analyzed to help determine the function of the missing gene. As described in Chapter 5, an especially useful type of transgenic animal takes advantage of a site-specific recombination system to excise—and thus disable—the target gene in a particular place or at a particular time (see Figure 5–66). In this case, the target gene in embryonic stem (ES) cells is replaced by a fully functional version of the gene that is flanked by a pair of the short DNA sequences, called *lox* sites, that are recognized by the *Cre recombinase* protein. The transgenic mice that result are phenotypically normal. They are then mated with transgenic mice that express the *Cre recombinase* gene under the control of an inducible promoter. In



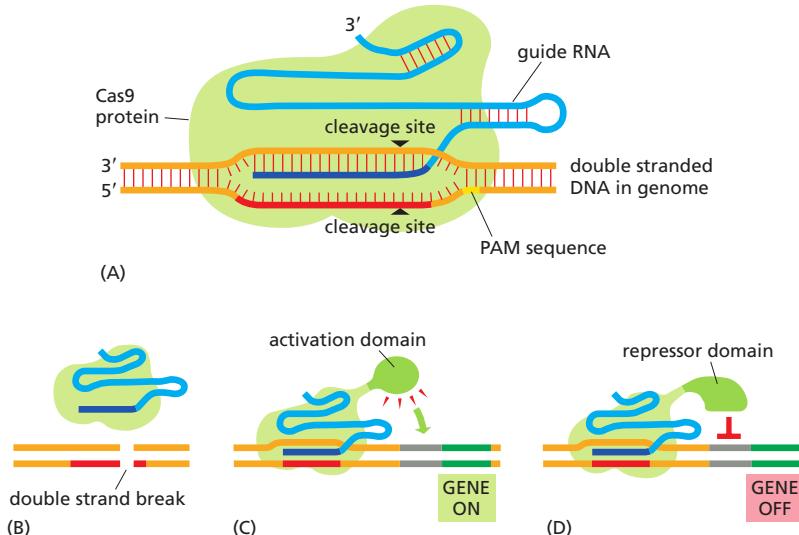
the specific cells or tissues in which Cre is switched on, it catalyzes recombination between the lox sequences—excising a target gene and eliminating its activity (see Figure 22–5).

### The Bacterial CRISPR System Has Been Adapted to Edit Genomes in a Wide Variety of Species

One of the difficulties in making transgenic mice by the procedure just described is that the introduced DNA molecule (bearing the experimentally altered gene) often inserts at random in the genome, and many ES cells must therefore be screened individually to find one that has the “correct” gene replacement.

Creative use of the CRISPR system, discovered in bacteria as a defense against viruses, has largely solved this problem. As described in Chapter 7, the CRISPR system uses a guide RNA sequence to target (through complementary base-pairing) double-stranded DNA, which it then cleaves (see Figure 7–78). The gene coding for the key component of this system, the bacterial Cas9 protein, has been transferred into a variety of organisms, where it greatly simplifies the process of making transgenic organisms (Figure 8–55A and B). The basic strategy is as follows: Cas9 protein is expressed in ES cells along with a guide RNA designed by the experimenter to target a particular location on the genome. The Cas9 and guide RNA associate, the complex is brought to the matching sequence on the genome, and the Cas9 protein makes a double-strand break. As we saw in Chapter 5, double-strand breaks are often repaired by homologous recombination; here, the template chosen by the cell to repair the damage is often the altered gene, which is introduced to ES cells by the experimenter. In this way, the normal gene can be selectively damaged by the CRISPR system and replaced at high efficiency by the experimentally altered gene.

The CRISPR system has a variety of other uses. Its particular power lies with its ability to target Cas9 to thousands of different positions across a genome through the simple rules of complementary base-pairing. Thus, if a catalytically inactive Cas9 protein is fused to a transcription activator or repressor, it is possible, in principle, to turn any gene on or off (Figure 8–55C and D).



**Figure 8–54** Transgenic mice engineered to express a mutant DNA helicase show premature aging. The helicase, encoded by the *Xpd* gene, is involved in both transcription and DNA repair. Compared with a wild-type mouse of the same age (A), a transgenic mouse that expresses a defective version of *Xpd* (B) exhibits many of the symptoms of premature aging, including osteoporosis, emaciation, early graying, infertility, and reduced life-span. The mutation in *Xpd* used here impairs the activity of the helicase and mimics a mutation that in humans causes trichothiodystrophy, a disorder characterized by brittle hair, skeletal abnormalities, and a very reduced life expectancy. These results indicate that an accumulation of DNA damage can contribute to the aging process in both humans and mice. (From J. de Boer et al., *Science* 296:1276–1279, 2002. With permission from AAAS.)

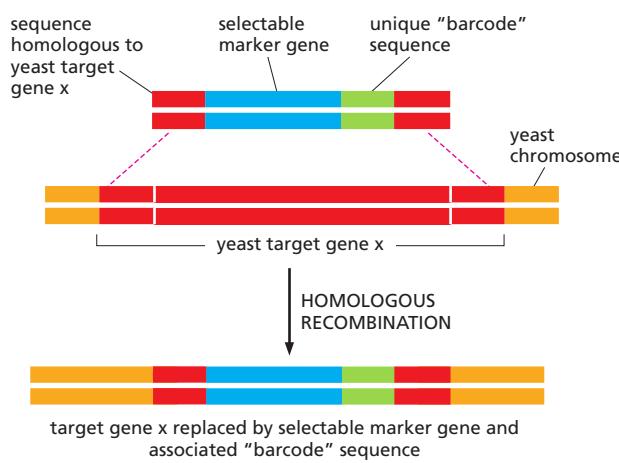
**Figure 8–55** Use of CRISPR to study gene function in a wide variety of species. (A) The Cas9 protein (artificially expressed in the species of interest) binds to a guide RNA, designed by the experimenter and also expressed. The portion of RNA in light blue is needed for associations with Cas9; that in dark blue is specified by the experimenter to match a position on the genome. The only other requirement is that the adjacent genome sequence includes a short PAM (protospacer adjacent motif) that is needed for Cas9 to cleave. As described in Chapter 7, this sequence is how the CRISPR system in bacteria distinguishes its own genome from that of invading viruses. (B) When directed to make double-strand breaks, the CRISPR system greatly improves the ability to replace an endogenous gene with an experimentally altered gene since the altered gene is used to “repair” the double-strand break (C, D). By using a mutant form of Cas9 that can no longer cleave DNA, Cas9 can be used to activate a normally dormant gene (C) or turn off an actively expressed gene (D). (Adapted from P. Mali et al., *Nat. Methods* 10:957–963, 2013. With permission from Macmillan Publishers Ltd.)

The CRISPR system has several advantages over other strategies for experimentally manipulating gene expression. First, it is relatively easy for the experimenter to design the guide RNA: it simply follows standard base pairing convention. Second, the gene to be controlled does not have to be modified; the CRISPR strategy exploits DNA sequences already present in the genome. Third, numerous genes can be controlled simultaneously. Cas9 has to be expressed only once, but many guide RNAs can be expressed in the same cell; this strategy allows the experimenter to turn on or off a whole set of genes at once.

The export of the CRISPR system from bacteria to virtually all other experimental organisms (including mice, zebrafish, worms, flies, rice, and wheat) has revolutionized the study of gene function. Like the earlier discovery of restriction enzymes, this breakthrough came from scientists studying a fascinating phenomenon in bacteria without—at first—realizing the enormous impact these discoveries would have on all aspects of biology.

### Large Collections of Engineered Mutations Provide a Tool for Examining the Function of Every Gene in an Organism

Extensive collaborative efforts have produced comprehensive libraries of mutations in a variety of model organisms, including *S. cerevisiae*, *C. elegans*, *Drosophila*, *Arabidopsis*, and even the mouse. The ultimate aim in each case is to produce a collection of mutant strains in which every gene in the organism has been systematically deleted or altered in such a way that it can be conditionally disrupted. Collections of this type provide an invaluable resource for investigating gene function on a genomic scale. For example, a large collection of mutant organisms can be screened for a particular phenotype. Like the classic genetic approaches described earlier, this is one of the most powerful ways to identify the genes responsible for a particular phenotype. Unlike the classical genetic approach, however, the set of mutants is “pre-engineered,” so that there is no need to rely on chance events such as spontaneous mutations or transposon insertions. In addition, each of the individual mutations within the collection is often engineered to contain a distinct molecular “barcode”—in the form of a unique DNA sequence—designed to make identification of the altered gene rapid and routine (Figure 8–56).



**Figure 8–56 Making barcoded collections of mutant organisms.** A deletion construct for use in yeast contains DNA sequences (red) homologous to each end of a target gene x, a selectable marker gene (blue), and a unique “barcode” sequence approximately 20 nucleotide pairs in length (green). This DNA is introduced into yeast cells, where it readily replaces the target gene by homologous recombination. Cells that carry a successful gene replacement are identified by expression of the selectable marker gene, typically a gene that provides resistance to a drug. By using a collection of such constructs, each specific for one gene, a library of yeast mutants was constructed containing a mutant for every gene. Essential genes cannot be studied this way, as their deletion from the genome causes the cells to die. In this case, the target gene is replaced by a version of the gene that can be regulated by the experimenter (see Figure 8–52). The gene can then be turned off and the effect of this can be monitored before the cells die.

**Figure 8–57** Genome-wide screens for fitness using a large pool of barcoded yeast deletion mutants. A large pool of yeast mutants, each with a different gene deleted and present in equal amounts, is grown under conditions selected by the experimenter. Some mutants (blue) grow normally, but others show reduced growth (orange and green) or no growth at all (red). The fitness of each mutant is experimentally determined in the following way. After the growth phase is completed, genomic DNA (isolated from the mixture of strains) is purified and the relative abundance of each mutant is determined by quantifying the level of the DNA barcode matched to each deletion. This can be done by sequencing the pooled genomic DNA or hybridizing it to microarrays (see Figure 8–64) that contain DNA oligonucleotides complementary to each barcode. In this way, the contribution of every gene to growth under the specified condition can be rapidly ascertained. This type of study has revealed that of the approximately 6000 coding genes in yeast, only about 1000 are essential under standard growth conditions.

In *S. cerevisiae*, the task of generating a complete set of 6000 mutants, each missing only one gene, was accomplished several years ago. Because each mutant strain has an individual barcode sequence embedded in its genome, a large mixture of engineered strains can be grown under various selective test conditions—such as nutritional deprivation, a temperature shift, or the presence of various drugs—and the cells that survive can be rapidly identified by the unique sequence tags present in their genomes. By assessing how well each mutant in the mixture fares, one can begin to discern which genes are essential, useful, or irrelevant for growth under the various conditions (Figure 8–57).

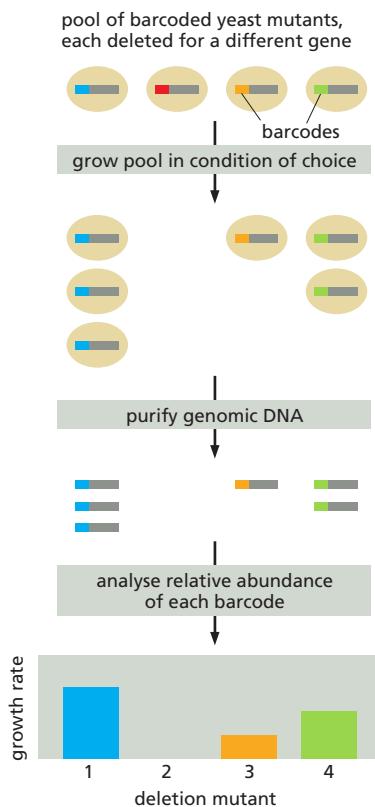
The insights generated by examining mutant libraries can be considerable. For example, studies of an extensive collection of mutants in *Mycoplasma genitalium*—the organism with the smallest known genome—have identified the minimum complement of genes essential for cellular life. Growth under laboratory conditions requires about three-quarters of the 480 protein-coding genes in *M. genitalium*. Approximately 100 of these essential genes are of unknown function, which suggests that a surprising number of the basic molecular mechanisms that underlie life have yet to be discovered.

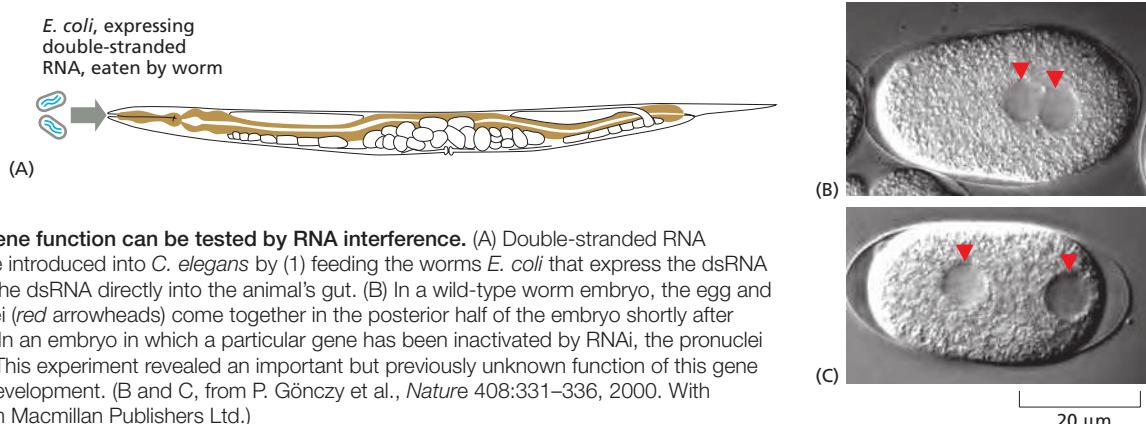
Collections of mutant organisms are also available for many animal and plant species. For example, it is possible to “order,” by phone or email from a consortium of investigators, a deletion or insertion mutant for almost all coding genes in *Drosophila*. Likewise, a nearly complete set of mutants exists for the “model” plant *Arabidopsis*. And the adaptation of the CRISPR system for use in mice means that, in the near future, we can expect to be able to turn on or off—at will—each gene in the mouse genome. Although we are still ignorant about the function of most genes in most organisms, these technologies allow an exploration of gene function on a scale that was unimaginable a decade ago.

### RNA Interference Is a Simple and Rapid Way to Test Gene Function

Although knocking out (or conditionally expressing) a gene in an organism and studying the consequences is the most powerful approach for understanding the functions of the gene, *RNA interference* (*RNAi*, for short), is an alternative, particularly convenient approach. As discussed in Chapter 7, this method exploits a natural mechanism used in many plants, animals, and fungi to protect themselves against viruses and transposable elements. The technique introduces into a cell or organism a double-stranded RNA molecule whose nucleotide sequence matches that of part of the gene to be inactivated. After the RNA is processed, it hybridizes with the target-gene RNA (either mRNA or noncoding RNA) and reduces its expression by the mechanisms shown in Figure 7–75.

*RNAi* is frequently used to inactivate genes in *Drosophila* and mammalian cell culture lines. Indeed, a set of 15,000 *Drosophila* *RNAi* molecules (one for every coding gene) allows researchers, in several months, to test the role of every fly gene in any process that can be monitored using cultured cells. *RNAi* has also been widely used to study gene function in whole organisms, including the nematode



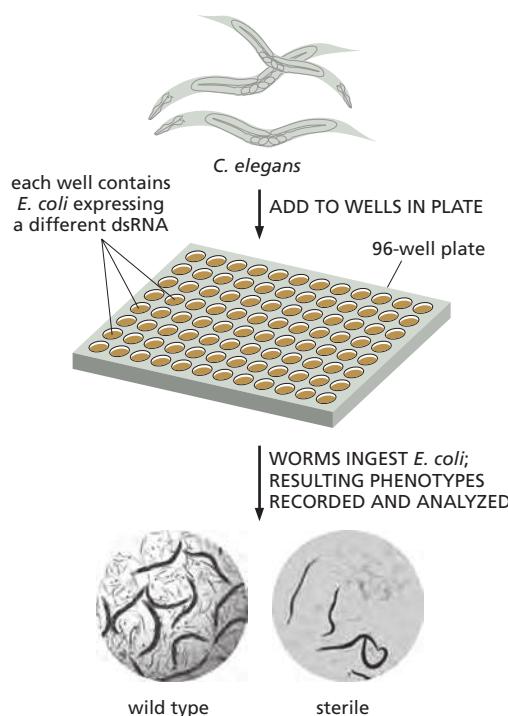


**Figure 8–58** Gene function can be tested by RNA interference. (A) Double-stranded RNA (dsRNA) can be introduced into *C. elegans* by (1) feeding the worms *E. coli* that express the dsRNA or (2) injecting the dsRNA directly into the animal's gut. (B) In a wild-type worm embryo, the egg and sperm pronuclei (red arrowheads) come together in the posterior half of the embryo shortly after fertilization. (C) In an embryo in which a particular gene has been inactivated by RNAi, the pronuclei fail to migrate. This experiment revealed an important but previously unknown function of this gene in embryonic development. (B and C, from P. Gönczy et al., *Nature* 408:331–336, 2000. With permission from Macmillan Publishers Ltd.)

*C. elegans*. When working with worms, introducing the double-stranded RNA is quite simple: the RNA can be injected directly into the intestine of the animal, or the worm can be fed with *E. coli* engineered to produce the RNA (Figure 8–58). The RNA is amplified (see p. 431) and distributed throughout the body of the worm, where it inhibits expression of the target gene in different tissue types. RNAi is being used to help in assigning functions to the entire complement of worm genes (Figure 8–59).

A related technique has also been applied to mice. In this case, the RNAi molecules are not injected or fed to the mouse; rather, recombinant DNA techniques are used to make transgenic animals that express the RNAi under the control of an inducible promoter. Often this is a specially designed RNA that can fold back on itself and, through base-pairing, produce a double-stranded region that is recognized by the RNAi machinery. In the simplest cases, the process inactivates only the genes that exactly match the RNAi sequence. Depending on the inducible promoter used, the RNAi can be produced only in a specified tissue or only at a particular time in development, allowing the functions of the target genes to be analyzed in elaborate detail.

RNAi has made reverse genetics simple and efficient in many organisms, but it has several potential limitations compared with true genetic knockouts. For



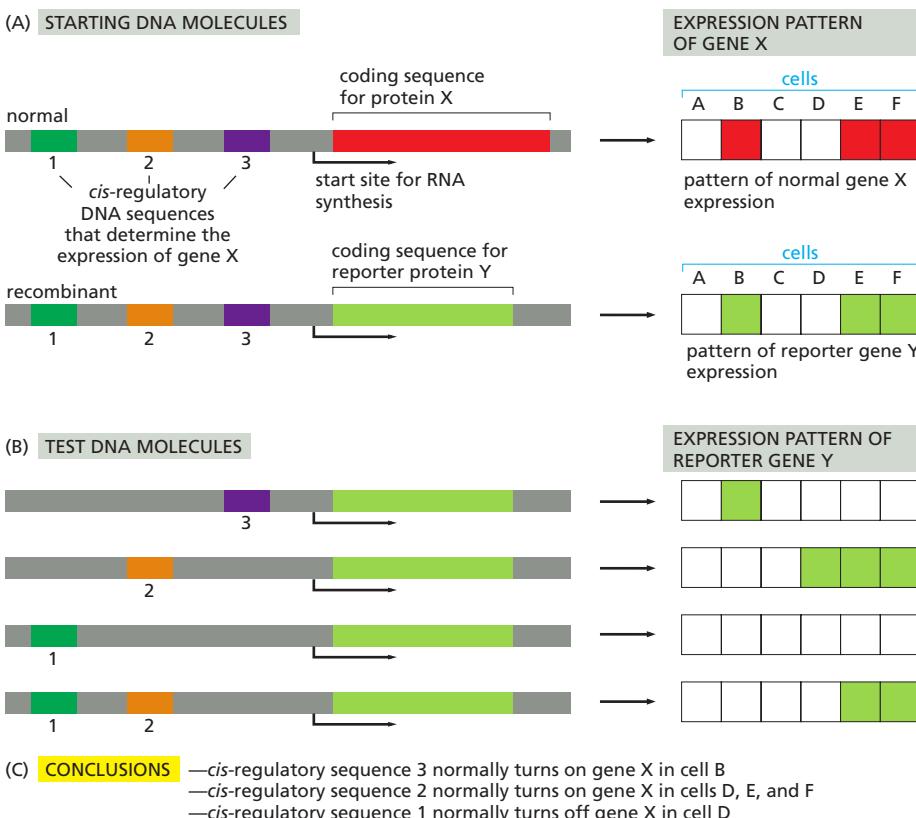
**Figure 8–59** RNA interference provides a convenient method for conducting genome-wide genetic screens. In this experiment, each well in this 96-well plate is filled with *E. coli* that produce a different double-stranded RNA. Each interfering RNA matches the nucleotide sequence of a single *C. elegans* gene, thereby inactivating it. About 10 worms are added to each well, where they ingest the genetically modified bacteria. The plate is incubated for several days, which gives the RNAs time to inactivate their target genes—and the worms time to grow, mate, and produce offspring. The plate is then examined in a microscope, which can be controlled robotically, to screen for genes that affect the worms' ability to survive, reproduce, develop, and behave. Shown here are normal worms alongside worms that show an impaired ability to reproduce due to inactivation of a particular "fertility" gene. (From B. Lehner et al., *Nat. Genet.* 38:896–903, 2006. With permission from Macmillan Publishers Ltd.)

unknown reasons, RNAi does not efficiently inactivate all genes. Moreover, within whole organisms, certain tissues may be resistant to the action of RNAi (for example, neurons in nematodes). Another problem arises because many organisms contain large gene families, the members of which exhibit sequence similarity. RNAi therefore sometimes produces “off-target” effects, inactivating related genes in addition to the targeted gene. One strategy to avoid such problems is to use multiple small RNA molecules matched to different regions of the same gene. Ultimately, the results of any RNAi experiment must be viewed as a strong clue to, but not necessarily a proof of, normal gene function.

### Reporter Genes Reveal When and Where a Gene Is Expressed

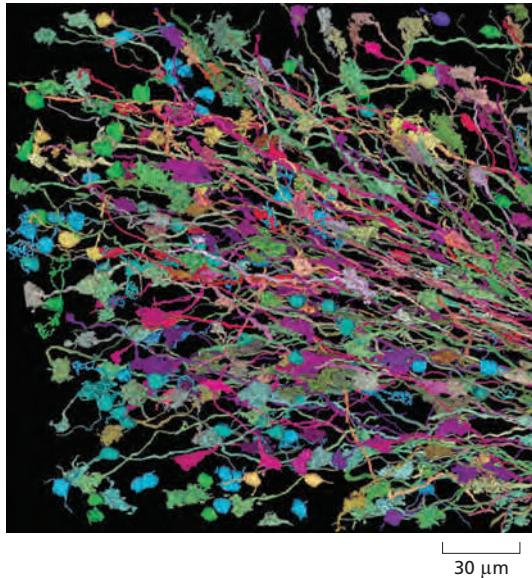
In the preceding section, we discussed how genetic approaches can be used to assess a gene’s function in cultured cells or, even better, in the intact organism. Although this information is crucial to understanding gene function, it does not generally reveal the molecular mechanisms through which the gene product works in the cell. For example, genetics on its own rarely tells us all the places in the organism where the gene is expressed, or how its expression is controlled. It does not necessarily reveal whether the gene acts in the nucleus, the cytosol, on the cell surface, or in one of the numerous other compartments of the cell. And it does not reveal how a gene product might change its location or its expression pattern when the external environment of the cell changes. Key insights into gene function can be obtained by simply observing when and where a gene is expressed. A variety of approaches, most involving some form of genetic engineering, can easily provide this critical information.

As discussed in detail in Chapter 7, *cis*-regulatory DNA sequences, located upstream or downstream of the coding region, control gene transcription. These regulatory sequences, which determine precisely when and where the gene is expressed, can be easily studied by placing a reporter gene under their control and introducing these recombinant DNA molecules into cells (Figure 8–60). In this way, the normal expression pattern of a gene can be determined, as well as



**Figure 8–60** Using a reporter protein to determine the pattern of a gene’s expression. (A) In this example, the coding sequence for protein X is replaced by the coding sequence for reporter protein Y. The expression patterns for X and Y are the same. (B) Various fragments of DNA containing candidate *cis*-regulatory sequences are added in combinations to produce test DNA molecules encoding reporter gene Y. These recombinant DNA molecules are then tested for expression after introducing them into a variety of different types of mammalian cells. The results are summarized in (C).

For experiments in eukaryotic cells, two commonly used reporter proteins are the enzyme  $\beta$ -galactosidase ( $\beta$ -gal) (see Figure 7–28) and green fluorescent protein (GFP) (see Figure 9–22).



the contribution of individual *cis*-regulatory sequences in establishing this pattern (see also Figure 7–29).

Reporter genes also allow any protein to be tracked over time in living cells. Here, the reporter gene typically encodes a fluorescent protein, often **green fluorescent protein (GFP)**, the molecule that gives luminescent jellyfish their greenish glow. The GFP is simply attached—in the coding frame—to the protein-coding gene of interest. The resulting *GFP fusion protein* often behaves in the same way the normal protein does and its location can be monitored by fluorescence microscopy, a topic that is discussed in the next chapter (see Figure 9–25). GFP fusion has become a standard strategy for tracking not only the location but also the movement of specific proteins in living cells. In addition, the use of multiple GFP variants that fluoresce at different wavelengths can provide insights into how different cells interact in a living tissue (Figure 8–61).

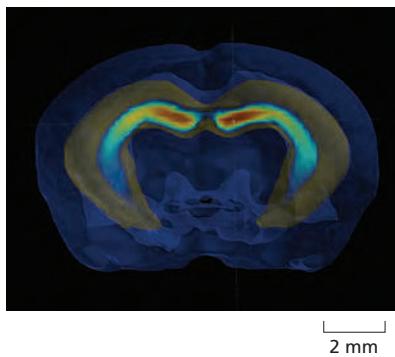
### *In situ Hybridization Can Reveal the Location of mRNAs and Noncoding RNAs*

It is also possible to directly observe the time and place that an RNA product of a gene is expressed using *in situ hybridization*. For protein-coding genes, this strategy often provides the same general information as the reporter gene approaches described above; however, it is crucial for genes whose final product is RNA rather than protein. We encountered *in situ* hybridization earlier in the chapter (see Figure 8–34); it relies on the basic principles of nucleic acid hybridization. Typically, tissues are gently fixed so that their RNA is retained in an exposed form that can hybridize with a labeled complementary DNA or RNA probe. In this way, the patterns of differential gene expression can be observed in tissues, and the location of specific RNAs can be determined (Figure 8–62). An advantage of *in situ* hybridization over other approaches is that genetic engineering is not required. Thus, it is often simpler and faster and can be used for genetically intractable species.

### Expression of Individual Genes Can Be Measured Using Quantitative RT-PCR

Although reporter genes and *in situ* hybridization accurately reveal patterns of gene expression, they are not the most powerful methods for quantifying amounts of individual RNAs in cells. We have seen that RNA sequencing can provide information about the relative abundance of different RNA molecules (see Figure 7–3). Here, the number of “sequence reads” (short bits of nucleotide sequence) is proportional to the abundance of the RNA species. But this method is limited to RNAs

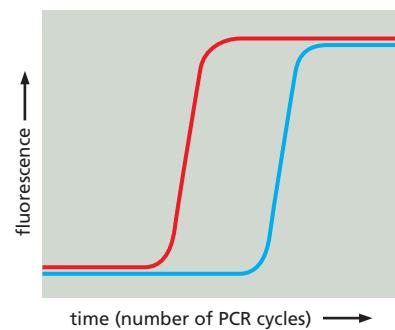
**Figure 8–61** GFPs that fluoresce at different wavelengths help reveal the connections that individual neurons make within the brain. This image shows differently colored neurons in one region of a mouse brain. The neurons randomly express different combinations of differently colored GFPs (see Figure 9–13), making it possible to distinguish and trace many individual neurons within a population. These images were obtained by genetically engineering the genes for four different fluorescent proteins, each flanked by loxP sites of recombination (see Figure 5–66), and integrating them into the mouse germ line. When crossed to a mouse that produced the Cre recombinase in neuronal cells, the fluorescent protein genes were randomly excised, producing neurons that express many different combinations of the four fluorescent proteins. Over 100 combinations of fluorescent protein can be produced, allowing scientists to distinguish one neuron from the next. The stunning appearance of these labeled neurons has earned these animals the colorful nickname “brainbow mice.” (From J. Livet et al., *Nature* 450:56–62, 2007. With permission from Macmillan Publishers Ltd.)



**Figure 8–62** *In situ* hybridization to mRNAs has been used to generate an atlas of gene expression in the mouse brain. This computer-generated image shows the expression of several different mRNAs specific to an area of the brain associated with learning and memory. Similar maps of expression patterns of all known genes in the mouse brain are compiled in the brain atlas project, which is available online. (From M. Hawrylycz et al., *PLoS Comput. Biol.* 7:e1001065, 2011.)

**Figure 8–63 RNA levels can be measured by quantitative RT-PCR.** The fluorescence measured is generated by a dye that fluoresces only when bound to the double-stranded DNA products of the RT-PCR (see Figure 8–36). The red sample has a higher concentration of the mRNA being measured than does the blue sample, since it requires fewer PCR cycles to reach the same half-maximal concentration of double-stranded DNA. Based on this difference, the relative amounts of the mRNA in the two samples can be precisely determined.

that are expressed at reasonably high levels, and it is difficult to quantify (or even identify) rare RNAs. A more accurate method is based on the principles of PCR (**Figure 8–63**). Called **quantitative RT-PCR** (reverse transcription-polymerase chain reaction), this method begins with the total population of RNA molecules purified from a tissue or a cell culture. It is important that no DNA be present in the preparation; it must be purified away or enzymatically degraded. Two DNA primers that specifically match the mRNA of interest are added, along with reverse transcriptase, DNA polymerase, and the four deoxyribonucleoside triphosphates needed for DNA synthesis. The first round of synthesis is the reverse transcription of the RNA into DNA using one of the primers. Next, a series of heating and cooling cycles allows the amplification of that DNA strand by PCR (see Figure 8–36). The quantitative part of this method relies on a direct relationship between the rate at which the PCR product is generated and the original concentration of the mRNA species of interest. By adding chemical dyes to the PCR that fluoresce only when bound to double-stranded DNA, a simple fluorescence measurement can be used to track the progress of the reaction and thereby accurately deduce the starting concentration of the mRNA that is amplified. Although it seems complicated, this quantitative RT-PCR technique is relatively fast and simple to perform in the laboratory; it is currently the method of choice for accurately quantifying mRNA levels from any given gene.



### Analysis of mRNAs by Microarray or RNA-seq Provides a Snapshot of Gene Expression

As discussed in Chapter 7, a cell expresses only a subset of the many thousands of genes available in its genome; moreover, this subset differs from one cell type to another or, in the same cell, from one environment to the next. One way to determine which genes are being expressed by a population of cells or a tissue is to analyze which mRNAs are being produced.

The first tool that allowed investigators to analyze simultaneously the thousands of different RNAs produced by cells or tissues was the **DNA microarray**. Developed in the 1990s, DNA microarrays are glass microscope slides that contain hundreds of thousands of DNA fragments, each of which serves as a probe for the mRNA produced by a specific gene. Such microarrays allow investigators to monitor the expression of every gene in a genome in a single experiment. To do the analysis, mRNAs are extracted from cells or tissues and converted to cDNAs (see Figure 8–31). The cDNAs are fluorescently labeled and allowed to hybridize to the fragments bound to the microarray. An automated fluorescence microscope then determines which mRNAs were present in the original sample based on the array positions to which the cDNAs are bound (**Figure 8–64**).

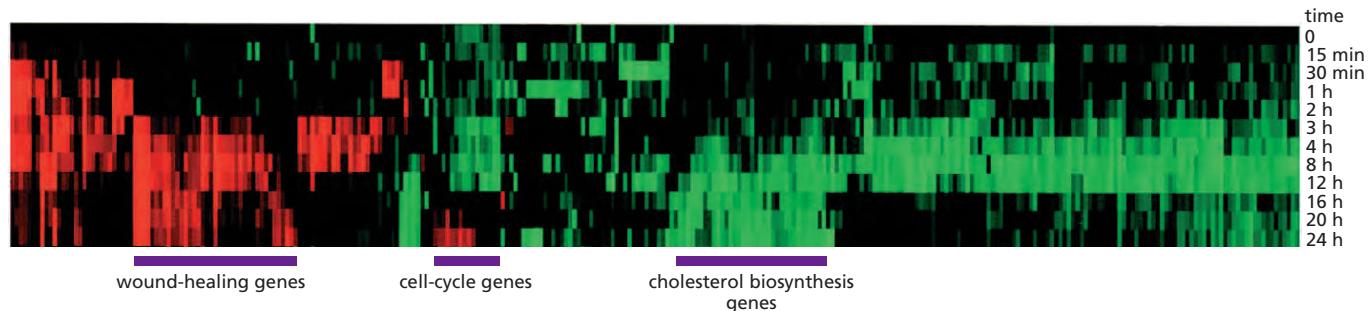
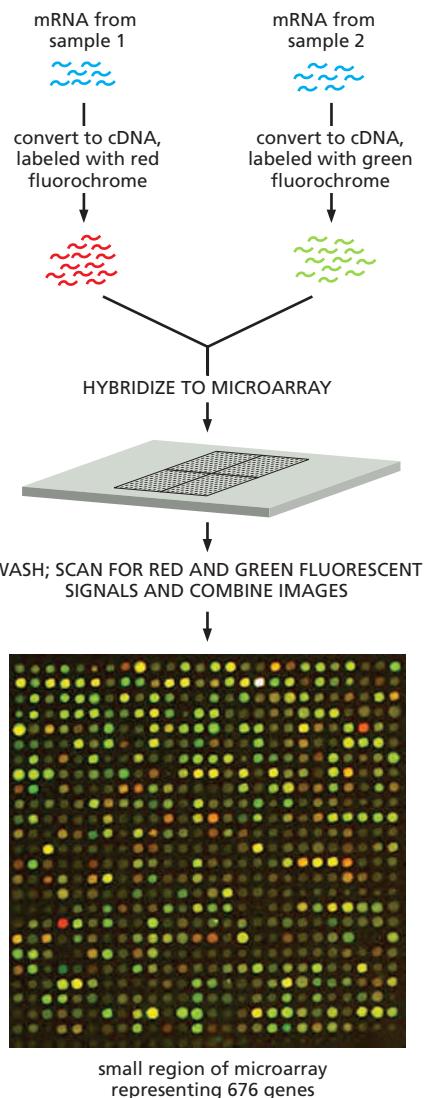
Although microarrays are relatively inexpensive and easy to use, they suffer from one obvious drawback: the sequences of the mRNA samples to be analyzed must be known in advance and represented by a corresponding probe on the array. With the development of improved sequencing technologies, investigators increasingly use **RNA-seq**, discussed earlier, as a more direct approach for cataloging the RNAs produced by a cell. For example, this approach can readily detect alternative RNA splicing, RNA editing, and the many noncoding RNAs produced from a complex genome.

DNA microarrays and RNA-seq analysis have been used to examine everything from the changes in gene expression that make strawberries ripen to the gene expression “signatures” of different types of human cancer cells; or from changes

**Figure 8–64 DNA microarrays are used to analyze the production of thousands of different mRNAs in a single experiment.** In this example, mRNA is collected from two different cell samples—for example, cells treated with a hormone and untreated cells of the same type—to allow for a direct comparison of the specific genes expressed under both conditions. The mRNAs are converted to cDNAs that are labeled with a red fluorescent dye for one sample and a green fluorescent dye for the other. The labeled samples are mixed and then allowed to hybridize to the microarray. Each microscopic spot on the microarray is a 50-nucleotide DNA molecule of defined sequence made by chemical synthesis and spotted on the array. The DNA sequence represented by each spot is different, and the hundreds of thousands of such spots are designed to span the sequence of the genome. The DNA sequence of each spot is kept track of by computer. After incubation, the array is washed and the fluorescence scanned. Only a small proportion of the microarray, representing 676 genes, is shown. Red spots indicate that the gene in sample 1 is expressed at a higher level than the corresponding gene in sample 2, and the green spots indicate the opposite. Yellow spots reveal genes that are expressed at about equal levels in both cell samples. The intensity of the fluorescence provides an estimate of how much RNA is present from a gene. Dark spots indicate little or no expression of the gene whose probe is located at that position in the array.

that occur as cells progress through the cell cycle to those made in response to sudden shifts in temperature. Indeed, because these approaches allow the simultaneous monitoring of large numbers of RNAs, they can detect subtle changes in a cell, changes that might not be manifested in its outward appearance or behavior.

Comprehensive studies of gene expression also provide information that is useful for predicting gene function. Earlier in this chapter, we discussed how identifying a protein's interaction partners can yield clues about that protein's function. A similar principle holds true for genes: information about a gene's function can be deduced by identifying genes that share its expression pattern. Using an approach called *cluster analysis*, one can identify sets of genes that are coordinately regulated. Genes that are turned on or turned off together under different circumstances are likely to work in concert in the cell: they may encode proteins that are part of the same multiprotein machine, or proteins that are involved in a complex coordinated activity, such as DNA replication or RNA splicing. Characterizing a gene whose function is unknown by grouping it with known genes that share its transcriptional behavior is sometimes called "guilt by association." Cluster analyses have been used to analyze the gene expression profiles that underlie many interesting biological processes, including wound healing in humans (Figure 8–65).



**Figure 8–65 Using cluster analysis to identify sets of genes that are coordinately regulated.** Genes that have the same expression pattern are likely to be involved in common pathways or processes. To perform a cluster analysis, RNA-seq or microarray data are obtained from cell samples exposed to a variety of different conditions, and genes that show coordinate changes in their expression pattern are grouped together. In this experiment, human fibroblasts were deprived of serum for 48 hours; serum was then added back to the cultures at time 0 and the cells were harvested for microarray analysis at different time points. Of the 8600 genes depicted here (each represented by a thin, vertical line), just over 300 showed threefold or greater variation in their expression patterns in response to serum reintroduction. Here, red indicates an increase in expression; green is a decrease in expression. On the basis of the results of many other experiments, the 8600 genes have been grouped in clusters based on similar patterns of expression. The results of this analysis show that genes involved in wound healing are turned on in response to serum, while genes involved in regulating cell-cycle progression and cholesterol biosynthesis are shut down. (From M.B. Eisen et al., Proc. Natl. Acad. Sci. USA 94:14863–14868, 1998. With permission from National Academy of Sciences.)

**Figure 8–66 Chromatin immunoprecipitation.** This method allows the identification of all the sites in a genome that a transcription regulator occupies *in vivo*. The identities of the precipitated, amplified DNA fragments are determined by DNA sequencing.

### Genome-wide Chromatin Immunoprecipitation Identifies Sites on the Genome Occupied by Transcription Regulators

We have discussed several strategies to measure the levels of individual RNAs in a cell and to monitor changes in their levels in response to external signals. But this information does not tell us how such changes are brought about. We saw in Chapter 7 that transcription regulators, by binding to *cis*-regulatory sequences in DNA, are responsible for establishing and changing patterns of transcription. Typically, these proteins do not occupy all of their potential *cis*-regulatory sequences in the genome under all conditions. For example, in some cell types, the regulatory protein may not be expressed, or it may be present but lack an obligatory partner protein, or it may be excluded from the nucleus until an appropriate signal is received from the cell's environment. Even if the protein is present in the nucleus and is competent to bind DNA, other transcription regulators or components of chromatin can occupy overlapping DNA sequences and thereby occlude some of its *cis*-regulatory sequences in the genome.

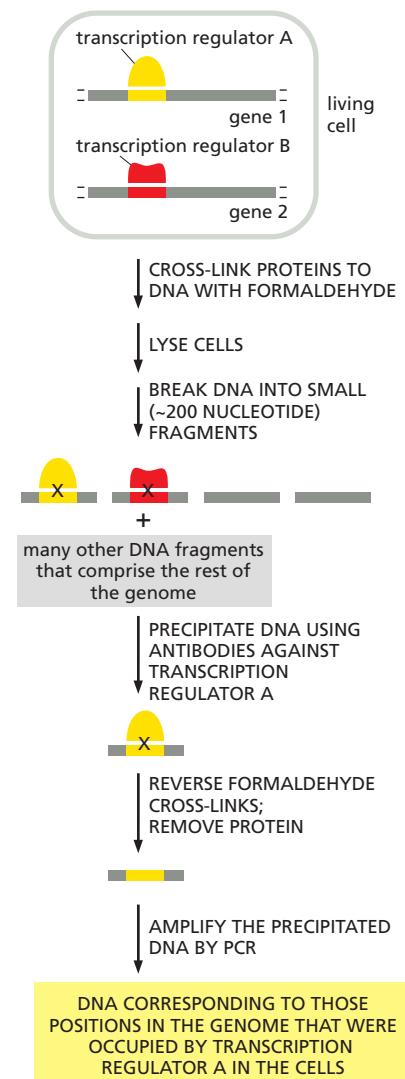
**Chromatin immunoprecipitation** provides a way to experimentally determine all the *cis*-regulatory sequences in a genome that are occupied by a given transcription regulator under a particular set of conditions (Figure 8–66). In this approach, proteins are covalently cross-linked to DNA in living cells, the cells are broken open, and the DNA is mechanically sheared into small fragments. Antibodies directed against a given transcription regulator are then used to purify the DNA that became covalently cross-linked to that protein in the cell. This DNA is then sequenced using the rapid methods discussed earlier; the precise location of each precipitated DNA fragment along the genome is determined by comparing its DNA sequence to that of the whole genome sequence (Figure 8–67). In this way, all of the sites occupied by the transcription regulator in the cell sample can be mapped across the cell's genome (see Figure 7–37). In combination with microarray or RNA-seq information, chromatin immunoprecipitation can identify the key transcriptional regulator responsible for specifying a particular pattern of gene expression.

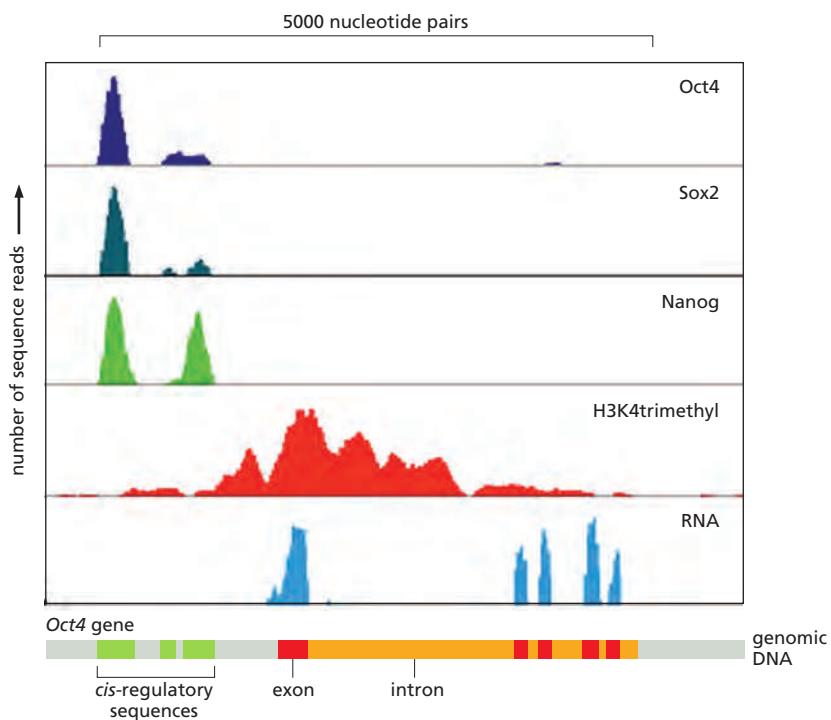
Chromatin immunoprecipitation can also be used to deduce the *cis*-regulatory sequences recognized by a given transcription regulator. Here, all the DNA sequences precipitated by the regulator are lined up (by computer) and features in common are tabulated to produce the spectrum of *cis*-regulatory sequences recognized by the protein (see Figure 7–9A). Chromatin immunoprecipitation is also used routinely to identify the positions along a genome that are bound by the various types of modified histones discussed in Chapter 4. In this case, antibodies specific to the particular histone modification are employed (see Figure 8–67). A variation of the technique can also be used to map positions of chromosomes that are in physical proximity (see Figure 4–48).

### Ribosome Profiling Reveals Which mRNAs Are Being Translated in the Cell

In preceding sections, we discussed several ways that RNA levels in the cell can be monitored. But for mRNAs, this represents only one step in gene expression, and we are often more interested in the final level of the protein produced by the gene. As described in the first part of this chapter, mass-spectroscopy methods can be used to monitor the levels of all proteins in the cell, including modified forms of the proteins. However, if we want to understand *how* synthesis of proteins is controlled by the cell, we need to consider the translation step of gene expression.

An approach called *ribosome profiling* provides an instantaneous map of the position of ribosomes on each mRNA in the cell and thereby identifies those





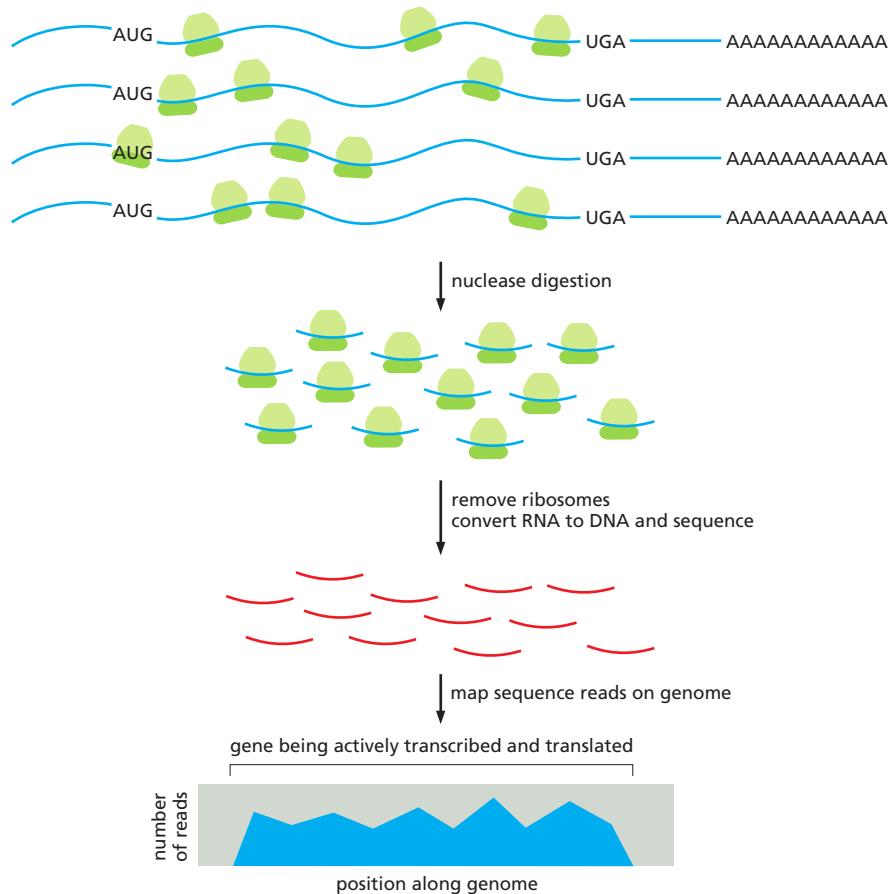
**Figure 8–67** Results of several chromatin immunoprecipitations showing proteins bound to the control region that control expression of the *Oct4* gene. In this series of chromatin immunoprecipitation experiments, antibodies directed against a transcription regulator (first three panels) or a particular histone modification (fourth panel) were used to precipitate bound, cross-linked DNA. Precipitated DNA was sequenced, and the positions across the genome were mapped. (Only the small part of the mouse genome containing the *Oct4* gene is shown.) The results show that, in the embryonic stem cells analyzed in these experiments, Oct4 binds upstream of its own gene and that Sox2 and Nanog are bound in close proximity. Oct4, Sox2, and Nanog are key regulators in embryonic stem cells (discussed in Chapter 22) and this experiment reveals the position on the genome through which they exert their effects on *Oct4* expression. In the fourth panel, the positions of a histone modification associated with actively transcribed genes is shown (see Figure 4–39). Finally, the bottom panel shows the RNA produced from the *Oct4* gene under the same conditions used for the chromatin immunoprecipitations. Note that the introns and exons are relatively easy to identify from these RNA-seq data.

mRNAs that are being actively translated. To accomplish this, total RNA from a cell line or tissue is exposed to RNases under conditions where only those RNA sequences covered by ribosomes are spared. The protected RNAs are released from ribosomes, converted to DNA, and the nucleotide sequence of each is determined (Figure 8–68). When these sequences are mapped on the genome, the position of ribosomes across each mRNA species can be ascertained.

Ribosome profiling has revealed many cases where mRNAs are abundant but are not translated until the cell receives an external signal. It has also shown that many open reading frames (ORFs) that were too short to be annotated as genes are actively translated and probably encode functional, albeit very small, proteins (Figure 8–69). Finally, ribosome profiling has revealed the ways that cells rapidly and globally change their translation patterns in response to sudden changes in temperature, nutrient availability, or chemical stress.

### Recombinant DNA Methods Have Revolutionized Human Health

We have seen that nucleic acid methodologies developed in the past 40 years have completely changed the way that cell and molecular biology is studied. But they have also had a profound effect on our day-to-day lives. Many human pharmaceuticals in routine use (insulin, human growth hormone, blood-clotting factors, and interferon, for example) are based on cloning human genes and expressing the encoded proteins in large amounts. As DNA sequencing continues to drop in cost, more and more individuals will elect to have their genome sequenced; this information can be used to predict susceptibility to diseases (often with the option of minimizing this possibility by appropriate behavior) or to predict the way an individual will respond to a given drug. The genomes of tumor cells from an individual can be sequenced to determine the best type of anticancer treatment. And mutations that cause or greatly increase the risk of disease continue to be identified at an unprecedented pace. Using the recombinant DNA technologies discussed in this chapter, these mutations can then be introduced into animals, such as mice, that can be studied in the laboratory. The resulting transgenic animals, which often mimic some of the phenotypic abnormalities associated with the condition in patients, can be used to explore the cellular and molecular basis of the disease and to screen for drugs that could potentially be used therapeutically in humans.

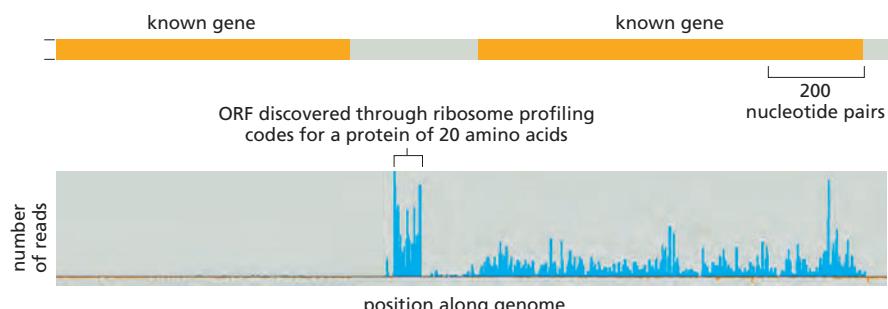


**Figure 8–68 Ribosome profiling.** RNA is purified from cells and digested with an RNase to leave only those portions of the mRNAs that are protected by a bound ribosome. These short pieces of protected RNA (approximately 20 nucleotides in length) are converted to DNA and sequenced. The resulting information is displayed as the number of sequence reads along each position of the genome. In the diagram here, the data for only one gene, whose mRNA is being efficiently translated, are shown. Ribosome profiling provides this type of information for every mRNA produced by the cell.

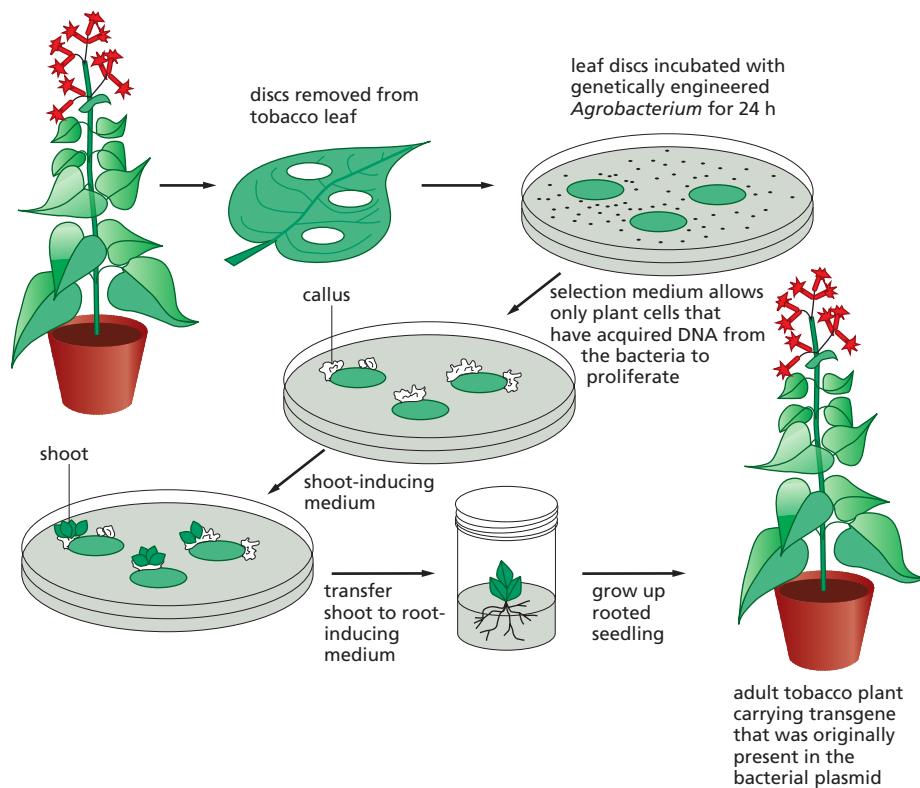
## Transgenic Plants Are Important for Agriculture

Although we tend to think of recombinant DNA research in terms of animal biology, these techniques have also had a profound impact on the study of plants. In fact, certain features of plants make them especially amenable to recombinant DNA methods.

When a piece of plant tissue is cultured in a sterile medium containing nutrients and appropriate growth regulators, some of the cells are stimulated to proliferate indefinitely in a disorganized manner, producing a mass of relatively undifferentiated cells called a *callus*. If the nutrients and growth regulators are carefully manipulated, one can induce the formation of a shoot within the callus, and in many species a whole new plant can be regenerated from such shoots. In a number of plants—including tobacco, petunia, carrot, potato, and *Arabidopsis*—a single cell from such a callus (known as a *totipotent cell*) can be grown into a small clump of cells from which a whole plant can be regenerated (see Figure 7–2B). Just as mutant mice can be derived by the genetic manipulation of embryonic stem



**Figure 8–69 Ribosome profiling can identify new genes.** This experiment shows the discovery of a previously unrecognized gene—one that encodes a protein of only 20 amino acids. At the top is shown a portion of a viral genome with two previously annotated genes. Below are the results of a ribosome profiling experiment, displayed across the same section of the genome, after the virus was infected into human cells. The results show that the left-hand gene is not expressed under these conditions, the right-hand gene is expressed at low levels, and a previously unrecognized gene that lies between them is expressed at high levels.



**Figure 8–70** Transgenic plants can be made using recombinant DNA techniques optimized for plants. A disc is cut out of a leaf and incubated in a culture of *Agrobacterium* that carries a recombinant plasmid with both a selectable marker and a desired genetically engineered gene. The wounded plant cells at the edge of the disc release substances that attract the bacteria, which inject their DNA into the plant cells. Only those plant cells that take up the appropriate DNA and express the selectable marker gene survive and proliferate and form a callus. The manipulation of growth factors supplied to the callus induces it to form shoots, which subsequently root and grow into adult plants carrying the engineered gene.

cells in culture, so transgenic plants can be created from plant cells transfected with DNA in culture (Figure 8–70).

The ability to produce transgenic plants has greatly accelerated progress in many areas of plant cell biology. It has played an important part, for example, in isolating receptors for growth regulators and in analyzing the mechanisms of morphogenesis and of gene expression in plants. These techniques have also opened up many new possibilities in agriculture that could benefit both the farmer and the consumer. They have made it possible, for example, to modify the ratio of lipid, starch, and protein in seeds, to impart pest and virus resistance to plants, and to create modified plants that tolerate extreme habitats such as salt marshes or water-stressed soil. One variety of rice has been genetically engineered to produce β-carotene, the precursor of vitamin A. Were it to replace conventional rice, this “golden rice”—so-called because of its faint yellow color—could help to alleviate severe vitamin A deficiency, which causes blindness in hundreds of thousands of children in the developing world each year.

## Summary

*Genetics and genetic engineering provide powerful tools for understanding the function of individual genes in cells and organisms. In the classical genetic approach, random mutagenesis is coupled with screening to identify mutants that are deficient in a particular biological process. These mutants are then used to locate and study the genes responsible for that process.*

*Gene function can also be ascertained by reverse genetic techniques. DNA engineering methods can be used to alter genes and to re-insert them into a cell's chromosomes so that they become a permanent part of the genome. If the cell used for this gene transfer is a fertilized egg (for an animal) or a totipotent plant cell in culture, transgenic organisms can be produced that express the mutant gene and pass it on to their progeny. Especially important for cell and molecular biology is the ability to alter cells and organisms in highly specific ways—allowing one to discern the effect on the cell or the organism of a designed change in a single protein or RNA molecule. For example, genomes can be altered so that the expression of any gene can be switched on or off by the experimenter.*

*Many of these methods are being expanded to investigate gene function on a genome-wide scale. The generation of mutant libraries in which every gene in an organism has been systematically deleted, disrupted, or made controllable by the experimenter provides invaluable tools for exploring the role of each gene in the elaborate molecular collaboration that gives rise to life. Technologies such as RNA-seq and DNA microarrays can monitor the expression of tens of thousands of genes simultaneously, providing detailed, comprehensive snapshots of the dynamic patterns of gene expression that underlie complex cell processes.*

## MATHEMATICAL ANALYSIS OF CELL FUNCTIONS

Quantitative experiments combined with mathematical theory mark the beginning of modern science. Galileo, Kepler, Newton, and their contemporaries did more than set out some rules of mechanics and offer an explanation of the movements of the planets around the Sun: they showed how a quantitative mathematical approach could provide a depth and precision of understanding, at least for physical systems, that had never before been dreamed to be possible.

What is it that gives mathematics this almost magical power to explain the natural world, and why has mathematics played so much more important a part in physical sciences than in biology? What do biologists need to know about mathematics?

Mathematics can be viewed as a tool for deriving logical consequences from propositions. It differs from ordinary intuitive reasoning in its insistence on rigorous, accurate logic and the precise treatment of quantitative information. If the initial propositions are correct, then the deductions drawn from them by mathematics will be true. The surprising power of mathematics comes from the length of the chains of reasoning that rigorous logic and mathematical arguments make possible, and from the unexpectedness of the conclusions that can be reached, often revealing connections that one would not otherwise have guessed at. Reversing the argument, mathematics provides a way to test experimental hypotheses: if mathematical reasoning from a given hypothesis leads to a prediction that is not true, then the hypothesis is not true.

Clearly, mathematics is not much use unless we can frame our ideas—our initial hypotheses—about the given system in a precise, quantitative form. A mathematical edifice raised on a rickety or—even worse—a vague or overcomplicated set of propositions is likely to lead us astray. For mathematics to be useful, we must focus our analysis on simple subsystems in which we can pick out key quantitative parameters and frame well-defined hypotheses. This approach has been used with great success in physics for centuries, but it has been less common in biology. But times are changing, and more and more it is becoming possible for biologists to exploit the power of quantitative mathematical analysis.

In this final section of our methods chapter, we do not attempt to teach readers every way in which mathematics can be fruitfully applied to biological problems. Rather, we simply aim to give a sense of what mathematics and quantitative approaches can do for us in modern biology. We focus primarily on the important principles that mathematics teaches us about the dynamics of molecular interactions, and how mathematics can unveil surprising and useful features of complex systems containing feedback. We will illustrate these principles using the regulation of gene expression by transcription regulators like those discussed in Chapter 7. The same principles apply to the post-transcriptional regulatory systems that govern cell signaling (Chapter 15), cell-cycle control (Chapter 17), and essentially all cell processes.

### Regulatory Networks Depend on Molecular Interactions

Cell function and regulation depend on transient interactions among thousands of different macromolecules in the cell. We often summarize these interactions in this book with schematic cartoons. These diagrams are useful, but a complete picture requires a deeper, more *quantitative* level of understanding. To meaningfully

assess the biological impact of any interaction in the cell, we need to know in precise terms how the molecules interact, how they catalyze reactions, and, most importantly, how the behaviors of the molecules change over time. If a cartoon shows that protein A activates protein B, for example, we cannot judge the importance of this relationship without quantitative details about the concentrations, affinities, and kinetic behaviors of proteins A and B.

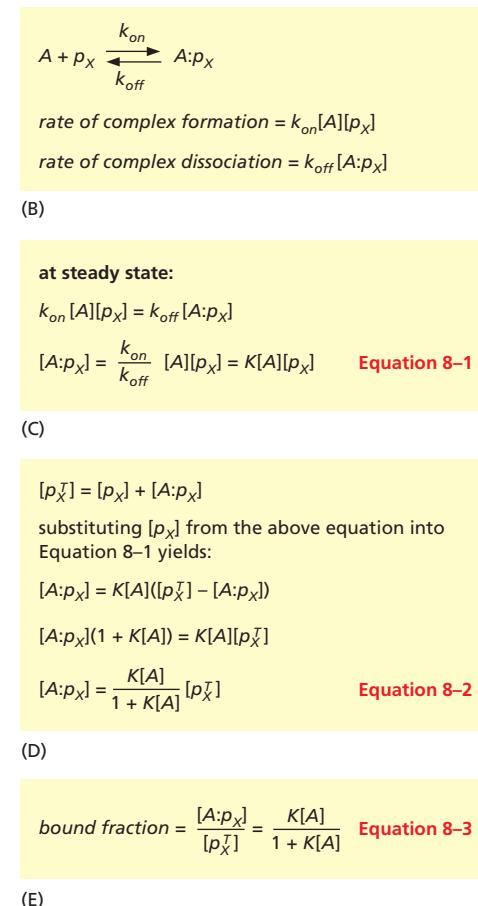
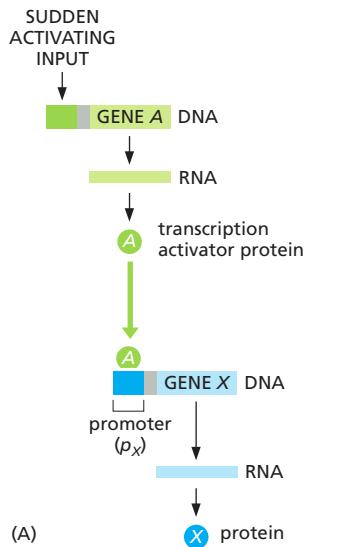
Let us begin by defining two different types of regulatory interaction in our cartoons: one designating inhibition and the other designating activation. If the protein product of gene X is a transcription repressor that inhibits the expression of gene Z, we depict the relationship as a *red bar-headed line* (→) drawn between genes X and Z (Figure 8–71). If the protein product of gene Y is a transcription activator that induces the expression of gene Z, then a *green arrow* (→) is drawn between genes Y and Z.

The regulation of one gene's expression by another is more complicated than a single arrow connecting them, and a complete understanding of this regulation requires that we tease apart the underlying biochemical processes. Figure 8–72A sketches some of the biochemical steps in the activation of gene expression by a transcription activator. A gene encoding the activator, designated as gene A, will produce its product, protein A, via an RNA intermediate. This protein A will then bind to  $p_X$ , the regulatory *promoter* of gene X, to form the complex  $A:p_X$ . Once the  $A:p_X$  complex forms, it stimulates the production of an RNA transcript that is subsequently translated to produce protein X.

We will focus here on the binding interaction that lies at the heart of this regulatory system: the interaction between protein A and the promoter  $p_X$ . Any molecule of protein A that is bound to  $p_X$  can also dissociate from it. The steps represented by the green activation arrow in Figure 8–72A include both the binding of A to  $p_X$  and the dissociation of the complex  $A:p_X$  to re-form A and  $p_X$ , as illustrated



**Figure 8–71** Diagrams that summarize biochemical relationships. Here, a simple cartoon indicates that gene X represses gene Z (left) whereas gene Y activates gene Z (right).



**Figure 8–72** A simple transcriptional interaction. (A) Genes A and X each produce a protein, with the product of gene A serving as a transcription activator to stimulate expression of gene X. As indicated by the green arrow, stimulation depends in part on the binding of protein A to the promoter region of gene X, designated as  $p_X$ . (B) The binding of protein A to the gene promoter is determined by the concentrations of the two binding partners (denoted as  $[A]$  and  $[p_X]$ , in units of mol/liter, or M), the association rate constant  $k_{on}$  (in units of sec<sup>-1</sup> M<sup>-1</sup>), and the dissociation rate constant  $k_{off}$  (in units of sec<sup>-1</sup>). (C) At steady state, the rates of association and dissociation are equal, and the concentration of the bound complex is determined by Equation 8–1, in which the two rate constants are combined in the equilibrium constant K. (D) Equation 8–2 can be derived to calculate the steady-state concentration of bound complex at a known total concentration of the promoter  $[p_X^T]$ . (E) Rearrangement of Equation 8–2 yields Equation 8–3, which allows calculation of the fraction of promoter  $p_X$  that is occupied by protein A.

by the notation in Figure 8–72B. This reaction notation is more informative than the diagrams in our figures, but has its own limitations. Suppose that the concentration of  $A$  increases by a factor of ten as a response to an environmental input. If  $A$  increases, we intuitively know that  $A:p_X$  should increase too, but we cannot determine the amount of the increase without additional information. We need to know the affinity of the binding interaction and the concentrations of the components. With this information in hand, we can rigorously derive the answer.

As discussed earlier and in Chapter 3 (see Figure 3–44), we know that the formation of a complex between two binding partners, such as  $A$  and  $p_X$ , depends on a rate constant  $k_{\text{on}}$ , which describes how many productive collisions occur per unit time per protein at a given concentration of  $p_X$ . The rate of complex formation equals the product of this rate constant  $k_{\text{on}}$  and the concentrations of  $A$  and  $p_X$  (see Figure 8–72B). Complex dissociation occurs at a rate  $k_{\text{off}}$  multiplied by the concentration of the complex. The rate constant  $k_{\text{off}}$  can differ by orders of magnitude for different DNA sequences because it depends on the strength of the non-covalent bonds formed between  $A$  and  $p_X$ .

We are primarily interested in understanding the amount of bound promoter complex at equilibrium or *steady state*, where the rate of complex formation equals the rate of complex dissociation. Under these conditions, the concentration of the promoter complex is specified by a simple equation that combines the two rate constants into a single equilibrium constant  $K = k_{\text{on}}/k_{\text{off}}$  (Equation 8–1; Figure 8–72C).  $K$  is sometimes called the association constant,  $K_a$ . The larger this constant  $K$ , the stronger the interaction between  $A$  and  $p_X$  (see Figure 3–44). The reciprocal of  $K$  is the dissociation constant,  $K_d$ .

To calculate the steady-state concentration of promoter complex using Equation 8–1, we need to account for another complication: both  $A$  and  $p_X$  exist in two forms—free in solution and bound to each other. In most cases, we know the total concentration of  $p_X$  and not the free or bound concentrations, so we must find a way to use the total concentration in our calculations. To do this, we first specify that the total concentration of  $p_X$  ( $[p_X^T]$ ) is the sum of the concentrations of free ( $[p_X]$ ) and bound ( $[A:p_X]$ ) forms (Figure 8–72D). This leads to a new equation that allows us to use  $[p_X^T]$  to calculate the steady-state concentration of the promoter complex ( $[A:p_X]$ ) (Equation 8–2, Figure 8–72D).

Protein  $A$  also exists in two forms: free ( $[A]$ ) and bound to  $p_X$  ( $[A:p_X]$ ). In a cell, there are typically one or two copies of  $p_X$  (assuming there is only one gene  $X$  per haploid genome) and multiple copies of  $A$ . As a result, we can safely assume that from the viewpoint of  $A$ ,  $[A:p_X]$  is negligible relative to the total  $[A^T]$ . This means that  $[A] \approx [A^T]$ , and we can just plug in the values of total  $[A^T]$  in Equation 8–2 without incurring appreciable error in the calculation of  $[A:p_X]$ .

Now, we are ready to determine the effects of increasing the concentration of  $A$ . Suppose that  $K = 10^8 \text{ M}^{-1}$ , which is a typical value for many such interactions. The starting concentration of  $A$  is  $[A^T] = 10^{-9} \text{ M}$ , and  $[p_X^T] = 10^{-10} \text{ M}$  (assuming there is one copy of gene  $X$  in a haploid yeast cell, for example, with a volume of around  $2 \times 10^{-14} \text{ L}$ ). Using Equation 8–2, we find that a tenfold increase in the concentration of  $A$  causes the amount of promoter complex  $[A:p_X]$  to increase 5.5-fold, from  $0.09 \times 10^{-10} \text{ M}$  to  $0.5 \times 10^{-10} \text{ M}$  at steady state. The effects of a tenfold increase in the concentration of  $A$  will vary dramatically depending on its starting concentration relative to the equilibrium constant. Only through this mathematical approach can we achieve a thorough understanding of what these effects will be and what impact they will have on the biological response.

To assess the biological impact of a change in transcription activator levels, it is also important in many cases to determine the fraction of the target gene promoter that is bound by the activator, since this number will be directly proportional to the activity of the gene's promoter. In our case, we can calculate the fraction of the gene  $X$  promoter,  $p_X$ , that has protein  $A$  bound to it by rearranging Equation 8–2 (Equation 8–3, Figure 8–72E). This fraction can be viewed as the probability that promoter  $p_X$  is occupied, averaged over time. It is also equal to the average occupancy across a large population of cells at any instant in time. When there is no protein  $A$  present,  $p_X$  is always free, the bound fraction is zero, and transcription is

off. When  $[A] = 1/K$ , the promoter  $p_X$  has a 50% chance of being occupied. When  $[A]$  greatly exceeds  $1/K$ , the bound fraction is almost equal to one, meaning that  $p_X$  is fully occupied and transcription is maximal.

### Differential Equations Help Us Predict Transient Behavior

The most important and basic insights for which we, as biologists, depend on mathematics concern the behavior of regulatory systems over time. This is the central theme of dynamics, and it was for the solution of problems in dynamics that the techniques of calculus were developed, by Newton and Leibniz, in the seventeenth century. Briefly, the general problem is this: if we are given the rates of change of a set of variables that characterize the system at any instant, how can we compute its future state? The problem becomes especially interesting, and the predictions often remarkable, when the rates of change themselves depend on the values of the state variables, as in systems with feedback.

Let us return to Equation 8-2 (Figure 8-72D), which tells us that when  $[A]$  changes,  $[A:p_X]$  at steady state will also change to a new concentration that we can calculate with precision. However,  $[A:p_X]$  does not change instantaneously to this value. If we hope to understand the behavior of this system in detail, we must also ask how long it takes  $[A:p_X]$  to get to its new steady-state value inside the cell. Equation 8-2 cannot answer this question. We need calculus.

The most common strategy for solving this problem is to use ordinary differential equations. The equations that describe biochemical reactions have a simple premise: the rate of change in the concentration of any molecular species  $X$  (that is,  $d[X]/dt$ ) is given by the balance of the rate of its appearance with that of its disappearance. For our example, the rate of change in the concentration of the bound promoter complex,  $[A:p_X]$ , is determined by the rates of complex assembly and disassembly. We can incorporate these rates into the differential equation shown in **Figure 8-73A** (Equation 8-4). When  $[A]$  changes, Equation 8-4 can be solved to generate the concentration of  $[A:p_X]$  as a function of time. Notice that when  $k_{\text{on}}[A][p_X] = k_{\text{off}}[A:p_X]$ , then  $d[A:p_X]/dt = 0$  and  $[A:p_X]$  stops changing. At this point, the system has reached the steady state.

Calculation of all  $[A:p_X]$  values as a function of time, using Equation 8-4, allows us to determine the rate at which  $[A:p_X]$  reaches its steady-state value. Because this value is attained asymptotically, it is often most useful to compare the times needed to get to 50, 90, or 99 percent of this new steady state. The simplest way to determine these values is to solve Equation 8-4 with a method called numerical integration, which involves plugging in values for all of the parameters ( $k_{\text{on}}$ ,  $k_{\text{off}}$ , etc.) and then using a computer to determine the values of  $[A:p_X]$  over time, starting from given initial concentrations of  $[A]$  and  $[p_X]$ . For  $k_{\text{on}} = 0.5 \times 10^7 \text{ sec}^{-1} \text{ M}^{-1}$ ,  $k_{\text{off}} = 0.5 \times 10^{-1} \text{ sec}^{-1}$  ( $K = 10^8 \text{ M}^{-1}$  as above), and  $[p_X^T] = 10^{-10} \text{ M}$ , it takes  $[A:p_X]$  about 5, 20, and 40 seconds to reach 50, 90, and 99 percent of the new steady-state value following a sudden tenfold change in  $[A]$  (Figure 8-73B). Thus, a sudden jump in  $[A]$  does not have instantaneous effects, as we might have assumed from looking at the cartoon in Figure 8-72A.

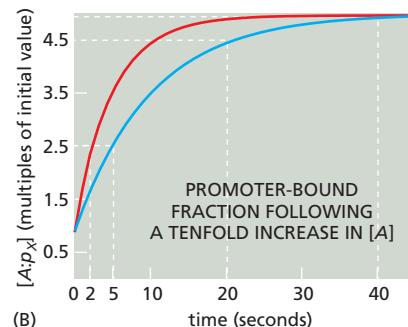
Differential equations therefore allow us to understand the transient dynamics of biochemical reactions. This tool is critical for achieving a deep understanding of cell behavior, in part because it allows us to determine the dependence of the dynamics inside cells on parameters that are specific to the particular molecules involved. For example, if we double the values of both  $k_{\text{on}}$  and  $k_{\text{off}}$ , then Equation 8-1 (Figure 8-72C) indicates that the steady-state value of  $[A:p_X]$  does not change. However, the time it takes to reach 50% of this steady state after a ten-fold

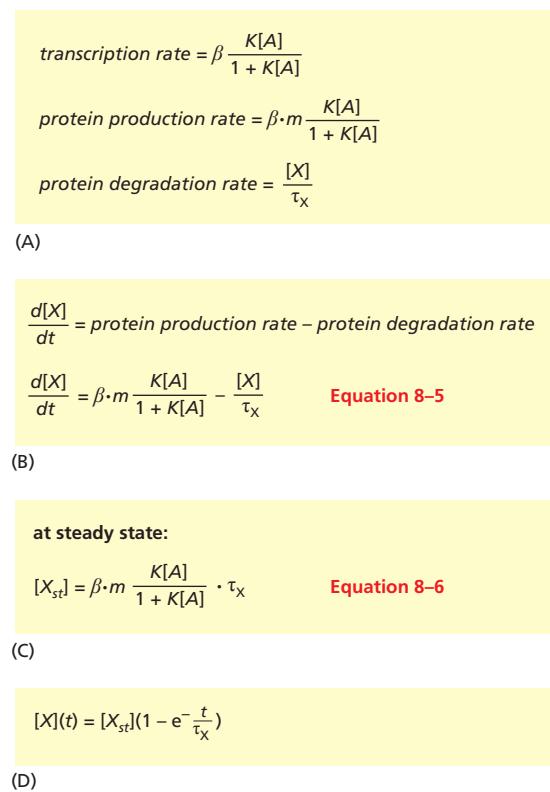
$$\frac{d[A:p_X]}{dt} = \text{rate of complex formation} - \text{rate of complex dissociation}$$

$$\frac{d[A:p_X]}{dt} = k_{\text{on}}[A][p_X] - k_{\text{off}}[A:p_X] \quad \text{Equation 8-4}$$

(A)

**Figure 8-73** Using differential equations to study the dynamics and steady-state behavior of a biological system. (A) Equation 8-4 is an ordinary differential equation for calculating the rate of change in the formation of bound promoter complex in response to a change in other components. (B) Formation of  $[A:p_X]$  after a tenfold increase in  $[A]$ , as determined by solving Equation 8-4. In blue is the solution corresponding to  $k_{\text{on}} = 0.5 \times 10^7 \text{ sec}^{-1} \text{ M}^{-1}$  and  $k_{\text{off}} = 0.5 \times 10^{-1} \text{ sec}^{-1}$ . In this case, it takes  $[A:p_X]$  about 5, 20, and 40 seconds to reach 50, 90, and 99 percent of the new steady-state value. For the red curve, the  $k_{\text{on}}$  and  $k_{\text{off}}$  values are doubled, and the system reaches the same steady state more rapidly.





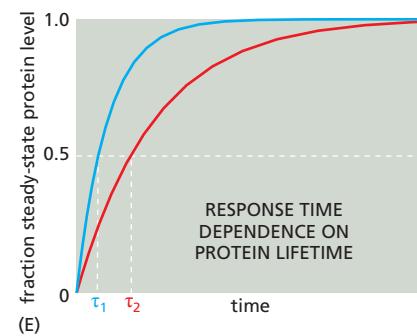
change in  $[A]$  in our example changes from about 5 seconds to 2 seconds (see Figure 8–73B). These insights are not accessible from either cartoons or equilibrium equations. This is an unusually simple example; mathematical descriptions such as differential equations become more indispensable for understanding biological interactions as the number of interactions increases.

### Both Promoter Activity and Protein Degradation Affect the Rate of Change of Protein Concentration

To understand our gene regulatory system further, we also need to describe the dynamics of protein X production in response to changes in the amount of transcription activator protein A. Here again, we use an ordinary differential equation for the rate of change of protein X concentration—determined by the balance of the rate of production of protein X through expression of gene X and the protein's rate of degradation.

Let us begin with the rate of protein X production, which is determined primarily by the occupancy of the promoter of gene X by protein A. The binding and dissociation of a transcription regulator at a promoter generally occurs on a much faster time scale than transcription initiation, causing many binding and unbinding events to occur before transcription proceeds. As a result, we can assume that the binding reaction is at equilibrium on the time scale of transcription, and we can calculate promoter occupancy by protein A using the equilibrium equation discussed earlier (Equation 8–3, Figure 8–72E). To determine transcription rate, we simply multiply the occupied promoter fraction by a *transcription rate constant*,  $\beta$ , that represents the binding of RNA polymerase and the subsequent steps that lead to production of mRNA and protein (Figure 8–74A). If each mRNA molecule produces, on average,  $m$  molecules of protein product, then we can determine protein production rate by multiplying the transcription rate by  $m$  (Figure 8–74A).

Now let us consider the factors that influence protein X degradation and its dilution due to cell growth. Degradation generally results in an exponential decline in protein levels, and the average time required for a specific protein to be



**Figure 8–74** Effect of protein lifetime on the timing of the response.

(A) Equations for calculation of the rates of gene X transcription, protein X production, and protein X degradation, as explained in the text. (B) Equation 8–5 is an ordinary differential equation for calculating the rate of change in protein X in response to changes in other components. (C) When the rate of change in protein X is zero (steady state), its concentration can be calculated with Equation 8–6, revealing a direct relationship with protein lifetime ( $\tau$ ). (D) The solution of Equation 8–5 specifies the concentration of protein X over time as it approaches its steady-state concentration. (E) Response time depends on protein lifetime. As described in the text, the time that it takes a protein to reach a new steady state is greater when the protein is more stable. Here, the blue line corresponds to a protein with a lifetime that is 2.5-fold shorter than the lifetime of the protein in red.

degraded is defined as its mean lifetime,  $\tau$ . In our current example, the rate of degradation of protein  $X$  depends on its mean lifetime  $\tau_X$ , which takes into account active degradation as well as its dilution as the cell grows. The degradation rate depends on the concentration of protein  $X$  and is calculated by dividing this concentration by the lifetime (Figure 8–74A).

With equations for rates of production and degradation in hand, we can now generate a differential equation to determine the rate of change of protein  $X$  as a function of time (Equation 8–5, Figure 8–74B). This equation can be solved by the numerical methods mentioned earlier. According to the solution of this equation, when transcription begins, the concentration of protein  $X$  rises to a steady-state level at which the concentration of  $X$  is not changing anymore; that is, its rate of change is zero. When this occurs, rearrangement of Equation 8–5 yields an equation that can be used to determine the steady-state value of  $X$ ,  $[X_{st}]$  (Equation 8–6, Figure 8–74C). An important concept emerges from the mathematics: the steady-state concentration of a gene product is directly proportional to its lifetime. If lifetime doubles, protein concentration doubles as well.

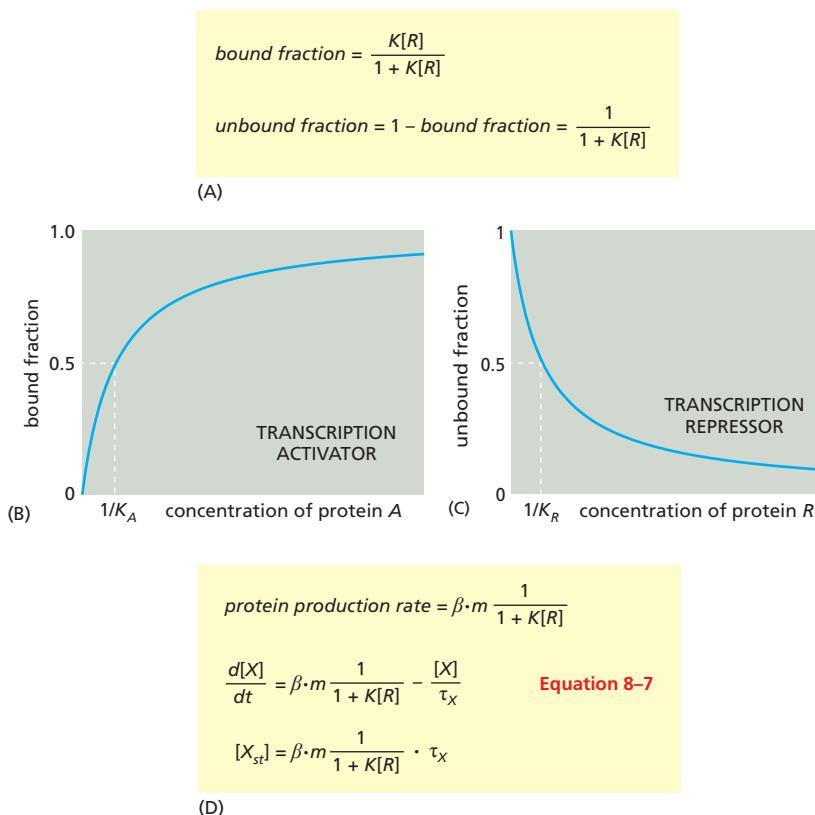
### The Time Required to Reach Steady State Depends on Protein Lifetime

We can see from Equation 8–6 (see Figure 8–74C) that when the concentration of protein  $A$  rises, protein  $X$  increases to a new steady-state value,  $[X_{st}]$ . But this cannot happen instantaneously. Instead,  $X$  changes dynamically according to the solution of its differential rate equation (Equation 8–5). The solution of this equation reveals that the concentration of  $X$  over time is related to its steady-state concentration according to the equation in Figure 8–74D. Once again, mathematics uncovers a simple but important concept that is not intuitively obvious: following a sudden increase in  $[A]$ ,  $[X]$  rises to a new steady state at an exponential rate that is inversely related to its lifetime; the faster  $X$  is degraded, the less time it takes it to reach its new steady-state value (Figure 8–74E). The faster response time comes at a higher metabolic cost, however, since proteins with a rapid response time must be produced and degraded at a high rate. For proteins that are not rapidly turned over, the response time is very long, and protein concentration is determined primarily by the dilution that results from cell growth and division.

### Quantitative Methods Are Similar for Transcription Repressors and Activators

Positive control is not the only mechanism that cells use to regulate the expression of their genes. As we discussed in Chapter 7, cells also actively shut off genes, often by employing transcription repressor proteins that bind to specific sites on target genes, thereby blocking access to RNA polymerase. We can analyze the function of these repressors by the same quantitative methods described above for transcription activators. If a repressor protein  $R$  binds to the regulatory region of gene  $X$  and represses its transcription, then the fraction of gene binding sites occupied by the repressor is specified by the same equation we used earlier for the transcription activator (Figure 8–75A). In this case, however, it is only when the DNA is free that RNA polymerase can bind to the promoter and transcribe the gene. Thus, the quantity of interest is the unbound fraction, which can be viewed as the probability that the site is free, averaged over multiple binding and unbinding events. When the repressor concentration is zero, the unbound fraction is 1 and the promoter is fully active; when the repressor concentration greatly exceeds  $1/K$ , the unbound fraction approaches zero. Figures 8–75B and C compare these relationships for a transcription activator and a transcription repressor.

We can create a differential equation that provides the rate of change in protein  $X$  when repressor concentrations change (Equation 8–7, Figure 8–75D). As in the case of the transcription activator, the steady-state concentration of protein  $X$  increases as its lifetime increases, but it decreases as the concentration of the transcription repressor increases.



**Figure 8–75** How promoter occupancy depends on the binding affinity of a transcription regulator protein. (A) The fraction of a binding site that is occupied by a transcription repressor  $R$  is determined by an equation that is similar to the one we used for a transcription activator (see Figure 8–72E), except that in the case of a repressor we are interested primarily in the unbound fraction. (B) For a transcription activator  $A$ , half of the promoters are occupied when  $[A] = 1/K_A$ . Gene activity is proportional to this bound fraction. (C) For a transcription repressor  $R$ , gene activity is proportional to the unbound fraction of promoters. As indicated, this fraction is reduced to half of its maximal value when  $[R] = 1/K_R$ . (D) As in the case of the transcription activator  $A$  (see Figure 8–74), we can derive equations to assess the timing of protein  $X$  production as a function of repressor concentrations.

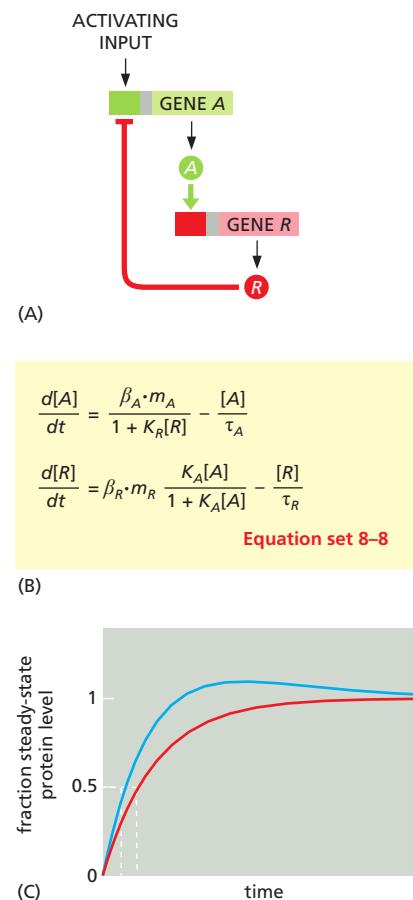
## Negative Feedback Is a Powerful Strategy in Cell Regulation

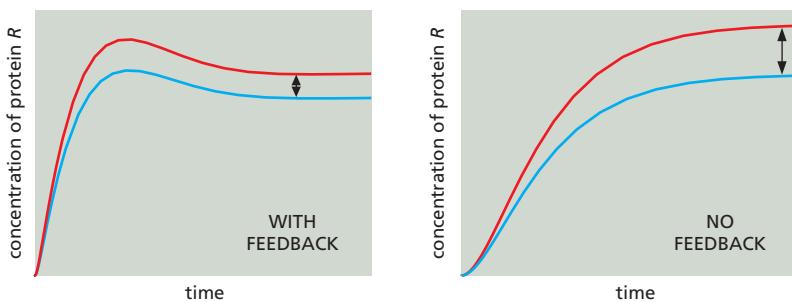
Thus far, we have considered simple regulatory systems of just a few components. In most of the complex regulatory systems that govern cell behaviors, multiple modules are linked to produce larger circuits that we call *network motifs*, which can produce surprisingly complex and biologically useful responses whose properties become apparent only through mathematical analysis. A particularly common and important network motif is the negative feedback loop, which can have dramatically different functions depending on how it is structured.

We take as a first example a network motif consisting of two linked modules (Figure 8–76A). Here, an input signal initiates the transcription of gene  $A$ , which produces a transcription activator protein  $A$ . This activates gene  $R$ , which synthesizes a transcription repressor protein  $R$ . Protein  $R$  in turn binds to the promoter of gene  $A$  to inhibit its expression. This cyclical organization creates a negative feedback loop that one can intuitively understand as a mechanism to prevent proteins from accumulating to high levels. But what can we learn about negative feedback loops, and their value in biology, by using mathematics to model them?

The negative feedback loop in Figure 8–76A can be modeled using Equation 8–7 (see Figure 8–75D) for the repression of gene  $A$  and Equation 8–5 (see Figure 8–74B) for the activation of gene  $R$ . Thus, for proteins  $A$  and  $R$ , we use the set of differential equations (Equation set 8–8) shown in Figure 8–76B. The two equations in this set are coupled, which means that they must be solved together to describe

**Figure 8–76** A simple negative feedback motif. (A) Gene  $A$  negatively regulates its own expression by activating gene  $R$ . The product of gene  $R$  is a transcription repressor that inhibits gene  $A$ . (B) Equation set 8–8 can be solved to determine the dynamics of system components over time. (C) A system with negative feedback (blue) reaches its steady state faster than a system with no feedback (red). The plots indicate the levels of protein  $A$ , expressed as a fraction of the steady-state level. The blue line reflects the solution of Equation set 8–8, which includes negative feedback of gene  $A$  by the repressor  $R$ . The red line represents the solution when the rate of synthesis of  $A$  was set to a constant value that is unaffected by the repressor  $R$ .





**Figure 8–77** The effect of fluctuations in kinetic rate constants on a system with negative feedback compared to one without feedback. The plot at left represents the levels of protein  $R$  after a sudden activating stimulus, according to the regulatory scheme in Figure 8–76A and determined by the solution of Equation set 8–8 (see Figure 8–76B). A perturbation was induced by changing  $\beta_A$  from 4 M/min (red line) to 3 M/min (blue line). The plot at right shows the results when negative feedback was removed. The system with negative feedback deviates less from its normal operation as  $\beta$  changes than does the system with no feedback. Notice that, as in Figure 8–76C, the system with negative feedback also reaches its steady state more rapidly.

the behavior of  $A$  and  $R$  over time for any value of the input. As before, we plug in values for the parameters ( $\beta_R$ ,  $\tau_R$ , etc.) and then use a computer to determine the values of  $[A]$  and  $[R]$  as a function of time after a sudden input activates gene  $A$ .

The results reveal several important properties of negative feedback. First, rather surprisingly, negative feedback increases the speed of the response to the activating input. As shown in Figure 8–76C, the system with negative feedback reaches its new steady state faster than the system with no feedback.

Second, negative feedback is useful for protecting cells from perturbations that continuously arise in the cell's internal environment—due either to random variations in the birth and death of molecules or to fluctuations in environmental variables such as temperature and nutritional supplies. Let us imagine, for example, that  $\beta_A$ , the transcription rate constant for gene  $A$ , fluctuates by 25% of its value and ask whether and how much the levels of protein  $R$  are affected. The results, shown in Figure 8–77, reveal that a change in  $\beta_A$  causes a smaller change in the steady-state value of  $R$  when the network has negative feedback.

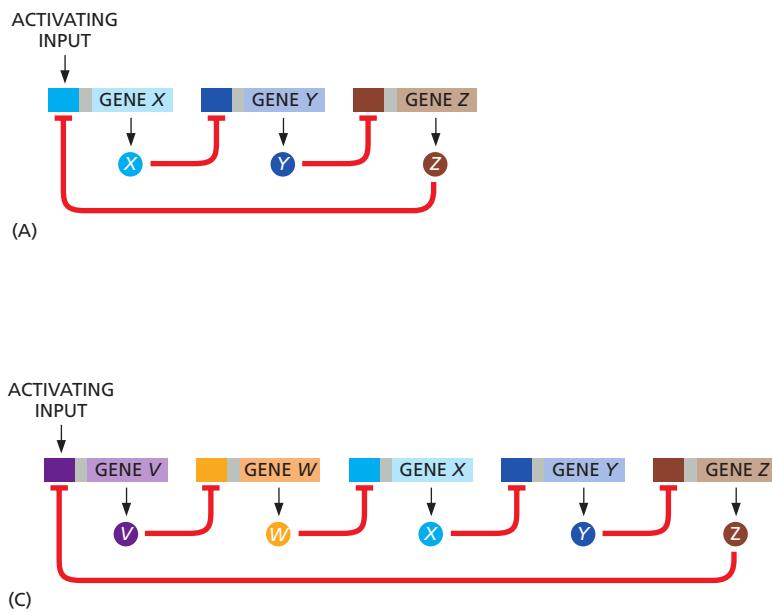
### Delayed Negative Feedback Can Induce Oscillations

A beautiful thing happens when a negative feedback loop contains some delay mechanism that slows the feedback signal through the loop: rather than generating a new stable state as in a rapid negative feedback loop, a delayed loop generates pulses, or *oscillations*, in the levels of its components. This can be seen, for example, if the number of components in a negative feedback loop increases, which leads to delays in the amount of time required for the cycle of signals to be completed. Figure 8–78 compares the behavior of two network motifs—one with a three-stage and one with a five-stage negative feedback loop. Using the same kinetic parameters at each stage in the two loops, one finds that stable oscillations arise in the longer loop, while in the shorter loop the same parameters lead to relatively rapid convergence to a stable steady state.

Changes in the parameters of a delayed negative feedback loop—binding affinities, transcription rates, or protein stabilities, for example—can change the amplitude and period of the oscillations, providing a remarkably versatile mechanism for generating all sorts of oscillators that can be used for various purposes in the cell. Indeed, many naturally occurring oscillators, including the calcium oscillators described in Chapter 15 and the cell-cycle network described in Chapter 17, use delayed negative feedback as the basis for biologically important oscillations. Not all of the oscillations observed in cells are thought to have a function, however. Oscillations become inevitable in a highly complex, multicompartment biochemical pathway like glycolysis, due simply to the large number of feedback loops that appear to be required for its regulation.

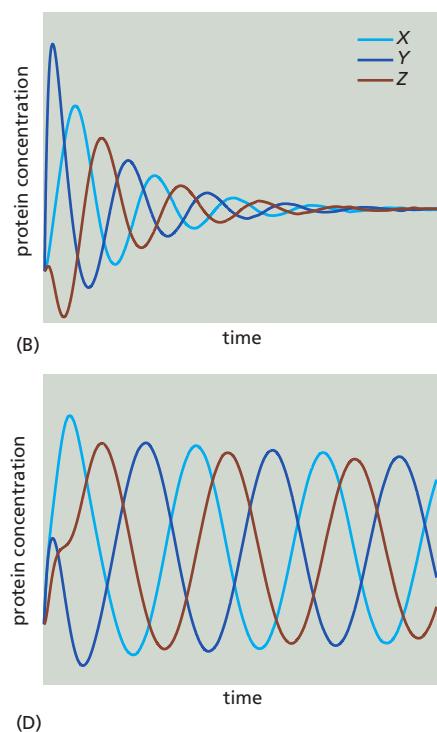
### DNA Binding By a Repressor or an Activator Can Be Cooperative

We have focused thus far on the binding of a single transcription regulator to a single site in a gene promoter. Many promoters, however, contain multiple adjacent binding sites for the same transcription regulator, and it is not uncommon for these regulators to interact with each other on the DNA to form dimers or larger oligomers. These interactions can result in a *cooperative* form of DNA binding,



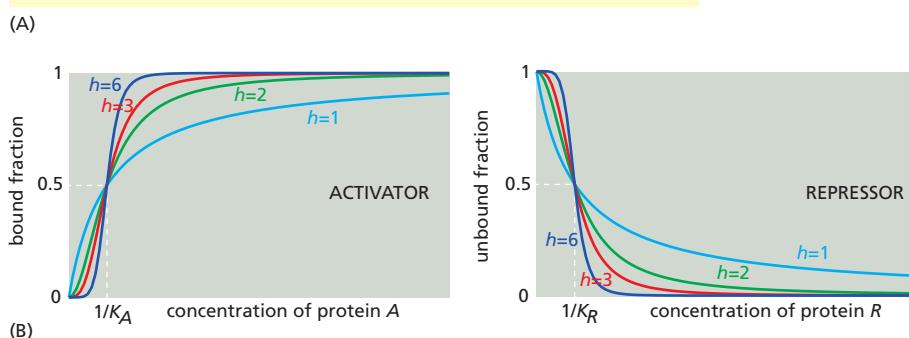
such that DNA-binding affinity increases at higher concentrations of the transcription regulator. Cooperativity produces a steeper transcriptional response to increasing regulator concentration than the response that can be generated by the binding of a monomeric protein to a single site. A steep transcriptional response of this sort, when present in conjunction with positive feedback, is an important ingredient for producing systems with the ability to switch between different discrete phenotypic states. To begin to understand how this occurs, we need to modify our equations to include cooperativity.

Cooperative binding events can produce steep S-shaped (or *sigmoidal*) relationships between the concentration of regulatory protein and the amount bound on the DNA (see Figure 15–16). In this case, a number called the *Hill coefficient* (*h*) describes the degree of cooperativity, and we can include this coefficient in our equations for calculating the bound fraction of promoter (Figure 8–79A). As the Hill coefficient increases, the dependence of binding on protein concentration becomes steeper (Figure 8–79B). In principle, the Hill coefficient is similar to the number of molecules that must come together to generate a reaction. In practice, however, cooperativity is rarely complete, and the Hill coefficient does not reach this number.



**Figure 8–78** Oscillations arising from delayed negative feedback. A transcriptional circuit with three components (A, B) is less likely to oscillate than a transcriptional circuit with five components (C, D). The *X* (light blue), *Y* (dark blue), and *Z* (brown) here represent transcription regulatory proteins. For the simulations in (B) and (D), the system was initiated from random initial conditions for *X*, *Y*, and *Z*. Oscillations are produced by a delay induced as the signal propagates through the loop.

$$\text{bound fraction} = \frac{(K_A[A])^h}{1 + (K_A[A])^h} \text{ for activators, or } \frac{(K_R[R])^h}{1 + (K_R[R])^h} \text{ for repressors}$$



**Figure 8–79** How the cooperative binding of transcription regulatory proteins affects the fraction of promoters bound. (A) Cooperativity is incorporated into our mathematical models by including a Hill coefficient (*h*) in the equations used previously to determine the fraction of bound promoter (see Figures 8–72E and 8–75A). When *h* is 1, the equations shown here become identical to the equations used previously, and there is no cooperativity. (B) The left panel depicts a cooperatively bound transcription activator and the right panel depicts a cooperatively bound transcription repressor. Recall from Figure 8–75B that gene activity is proportional to bound activator (left panel) or unbound repressor (right panel). Note that the plots get steeper as the Hill coefficient increases.

## Positive Feedback Is Important for Switchlike Responses and Bistability

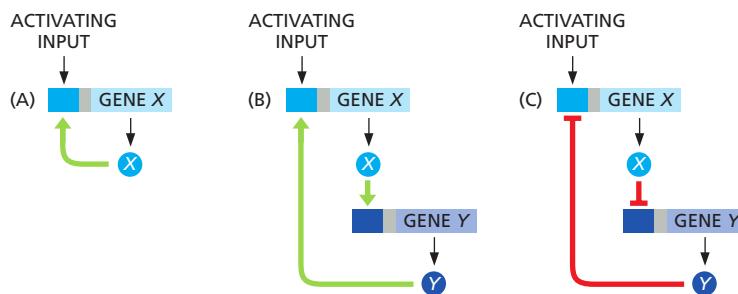
We turn now to positive feedback and its very important consequences. First and foremost, positive feedback can make a system *bistable*, enabling it to persist in either of two (or more) alternative steady states. The idea is simple and can be conveyed by drawing an analogy with a candle, which can exist either in a burning state or in an unlit state. The burning state is maintained by positive feedback: the heat generated by burning keeps the flame alight. The unlit state is maintained by the absence of this feedback signal: so long as sufficient heat has never been applied, the candle will stay unlit.

For the biological system, as for the candle, bistability has an important corollary: it means that the system has a memory, such that its present state depends on its history. If we start with the system in an Off state and gradually rack up the concentration of the activator protein, there will come a point where autostimulation becomes self-sustaining (the candle lights), and the system moves rapidly to an On state. If we now intervene to decrease the level of activator, there will come a point where the same thing happens in reverse, and the system moves rapidly back to an Off state. But the transition points for switching on and switching off are different, and so the current state of the system depends on the route by which it has been taken in the past—a phenomenon called *hysteresis*.

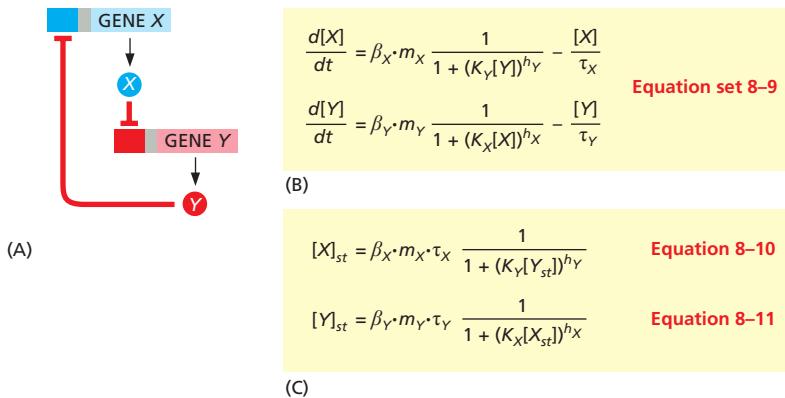
A simple case of positive feedback can be seen in a regulatory system in which a transcription regulator activates (directly or indirectly) its own expression, as in **Figure 8–80A**. Positive feedback can also arise in a circuit with many intervening repressors or activators, so long as the net overall effect of the interactions is activation (Figure 8–80B and C).

To illustrate how positive feedback can generate stable states, let us focus on a simple positive feedback loop containing two repressors, *X* and *Y*, each of which inhibits expression of the other (**Figure 8–81A**). As we saw with Equation set 8–8 (Figure 8–76B) earlier, we can create differential equations describing the rate of change of  $[X]$  and  $[Y]$  (Equation set 8–9, Figure 8–81B). We can further modify these equations to include cooperativity by adding Hill coefficients. As we did earlier, we can then create equations for calculating the concentrations of  $[X]$  and  $[Y]$  when the system reaches a steady state (that is, when  $(d[X]/dt) = 0$  and  $(d[Y]/dt) = 0$ ; Equations 8–10 and 8–11, Figure 8–81C).

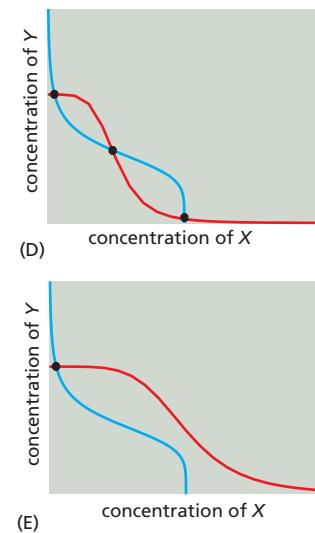
Equations 8–10 and 8–11 can be used to carry out an intriguing mathematical procedure called a *nullcline* analysis. These equations define the relationships between the concentration of *X* at steady state,  $[X_{st}]$ , and the concentration of *Y* at steady state,  $[Y_{st}]$ , which must be simultaneously satisfied. We can plug in different values for  $[Y_{st}]$  in Equation 8–10, and calculate the corresponding  $[X_{st}]$  for each of these values. We can then graph  $[X_{st}]$  as a function of  $[Y_{st}]$ . Next, we repeat the process by varying  $[X_{st}]$  in Equation 8–11 to graph the resulting  $[Y_{st}]$ . The intersections of these two graphs determine the theoretically possible steady states of the system. For systems in which the Hill coefficients  $h_X$  and  $h_Y$  are much larger than 1, the lines in the two graphs intersect at three locations (Figure 8–81D). In other systems that have the same arrangement of regulators but different parameters, there might only be one intersection, indicating the presence of only a single



**Figure 8–80 Positive feedback of a gene onto itself through serially connected interactions.** A sequence of activators and repressors of any length can be connected to produce a positive feedback loop, as long as the overall sign is positive. Because the negative of a negative is positive, not only circuit (A) and (B) but also circuit (C) create positive feedback.



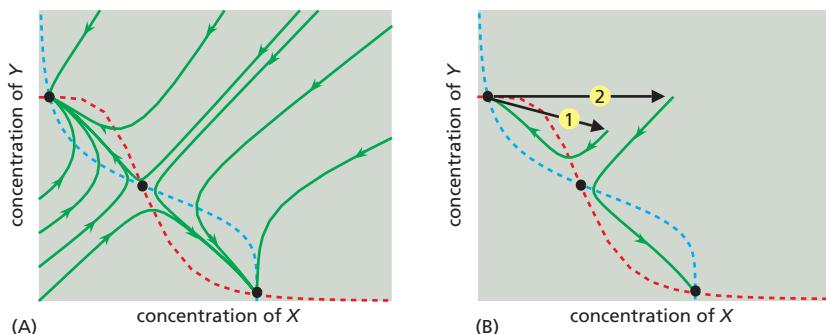
**Figure 8–81** A graphical nullcline analysis. (A)  $X$  inhibits  $Y$  and  $Y$  inhibits  $X$ , resulting in a positive feedback loop. (B) Equation set 8–9 can be used to determine the rate of change in the concentrations of proteins  $X$  and  $Y$ . (C) Equations 8–10 and 8–11 provide the concentrations of proteins  $X$  and  $Y$ , respectively, when these concentrations reach a steady state. (D, E) Blue curves (called nullclines) are plots of  $[X]_{st}$  calculated from Equation 8–10 over a range of concentrations of  $[Y]_{st}$ . Red curves indicate values of  $[Y]_{st}$  calculated from Equation 8–11 over a range of concentrations of  $[X]_{st}$ . At an intersection of the two lines, both  $[X]$  and  $[Y]$  are at steady state. For plot (D), the binding of both proteins to their target gene promoters was cooperative ( $h_X$  and  $h_Y$  much larger than 1), resulting in the presence of multiple intersections of the nullclines—suggesting that the system can assume multiple discrete steady states. In plot (E), the binding of protein  $X$  to the promoter of gene  $Y$  was not cooperative ( $h_X$  close to 1), resulting in only one nullcline intersection and thus just one likely steady state.



steady state. For example, when there is a low cooperativity of protein  $X$  binding to the promoter of gene  $Y$  (that is, a small Hill coefficient,  $h_X$ , in Equation 8–11), the plot of  $[Y]$  is less curved (Figure 8–81E), and it is less likely that there will be multiple intersections of the two curves.

We emphasized earlier that positive feedback typically generates a bistable system with two stable steady states. Why does the system modeled in Figure 8–81D have three? This conundrum can be explained by solving the reaction rate equations (Equation set 8–9, Figure 8–81B) for various different starting conditions of  $[X]$  and  $[Y]$ , determining all values of  $[X]$  and  $[Y]$  as a function of time. Starting with each set of initial concentrations of  $[X]$  and  $[Y]$ , these calculations produce a so-called *trajectory* of points, each indicated by a curved green line on Figure 8–82A. A fascinating pattern emerges: each trajectory moves across the plot and settles in one of two steady states, but never in the third (middle steady state). We conclude that the middle steady state is *unstable* because it cannot “attract” any trajectories. The system therefore has only two *stable* steady states. Thus, the number of stable steady states in a system need not be equal to the total number of its theoretically possible steady states. In fact, stable steady states are usually separated by unstable ones, as in our example.

Once this system adopts a fate by settling in one of the two steady states, does it have the ability to switch to the other state? The numerical solution of Equation



**Figure 8–82** Analysis of the stability of a system’s steady states. (A) The dotted lines are the nullclines for the system shown in Figure 8–81. Also shown are dynamic trajectories (green) that show the changes over time in  $[X]$  and  $[Y]$ , starting at a variety of different initial concentrations (determined by solution of Equation set 8–9; see Figure 8–81B). By plotting  $[X]$  versus  $[Y]$  at each time point, we find that, although there are three possible steady states in this system, the dynamic trajectories converge on only two of them. The middle steady state is avoided: it is unstable, being unable to attract any trajectories. (B) Imagine that the system is at the upper-left steady state and experiences a perturbation (black arrows), such as a random fluctuation in the production rates of  $X$  and/or  $Y$ . If the perturbation is small (arrow 1), the system will return to the same steady state. On the other hand, a perturbation that drives the system beyond the unstable (middle) steady state (arrow 2) causes it to switch to the lower-right steady state. The set of perturbations that a system can withstand without switching from one steady state to the other is known as the region of attraction of that steady state.

set 8–9 can again provide an answer. In Figure 8–82B, we show the solution of this equation set for two perturbations from the upper-left steady state. For a small perturbation, the system returns to its original steady state. But the larger perturbation causes the system to switch to the alternate steady state. Thus, this system can be switched from one stable steady state to the other by subjecting it to an input (or a perturbation) that is large enough to make the other steady state more attractive. More generally, every stable steady state has a corresponding *region of attraction*, which can be intuitively thought of as the range of perturbations (of  $[X]$  or  $[Y]$  in this example) for which the dynamic trajectories converge back to that particular steady state, rather than switch to the other one.

The concept of a region of attraction has interesting implications for the heritability of transcriptional states and the transition rates between them. If the region of attraction around one steady state is large, for example, then most cells in the population will assume this particular state. Furthermore, this state is likely to be inherited by daughter cells, since minor perturbations, like those ensuing from an asymmetric distribution of molecules during cell division, will rarely be sufficient to induce switching to the other steady state. We should expect that the use of positive feedback, coupled to cooperativity, will quite often be associated with systems requiring stable cell memory.

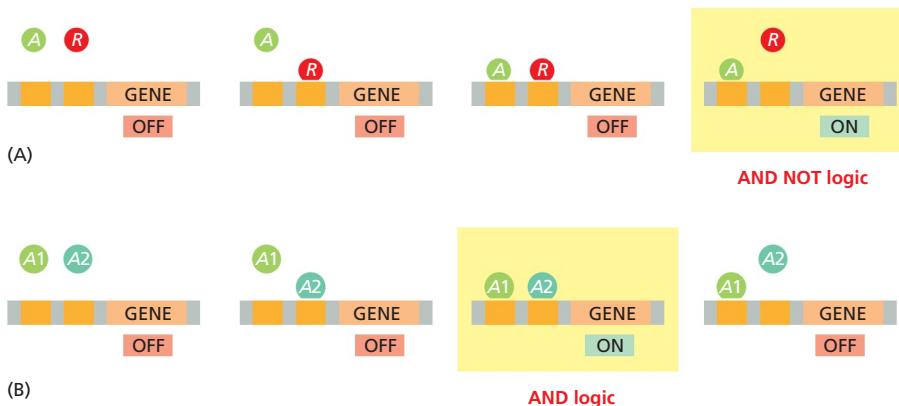
### Robustness Is an Important Characteristic of Biological Networks

Biological regulatory systems are exposed to frequent and sometimes extreme variations in external conditions or the concentrations or activities of key components. The ability of these systems to function normally in the face of such perturbations is called **robustness**. If we understand a complex system to the extent that we can reproduce its behavior with a computational model, then the robustness of the system can be assessed by determining how well its normal function persists following changes in various parameters, such as rate constants and component concentrations. We have already seen, for example, how the presence of negative feedback reduces the sensitivity of the steady state to changes in the values of the system's parameters (see Figure 8–77). Considerations of robustness also apply to dynamic behaviors. Thus, for example, when discussing negative feedback, we described how the behavior of a system tends to become more oscillatory as the number of components that constitute the feedback loop increases. If we use different values of the parameters in models derived for systems like those in Figure 8–78, we find that the system with the longer loop tends to exhibit stable oscillations within a much broader range of parameters, indicating that this system provides a more robust oscillator. We can perform similar calculations to determine the ability of different systems to achieve robust bistability arising from positive feedback. Thus, one benefit of computational models is that they allow us to probe the robustness of biological networks in a systematic and rigorous way.

### Two Transcription Regulators That Bind to the Same Gene Promoter Can Exert Combinatorial Control

Thus far, we have discussed how one transcription regulator can modulate the expression level of a gene. Most genes, however, are controlled by more than one type of transcription regulator, providing *combinatorial control* that allows two or more inputs to influence the expression of one gene. We can use computational methods to unveil some of the important regulatory features of combinatorial control systems.

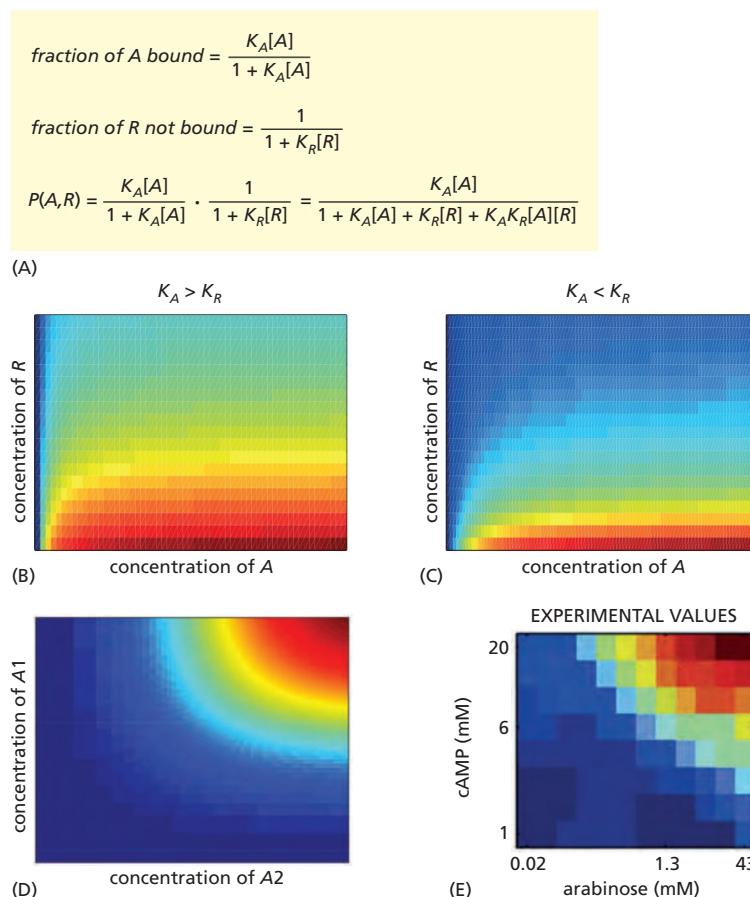
Consider a gene whose promoter contains binding sites for two regulatory proteins,  $A$  and  $R$ , which bind to their individual sites independently. There are four possible binding configurations (Figure 8–83A). Suppose that  $A$  is a transcription activator,  $R$  is a transcription repressor, and the gene is only active when  $A$  is bound and  $R$  is not bound. We learned earlier that the probability that  $A$  is bound and the probability that  $R$  is not bound can be determined by the equations in Figure 8–84A. The product of these two probabilities gives us the probability of gene activation.



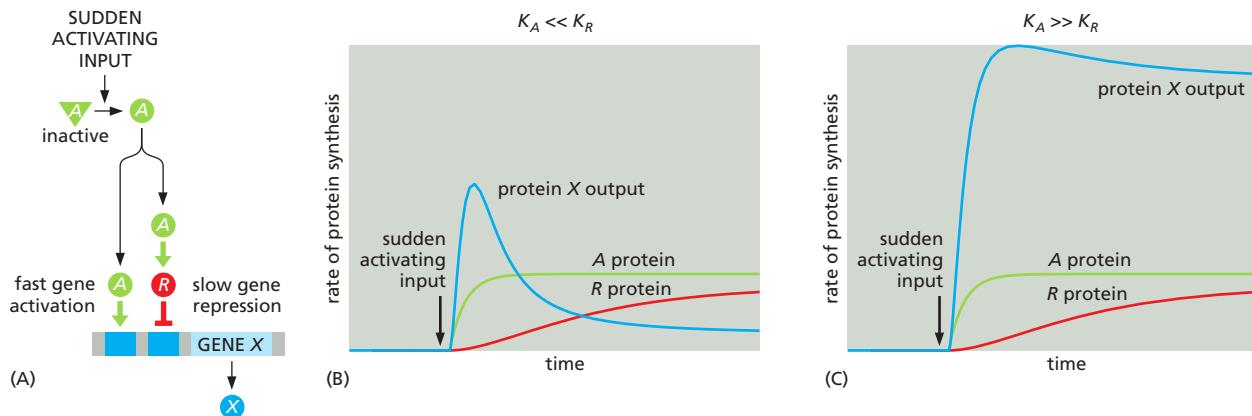
This example illustrates an AND NOT logic function ( $A$  and not  $R$ ) (see Figure 8–83A). Maximal activation of this gene is accomplished when  $[A]$  is high and  $[R]$  is zero. However, intermediate levels of gene activation are also possible depending on the levels of  $A$  and  $R$  and also on the binding affinities of  $[A]$  and  $[R]$  for their respective sites (that is,  $K_A$  and  $K_R$ ). When  $K_A \gg K_R$ , even a small concentration of  $[A]$  is capable of overcoming repression by  $R$ . Conversely, if  $K_A \ll K_R$ , then much more  $[A]$  is needed to activate the gene (Figure 8–84B and C).

Many other logic functions can govern combinatorial gene regulation. For example, an AND logic gate results when two activators,  $A_1$  and  $A_2$ , are both required for a gene to be transcribed (Figures 8–83B and 8–84D). In *E. coli* cells, the *AraJ* gene controls some aspects of arabinose sugar metabolism: its expression requires two transcription regulators, one activated by arabinose and the other activated by the small molecule cAMP (Figure 8–84E).

**Figure 8–83 Combinatorial control of gene expression.** There are many ways in which gene expression can be controlled by two transcription regulators. To define precisely the relationship between the two inputs and the gene expression output, a regulatory circuit is often described as a specific type of *logic gate*, a term borrowed from electronic circuit design. A simple example is the OR logic gate (not shown here), in which a gene is controlled by two transcription activators, and one or the other can activate gene expression. (A) In a system with an activator  $A$  and repressor  $R$ , if transcription is turned on only when  $A$  is bound and  $R$  is not, then the result is an AND NOT logic gate. We saw an example of this logic in Chapter 7 (Figure 7–15). (B) An AND gate results when two transcription activators,  $A_1$  and  $A_2$ , are both required to turn on a gene.



**Figure 8–84 How the quantitative output of a gene depends on both its combinatorial logic and the affinities of transcription regulators.** (A) In a combinatorial gene regulatory system like that illustrated in Figure 8–83A, the fraction of promoters bound by activator  $A$  and not bound by repressor  $R$  are each determined as shown here. The product of these probabilities provides the probability,  $P(A, R)$ , that a gene promoter is active. (B–E) In these four panels, red indicates high gene expression and blue indicates low gene expression. (B) and (C) depict gene expression from the system described in panel (A). The two panels demonstrate how the system behaves when the relative affinities of the two transcription regulators change as indicated above each panel. (D) Gene expression in a case where the gene turns on only at high levels of both activating inputs ( $A_1$  and  $A_2$ ), as shown in Figure 8–83B. (E) Experimental data showing measured expression of a gene in *E. coli* that is combinatorially regulated by two inputs: arabinose and cAMP. Note the close resemblance to panel (D). (E, adapted from S. Kaplan et al., *Mol. Cell* 29:786–792, 2008.)



### An Incoherent Feed-forward Interaction Generates Pulses

Imagine that a sudden input signal immediately activates a transcription activator  $A$  and that the same input signal induces the much slower synthesis of a transcription repressor protein  $R$  that acts on the same gene  $X$ . If  $A$  and  $R$  control gene expression by an AND NOT logic function like that described above, our intuition tells us that this system should be able to generate a pulse of transcription: when  $A$  is activated (and  $R$  is absent), the transcription of gene  $X$  will begin and cause an increase in the concentration of protein  $X$ , but then transcription will shut off when the concentration of  $R$  increases to a sufficiently high value.

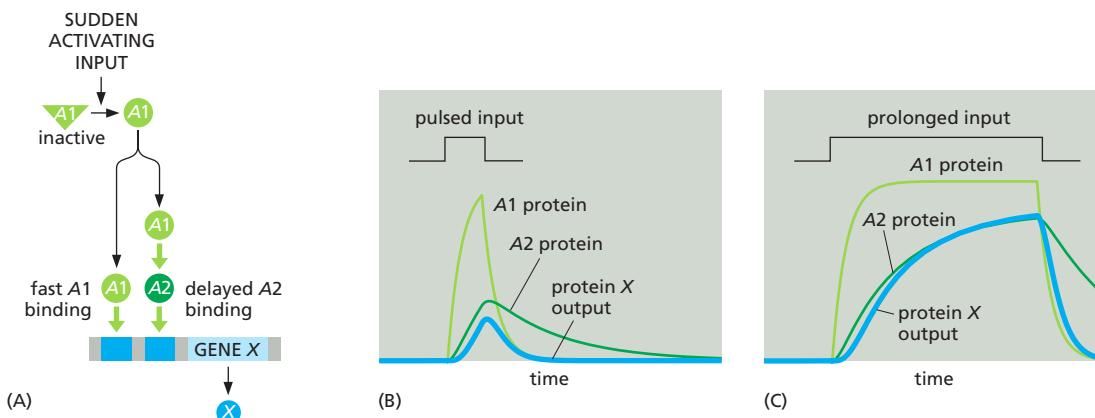
Arrangements of this type are common in the cell. In *E. coli*, for example, galactose metabolic genes are positively regulated by the catabolite activator protein (CAP), which is activated at high levels of cAMP. The same genes are repressed by the GalS repressor protein, which is encoded by a gene whose transcription is likewise activated by CAP. Thus, an increase in input (cAMP) activates  $A$  (CAP), and transcription of the galactose genes begins. But activation of  $A$  also causes a subsequent buildup of  $R$  (GalS), which causes the same genes to be repressed after a delay. This results in an *incoherent feed-forward motif* (Figure 8-85A).

The response of the incoherent feed-forward motif will vary, depending on the parameters of the system. Suppose, for example, that the transcription activator protein  $A$  binds more weakly to the gene regulatory region than does the transcription repressor protein  $R$  ( $K_A < K_R$ ). In this case, there will be a transient burst of protein synthesized by the affected gene (gene  $X$ ) in response to a sudden activating input (Figure 8-85B). In contrast, the output will be more sustained if  $K_A$  is much larger than  $K_R$ , because the repression will be too weak to overcome the gene activation (Figure 8-85C). Other properties of this network, such as the dependence of the amplitude of the pulse on the various rate constants in the system, can be explored with the same computational tools. Thus, our intuitive guess about how this system would behave was only partially correct; even the simplest of networks depends on precise interaction strengths, demonstrating yet again why mathematics is needed to complement cartoon drawings.

### A Coherent Feed-forward Interaction Detects Persistent Inputs

In the bacterium *E. coli*, the sugar arabinose is only consumed when the preferred sugar, glucose, is scarce. The strategy that cells use to assess the presence of arabinose and absence of glucose involves a feed-forward arrangement that is different from the one just described. In this case, depletion of glucose causes an increase of cAMP, which is sensed by the CAP transcription activator protein, as described previously. In this case, however, CAP also induces the synthesis of a second transcription activator, AraC. Both activator proteins are necessary to activate arabinose metabolic genes (the AND logic function in Figure 8-83B).

**Figure 8–85** How an incoherent feed-forward motif can generate a brief pulse of gene activation in response to a sustained input. (A) Diagram of an incoherent feed-forward motif in which the transcription activator  $A$  and the repressor  $R$  control the expression of gene  $X$  using the AND NOT logic of Figure 8–83A. (B) When  $K_A \ll K_R$ , this motif generates a pulse of protein  $X$  expression, such that the output goes back down even if the input remains high. (C) When  $K_A \gg K_R$ , the same motif responds to a sustained input by generating a sustained output.



This arrangement, known as a *coherent feed-forward motif*, has the interesting characteristics illustrated in **Figure 8–86**. Imagine that two activators, A1 and A2, are both required to initiate transcription of a gene. The input to the network activates A1 directly, but only activates A2 through this A1 activation. Thus, for a protein to be synthesized from this gene, long-term inputs are required that allow both A1 and A2 to be produced in active form. Brief input pulses are either ignored or produce small outputs. The requirement for a long input is important if assurances about a signal are needed before a costly cellular program is triggered. For example, glucose is the sugar on which *E. coli* cells grow best. Before cells trigger arabinose metabolism in the example above, it might be beneficial to be sure that glucose has been depleted (a sustained CAP pulse), rather than inducing the arabinose program during a transient glucose fluctuation.

### The Same Network Can Behave Differently in Different Cells Due to Stochastic Effects

Up to this point, we have assumed that all cells in a population produce identical behaviors if they contain the same network. It is important, however, to account for the fact that cells often show considerable individuality in their responses. Consider a situation in which a single mother cell divides into two daughter cells of equal volume. If the mother cell has only one molecule of a given protein, then only one daughter will inherit it. The daughters, though genetically identical, are already different. This variability is most pronounced for molecules that are present in small numbers. Nevertheless, even when there are many copies of a particular protein (or RNA), it is very unlikely that both daughter cells will end up with exactly the same number of molecules.

This is just one illustration of a universal feature of cells: their behaviors are often **stochastic**, meaning that they display variability in their protein content and therefore exhibit variations in phenotypes. In addition to the asymmetric partitioning of molecules following cell division, variability can originate from many chemical reactions. Imagine, for example, that our mother cell contains a simple gene regulatory circuit with a positive feedback loop like that shown in Figure 8–80B. Even if both daughter cells receive a copy of this circuit, including one copy of the initial transcription activator protein, there will be variability in the time required for promoter binding—and it will be statistically nearly impossible for the genes in the two daughter cells to become activated at precisely the same time. If the system is bistable and poised near a switching point, then variability in the response might flip the switch in only one daughter cell. Two daughter cells that were born identical can thereby acquire, by chance, a dramatic difference in phenotype.

More generally, isogenic populations of cells grown in the same environment display diversity in size, shape, cell-cycle position, and gene expression. These differences arise because biochemical reactions require probabilistic collisions

**Figure 8–86** How a coherent feed-forward motif responds to various inputs. (A) Diagram of a coherent feed-forward motif in which the transcription activators A1 and A2 together activate expression of gene X using the AND logic of Figure 8–83B. (B) The response to a brief input can be either weak (as shown) or nonexistent. This allows the motif to ignore random fluctuations in the concentration of signaling molecules. (C) A prolonged input produces a strong response that can turn off rapidly.

between randomly moving molecules, with each event resulting in changes in the number of molecular species by integer amounts. The amplified effect of fluctuations in a molecular reactant, or the compounded effects of fluctuations across many molecular reactants, often accumulates as an observable phenotype. This can endow a cell with individuality and generate non-genetic cell-to-cell variability in a population.

Non-genetic variability can be studied in the laboratory by single-cell measurements of fluorescent proteins expressed from genes under the control of a specific promoter. Live cells can be mounted on a slide and viewed through a fluorescence microscope, revealing the striking variability in protein expression levels (**Figure 8–87**). Another approach is to use flow cytometry, which works by streaming a dilute suspension of cells past an illuminator and measuring the fluorescence of individual cells as they flow past the detector (see Figure 8–2). Fluorescence values can be used to build histograms that reveal the variability in a process across a population of cells, with a broad histogram indicating higher variability.

### Several Computational Approaches Can Be Used to Model the Reactions in Cells

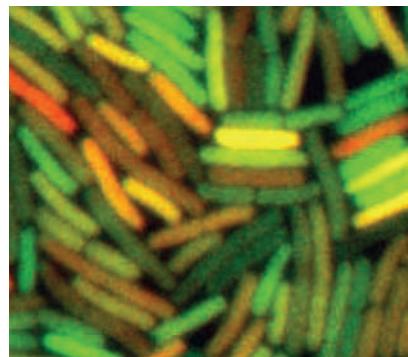
We have focused primarily on the use of ordinary differential equations to model the dynamics of simple regulatory circuits. These models are called *deterministic*, because they do not incorporate stochastic variability and will always produce the same result from a specific set of parameters. As we have seen, such models can provide useful insights, particularly in the detailed mechanistic analysis of small regulatory circuits. However, other types of computational approaches are also needed to comprehend the great complexity of cell behavior. *Stochastic models*, for example, attempt to account for the very important problem of random variability in molecular networks. These models do not provide deterministic predictions about the behavior of molecules; instead, they incorporate random variation into molecule numbers and interactions, and the purpose of these models is to obtain a better understanding of the probability that a system will exist in a certain state over time.

Numerous other modeling strategies have been or are being developed. *Boolean networks* are used for the qualitative analysis of complex gene regulatory networks containing large numbers of interacting components. In these models, each molecule is a node that can exist in either the active or inactive state, thereby affecting the state of the nodes it is linked to. Models of this sort provide insights into the flow of information through a network, and they were useful in helping us understand the complex gene regulatory network that controls the early development of the sea urchin (see Figure 7–43). Boolean networks therefore reduce complex networks to a highly simplified (and potentially inaccurate) form. At the other extreme are *agent-based simulations*, in which thousands of molecules (or “agents”) in a system are modeled individually, and their probable behaviors and interactions with each other over time are calculated on the basis of predicted physical and chemical behaviors, often while taking stochastic variation into account. Agent-based approaches are computationally demanding but have the potential to generate highly lifelike simulations of real biological systems.

### Statistical Methods Are Critical For the Analysis of Biological Data

Dynamics, differential equations, and theoretical modeling are not the be-all and end-all of mathematics. Other branches of the subject are no less important for biologists. Statistics—the mathematics of probabilistic processes and noisy datasets—is an inescapable part of every biologist’s life.

This is true in two main ways. First, imperfect measurement devices and other errors generate experimental noise in our data. Second, all cell-biological processes depend on the stochastic behavior of individual molecules, as we just discussed, and this results in biological noise in our results. How, in the face of all this noise, do we come to conclusions about the truth of hypotheses? The answer is statistical analysis, which shows how to move from one level of description to



**Figure 8–87** Different levels of gene expression in individual cells within a population of *E. coli* bacteria. For this experiment, two different reporter proteins (one fluorescing green, the other red), controlled by a copy of the same promoter, have been introduced into all of the bacteria. Some cells express only one gene copy, and so appear either red or green, while others express both gene copies, and so appear yellow. This experiment reveals variable levels of fluorescence, indicating variable levels of gene expression within an apparently uniform population of cells. (From M.B. Elowitz et al., *Science* 297:1183–1186, 2002. With permission from AAAS.)

another: from a set of erratic individual data points to a simpler description of the key features of the data.

Statistics teaches us that the more times we repeat our measurements, the better and more refined the conclusions we can draw from them. Given many repetitions, it becomes possible to describe our data in terms of variables that summarize the features that matter: the mean value of the measured variable, taken over the set of data points; the magnitude of the noise (the standard deviation of the set of data points); the likely error in our estimate of the mean value (the standard error of the mean); and, for specialists, the details of the probability distribution describing the likelihood that an individual measurement will yield a given value. For all these things, statistics provides recipes and quantitative formulas that biologists must understand if they are to make rigorous conclusions on the basis of variable results.

## Summary

*Quantitative mathematical analysis can provide a powerful extra dimension in our understanding of cell regulation and function. Cell regulatory systems often depend on macromolecular interactions, and mathematical analysis of the dynamics of these interactions can unveil important insights into the importance of binding affinities and protein stability in the generation of transcriptional or other signals. Regulatory systems often employ network motifs that generate useful behaviors: a rapid negative feedback loop dampens the response to input signals; a delayed negative feedback loop creates a biochemical oscillator; positive feedback yields a system that alternates between two stable states; and feed-forward motifs provide systems that generate transient signal pulses or respond only to sustained inputs. The dynamic behavior of these network motifs can be dissected in detail with deterministic and stochastic mathematical modeling.*

## WHAT WE DON'T KNOW

- Many of the tools that revolutionized DNA technology were discovered by scientists studying basic biological problems that had no obvious applications. What are the best strategies to ensure that such crucially important technologies will continue to be discovered?
- As the cost of DNA sequencing decreases and the amount of sequence data accumulates, how are we going to keep track of and meaningfully analyze this vast amount of information? What new questions will this information allow us to answer?
- Can we develop tools to analyze each of the post-transcriptional modifications on the proteins in living cells, so as to follow all of their changes in real time?
- Can we develop mathematical models to accurately describe the enormous complexity of cellular networks and to predict undiscovered components and mechanisms?

## PROBLEMS

Which statements are true? Explain why or why not.

**8–1** Because a monoclonal antibody recognizes a specific antigenic site (epitope), it binds only to the specific protein against which it was made.

**8–2** Given the inexorable march of technology, it seems inevitable that the sensitivity of detection of molecules will ultimately be pushed beyond the yoctomole level ( $10^{-24}$  mole).

**8–3** If each cycle of PCR doubles the amount of DNA synthesized in the previous cycle, then 10 cycles will give a  $10^3$ -fold amplification, 20 cycles will give a  $10^6$ -fold amplification, and 30 cycles will give a  $10^9$ -fold amplification.

**8–4** To judge the biological importance of an interaction between protein A and protein B, we need to know quantitative details about their concentrations, affinities, and kinetic behaviors.

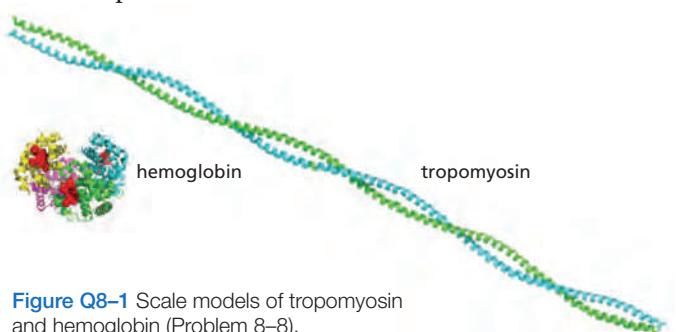
**8–5** The rate of change in the concentration of any molecular species X is given by the balance between its rate of appearance and its rate of disappearance.

**8–6** After a sudden increase in transcription, a protein with a slow rate of degradation will reach a new steady state level more quickly than a protein with a rapid rate of degradation.

Discuss the following problems.

**8–7** A common step in the isolation of cells from a sample of animal tissue is to treat the tissue with trypsin, collagenase, and EDTA. Why is such a treatment necessary, and what does each component accomplish? And why does this treatment not kill the cells?

**8–8** Tropomyosin, at 93 kd, sediments at 2.6S, whereas the 65-kd protein, hemoglobin, sediments at 4.3S. (The sedimentation coefficient S is a linear measure of the rate of sedimentation.) These two proteins are drawn to scale in **Figure Q8–1**. How is it that the bigger protein sediments more slowly than the smaller one? Can you think of an analogy from everyday experience that might help you with this problem?

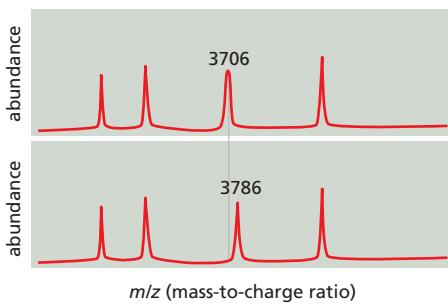


**Figure Q8–1** Scale models of tropomyosin and hemoglobin (Problem 8–8).

**8–9** Hybridoma technology allows one to generate monoclonal antibodies to virtually any protein. Why is it, then, that genetically tagging proteins with epitopes is such a commonly used technique, especially since an epitope tag has the potential to interfere with the function of the protein?

**8–10** How many copies of a protein need to be present in a cell in order for it to be visible as a band on an SDS gel? Assume that you can load 100 µg of cell extract onto a gel and that you can detect 10 ng in a single band by silver staining the gel. The concentration of protein in cells is about 200 mg/mL, and a typical mammalian cell has a volume of about 1000 µm<sup>3</sup> and a typical bacterium a volume of about 1 µm<sup>3</sup>. Given these parameters, calculate the number of copies of a 120-kd protein that would need to be present in a mammalian cell and in a bacterium in order to give a detectable band on a gel. You might try an order-of-magnitude guess before you make the calculations.

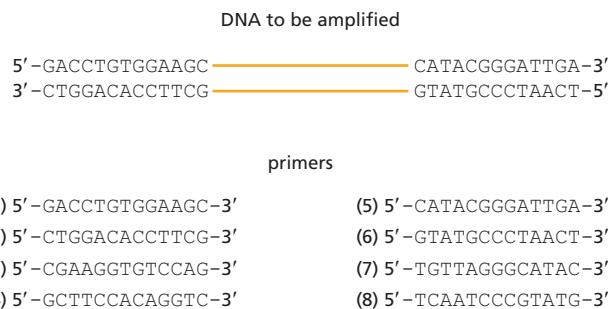
**8–11** You have isolated the proteins from two adjacent spots after two-dimensional polyacrylamide-gel electrophoresis and digested them with trypsin. When the masses of the peptides were measured by MALDI-TOF mass spectrometry, the peptides from the two proteins were found to be identical except for one (Figure Q8–2). For this peptide, the mass-to-charge (*m/z*) values differed by 80, a value that does not correspond to a difference in amino acid sequence. (For example, glutamic acid instead of valine at one position would give an *m/z* difference of around 30.) Can you suggest a possible difference between the two peptides that might account for the observed *m/z* difference?



**Figure Q8–2**  
Masses of peptides measured by MALDI-TOF mass spectrometry (Problem 8–11). Only the numbered peaks differ between the two protein samples.

**8–12** You want to amplify the DNA between the two stretches of sequence shown in Figure Q8–3. Of the listed primers, choose the pair that will allow you to amplify the DNA by PCR.

**8–13** In the very first round of PCR using genomic DNA, the DNA primers prime synthesis that terminates only when the cycle ends (or when a random end of DNA is encountered). Yet, by the end of 20 to 30 cycles—a typical amplification—the only visible product is defined precisely by the ends of the DNA primers. In what cycle is a double-stranded fragment of the correct size first generated?



**Figure Q8–3** DNA to be amplified and potential PCR primers (Problem 8–12).

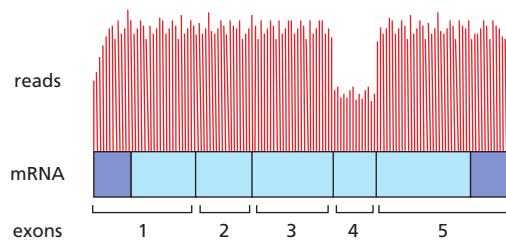
**8–14** Explain the difference between a gain-of-function mutation and a dominant-negative mutation. Why are both these types of mutation usually dominant?

**8–15** Discuss the following statement: “We would have no idea today of the importance of insulin as a regulatory hormone if its absence were not associated with the human disease diabetes. It is the dramatic consequences of its absence that focused early efforts on the identification of insulin and the study of its normal role in physiology.”

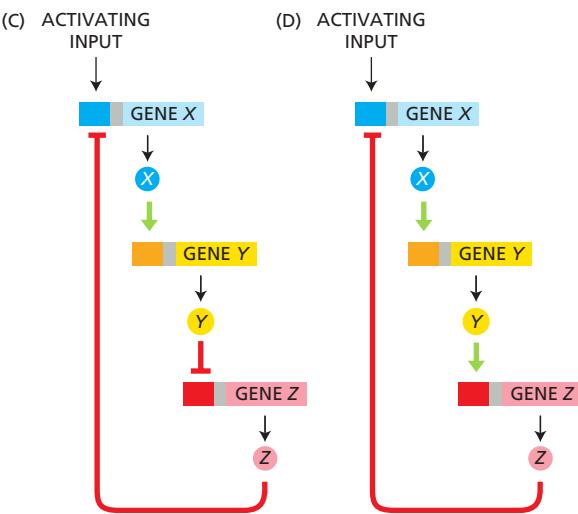
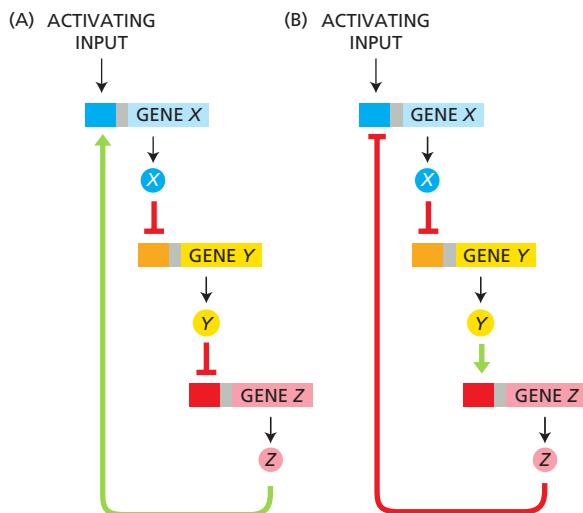
**8–16** You have just gotten back the results from an RNA-seq analysis of mRNAs from liver. You had anticipated counting the number of reads of each mRNA to determine the relative abundance of different mRNAs. But you are puzzled because many of the mRNAs have given you results like those shown in Figure Q8–4. How is it that different parts of an mRNA can be represented at different levels?

**8–17** Examine the network motifs in Figure Q8–5. Decide which ones are negative feedback loops and which are positive. Explain your reasoning.

**8–18** Imagine that a random perturbation positions a bistable system precisely at the boundary between two stable states (at the orange dot in Figure Q8–6). How would the system respond?

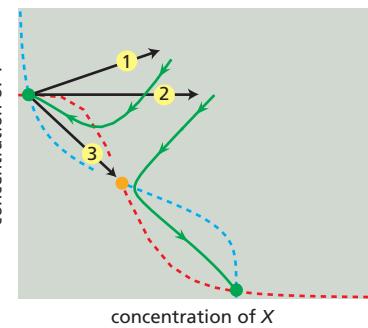


**Figure Q8–4** RNA-seq reads for a liver mRNA (Problem 8–16). The exon structure of the mRNA is indicated, with protein-coding segments indicated in light blue and untranslated regions in dark blue. The numbers of sequencing reads are indicated by the heights of the vertical lines above the mRNA.



**Figure Q8-5** Network motifs composed of transcription activators and repressors (Problem 8-17).

**8-19** Detailed analysis of the regulatory region of the *Lac* operon has revealed surprising complexity. Instead of a single binding site for the *Lac* repressor, as might be expected, there are three sites termed operators: O<sub>1</sub>, O<sub>2</sub>, and O<sub>3</sub>, arrayed along the DNA as shown in **Figure Q8-7**. To probe the functions of these three sites, you make a series of constructs in which various combinations of operator sites are present. You examine their ability to repress expression of  $\beta$ -galactosidase, using either tetrameric (wild type) or dimeric (mutant) forms of the *Lac* repressor. The dimeric form of the repressor can bind to a single operator (with the same affinity as the tetramer) with each monomer binding to half the site. The tetramer, the form normally expressed in cells, can bind to two sites simultaneously. When you measure repression of  $\beta$ -galactosidase expression, you find the results shown in **Figure Q8-7**, with higher numbers indicating more effective repression.



**Figure Q8-6** Perturbations of a bistable system (Problem 8-18). As shown by the green lines, after perturbation 1 the system returns to its original stable state (green dot at left), and after perturbation 2, the system moves to the other stable state (green dot at right). Perturbation 3 moves the system to the precise boundary between the two stable states (orange dot).

**A.** Which single operator site is the most important for repression? How can you tell?

**B.** Do combinations of operator sites (Figure Q8-7, constructs 1, 2, 3, and 5) substantially increase repression by the dimeric repressor? Do combinations of operator sites substantially increase repression by the tetrameric repressor? If the two repressors behave differently, offer an explanation for the difference.

**C.** The wild-type repressor binds O<sub>3</sub> very weakly when it is by itself on a segment of DNA. However, if O<sub>1</sub> is included on the same segment of DNA, the repressor binds O<sub>3</sub> quite well. How can that be?

	92 bp	401 bp	2-mer	4-mer
1	O <sub>3</sub> O <sub>1</sub>	O <sub>2</sub>	110	6700
2	O <sub>3</sub> O <sub>1</sub>		90	3900
3	O <sub>1</sub>	O <sub>2</sub>	80	1400
4	O <sub>1</sub>		60	140
5	O <sub>3</sub>	O <sub>2</sub>	1	5
6	O <sub>3</sub>		1	2
7		O <sub>2</sub>	1	1
8			1	1

**Figure Q8-7** Repression of  $\beta$ -galactosidase by promoter regions that contain different combinations of *Lac* repressor binding sites (Problem 8-19). The base-pair (bp) separation of the three operator sites is shown. Numbers at right refer to the level of repression, with higher numbers indicating more effective repression by dimeric (2-mer) or tetrameric (4-mer) repressors. (From S. Oehler et al., *EMBO J.* 9:973–979, 1990. With permission from John Wiley and Sons.)

## REFERENCES

### General

- Ausubel FM, Brent R, Kingston RE et al. (eds) (2002) Short Protocols in Molecular Biology, 5th ed. New York: Wiley.
- Brown TA (2007) Genomes 3. New York: Garland Science Publishing.
- Spector DL, Goldman RD & Leinwand LA (eds) (1998) Cells: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Watson JD & Berry A (2008) DNA: The Secret of Life. New York: Alfred A Knopf.
- Watson JD, Myers RM & Caudy AA (2007) Recombinant DNA: Genes and Genomes – A Short Course, 3rd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

### Isolating Cells and Growing Them in Culture

- Ham RG (1965) Clonal growth of mammalian cells in a chemically defined, synthetic medium. *Proc. Natl Acad. Sci. USA* 53, 288–293.
- Harlow E & Lane D (1999) Using Antibodies: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Herzenberg LA, Sweet RG & Herzenberg LA (1976) Fluorescence-activated cell sorting. *Sci. Am.* 234, 108–116.
- Milstein C (1980) Monoclonal antibodies. *Sci. Am.* 243, 66–74.

### Purifying Proteins

- de Duve C & Beaufay H (1981) A short history of tissue fractionation. *J. Cell Biol.* 91, 293s–299s.
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Scopes RK (1994) Protein Purification: Principles and Practice, 3rd ed. New York: Springer-Verlag.
- Simpson RJ, Adams PD & Golemis EA (2008) Basic Methods in Protein Purification and Analysis: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Wood DW (2014) New trends and affinity tag designs for recombinant protein purification. *Curr. Opin. Struct. Biol.* 26, 54–61.

### Analyzing Proteins

- Choudhary C & Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.* 11, 427–439.
- Domon B & Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312, 212–217.
- Goodrich JA & Kugel JF (2007) Binding and Kinetics for Molecular Biologists. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Kendrew JC (1961) The three-dimensional structure of a protein molecule. *Sci. Am.* 205, 96–111.
- Knight ZA & Shokat KM (2007) Chemical genetics: where genetics and pharmacology meet. *Cell* 128, 425–430.
- O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021.
- Pollard TD (2010) A guide to simple and informative binding assays. *Mol. Biol. Cell* 21, 4061–4067.
- Wüthrich K (1989) Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* 243, 45–50.

### Analyzing and Manipulating DNA

- Cohen S, Chang A, Boyer H & Helling R (1973) Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* 70, 3240–3244.
- Green MR & Sambrook J (2012) Molecular Cloning: A Laboratory Manual, 4th ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Jackson D, Symons R & Berg P (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 69, 2904–2909.

Kosuri S & Church GM (2014) Large-scale *de novo* DNA synthesis: technologies and applications. *Nat. Methods* 11, 499–507.

Maniatis T, Hardison RC, Lacy E et al. (1978) The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15, 687–701.

Mullis KB (1990) The unusual origin of the polymerase chain reaction. *Sci. Am.* 262, 56–61.

Nathans D & Smith HO (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. *Annu. Rev. Biochem.* 44, 273–293.

Saiki RK, Gelfand DH, Stoffel S et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* 74, 5463–5467.

Shendure J & Lieberman Aiden E (2012) The expanding scope of DNA sequencing. *Nat. Biotechnol.* 30, 1084–1094.

### Studying Gene Expression and Function

Botstein D, White RL, Skolnick M & Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

DeRisi JL, Iyer VR & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.

Esveld KM, Mali P, Braff JL et al. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10, 1116–1121.

Fellmann C & Lowe SW (2014) Stable RNA interference rules for silencing. *Nat. Cell Biol.* 16, 10–18.

Mello CC & Conte D (2004) Revealing the world of RNA interference. *Nature* 431, 338–342.

Nüsslein-Volhard C & Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801.

Palmiter RD & Brinster RL (1985) Transgenic mice. *Cell* 41, 343–345.

Weigel D & Glazebrook J (2002) *Arabidopsis*: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Wilson RC & Doudna JA (2013) Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* 42, 217–239.

### Mathematical Analysis of Cell Functions

Alon U (2006) An Introduction to Systems Biology: Design Principles of Biological Circuits. Boca Raton, FL: Chapman & Hall/CRC.

Alon U (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.

Ferrell JE Jr (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* 14, 140–148.

Ferrell JE Jr, Tsai TY & Yang Q (2011) Modeling the cell cycle: why do certain circuits oscillate? *Cell* 144, 874–885.

Gunawardena J (2014) Models in biology: ‘accurate descriptions of our pathetic thinking’. *BMC Biol.* 12, 29.

Lewis J (2008) From signals to patterns: space, time, and mathematics in developmental biology. *Science* 322, 399–403.

Mogilner A, Allard J & Wollman R (2012) Cell polarity: quantitative modeling as a tool in cell biology. *Science* 336, 175–179.

Novak B & Tyson JJ (2008) Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* 9, 981–991.

Silva-Rocha R & de Lorenzo V (2008) Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Lett.* 582, 1237–1244.

Tyson JJ, Chen KC & Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231.

## CHAPTER

## 9

# Visualizing Cells

Understanding the structural organization of cells is essential for learning how they function. In this chapter, we briefly describe some of the principal microscopy methods used to study cells. Optical microscopy will be our starting point because cell biology began with the light microscope, and it is still an indispensable tool. The development of methods for the specific labeling and imaging of individual cellular constituents and the reconstruction of their three-dimensional architecture has meant that, far from falling into disuse, optical microscopy continues to increase in importance. One advantage of optical microscopy is that light is relatively nondestructive. By tagging specific cell components with fluorescent probes, such as intrinsically fluorescent proteins, we can watch their movement, dynamics, and interactions in living cells. Although conventional optical microscopy is limited in resolution by the wavelength of visible light, new methods cleverly bypass this limitation and allow the position of even single molecules to be mapped. By using a beam of electrons instead of visible light, electron microscopy can image the interior of cells, and their macromolecular components, at almost atomic resolution, and in three dimensions.

This chapter is intended as a companion, rather than an introduction, to the chapters that follow; readers may wish to refer back to it as they encounter applications of microscopy to basic biological problems in the later pages of the book.

## LOOKING AT CELLS IN THE LIGHT MICROSCOPE

A typical animal cell is 10–20  $\mu\text{m}$  in diameter, which is about one-fifth the size of the smallest object that we can normally see with the naked eye. Only after good light microscopes became available in the early part of the nineteenth century did Schleiden and Schwann propose that all plant and animal tissues were aggregates of individual cells. Their proposal in 1838, known as the **cell doctrine**, marks the formal birth of cell biology.

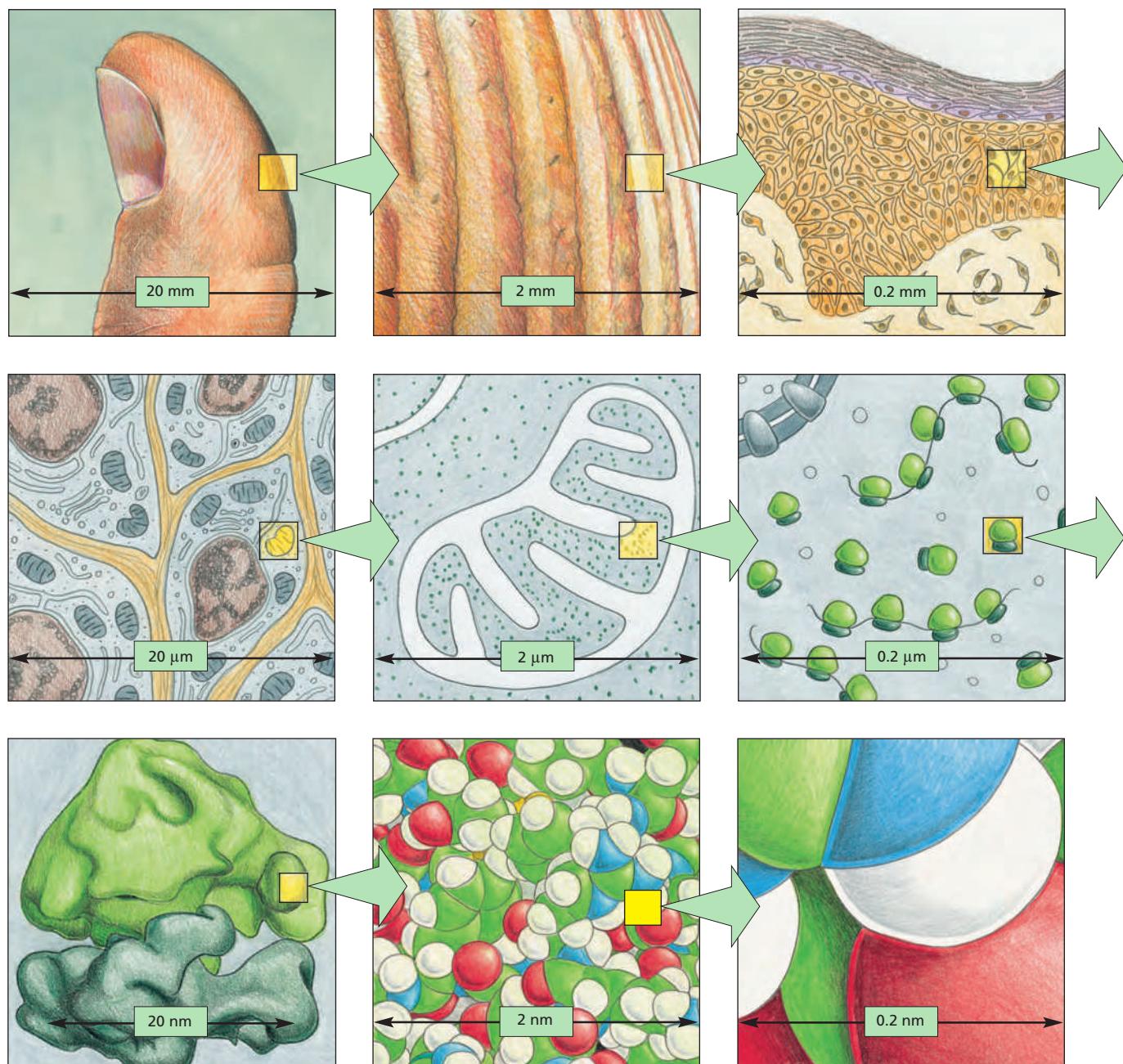
Animal cells are not only tiny, but they are also colorless and translucent. The discovery of their main internal features, therefore, depended on the development, in the late nineteenth century, of a variety of stains that provided sufficient contrast to make those features visible. Similarly, the far more powerful electron microscope introduced in the early 1940s required the development of new techniques for preserving and staining cells before the full complexities of their internal fine structure could begin to emerge. To this day, microscopy often relies as much on techniques for preparing the specimen as on the performance of the microscope itself. In the following discussions, we therefore consider both instruments and specimen preparation, beginning with the light microscope.

The images in **Figure 9–1** illustrate a stepwise progression from a thumb to a cluster of atoms. Each successive image represents a tenfold increase in magnification. The naked eye can see features in the first two panels, the light microscope allows us to see details corresponding to about the fourth or fifth panel, and the electron microscope takes us to about the seventh or eighth panel. **Figure 9–2** shows the sizes of various cellular and subcellular structures and the ranges of size that different types of microscopes can visualize.

## IN THIS CHAPTER

LOOKING AT CELLS IN THE LIGHT MICROSCOPE

LOOKING AT CELLS AND MOLECULES IN THE ELECTRON MICROSCOPE

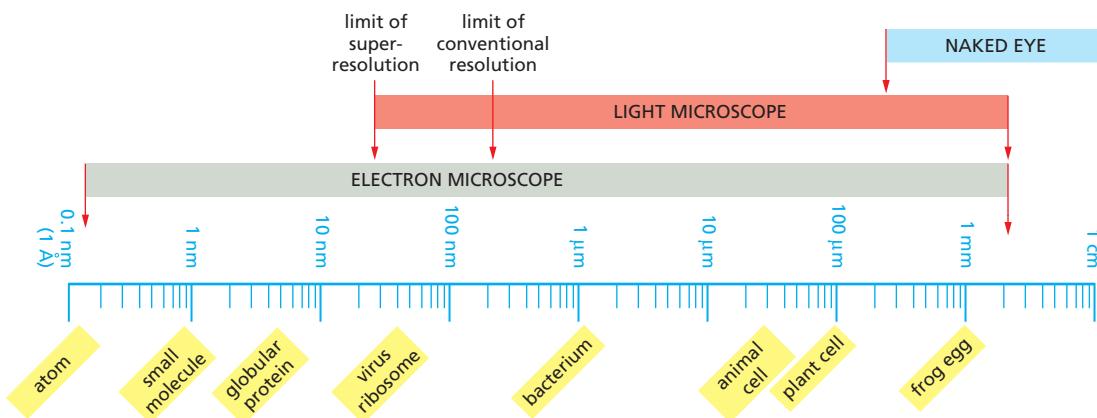


### The Light Microscope Can Resolve Details 0.2 $\mu\text{m}$ Apart

For well over 100 years, all microscopes were constrained by a fundamental limitation: that a given type of radiation cannot be used to probe structural details much smaller than its own wavelength. A limit to the resolution of a light microscope was therefore set by the wavelength of visible light, which ranges from about 0.4  $\mu\text{m}$  (for violet) to 0.7  $\mu\text{m}$  (for deep red). In practical terms, bacteria and mitochondria, which are about 500 nm (0.5  $\mu\text{m}$ ) wide, are generally the smallest objects whose shape we can clearly discern in the **light microscope**; details smaller than this are obscured by effects resulting from the wavelike nature of light. To understand why this occurs, we must follow the behavior of a beam of light as it passes through the lenses of a microscope (**Figure 9-3**).

Because of its wave nature, light does not follow the idealized straight ray paths that geometrical optics predicts. Instead, light waves travel through an optical system by many slightly different routes, like ripples in water, so that they

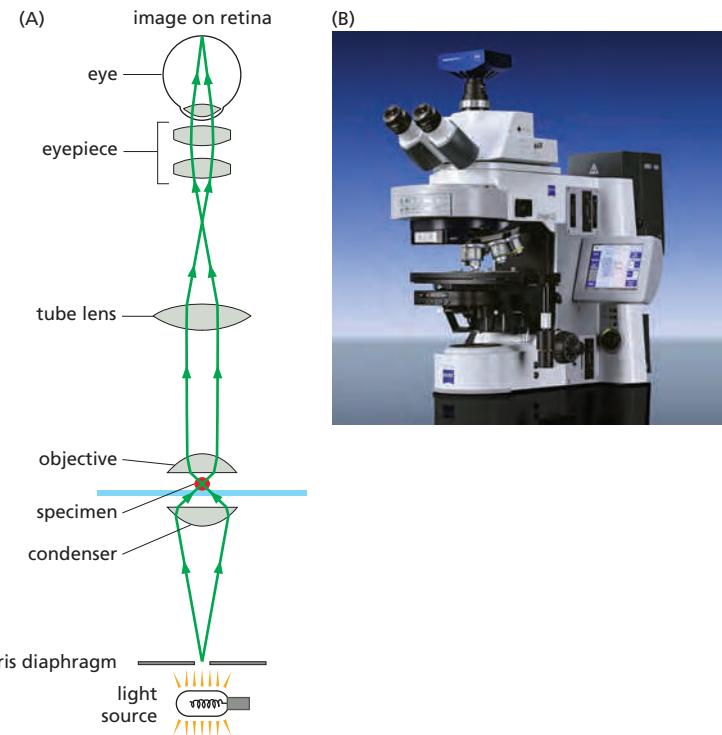
**Figure 9-1** A sense of scale between living cells and atoms. Each diagram shows an image magnified by a factor of ten in an imaginary progression from a thumb, through skin cells, to a ribosome, to a cluster of atoms forming part of one of the many protein molecules in our body. Atomic details of biological macromolecules, as shown in the last two panels, are usually beyond the power of the electron microscope. While color has been used here in all the panels, it is not a feature of objects much smaller than the wavelength of light, so the last five panels should really be in black and white.



**Figure 9–2 Resolving power.** Sizes of cells and their components are drawn on a logarithmic scale, indicating the range of objects that can be readily resolved by the naked eye and in the light and electron microscopes. Note that new superresolution microscopy techniques, discussed in detail later, allow an improvement in resolution by an order of magnitude compared with conventional light microscopy.

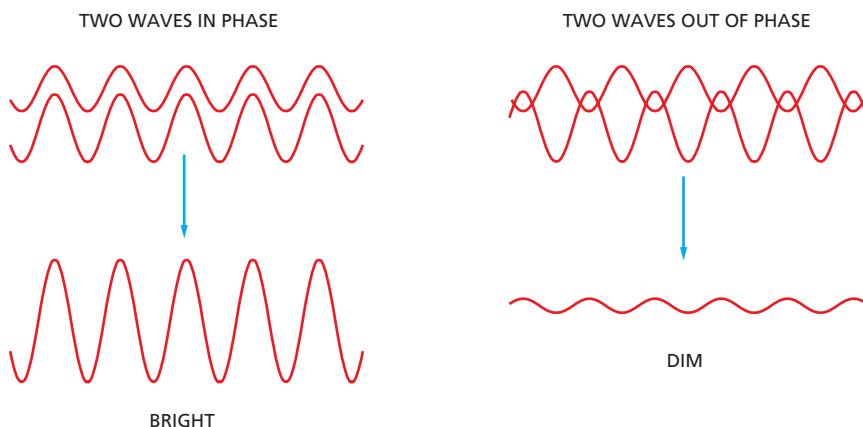
interfere with one another and cause *optical diffraction* effects. If two trains of waves reaching the same point by different paths are precisely *in phase*, with crest matching crest and trough matching trough, they will reinforce each other so as to increase brightness. In contrast, if the trains of waves are *out of phase*, they will interfere with each other in such a way as to cancel each other partly or entirely (Figure 9–4). The interaction of light with an object changes the phase relationships of the light waves in a way that produces complex interference effects. At high magnification, for example, the shadow of an edge that is evenly illuminated with light of uniform wavelength appears as a set of parallel lines (Figure 9–5), whereas that of a circular spot appears as a set of concentric rings. For the same reason, a single point seen through a microscope appears as a blurred disc, and two point objects close together give overlapping images and may merge into one.

The following units of length are commonly employed in microscopy:  
 $\mu\text{m}$  (micrometer) =  $10^{-6}\text{ m}$   
 $\text{nm}$  (nanometer) =  $10^{-9}\text{ m}$   
 $\text{\AA}$  (Ångström unit) =  $10^{-10}\text{ m}$



**Figure 9–3 A light microscope.**

(A) Diagram showing the light path in a compound microscope. Light is focused on the specimen by lenses in the condenser. A combination of objective lenses, tube lenses, and eyepiece lenses is arranged to focus an image of the illuminated specimen in the eye. (B) A modern research light microscope. (B, courtesy of Carl Zeiss Microscopy, GmbH.)



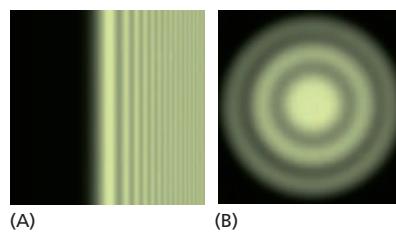
**Figure 9–4 Interference between light waves.** When two light waves combine in phase, the amplitude of the resultant wave is larger and the brightness is increased. Two light waves that are out of phase cancel each other partly and produce a wave whose amplitude, and therefore brightness, is decreased.

Although no amount of refinement of the lenses can overcome the diffraction limit imposed by the wavelike nature of light, other ways of cleverly bypassing this limit have emerged, creating so-called superresolution imaging techniques that can even detect the position of single molecules.

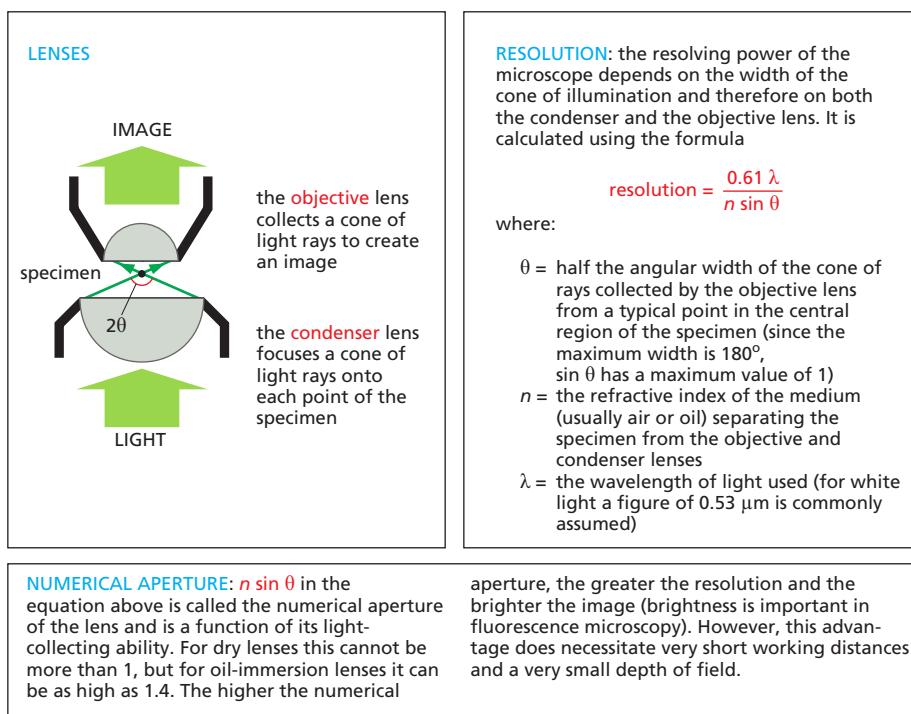
The limiting separation at which two objects appear distinct—the so-called **limit of resolution**—depends on both the wavelength of the light and the *numerical aperture* of the lens system used. The numerical aperture affects the light-gathering ability of the lens and is related both to the angle of the cone of light that can enter it and to the refractive index of the medium the lens is operating in; the wider the microscope opens its eye, so to speak, the more sharply it can see (**Figure 9–6**). The *refractive index* is the ratio of the speed of light in a vacuum to the speed of light in a particular transparent medium. For example, for water this is 1.33, meaning that light travels 1.33 times slower in water than in a vacuum. Under the best conditions, with violet light (wavelength = 0.4  $\mu\text{m}$ ) and a numerical aperture of 1.4, the basic light microscope can theoretically achieve a limit of resolution of about 0.2  $\mu\text{m}$ , or 200 nm. Some microscope makers at the end of the nineteenth century achieved this resolution, but it is routinely matched in contemporary, factory-produced microscopes. Although it is possible to *enlarge* an image as much as we want—for example, by projecting it onto a screen—it is not possible, in a conventional light microscope, to resolve two objects in the light microscope that are separated by less than about 0.2  $\mu\text{m}$ ; they will appear as a single object. It is important, however, to distinguish between *resolution* and *detection*. If a small object, below the resolution limit, itself emits light, then we may still be able to see or detect it. Thus, we can see a single fluorescently labeled microtubule even though it is about ten times thinner than the resolution limit of the light microscope. Diffraction effects, however, will cause it to appear blurred and at least 0.2  $\mu\text{m}$  thick (see Figure 9–16). In a similar way, we can see the stars in the night sky, even though their diameters are far below the angular resolution of our unaided eyes: they all appear as similar, slightly blurred points of light, differing only in their color and brightness.

### Photon Noise Creates Additional Limits to Resolution When Light Levels Are Low

Any image, whether produced by an electron microscope or by an optical microscope, is made by particles—electrons or photons—striking a detector of some sort. But these particles are governed by quantum mechanics, so the numbers reaching the detector are predictable only in a statistical sense. Finite samples, collected by imaging for a limited period of time (that is, by taking a snapshot), will show random variation: successive snapshots of the same scene will not be exactly identical. Moreover, every detection method has some level of background signal or noise, adding to the statistical uncertainty. With bright illumination, corresponding to very large numbers of photons or electrons, the features of the imaged



**Figure 9–5 Images of an edge and of a point of light.** (A) The interference effects, or fringes, seen at high magnification when light of a specific wavelength passes the edge of a solid object placed between the light source and the observer. (B) The image of a point source of light. Diffraction spreads this out into a complex, circular pattern, whose width depends on the numerical aperture of the optical system: the smaller the aperture, the bigger (more blurred) the diffracted image. Two point sources can be just resolved when the center of the image of one lies on the first dark ring in the image of the other: this is used to define the limit of resolution.



specimen are accurately determined based on the distribution of these particles at the detector. However, with smaller numbers of particles, the structural details of the specimen are obscured by the statistical fluctuations in the numbers of particles detected in each region, which give the image a speckled appearance and limit its precision. The term *noise* describes this random variability.

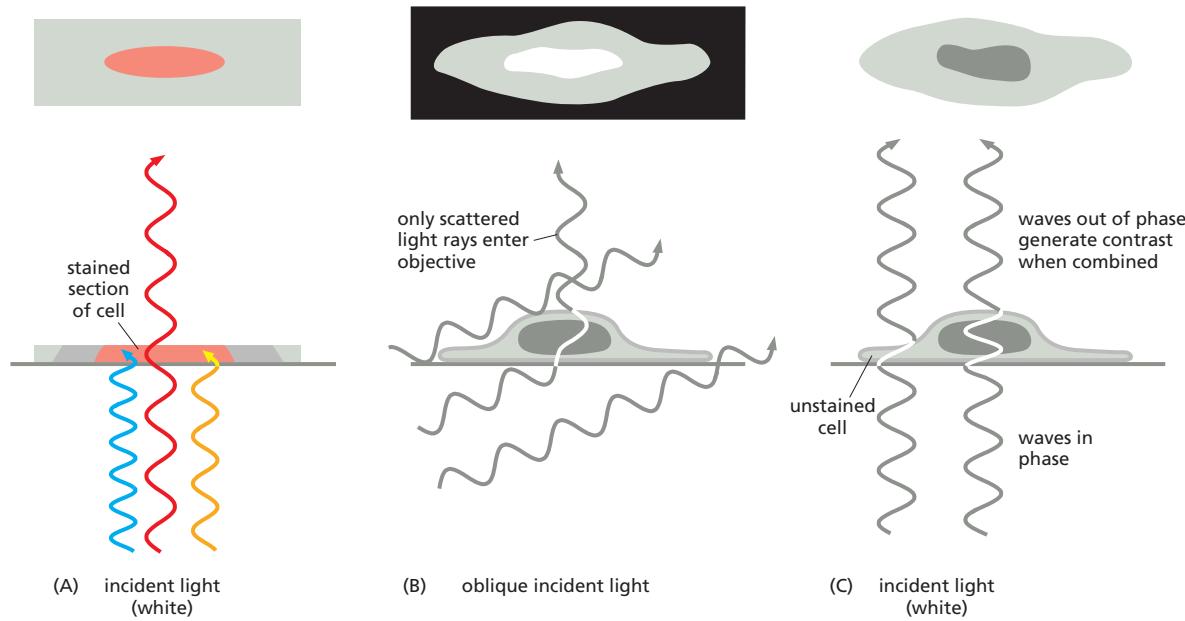
### Living Cells Are Seen Clearly in a Phase-Contrast or a Differential-Interference-Contrast Microscope

There are many ways in which contrast in a specimen can be generated (**Figure 9–7A**). While fixing and staining a specimen can generate contrast through color, microscopists have always been challenged by the possibility that some components of the cell may be lost or distorted during specimen preparation. The only certain way to avoid the problem is to examine cells while they are alive, without fixing or freezing. For this purpose, light microscopes with special optical systems are especially useful.

In the normal **bright-field microscope**, light passing through a cell in culture forms the image directly. Another system, **dark-field microscopy**, exploits the fact that light rays can be scattered in all directions by small objects in their path. If oblique lighting from the condenser is arranged, which does not directly enter the objective, focused but unstained objects in a living cell can scatter the rays, some of which then enter the objective to create a bright image against a black background (Figure 9–7B).

When light passes through a living cell, the phase of the light wave is changed according to the cell's refractive index: a relatively thick or dense part of the cell, such as a nucleus, slows the light passing through it. The phase of the light, consequently, is shifted relative to light that has passed through an adjacent thinner region of the cytoplasm (Figure 9–7C). The **phase-contrast microscope** and, in a more complex way, the **differential-interference-contrast microscope** increase these phase differences so that the waves are more nearly out of phase, producing amplitude differences when the sets of waves recombine, thereby creating an image of the cell's structure. Both types of light microscopy are widely used to visualize living cells (see Movie 17.2). **Figure 9–8** compares images of the same cell obtained by four kinds of light microscopy.

**Figure 9–6 Numerical aperture.** The path of light rays passing through a transparent specimen in a microscope illustrates the concept of numerical aperture and its relation to the limit of resolution.



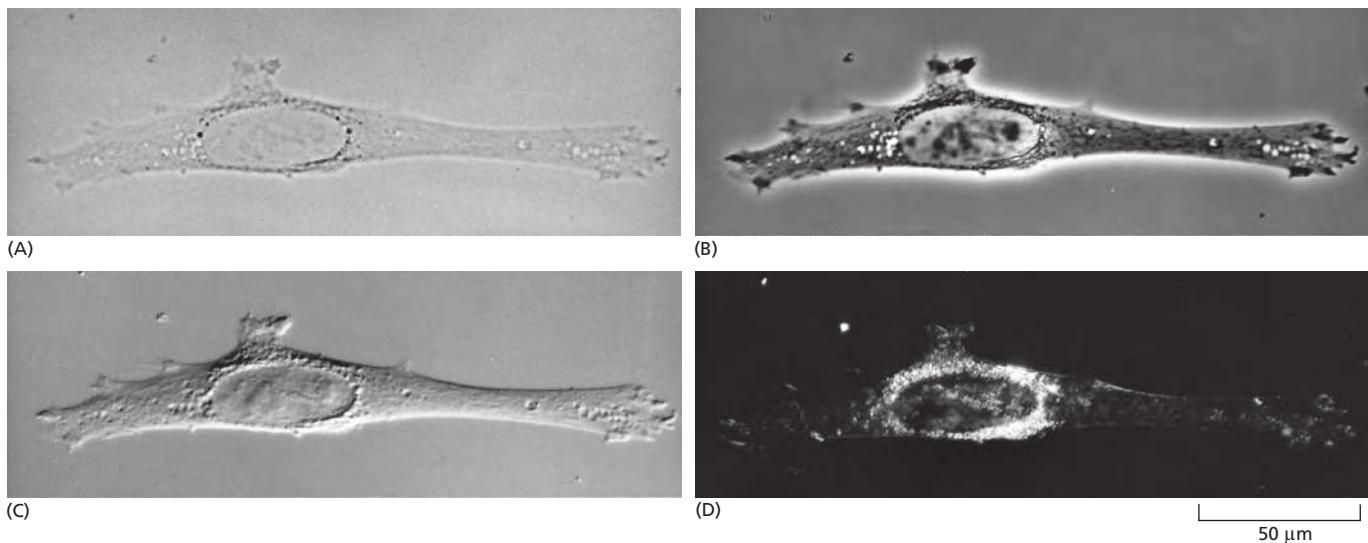
**Figure 9–7 Contrast in light microscopy.** (A) The stained portion of the cell will absorb light of some wavelengths, which depends on the stain, but will allow other wavelengths to pass through it. A colored image of the cell is thereby obtained that is visible in the normal bright-field light microscope. (B) In the dark-field microscope, oblique rays of light focused on the specimen do not enter the objective lens, but light that is scattered by components in the living cell can be collected to produce a bright image on a dark background. (C) Light passing through the unstained living cell experiences very little change in amplitude, and the structural details cannot be seen even if the image is highly magnified. The phase of the light, however, is altered by its passage through either thicker or denser parts of the cell, and small phase differences can be made visible by exploiting interference effects using a phase-contrast or a differential-interference-contrast microscope.

Phase-contrast, differential-interference-contrast, and dark-field microscopy make it possible to watch the movements involved in such processes as mitosis and cell migration. Since many cellular motions are too slow to be seen in real time, it is often helpful to make time-lapse movies in which the camera records successive frames separated by a short time delay, so that when the resulting picture series is played at normal speed, events appear greatly sped up.

### Images Can Be Enhanced and Analyzed by Digital Techniques

In recent years, electronic, or digital, imaging systems, and the associated technology of **image processing**, have had a major impact on light microscopy. Certain practical limitations of microscopes relating to imperfections in the optical system have been largely overcome. Electronic imaging systems have also circumvented two fundamental limitations of the human eye: the eye cannot see well in extremely dim light, and it cannot perceive small differences in light intensity against a bright background. To increase our ability to observe cells in these difficult conditions, we can attach a sensitive digital camera to a microscope. These cameras detect light by means of charge-coupled devices (CCDs), or high sensitivity complementary metal-oxide semiconductor (CMOS) sensors, similar to those found in digital cameras. Such image sensors are 10 times more sensitive than the human eye and can detect 100 times more intensity levels. It is therefore possible to observe cells for long periods at very low light levels, thereby avoiding the damaging effects of prolonged bright light (and heat). Such low-light cameras are especially important for viewing fluorescent molecules in living cells, as explained below.

Because images produced by digital cameras are in electronic form, they can be processed in various ways to extract latent information. Such image processing makes it possible to compensate for several optical faults in microscopes. Moreover, by digital image processing, contrast can be greatly enhanced to overcome



**Figure 9-8 Four types of light microscopy.** Four images are shown of the same fibroblast cell in culture. All images can be obtained with most modern microscopes by interchanging optical components. (A) Bright-field microscopy, in which light is transmitted straight through the specimen. (B) Phase-contrast microscopy, in which phase alterations of light transmitted through the specimen are translated into brightness changes. (C) Differential-interference-contrast microscopy, which highlights edges where there is a steep change of refractive index. (D) Dark-field microscopy, in which the specimen is lit from the side and only the scattered light is seen.

the eye's limitations in detecting small differences in light intensity, and background irregularities in the optical system can be digitally subtracted. This procedure reveals small transparent objects that were previously impossible to distinguish from the background.

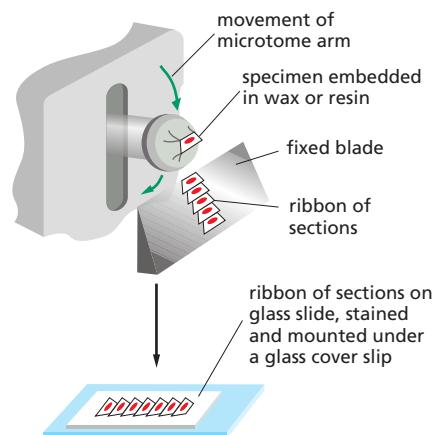
### Intact Tissues Are Usually Fixed and Sectioned Before Microscopy

Because most tissue samples are too thick for their individual cells to be examined directly at high resolution, they are often cut into very thin transparent slices, or *sections*. To preserve the cells within the tissue they must be treated with a *fixative*. Common fixatives include glutaraldehyde, which forms covalent bonds with the free amino groups of proteins, cross-linking them so they are stabilized and locked into position.

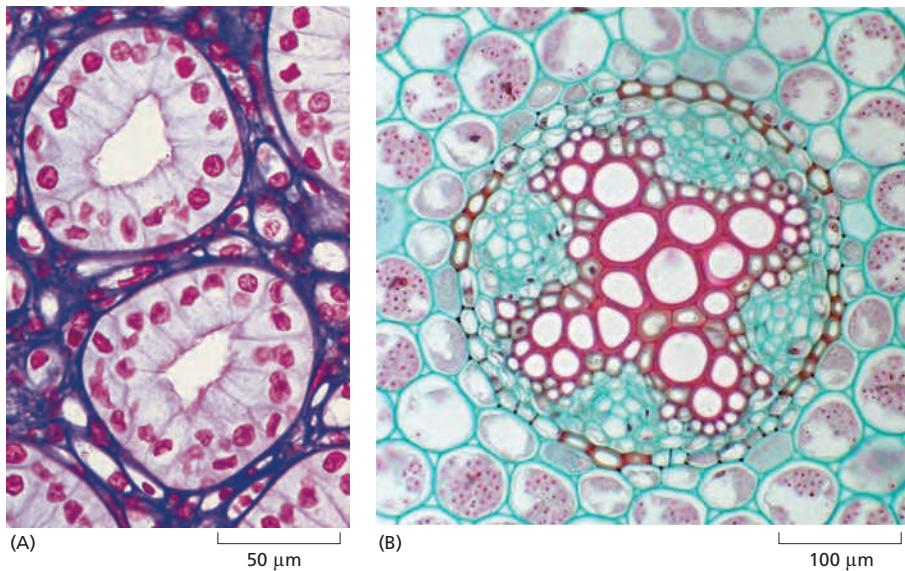
Because tissues are generally soft and fragile, even after fixation, they need to be either frozen or embedded in a supporting medium before being sectioned. The usual embedding media are waxes or resins. In liquid form, these media both permeate and surround the fixed tissue; they can then be hardened (by cooling or by polymerization) to form a solid block, which is readily sectioned with a microtome. This is a machine with a sharp blade, usually of steel or glass, which operates like a meat-slicer (Figure 9-9). The sections (typically 0.5–10  $\mu\text{m}$  thick) are then laid flat on the surface of a glass microscope slide.

There is little in the contents of most cells (which are 70% water by weight) to impede the passage of light rays. Thus, most cells in their natural state, even if fixed and sectioned, are almost invisible in an ordinary light microscope. We have seen that cellular components can be made visible by techniques such as phase-contrast and differential-interference-contrast microscopy, but these methods tell us almost nothing about the underlying chemistry. There are three main approaches to working with thin tissue sections that reveal differences in types of molecules that are present.

First, and traditionally, sections can be stained with organic dyes that have some specific affinity for particular subcellular components. The dye hematoxylin, for example, has an affinity for negatively charged molecules and therefore reveals the distribution of DNA, RNA, and acidic proteins in a cell (Figure 9-10). The chemical basis for the specificity of many dyes, however, is not known.



**Figure 9-9 Making tissue sections.** This illustration shows how an embedded tissue is sectioned with a microtome in preparation for examination in the light microscope.



**Figure 9–10** Staining of cell components.

(A) This section of cells in the urine-collecting ducts of the kidney was stained with hematoxylin and eosin, two dyes commonly used in histology. Each duct is made of closely packed cells (with nuclei stained red) that form a ring. The ring is surrounded by extracellular matrix, stained purple. (B) This section of a young plant root is stained with two dyes, safranin and fast green. The fast green stains the cellulosic cell walls while the safranin stains the lignified xylem cell walls bright red. (A, from P.R. Wheater et al., *Functional Histology*, 2nd ed. London: Churchill Livingstone, 1987; B, courtesy of Stephen Grace.)

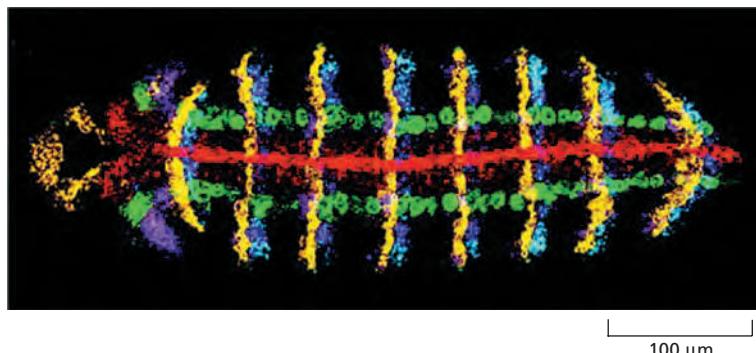
Second, sectioned tissues can be used to visualize specific patterns of differential gene expression. *In situ* hybridization, discussed earlier (see Figure 8–34), reveals the cellular distribution and abundance of specific expressed RNA molecules in sectioned material or in whole mounts of small organisms or organs. This is particularly effective when used in conjunction with fluorescent probes (Figure 9–11).

A third and very sensitive approach, generally and widely applicable for localizing proteins of interest, also depends on using fluorescent probes and markers, as we explain next.

### Specific Molecules Can Be Located in Cells by Fluorescence Microscopy

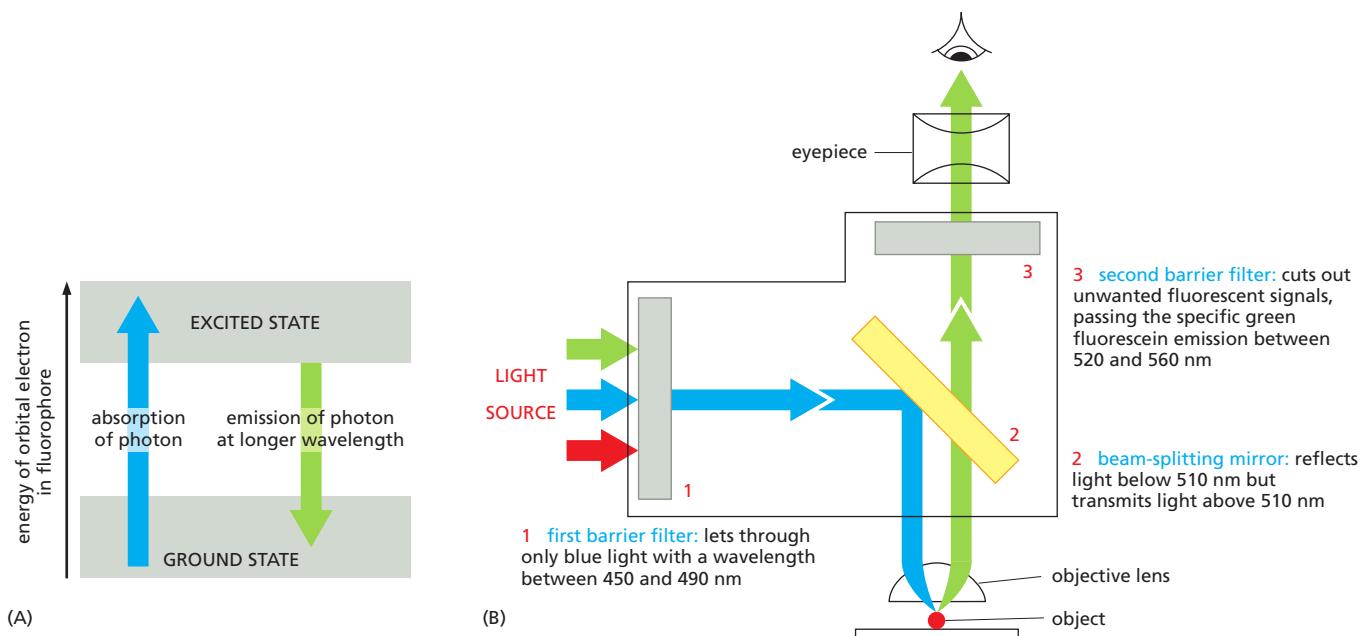
Fluorescent molecules absorb light at one wavelength and emit it at another, longer wavelength (Figure 9–12A). If we illuminate such a molecule at its absorbing wavelength and then view it through a filter that allows only light of the emitted wavelength to pass, it will glow against a dark background. Because the background is dark, even a minute amount of the glowing fluorescent dye can be detected. In contrast, the same number of molecules of a nonfluorescent stain, viewed conventionally, would be practically indiscernible because the absorption of light by molecules in the stain would result in only the faintest tinge of color in the light transmitted through that part of the specimen.

The fluorescent dyes used for staining cells are visualized with a **fluorescence microscope**. This microscope is similar to an ordinary light microscope except that the illuminating light, from a very powerful source, is passed through two sets of filters—one to filter the light before it reaches the specimen and one to



**Figure 9–11** RNA *in situ* hybridization.

As described in Chapter 8 (see Figure 8–62), it is possible to visualize the distribution of different RNAs in tissues using *in situ* hybridization. Here, the transcription pattern of five different genes involved in patterning the early fly embryo is revealed in a single embryo. Each RNA probe has been fluorescently labeled in a different way, some directly and some indirectly; the resulting images are displayed each in a different color ("false-colored") and combined to give an image where different color combinations represent different sets of genes expressed. The genes whose expression pattern is revealed here are *wingless* (yellow), *engrailed* (blue), *short gastrulation* (red), *intermediate neuroblasts defective* (green), and *muscle specific homeobox* (purple). (From D. Kosman et al., *Science* 305:846, 2004. With permission from AAAS.)

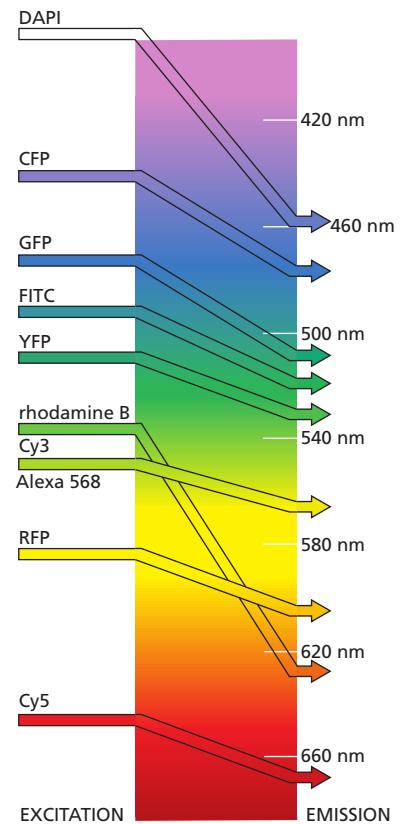


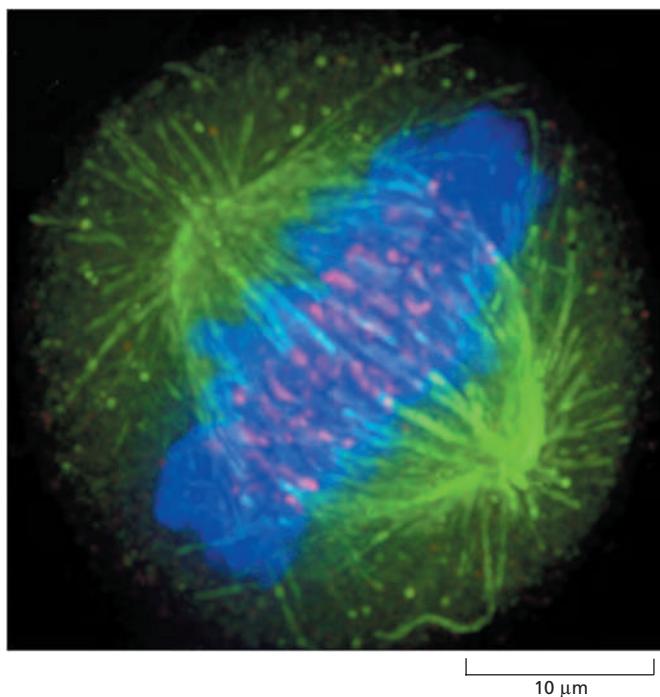
**Figure 9–12 Fluorescence and the fluorescence microscope.** (A) An orbital electron of a fluorochrome molecule can be raised to an excited state following the absorption of a photon. **Fluorescence** occurs when the electron returns to its ground state and emits a photon of light at a longer wavelength. Too much exposure to light, or too bright a light, can also destroy the fluorochrome molecule, in a process called *photobleaching*. (B) In the fluorescence microscope, a filter set consists of two barrier filters (1 and 3) and a dichroic (beam-splitting) mirror (2). This example shows the filter set for detection of the fluorescent molecule fluorescein. High-numerical-aperture objective lenses are especially important in this type of microscopy because, for a given magnification, the brightness of the fluorescent image is proportional to the fourth power of the numerical aperture (see also Figure 9–6).

filter the light obtained from the specimen. The first filter passes only the wavelengths that excite the particular fluorescent dye, while the second filter blocks out this light and passes only those wavelengths emitted when the dye fluoresces (Figure 9–12B).

Fluorescence microscopy is most often used to detect specific proteins or other molecules in cells and tissues. A very powerful and widely used technique is to couple fluorescent dyes to antibody molecules, which then serve as highly specific and versatile staining reagents that bind selectively to the particular macromolecules they recognize in cells or in the extracellular matrix. Two fluorescent dyes that have been commonly used for this purpose are *fluorescein*, which emits an intense green fluorescence when excited with blue light, and *rhodamine*, which emits deep red fluorescence when excited with green-yellow light (Figure 9–13). By coupling one antibody to fluorescein and another to rhodamine, the distributions of different molecules can be compared in the same cell; the two molecules are visualized separately in the microscope by switching back and forth between two sets of filters, each specific for one dye. As shown in Figure 9–14, three fluorescent dyes can be used in the same way to distinguish among three types of molecules in the same cell. Many newer fluorescent dyes, such as Cy3, Cy5, and the Alexa dyes, have been specifically developed for fluorescence microscopy

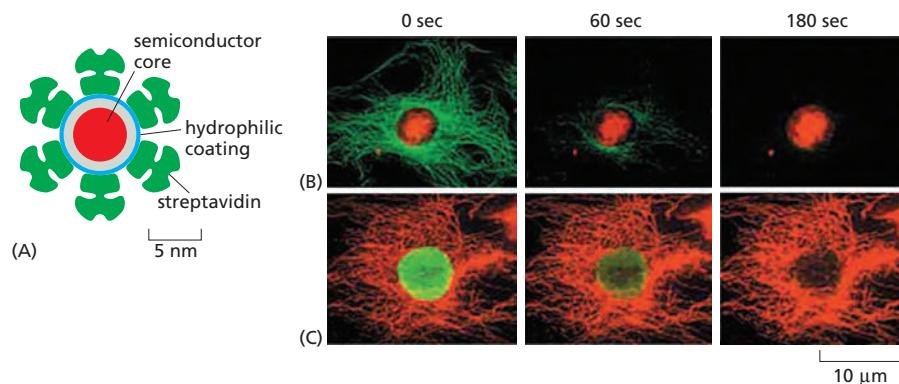
**Figure 9–13 Fluorescent probes.** The maximum excitation and emission wavelengths of several commonly used fluorescent probes are shown in relation to the corresponding colors of the spectrum. The photon emitted by a fluorescent molecule is necessarily of lower energy (longer wavelength) than the absorbed photon and this accounts for the difference between the excitation and emission peaks. CFP, GFP, YFP, and RFP are cyan, green, yellow, and red fluorescent proteins, respectively. DAPI is widely used as a general fluorescent DNA probe, which absorbs ultraviolet light and fluoresces bright blue. FITC is an abbreviation for fluorescein isothiocyanate, a widely used derivative of fluorescein, which fluoresces bright green. The other probes are all commonly used to fluorescently label antibodies and other proteins. The use of fluorescent proteins will be discussed later in the chapter.



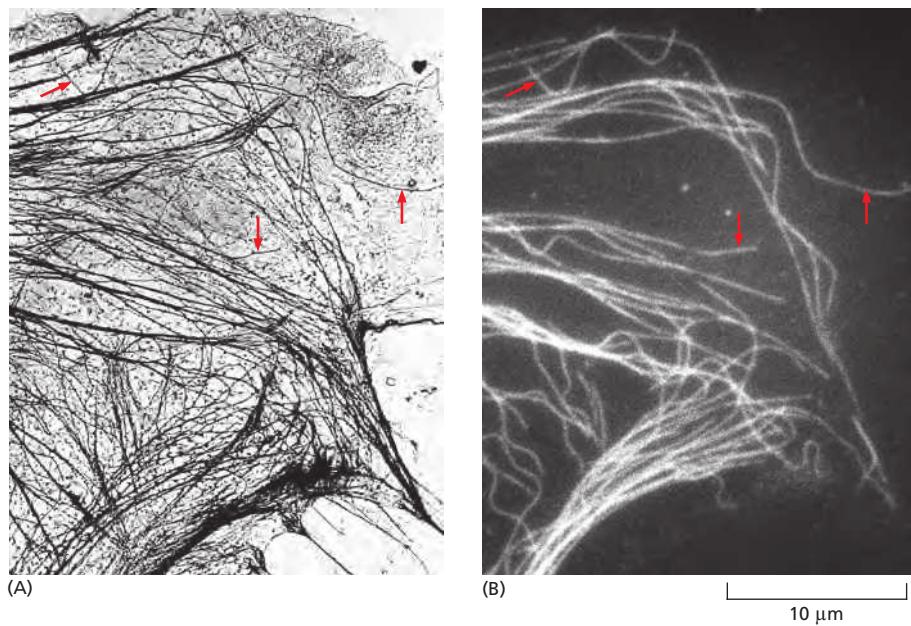


**Figure 9–14** Different fluorescent probes can be visualized in the same cell. In this composite micrograph of a cell in mitosis, three different fluorescent probes have been used to label three different cellular components (**Movie 9.1**). The spindle microtubules are revealed with a green fluorescent antibody, centromeres with a red fluorescent antibody, and the DNA of the condensed chromosomes with the blue fluorescent dye DAPI. (Courtesy of Kevin F. Sullivan.)

(see Figure 9–13) but, like many organic fluorochromes, they fade fairly rapidly when continuously illuminated. More stable fluorochromes have been developed based on inorganic chemistry. Tiny crystals of semiconductor material, called nanoparticles, or *quantum dots*, can be excited to fluoresce by a broad spectrum of blue light. Their emitted light has a color that depends on the exact size of the nanocrystal, between 2 and 10 nm in diameter, and additionally the fluorescence fades only slowly with time (Figure 9–15). These nanoparticles, when coupled to other probes such as antibodies, are therefore ideal for tracking molecules over time. If introduced into a living cell, in an embryo for example, the progeny of that cell can be followed many days later by their fluorescence, allowing cell lineages to be tracked.



**Figure 9–15** Fluorescent nanoparticles or quantum dots. (A) Quantum dots are tiny particles of cadmium selenide, a semiconductor, with a coating to make them water-soluble. They can be coupled to protein molecules such as antibodies or streptavidin and, when introduced into a cell, will bind to a target protein of interest. Different-sized quantum dots emit light of different colors—the larger the dot, the longer the wavelength—but they are all excited by the same blue light. Quantum dots can keep shining for weeks, unlike most fluorescent organic dyes. (B) In this cell, microtubules are labeled (green) with an organic fluorescent dye (Alexa 488), while a nuclear protein is stained (red) with quantum dots bound to streptavidin. On continuous exposure to strong blue light, the fluorescent dye fades quickly while the quantum dots continue to shine. (C) In this cell, the labeling pattern is reversed; a nuclear protein is labeled (green) with an organic fluorescent dye (Alexa 488), while microtubules are labeled (red) with quantum dots. Again, the quantum dots far outlast the fluorescent dye. (B and C, from L. Medintz et al., *Nat. Mater.* 4:435–446, 2005. With permission from Macmillan Publishers Ltd.)

**Figure 9-16 Immunofluorescence.**

(A) A transmission electron micrograph of the periphery of a cultured epithelial cell showing the distribution of microtubules and other filaments. (B) The same area stained with fluorescent antibodies against tubulin, the protein that assembles to form microtubules, using the technique of indirect immunocytochemistry (see Figure 9-17). Red arrows indicate individual microtubules that are readily recognizable in both images. Note that, because of diffraction effects, the microtubules in the light microscope appear 0.2  $\mu\text{m}$  wide rather than their true width of 0.025  $\mu\text{m}$ . (From M. Osborn, R. Webster and K. Weber, *J. Cell Biol.* 77:R27–R34, 1978. With permission from The Rockefeller University Press.)

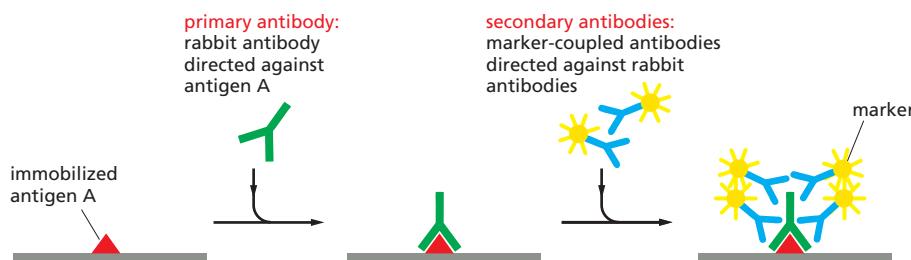
Later in the chapter, additional fluorescence microscopy methods will be discussed that can be used to monitor changes in the concentration and location of specific molecules inside *living* cells.

### Antibodies Can Be Used to Detect Specific Molecules

Antibodies are proteins produced by the vertebrate immune system as a defense against infection (discussed in Chapter 24). They are unique among proteins in that they are made in billions of different forms, each with a different binding site that recognizes a specific target molecule (or *antigen*). The precise antigen specificity of antibodies makes them powerful tools for the cell biologist. When labeled with fluorescent dyes, antibodies are invaluable for locating specific molecules in cells by fluorescence microscopy (Figure 9-16); labeled with electron-dense particles such as colloidal gold spheres, they are used for similar purposes in the electron microscope (discussed below). The antibodies employed in microscopy are commonly either purified from antiserum so as to remove all nonspecific antibodies, or they are specific monoclonal antibodies that only recognize the target molecule.

When we use antibodies as probes to detect and assay specific molecules in cells, we frequently use chemical methods to amplify the fluorescent signal they produce. For example, although a marker molecule such as a fluorescent dye can be linked directly to an antibody—the *primary antibody*—a stronger signal is achieved by using an unlabeled primary antibody and then detecting it with a group of labeled *secondary antibodies* that bind to it (Figure 9-17). This process is called *indirect immunocytochemistry*.

Some amplification methods use an enzyme as a marker molecule attached to the secondary antibody. The enzyme alkaline phosphatase, for example, in the presence of appropriate chemicals, produces inorganic phosphate that in turn

**Figure 9-17 Indirect immunocytochemistry.**

This detection method is very sensitive because many molecules of the secondary antibody recognize each primary antibody. The secondary antibody is covalently coupled to a marker molecule that makes it readily detectable. Commonly used marker molecules include fluorescent probes (for fluorescence microscopy), the enzyme horseradish peroxidase (for either conventional light microscopy or electron microscopy), colloidal gold spheres (for electron microscopy), and the enzymes alkaline phosphatase or peroxidase (for biochemical detection).

leads to the local formation of a colored precipitate. This reveals the location of the secondary antibody and hence the location of the antibody–antigen complex. Since each enzyme molecule acts catalytically to generate many thousands of molecules of product, even tiny amounts of antigen can be detected. Although the enzyme amplification makes enzyme-linked methods sensitive, diffusion of the colored precipitate away from the enzyme limits the spatial resolution of this method for microscopy, and fluorescent labels are usually used for the most sensitive and precise optical localization.

### Imaging of Complex Three-Dimensional Objects Is Possible with the Optical Microscope

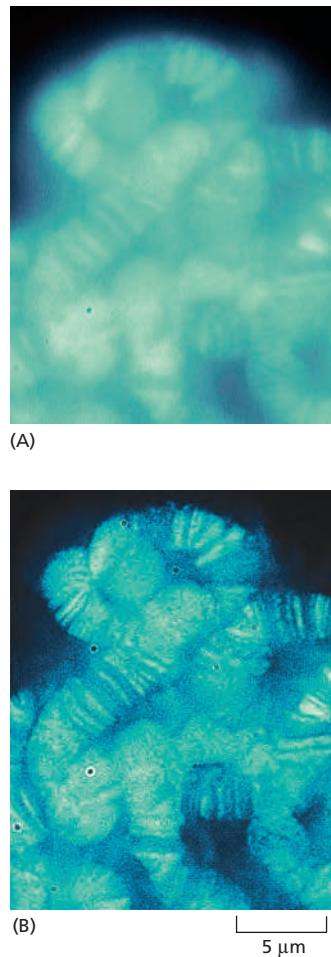
For ordinary light microscopy, as we have seen, a tissue has to be sliced into thin sections to be examined; the thinner the section, the crisper the image. Since information about the third dimension is lost upon sectioning, how, then, can we get a picture of the three-dimensional architecture of a cell or tissue, and how can we view the microscopic structure of a specimen that, for one reason or another, cannot first be sliced into sections? Although an optical microscope is focused on a particular focal plane within a three-dimensional specimen, all the other parts of the specimen, above and below the plane of focus, are also illuminated and the light originating from these regions contributes to the image as “out-of-focus” blur. This can make it very hard to interpret the image in detail and can lead to fine image structure being obscured by the out-of-focus light.

Two distinct but complementary approaches solve this problem: one is computational, the other optical. These three-dimensional microscopic imaging methods make it possible to focus on a chosen plane in a thick specimen while rejecting the light that comes from out-of-focus regions above and below that plane. Thus one sees a crisp, thin *optical section*. From a series of such optical sections taken at different depths and stored in a computer, a three-dimensional image can be reconstructed. The methods do for the microscopist what the computed tomography (CT) scanner does (by different means) for the radiologist investigating a human body: both machines give detailed sectional views of the interior of an intact structure.

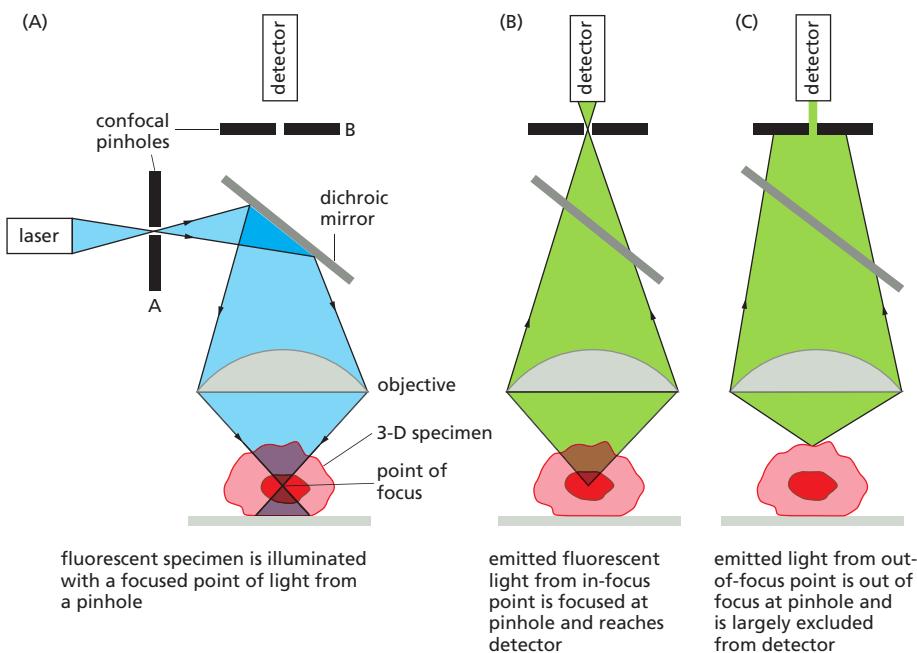
The computational approach is often called *image deconvolution*. To understand how it works, remember that the wavelike nature of light means that the microscope lens system produces a small blurred disc as the image of a point light source (see Figure 9–5), with increased blurring if the point source lies above or below the focal plane. This blurred image of a point source is called the *point spread function* (see Figure 9–36). An image of a complex object can then be thought of as being built up by replacing each point of the specimen by a corresponding blurred disc, resulting in an image that is blurred overall. For deconvolution, we first obtain a series of (blurred) images, usually with a cooled CCD camera or more recently a CMOS camera, focusing the microscope in turn on a series of focal planes—in effect, a (blurred) three-dimensional image. Digital processing of the stack of digital images then removes as much of the blur as possible. In essence, the computer program uses the measured point spread function of a point source of light from that microscope to determine what the effect of the blurring would have been on the image, and then applies an equivalent “deblurring” (deconvolution), turning the blurred three-dimensional image into a series of clean optical sections, albeit still constrained by the diffraction limit. **Figure 9–18** shows an example.

### The Confocal Microscope Produces Optical Sections by Excluding Out-of-Focus Light

The confocal microscope achieves a result similar to that of deconvolution, but does so by manipulating the light before it is measured; it is an analog technique rather than a digital one. The optical details of the **confocal microscope** are complex, but the basic idea is simple, as illustrated in **Figure 9–19**, and the results are



**Figure 9–18** Image deconvolution.  
 (A) A light micrograph of the large polytene chromosomes from *Drosophila*, stained with a fluorescent DNA-binding dye.  
 (B) The same field of view after image deconvolution clearly reveals the banding pattern on the chromosomes. Each band is about 0.25  $\mu\text{m}$  thick, approaching the diffraction limit of the light microscope.  
 (Courtesy of the John Sedat Laboratory.)

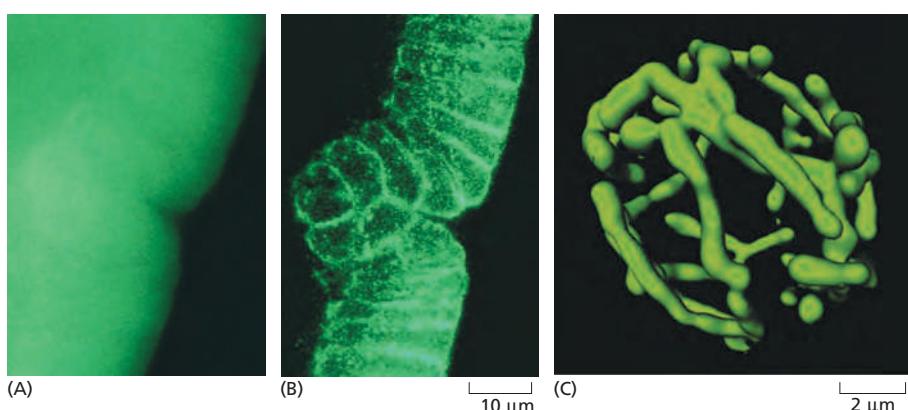


**Figure 9–19 The confocal fluorescence microscope.** (A) This simplified diagram shows that the basic arrangement of optical components is similar to that of the standard fluorescence microscope shown in Figure 9–12, except that a laser is used to illuminate a small pinhole whose image is focused at a single point in the three-dimensional (3-D) specimen. (B) Emitted fluorescence from this focal point in the specimen is focused at a second (confocal) pinhole. (C) Emitted light from elsewhere in the specimen is not focused at the pinhole and therefore does not contribute to the final image. By scanning the beam of light across the specimen, a very sharp two-dimensional image of the exact plane of focus is built up that is not significantly degraded by light from other regions of the specimen.

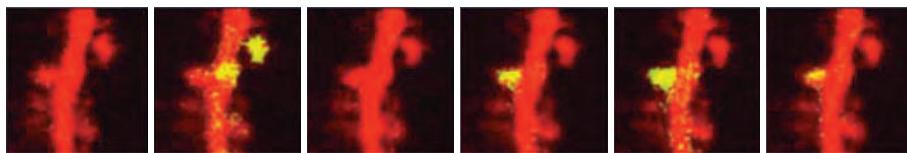
far superior to those obtained by conventional light microscopy (Figure 9–20A and B).

The confocal microscope is generally used with fluorescence optics (see Figure 9–12), but instead of illuminating the whole specimen at once, in the usual way, the optical system at any instant focuses a spot of light onto a single point at a specific depth in the specimen. This requires a source of pinpoint illumination that is usually supplied by a laser whose light has been passed through a pinhole. The fluorescence emitted from the illuminated material is collected at a suitable light detector and used to generate an image. A pinhole aperture is placed in front of the detector, at a position that is *confocal* with the illuminating pinhole—that is, precisely where the rays emitted from the illuminated point in the specimen come to a focus. Thus, the light from this point in the specimen converges on this aperture and enters the detector.

By contrast, the light from regions out of the plane of focus of the spotlight is also out of focus at the pinhole aperture and is therefore largely excluded from the detector (see Figure 9–19). To build up a two-dimensional image, data from each point in the plane of focus are collected sequentially by scanning across the field from left to right in a regular pattern of pixels and are displayed on a computer screen. Although not shown in Figure 9–19, the scanning is usually done by deflecting the beam with an oscillating mirror placed between the dichroic mirror and the objective lens in such a way that the illuminating spotlight and the



**Figure 9–20 Confocal fluorescence microscopy produces clear optical sections and three-dimensional data sets.** The first two micrographs are of the same intact gastrula-stage *Drosophila* embryo, which has been stained with a fluorescent probe for actin filaments. (A) The conventional, unprocessed image is blurred by the presence of fluorescent structures above and below the plane of focus. (B) In the confocal image, this out-of-focus information is removed, resulting in a crisp optical section of the cells in the embryo. (C) A three-dimensional reconstruction of an object can be assembled from a stack of such optical sections. In this case, the complex branching structure of the mitochondrial compartment in a single live yeast cell is shown. (A and B, courtesy of Richard Warn and Peter Shaw; C, courtesy of Stefan Hell.)



**Figure 9–21 Multiphoton imaging.** Infrared laser light causes less damage to living cells than visible light and can also penetrate further, allowing microscopists to peer deeper into living tissues. The two-photon effect, in which a fluorochrome can be excited by two coincident infrared photons instead of a single high-energy photon, allows us to see nearly 0.5 mm inside the cortex of a live mouse brain. A dye, whose fluorescence changes with the calcium concentration, reveals active synapses (yellow) on the dendritic spines (red) that change as a function of time; in this case, there is a day between each image. (Courtesy of Thomas Oertner and Karel Svoboda.)

confocal pinhole at the detector remain strictly in register. Variations in design now allow the rapid collection of data at video rates.

The confocal microscope has been used to resolve the structure of numerous complex three-dimensional objects (Figure 9–20C) including the networks of cytoskeletal fibers in the cytoplasm and the arrangements of chromosomes and genes in the nucleus.

The relative merits of deconvolution methods and confocal microscopy for three-dimensional optical microscopy depend on the specimen being imaged. Confocal microscopes tend to be better for thicker specimens with high levels of out-of-focus light. They are also generally easier to use than deconvolution systems and the final optical sections can be seen quickly. In contrast, the cooled CCD or CMOS cameras used for deconvolution systems are extremely efficient at collecting small amounts of light, and they can be used to make detailed three-dimensional images from specimens that are too weakly stained or too easily damaged by the bright light used for confocal microscopy.

Both methods, however, have another drawback; neither is good at coping with very thick specimens. Deconvolution methods quickly become ineffective any deeper than about 40  $\mu\text{m}$  into a specimen, while confocal microscopes can only obtain images up to a depth of about 150  $\mu\text{m}$ . Special microscopes can now take advantage of the way in which fluorescent molecules are excited, to probe even deeper into a specimen. Fluorescent molecules are usually excited by a single high-energy photon, of shorter wavelength than the emitted light, but they can in addition be excited by the absorption of two (or more) photons of lower energy, as long as they both arrive within a femtosecond or so of each other. The use of this longer-wavelength excitation has some important advantages. In addition to reducing background noise, red or near-infrared light can penetrate deeper into a specimen. Multiphoton microscopes, constructed to take advantage of this *two-photon* effect, can obtain sharp images, sometimes even at a depth of 250  $\mu\text{m}$  within a specimen. This is particularly valuable for studies of living tissues, notably in imaging the dynamic activity of synapses and neurons just below the surface of living brains (**Figure 9–21**).

### Individual Proteins Can Be Fluorescently Tagged in Living Cells and Organisms

Even the most stable cell structures must be assembled, disassembled, and reorganized during the cell's life cycle. Other structures, often enormous on the molecular scale, rapidly change, move, and reorganize themselves as the cell conducts its internal affairs and responds to its environment. Complex, highly organized pieces of molecular machinery move components around the cell, controlling traffic into and out of the nucleus, from one organelle to another, and into and out of the cell itself.

Various techniques have been developed to visualize the specific components involved in such dynamic phenomena. Many of these methods use fluorescent

proteins, and they require a trade-off between structural preservation and efficient labeling. All of the fluorescent molecules discussed so far are made outside the cell and then artificially introduced into it. But use of genes coding for protein molecules that are themselves inherently fluorescent also enables the creation of organisms and cell lines that make their own visible tags and labels, without the introduction of foreign molecules. These cellular exhibitionists display their inner workings in glowing fluorescent color.

Foremost among the fluorescent proteins used for these purposes by cell biologists is the **green fluorescent protein (GFP)**, isolated from the jellyfish *Aequorea victoria*. This protein is encoded by a single gene, which can be cloned and introduced into cells of other species. The freshly translated protein is not fluorescent, but within an hour or so (less for some alleles of the gene, more for others) it undergoes a self-catalyzed post-translational modification to generate an efficient fluorochrome, shielded within the interior of a barrel-like protein, which will now fluoresce when illuminated appropriately with blue light (Figure 9–22). Extensive site-directed mutagenesis performed on the original gene sequence has resulted in multiple variants that can be used effectively in organisms ranging from animals and plants to fungi and microbes. The fluorescence efficiency has also been improved, and variants have been generated with altered absorption and emission spectra from the blue-green, like blue fluorescent protein or BFP, to the far visible red. Other, related fluorescent proteins have since been discovered (for example, in corals) that also extend the range into the red region of the spectrum, like red fluorescent protein or RFP.

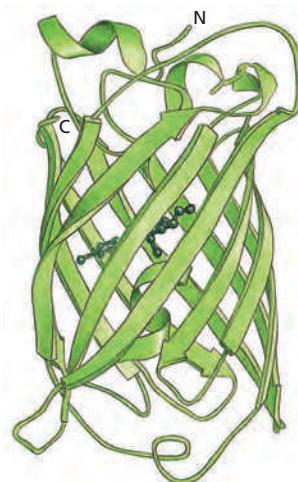
One of the simplest uses of GFP is as a reporter molecule, a fluorescent probe to monitor gene expression. A transgenic organism can be made with the GFP-coding sequence placed under the transcriptional control of the promoter belonging to a gene of interest, giving a directly visible readout of the gene's expression pattern in the living organism (Figure 9–23). In another application, a peptide location signal can be added to the GFP to direct it to a particular cell compartment, such as the endoplasmic reticulum or a mitochondrion, lighting up these organelles so they can be observed in the living state (see Figure 12–31).

The GFP DNA coding sequence can also be inserted at the beginning or end of the gene for another protein, yielding a chimeric product consisting of that protein with a GFP domain attached. In many cases, this GFP fusion protein behaves in the same way as the original protein, directly revealing its location and activities by means of its genetically encoded fluorescence (Figure 9–24). It is often possible to prove that the GFP fusion protein is functionally equivalent to the untagged protein, for example by using it to rescue a mutant lacking that protein. GFP tagging is the clearest and most unequivocal way of showing the distribution and dynamics of a protein in a living organism (Figure 9–25 and see Movie 16.8).

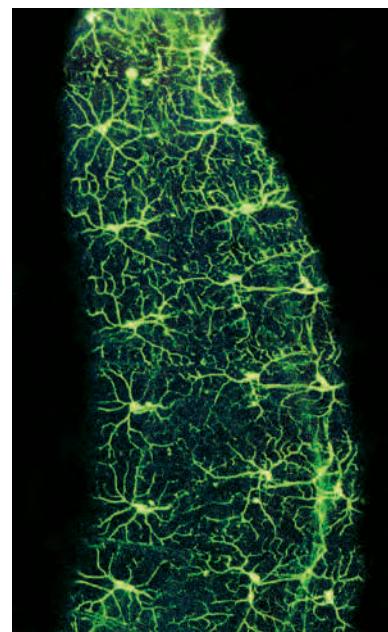
### Protein Dynamics Can Be Followed in Living Cells

Fluorescent proteins are now exploited not just to see where in a cell a particular protein is located, but also to uncover its kinetic properties and to find out whether it might interact with other molecules. We now describe three techniques in which fluorescent proteins are used in this way.

First, interactions between one protein and another can be monitored by **fluorescence resonance energy transfer**, also called **Förster resonance energy**

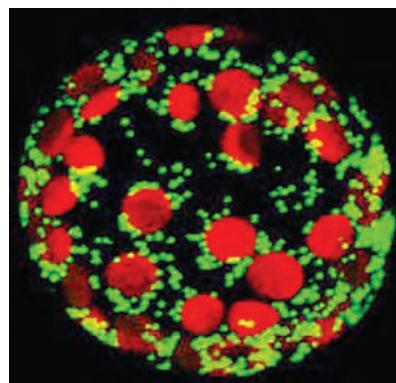


**Figure 9–22 Green fluorescent protein (GFP).** The structure of GFP, shown here schematically, highlights the eleven  $\beta$  strands that form the staves of a barrel. Buried within the barrel is the active chromophore (dark green) that is formed post-translationally from the protruding side chains of three amino acid residues. (From M. Ormö et al., *Science* 273:1392–1395, 1996. With permission from AAAS.)



**Figure 9–23 Green fluorescent protein (GFP) as a reporter.** For this experiment, carried out in the fruit fly, the GFP gene was joined (using recombinant DNA techniques) to a fly promoter that is active only in a specialized set of neurons. This image of a live fly embryo was captured by a fluorescence microscope and shows approximately 20 neurons, each with long projections (axons and dendrites) that communicate with other (nonfluorescent) cells. These neurons are located just under the surface of the animal and allow it to sense its immediate environment. (From W.B. Grueber et al., *Curr. Biol.* 13:618–626, 2003. With permission from Elsevier.)

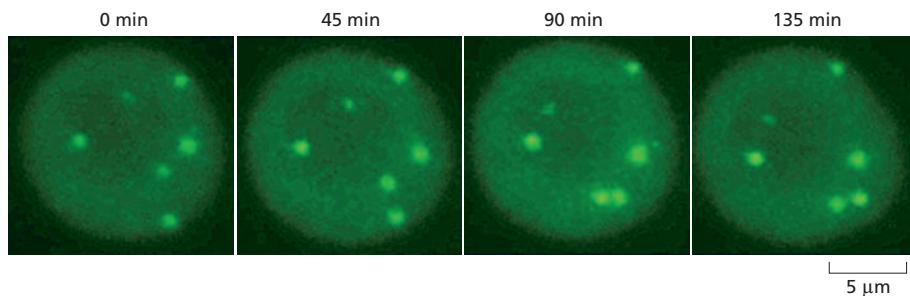
**Figure 9–24 GFP-tagged proteins.** This living cell from a tobacco plant is expressing high levels of green fluorescent protein, fused to a protein that is targeted to mitochondria, which accordingly appear green. The mitochondria are seen to cluster around the chloroplasts, whose chlorophyll autofluorescence marks them out in red. (Courtesy of Olivier Grandjean.)



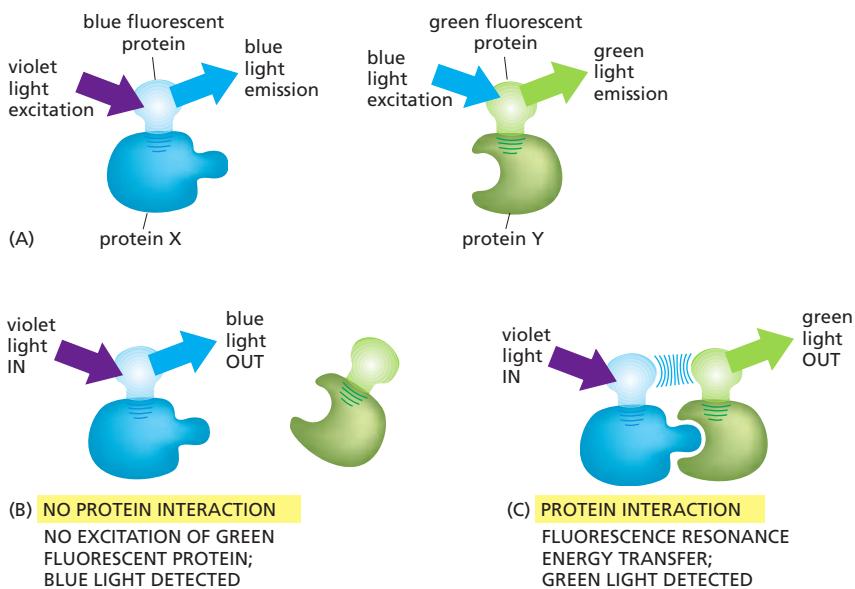
transfer, both abbreviated FRET. In this technique, two molecules of interest are each labeled with a different fluorochrome, chosen so that the emission spectrum of one fluorochrome, the donor, overlaps with the absorption spectrum of the other, the acceptor. If the two proteins bind so as to bring their fluorochromes into very close proximity (closer than about 5 nm), one fluorochrome, when excited, can transfer energy from the absorbed light directly (by resonance, nonradiatively) to the other. Thus, when the complex is illuminated at the excitation wavelength of the first fluorochrome, fluorescent light is produced at the emission wavelength of the second. This method can be used with two different spectral variants of GFP as fluorochromes to monitor processes such as the interaction of signaling molecules with their receptors, or proteins in macromolecular complexes at specific locations inside living cells (Figure 9–26). The FRET can be measured by quantifying the reduction of the donor fluorescence in the presence of the acceptor.

A second example of a fluorescence-tagging technique that allows detailed observations of proteins within cells involves synthesizing an inactive form of the fluorescent molecule of interest, introducing it into the cell, and then activating it suddenly at a chosen site in the cell by focusing a spot of light on it. This process is referred to as **photoactivation**. Many inactive photosensitive precursors of this type, often called *caged molecules*, have been made based on a variety of fluorescent molecules. A microscope can be used to focus a strong pulse of light from a laser on any tiny region of the cell, so that the experimenter can control exactly where and when the fluorescent molecule is photoactivated. The technique allows us to follow complex and rapid intracellular processes, such as the actions of signaling molecules or the movements of cytoskeletal proteins.

When a photoactivatable fluorescent tag is attached to a purified protein, it is important that the modified protein remain biologically active: labeling with a caged fluorescent dye adds a bulky group to the surface of a protein, which can easily change the protein's properties. A satisfactory labeling protocol is usually found by trial and error. Once a biologically active labeled protein has been produced, it needs to be introduced into the living cell where its behavior can be followed. Tubulin labeled with caged fluorescein, for example, can be injected into a dividing cell, where it is incorporated into microtubules of the mitotic spindle. When a small region of the spindle is illuminated with a laser, the labeled tubulin



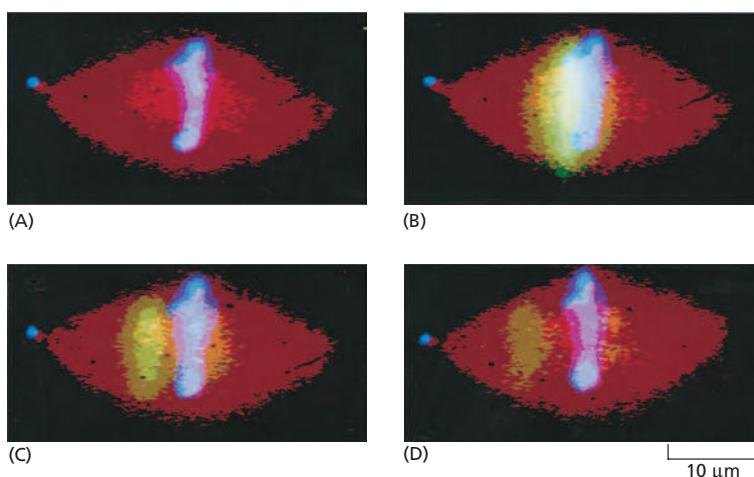
**Figure 9–25 Dynamics of GFP tagging.** This sequence of micrographs shows a set of three-dimensional images of a living nucleus taken over the course of 135 minutes. Tobacco cells have been stably transformed with GFP fused to a spliceosomal protein that is concentrated in small nuclear bodies called Cajal bodies (see Figure 6–46). The fluorescent Cajal bodies, easily visible in a living cell with confocal microscopy, are dynamic structures that move around within the nucleus. (Courtesy of Kurt Boudonck, Liam Dolan, and Peter Shaw.)



becomes fluorescent, so that its movement along the spindle microtubules can be readily followed (**Figure 9–27**).

A further development in photoactivation is the discovery that the genes encoding GFP and related fluorescent proteins can be engineered to produce protein variants, usually with one or more amino acid changes, that fluoresce only weakly under normal excitation conditions, but can be induced to fluoresce either more strongly or with a color shift (for example, from green to red) by activating them with a strong pulse of light at a different wavelength. In principle, the microscopist can then follow the local *in vivo* behavior of any protein that can be expressed as a fusion with one of these GFP variants. These genetically encoded, photoactivatable fluorescent proteins allow the lifetime and behavior of any protein to be studied independently of other newly synthesized proteins (**Figure 9–28**).

A third way to exploit GFP fused to a protein of interest is known as **fluorescence recovery after photobleaching (FRAP)**. Here, one uses a strong focused beam of light from a laser to extinguish the GFP fluorescence in a specified region of the cell, after which one can analyze the way in which remaining unbleached fluorescent protein molecules move into the bleached area as a function of time.

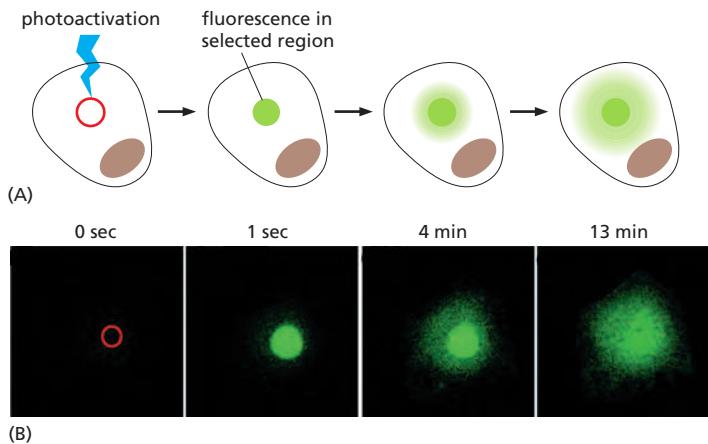


**Figure 9–26 Fluorescence resonance energy transfer (FRET).** To determine whether (and when) two proteins interact inside a cell, the proteins are first produced as fusion proteins attached to different color variants of green fluorescent protein (GFP). (A) In this example, protein X is coupled to a blue fluorescent protein, which is excited by violet light (370–440 nm) and emits blue light (440–480 nm); protein Y is coupled to a green fluorescent protein, which is excited by blue light (440–480 nm) and emits green light (510 nm). (B) If protein X and Y do not interact, illuminating the sample with violet light yields fluorescence from the blue fluorescent protein only. (C) When protein X and protein Y interact, the resonance transfer of energy, FRET, can now occur. Illuminating the sample with violet light excites the blue fluorescent protein, which transfers its energy to the green fluorescent protein, resulting in an emission of green light. The fluorochromes must be quite close together—within about 1–5 nm of one another—for FRET to occur. Because not every molecule of protein X and protein Y is bound at all times, some blue light may still be detected. But as the two proteins begin to interact, emission from the donor blue fluorescent protein falls as the emission from the acceptor GFP rises.

**Figure 9–27 Determining microtubule flux in the mitotic spindle with caged fluorescein linked to tubulin.**

(A) A metaphase spindle formed *in vitro* from an extract of *Xenopus* eggs has incorporated three fluorescent markers: rhodamine-labeled tubulin (red) to mark all the microtubules, a blue DNA-binding dye that labels the chromosomes, and caged-fluorescein-labeled tubulin, which is also incorporated into all the microtubules but is invisible because it is nonfluorescent until activated by ultraviolet (UV) light.

(B) A beam of UV light activates, or “uncages,” the caged-fluorescein-labeled tubulin locally, mainly just to the left side of the metaphase plate. Over the next few minutes—after 1.5 minutes in (C) and after 2.5 minutes in (D)—the uncaged-fluorescein-tubulin signal moves toward the left spindle pole, indicating that tubulin is continuously moving poleward even though the spindle (visualized by the red rhodamine-labeled tubulin fluorescence) remains largely unchanged. (From K.E. Sawin and T.J. Mitchison, *J. Cell Biol.* 112:941–954, 1991. With permission from The Rockefeller University Press.)

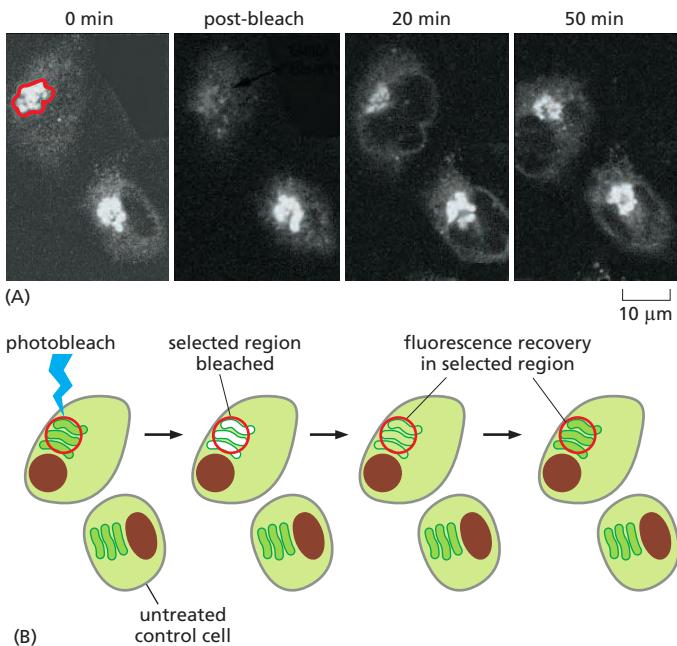
**Figure 9–28 Photoactivation.**

Photoactivation is the light-induced activation of an inert molecule to an active state. In this experiment, shown schematically in (A), a photoactivatable variant of GFP is expressed in a cultured animal cell. (B) Before activation (time 0 sec), little or no GFP fluorescence is detected in the selected region (red circle) when excited by blue light at 488 nm. After activation of the GFP with an ultraviolet laser pulse at 413 nm, it rapidly fluoresces brightly in the selected region (green). The movement of GFP, as it diffuses out of this region, can be measured. Since only the photoactivated proteins are fluorescent within the cell, the trafficking, turnover, and degradative pathways of proteins can be monitored. (B, from J. Lippincott-Schwartz and G.H. Patterson, *Science* 300:87–91, 2003.)

This technique is usually carried out with a confocal microscope and, like photoactivation, can deliver valuable quantitative data about a protein's kinetic parameters, such as diffusion coefficients, active transport rates, or binding and dissociation rates from other proteins (Figure 9–29).

### Light-Emitting Indicators Can Measure Rapidly Changing Intracellular Ion Concentrations

One way to study the chemistry of a single living cell is to insert the tip of a fine, glass, ion-sensitive **microelectrode** directly into the cell interior through the plasma membrane. This technique is used to measure the intracellular concentrations of common inorganic ions, such as  $H^+$ ,  $Na^+$ ,  $K^+$ ,  $Cl^-$ , and  $Ca^{2+}$ . However, ion-sensitive microelectrodes reveal the ion concentration only at one point in a cell, and for an ion present at a very low concentration, such as  $Ca^{2+}$ , their responses are slow and somewhat erratic. Thus, these microelectrodes are not ideally suited to record the rapid and transient changes in the concentration of cytosolic  $Ca^{2+}$  that have an important role in allowing cells to respond to extracellular signals. Such changes can be analyzed with **ion-sensitive indicators**, whose light emission reflects the local concentration of the ion. Some of these indicators are luminescent (emitting light spontaneously), while others are fluorescent (emitting light on exposure to light).

**Figure 9–29 Fluorescence recovery**

**after photobleaching (FRAP).** A strong focused pulse of laser light will extinguish, or bleach, the fluorescence of GFP. By selectively photobleaching a set of fluorescently tagged protein molecules within a defined region of a cell, the microscopist can monitor recovery over time, as the remaining fluorescent molecules move into the bleached region (see Movie 10.6). (A) The experiment shown uses monkey cells in culture that express galactosyltransferase, an enzyme that constantly recycles between the Golgi apparatus and the endoplasmic reticulum (ER). The Golgi apparatus in one of the two cells is selectively photobleached, while the production of new fluorescent protein is blocked by treating the cells with cycloheximide. The recovery, resulting from fluorescent enzyme molecules moving from the ER to the Golgi, can then be followed over a period of time. (B) Schematic diagram of the experiment shown in (A). (A, from J. Lippincott-Schwartz, *Histochem. Cell Biol.* 116:97–107, 2001. With permission from Springer-Verlag.)

**Figure 9-30 Aequorin, a luminescent protein.** The luminescent protein aequorin emits blue light in the presence of free  $\text{Ca}^{2+}$ . Here, an egg of the medaka fish has been injected with aequorin, which has diffused throughout the cytosol, and the egg has then been fertilized with a sperm and examined with the help of a very sensitive camera. The four photographs were taken looking down on the site of sperm entry at intervals of 10 seconds and reveal a wave of release of free  $\text{Ca}^{2+}$  into the cytosol from internal stores just beneath the plasma membrane. This wave sweeps across the egg starting from the site of sperm entry, as indicated in the diagrams on the left. (Photographs reproduced from J.C. Gilkey, L.F. Jaffe, E.B. Ridgway and G.T. Reynolds, *J. Cell Biol.* 76:448–466, 1978. With permission from The Rockefeller University Press.)

Aequorin is a luminescent protein isolated from the same marine jellyfish that produces GFP. It emits blue light in the presence of  $\text{Ca}^{2+}$  and responds to changes in  $\text{Ca}^{2+}$  concentration in the range of 0.5–10  $\mu\text{M}$ . If microinjected into an egg, for example, aequorin emits a flash of light in response to the sudden localized release of free  $\text{Ca}^{2+}$  into the cytoplasm that occurs when the egg is fertilized (Figure 9-30). Aequorin has also been expressed transgenically in plants and other organisms to provide a method of monitoring  $\text{Ca}^{2+}$  in all their cells without the need for microinjection, which can be a difficult procedure.

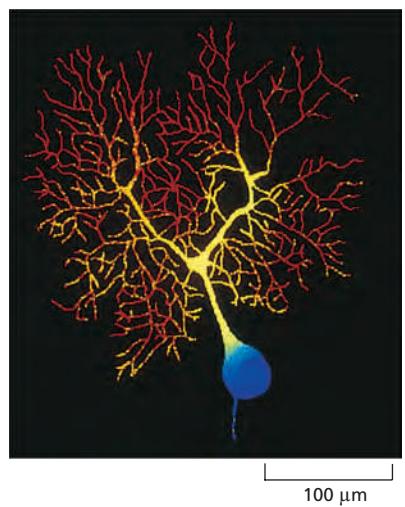
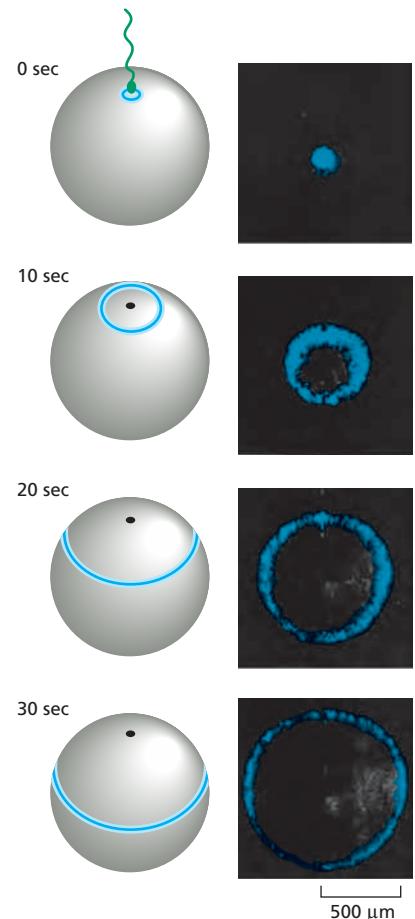
Bioluminescent molecules like aequorin emit tiny amounts of light—at best, a few photons per indicator molecule—that are difficult to measure. Fluorescent indicators produce orders of magnitude more photons per molecule; they are therefore easier to measure and can give better spatial resolution. Genetically encoded fluorescent  $\text{Ca}^{2+}$  indicators have been synthesized that bind  $\text{Ca}^{2+}$  tightly and are excited by or emit light at slightly different wavelengths when they are free of  $\text{Ca}^{2+}$  than when they are in their  $\text{Ca}^{2+}$ -bound form. By measuring the ratio of fluorescence intensity at two excitation or emission wavelengths, we can determine the concentration ratio of the  $\text{Ca}^{2+}$ -bound indicator to the  $\text{Ca}^{2+}$ -free indicator, thereby providing an accurate measurement of the free  $\text{Ca}^{2+}$  concentration (see Movie 15.4). Indicators of this type are widely used for second-by-second monitoring of changes in intracellular  $\text{Ca}^{2+}$  concentration, or other ion concentrations, in the different parts of a cell viewed in a fluorescence microscope (Figure 9-31).

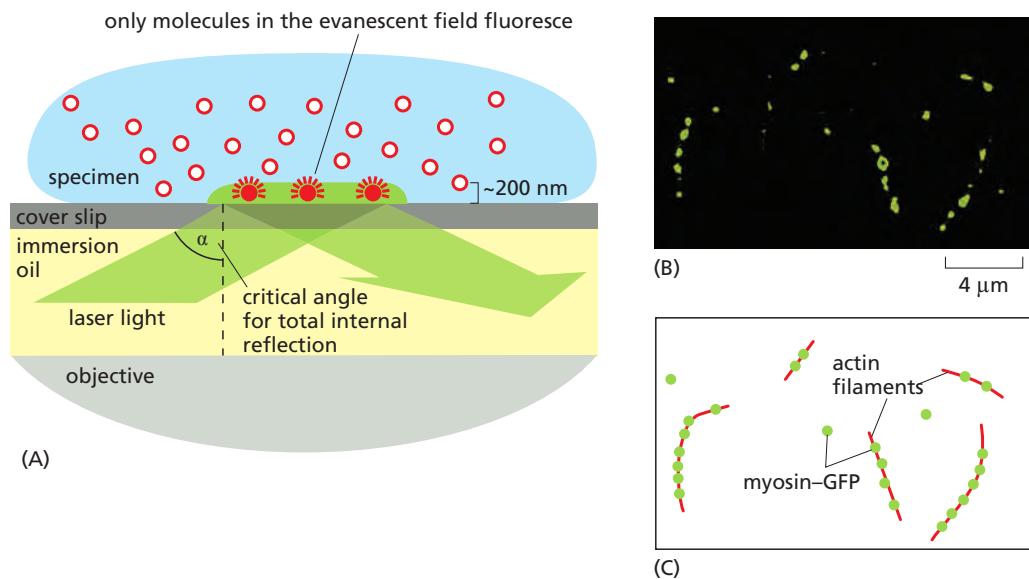
Similar fluorescent indicators measure other ions; some detect  $\text{H}^+$ , for example, and hence measure intracellular pH. Some of these indicators can enter cells by diffusion and thus need not be microinjected; this makes it possible to monitor large numbers of individual cells simultaneously in a fluorescence microscope. New types of indicators, used in conjunction with modern image-processing methods, make possible similarly rapid and precise methods for analyzing changes in the concentrations of many types of small molecules in cells.

### Single Molecules Can Be Visualized by Total Internal Reflection Fluorescence Microscopy

In ordinary microscopes, single fluorescent molecules such as tagged proteins cannot be reliably detected. The limitation has nothing to do with the resolution limit, but instead arises from the strong background due to light emitted or scattered by out-of-focus molecules. This tends to blot out the fluorescence from the

**Figure 9-31 Visualizing intracellular  $\text{Ca}^{2+}$  concentrations by using a fluorescent indicator.** The branching tree of dendrites of a Purkinje cell in the cerebellum receives more than 100,000 synapses from other neurons. The output from the cell is conveyed along the single axon seen leaving the cell body at the bottom of the picture. This image of the intracellular  $\text{Ca}^{2+}$  concentration in a single Purkinje cell (from the brain of a guinea pig) was taken with a low-light camera and the  $\text{Ca}^{2+}$ -sensitive fluorescent indicator fura-2. The concentration of free  $\text{Ca}^{2+}$  is represented by different colors, red being the highest and blue the lowest. The highest  $\text{Ca}^{2+}$  levels are present in the thousands of dendritic branches. (Courtesy of D.W. Tank, J.A. Connor, M. Sugimori, and R.R. Llinás.)



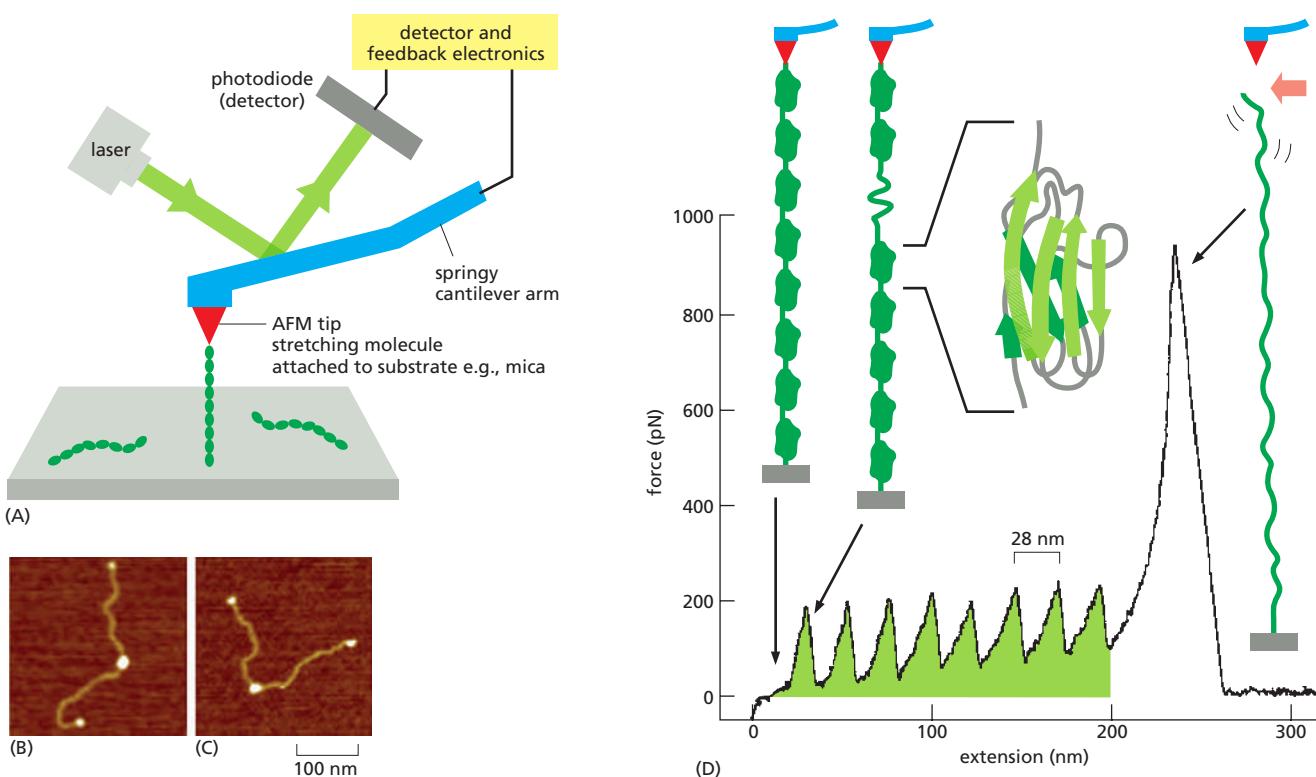


**Figure 9–32** TIRF microscopy allows the detection of single fluorescent molecules. (A) TIRF microscopy uses excitatory laser light to illuminate the cover-slip surface at the critical angle at which all the light is reflected by the glass–water interface. Some electromagnetic energy extends a short distance across the interface as an evanescent wave that excites just those molecules that are attached to the cover slip or are very close to its surface. (B) TIRF microscopy is used here to image individual myosin–GFP molecules (green dots) attached to nonfluorescent actin filaments (C), which are invisible but stuck to the surface of the cover slip. (Courtesy of Dmitry Cherny and Clive R. Bagshaw.)

particular molecule of interest. This problem can be solved by the use of a special optical technique called *total internal reflection fluorescence (TIRF)* microscopy. In a TIRF microscope, laser light shines onto the cover-slip surface at the precise critical angle at which total internal reflection occurs (Figure 9–32A). Because of total internal reflection, the light does not enter the sample, and the majority of fluorescent molecules are not, therefore, illuminated. However, electromagnetic energy does extend, as an evanescent field, for a very short distance beyond the surface of the cover slip and into the specimen, allowing just those molecules in the layer closest to the surface to become excited. When these molecules fluoresce, their emitted light is no longer competing with out-of-focus light from the overlying molecules, and can now be detected. TIRF has allowed several dramatic experiments, for instance imaging of single motor proteins moving along microtubules or single actin filaments forming and branching. At present, the technique is restricted to a thin layer within only 100–200 nm of the cell surface (Figure 9–32B and C).

### Individual Molecules Can Be Touched, Imaged, and Moved Using Atomic Force Microscopy

While TIRF allows single molecules to be visualized under certain conditions, it is strictly a passive observation method. In order to probe molecular function, it is ultimately useful to be able to manipulate individual molecules themselves, and *atomic force microscopy (AFM)* provides a method to do just that. In an AFM device, an extremely small and sharply pointed tip, often of silicon or silicon nitride, is made using nanofabrication methods similar to those used in the semiconductor industry. The tip of the AFM probe is attached to a springy cantilever arm mounted on a highly precise positioning system that allows it to be moved over very small distances. In addition to this precise movement capability, the AFM device is able to collect information about a variety of forces that it encounters—including electrostatic, van der Waals, and mechanical forces—which are felt by its tip as it moves close to or touches the surface (Figure 9–33A). When AFM was first developed, it was intended as an imaging technology to measure molecular-scale

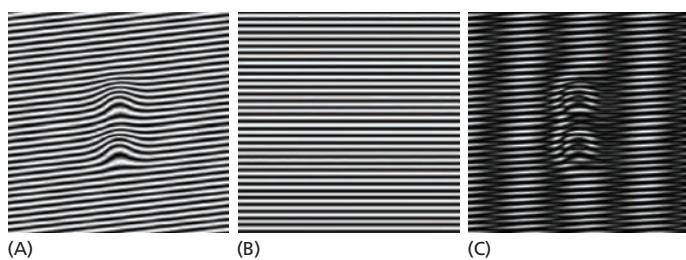


**Figure 9-33 Single molecules can be imaged and manipulated by atomic force microscopy.** (A) Schematic diagram of the key components of an atomic force microscope (AFM), showing the force-sensing tip attached to one end of a single protein molecule, as in the experiment described in (D). (B) and (C) An AFM in imaging mode created these images of a single heteroduplex DNA molecule with a MutS protein dimer (larger white regions) bound near its center, at the point of a mismatched base pair. MutS is the first protein that binds to DNA when the mismatch repair process is initiated (see Figure 5-19). The smaller white dots are single streptavidin molecules, used to label the two ends of each DNA molecule. (D) Titin is an enormous protein molecule that provides muscle with its passive elasticity (see Figure 16-34). The extensibility of this protein can be directly tested using a short, artificially produced protein that contains eight repeated immunoglobulin (Ig) domains from one region of the titin protein. In this experiment, the tip of the AFM is used to pick up, and progressively stretch, a single molecule until it eventually ruptures. As force is applied, each Ig domain suddenly begins to unfold, and the force needed in each case (about 200 pN) can be recorded. The region of the force-extension curve shaded green records the sequential unfolding event for each of the eight protein domains. (B and C, from Y. Jiang and P.E. Marszalek, *EMBO J.* 30:2881–2893, 2011. Reprinted with permission of John Wiley & Sons; D, adapted from W.A. Linke et al., *J. Struct. Biol.* 137:194–205, 2002. With permission from Elsevier.)

features on a surface. When used in this mode, the probe is scanned over the surface, moving up and down as necessary to maintain a constant interaction force with the surface, thus revealing any objects such as proteins or other molecules that might be present on the otherwise flat surface (Figure 9-33B and C). AFM is not limited to simply imaging surfaces, however, and can also be used to pick up and move single molecules that adsorb strongly to the tip. Using this technology, the mechanical properties of individual protein molecules can be measured in detail. For example, AFM has been used to unfold a single protein molecule in order to measure the energetics of domain folding (Figure 9-33 D).

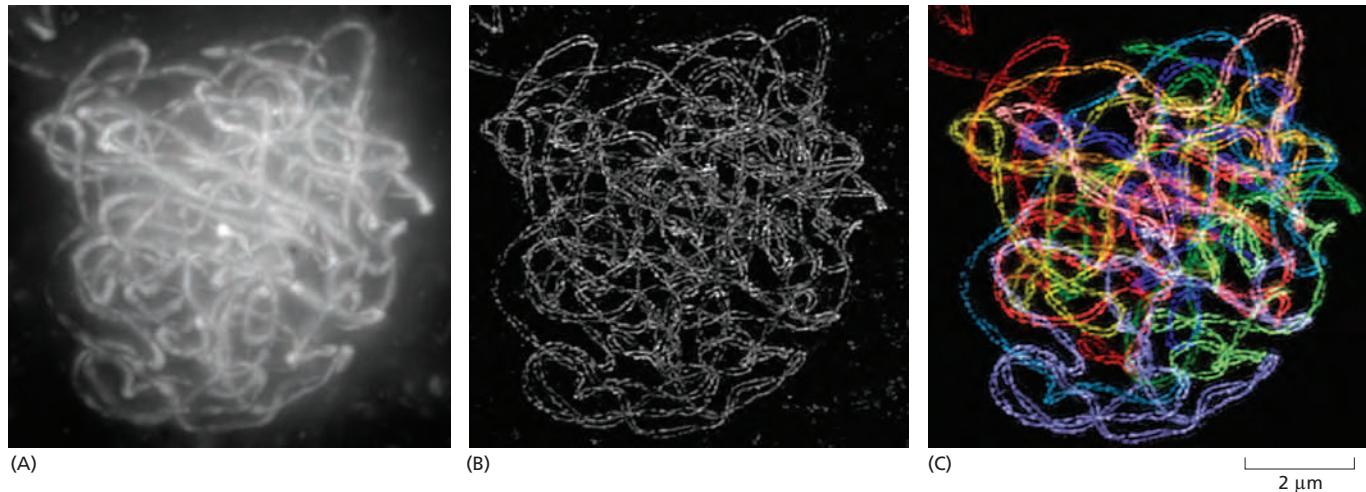
### Superresolution Fluorescence Techniques Can Overcome Diffraction-Limited Resolution

The variations on light microscopy we have described so far are all constrained by the classic diffraction limit to resolution described earlier; that is, to about 200 nm (see Figure 9-6). Yet many cellular structures—from nuclear pores to nucleosomes and clathrin-coated pits—are much smaller than this and so are unresolvable by conventional light microscopy. Several approaches, however, are now available that bypass the limit imposed by the diffraction of light, and successfully allow objects as small as 20 nm to be imaged and clearly resolved: a remarkable, order-of-magnitude improvement.

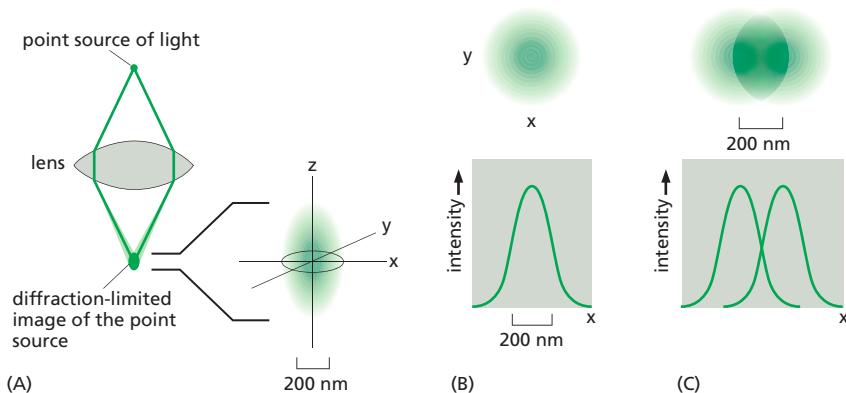


The first of these so-called **superresolution** approaches, **structured illumination microscopy (SIM)**, is a fluorescence imaging method with a resolution of about 100 nm, or twice the resolution of conventional bright-field and confocal microscopy. SIM overcomes the diffraction limit by using a grated or structured pattern of light to illuminate the sample. The microscope's physical set-up and operation is quite complex, but the general principle can be thought of as similar to creating a moiré pattern, an interference pattern created by overlaying two grids with different angles or mesh sizes (Figure 9-34). In a similar way to creating a moiré pattern, the illuminating grid and the sample features combine into an interference pattern, from which the original high-resolution contributions to the image of features beyond the classical resolution limit can be calculated. Illumination by a grid means that the parts of the sample in the dark stripes of the grid are not illuminated and therefore not imaged, so the imaging is repeated several times (usually three) after translating the grid through a fraction of the grid spacing between each image. As the interference effect is strongest for image components close to the direction of the grid bars, the whole process is repeated with the grid pattern rotated through a series of angles to obtain an equivalent enhancement in all directions. Finally, mathematically combining all these separate images by computer creates an enhanced superresolution image. SIM is versatile because it can be used with any fluorescent dye or protein, and combining SIM images captured at consecutive focal planes can create three-dimensional data sets (Figure 9-35).

**Figure 9-34 Structured illumination microscopy.** The principle, illustrated here, is to illuminate a sample with patterned light and measure the moiré pattern. Shown are (A) the pattern from an unknown structure and (B) a known pattern. (C) When these are combined, the resulting moiré pattern contains more information than is easily seen in (A), the original pattern. If the known pattern (B) has higher spatial frequencies, then better resolution will result. However, because the spatial patterns that can be created optically are also diffraction-limited, SIM can only improve the resolution by about a factor of two. (From B.O. Leung and K.C. Chou, *Appl. Spectrosc.* 65:967–980, 2011.)



**Figure 9-35 Structured illumination microscopy can be used to create three-dimensional data.** These three-dimensional projections of the meiotic chromosomes at pachytene in a maize cell show the paired lateral elements of the synaptonemal complexes. (A) The chromosome set has been stained with a fluorescent antibody to cohesin and is viewed here by conventional fluorescence microscopy. Because the distance between the two lateral elements is about 200 nm, the diffraction limit, the two lateral elements that make up each complex are not resolved. (B) In the three-dimensional SIM image, the improved resolution enables each lateral element, about 100 nm across, to be clearly resolved, and the two chromosomes can clearly be seen to coil around each other. (C) Because the complete three-dimensional data set for the whole nucleus is available, the path of each separate pair of chromosomes can be traced and artificially assigned a different color. (Courtesy of C.J. Rachel Wang, Peter Carlton and Zacheus Cande.)



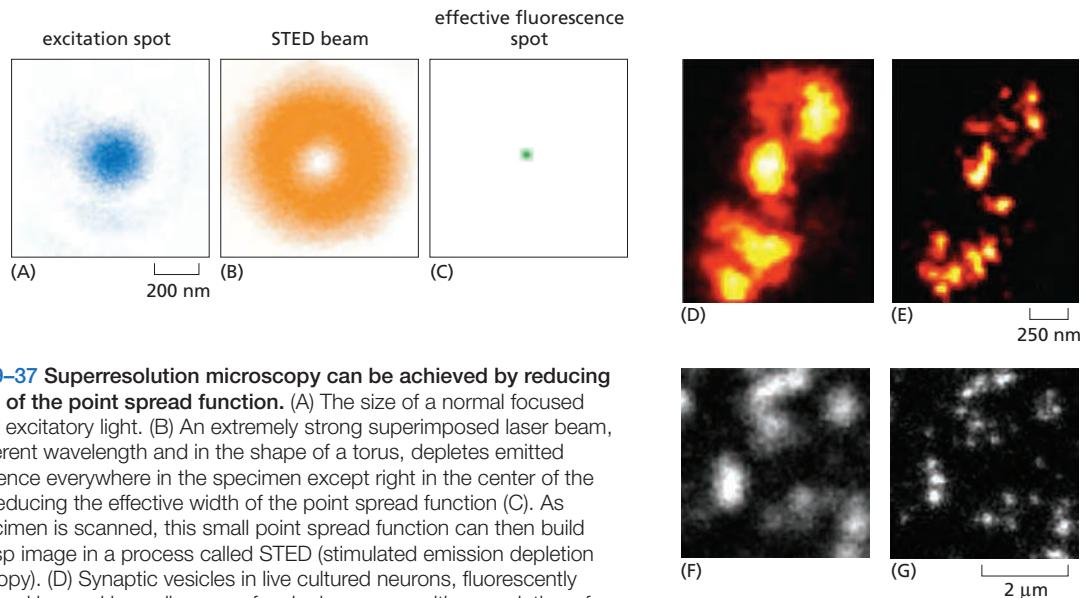
**Figure 9–36** The point spread function of a lens determines resolution. (A) When a point source of light is brought to a focus by a lens system, diffraction effects mean that, instead of being imaged as a point, it is blurred in all dimensions. (B) In the plane of the image, the distribution of light approximates a Gaussian distribution, whose width at half-maximum height under ideal conditions is about 200 nm. (C) Two point sources that are about 200 nm apart can still just be distinguished as separate objects in the image, but if they are any nearer than that, their images will overlap and not be resolvable.

To get around the diffraction limit, the other two superresolution techniques exploit aspects of the point spread function, a property of the optical system mentioned earlier. The *point spread function* is the distribution of light intensity within the three-dimensional, blurred image that is formed when a single point source of light is brought to a focus with a lens. Instead of being identical to the point source, the image has an intensity distribution that is approximately described by a Gaussian distribution, which in turn determines the resolution of the lens system (Figure 9–36). Two points that are closer than the width at half-maximum height of this distribution will become hard to resolve because their images overlap too much (see Figure 9–36C).

In fluorescence microscopy, the excitation light is focused to a spot on the specimen by the objective lens, which then captures the photons emitted by any fluorescent molecule that the beam has raised from a ground state to an excited state. Because the excitation spot is blurred according to the point spread function, fluorescent molecules that are closer than about 200 nm will be imaged as a single blurred spot. One approach to increasing the resolution is to switch all the fluorescent molecules at the periphery of the blurry excitation spot back to their ground state, or to a state where they no longer fluoresce in the normal way, leaving only those at the very center to be recorded. This can be done in practice by adding a second, very bright laser beam that wraps around the excitation beam like a torus. The wavelength and intensity of this second beam are adjusted so as to switch the fluorescent molecules off everywhere except at the very center of the point spread function, a region that can be as small as 20 nm across (Figure 9–37). The fluorescent probes used must be in a special class that is photoswitchable: their emission can be reversibly switched on and off with lights of different wavelengths. As the specimen is scanned with this arrangement of lasers, fluorescent molecules are switched on and off, and the small point spread function at each location is recorded. The diffraction limit is breached because the technique ensures that similar but very closely spaced molecules are in one of two different states, either fluorescing or dark. This approach is called *STED* (*stimulated emission depletion microscopy*) and various microscopes using versions of the general method are now in wide use. Resolutions of 20 nm have been achieved in biological specimens, and even higher resolution attained with nonbiological specimens (see Figure 9–37).

### Superresolution Can Also be Achieved Using Single-Molecule Localization Methods

If a single fluorescent molecule is imaged, it appears as a circular blurry disc, but if sufficient photons have contributed to this image, the precise mathematical center of the dislike image can be determined very accurately, often to within a few nanometers. But the problem with a specimen that contains a large number

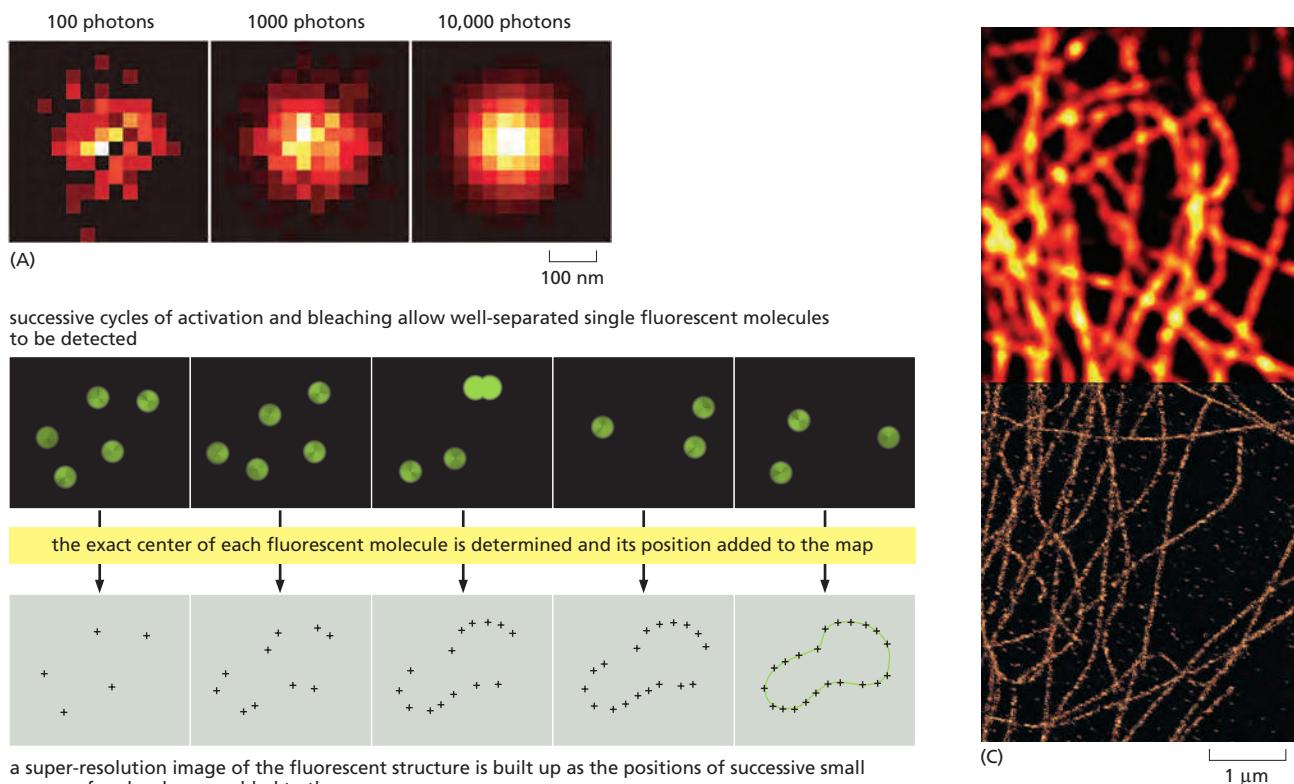


**Figure 9–37** Superresolution microscopy can be achieved by reducing the size of the point spread function. (A) The size of a normal focused beam of excitatory light. (B) An extremely strong superimposed laser beam, at a different wavelength and in the shape of a torus, depletes emitted fluorescence everywhere in the specimen except right in the center of the beam, reducing the effective width of the point spread function (C). As the specimen is scanned, this small point spread function can then build up a crisp image in a process called STED (stimulated emission depletion microscopy). (D) Synaptic vesicles in live cultured neurons, fluorescently labeled and imaged by ordinary confocal microscopy, with a resolution of 260 nm. (E) The same vesicles imaged by STED, with a resolution of 60 nm, which allows single vesicles to be resolved. (F) Fluorescently labeled replication factories in the nucleus of a cultured cell, imaged by ordinary confocal microscopy. (G) The same replication factories imaged by STED. Single, discrete replication sites can be resolved by STED that cannot be seen in the confocal image. (A, B, and C, from G. Donnert et al., *Proc. Natl Acad. Sci. USA* 103:11440–11445, 2006. With permission from National Academy of Sciences; D and E, from V. Westphal et al., *Science* 320:246–249, 2008. With permission from AAAS; F and G, from Z. Cseresnyes, U. Schwarz and C.M. Green, *BMC Cell Biol.* 10:88, 2009.)

of adjacent fluorescent molecules, as we saw earlier, is that they each contribute blurry, overlapping point spread functions to the image, making the exact position of any one molecule impossible to resolve. Another way round this limitation is to arrange for only a very few, clearly separated molecules to actively fluoresce at any one moment. The exact position of each of these can then be computed, before subsequent sets of molecules are examined.

In practice, this can be achieved by using lasers to sequentially switch on a sparse subset of fluorescent molecules in a specimen containing photoactivatable or photoswitchable fluorescent labels. Labels are activated, for example, by illumination with near-ultraviolet light, which modifies a small subset of molecules so that they fluoresce when exposed to an excitation beam at another wavelength. These are then imaged before bleaching quenches their fluorescence and a new subset is activated. Each molecule emits a few thousand photons in response to the excitation before switching off, and the switching process can be repeated hundreds or even thousands of times, allowing the exact coordinates of a very large set of single molecules to be determined. The full set can be combined and digitally displayed as an image in which the computed location of each individual molecule is exactly marked (**Figure 9–38**). This class of methods has been variously termed *photoactivated localization microscopy (PALM)* or *stochastic optical reconstruction microscopy (STORM)*.

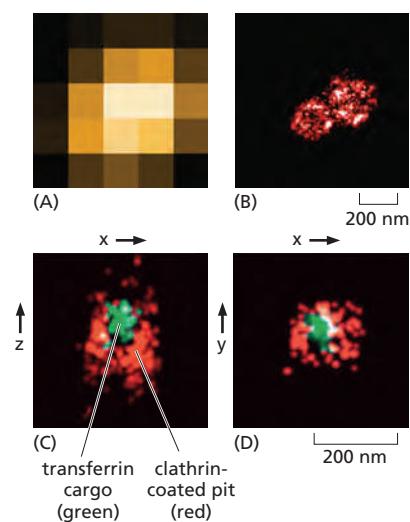
By switching the fluorophores off and on sequentially in different regions of the specimen as a function of time, all the superresolution imaging methods described above allow the resolution of molecules that are much closer together than the 200 nm diffraction limit. In STED, the locations of the molecules are determined by using optical methods to define exactly where their fluorescence will be on or off. In PALM and STORM, individual fluorescent molecules are switched on and off at random over a period of time, allowing their positions to be accurately determined. PALM and STORM techniques have depended on the



**Figure 9–38** Single fluorescent molecules can be located with great accuracy. (A) Determining the exact mathematical center of the blurred image of a single fluorescent molecule becomes more accurate the more photons contribute to the final image. The point spread function described in the text dictates that the size of the molecular image is about 200 nm across, but in very bright specimens, the position of its center can be pinpointed to within a nanometer. (B) In this imaginary specimen, sparse subsets of fluorescent molecules are individually switched on briefly and then bleached. The exact positions of all these well-spaced molecules can be gradually built up into an image at superresolution. (C) In this portion of a cell, the microtubules have been fluorescently labeled and imaged at the top in a TIRF microscope (see Figure 9–32) and below, at superresolution, in a PALM microscope. The diameter of the microtubules in the lower panel now resembles their true size, about 25 nm, rather than the 250 nm in the blurred image at the top. (A, from A.L. McEvoy et al., *BMC Biol.* 8:106, 2010; C, courtesy of Carl Zeiss Ltd.)

development of novel fluorescent probes that exhibit the appropriate switching behavior. All these methods are now being extended to incorporate multicolor imaging, three-dimensional imaging (Figure 9–39), and live-cell imaging in real time. Ending the long reign of the diffraction limit has certainly reinvigorated light microscopy and its place in cell biology research.

**Figure 9–39** Small fluorescent structures can be imaged in three dimensions with superresolution. (A) The image of two touching 180-nm-diameter clathrin-coated pits on the plasma membrane of a cultured cell is diffraction-limited, and the individual pits cannot be distinguished in this conventional fluorescence image. (B) Using STORM superresolution microscopy, however, the pits are clearly resolvable. Not only can such pits be imaged using probes of different colors, but additional three-dimensional information can also be obtained. (C) and (D) Shown are two different orthogonal views of one single coated pit. The clathrin is labeled red and transferrin—the cargo within the pit—is labeled green. Images of this sort can be acquired in less than one second, making possible dynamic observations on living cells. These techniques depend heavily on the development of new, very fast-switching, and extremely bright fluorescent probes. (A and B, from M. Bates et al., *Science* 317:1749–1753, 2007; C and D, from S.A. Jones et al., *Nat. Methods* 8:499–508, 2011. With permission from Macmillan Publishers Ltd.)



## Summary

Many light-microscope techniques are available for observing cells. Cells that have been fixed and stained can be studied in a conventional light microscope, whereas antibodies coupled to fluorescent dyes can be used to locate specific molecules in cells in a fluorescence microscope. Living cells can be seen with phase-contrast, differential-interference-contrast, dark-field, or bright-field microscopes. All forms of light microscopy are facilitated by digital image-processing techniques, which enhance sensitivity and refine the image. Confocal microscopy and image deconvolution both provide thin optical sections and can be used to reconstruct three-dimensional images.

Techniques are now available for detecting, measuring, and following almost any desired molecule in a living cell. Fluorescent indicator dyes can be introduced to measure the concentrations of specific ions in individual cells or in different parts of a cell. Virtually any protein of interest can be genetically engineered as a fluorescent fusion protein, and then imaged in living cells by fluorescence microscopy. The dynamic behavior and interactions of many molecules can be followed in living cells by variations on the use of fluorescent protein tags, in some cases at the level of single molecules. Various superresolution techniques can circumvent the diffraction limit and resolve molecules separated by distances as small as 20 nm.

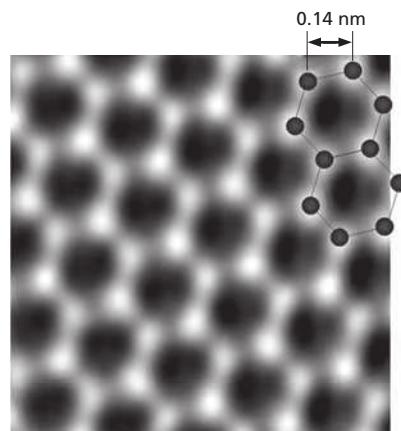
## LOOKING AT CELLS AND MOLECULES IN THE ELECTRON MICROSCOPE

Light microscopy is limited in the fineness of detail that it can reveal. Microscopes using other types of radiation—in particular, electron microscopes—can resolve much smaller structures than is possible with visible light. This higher resolution comes at a cost: specimen preparation for electron microscopy is complex and it is harder to be sure that what we see in the image corresponds precisely to the original living structure. It is possible, however, to use very rapid freezing to preserve structures faithfully for electron microscopy. Digital image analysis can be used to reconstruct three-dimensional objects by combining information either from many individual particles or from multiple tilted views of a single object. Together, these approaches extend the resolution and scope of electron microscopy to the point at which we can faithfully image the structures of individual macromolecules and the complexes they form.

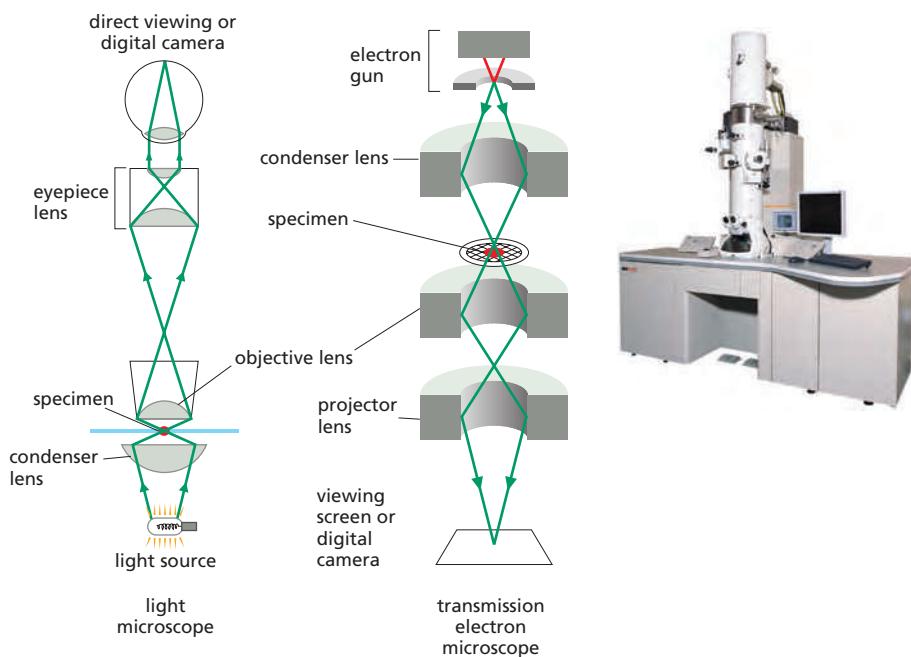
### The Electron Microscope Resolves the Fine Structure of the Cell

The formal relationship between the diffraction limit to resolution and the wavelength of the illuminating radiation (see Figure 9–6) holds true for any form of radiation, whether it is a beam of light or a beam of electrons. With electrons, however, the limit of resolution is very small. The wavelength of an electron decreases as its velocity increases. In an **electron microscope** with an accelerating voltage of 100,000 V, the wavelength of an electron is 0.004 nm. In theory, the resolution of such a microscope should be about 0.002 nm, which is 100,000 times that of the light microscope. Because the aberrations of an electron lens are considerably harder to correct than those of a glass lens, however, the practical resolving power of modern electron microscopes is, even with careful image processing to correct for lens aberrations, about 0.05 nm (0.5 Å) (Figure 9–40). This is because only the very center of the electron lenses can be used, and the effective numerical aperture is tiny. Furthermore, problems of specimen preparation, contrast, and radiation damage have generally limited the normal effective resolution for biological objects to 1 nm (10 Å). This is nonetheless about 200 times better than the resolution of the light microscope. Moreover, the performance of electron microscopes is improved by electron illumination sources called field emission guns. These very bright and coherent sources substantially improve the resolution achieved.

In overall design, the transmission electron microscope (TEM) is similar to a light microscope, although it is much larger and “upside down” (Figure 9–41).



**Figure 9–40** The resolution of the electron microscope. This transmission electron micrograph of a monolayer of graphene resolves the individual carbon atoms as bright spots in a hexagonal lattice. Graphene is a single isolated atomic plane of graphite and forms the basis of carbon nanotubes. The distance between adjacent bonded carbon atoms is 0.14 nm (1.4 Å). Such resolution can only be obtained in a specially built transmission electron microscope in which all lens aberrations are carefully corrected, and with optimal specimens; it cannot be achieved with most conventional biological specimens. (From A. Dato et al., *Chem. Commun.* 40:6095–6097, 2009. With permission from The Royal Society of Chemistry.)



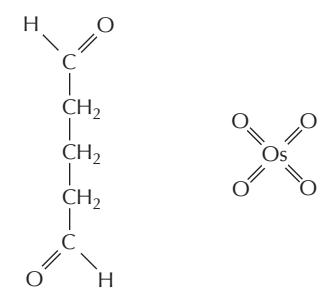
**Figure 9–41** The principal features of a light microscope and a transmission electron microscope. These drawings emphasize the similarities of overall design. Whereas the lenses in the light microscope are made of glass, those in the electron microscope are magnetic coils. The electron microscope requires that the specimen be placed in a vacuum. The inset shows a transmission electron microscope in use. (Photograph courtesy of JEOL Ltd.)

The source of illumination is a filament or cathode that emits electrons at the top of a cylindrical column about 2 m high. Since electrons are scattered by collisions with air molecules, air must first be pumped out of the column to create a vacuum. The electrons are then accelerated from the filament by a nearby anode and allowed to pass through a tiny hole to form an electron beam that travels down the column. Magnetic coils placed at intervals along the column focus the electron beam, just as glass lenses focus the light in a light microscope. The specimen is put into the vacuum, through an airlock, into the path of the electron beam. As in light microscopy, the specimen is usually stained—in this case, with *electron-dense* material. Some of the electrons passing through the specimen are scattered by structures stained with the electron-dense material; the remainder are focused to form an image, in a manner analogous to the way an image is formed in a light microscope. The image can be observed on a phosphorescent screen or recorded with a high-resolution digital camera. Because the scattered electrons are lost from the beam, the dense regions of the specimen show up in the image as areas of reduced electron flux, which look dark.

### Biological Specimens Require Special Preparation for Electron Microscopy

In the early days of its application to biological materials, the electron microscope revealed many previously unimagined structures in cells. But before these discoveries could be made, electron microscopists had to develop new procedures for embedding, cutting, and staining tissues.

Since the specimen is exposed to a very high vacuum in the electron microscope, living tissue is usually killed and preserved by fixation—first with *glutaraldehyde*, which covalently cross-links protein molecules to their neighbors, and then with *osmium tetroxide*, which binds to and stabilizes lipid bilayers as well as proteins (Figure 9–42). Because electrons have very limited penetrating power, the fixed tissues normally have to be cut into extremely thin sections (25–100 nm thick, about 1/200 the thickness of a single cell) before they are viewed. This is achieved by dehydrating the specimen, permeating it with a monomeric resin that polymerizes to form a solid block of plastic, then cutting the block with a fine glass or diamond knife on a special microtome. The resulting *thin sections*, free of water and other volatile solvents, are supported on a small metal grid for viewing in the microscope (Figure 9–43).



glutaraldehyde      osmium tetroxide

**Figure 9–42** Two common chemical fixatives used for electron microscopy. The two reactive aldehyde groups of glutaraldehyde enable it to cross-link various types of molecules, forming covalent bonds between them. Osmium tetroxide forms cross-linked complexes with many organic compounds, and in the process becomes reduced. This reaction is especially useful for fixing cell membranes, since the C=C double bonds present in many fatty acids react with osmium tetroxide.

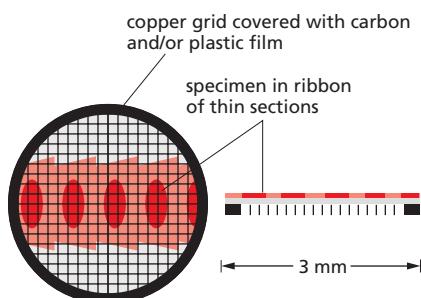
The steps required to prepare biological material for electron microscopy are challenging. How can we be sure that the image of the fixed, dehydrated, resin-embedded specimen bears any relation to the delicate, aqueous biological system present in the living cell? The best current approaches to this problem depend on rapid freezing. If an aqueous system is cooled fast enough and to a low enough temperature, the water and other components in it do not have time to rearrange themselves or crystallize into ice. Instead, the water is supercooled into a rigid but noncrystalline state—a “glass”—called vitreous ice. This state can be achieved by slamming the specimen onto a polished copper block cooled by liquid helium, by plunging it into or spraying it with a jet of a coolant such as liquid propane, or by cooling it at high pressure.

Some rapidly frozen specimens can be examined directly in the electron microscope using a special cooled specimen holder. In other cases, the frozen block can be fractured to reveal interior cell surfaces, or the surrounding ice can be sublimed away to expose external surfaces. However, we often want to examine thin sections. A compromise is therefore to rapid-freeze the tissue, replace the water with organic solvents, embed the tissue in plastic resin, and finally cut sections and stain. Although technically still difficult, this approach stabilizes and preserves the tissue in a condition very close to its original living state (Figure 9–44).

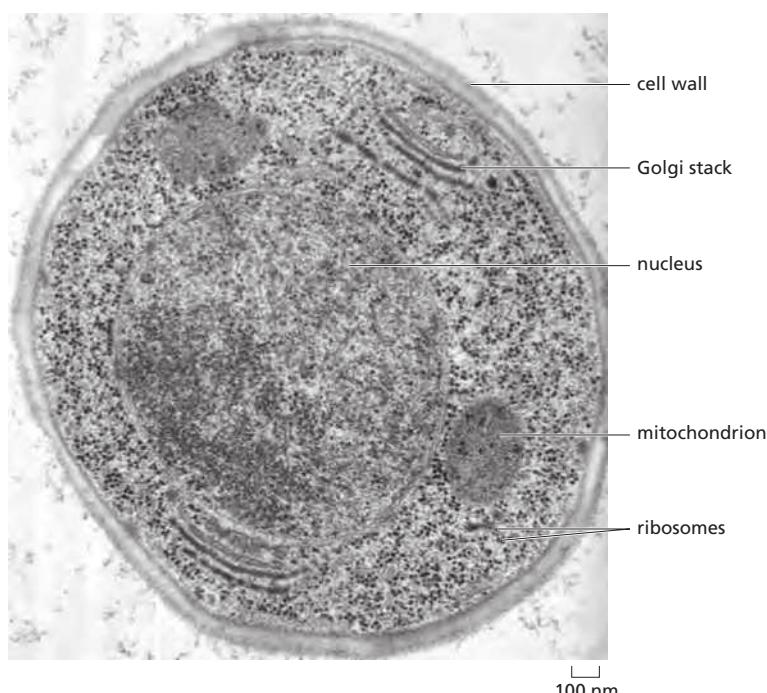
Image clarity in an electron micrograph depends upon having a range of contrasting electron densities within the specimen. Electron density in turn depends on the atomic number of the atoms that are present: the higher the atomic number, the more electrons are scattered and the darker that part of the image. Biological tissues are composed mainly of atoms of very low atomic number (primarily carbon, oxygen, nitrogen, and hydrogen). To make them visible, tissues are usually impregnated (before or after sectioning) with the salts of heavy metals such as uranium, lead, and osmium. The degree of impregnation, or “staining,” with these salts will vary for different cell constituents. Lipids, for example, tend to stain darkly after osmium fixation, revealing the location of cell membranes.

### Specific Macromolecules Can Be Localized by Immunogold Electron Microscopy

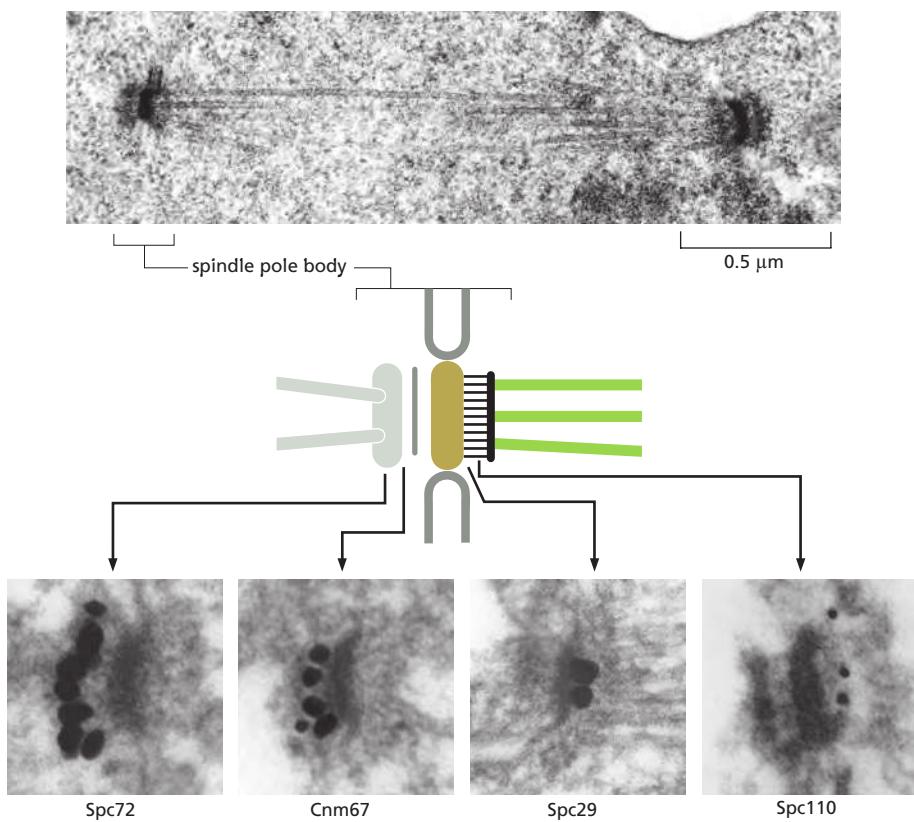
We have seen how antibodies can be used in conjunction with fluorescence microscopy to localize specific macromolecules. An analogous method—**immunogold**



**Figure 9–43** The metal grid that supports the thin sections of a specimen in a transmission electron microscope.



**Figure 9–44** Thin section of a cell. This thin section is of a yeast cell that has been very rapidly frozen and the vitreous ice replaced by organic solvents and then by plastic resin. The nucleus, mitochondria, cell wall, Golgi stacks, and ribosomes can all be readily seen in a state that is presumed to be as lifelike as possible. (Courtesy of Andrew Staehelin.)



**Figure 9–45 Localizing proteins in electron microscopy.** Immunogold electron microscopy is used here to find the specific location of four different protein components within the spindle pole body of yeast. At the top is a thin section of a yeast mitotic spindle showing the spindle microtubules that cross the nucleus and connect at each end to spindle pole bodies embedded in the nuclear envelope. A diagram of the components of a single spindle pole body is shown below. On separate sections, antibodies against four different proteins of the spindle pole body are used, together with colloidal gold particles (black dots), to reveal where within the complex structure each protein is located. (Courtesy of John Kilmartin.)

**electron microscopy**—can be used in the electron microscope. The usual procedure is to incubate a thin section first with a specific primary antibody, and then with a secondary antibody to which a colloidal gold particle has been attached. The gold particle is electron-dense and can be seen as a black dot in the electron microscope (Figure 9–45). Different antibodies can be conjugated to different sized gold particles so multiple proteins can be localized in a single sample.

A complication for immunogold labeling is that the antibodies and colloidal gold particles do not penetrate into the resin used for embedding; therefore, they detect antigens only at the surface of the section. This means that the method's sensitivity is low, since antigen molecules in the deeper parts of the section are not detected. Furthermore, we may get a false impression regarding which structures contain the antigen and which do not. One solution is to label the specimen before embedding it in plastic, when cells and tissues are still fully accessible to labeling reagents. Extremely small gold particles, about 1 nm in diameter, work best for this procedure. Such small gold particles are usually not easily visible in the final sections, so additional silver or gold is nucleated around the tiny 1 nm gold particles in a chemical process very much like photographic development.

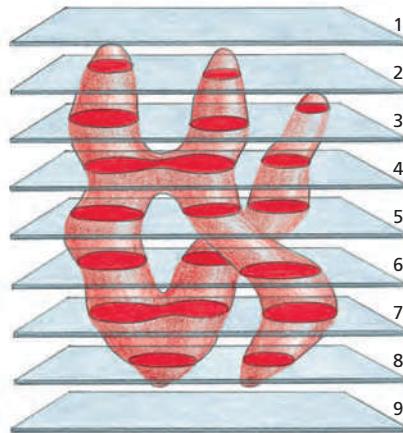
#### Different Views of a Single Object Can Be Combined to Give a Three-Dimensional Reconstruction

Thin sections often fail to convey the three-dimensional arrangement of cellular components viewed in a TEM, and the image can be very misleading: a linear structure such as a microtubule may appear in section as a pointlike object, for example, and a section through protruding parts of a single irregularly shaped solid body may give the appearance of two or more separate objects (Figure 9–46). The third dimension can be reconstructed from serial sections, but this is a lengthy and tedious process. Even thin sections, however, have a significant depth compared with the resolution of the electron microscope, so the TEM image can also be misleading in an opposite way, through the superimposition of objects that lie at different depths.

Because of the large depth of field of electron microscopes, all the parts of the three-dimensional specimen are in focus, and the resulting image is a projection (a superimposition of layers) of the structure along the viewing direction. The lost information in the third dimension can be recovered if we have views of the same specimen from many different directions. The computational methods for this technique are widely used in medical CT scans. In a CT scan, the imaging equipment is moved around the patient to generate the different views. In **electron-microscope (EM) tomography**, the specimen holder is tilted in the microscope, which achieves the same result. In this way, we can arrive at a three-dimensional reconstruction, in a chosen standard orientation, by combining different views of a single object. Each individual view will be very noisy but by combining them in three dimensions and taking an average, the noise can be largely eliminated. Starting with thick plastic sections of embedded material, three-dimensional reconstructions, or *tomograms*, are used extensively to describe the detailed anatomy of specific regions of the cell, such as the Golgi apparatus (Figure 9–47) or the cytoskeleton. Increasingly, microscopists are also applying EM tomography to unstained frozen, hydrated sections, and even to rapidly frozen whole cells or organelles (Figure 9–48). Electron microscopy now provides a robust bridge between the scale of the single molecule and that of the whole cell.

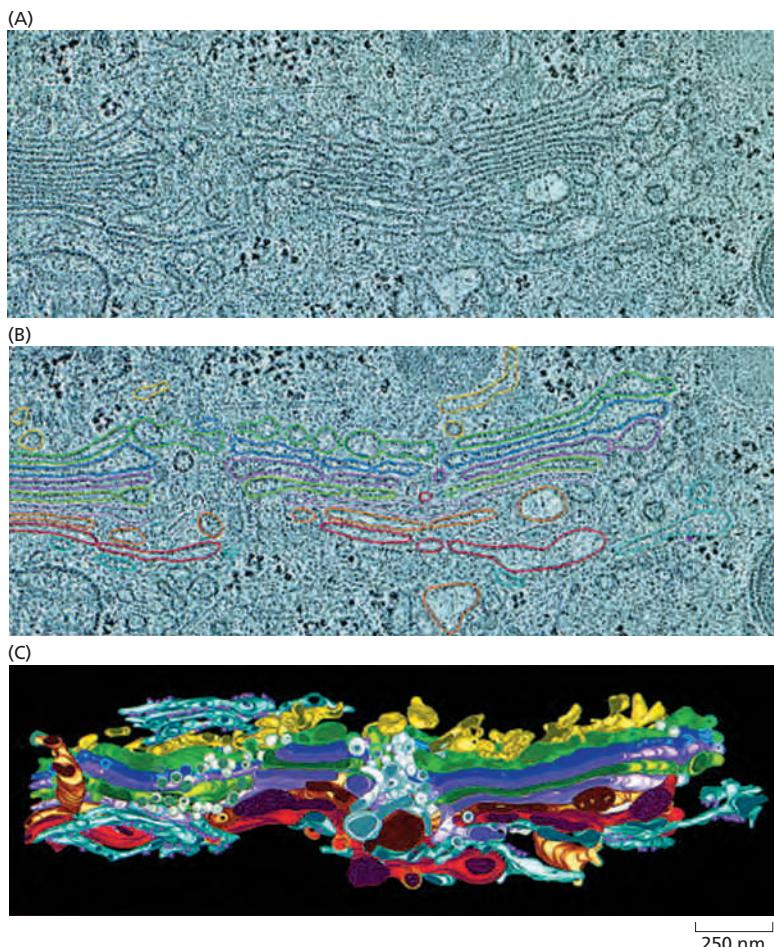
### Images of Surfaces Can Be Obtained by Scanning Electron Microscopy

A **scanning electron microscope (SEM)** directly produces an image of the three-dimensional structure of the surface of a specimen. The SEM is usually smaller, simpler, and cheaper than a transmission electron microscope. Whereas the TEM uses the electrons that have passed through the specimen to form an



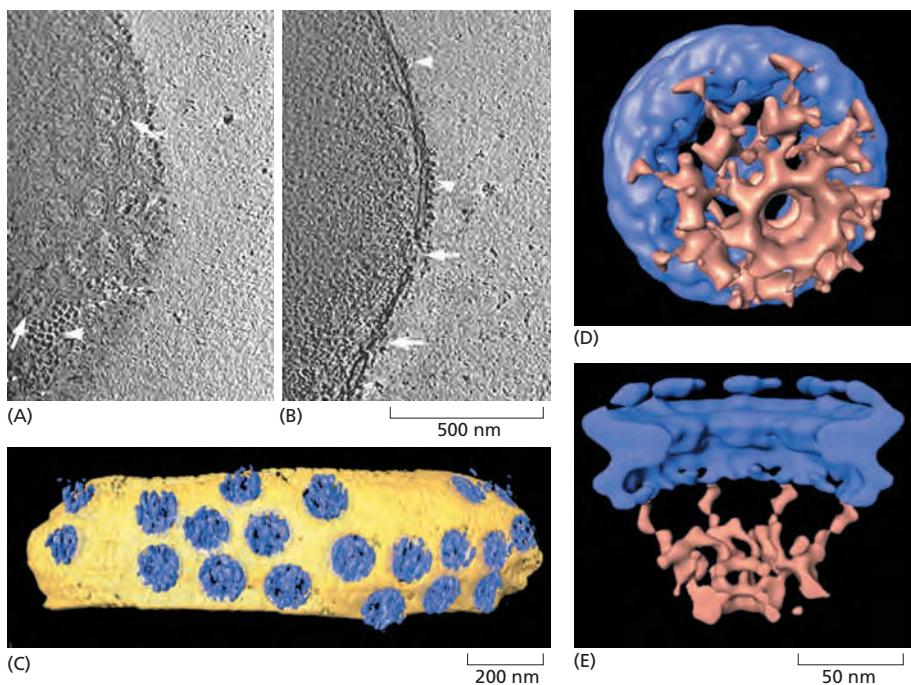
**Figure 9–46** A three-dimensional reconstruction from serial sections.

Single thin sections in the electron microscope sometimes give misleading impressions. In this example, most sections through a cell containing a branched mitochondrion seem to contain two or three separate mitochondria (compare Figure 9–44). Sections 4 and 7, moreover, might be interpreted as showing a mitochondrion in the process of dividing. The true three-dimensional shape can be reconstructed from a complete set of serial sections.



**Figure 9–47** Electron-microscope (EM) tomography.

**tomography.** Samples that have been rapidly frozen, and then freeze-substituted and embedded in plastic, preserve their structure in a condition that is very close to their original living state (Movie 9.2). This example shows the three-dimensional structure of the Golgi apparatus from a rat kidney cell. Several thick sections (250 nm) of the cell were tilted in a high-voltage electron microscope, along two different axes, and about 160 different views recorded. The digital data allow individual thin slices of the complete three-dimensional data set, or tomogram, to be viewed; for example, the serial slices, each only 4 nm thick, are shown in (A) and (B). Very little changes from one slice to the next, but using the full data set, and manually color-coding the membranes (B), one can obtain a full three-dimensional reconstruction, at a resolution of about 7 nm, of the complete Golgi complex and its associated vesicles (C). (From M.S. Ladinsky et al., *J. Cell Biol.* 144:1135–1149, 1999. With permission from the authors.)



**Figure 9–48** Combining cryoelectron-microscope tomography and single-particle reconstruction. Small, unfixed, rapidly frozen specimens can be examined while still frozen. In this example, the small nuclei of the amoeba *Dictyostelium* were gently isolated and then very rapidly frozen before a series of angled views were recorded with the aid of a tilting microscope stage. These digital views are combined by EM tomographs to produce a three-dimensional tomogram. Two thin digital slices (10 nm) through this tomogram show (A) top views and (B) side views of individual nuclear pores (white arrows). (C) In the three-dimensional model, a surface rendering of the pores (blue) is seen embedded in the nuclear envelope (yellow). From a series of tomograms it was possible to extract data sets for nearly 300 separate nuclear pores, whose structures could then be averaged using the techniques of single-particle reconstruction. The surface-rendered view of one of these reconstructed pores is shown (D) from the nuclear face and (E) in cross section (compare with Figure 12–8). The pore complex is colored blue and the nuclear basket brown. (From M. Beck et al., *Science* 306:1387–1390, 2004. With permission from AAAS.)

image, the SEM uses electrons that are scattered or emitted from the specimen's surface. The specimen to be examined is fixed, dried, and coated with a thin layer of heavy metal. Alternatively, it can be rapidly frozen, and then transferred to a cooled specimen stage for direct examination in the microscope. Often an entire plant part or small animal can be put into the microscope with very little preparation (**Figure 9–49**). The specimen is scanned with a very narrow beam of electrons. The quantity of electrons scattered or emitted as this primary beam bombards each successive point of the metallic surface is measured and used to control the intensity of a second beam, which moves in synchrony with the primary beam and forms an image on a computer screen. Eventually a highly enlarged image of the surface as a whole is built up (**Figure 9–50**).

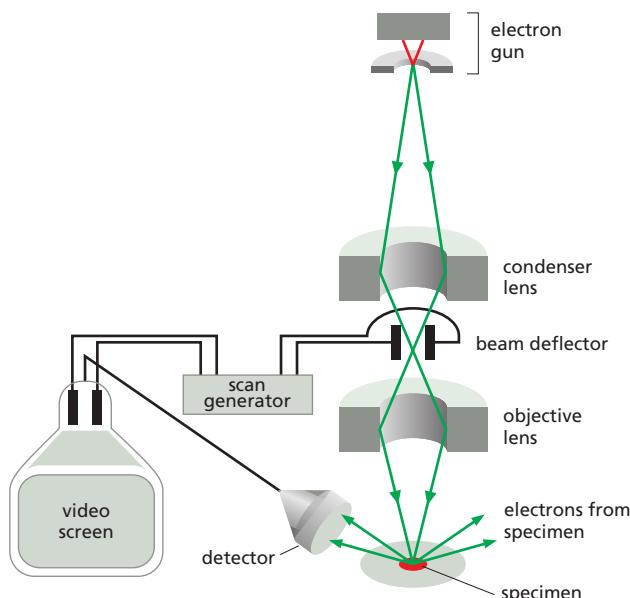
The SEM technique provides great depth of field; moreover, since the amount of electron scattering depends on the angle of the surface relative to the beam, the image has highlights and shadows that give it a three-dimensional appearance (see Figure 9–49 and **Figure 9–51**). Only surface features can be examined, however, and in most forms of SEM, the resolution attainable is not very high (about 10 nm, with an effective magnification of up to 20,000 times). As a result, the technique is usually used to study whole cells and tissues rather than subcellular organelles (see Movie 21.3). Very-high-resolution SEMs have, however, been developed with a bright coherent-field emission gun as the electron source. This type of SEM can produce images that rival the resolution possible with a TEM (**Figure 9–52**).

### Negative Staining and Cryoelectron Microscopy Both Allow Macromolecules to Be Viewed at High Resolution

If they are shadowed with a heavy metal to provide contrast, isolated macromolecules such as DNA or large proteins can be visualized readily in the electron microscope, but **negative staining** allows finer detail to be seen. In this technique, the molecules are supported on a thin film of carbon and mixed with a solution of a heavy-metal salt such as uranyl acetate. After the sample has dried, a very thin film of metal salt covers the carbon film everywhere except where it has been excluded by the presence of an adsorbed macromolecule. Because the macromolecule allows electrons to pass through it much more readily than does the surrounding heavy-metal stain, a reverse or negative image of the molecule is



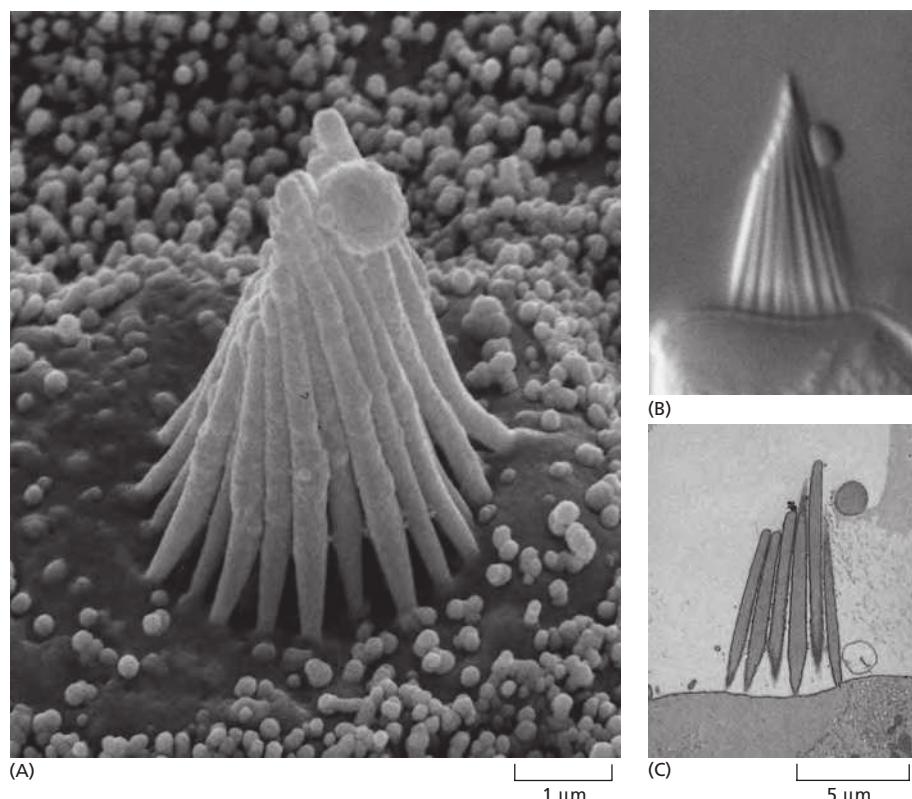
**Figure 9–49** A developing wheat flower, or spike. This delicate flower spike was rapidly frozen, coated with a thin metal film, and examined in the frozen state with an SEM. This low-magnification micrograph demonstrates the large depth of focus of an SEM. (Courtesy of Kim Findlay.)



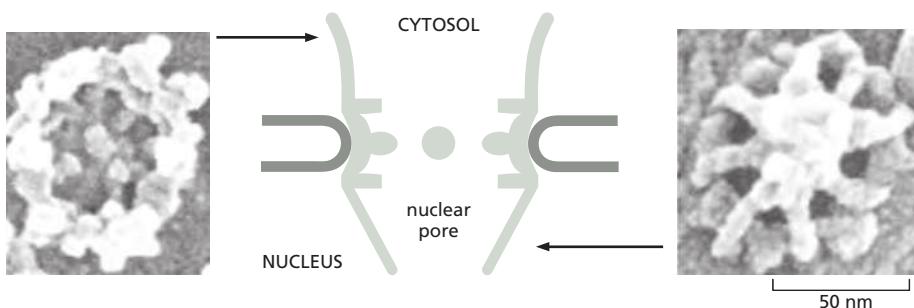
created. Negative staining is especially useful for viewing large macromolecular aggregates such as viruses or ribosomes, and for seeing the subunit structure of protein filaments (Figure 9–53).

Shadowing and negative staining can provide high-contrast surface views of small macromolecular assemblies, but the size of the smallest metal particles in the shadow or stain used limits the resolution of both techniques. An alternative that allows us to visualize directly at high resolution even the interior features of three-dimensional structures such as viruses and organelles is **cryoelectron microscopy**, in which rapid freezing to form vitreous ice is again the key. A very thin (about 100 nm) film of an aqueous suspension of virus or purified macromolecular complex is prepared on a microscope grid and is then rapidly frozen by

**Figure 9–50** The scanning electron microscope. In an SEM, the specimen is scanned by a beam of electrons brought to a focus on the specimen by the electromagnetic coils that act as lenses. The detector measures the quantity of electrons scattered or emitted as the beam bombards each successive point on the surface of the specimen and controls the intensity of successive points in an image built up on a screen. The SEM creates striking images of three-dimensional objects with great depth of focus and a resolution between 3 nm and 20 nm depending on the instrument. (Photograph courtesy of Andrew Davies.)



**Figure 9–51** Scanning electron microscopy. (A) A scanning electron micrograph of the stereocilia projecting from a hair cell in the inner ear of a bullfrog. For comparison, the same structure is shown by (B) differential-interference-contrast light microscopy (Movie 9.3) and (C) thin-section transmission electron microscopy. (Courtesy of Richard Jacobs and James Hudspeth.)



**Figure 9–52 The nuclear pore.** Rapidly frozen nuclear envelopes were imaged in a high-resolution SEM, equipped with a field emission gun as the electron source. These views of each side of a nuclear pore represent the limit of resolution of the SEM (compare with Figure 12–8). (Courtesy of Martin Goldberg and Terry Allen.)

being plunged into a coolant. A special sample holder keeps this hydrated specimen at  $-160^{\circ}\text{C}$  in the vacuum of the microscope, where it can be viewed directly without fixation, staining, or drying. Unlike negative staining, in which what we see is the envelope of stain exclusion around the particle, hydrated cryoelectron microscopy produces an image from the macromolecular structure itself. However, the contrast in this image is very low, and to extract the maximum amount of structural information, special image-processing techniques must be used, as we describe next.

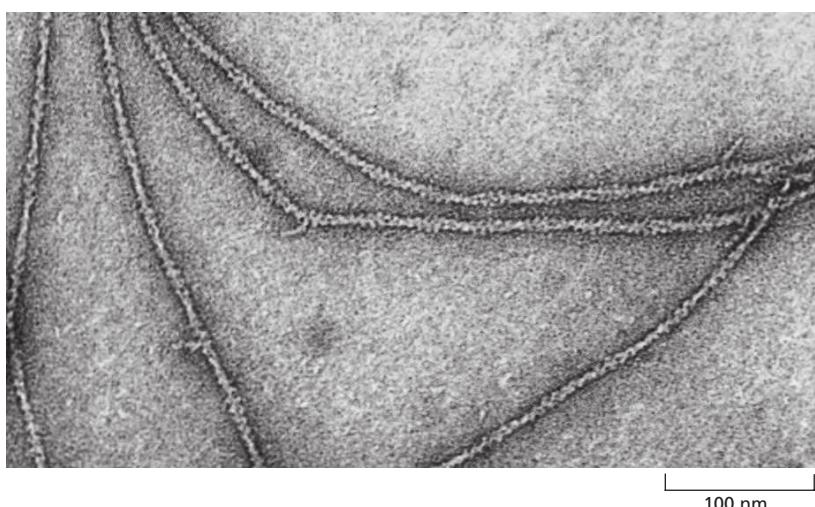
### Multiple Images Can Be Combined to Increase Resolution

As we saw earlier (p. 532), noise is important in light microscopy at low light levels, but it is a particularly severe problem for electron microscopy of unstained macromolecules. A protein molecule can tolerate a dose of only a few tens of electrons per square nanometer without damage, and this dose is orders of magnitude below what is needed to define an image at atomic resolution.

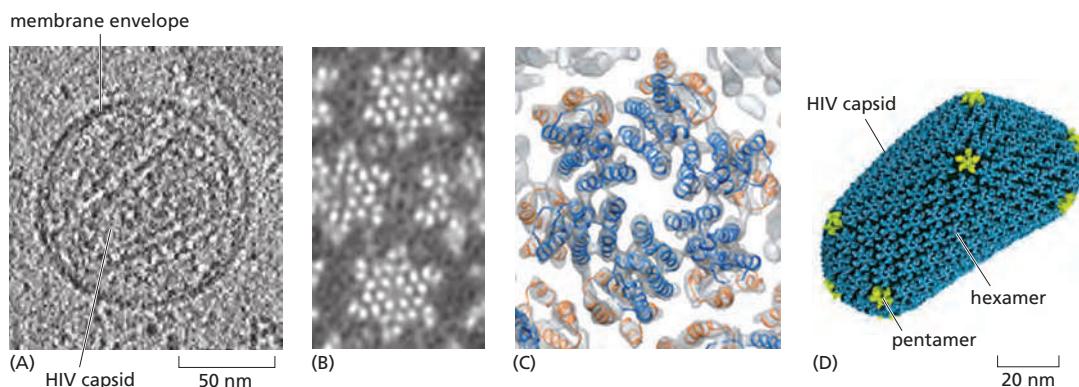
The solution is to obtain images of many identical molecules—perhaps tens of thousands of individual images—and combine them to produce an averaged image, revealing structural details that are hidden by the noise in the original images. This procedure is called **single-particle reconstruction**. Before combining all the individual images, however, they must be aligned with each other. Sometimes it is possible to induce proteins and complexes to form crystalline arrays, in which each molecule is held in the same orientation in a regular lattice. In this case, the alignment problem is easily solved, and several protein structures have been determined at atomic resolution by this type of electron crystallography. In principle, however, crystalline arrays are not absolutely required. With the help of a computer, the digital images of randomly distributed and unaligned molecules can be processed and combined to yield high-resolution reconstructions (see Movie 13.1). Although structures that have some intrinsic symmetry make the task of alignment easier and more accurate, this technique has also been used for objects like ribosomes, with no symmetry. **Figure 9–54** shows the structure of

### WHAT WE DON'T KNOW

- We know in detail about many cell processes, such as DNA replication and transcription and RNA translation, but will we ever be able to visualize such rapid molecular processes in action in cells?
- Will we ever be able to image intracellular structures at the resolution of the electron microscope in living cells?
- How can we improve crystallization and single-particle cryoelectron microscopy techniques to obtain high-resolution structures of all important membrane channels and transporters? What new concepts might these structures reveal?



**Figure 9–53 Negatively stained actin filaments.** In this transmission electron micrograph, each filament is about 8 nm in diameter and is seen, on close inspection, to be composed of a helical chain of globular actin molecules. (Courtesy of Roger Craig.)



the protein capsid inside a human immunodeficiency virus (HIV) that has been determined at high resolution by the combination of many particles, multiple views, and molecular modeling.

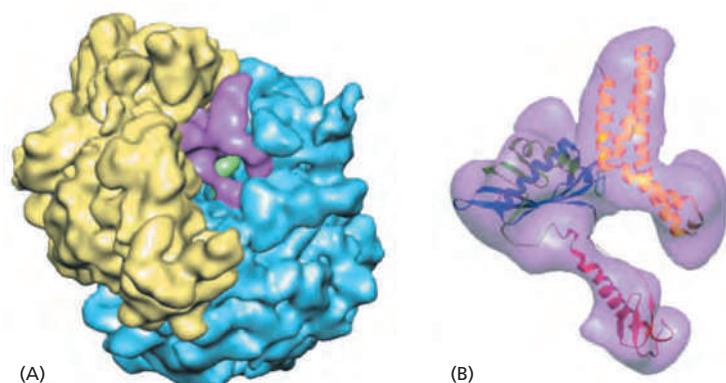
A resolution of 0.3 nm has been achieved by electron microscopy—enough to begin to see the internal atomic arrangements in a protein and to rival x-ray crystallography in resolution. Although electron microscopy is unlikely to supersede x-ray crystallography (discussed in Chapter 8) as a method for macromolecular structure determination, it has some very clear advantages. First, it does not absolutely require crystalline specimens. Second, it can deal with extremely large complexes—structures that may be too large or too variable to crystallize satisfactorily. Third, it allows the rapid analysis of different conformations of protein complexes.

The analysis of large and complex macromolecular structures is helped considerably if the atomic structure of one or more of the subunits is known, for example from x-ray crystallography. Molecular models can then be mathematically “fitted” into the envelope of the structure determined at lower resolution using the electron microscope (see Figures 16–16D and 16–46). **Figure 9–55** shows the structure of a ribosome with the location of a bound release factor displayed in this way (see also Figure 6–72).

## Summary

*Discovering the detailed structure of membranes and organelles requires the higher resolution attainable in a transmission electron microscope. Specific macromolecules can be localized after being labeled with colloidal gold linked to antibodies. Three-dimensional views of the surfaces of cells and tissues are obtained by scanning electron microscopy. The shapes of isolated molecules can be readily determined by electron microscopy techniques involving fast freezing or negative staining. Electron tomography and single-particle reconstruction use computational manipulations of data obtained from multiple images and multiple viewing angles to produce detailed reconstructions of macromolecules and molecular complexes. The resolution obtained with these methods means that atomic structures of individual*

**Figure 9–54 Single-particle reconstruction.** The structure of a complete human immunodeficiency virus (HIV) capsid has been determined by a combination of cryoelectron microscopy, protein structure determination, and modeling. (A) A single 4 nm slice from an EM tomographic model (see also Figure 9–48) of an intact HIV particle with its membrane outer envelope and its internal, irregularly shaped protein capsid that houses its RNA genome. (B) Electron microscopy of capsid subunits that have self-assembled into a helical tube can be used to derive an electron-density map at a resolution of 8 nm, in which details of the hexamers are clearly visible. (C) Using the known atomic coordinates of a single subunit of the hexamer, the structure has been modeled into the electron-density map from (B). (D) A molecular reconstruction of the entire HIV capsid, based on the detailed structures shown in (A) and (C). This capsid contains 216 hexamers (blue) and 12 pentamers (yellow). (Adapted from G. Zhao et al., *Nature* 497:643–646, 2013. With permission from Macmillan Publishers Ltd. C, PDB code: 3J34.)



**Figure 9–55 Single-particle reconstruction and molecular model fitting.** Bacterial ribosomes, with and without the release factor required for peptide release from the ribosome, were used to derive high-resolution, three-dimensional cryoelectron microscopy maps at a resolution of better than 1 nm. Images of nearly 20,000 separate ribosomes preserved in ice were used to produce single-particle reconstructions. (A) The 30S ribosomal subunit (yellow) and the 50S subunit (blue) can be distinguished from the additional electron density that can be attributed to the release factor RF2 (purple). (B) The known molecular structure of RF2 modeled into the electron density from (A). (From U.B.S. Rawat et al., *Nature* 421:87–90, 2003. With permission from Macmillan Publishers Ltd.)

*macromolecules can often be “fitted” to the images derived by electron microscopy. In this way, the TEM is increasingly able to bridge the gap between structures discovered by x-ray crystallography and those discovered with the light microscope.*

## PROBLEMS

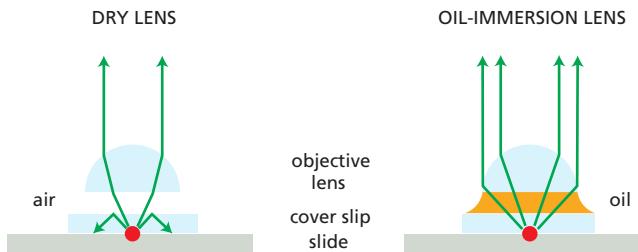
Which statements are true? Explain why or why not.

**9–1** Because the DNA double helix is only 2 nm wide—well below the limit of resolution of the light microscope—it is impossible to see chromosomes in living cells without special stains.

**9–2** A fluorescent molecule, having absorbed a single photon of light at one wavelength, always emits it at a longer wavelength.

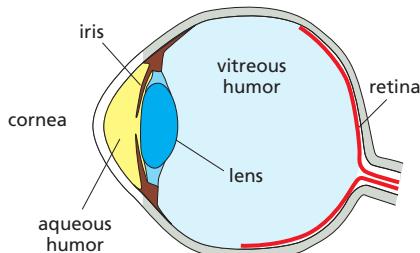
Discuss the following problems.

**9–3** The diagrams in **Figure Q9–1** show the paths of light rays passing through a specimen with a dry lens and with an oil-immersion lens. Offer an explanation for why oil-immersion lenses should give better resolution. Air, glass, and oil have refractive indices of 1.00, 1.51, and 1.51, respectively.



**Figure Q9–1** Paths of light rays through dry and oil-immersion lenses (Problem 9–3). The red circle at the origin of the light rays is the specimen.

**9–4** **Figure Q9–2** shows a diagram of the human eye. The refractive indices of the components in the light path are: cornea 1.38, aqueous humor 1.33, crystalline lens 1.41, and vitreous humor 1.38. Where does the main refraction—the main focusing—occur? What role do you suppose the lens plays?



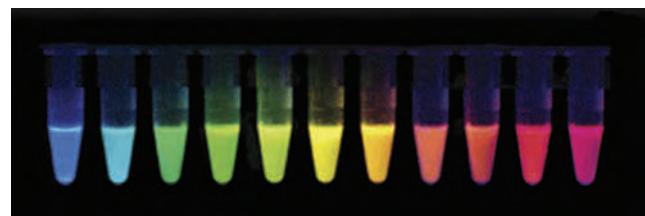
**Figure Q9–2** Diagram of the human eye (Problem 9–4).

**9–5** Why do humans see so poorly under water? And why do goggles help?

**9–6** Explain the difference between resolution and magnification.

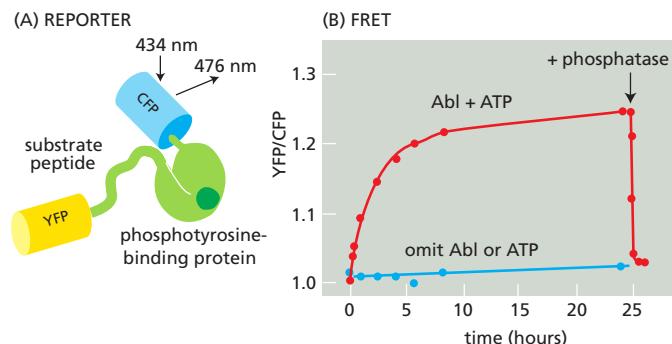
**9–7** Antibodies that bind to specific proteins are important tools for defining the locations of molecules in cells. The sensitivity of the primary antibody—the antibody that reacts with the target molecule—is often enhanced by using labeled secondary antibodies that bind to it. What are the advantages and disadvantages of using secondary antibodies that carry fluorescent tags versus those that carry bound enzymes?

**9–8** **Figure Q9–3** shows a series of modified fluorescent proteins that emit light in a range of colors. How do you suppose the exact same chromophore can fluoresce at so many different wavelengths?



**Figure Q9–3** A rainbow of colors produced by modified fluorescent proteins (Problem 9–8). (Courtesy of Nathan Shaner, Paul Steinbach and Roger Tsien.)

**9–9** Consider a fluorescent detector designed to report the cellular location of active protein tyrosine kinases. A blue (cyan) fluorescent protein (CFP) and a yellow fluorescent protein (YFP) were fused to either end of a hybrid protein domain. The hybrid protein segment consisted of a substrate peptide recognized by the Abl protein tyrosine kinase and a phosphotyrosine-binding domain (**Figure Q9–4A**). Stimulation of the CFP domain does not cause emission by the YFP domain when the domains are



**Figure Q9–4** Fluorescent reporter protein designed to detect tyrosine phosphorylation (Problem 9–9). (A) Domain structure of reporter protein. Four domains are indicated: CFP, YFP, tyrosine kinase substrate peptide, and a phosphotyrosine-binding domain. (B) FRET assay. YFP/CFP is normalized to 1.0 at time zero. The reporter was incubated in the presence (or absence) of Abl and ATP for the indicated times. Arrow indicates time of addition of a tyrosine phosphatase. (From A.Y. Ting, K.H. Kain, R.L. Klemke and R.Y. Tsien, *Proc. Natl Acad. Sci. USA* 98:15003–15008, 2001. With permission from National Academy of Sciences.)

separated. When the CFP and YFP domains are brought close together, however, fluorescence resonance energy transfer (FRET) allows excitation of CFP to stimulate emission by YFP. FRET shows up experimentally as an increase in the ratio of emission at 526 nm versus 476 nm (YFP/CFP) when CFP is excited by 434 nm light.

## REFERENCES

### General

- Celis JE, Carter N, Simons K et al. (eds) (2005) Cell Biology: A Laboratory Handbook, 3rd ed. San Diego: Academic Press. (Volume 3 of this four-volume set covers the practicalities of most of the current light and electron imaging methods that are used in cell biology.)  
 Pawley BP (ed) (2006) Handbook of Biological Confocal Microscopy, 3rd ed. New York: Springer Science.  
 Wayne R (2014) Light and Video Microscopy. San Diego: Academic Press.

### Looking at Cells in the Light Microscope

- Adams MC, Salmon WC, Gupton SL et al. (2003) A high-speed multispectral spinning-disk confocal microscope system for fluorescent speckle microscopy of living cells. *Methods* 29, 29–41.  
 Agard DA, Hiraoka Y, Shaw P & Sedat JW (1989) Fluorescence microscopy in three dimensions. In *Methods in Cell Biology*, Vol. 30: Fluorescence Microscopy of Living Cells in Culture, part B (DL Taylor, Y-L Wang eds). San Diego: Academic Press.  
 Burnette DT, Sengupta P, Dai Y et al. (2011) Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proc. Natl Acad. Sci. USA* 108, 21081–21086.  
 Chalfie M, Tu Y, Euskirchen G et al. (1994) Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805.  
 Giepmans BN, Adams SR, Ellisman MH & Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312, 217–224.  
 Harlow E & Lane D (1998) Using Antibodies: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.  
 Hell S (2009) Microscopy and its focal switch. *Nat. Methods* 6, 24–32.  
 Huang B, Babcock H & Zhuang X (2010) Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* 143, 1047–1058.  
 Huang B, Bates M & Zhuang X (2009) Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* 78, 993–1016.  
 Jaiswai JK & Simon SM (2004) Potentials and pitfalls of fluorescent quantum dots for biological imaging. *Trends Cell Biol.* 14, 497–504.  
 Klar TA, Jakobs S, Dyba M et al. (2000) Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl Acad. Sci. USA* 97, 8206–8210.  
 Lippincott-Schwartz J & Patterson GH (2003) Development and use of fluorescent protein markers in living cells. *Science* 300, 87–91.  
 Lippincott-Schwartz J, Altan-Bonnet N & Patterson G (2003) Photobleaching and photoactivation: following protein dynamics in living cells. *Nat. Cell Biol.* 5(Suppl), S7–S14.  
 McEvoy AL, Greenfield D, Bates M & Liphardt J (2010) Q&A: Single-molecule localization microscopy for biological imaging. *BMC Biol.* 8, 106.  
 Minsky M (1988) Memoir on inventing the confocal scanning microscope. *Scanning* 10, 128–138.  
 Miyawaki A, Sawano A & Kogure T (2003) Lighting up cells: labelling proteins with fluorophores. *Nat. Cell Biol.* 5(Suppl), S1–S7.  
 Parton RM & Read ND (1999) Calcium and pH imaging in living cells. In *Light Microscopy in Biology. A Practical Approach*, 2nd ed. (Lacey AJ ed.). Oxford: Oxford University Press.  
 Patterson G, Davidson M, Manley S & Lippincott-Schwartz J (2010) Superresolution imaging using single-molecule localization. *Annu. Rev. Phys. Chem.* 61, 345–367.  
 Rust MJ, Bates M & Zhuang X (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* 3, 793–795.

Incubation of the reporter protein with Abl protein tyrosine kinase in the presence of ATP gave an increase in YFP/CFP emission (**Figure Q9–4B**). In the absence of ATP or the Abl protein, no FRET occurred. FRET was also eliminated by addition of a tyrosine phosphatase (Figure Q9–4B). Describe as best you can how the reporter protein detects active Abl protein tyrosine kinase.

- Sako Y & Yanagida T (2003) Single-molecule visualization in cell biology. *Nat. Rev. Mol. Cell Biol.* 4(Suppl), SS1–SS5.  
 Schermelleh L, Heintzmann R & Leonhardt H (2010) A guide to super-resolution fluorescence microscopy. *J. Cell Biol.* 190, 165–175.  
 Shaner NC, Steinbach PA & Tsien RY (2005) A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909.  
 Sluder G & Wolf DE (2007) Digital Microscopy, 3rd ed: *Methods in Cell Biology*, Vol 81. San Diego: Academic Press.  
 Stephens DJ & Allan VJ (2003) Light microscopy techniques for live cell imaging. *Science* 300, 82–86.  
 Tsien RY (2008) Constructing and exploiting the fluorescent protein paintbox. Nobel Prize Lecture. [www.nobelprize.org](http://www.nobelprize.org)  
 White JG, Amos WB & Fordham M (1987) An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *J. Cell Biol.* 105, 41–48.  
 Zernike F (1955) How I discovered phase contrast. *Science* 121, 345–349.

### Looking at Cells and Molecules in the Electron Microscope

- Allen TD & Goldberg MW (1993) High-resolution SEM in cell biology. *Trends Cell Biol.* 3, 205–208.  
 Baumeister W (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Curr. Opin. Struct. Biol.* 12, 679–684.  
 Böttcher B, Wynne SA & Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386, 88–91.  
 Dubochet J, Adrian M, Chang J-J et al. (1988) Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.  
 Frank J (2003) Electron microscopy of functional ribosome complexes. *Biopolymers* 68, 223–233.  
 Frank J (2009) Single-particle reconstruction of biological macromolecules in electron microscopy—30 years. *Quart. Rev. Biophys.* 42, 139–158.  
 Hayat MA (2000) Principles and Techniques of Electron Microscopy, 4th ed. Cambridge: Cambridge University Press.  
 Heuser J (1981) Quick-freeze, deep-etch preparation of samples for 3-D electron microscopy. *Trends Biochem. Sci.* 6, 64–68.  
 Liao M, Cao E, Julius D & Cheng Y (2014) Single particle electron cryo-microscopy of a mammalian ion channel. *Curr. Opin. Struct. Biol.* 27, 1–7.  
 Lucic V, Förster F & Baumeister W (2005) Structural studies by electron tomography: from cells to molecules. *Annu. Rev. Biochem.* 74, 833–865.  
 McDonald KL & Auer M (2006) High-pressure freezing, cellular tomography, and structural cell biology. *Biotechniques* 41, 137–139.  
 McIntosh JR (2007) Cellular Electron Microscopy. 3rd ed: *Methods in Cell Biology*, Vol 79. San Diego: Academic Press.  
 McIntosh R, Nicastro D & Mastronarde D (2005) New views of cells in 3D: an introduction to electron tomography. *Trends Cell Biol.* 15, 43–51.  
 Pease DC & Porter KR (1981) Electron microscopy and ultramicrotomy. *J. Cell Biol.* 91, 287s–292s.  
 Unwin PNT & Henderson R (1975) Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* 94, 425–440.  
 Zhou ZH (2008) Towards atomic resolution structural determination by single particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.* 18, 218–228.

# INTERNAL ORGANIZATION OF THE CELL

CHAPTER

## 10

### Membrane Structure

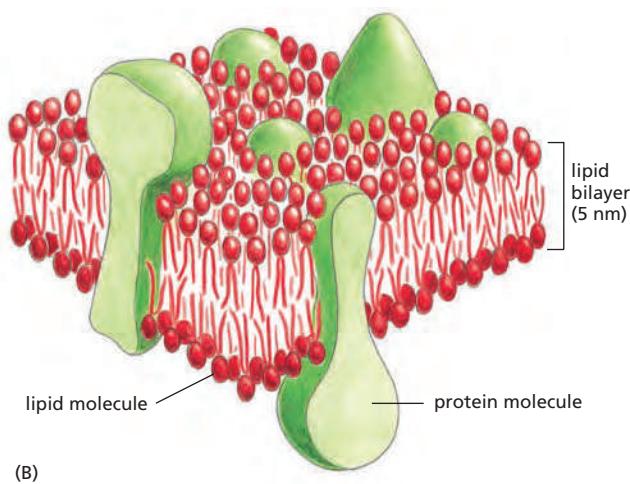
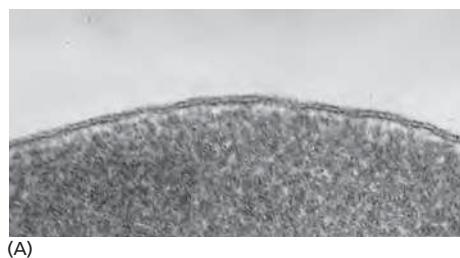
Cell membranes are crucial to the life of the cell. The **plasma membrane** encloses the cell, defines its boundaries, and maintains the essential differences between the cytosol and the extracellular environment. Inside eukaryotic cells, the membranes of the nucleus, endoplasmic reticulum, Golgi apparatus, mitochondria, and other membrane-enclosed organelles maintain the characteristic differences between the contents of each organelle and the cytosol. Ion gradients across membranes, established by the activities of specialized membrane proteins, can be used to synthesize ATP, to drive the transport of selected solutes across the membrane, or, as in nerve and muscle cells, to produce and transmit electrical signals. In all cells, the plasma membrane also contains proteins that act as sensors of external signals, allowing the cell to change its behavior in response to environmental cues, including signals from other cells; these protein sensors, or *receptors*, transfer information—rather than molecules—across the membrane.

Despite their differing functions, all biological membranes have a common general structure: each is a very thin film of lipid and protein molecules, held together mainly by noncovalent interactions (Figure 10–1). Cell membranes

#### IN THIS CHAPTER

#### THE LIPID BILAYER

#### MEMBRANE PROTEINS



**Figure 10–1** Two views of a cell membrane. (A) An electron micrograph of a segment of the plasma membrane of a human red blood cell seen in cross section, showing its bilayer structure. (B) A three-dimensional schematic view of a cell membrane and the general disposition of its lipid and protein constituents. (A, courtesy of Daniel S. Friend.)

are dynamic, fluid structures, and most of their molecules move about in the plane of the membrane. The lipid molecules are arranged as a continuous double layer about 5 nm thick. This *lipid bilayer* provides the basic fluid structure of the membrane and serves as a relatively impermeable barrier to the passage of most water-soluble molecules. Most *membrane proteins* span the lipid bilayer and mediate nearly all of the other functions of the membrane, including the transport of specific molecules across it, and the catalysis of membrane-associated reactions such as ATP synthesis. In the plasma membrane, some transmembrane proteins serve as structural links that connect the cytoskeleton through the lipid bilayer to either the extracellular matrix or an adjacent cell, while others serve as receptors to detect and transduce chemical signals in the cell's environment. It takes many kinds of membrane proteins to enable a cell to function and interact with its environment, and it is estimated that about 30% of the proteins encoded in an animal's genome are membrane proteins.

In this chapter, we consider the structure and organization of the two main constituents of biological membranes—the lipids and the proteins. Although we focus mainly on the plasma membrane, most concepts discussed apply to the various internal membranes of eukaryotic cells as well. The functions of cell membranes are considered in later chapters: their role in energy conversion and ATP synthesis, for example, is discussed in Chapter 14; their role in the transmembrane transport of small molecules in Chapter 11; and their roles in cell signaling and cell adhesion in Chapters 15 and 19, respectively. In Chapters 12 and 13, we discuss the internal membranes of the cell and the protein traffic through and between them.

## THE LIPID BILAYER

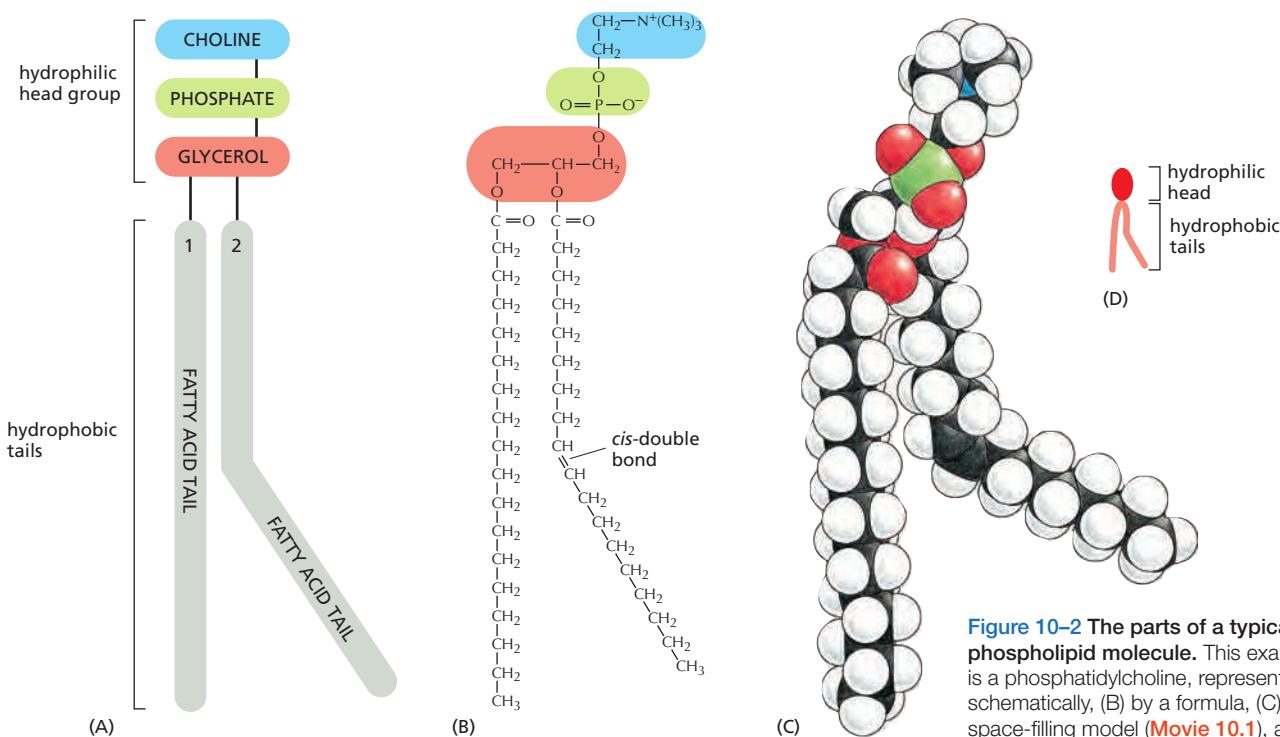
The **lipid bilayer** provides the basic structure for all cell membranes. It is easily seen by electron microscopy, and its bilayer structure is attributable exclusively to the special properties of the lipid molecules, which assemble spontaneously into bilayers even under simple artificial conditions. In this section, we discuss the different types of lipid molecules found in cell membranes and the general properties of lipid bilayers.

### Phosphoglycerides, Sphingolipids, and Sterols Are the Major Lipids in Cell Membranes

Lipid molecules constitute about 50% of the mass of most animal cell membranes, nearly all of the remainder being protein. There are approximately  $5 \times 10^6$  lipid molecules in a  $1 \mu\text{m} \times 1 \mu\text{m}$  area of lipid bilayer, or about  $10^9$  lipid molecules in the plasma membrane of a small animal cell. All of the lipid molecules in cell membranes are **amphiphilic**—that is, they have a **hydrophilic** (“water-loving”) or **polar** end and a **hydrophobic** (“water-fearing”) or **nonpolar** end.

The most abundant membrane lipids are the **phospholipids**. These have a polar head group containing a phosphate group and two hydrophobic *hydrocarbon tails*. In animal, plant, and bacterial cells, the tails are usually fatty acids, and they can differ in length (they normally contain between 14 and 24 carbon atoms). One tail typically has one or more *cis*-double bonds (that is, it is *unsaturated*), while the other tail does not (that is, it is *saturated*). As shown in **Figure 10–2**, each *cis*-double bond creates a kink in the tail. Differences in the length and saturation of the fatty acid tails influence how phospholipid molecules pack against one another, thereby affecting the fluidity of the membrane, as we discuss later.

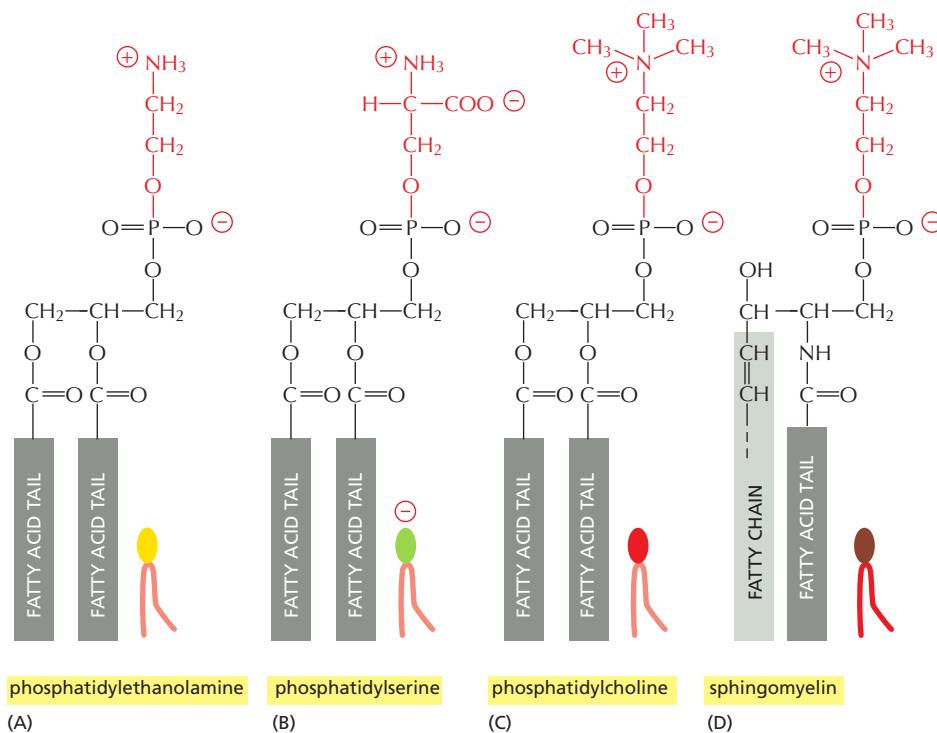
The main phospholipids in most animal cell membranes are the **phosphoglycerides**, which have a three-carbon *glycerol* backbone (see Figure 10–2). Two long-chain fatty acids are linked through ester bonds to adjacent carbon atoms of the glycerol, and the third carbon atom of the glycerol is attached to a phosphate group, which in turn is linked to one of several types of head group. By combining several different fatty acids and head groups, cells make many different phosphoglycerides. *Phosphatidylethanolamine*, *phosphatidylserine*, and



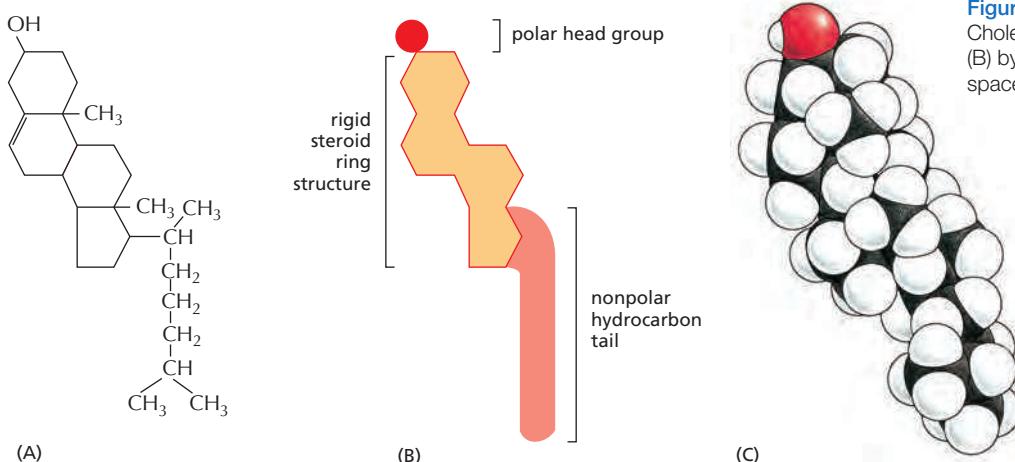
**Figure 10-2** The parts of a typical phospholipid molecule. This example is a phosphatidylcholine, represented (A) schematically, (B) by a formula, (C) as a space-filling model (Movie 10.1), and (D) as a symbol.

phosphatidylcholine are the most abundant ones in mammalian cell membranes (Figure 10-3A–C).

Another important class of phospholipids are the *sphingolipids*, which are built from *sphingosine* rather than glycerol (Figure 10-3D–E). Sphingosine is a long acyl chain with an amino group ( $\text{NH}_2$ ) and two hydroxyl groups ( $\text{OH}$ ) at one end. In sphingomyelin, the most common sphingolipid, a fatty acid tail is attached to the amino group, and a phosphocholine group is attached to the terminal hydroxyl group. Together, the phospholipids phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and sphingomyelin constitute more than half the mass of lipid in most mammalian cell membranes (see Table 10-1, p. 571).



**Figure 10-3** Four major phospholipids in mammalian plasma membranes. Different head groups are represented by different colors in the symbols. The lipid molecules shown in (A–C) are phosphoglycerides, which are derived from glycerol. The molecule in (D) is sphingomyelin, which is derived from sphingosine (E) and is therefore a sphingolipid. Note that only phosphatidylserine carries a net negative charge, the importance of which we discuss later; the other three are electrically neutral at physiological pH, carrying one positive and one negative charge.



**Figure 10-4** The structure of cholesterol. Cholesterol is represented (A) by a formula, (B) by a schematic drawing, and (C) as a space-filling model.

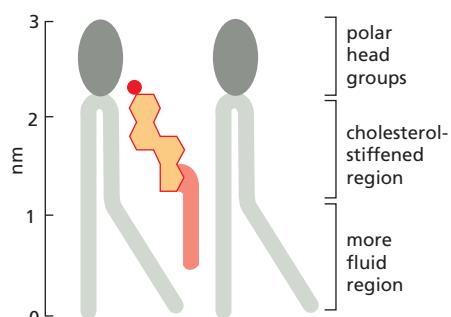
In addition to phospholipids, the lipid bilayers in many cell membranes contain *glycolipids* and *cholesterol*. Glycolipids resemble sphingolipids, but, instead of a phosphate-linked head group, they have sugars attached. We discuss glycolipids later. Eukaryotic plasma membranes contain especially large amounts of **cholesterol**—up to one molecule for every phospholipid molecule. Cholesterol is a sterol. It contains a rigid ring structure, to which is attached a single polar hydroxyl group and a short nonpolar hydrocarbon chain (Figure 10-4). The cholesterol molecules orient themselves in the bilayer with their hydroxyl group close to the polar head groups of adjacent phospholipid molecules (Figure 10-5).

### Phospholipids Spontaneously Form Bilayers

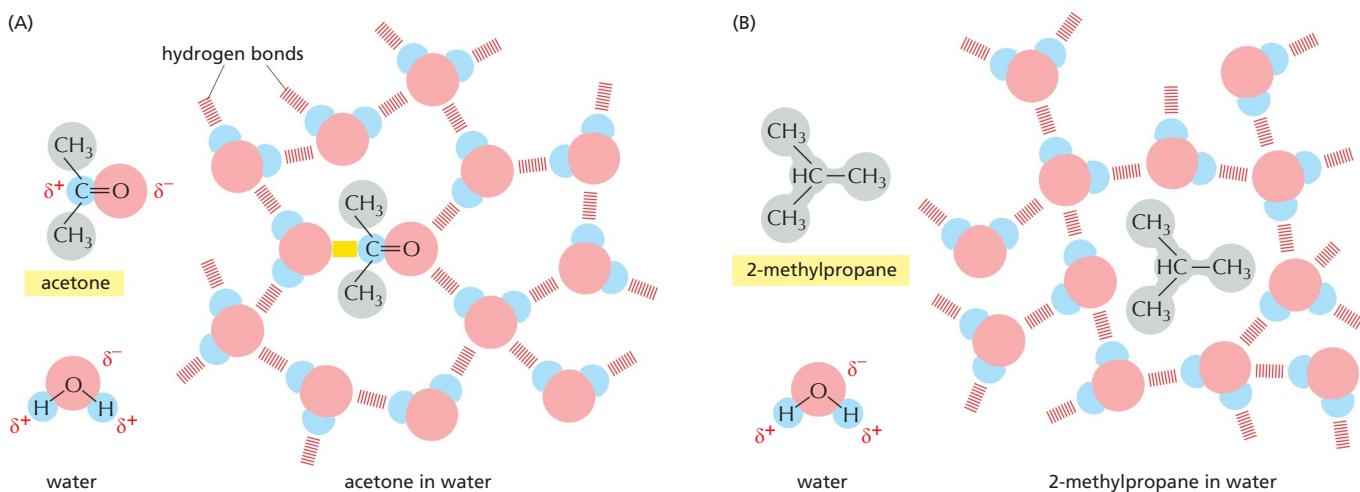
The shape and amphiphilic nature of the phospholipid molecules cause them to form bilayers spontaneously in aqueous environments. As discussed in Chapter 2, hydrophilic molecules dissolve readily in water because they contain charged groups or uncharged polar groups that can form either favorable electrostatic interactions or hydrogen bonds with water molecules (Figure 10-6A). Hydrophobic molecules, by contrast, are insoluble in water because all, or almost all, of their atoms are uncharged and nonpolar and therefore cannot form energetically favorable interactions with water molecules. If dispersed in water, they force the adjacent water molecules to reorganize into icelike cages that surround the hydrophobic molecule (Figure 10-6B). Because these cage structures are more ordered than the surrounding water, their formation increases the free energy. This free-energy cost is minimized, however, if the hydrophobic molecules (or the hydrophobic portions of amphiphilic molecules) cluster together so that the smallest number of water molecules is affected.

When amphiphilic molecules are exposed to an aqueous environment, they behave as you would expect from the above discussion. They spontaneously aggregate to bury their hydrophobic tails in the interior, where they are shielded from the water, and they expose their hydrophilic heads to water. Depending on their shape, they can do this in either of two ways: they can form spherical *micelles*, with the tails inward, or they can form double-layered sheets, or *bilayers*, with the hydrophobic tails sandwiched between the hydrophilic head groups (Figure 10-7).

The same forces that drive phospholipids to form bilayers also provide a self-sealing property. A small tear in the bilayer creates a free edge with water; because this is energetically unfavorable, the lipids tend to rearrange spontaneously to eliminate the free edge. (In eukaryotic plasma membranes, the fusion of intracellular vesicles repairs larger tears.) The prohibition of free edges has a profound consequence: the only way for a bilayer to avoid having edges is by closing in on itself and forming a sealed compartment (Figure 10-8). This remarkable



**Figure 10-5** Cholesterol in a lipid bilayer. Schematic drawing (to scale) of a cholesterol molecule interacting with two phospholipid molecules in one monolayer of a lipid bilayer.



behavior, fundamental to the creation of a living cell, follows directly from the shape and amphiphilic nature of the phospholipid molecule.

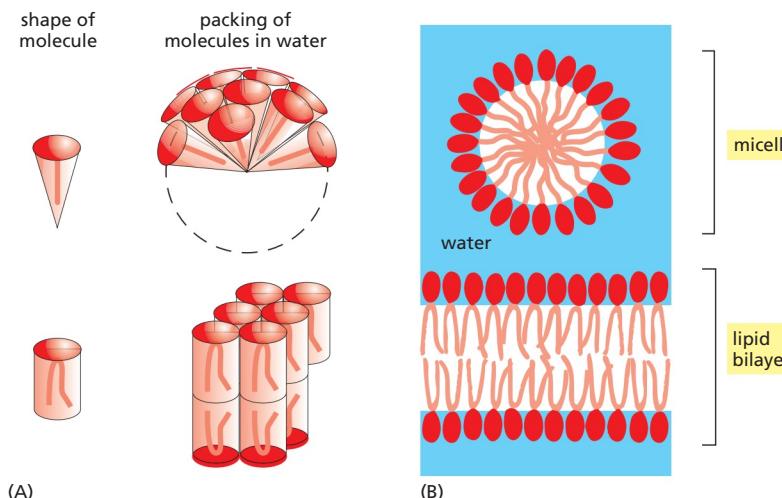
A lipid bilayer also has other characteristics that make it an ideal structure for cell membranes. One of the most important of these is its fluidity, which is crucial to many membrane functions ([Movie 10.2](#)).

### The Lipid Bilayer Is a Two-dimensional Fluid

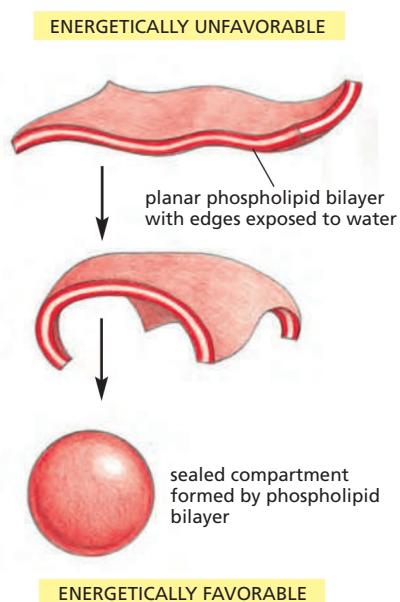
Around 1970, researchers first recognized that individual lipid molecules are able to diffuse freely within the plane of a lipid bilayer. The initial demonstration came from studies of synthetic (artificial) lipid bilayers, which can be made in the form of spherical vesicles, called **liposomes** ([Figure 10-9](#)); or in the form of planar bilayers formed across a hole in a partition between two aqueous compartments or on a solid support.

Various techniques have been used to measure the motion of individual lipid molecules and their components. One can construct a lipid molecule, for example, with a fluorescent dye or a small gold particle attached to its polar head group and follow the diffusion of even individual molecules in a membrane. Alternatively, one can modify a lipid head group to carry a “spin label,” such as a nitroxide

**Figure 10-6 How hydrophilic and hydrophobic molecules interact differently with water.** (A) Because acetone is polar, it can form hydrogen bonds (red) and favorable electrostatic interactions (yellow) with water molecules, which are also polar. Thus, acetone readily dissolves in water. (B) By contrast, 2-methylpropane is entirely hydrophobic. Because it cannot form favorable interactions with water, it forces adjacent water molecules to reorganize into icelike cage structures, which increases the free energy. This compound is therefore virtually insoluble in water. The symbol  $\delta^-$  indicates a partial negative charge, and  $\delta^+$  indicates a partial positive charge. Polar atoms are shown in color and nonpolar groups are shown in gray.



**Figure 10-7** Packing arrangements of amphiphilic molecules in an aqueous environment. (A) These molecules spontaneously form micelles or bilayers in water, depending on their shape. Cone-shaped amphiphilic molecules (above) form micelles, whereas cylinder-shaped amphiphilic molecules such as phospholipids (below) form bilayers. (B) A micelle and a lipid bilayer seen in cross section. Note that micelles of amphiphilic molecules are thought to be much more irregular than drawn here (see Figure 10-26C).



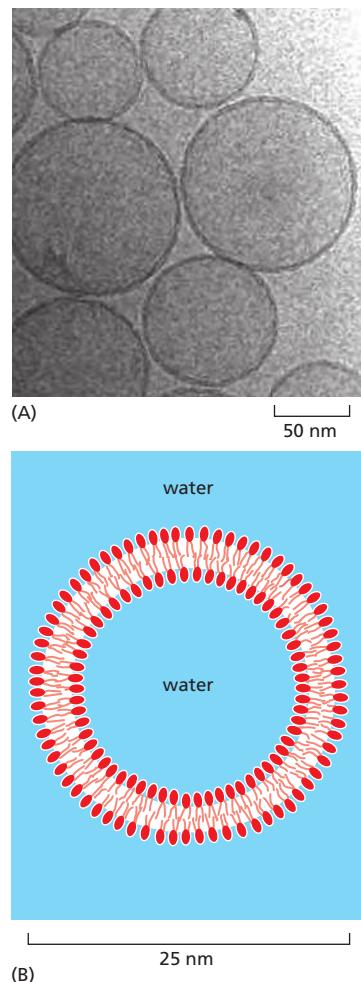
**Figure 10-8** The spontaneous closure of a phospholipid bilayer to form a sealed compartment. The closed structure is stable because it avoids the exposure of the hydrophobic hydrocarbon tails to water, which would be energetically unfavorable.

**Figure 10–9 Liposomes.** (A) An electron micrograph of unfixed, unstained, synthetic phospholipid vesicles—liposomes—in water, which have been rapidly frozen at liquid-nitrogen temperature. (B) A drawing of a small spherical liposome seen in cross section. Liposomes are commonly used as model membranes in experimental studies, especially to study incorporated membrane proteins. (A, from P. Frederik and D. Hubert, *Methods Enzymol.* 391:431–448, 2005. With permission from Elsevier.)

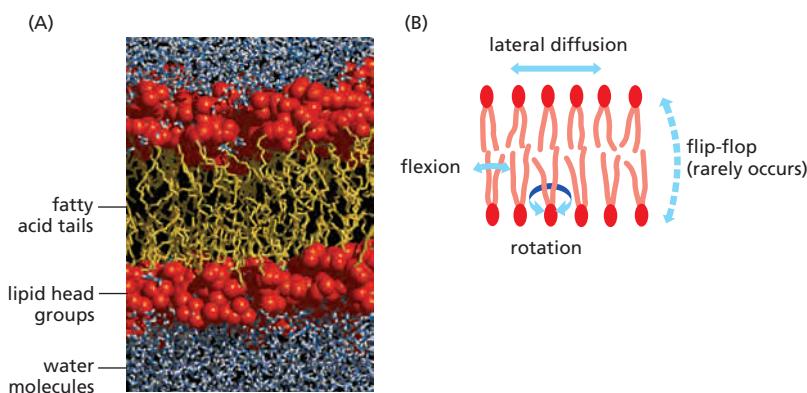
group (=N–O); this contains an unpaired electron whose spin creates a paramagnetic signal that can be detected by electron spin resonance (ESR) spectroscopy, the principles of which are similar to those of nuclear magnetic resonance (NMR), discussed in Chapter 8. The motion and orientation of a spin-labeled lipid in a bilayer can be deduced from the ESR spectrum. Such studies show that phospholipid molecules in synthetic bilayers very rarely migrate from the monolayer (also called a *leaflet*) on one side to that on the other. This process, known as “flip-flop,” occurs on a time scale of hours for any individual molecule, although cholesterol is an exception to this rule and can flip-flop rapidly. In contrast, lipid molecules rapidly exchange places with their neighbors *within* a monolayer (~10<sup>7</sup> times per second). This gives rise to a rapid lateral diffusion, with a diffusion coefficient (*D*) of about 10<sup>-8</sup> cm<sup>2</sup>/sec, which means that an average lipid molecule diffuses the length of a large bacterial cell (~2 μm) in about 1 second. These studies have also shown that individual lipid molecules rotate very rapidly about their long axis and have flexible hydrocarbon chains. Computer simulations show that lipid molecules in synthetic bilayers are very disordered, presenting an irregular surface of variously spaced and oriented head groups to the water phase on either side of the bilayer (Figure 10–10).

Similar mobility studies on labeled lipid molecules in isolated biological membranes and in living cells give results similar to those in synthetic bilayers. They demonstrate that the lipid component of a biological membrane is a two-dimensional liquid in which the constituent molecules are free to move laterally. As in synthetic bilayers, individual phospholipid molecules are normally confined to their own monolayer. This confinement creates a problem for their synthesis. Phospholipid molecules are manufactured in only one monolayer of a membrane, mainly in the cytosolic monolayer of the endoplasmic reticulum membrane. If none of these newly made molecules could migrate reasonably promptly to the noncytosolic monolayer, new lipid bilayer could not be made. The problem is solved by a special class of membrane proteins called *phospholipid translocators*, or *flippases*, which catalyze the rapid flip-flop of phospholipids from one monolayer to the other, as discussed in Chapter 12.

Despite the fluidity of the lipid bilayer, liposomes do not fuse spontaneously with one another when suspended in water. Fusion does not occur because the polar lipid head groups bind water molecules that need to be displaced for the bilayers of two different liposomes to fuse. The hydration shell that keeps liposomes apart also insulates the many internal membranes in a eukaryotic cell and prevents their uncontrolled fusion, thereby maintaining the compartmental integrity of membrane-enclosed organelles. All cell membrane fusion events



**Figure 10–10 The mobility of phospholipid molecules in an artificial lipid bilayer.** Starting with a model of 100 phosphatidylcholine molecules arranged in a regular bilayer, a computer calculated the position of every atom after 300 picoseconds of simulated time. From these theoretical calculations, a model of the lipid bilayer emerges that accounts for almost all of the measurable properties of a synthetic lipid bilayer, including its thickness, number of lipid molecules per membrane area, depth of water penetration, and unevenness of the two surfaces. Note that the tails in one monolayer can interact with those in the other monolayer, if the tails are long enough. (B) The different motions of a lipid molecule in a bilayer. (A, based on S.W. Chiu et al., *Biophys. J.* 69:1230–1245, 1995. With permission from the Biophysical Society.)



are catalyzed by tightly regulated fusion proteins, which force appropriate membranes into tight proximity, squeezing out the water layer that keeps the bilayers apart, as we discuss in Chapter 13.

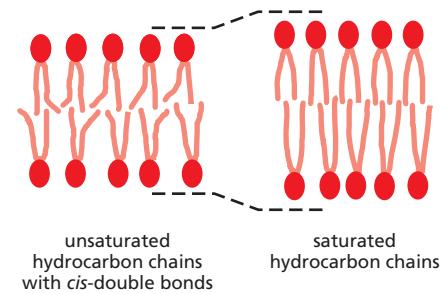
### The Fluidity of a Lipid Bilayer Depends on Its Composition

The fluidity of cell membranes has to be precisely regulated. Certain membrane transport processes and enzyme activities, for example, cease when the bilayer viscosity is experimentally increased beyond a threshold level.

The fluidity of a lipid bilayer depends on both its composition and its temperature, as is readily demonstrated in studies of synthetic lipid bilayers. A synthetic bilayer made from a single type of phospholipid changes from a liquid state to a two-dimensional rigid crystalline (or gel) state at a characteristic temperature. This change of state is called a *phase transition*, and the temperature at which it occurs is lower (that is, the membrane becomes more difficult to freeze) if the hydrocarbon chains are short or have double bonds. A shorter chain length reduces the tendency of the hydrocarbon tails to interact with one another, in both the same and opposite monolayer, and *cis*-double bonds produce kinks in the chains that make them more difficult to pack together, so that the membrane remains fluid at lower temperatures (Figure 10–11). Bacteria, yeasts, and other organisms whose temperature fluctuates with that of their environment adjust the fatty acid composition of their membrane lipids to maintain a relatively constant fluidity. As the temperature falls, for instance, the cells of those organisms synthesize fatty acids with more *cis*-double bonds, thereby avoiding the decrease in bilayer fluidity that would otherwise result from the temperature drop.

Cholesterol modulates the properties of lipid bilayers. When mixed with phospholipids, it enhances the permeability-barrier properties of the lipid bilayer. Cholesterol inserts into the bilayer with its hydroxyl group close to the polar head groups of the phospholipids, so that its rigid, platelike steroid rings interact with—and partly immobilize—those regions of the hydrocarbon chains closest to the polar head groups (see Figure 10–5 and Movie 10.3). By decreasing the mobility of the first few CH<sub>2</sub> groups of the chains of the phospholipid molecules, cholesterol makes the lipid bilayer less deformable in this region and thereby decreases the permeability of the bilayer to small water-soluble molecules. Although cholesterol tightens the packing of the lipids in a bilayer, it does not make membranes any less fluid. At the high concentrations found in most eukaryotic plasma membranes, cholesterol also prevents the hydrocarbon chains from coming together and crystallizing.

**Table 10–1** compares the lipid compositions of several biological membranes. Note that bacterial plasma membranes are often composed of one main type of phospholipid and contain no cholesterol. In archaea, lipids usually contain



**Figure 10–11** The influence of *cis*-double bonds in hydrocarbon chains.

The double bonds make it more difficult to pack the chains together, thereby making the lipid bilayer more difficult to freeze. In addition, because the hydrocarbon chains of unsaturated lipids are more spread apart, lipid bilayers containing them are thinner than bilayers formed exclusively from saturated lipids.

**TABLE 10–1** Approximate Lipid Compositions of Different Cell Membranes

Lipid	Percentage of total lipid by weight					
	Liver cell plasma membrane	Red blood cell plasma membrane	Myelin	Mitochondrion (inner and outer membranes)	Endoplasmic reticulum	<i>E. coli</i> bacterium
Cholesterol	17	23	22	3	6	0
Phosphatidylethanolamine	7	18	15	28	17	70
Phosphatidylserine	4	7	9	2	5	trace
Phosphatidylcholine	24	17	10	44	40	0
Sphingomyelin	19	18	8	0	5	0
Glycolipids	7	3	28	trace	trace	0
Others	22	14	8	23	27	30

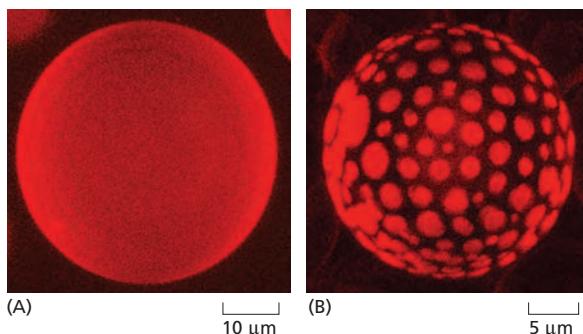
20–25-carbon-long prenyl chains instead of fatty acids; prenyl and fatty acid chains are similarly hydrophobic and flexible (see Figure 10–20F); in thermo-philic archaea, the longest lipid chains span both leaflets, making the membrane particularly stable to heat. Thus, lipid bilayers can be built from molecules with similar features but different molecular designs. The plasma membranes of most eukaryotic cells are more varied than those of prokaryotes and archaea, not only in containing large amounts of cholesterol but also in containing a mixture of different phospholipids.

Analysis of membrane lipids by mass spectrometry has revealed that the lipid composition of a typical eukaryotic cell membrane is much more complex than originally thought. These membranes contain a bewildering variety of perhaps 500–2000 different lipid species with even the simple plasma membrane of a red blood cell containing well over 150. While some of this complexity reflects the combinatorial variation in head groups, hydrocarbon chain lengths, and desaturation of the major phospholipid classes, some membranes also contain many structurally distinct minor lipids, at least some of which have important functions. The *inositol phospholipids*, for example, are present in small quantities in animal cell membranes and have crucial functions in guiding membrane traffic and in cell signaling (discussed in Chapters 13 and 15, respectively). Their local synthesis and destruction are regulated by a large number of enzymes, which create both small intracellular signaling molecules and lipid docking sites on membranes that recruit specific proteins from the cytosol, as we discuss later.

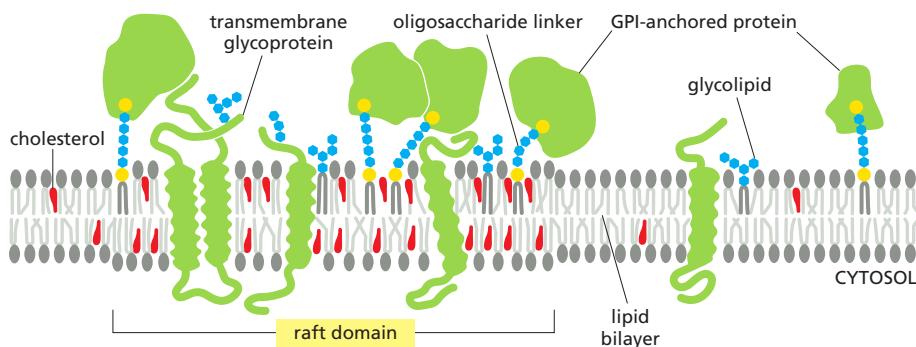
### Despite Their Fluidity, Lipid Bilayers Can Form Domains of Different Compositions

Because a lipid bilayer is a two-dimensional fluid, we might expect most types of lipid molecules in it to be well mixed and randomly distributed in their own monolayer. The van der Waals attractive forces between neighboring hydrocarbon tails are not selective enough to hold groups of phospholipid molecules together. With certain lipid mixtures in artificial bilayers, however, one can observe phase segregations in which specific lipids come together in separate domains (**Figure 10–12**).

There has been a long debate among cell biologists about whether the lipid molecules in the plasma membrane of living cells similarly segregate into specialized domains, called **lipid rafts**. Although many lipids and membrane proteins are not distributed uniformly, large-scale lipid phase segregations are rarely seen in living cell membranes. Instead, specific membrane proteins and lipids are seen to concentrate in a more temporary, dynamic fashion facilitated by protein-protein interactions that allow the transient formation of specialized membrane regions (**Figure 10–13**). Such clusters can be tiny nanoclusters on a scale of a few molecules, or larger assemblies that can be seen with electron microscopy, such as the *caveolae* involved in endocytosis (discussed in Chapter 13). The tendency of mixtures of lipids to undergo phase partitioning, as seen in artificial bilayers (see **Figure 10–12**), may help create rafts in living cell membranes—organizing and concentrating membrane proteins either for transport in membrane vesicles



**Figure 10–12** Lateral phase separation in artificial lipid bilayers. (A) Giant liposomes produced from a 1:1 mixture of phosphatidylcholine and sphingomyelin form uniform bilayers. (B) By contrast, liposomes produced from a 1:1:1 mixture of phosphatidylcholine, sphingomyelin, and cholesterol form bilayers with two separate phases. The liposomes are stained with trace concentrations of a fluorescent dye that preferentially partitions into one of the two phases. The average size of the domains formed in these giant artificial liposomes is much larger than that expected in cell membranes, where “lipid rafts” (see text) may be as small as a few nanometers in diameter. (A, from N. Kahya et al., *J. Struct. Biol.* 147:77–89, 2004. With permission from Elsevier; B, courtesy of Petra Schwille.)



**Figure 10–13 A model of a raft domain.** Weak protein–protein, protein–lipid, and lipid–lipid interactions reinforce one another to partition the interacting components into raft domains. Cholesterol, sphingolipids, glycolipids, glycosylphosphatidylinositol (GPI)-anchored proteins, and some transmembrane proteins are enriched in these domains. Note that because of their composition, raft domains have an increased membrane thickness. We discuss glycolipids, GPI-anchored proteins, and oligosaccharide linkers later. (Adapted from D. Lingwood and K. Simons, *Science* 327:46–50, 2010.)

(discussed in Chapter 13) or for working together in protein assemblies, such as when they convert extracellular signals into intracellular ones (discussed in Chapter 15).

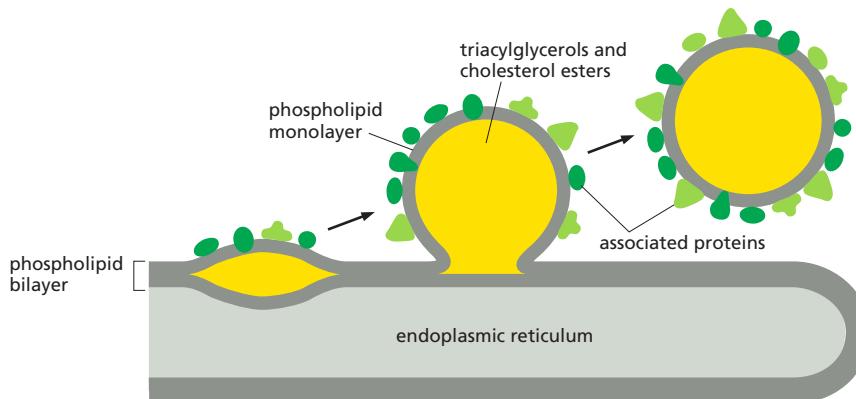
### Lipid Droplets Are Surrounded by a Phospholipid Monolayer

Most cells store an excess of lipids in **lipid droplets**, from where they can be retrieved as building blocks for membrane synthesis or as a food source. Fat cells, or *adipocytes*, are specialized for lipid storage. They contain a giant lipid droplet that fills up most of their cytoplasm. Most other cells have many smaller lipid droplets, the number and size varying with the cell's metabolic state. Fatty acids can be liberated from lipid droplets on demand and exported to other cells through the bloodstream. Lipid droplets store neutral lipids, such as triacylglycerols and cholesterol esters, which are synthesized from fatty acids and cholesterol by enzymes in the endoplasmic reticulum membrane. Because these lipids do not contain hydrophilic head groups, they are exclusively hydrophobic molecules, and therefore aggregate into three-dimensional droplets rather than into bilayers.

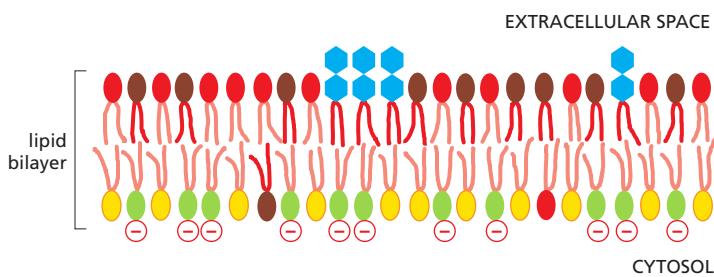
Lipid droplets are unique organelles in that they are surrounded by a single monolayer of phospholipids, which contains a large variety of proteins. Some of the proteins are enzymes involved in lipid metabolism, but the functions of most are unknown. Lipid droplets form rapidly when cells are exposed to high concentrations of fatty acids. They are thought to form from discrete regions of the endoplasmic reticulum membrane where many enzymes of lipid metabolism are concentrated. **Figure 10–14** shows one model of how lipid droplets may form and acquire their surrounding monolayer of phospholipids and proteins.

### The Asymmetry of the Lipid Bilayer Is Functionally Important

The lipid compositions of the two monolayers of the lipid bilayer in many membranes are strikingly different. In the human red blood cell (erythrocyte) membrane, for example, almost all of the phospholipid molecules that have choline— $(\text{CH}_3)_3\text{N}^+\text{CH}_2\text{CH}_2\text{OH}$ —in their head group (phosphatidylcholine and



**Figure 10–14 A model for the formation of lipid droplets.** Neutral lipids are deposited between the two monolayers of the endoplasmic reticulum membrane. There, they aggregate into a three-dimensional droplet, which buds and pinches off from the endoplasmic reticulum membrane as a unique organelle, surrounded by a single monolayer of phospholipids and associated proteins. (Adapted from S. Martin and R.G. Parton, *Nat. Rev. Mol. Cell Biol.* 7:373–378, 2006. With permission from Macmillan Publishers Ltd.)



**Figure 10–15** The asymmetrical distribution of phospholipids and glycolipids in the lipid bilayer of human red blood cells. The colors used for the phospholipid head groups are those introduced in Figure 10–3. In addition, glycolipids are drawn with hexagonal polar head groups (blue). Cholesterol (not shown) is distributed roughly equally in both monolayers.

sphingomyelin) are in the outer monolayer, whereas almost all that contain a terminal primary amino group (phosphatidylethanolamine and phosphatidylserine) are in the inner monolayer (Figure 10–15). Because the negatively charged phosphatidylserine is located in the inner monolayer, there is a significant difference in charge between the two halves of the bilayer. We discuss in Chapter 12 how membrane-bound phospholipid translocators generate and maintain lipid asymmetry.

Lipid asymmetry is functionally important, especially in converting extracellular signals into intracellular ones (discussed in Chapter 15). Many cytosolic proteins bind to specific lipid head groups found in the cytosolic monolayer of the lipid bilayer. The enzyme *protein kinase C* (PKC), for example, which is activated in response to various extracellular signals, binds to the cytosolic face of the plasma membrane, where phosphatidylserine is concentrated, and requires this negatively charged phospholipid for its activity.

In other cases, specific lipid head groups must first be modified to create protein-binding sites at a particular time and place. One example is *phosphatidylinositol* (PI), one of the minor phospholipids that are concentrated in the cytosolic monolayer of cell membranes (see Figure 13–10A–C). Various lipid kinases can add phosphate groups at distinct positions on the inositol ring, creating binding sites that recruit specific proteins from the cytosol to the membrane. An important example of such a lipid kinase is *phosphoinositide 3-kinase* (PI 3-kinase), which is activated in response to extracellular signals and helps to recruit specific intracellular signaling proteins to the cytosolic face of the plasma membrane (see Figure 15–53). Similar lipid kinases phosphorylate inositol phospholipids in intracellular membranes and thereby help to recruit proteins that guide membrane transport.

Phospholipids in the plasma membrane are used in yet another way to convert extracellular signals into intracellular ones. The plasma membrane contains various *phospholipases* that are activated by extracellular signals to cleave specific phospholipid molecules, generating fragments of these molecules that act as short-lived intracellular mediators. *Phospholipase C*, for example, cleaves an inositol phospholipid in the cytosolic monolayer of the plasma membrane to generate two fragments, one of which remains in the membrane and helps activate protein kinase C, while the other is released into the cytosol and stimulates the release of  $\text{Ca}^{2+}$  from the endoplasmic reticulum (see Figure 15–28).

Animals exploit the phospholipid asymmetry of their plasma membranes to distinguish between live and dead cells. When animal cells undergo apoptosis (a form of programmed cell death, discussed in Chapter 18), phosphatidylserine, which is normally confined to the cytosolic (or inner) monolayer of the plasma membrane lipid bilayer, rapidly translocates to the extracellular (or outer) monolayer. The phosphatidylserine exposed on the cell surface signals neighboring cells, such as macrophages, to phagocytose the dead cell and digest it. The translocation of the phosphatidylserine in apoptotic cells is thought to occur by two mechanisms:

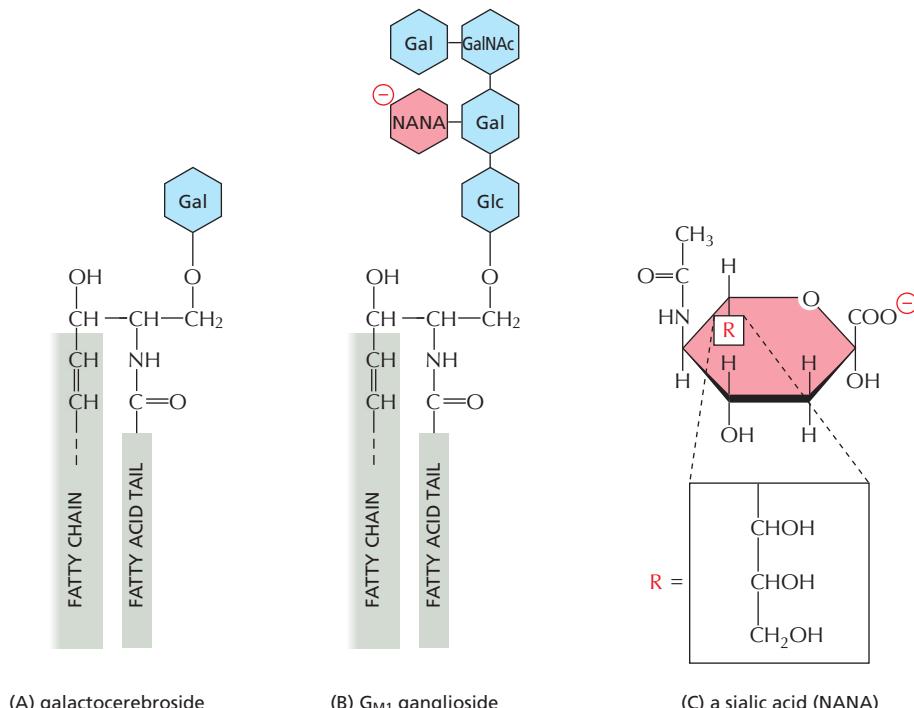
1. The phospholipid translocator that normally transports this lipid from the outer monolayer to the inner monolayer is inactivated.
2. A “scramblase” that transfers phospholipids nonspecifically in both directions between the two monolayers is activated.

## Glycolipids Are Found on the Surface of All Eukaryotic Plasma Membranes

Sugar-containing lipid molecules called **glycolipids** have the most extreme asymmetry in their membrane distribution: these molecules, whether in the plasma membrane or in intracellular membranes, are found exclusively in the monolayer facing away from the cytosol. In animal cells, they are made from sphingosine, just like sphingomyelin (see Figure 10–3). These intriguing molecules tend to self-assemble, partly through hydrogen bonds between their sugars and partly through van der Waals forces between their long and straight hydrocarbon chains, which causes them to partition preferentially into lipid raft phases (see Figure 10–13). The asymmetric distribution of glycolipids in the bilayer results from the addition of sugar groups to the lipid molecules in the lumen of the Golgi apparatus. Thus, the compartment in which they are manufactured is topologically equivalent to the exterior of the cell (discussed in Chapter 12). As they are delivered to the plasma membrane, the sugar groups are exposed at the cell surface (see Figure 10–15), where they have important roles in interactions of the cell with its surroundings.

Glycolipids probably occur in all eukaryotic cell plasma membranes, where they generally constitute about 5% of the lipid molecules in the outer monolayer. They are also found in some intracellular membranes. The most complex of the glycolipids, the **gangliosides**, contain oligosaccharides with one or more sialic acid moieties, which give gangliosides a net negative charge (Figure 10–16). The most abundant of the more than 40 different gangliosides that have been identified are in the plasma membrane of nerve cells, where gangliosides constitute 5–10% of the total lipid mass; they are also found in much smaller quantities in other cell types.

Hints as to the functions of glycolipids come from their localization. In the plasma membrane of epithelial cells, for example, glycolipids are confined to the exposed apical surface, where they may help to protect the membrane against the harsh conditions frequently found there (such as low pH and high concentrations of degradative enzymes). Charged glycolipids, such as gangliosides, may be important because of their electrical effects: their presence alters the electrical field across the membrane and the concentrations of ions—especially  $\text{Ca}^{2+}$ —at the membrane surface. Glycolipids also function in cell-recognition processes,



**Figure 10–16 Glycolipid molecules.**  
 (A) Galactocerebroside is called a *neutral glycolipid* because the sugar that forms its head group is uncharged. (B) A ganglioside always contains one or more negatively charged sialic acid moiety. There are various types of sialic acid; in human cells, it is mostly *N*-acetylneurameric acid, or NANA, whose structure is shown in (C). Whereas in bacteria and plants almost all glycolipids are derived from glycerol, as are most phospholipids, in animal cells almost all glycolipids are based on sphingosine, as is the case for sphingomyelin (see Figure 10–3). Gal = galactose; Glc = glucose, GalNAc = *N*-acetylgalactosamine; these three sugars are uncharged.

in which membrane-bound carbohydrate-binding proteins (*lectins*) bind to the sugar groups on both glycolipids and glycoproteins in the process of cell-cell adhesion (discussed in Chapter 19). Mutant mice that are deficient in all of their complex gangliosides show abnormalities in the nervous system, including axonal degeneration and reduced myelination.

Some glycolipids provide entry points for certain bacterial toxins and viruses. The ganglioside G<sub>M1</sub> (see Figure 10–16), for example, acts as a cell-surface receptor for the bacterial toxin that causes the debilitating diarrhea of cholera. Cholera toxin binds to and enters only those cells that have G<sub>M1</sub> on their surface, including intestinal epithelial cells. Its entry into a cell leads to a prolonged increase in the concentration of intracellular cyclic AMP (discussed in Chapter 15), which in turn causes a large efflux of Cl<sup>−</sup>, leading to the secretion of Na<sup>+</sup>, K<sup>+</sup>, HCO<sub>3</sub><sup>−</sup>, and water into the intestine. Polyomaviruses also enter the cell after binding initially to gangliosides.

### Summary

*Biological membranes consist of a continuous double layer of lipid molecules in which membrane proteins are embedded. This lipid bilayer is fluid, with individual lipid molecules able to diffuse rapidly within their own monolayer. The membrane lipid molecules are amphiphilic. When placed in water, they assemble spontaneously into bilayers, which form sealed compartments.*

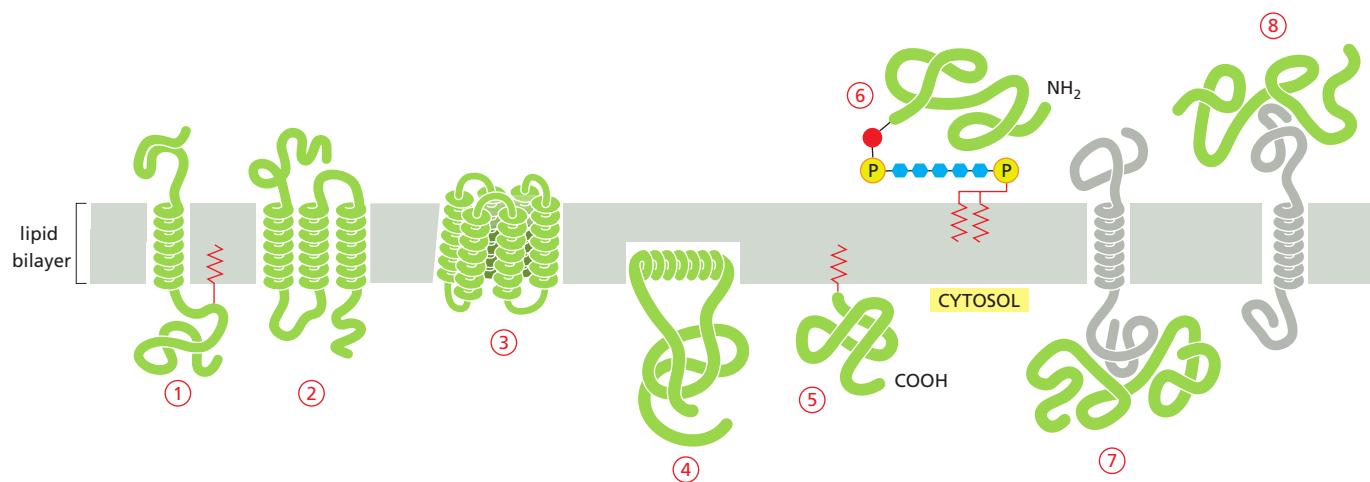
*Although cell membranes can contain hundreds of different lipid species, the plasma membrane in animal cells contains three major classes—phospholipids, cholesterol, and glycolipids. Because of their different backbone structure, phospholipids fall into two subclasses—phosphoglycerides and sphingolipids. The lipid compositions of the inner and outer monolayers are different, reflecting the different functions of the two faces of a cell membrane. Different mixtures of lipids are found in the membranes of cells of different types, as well as in the various membranes of a single eukaryotic cell. Inositol phospholipids are a minor class of phospholipids, which in the cytosolic leaflet of the plasma membrane lipid bilayer play an important part in cell signaling: in response to extracellular signals, specific lipid kinases phosphorylate the head groups of these lipids to form docking sites for cytosolic signaling proteins, whereas specific phospholipases cleave certain inositol phospholipids to generate small intracellular signaling molecules.*

## MEMBRANE PROTEINS

Although the lipid bilayer provides the basic structure of biological membranes, the membrane proteins perform most of the membrane's specific tasks and therefore give each type of cell membrane its characteristic functional properties. Accordingly, the amounts and types of proteins in a membrane are highly variable. In the myelin membrane, which serves mainly as electrical insulation for nerve-cell axons, less than 25% of the membrane mass is protein. By contrast, in the membranes involved in ATP production (such as the internal membranes of mitochondria and chloroplasts), approximately 75% is protein. A typical plasma membrane is somewhere in between, with protein accounting for about half of its mass. Because lipid molecules are small compared with protein molecules, however, there are always many more lipid molecules than protein molecules in cell membranes—about 50 lipid molecules for each protein molecule in cell membranes that are 50% protein by mass. Membrane proteins vary widely in structure and in the way they associate with the lipid bilayer, which reflects their diverse functions.

### Membrane Proteins Can Be Associated with the Lipid Bilayer in Various Ways

**Figure 10–17** shows the different ways in which proteins can associate with the membrane. Like their lipid neighbors, **membrane proteins** are amphiphilic,



having hydrophobic and hydrophilic regions. Many membrane proteins extend through the lipid bilayer, and hence are called **transmembrane proteins**, with part of their mass on either side (Figure 10–17, examples 1, 2, and 3). Their hydrophobic regions pass through the membrane and interact with the hydrophobic tails of the lipid molecules in the interior of the bilayer, where they are sequestered away from water. Their hydrophilic regions are exposed to water on either side of the membrane. The covalent attachment of a fatty acid chain that inserts into the cytosolic monolayer of the lipid bilayer increases the hydrophobicity of some of these transmembrane proteins (see Figure 10–17, example 1).

Other membrane proteins are located entirely in the cytosol and are attached to the cytosolic monolayer of the lipid bilayer, either by an amphiphilic  $\alpha$  helix exposed on the surface of the protein (Figure 10–17, example 4) or by one or more covalently attached lipid chains (Figure 10–17, example 5). Yet other membrane proteins are entirely exposed at the external cell surface, being attached to the lipid bilayer only by a covalent linkage (via a specific oligosaccharide) to a lipid anchor in the outer monolayer of the plasma membrane (Figure 10–17, example 6).

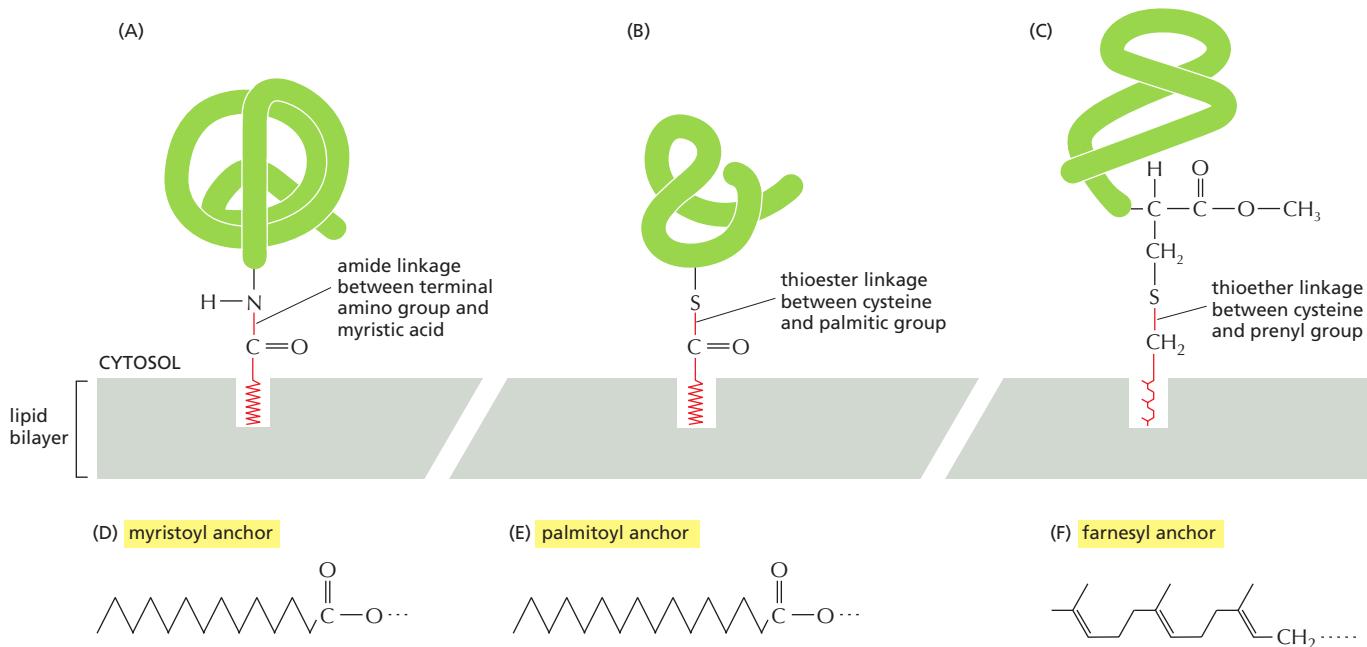
The lipid-linked proteins in example 5 in Figure 10–17 are made as soluble proteins in the cytosol and are subsequently anchored to the membrane by the covalent attachment of the lipid group. The proteins in example 6, however, are made as single-pass membrane proteins in the endoplasmic reticulum (ER). While still in the ER, the transmembrane segment of the protein is cleaved off and a **glycosylphosphatidylinositol (GPI) anchor** is added, leaving the protein bound to the noncytosolic surface of the ER membrane solely by this anchor (discussed in Chapter 12); transport vesicles eventually deliver the protein to the plasma membrane (discussed in Chapter 13).

By contrast to these examples, **membrane-associated proteins** do not extend into the hydrophobic interior of the lipid bilayer at all; they are instead bound to either face of the membrane by noncovalent interactions with other membrane proteins (Figure 10–17, examples 7 and 8). Many of the proteins of this type can be released from the membrane by relatively gentle extraction procedures, such as exposure to solutions of very high or low ionic strength or of extreme pH, which interfere with protein–protein interactions but leave the lipid bilayer intact; these proteins are often referred to as *peripheral membrane proteins*. Transmembrane proteins and many proteins held in the bilayer by lipid groups or hydrophobic polypeptide regions that insert into the hydrophobic core of the lipid bilayer cannot be released in these ways.

### Lipid Anchors Control the Membrane Localization of Some Signaling Proteins

How a membrane protein is associated with the lipid bilayer reflects the function of the protein. Only transmembrane proteins can function on both sides of

**Figure 10–17** Various ways in which proteins associate with the lipid bilayer. Most membrane proteins are thought to extend across the bilayer as (1) a single  $\alpha$  helix, (2) as multiple  $\alpha$  helices, or (3) as a rolled-up  $\beta$  sheet ( $\beta$  barrel). Some of these “single-pass” and “multipass” proteins have a covalently attached fatty acid chain inserted in the cytosolic lipid monolayer (1). Other membrane proteins are exposed at only one side of the membrane. (4) Some of these are anchored to the cytosolic surface by an amphiphilic  $\alpha$  helix that partitions into the cytosolic monolayer of the lipid bilayer through the hydrophobic face of the helix. (5) Others are attached to the bilayer solely by a covalently bound lipid chain—either a fatty acid chain or a prenyl group (see Figure 10–18)—in the cytosolic monolayer or, (6) via an oligosaccharide linker, to phosphatidylinositol in the noncytosolic monolayer—called a GPI anchor. (7, 8) Finally, membrane-associated proteins are attached to the membrane only by noncovalent interactions with other membrane proteins. The way in which the structure in (5) is formed is illustrated in Figure 10–18, while the way in which the GPI anchor shown in (6) is formed is illustrated in Figure 12–52. The details of how membrane proteins become associated with the lipid bilayer are discussed in Chapter 12.



the bilayer or transport molecules across it. Cell-surface receptors, for example, are usually transmembrane proteins that bind signal molecules in the extracellular space and generate different intracellular signals on the opposite side of the plasma membrane. To transfer small hydrophilic molecules across a membrane, a membrane transport protein must provide a path for the molecules to cross the hydrophobic permeability barrier of the lipid bilayer; the molecular architecture of multipass transmembrane proteins (Figure 10–17, examples 2 and 3) is ideally suited for this task, as we discuss in Chapter 11.

Proteins that function on only one side of the lipid bilayer, by contrast, are often associated exclusively with either the lipid monolayer or a protein domain on that side. Some intracellular signaling proteins, for example, that help relay extracellular signals into the cell interior are bound to the cytosolic half of the plasma membrane by one or more covalently attached lipid groups, which can be fatty acid chains or *prenyl groups* (Figure 10–18). In some cases, myristic acid, a saturated 14-carbon fatty acid, is added to the N-terminal amino group of the protein during its synthesis on a ribosome. All members of the *Src family* of cytoplasmic protein tyrosine kinases (discussed in Chapter 15) are myristoylated in this way. Membrane attachment through a single lipid anchor is not very strong, however, and a second lipid group is often added to anchor proteins more firmly to a membrane. For most *Src* kinases, the second lipid modification is the attachment of palmitic acid, a saturated 16-carbon fatty acid, to a cysteine side chain of the protein. This modification occurs in response to an extracellular signal and helps recruit the kinases to the plasma membrane. When the signaling pathway is turned off, the palmitic acid is removed, allowing the kinase to return to the cytosol. Other intracellular signaling proteins, such as the *Ras* family small GTPases (discussed in Chapter 15), use a combination of prenyl group and palmitic acid attachment to recruit the proteins to the plasma membrane.

Many proteins attach to membranes transiently. Some are classical peripheral membrane proteins that associate with membranes by regulated protein-protein interactions. Others undergo a transition from soluble to membrane protein by a conformational change that exposes a hydrophobic peptide or covalently attached lipid anchor. Many of the small GTPases of the *Rab* protein family that regulate intracellular membrane traffic (discussed in Chapter 13), for example, switch depending on the nucleotide that is bound to the protein. In their GDP-bound state they are soluble and free in the cytosol, whereas in their GTP-bound state their lipid anchor is exposed and tethers them to membranes. They are

**Figure 10–18** Membrane protein attachment by a fatty acid chain or a prenyl group. The covalent attachment of either type of lipid can help localize a water-soluble protein to a membrane after its synthesis in the cytosol. (A) A fatty acid chain (myristic acid) is attached via an amide linkage to an N-terminal glycine. (B) A fatty acid chain (palmitic acid) is attached via a thioester linkage to a cysteine. (C) A prenyl chain (either farnesyl or a longer geranylgeranyl chain) is attached via a thioether linkage to a cysteine residue that is initially located four residues from the protein's C-terminus. After prenylation, the terminal three amino acids are cleaved off, and the new C-terminus is methylated before insertion of the anchor into the membrane (not shown). The structures of the lipid anchors are shown below: (D) a myristoyl anchor (derived from a 14-carbon saturated fatty acid chain), (E) a palmitoyl anchor (a 16-carbon saturated fatty acid chain), and (F) a farnesyl anchor (a 15-carbon unsaturated hydrocarbon chain).

**Figure 10–19** A segment of a membrane-spanning polypeptide chain crossing the lipid bilayer as an  $\alpha$  helix. Only the  $\alpha$ -carbon backbone of the polypeptide chain is shown, with the hydrophobic amino acids in green and yellow. The polypeptide segment shown is part of the bacterial photosynthetic reaction center, the structure of which was determined by x-ray diffraction. (Based on data from J. Deisenhofer et al., *Nature* 318:618–624, 1985, and H. Michel et al., *EMBO J.* 5:1149–1158, 1986.)

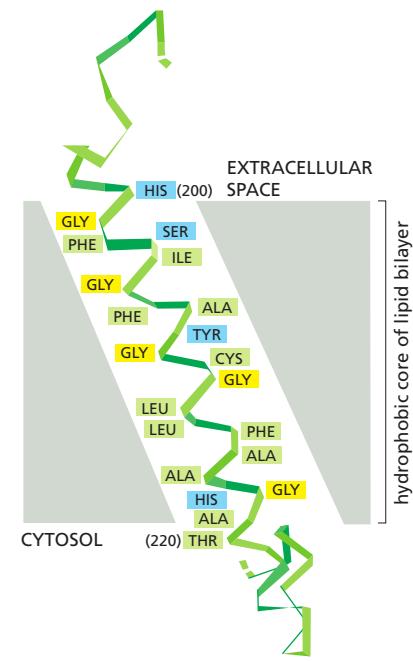
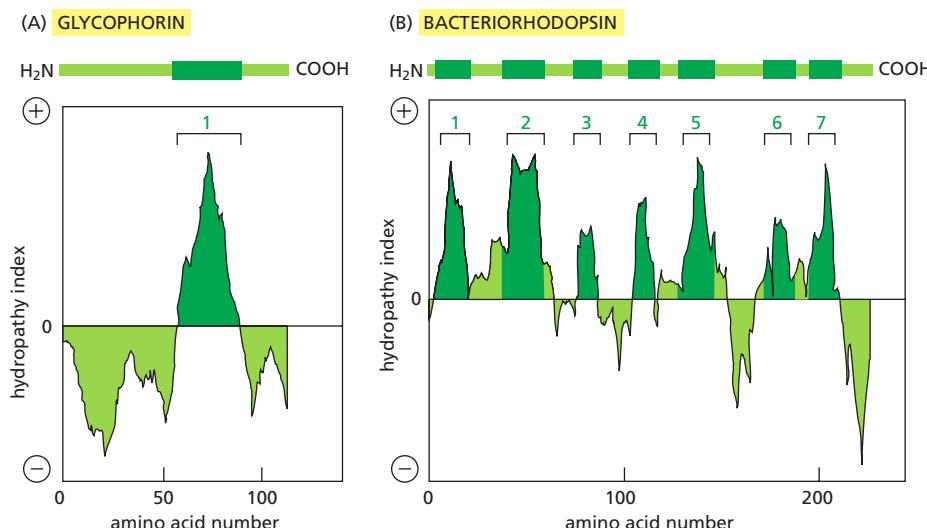
membrane proteins at one moment and soluble proteins at the next. Such highly dynamic interactions greatly expand the repertoire of membrane functions.

### In Most Transmembrane Proteins, the Polypeptide Chain Crosses the Lipid Bilayer in an $\alpha$ -Helical Conformation

A transmembrane protein always has a unique orientation in the membrane. This reflects both the asymmetric manner in which it is inserted into the lipid bilayer in the ER during its biosynthesis (discussed in Chapter 12) and the different functions of its cytosolic and noncytosolic domains. These domains are separated by the membrane-spanning segments of the polypeptide chain, which contact the hydrophobic environment of the lipid bilayer and are composed largely of amino acids with nonpolar side chains. Because the peptide bonds themselves are polar and because water is absent, all peptide bonds in the bilayer are driven to form hydrogen bonds with one another. The hydrogen-bonding between peptide bonds is maximized if the polypeptide chain forms a regular  $\alpha$  helix as it crosses the bilayer, and this is how most membrane-spanning segments of polypeptide chains traverse the bilayer (Figure 10–19).

In **single-pass transmembrane proteins**, the polypeptide chain crosses only once (see Figure 10–17, example 1), whereas in **multipass transmembrane proteins**, the polypeptide chain crosses multiple times (see Figure 10–17, example 2). An alternative way for the peptide bonds in the lipid bilayer to satisfy their hydrogen-bonding requirements is for multiple transmembrane strands of a polypeptide chain to be arranged as a  $\beta$  sheet that is rolled up into a cylinder (a so-called  $\beta$  barrel; see Figure 10–17, example 3). This protein architecture is seen in the *porin proteins* that we discuss later.

Progress in the x-ray crystallography of membrane proteins has enabled the determination of the three-dimensional structure of many of them. The structures confirm that it is often possible to predict from the protein's amino acid sequence which parts of the polypeptide chain extend across the lipid bilayer. Segments containing about 20–30 amino acids, with a high degree of hydrophobicity, are long enough to span a lipid bilayer as an  $\alpha$  helix, and they can often be identified in *hydropathy plots* (Figure 10–20). From such plots, it is estimated that about 30%



**Figure 10–20** Using hydropathy plots to localize potential  $\alpha$ -helical membrane-spanning segments in a polypeptide chain. The free energy needed to transfer successive segments of a polypeptide chain from a nonpolar solvent to water is calculated from the amino acid composition of each segment using data obtained from model compounds. This calculation is made for segments of a fixed size (usually around 10–20 amino acids), beginning with each successive amino acid in the chain. The “hydropathy index” of the segment is plotted on the Y axis as a function of its location in the chain. A positive value indicates that free energy is required for transfer to water (i.e., the segment is hydrophobic), and the value assigned is an index of the amount of energy needed. Peaks in the hydropathy index appear at the positions of hydrophobic segments in the amino acid sequence. (A and B) Hydropathy plots for two membrane proteins that are discussed later in this chapter. Glycophorin (A) has a single membrane-spanning  $\alpha$  helix, and one corresponding peak in the hydropathy plot. Bacteriorhodopsin (B) has seven membrane-spanning  $\alpha$  helices and seven corresponding peaks in the hydropathy plot. (A, adapted from D. Eisenberg, *Annu. Rev. Biochem.* 53:595–624, 1984. With permission from Annual Reviews.)

of an organism's proteins are transmembrane proteins, emphasizing their importance. Hydropathy plots cannot identify the membrane-spanning segments of a  $\beta$  barrel, as 10 amino acids or fewer are sufficient to traverse a lipid bilayer as an extended  $\beta$  strand and only every other amino acid side chain is hydrophobic.

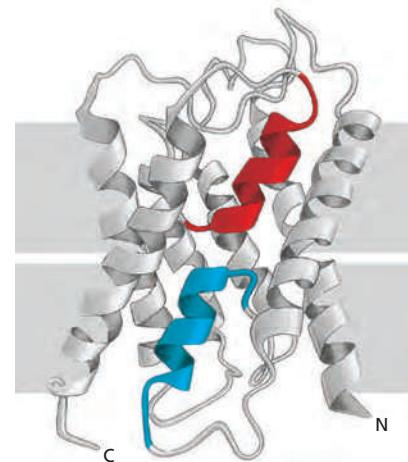
The strong drive to maximize hydrogen-bonding in the absence of water means that a polypeptide chain that enters the lipid bilayer is likely to pass entirely through it before changing direction, since chain bending requires a loss of regular hydrogen-bonding interactions. But multipass transmembrane proteins can also contain regions that fold into the membrane from either side, squeezing into spaces between transmembrane  $\alpha$  helices without contacting the hydrophobic core of the lipid bilayer. Because such regions interact only with other polypeptide regions, they do not need to maximize hydrogen-bonding; they can therefore have a variety of secondary structures, including helices that extend only part way across the lipid bilayer (Figure 10-21). Such regions are important for the function of some membrane proteins, including water channel and ion channel proteins, in which the regions contribute to the walls of the pores traversing the membrane and confer substrate specificity on the channels, as we discuss in Chapter 11. These regions cannot be identified in hydropathy plots and are only revealed by x-ray crystallography or electron crystallography (a technique similar to x-ray diffraction but performed on two-dimensional arrays of proteins) of the protein's three-dimensional structure.

### Transmembrane $\alpha$ Helices Often Interact with One Another

The transmembrane  $\alpha$  helices of many single-pass membrane proteins do not contribute to the folding of the protein domains on either side of the membrane. As a consequence, it is often possible to engineer cells to produce just the cytosolic or extracellular domains of these proteins as water-soluble molecules. This approach has been invaluable for studying the structure and function of these domains, especially the domains of transmembrane receptor proteins (discussed in Chapter 15). A transmembrane  $\alpha$  helix, even in a single-pass membrane protein, however, often does more than just anchor the protein to the lipid bilayer. Many single-pass membrane proteins form homo- or heterodimers that are held together by noncovalent, but strong and highly specific, interactions between the two transmembrane  $\alpha$  helices; the sequence of the hydrophobic amino acids of these helices contains the information that directs the protein–protein interaction.

Similarly, the transmembrane  $\alpha$  helices in multipass membrane proteins occupy specific positions in the folded protein structure that are determined by interactions between the neighboring helices. These interactions are crucial for the structure and function of the many channels and transporters that move molecules across cell membranes.

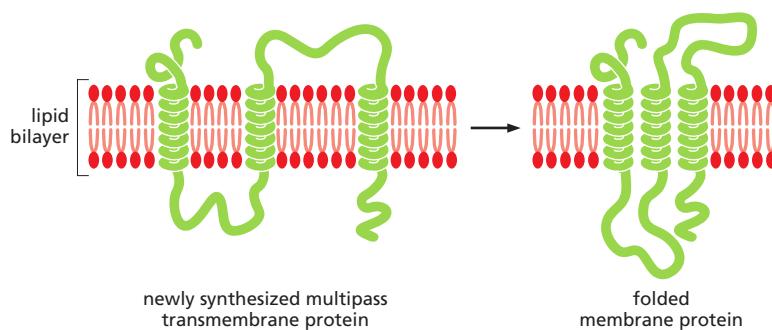
In these proteins, neighboring transmembrane helices in the folded structure of the protein shield many of the other transmembrane helices from the membrane lipids. Why, then, are these shielded helices nevertheless composed primarily of hydrophobic amino acids? The answer lies in the way in which multipass proteins are integrated into the membrane during their biosynthesis. As we discuss in Chapter 12, transmembrane  $\alpha$  helices are inserted into the lipid bilayer sequentially by a protein translocator. After leaving the translocator, each helix is transiently surrounded by lipids, which requires that the helix be hydrophobic. It is only as the protein folds up into its final structure that contacts are made between adjacent helices, and protein–protein contacts replace some of the protein–lipid contacts (Figure 10-22).



**Figure 10-21** Two short  $\alpha$  helices in the aquaporin water channel, each of which spans only halfway through the lipid bilayer. In the plasma membrane, four monomers, one of which is shown here, form a tetramer. Each monomer has a hydrophilic pore at its center, which allows water molecules to cross the membrane in single file (see Figure 11-20 and Movie 11.6). The two short colored helices are buried at an interface formed by protein–protein interactions. The mechanism by which the channel allows the passage of water molecules is discussed in more detail in Chapter 11.

### Some $\beta$ Barrels Form Large Channels

Multipass membrane proteins that have their transmembrane segments arranged as  $\beta$  barrels rather than as  $\alpha$  helices are comparatively rigid and therefore tend to form crystals readily when isolated. Thus, some of them were among the first



**Figure 10–22** Steps in the folding of a multipass transmembrane protein.

When a newly synthesized transmembrane  $\alpha$  helix is released into the lipid bilayer, it is initially surrounded by lipid molecules. As the protein folds, contacts between the helices displace some of the lipid molecules surrounding the helices.

multipass membrane protein structures to be determined by x-ray crystallography. The number of  $\beta$  strands in a  $\beta$  barrel varies widely, from as few as 8 strands to as many as 22 (Figure 10–23).

$\beta$ -barrel proteins are abundant in the outer membranes of bacteria, mitochondria, and chloroplasts. Some are pore-forming proteins, which create water-filled channels that allow selected small hydrophilic molecules to cross the membrane. The porins are well-studied examples (example 3 in Figure 10–23C). Many porin barrels are formed from a 16-strand, antiparallel  $\beta$  sheet rolled up into a cylindrical structure. Polar amino acid side chains line the aqueous channel on the inside, while nonpolar side chains project from the outside of the barrel to interact with the hydrophobic core of the lipid bilayer. Loops of the polypeptide chain often protrude into the lumen of the channel, narrowing it so that only certain solutes can pass. Some porins are therefore highly selective: *maltoxin*, for example, preferentially allows maltose and maltose oligomers to cross the outer membrane of *E. coli*.

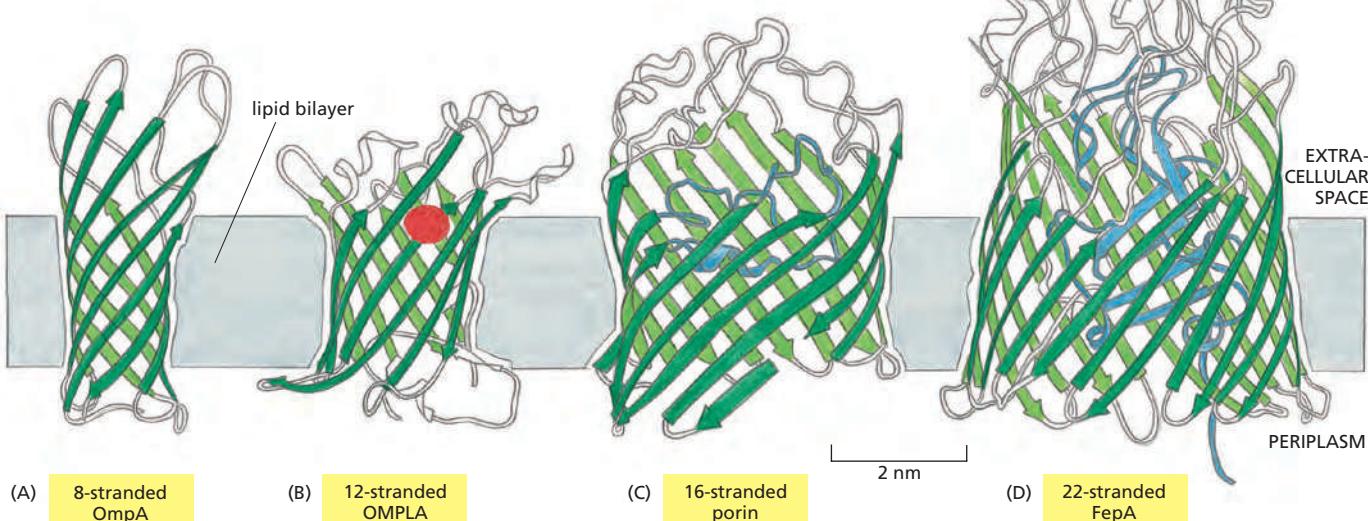
The *FepA* protein is a more complex example of a  $\beta$  barrel transport protein (Figure 10–23D). It transports iron ions across the bacterial outer membrane. It is constructed from 22  $\beta$  strands, and a large globular domain completely fills the inside of the barrel. Iron ions bind to this domain, which by an unknown mechanism moves or changes its conformation to transfer the iron across the membrane.

Not all  $\beta$ -barrel proteins are transport proteins. Some form smaller barrels that are completely filled by amino acid side chains that project into the center of the barrel. These proteins function as receptors or enzymes (Figure 10–23A and B); the barrel serves as a rigid anchor, which holds the protein in the membrane and orients the cytosolic loops that form binding sites for specific intracellular molecules.

Most multipass membrane proteins in eukaryotic cells and in the bacterial plasma membrane are constructed from transmembrane  $\alpha$  helices. The helices

**Figure 10–23**  $\beta$  barrels formed from different numbers of  $\beta$  strands.

(A) The *E. coli* OmpA protein serves as a receptor for a bacterial virus. (B) The *E. coli* OMPLA protein is an enzyme (a lipase) that hydrolyzes lipid molecules. The amino acids that catalyze the enzymatic reaction (shown in red) protrude from the outside surface of the barrel. (C) A porin from the bacterium *Rhodobacter capsulatus* forms a water-filled pore across the outer membrane. The diameter of the channel is restricted by loops (shown in blue) that protrude into the channel. (D) The *E. coli* FepA protein transports iron ions. The inside of the barrel is completely filled by a globular protein domain (shown in blue) that contains an iron-binding site (not shown).



**Figure 10–24 A single-pass transmembrane protein.** Note that the polypeptide chain traverses the lipid bilayer as a right-handed  $\alpha$  helix and that the oligosaccharide chains and disulfide bonds are all on the noncytosolic surface of the membrane. The sulfhydryl groups in the cytosolic domain of the protein do not normally form disulfide bonds because the reducing environment in the cytosol maintains these groups in their reduced ( $-SH$ ) form.

can slide against each other, allowing conformational changes in the protein that can open and shut ion channels, transport solutes, or transduce extracellular signals into intracellular ones. In  $\beta$ -barrel proteins, by contrast, hydrogen bonds bind each  $\beta$  strand rigidly to its neighbors, making conformational changes within the wall of the barrel unlikely.

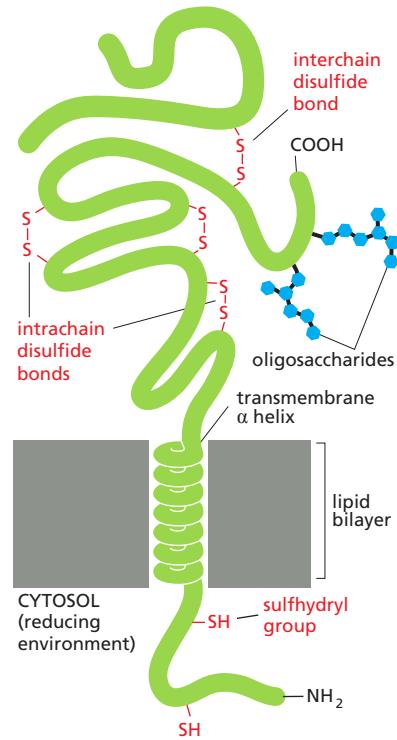
### Many Membrane Proteins Are Glycosylated

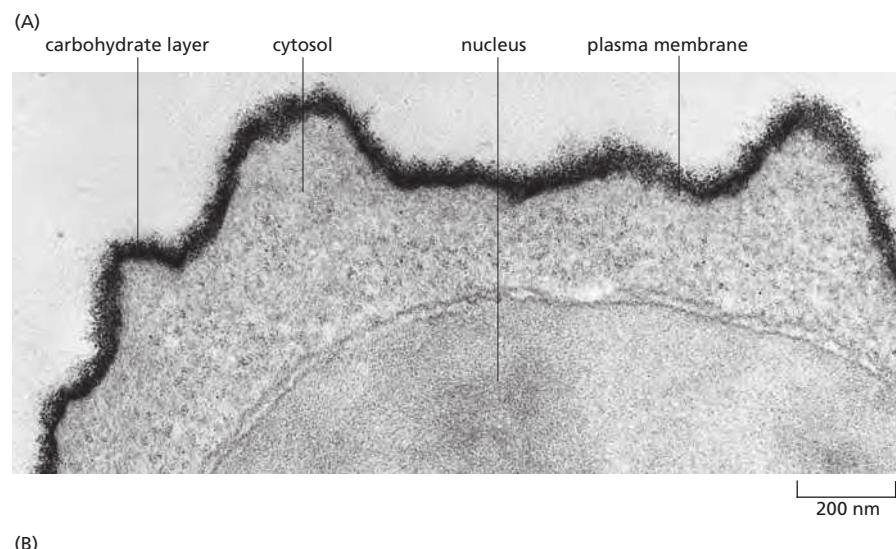
Most transmembrane proteins in animal cells are glycosylated. As in glycolipids, the sugar residues are added in the lumen of the ER and the Golgi apparatus (discussed in Chapters 12 and 13). For this reason, the oligosaccharide chains are always present on the noncytosolic side of the membrane. Another important difference between proteins (or parts of proteins) on the two sides of the membrane results from the reducing environment of the cytosol. This environment decreases the likelihood that intrachain or interchain disulfide (S-S) bonds will form between cysteines on the cytosolic side of membranes. These bonds form on the noncytosolic side, where they can help stabilize either the folded structure of the polypeptide chain or its association with other polypeptide chains (Figure 10–24).

Because the extracellular part of most plasma membrane proteins are glycosylated, carbohydrates extensively coat the surface of all eukaryotic cells. These carbohydrates occur as oligosaccharide chains covalently bound to membrane proteins (glycoproteins) and lipids (glycolipids). They also occur as the polysaccharide chains of integral membrane *proteoglycan* molecules. Proteoglycans, which consist of long polysaccharide chains linked covalently to a protein core, are found mainly outside the cell, as part of the extracellular matrix (discussed in Chapter 19). But, for some proteoglycans, the protein core either extends across the lipid bilayer or is attached to the bilayer by a glycosylphosphatidylinositol (GPI) anchor.

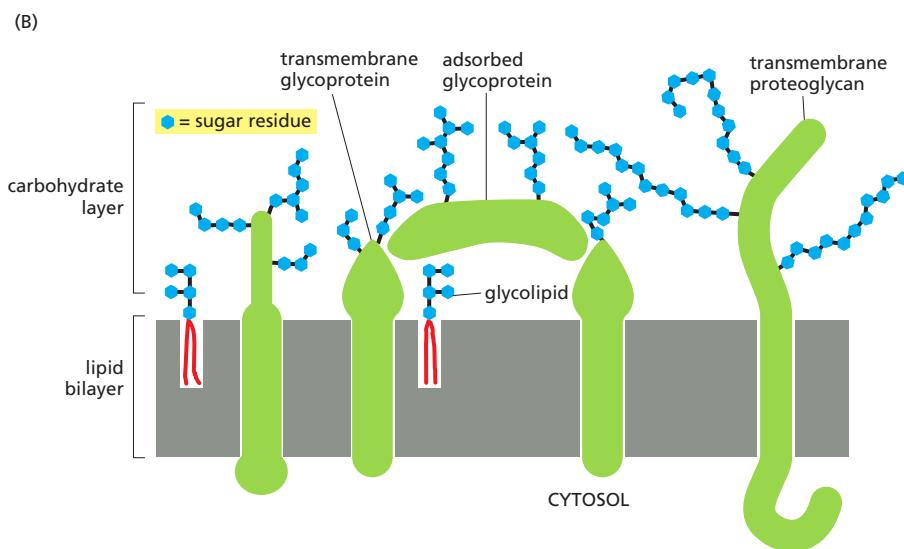
The terms cell coat or glycocalyx are sometimes used to describe the carbohydrate-rich zone on the cell surface. This **carbohydrate layer** can be visualized by various stains, such as ruthenium red (Figure 10–25A), as well as by its affinity for carbohydrate-binding proteins called **lectins**, which can be labeled with a fluorescent dye or some other visible marker. Although most of the sugar groups are attached to intrinsic plasma membrane molecules, the carbohydrate layer also contains both glycoproteins and proteoglycans that have been secreted into the extracellular space and then adsorbed onto the cell surface (Figure 10–25B). Many of these adsorbed macromolecules are components of the extracellular matrix, so that the boundary between the plasma membrane and the extracellular matrix is often not sharply defined. One of the many functions of the carbohydrate layer is to protect cells against mechanical and chemical damage; it also keeps various other cells at a distance, preventing unwanted cell-cell interactions.

The oligosaccharide side chains of glycoproteins and glycolipids are enormously diverse in their arrangement of sugars. Although they usually contain fewer than 15 sugars, the chains are often branched, and the sugars can be bonded together by various kinds of covalent linkages—unlike the amino acids in a polypeptide chain, which are all linked by identical peptide bonds. Even three sugars can be put together to form hundreds of different trisaccharides. Both the diversity and the exposed position of the oligosaccharides on the cell surface make them especially well suited to function in specific cell-recognition processes. As we discuss in Chapter 19, plasma-membrane-bound lectins that recognize specific oligosaccharides on cell-surface glycolipids and glycoproteins mediate a variety of





**Figure 10-25 The carbohydrate layer on the cell surface.** (A) This electron micrograph of the surface of a lymphocyte stained with ruthenium red emphasizes the thick carbohydrate-rich layer surrounding the cell. (B) The carbohydrate layer is made up of the oligosaccharide side chains of membrane glycolipids and membrane glycoproteins and the polysaccharide chains on membrane proteoglycans. In addition, adsorbed glycoproteins, and adsorbed proteoglycans (not shown), contribute to the carbohydrate layer in many cells. Note that all of the carbohydrate is on the extracellular surface of the membrane. (A, courtesy of Audrey M. Glauert and G.M.W. Cook.)



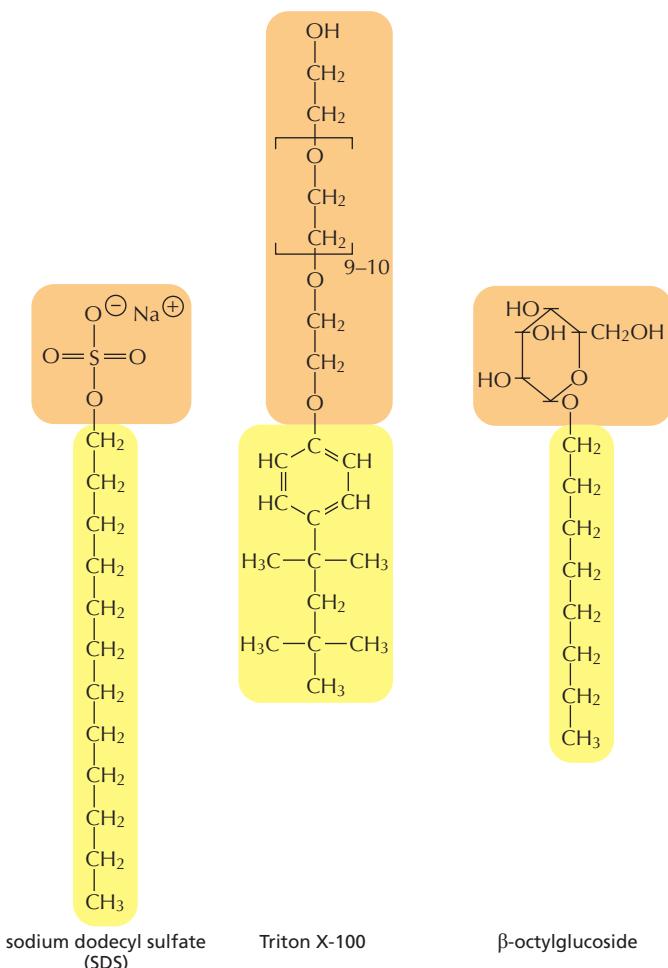
transient cell-cell adhesion processes, including those occurring in lymphocyte recirculation and inflammatory responses (see Figure 19-28).

### Membrane Proteins Can Be Solubilized and Purified in Detergents

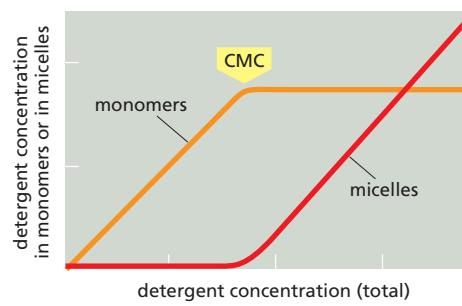
In general, only agents that disrupt hydrophobic associations and destroy the lipid bilayer can solubilize membrane proteins. The most useful of these for the membrane biochemist are **detergents**, which are small amphiphilic molecules of variable structure ([Movie 10.4](#)). Detergents are much more soluble in water than lipids. Their polar (hydrophilic) ends can be either charged (ionic), as in *sodium dodecyl sulfate (SDS)*, or uncharged (nonionic), as in *octylglucoside* and Triton ([Figure 10-26A](#)). At low concentration, detergents are monomeric in solution, but when their concentration is increased above a threshold, called the *critical micelle concentration (CMC)*, they aggregate to form micelles ([Figure 10-26B-D](#)). Above the CMC, detergent molecules rapidly diffuse in and out of micelles, keeping the concentration of monomer in the solution constant, no matter how many micelles are present. Both the CMC and the average number of detergent molecules in a micelle are characteristic properties of each detergent, but they also depend on the temperature, pH, and salt concentration. Detergent solutions are therefore complex systems and are difficult to study.

When mixed with membranes, the hydrophobic ends of detergents bind to the hydrophobic regions of the membrane proteins, where they displace lipid molecules with a collar of detergent molecules. Since the other end of the detergent

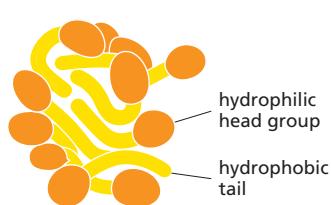
(A)



(B)



(C)



(D)

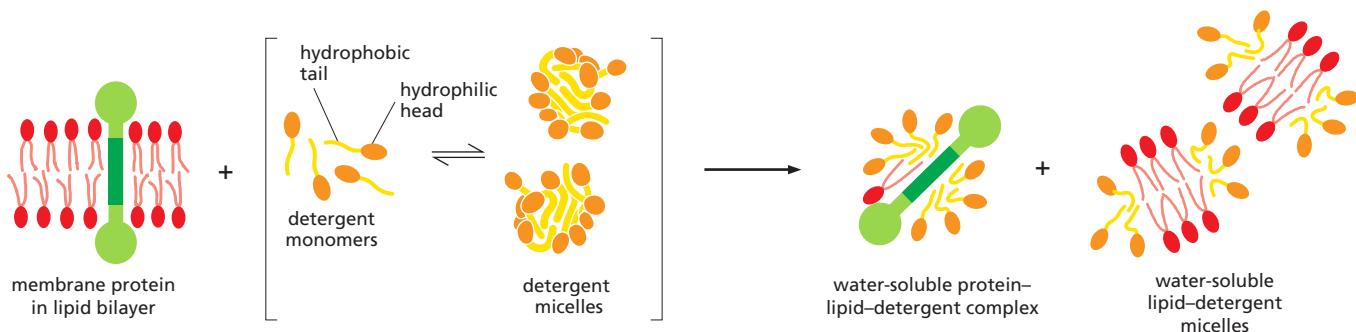


**Figure 10–26 The structure and function of detergents.** (A) Three commonly used detergents are sodium dodecyl sulfate (SDS), an anionic detergent, and Triton X-100 and  $\beta$ -octylglucoside, two nonionic detergents. Triton X-100 is a mixture of compounds in which the region in brackets is repeated between 9 and 10 times. The hydrophobic portion of each detergent is shown in yellow, and the hydrophilic portion is shown in orange. (B) At low concentration, detergent molecules are monomeric in solution. As their concentration is increased beyond the critical micelle concentration (CMC), some of the detergent molecules form micelles. Note that the concentration of detergent monomer stays constant above the CMC. (C) Because they have both polar and nonpolar ends, detergent molecules are amphiphilic; and because they are cone-shaped, they form micelles rather than bilayers (see Figure 10–7). Detergent micelles are thought to have irregular shapes, and, due to packing constraints, the hydrophobic tails are partially exposed to water. (D) The space-filling model shows the structure of a micelle composed of 20  $\beta$ -octylglucoside molecules, predicted by molecular dynamics calculations. The head groups are shown in red and the hydrophobic tails in gray. (B, adapted from G. Gunnarsson, B. Jönsson and H. Wennerström, *J. Phys. Chem.* 84:3114–3121, 1980; C, from S. Bogusz, R.M. Venable and R.W. Pastor, *J. Phys. Chem. B* 104:5462–5470, 2000.)

molecule is polar, this binding tends to bring the membrane proteins into solution as detergent-protein complexes (Figure 10–27). Usually, some lipid molecules also remain attached to the protein.

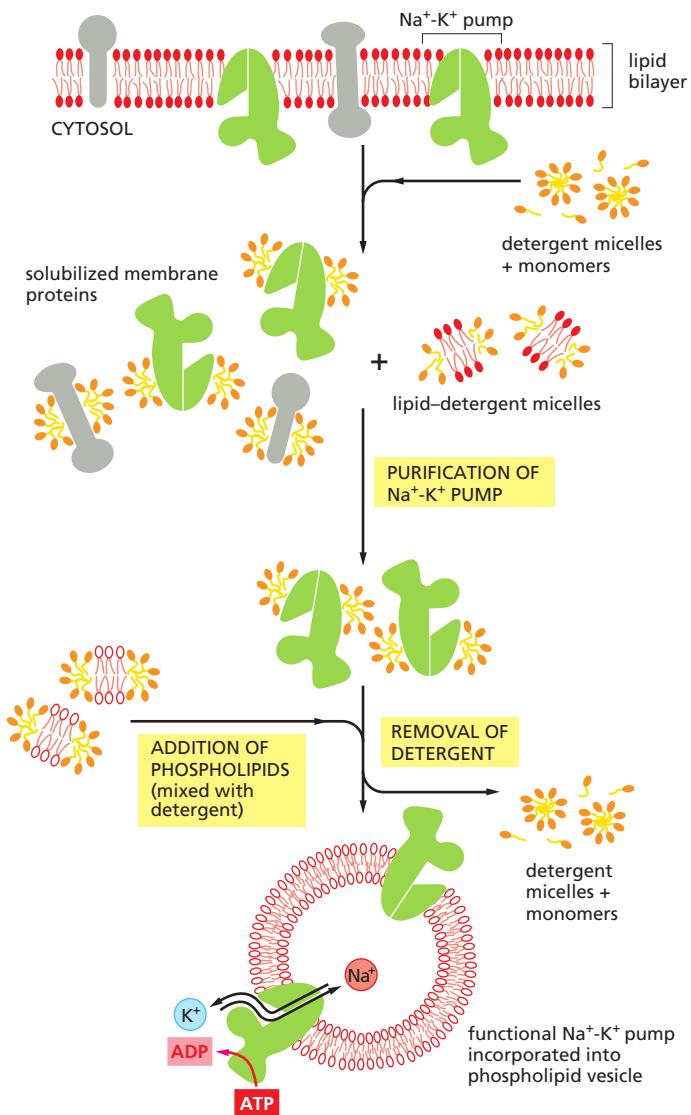
Strong ionic detergents, such as SDS, can solubilize even the most hydrophobic membrane proteins. This allows the proteins to be analyzed by *SDS polyacrylamide-gel electrophoresis* (discussed in Chapter 8), a procedure that has revolutionized the study of proteins. Such strong detergents, however, unfold (denature) proteins by binding to their internal “hydrophobic cores,” thereby rendering the proteins inactive and unusable for functional studies. Nonetheless, proteins can be readily separated and purified in their SDS-denatured form. In some cases, removal of the SDS allows the purified protein to renature, with recovery of functional activity.

Many membrane proteins can be solubilized and then purified in an active form by the use of mild detergents. These detergents cover the hydrophobic regions on membrane-spanning segments that become exposed after lipid

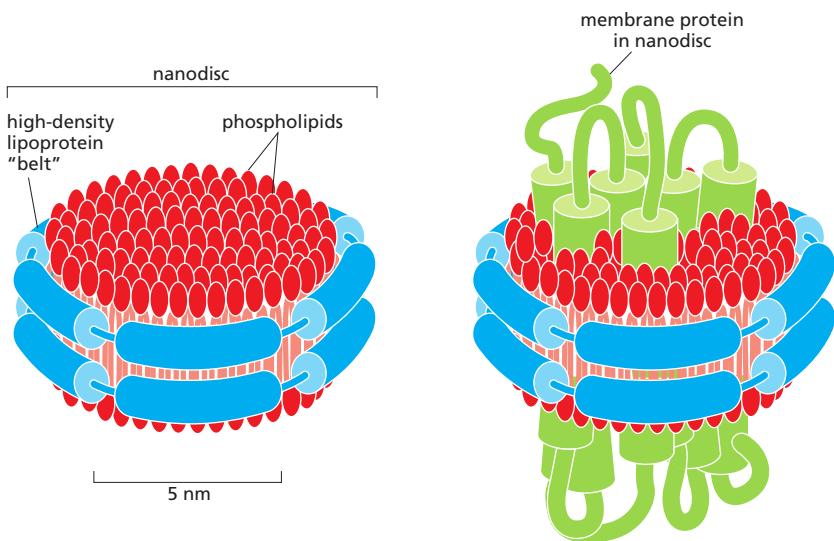


removal but do not unfold the protein. If the detergent concentration of a solution of solubilized membrane proteins is reduced (by dilution, for example), membrane proteins do not remain soluble. In the presence of an excess of phospholipid molecules in such a solution, however, membrane proteins incorporate into small liposomes that form spontaneously. In this way, functionally active membrane protein systems can be reconstituted from purified components, providing a powerful means of analyzing the activities of membrane transporters, ion channels, signaling receptors, and so on (Figure 10–28). Such functional reconstitution, for example, provided proof for the hypothesis that the enzymes that make

**Figure 10–27 Solubilizing a membrane protein with a mild nonionic detergent.** The detergent disrupts the lipid bilayer and brings the protein into solution as protein–lipid–detergent complexes. The phospholipids in the membrane are also solubilized by the detergent, as lipid–detergent micelles.



**Figure 10–28 The use of mild nonionic detergents for solubilizing, purifying, and reconstituting functional membrane protein systems.** In this example, functional  $\text{Na}^+-\text{K}^+$  pump molecules are purified and incorporated into phospholipid vesicles. This pump is present in the plasma membrane of most animal cells, where it uses the energy of ATP hydrolysis to pump  $\text{Na}^+$  out of the cell and  $\text{K}^+$  in, as discussed in Chapter 11.



**Figure 10–29 Model of a membrane protein reconstituted into a nanodisc.** When detergent is removed from a solution containing a multipass membrane protein, lipids, and a protein subunit of the high-density lipoprotein (HDL), the membrane protein becomes embedded in a small patch of lipid bilayer, which is surrounded by a belt of the HDL protein. In such nanodiscs, the hydrophobic edges of the bilayer patch are shielded by the protein belt, which renders the assembly water-soluble.

ATP (ATP synthases) use H<sup>+</sup> gradients in mitochondrial, chloroplast, and bacterial membranes to produce ATP.

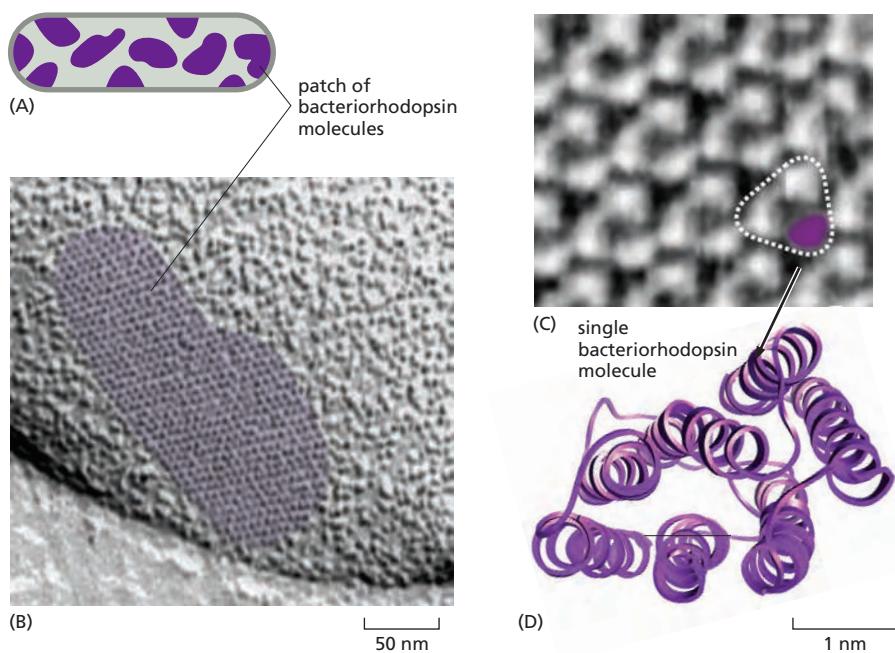
Membrane proteins can also be reconstituted from detergent solution into nanodiscs, which are small, uniformly sized patches of membrane that are surrounded by a belt of protein, which covers the exposed edge of the bilayer to keep the patch in solution (Figure 10–29). The belt is derived from high-density lipoproteins (HDL), which keep lipids soluble for transport in the blood. In nanodiscs the membrane protein of interest can be studied in its native lipid environment and is experimentally accessible from both sides of the bilayer, which is useful, for example, for ligand-binding experiments. Proteins contained in nanodiscs can also be analyzed by single particle electron microscopy techniques to determine their structure. By this rapidly improving technique (discussed in Chapter 9), the structure of a membrane protein can be determined to high resolution without a requirement of the protein of interest to crystallize into a regular lattice, which is often hard to achieve for membrane proteins.

Detergents have also played a crucial part in the purification and crystallization of membrane proteins. The development of new detergents and new expression systems that produce large quantities of membrane proteins from cDNA clones has led to a rapid increase in the number of three-dimensional structures of membrane proteins and protein complexes that are known, although they are still few compared to the known structures of water-soluble proteins and protein complexes.

### Bacteriorhodopsin Is a Light-driven Proton (H<sup>+</sup>) Pump That Traverses the Lipid Bilayer as Seven $\alpha$ Helices

In Chapter 11, we consider how multipass transmembrane proteins mediate the selective transport of small hydrophilic molecules across cell membranes. But a detailed understanding of how such a membrane transport protein works requires precise information about its three-dimensional structure in the bilayer. *Bacteriorhodopsin* was the first membrane transport protein whose structure was determined, and it has remained the prototype of many multipass membrane proteins with a similar structure.

The “purple membrane” of the archaeon *Halobacterium salinarum* is a specialized patch in the plasma membrane that contains a single species of protein molecule, **bacteriorhodopsin** (Figure 10–30A). The protein functions as a light-activated H<sup>+</sup> pump that transfers H<sup>+</sup> out of the archaeal cell. Because the bacteriorhodopsin molecules are tightly packed and arranged as a planar two-dimensional crystal (Figure 10–30B and C), it was possible to determine their three-dimensional structure by combining electron microscopy and electron diffraction analysis—a procedure called *electron crystallography*, which we

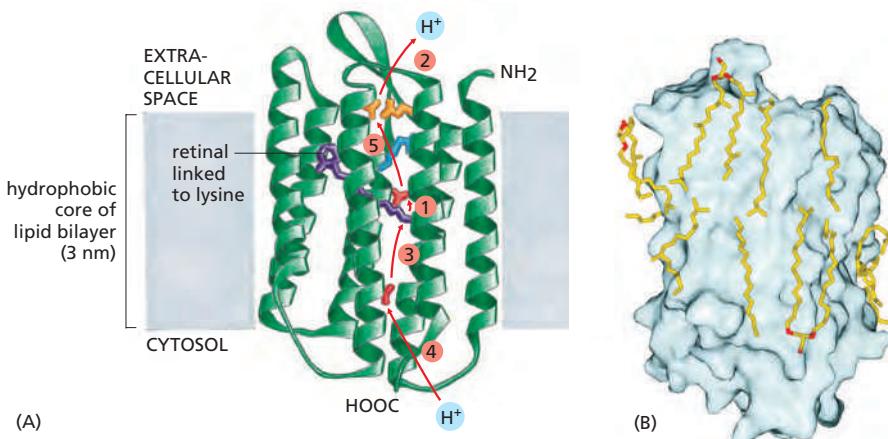


**Figure 10-30** Patches of purple membrane, which contain bacteriorhodopsin in the archaeon *Halobacterium salinarum*. (A) These archaea live in saltwater pools, where they are exposed to sunlight. They have evolved a variety of light-activated proteins, including bacteriorhodopsin, which is a light-activated H<sup>+</sup> pump in the plasma membrane. (B) The bacteriorhodopsin molecules in the purple membrane patches are tightly packed into two-dimensional crystalline arrays. (C) Details of the molecular surface visualized by atomic force microscopy. With this technique, individual bacteriorhodopsin molecules can be seen. (D) Outline of the approximate location of the bacteriorhodopsin monomer and the individual  $\alpha$  helices in the image shown in (C). (B–C, courtesy of Dieter Oesterhelt; D, PDB code: 2BRD.)

mentioned earlier. This method has provided the first structural views of many membrane proteins that were found to be difficult to crystallize from detergent solutions. For bacteriorhodopsin, the structure was later confirmed and extended to very high resolution by x-ray crystallography.

Each bacteriorhodopsin molecule is folded into seven closely packed transmembrane  $\alpha$  helices and contains a single light-absorbing group, or chromophore (in this case, *retinal*), which gives the protein its purple color. Retinal is vitamin A in its aldehyde form and is identical to the chromophore found in *rhodopsin* of the photoreceptor cells of the vertebrate eye (discussed in Chapter 15). Retinal is covalently linked to a lysine side chain of the bacteriorhodopsin protein. When activated by a single photon of light, the excited chromophore changes its shape and causes a series of small conformational changes in the protein, resulting in the transfer of one H<sup>+</sup> from the inside to the outside of the cell (Figure 10-31A). In bright light, each bacteriorhodopsin molecule can pump several hundred protons per second. The light-driven proton transfer establishes an H<sup>+</sup> gradient across the plasma membrane, which in turn drives the production of ATP by a second protein in the cell's plasma membrane. The energy stored in the H<sup>+</sup> gradient also drives other energy-requiring processes in the cell. Thus, bacteriorhodopsin converts solar energy into a H<sup>+</sup> gradient, which provides energy to the archaeal cell.

The high-resolution crystal structure of bacteriorhodopsin reveals many lipid molecules bound in specific places on the protein surface (Figure 10-31B).



**Figure 10-31** The three-dimensional structure of a bacteriorhodopsin molecule. (Movie 10.5) (A) The polypeptide chain crosses the lipid bilayer seven times as  $\alpha$  helices. The location of the retinal chromophore (purple) and the probable pathway taken by H<sup>+</sup> during the light-activated pumping cycle are shown. The first and key step is the passing of an H<sup>+</sup> from the chromophore to the side chain of aspartic acid 85 (red, located next to the chromophore) that occurs upon absorption of a photon by the chromophore. Subsequently, other H<sup>+</sup> transfers—in the numerical order indicated and utilizing the hydrophilic amino acid side chains that line a path through the membrane—complete the pumping cycle and return the enzyme to its starting state. Color code: glutamic acid (orange), aspartic acid (red), arginine (blue). (B) The high-resolution crystal structure of bacteriorhodopsin shows many lipid molecules (yellow with red head groups) that are tightly bound to specific places on the surface of the protein. (A, adapted from H. Luecke et al., *Science* 286:255–261, 1999. With permission from AAAS; B, from H. Luecke et al., *J. Mol. Biol.* 291:899–911, 1999. With permission from Academic Press.)

Interactions with specific lipids are thought to help stabilize many membrane proteins, which work best and sometimes crystallize more readily if some of the lipids remain bound during detergent extraction, or if specific lipids are added back to the proteins in detergent solutions. The specificity of these lipid–protein interactions helps explain why eukaryotic membranes contain such a variety of lipids, with head groups that differ in size, shape, and charge. We can think of the membrane lipids as constituting a two-dimensional solvent for the proteins in the membrane, just as water constitutes a three-dimensional solvent for proteins in an aqueous solution: some membrane proteins can function only in the presence of specific lipid head groups, just as many enzymes in aqueous solution require a particular ion for activity.

Bacteriorhodopsin is a member of a large superfamily of membrane proteins with similar structures but different functions. For example, rhodopsin in rod cells of the vertebrate retina and many cell-surface receptor proteins that bind extracellular signal molecules are also built from seven transmembrane  $\alpha$  helices. These proteins function as signal transducers rather than as transporters: each responds to an extracellular signal by activating a GTP-binding protein (G protein) inside the cell and they are therefore called *G-protein-coupled receptors (GPCRs)*, as we discuss in Chapter 15 (see Figure 15–6B). Although the structures of bacteriorhodopsins and GPCRs are strikingly similar, they show no sequence similarity and thus probably belong to two evolutionarily distant branches of an ancient protein family. A related class of membrane proteins, the *channelrhodopsins* that green algae use to detect light, form ion channels when they absorb a photon. When engineered so that they are expressed in animal brains, these proteins have become invaluable tools in neurobiology because they allow specific neurons to be stimulated experimentally by shining light on them, as we discuss in Chapter 11 (Figure 11–32).

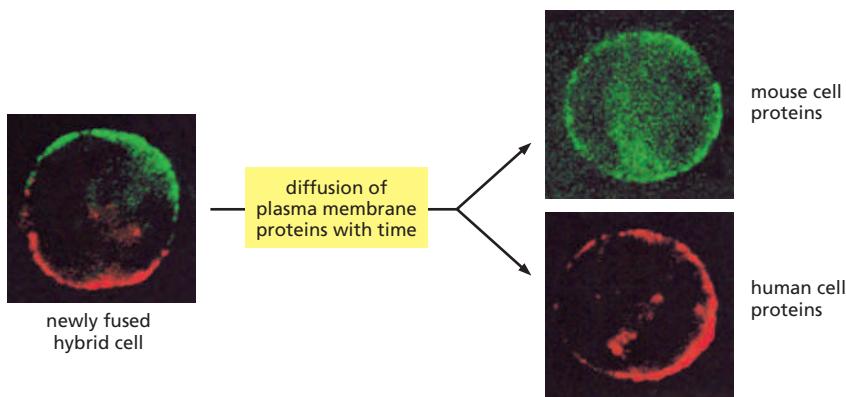
### Membrane Proteins Often Function as Large Complexes

Many membrane proteins function as part of multicomponent complexes, several of which have been studied by x-ray crystallography. One is a bacterial *photosynthetic reaction center*, which was the first membrane protein complex to be crystallized and analyzed by x-ray diffraction. In Chapter 14, we discuss how such photosynthetic complexes function to capture light energy and use it to pump H<sup>+</sup> across the membrane. Many of the membrane protein complexes involved in photosynthesis, proton pumping, and electron transport are even larger than the photosynthetic reaction center. The enormous photosystem II complex from cyanobacteria, for example, contains 19 protein subunits and well over 60 transmembrane helices (see Figure 14–49). Membrane proteins are often arranged in large complexes, not only for harvesting various forms of energy, but also for transducing extracellular signals into intracellular ones (discussed in Chapter 15).

### Many Membrane Proteins Diffuse in the Plane of the Membrane

Like most membrane lipids, membrane proteins do not tumble (*flip-flop*) across the lipid bilayer, but they do rotate about an axis perpendicular to the plane of the bilayer (*rotational diffusion*). In addition, many membrane proteins are able to move laterally within the membrane (*lateral diffusion*). An experiment in which mouse cells were artificially fused with human cells to produce hybrid cells (*heterokaryons*) provided the first direct evidence that some plasma membrane proteins are mobile in the plane of the membrane. Two differently labeled antibodies were used to distinguish selected mouse and human plasma membrane proteins. Although at first the mouse and human proteins were confined to their own halves of the newly formed heterokaryon, the two sets of proteins diffused and mixed over the entire cell surface in about half an hour (Figure 10–32).

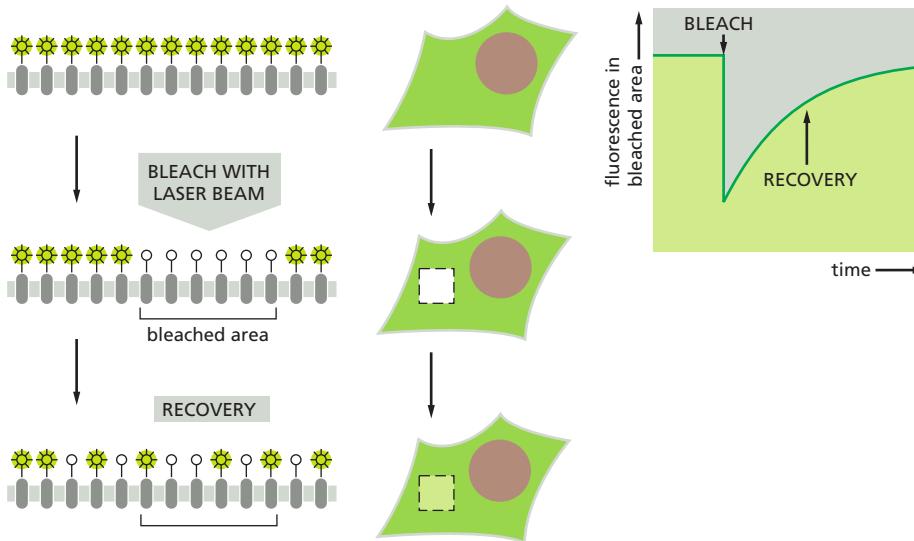
The lateral diffusion rates of membrane proteins can be measured by using the technique of *fluorescence recovery after photobleaching (FRAP)*. The method usually involves marking the membrane protein of interest with a specific fluorescent group. This can be done either with a fluorescent ligand such as a



**Figure 10–32** An experiment demonstrating the diffusion of proteins in the plasma membrane of mouse-human hybrid cells. In this experiment, a mouse and a human cell were fused to create a hybrid cell, which was then stained with two fluorescently labeled antibodies. One antibody (labeled with a green dye) detects mouse plasma membrane proteins, the other antibody (labeled with a red dye) detects human plasma membrane proteins. When cells were stained immediately after fusion, mouse and human plasma membrane proteins are still found in the membrane domains originating from the mouse and human cell, respectively. After a short time, however, the plasma membrane proteins diffuse over the entire cell surface and completely intermix. (From L.D. Frye and M. Edidine, *J. Cell Sci.* 7:319–335, 1970. With permission from The Company of Biologists.)

fluorophore-labeled antibody that binds to the protein or with recombinant DNA technology to express the protein fused to a fluorescent protein such as green fluorescent protein (GFP) (discussed in Chapter 9). The fluorescent group is then bleached in a small area of membrane by a laser beam, and the time taken for adjacent membrane proteins carrying unbleached ligand or GFP to diffuse into the bleached area is measured (Figure 10–33). From FRAP measurements, we can estimate the diffusion coefficient for the marked cell-surface protein. The values of the diffusion coefficients for different membrane proteins in different cells are highly variable, because interactions with other proteins impede the diffusion of the proteins to varying degrees. Measurements of proteins that are minimally impeded in this way indicate that cell membranes have a viscosity comparable to that of olive oil.

One drawback to the FRAP technique is that it monitors the movement of large populations of molecules in a relatively large area of membrane; one cannot follow individual protein molecules. If a protein fails to migrate into a bleached area, for example, one cannot tell whether the molecule is truly immobile or just restricted in its movement to a very small region of membrane—perhaps by cytoskeletal proteins. *Single-particle tracking* techniques overcome this problem by labeling individual membrane molecules with antibodies coupled to fluorescent dyes or tiny gold particles and tracking their movement by video microscopy. Using single-particle tracking, one can record the diffusion path of a single membrane protein molecule over time. Results from all of these techniques indicate that plasma membrane proteins differ widely in their diffusion characteristics, as we now discuss.



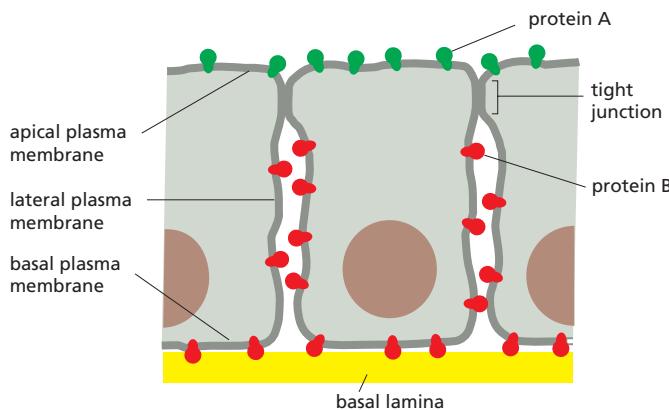
**Figure 10–33** Measuring the rate of lateral diffusion of a membrane protein by fluorescence recovery after photobleaching. A specific protein of interest can be expressed as a fusion protein with green fluorescent protein (GFP), which is intrinsically fluorescent. The fluorescent molecules are bleached in a small area using a laser beam. The fluorescence intensity recovers as the bleached molecules diffuse away and unbleached molecules diffuse into the irradiated area (shown here in side and top views). The diffusion coefficient is calculated from a graph of the rate of recovery: the greater the diffusion coefficient of the membrane protein, the faster the recovery (Movie 10.6).

## Cells Can Confine Proteins and Lipids to Specific Domains Within a Membrane

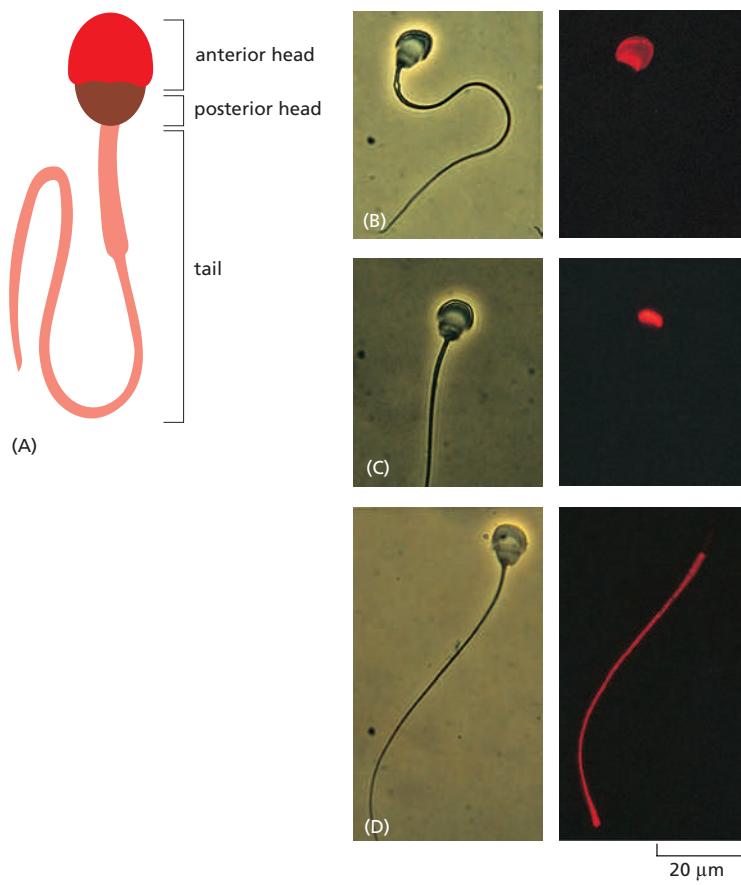
The recognition that biological membranes are two-dimensional fluids was a major advance in understanding membrane structure and function. It has become clear, however, that the picture of a membrane as a lipid sea in which all proteins float freely is greatly oversimplified. Most cells confine membrane proteins to specific regions in a continuous lipid bilayer. We have already discussed how bacteriorhodopsin molecules in the purple membrane of *Halobacterium* assemble into large two-dimensional crystals, in which individual protein molecules are relatively fixed in relationship to one another (see Figure 10–30). ATP synthase complexes in the inner mitochondrial membrane also associate into long double rows, as we discuss in Chapter 14 (see Figure 14–32). Large aggregates of this kind diffuse very slowly.

In epithelial cells, such as those that line the gut or the tubules of the kidney, certain plasma membrane enzymes and transport proteins are confined to the apical surface of the cells, whereas others are confined to the basal and lateral surfaces (Figure 10–34). This asymmetric distribution of membrane proteins is often essential for the function of the epithelium, as we discuss in Chapter 11 (see Figure 11–11). The lipid compositions of these two membrane domains are also different, demonstrating that epithelial cells can prevent the diffusion of lipid as well as protein molecules between the domains. The barriers set up by a specific type of intercellular junction (called a *tight junction*, discussed in Chapter 19; see Figure 19–18) maintain the separation of both protein and lipid molecules. Clearly, the membrane proteins that form these intercellular junctions cannot be allowed to diffuse laterally in the interacting membranes.

A cell can also create membrane domains without using intercellular junctions. As we already discussed, regulated protein–protein interactions in membranes are thought to create nanoscale raft domains that function in signaling and membrane trafficking. A more extreme example is seen in the mammalian spermatozoon, a single cell that consists of several structurally and functionally distinct parts covered by a continuous plasma membrane. When a sperm cell is examined by immunofluorescence microscopy with a variety of antibodies, each of which reacts with a specific cell-surface molecule, the plasma membrane is found to consist of at least three distinct domains (Figure 10–35). Some of the membrane molecules are able to diffuse freely within the confines of their own domain. The molecular nature of the “fence” that prevents the molecules from



**Figure 10–34 How membrane molecules can be restricted to a particular membrane domain.** In this drawing of an epithelial cell, protein A (in the apical domain of the plasma membrane) and protein B (in the basal and lateral domains) can diffuse laterally in their own domains but are prevented from entering the other domain, at least partly by the specialized cell–cell junction called a tight junction. Lipid molecules in the outer (extracellular) monolayer of the plasma membrane are likewise unable to diffuse between the two domains; lipids in the inner (cytosolic) monolayer, however, are able to do so (not shown). The basal lamina is a thin mat of extracellular matrix that separates epithelial sheets from other tissues (discussed in Chapter 19).



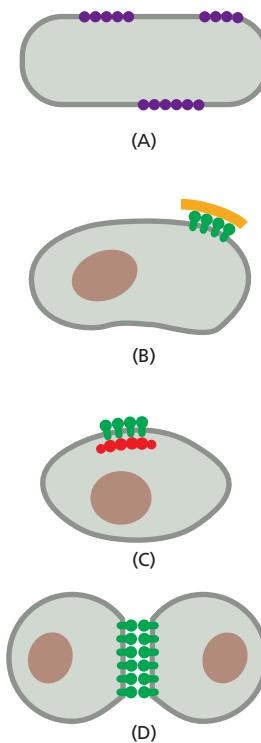
**Figure 10-35** Three domains in the plasma membrane of a guinea pig sperm. (A) A drawing of a guinea pig sperm. (B–D) In the three pairs of micrographs, phase-contrast micrographs are on the left, and the same cell is shown with cell-surface immunofluorescence staining on the right. Different monoclonal antibodies selectively label cell-surface molecules on (B) the anterior head, (C) the posterior head, and (D) the tail. (Micrographs courtesy of Selena Carroll and Diana Myles.)

leaving their domain is not known. Many other cells have similar membrane fences that confine membrane protein diffusion to certain membrane domains. The plasma membrane of nerve cells, for example, contains a domain enclosing the cell body and dendrites, and another enclosing the axon; it is thought that a belt of actin filaments tightly associated with the plasma membrane at the cell-body–axon junction forms part of the barrier.

**Figure 10-36** shows four common ways of immobilizing specific membrane proteins through protein–protein interactions.

### The Cortical Cytoskeleton Gives Membranes Mechanical Strength and Restricts Membrane Protein Diffusion

As shown in Figure 10-36B and C, a common way in which a cell restricts the lateral mobility of specific membrane proteins is to tether them to macromolecular assemblies on either side of the membrane. The characteristic biconcave shape of a red blood cell (**Figure 10-37**), for example, results from interactions of its plasma membrane proteins with an underlying *cytoskeleton*, which consists mainly of a meshwork of the filamentous protein **spectrin**. Spectrin is a long, thin, flexible rod about 100 nm in length. As the principal component of the red cell cytoskeleton, it maintains the structural integrity and shape of the plasma membrane, which is the red cell's only membrane, as the cell has no nucleus or other organelles. The spectrin cytoskeleton is riveted to the membrane through various membrane proteins. The final result is a deformable, netlike meshwork that covers the entire cytosolic surface of the red cell membrane (**Figure 10-38**). This spectrin-based cytoskeleton enables the red cell to withstand the stress on its membrane as it is forced through narrow capillaries. Mice and humans with genetic abnormalities in spectrin are anemic and have red cells that are spherical (instead of concave) and fragile; the severity of the anemia increases with the degree of spectrin deficiency.

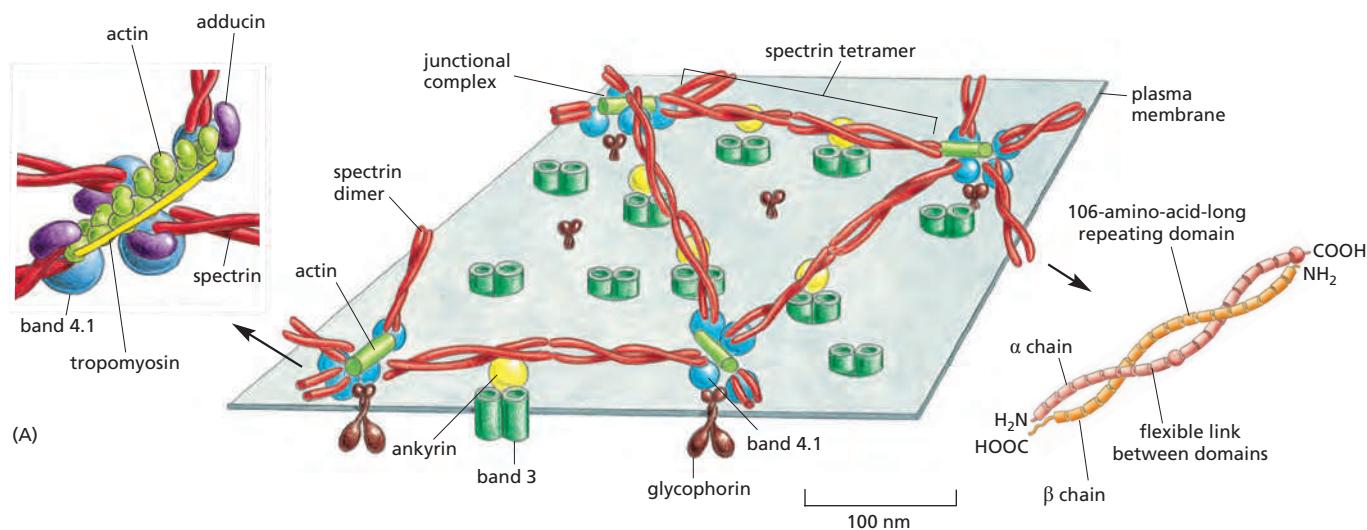


**Figure 10-36** Four ways of restricting the lateral mobility of specific plasma membrane proteins. (A) The proteins can self-assemble into large aggregates (as seen for bacteriorhodopsin in the purple membrane of *Halobacterium salinarum*); they can be tethered by interactions with assemblies of macromolecules (B) outside or (C) inside the cell; or (D) they can interact with proteins on the surface of another cell.



**Figure 10–37** A scanning electron micrograph of human red blood cells. The cells have a biconcave shape and lack a nucleus and other organelles (**Movie 10.7**). (Courtesy of Bernadette Chailley.)

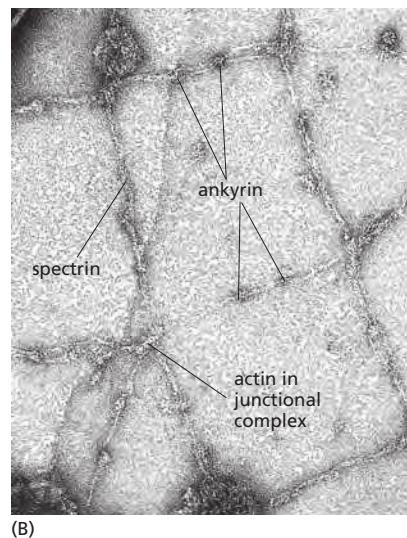
An analogous but much more elaborate and highly dynamic cytoskeletal network exists beneath the plasma membrane of most other cells in our body. This network, which constitutes the **cortex** of the cell, is rich in actin filaments, which are attached to the plasma membrane in numerous ways. The dynamic remodeling of the cortical actin network provides a driving force for many essential cell functions, including cell movement, endocytosis, and the formation of transient, mobile plasma membrane structures such as filopodia and lamellopodia.

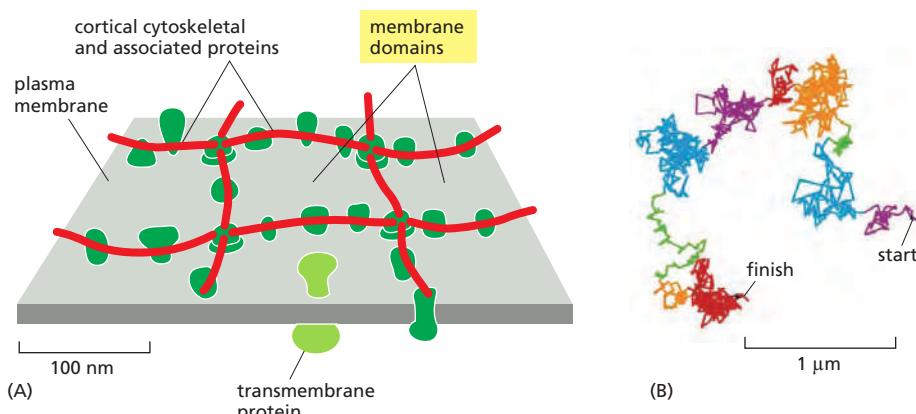


**Figure 10–38** The spectrin-based cytoskeleton on the cytosolic side of the human red blood cell plasma membrane. (A) The arrangement shown in the drawing has been deduced mainly from studies on the interactions of purified proteins *in vitro*. Spectrin heterodimers (enlarged in the drawing on the right) are linked together into a netlike meshwork by “junctional complexes” (enlarged in the drawing on the left). Each spectrin heterodimer consists of two antiparallel, loosely intertwined, flexible polypeptide chains called  $\alpha$  and  $\beta$ . The two spectrin chains are attached noncovalently to each other at multiple points, including at both ends. Both the  $\alpha$  and  $\beta$  chains are composed largely of repeating domains. Two spectrin heterodimers join end-to-end to form tetramers.

The junctional complexes are composed of short actin filaments (containing 13 actin monomers) and these proteins—band 4.1, adducin, and a tropomyosin molecule that probably determines the length of the actin filaments. The cytoskeleton is linked to the membrane through two transmembrane proteins—a multipass protein called band 3 and a single-pass protein called glycophorin. The spectrin tetramers bind to some band 3 proteins via ankyrin molecules, and to glycophorin and band 3 (not shown) via band 4.1 proteins.

(B) The electron micrograph shows the cytoskeleton on the cytosolic side of a red blood cell membrane after fixation and negative staining. The spectrin meshwork has been purposely stretched out to allow the details of its structure to be seen. In a normal cell, the meshwork shown would be much more crowded and occupy only about one-tenth of this area. (B, courtesy of T. Byers and D. Branton, *Proc. Natl. Acad. Sci. USA* 82:6153–6157, 1985. With permission from The National Academy of Sciences.)





**Figure 10–39** Corraling plasma membrane proteins by cortical cytoskeletal filaments. (A) The filaments are thought to provide diffusion barriers that divide the membrane into small domains, or corrals. (B) High-speed, single-particle tracking was used to follow the path of single fluorescently labeled membrane protein of one type over time. The trace shows that the individual protein molecules (the movement of each shown in a different color) diffuse within a tightly delimited membrane domain and only infrequently escape into a neighboring domain. (Adapted from A. Kusumi et al., *Annu. Rev. Biophys. Biomol. Struct.* 34:351–378, 2005. With permission from Annual Reviews.)

discussed in Chapter 16. The cortex of nucleated cells also contains proteins that are structurally homologous to spectrin and the other components of the red cell cytoskeleton. We discuss the cortical cytoskeleton in nucleated cells and its interactions with the plasma membrane in Chapter 16.

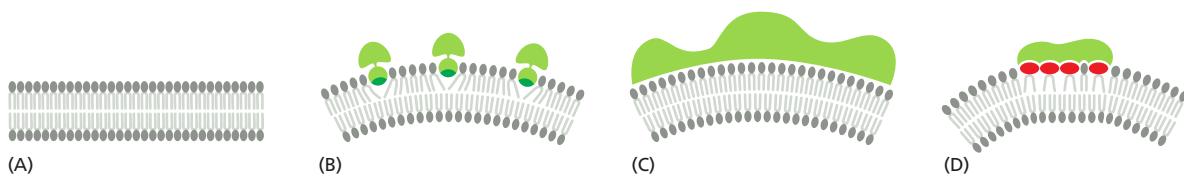
The cortical cytoskeletal network restricts diffusion of not only the plasma membrane proteins that are directly anchored to it. Because the cytoskeletal filaments are often closely apposed to the cytosolic surface of the plasma membrane, they can form mechanical barriers that obstruct the free diffusion of proteins in the membrane. These barriers partition the membrane into small domains, or *corrals* (Figure 10–39A), which can be either permanent, as in the sperm (see Figure 10–35), or transient. The barriers can be detected when the diffusion of individual membrane proteins is followed by high-speed, single-particle tracking. The proteins diffuse rapidly but are confined within an individual corral (Figure 10–39B); occasionally, however, thermal motions cause a few cortical filaments to detach transiently from the membrane, allowing the protein to escape into an adjacent corral.

The extent to which a transmembrane protein is confined within a corral depends on its association with other proteins and the size of its cytoplasmic domain; proteins with a large cytosolic domain will have a harder time passing through cytoskeletal barriers. When a cell-surface receptor binds its extracellular signal molecules, for example, large protein complexes build up on the cytosolic domain of the receptor, making it more difficult for the receptor to escape from its corral. It is thought that corraling helps concentrate such signaling complexes, increasing the speed and efficiency of the signaling process (discussed in Chapter 15).

### Membrane-bending Proteins Deform Bilayers

Cell membranes assume many different shapes, as illustrated by the elaborate and varied structures of cell-surface protrusions and membrane-enclosed organelles in eukaryotic cells. Flat sheets, narrow tubules, round vesicles, fenestrated sheets, and pitta bread-shaped cisternae are all part of the repertoire: often, a variety of shapes will be present in different regions of the same continuous bilayer. Membrane shape is controlled dynamically, as many essential cell processes—including vesicle budding, cell movement, and cell division—require elaborate transient membrane deformations. In many cases, membrane shape is influenced by dynamic pushing and pulling forces exerted by cytoskeletal or extracellular structures, as we discuss in Chapters 13 and 16). A crucial part in producing these deformations is played by **membrane-bending proteins**, which control local membrane curvature. Often, cytoskeletal dynamics and membrane-bending-protein forces work together. Membrane-bending proteins attach to specific membrane regions as needed and act by one or more of three principal mechanisms:

1. Some insert hydrophobic protein domains or attached lipid anchors into one of the leaflets of a lipid bilayer. Increasing the area of only one leaflet



**Figure 10–40** Three ways in which membrane-bending proteins shape membranes. Lipid bilayers are blue and proteins are green. (A) Bilayer without protein bound. (B) A hydrophobic region of the protein can insert as a wedge into one monolayer to pry lipid head groups apart. Such regions can either be amphiphilic helices as shown or hydrophobic hairpins. (C) The curved surface of the protein can bind to lipid head groups and deform the membrane or stabilize its curvature. (D) A protein can bind to and cluster lipids that have large head groups and thereby bend the membrane. (Adapted from W.A. Prinz and J.E. Hinshaw, *Crit. Rev. Biochem. Mol. Biol.* 44:278–291, 2009.)

causes the membrane to bend (Figure 10–40B). The proteins that shape the convoluted network of narrow ER tubules are thought to work in this way.

2. Some membrane-bending proteins form rigid scaffolds that deform the membrane or stabilize an already bent membrane (Figure 10–40C). The coat proteins that shape the budding vesicles in intracellular transport fall into this class.
3. Some membrane-bending proteins cause particular membrane lipids to cluster together, thereby inducing membrane curvature. The ability of a lipid to induce positive or negative membrane curvature is determined by the relative cross-sectional areas of its head group and its hydrocarbon tails. For example, the large head group of phosphoinositides make these lipid molecules wedge-shaped, and their accumulation in a domain of one leaflet of a bilayer therefore induces positive curvature (Figure 10–40D). By contrast, phospholipases that remove lipid head groups produce inversely shaped lipid molecules that induce negative curvature.

Often, different membrane-bending proteins collaborate to achieve a particular curvature, as in shaping a budding transport vesicle, as we discuss in Chapter 13.

## Summary

Whereas the lipid bilayer determines the basic structure of biological membranes, proteins are responsible for most membrane functions, serving as specific receptors, enzymes, transporters, and so on. Transmembrane proteins extend across the lipid bilayer. Some of these membrane proteins are single-pass proteins, in which the polypeptide chain crosses the bilayer as a single  $\alpha$  helix. Others are multipass proteins, in which the polypeptide chain crosses the bilayer multiple times—either as a series of  $\alpha$  helices or as a  $\beta$  sheet rolled up into the shape of a barrel. All proteins responsible for the transport of ions and other small water-soluble molecules through the membrane are multipass proteins. Some membrane proteins do not span the bilayer but instead are attached to either side of the membrane: some are attached to the cytosolic side by an amphipathic  $\alpha$  helix on the protein surface or by the covalent attachment of one or more lipid chains, others are attached to the non-cytosolic side by a GPI anchor. Some membrane-associated proteins are bound by noncovalent interactions with transmembrane proteins. In the plasma membrane of all eukaryotic cells, most of the proteins exposed on the cell surface and some of the lipid molecules in the outer lipid monolayer have oligosaccharide chains covalently attached to them. Like the lipid molecules in the bilayer, many membrane proteins are able to diffuse rapidly in the plane of the membrane. However, cells have ways of immobilizing specific membrane proteins, as well as ways of confining both membrane protein and lipid molecules to particular domains in a continuous lipid bilayer. The dynamic association of membrane-bending proteins confers on membranes their characteristic three-dimensional shapes.

## WHAT WE DON'T KNOW

- Given the highly complex lipid composition of cell membranes, what are the variations within different organelle membranes in an animal cell? What are the functional consequences of these differences, and what are the roles of the minor lipid species?
- Is the biophysical tendency of lipids to partition into separate phases within a lipid bilayer functionally utilized in cell membranes? If so, how is it regulated and what membrane functions does it control?
- How commonly do specific lipid molecules associate with membrane proteins to regulate their function?
- Given that the structure of only a tiny fraction of all membrane proteins has been determined, what new principles of membrane protein structure remain to be discovered?

## PROBLEMS

Which statements are true? Explain why or why not.

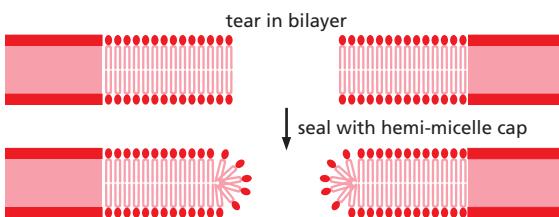
**10–1** Although lipid molecules are free to diffuse in the plane of the bilayer, they cannot flip-flop across the bilayer unless enzyme catalysts called phospholipid translocators are present in the membrane.

**10–2** Whereas all the carbohydrate in the plasma membrane faces outward on the external surface of the cell, all the carbohydrate on internal membranes faces toward the cytosol.

**10–3** Although membrane domains with different protein compositions are well known, there are at present no examples of membrane domains that differ in lipid composition.

Discuss the following problems.

**10–4** When a lipid bilayer is torn, why does it not seal itself by forming a “hemi-micelle” cap at the edges, as shown in **Figure Q10–1**?



**Figure Q10–1** A torn lipid bilayer sealed with a hypothetical “hemi-micelle” cap (Problem 10–4).

**10–5** Margarine is made from vegetable oil by a chemical process. Do you suppose this process converts saturated fatty acids to unsaturated ones, or vice versa? Explain your answer.

**10–7** Monomeric single-pass transmembrane proteins span a membrane with a single  $\alpha$  helix that has characteristic chemical properties in the region of the bilayer. Which of the three 20-amino-acid sequences listed below is the most likely candidate for such a transmembrane segment? Explain the reasons for your choice. (See back of book for one-letter amino acid code; FAMILY VW is a convenient mnemonic for hydrophobic amino acids.)

- A. I T L I Y F G V M A G V I G T I L L I S
- B. I T P I Y F G P M A G V I G T P L L I S
- C. I T E I Y F G R M A G V I G T D L L I S

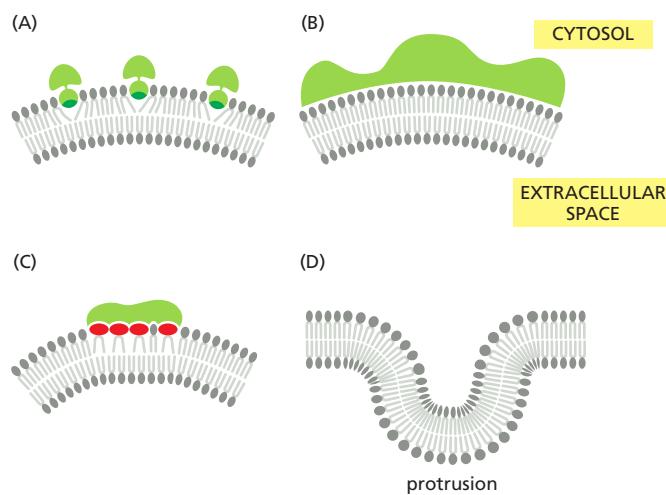
**10–6** If a lipid raft is typically 70 nm in diameter and each lipid molecule has a diameter of 0.5 nm, about how many lipid molecules would there be in a lipid raft composed entirely of lipid? At a ratio of 50 lipid molecules per protein molecule (50% protein by mass), how many

proteins would be in a typical raft? (Neglect the loss of lipid from the raft that would be required to accommodate the protein.)

**10–8** You are studying the binding of proteins to the cytoplasmic face of cultured neuroblastoma cells and have found a method that gives a good yield of inside-out vesicles from the plasma membrane. Unfortunately, your preparations are contaminated with variable amounts of right-side-out vesicles. Nothing you have tried avoids this problem. A friend suggests that you pass your vesicles over an affinity column made of lectin coupled to solid beads. What is the point of your friend’s suggestion?

**10–9** Glycophorin, a protein in the plasma membrane of the red blood cell, normally exists as a homodimer that is held together entirely by interactions between its transmembrane domains. Since transmembrane domains are hydrophobic, how is it that they can associate with one another so specifically?

**10–10** Three mechanisms by which membrane-binding proteins bend a membrane are illustrated in **Figure Q10–2A, B, and C**. As shown, each of these cytosolic membrane-bending proteins would induce an invagination of the plasma membrane. Could similar kinds of cytosolic proteins induce a protrusion of the plasma membrane (**Figure Q10–2D**)? Which ones? Explain how they might work.



**Figure Q10–2** Bending of the plasma membrane by cytosolic proteins (Problem 10–10). (A) Insertion of a protein “finger” into the cytosolic leaflet of the membrane. (B) Binding of lipids to the curved surface of a membrane-binding protein. (C) Binding of membrane proteins to membrane lipids with large head groups. (D) A segment of the plasma membrane showing a protrusion.

## REFERENCES

### General

- Bretscher MS (1973) Membrane structure: some general principles. *Science* 181, 622–629.
- Edidin M (2003) Lipids on the frontier: a century of cell-membrane bilayers. *Nat. Rev. Mol. Cell Biol.* 4, 414–418.
- Goñi FM (2014) The basic structure and dynamics of cell membranes: an update of the Singer-Nicolson model. *Biochim. Biophys. Acta* 1838, 1467–1476.
- Lipowsky R & Sackmann E (eds) (1995) *Structure and Dynamics of Membranes*. Amsterdam: Elsevier.
- Singer SJ & Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175, 720–731.
- Tanford C (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. New York: Wiley.

### The Lipid Bilayer

- Bevers EM, Comfurius P, Dekkers DW & Zwaal RF (1999) Lipid translocation across the plasma membrane of mammalian cells. *Biochim. Biophys. Acta* 1439, 317–330.
- Brügger B (2014) Lipidomics: analysis of the lipid composition of cells and subcellular organelles by electrospray ionization mass spectrometry. *Annu. Rev. Biochem.* 83, 79–98.
- Contreras FX, Sánchez-Magraner L, Alonso A & Goñi FM (2010) Transbilayer (flip-flop) lipid motion and lipid scrambling in membranes. *FEBS Lett.* 584, 1779–1786.
- Hakomori SI (2002) The glycosynapse. *Proc. Natl Acad. Sci. USA* 99, 225–232.
- Ichikawa S & Hirabayashi Y (1998) Glucosylceramide synthase and glycosphingolipid synthesis. *Trends Cell Biol.* 8, 198–202.
- Klose C, Surma MA & Simons K (2013) Organellar lipidomics—background and perspectives. *Curr. Opin. Cell Biol.* 25, 406–413.
- Kornberg RD & McConnell HM (1971) Lateral diffusion of phospholipids in a vesicle membrane. *Proc. Natl Acad. Sci. USA* 68, 2564–2568.
- Lingwood D & Simons K (2010) Lipid rafts as a membrane-organizing principle. *Science* 327, 46–50.
- Mansilla MC, Cybulski LE, Albanesi D & de Mendoza D (2004) Control of membrane lipid fluidity by molecular thermosensors. *J. Bacteriol.* 186, 6681–6688.
- Maxfield FR & van Meer G (2010) Cholesterol, the central lipid of mammalian cells. *Curr. Opin. Cell Biol.* 22, 422–429.
- McConnell HM & Radhakrishnan A (2003) Condensed complexes of cholesterol and phospholipids. *Biochim. Biophys. Acta* 1610, 159–173.
- Pomorski T & Menon AK (2006) Lipid flippases and their biological functions. *Cell. Mol. Life Sci.* 63, 2908–2921.
- Rothman JE & Lenard J (1977) Membrane asymmetry. *Science* 195, 743–753.
- Walther TC & Farese RV Jr (2012) Lipid droplets and cellular lipid metabolism. *Annu. Rev. Biochem.* 81, 687–714.

### Membrane Proteins

- Bennett V & Baines AJ (2001) Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol. Rev.* 81, 1353–1392.
- Bijlsma MJ & Marsh M (2003) The on-off story of protein palmitoylation. *Trends Cell Biol.* 13, 32–42.
- Branden C & Tooze J (1999) *Introduction to Protein Structure*, 2nd ed. New York: Garland Science.
- Bretscher MS & Raff MC (1975) Mammalian plasma membranes. *Nature* 258, 43–49.

Buchanan SK (1999) Beta-barrel proteins from bacterial outer membranes: structure, function and refolding. *Curr. Opin. Struct. Biol.* 9, 455–461.

Chen Y, Lagerholm BC, Yang B & Jacobson K (2006) Methods to measure the lateral diffusion of membrane lipids and proteins. *Methods* 39, 147–153.

Curran AR & Engelman DM (2003) Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* 13, 412–417.

Deisenhofer J & Michel H (1991) Structures of bacterial photosynthetic reaction centers. *Annu. Rev. Cell Biol.* 7, 1–23.

Drickamer K & Taylor ME (1993) Biology of animal lectins. *Annu. Rev. Cell Biol.* 9, 237–264.

Drickamer K & Taylor ME (1998) Evolving views of protein glycosylation. *Trends Biochem. Sci.* 23, 321–324.

Frye LD & Edidin M (1970) The rapid intermixing of cell surface antigens after formation of mouse-human heterokaryons. *J. Cell Sci.* 7, 319–335.

Helenius A & Simons K (1975) Solubilization of membranes by detergents. *Biochim. Biophys. Acta* 415, 29–79.

Henderson R & Unwin PN (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* 257, 28–32.

Kyte J & Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157, 105–132.

Lee AG (2003) Lipid-protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta* 1612, 1–40.

Marchesi VT, Furthmayr H & Tomita M (1976) The red cell membrane. *Annu. Rev. Biochem.* 45, 667–698.

Nakada C, Ritchie K, Oba Y et al. (2003) Accumulation of anchored proteins forms membrane diffusion barriers during neuronal polarization. *Nat. Cell Biol.* 5, 626–632.

Oesterhelt D (1998) The structure and mechanism of the family of retinal proteins from halophilic archaea. *Curr. Opin. Struct. Biol.* 8, 489–500.

Popot J-L (2010) Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solution. *Annu. Rev. Biochem.* 79, 737–775.

Prinz WA & Hinshaw JE (2009) Membrane-bending proteins. *Crit. Rev. Biochem. Mol. Biol.* 44, 278–291.

Rao M & Mayor S (2014) Active organization of membrane constituents in living cells. *Curr. Opin. Cell Biol.* 29, 126–132.

Reig N & van der Goot FG (2006) About lipids and toxins. *FEBS Lett.* 580, 5572–5579.

Sharon N & Lis H (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* 14, 53R–62R.

Sheetz MP (2001) Cell control by membrane-cytoskeleton adhesion. *Nat. Rev. Mol. Cell Biol.* 2, 392–396.

Shibata Y, Hu J, Kozlov MM & Rapoport TA (2009) Mechanisms shaping the membranes of cellular organelles. *Annu. Rev. Cell Dev. Biol.* 25, 329–354.

Steck TL (1974) The organization of proteins in the human red blood cell membrane. A review. *J. Cell Biol.* 62, 1–19.

Subramanian S (1999) The structure of bacteriorhodopsin: an emerging consensus. *Curr. Opin. Struct. Biol.* 9, 462–468.

Viel A & Branton D (1996) Spectrin: on the path from structure to function. *Curr. Opin. Cell Biol.* 8, 49–55.

Vinothkumar KR & Henderson R (2010) Structures of membrane proteins. *Q. Rev. Biophys.* 43, 65–158.

von Heijne G (2011) Membrane proteins: from bench to bits. *Biochem. Soc. Trans.* 39, 747–750.

White SH & Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28, 319–365.

# Membrane Transport of Small Molecules and the Electrical Properties of Membranes

CHAPTER  
11

Because of its hydrophobic interior, the lipid bilayer of cell membranes restricts the passage of most polar molecules. This barrier function allows the cell to maintain concentrations of solutes in its cytosol that differ from those in the extracellular fluid and in each of the intracellular membrane-enclosed compartments. To benefit from this barrier, however, cells have had to evolve ways of transferring specific water-soluble molecules and ions across their membranes in order to ingest essential nutrients, excrete metabolic waste products, and regulate intracellular ion concentrations. Cells use specialized *membrane transport proteins* to accomplish this goal. The importance of such small molecule transport is reflected in the large number of genes in all organisms that code for the transmembrane transport proteins involved, which make up 15–30% of the membrane proteins in all cells. Some mammalian cells, such as nerve and kidney cells, devote up to two-thirds of their total metabolic energy consumption to such transport processes.

Cells can also transfer macromolecules and even large particles across their membranes, but the mechanisms involved in most of these cases differ from those used for transferring small molecules, and they are discussed in Chapters 12 and 13.

We begin this chapter by describing some general principles of how small water-soluble molecules traverse cell membranes. We then consider, in turn, the two main classes of membrane proteins that mediate this transmembrane traffic: *transporters*, which undergo sequential conformational changes to transport specific small molecules across membranes, and *channels*, which form narrow pores, allowing passive transmembrane movement, primarily of water and small inorganic ions. Transporters can be coupled to a source of energy to catalyze active transport, which together with selective passive permeability, creates large differences in the composition of the cytosol compared with that of either the extracellular fluid (Table 11-1) or the fluid within membrane-enclosed organelles. By generating inorganic ion-concentration differences across the lipid bilayer, cell membranes can store potential energy in the form of electrochemical gradients, which drive various transport processes, convey electrical signals in electrically excitable cells, and (in mitochondria, chloroplasts, and bacteria) make most of the cell's ATP. We focus our discussion mainly on transport across the plasma membrane, but similar mechanisms operate across the other membranes of the eukaryotic cell, as discussed in later chapters.

In the last part of the chapter, we concentrate mainly on the functions of ion channels in neurons (nerve cells). In these cells, channel proteins perform at their highest level of sophistication, enabling networks of neurons to carry out all the astonishing feats your brain is capable of.

## PRINCIPLES OF MEMBRANE TRANSPORT

We begin this section by describing the permeability properties of protein-free, synthetic lipid bilayers. We then introduce some of the terms used to describe the various forms of membrane transport and some strategies for characterizing the proteins and processes involved.

### IN THIS CHAPTER

PRINCIPLES OF MEMBRANE TRANSPORT

TRANSPORTERS AND ACTIVE MEMBRANE TRANSPORT

CHANNELS AND THE ELECTRICAL PROPERTIES OF MEMBRANES

TABLE 11–1 A Comparison of Inorganic Ion Concentrations Inside and Outside a Typical Mammalian Cell*		
Component	Cytoplasmic concentration (mM)	Extracellular concentration (mM)
<b>Cations</b>		
Na <sup>+</sup>	5–15	145
K <sup>+</sup>	140	5
Mg <sup>2+</sup>	0.5	1–2
Ca <sup>2+</sup>	10 <sup>-4</sup>	1–2
H <sup>+</sup>	$7 \times 10^{-5}$ ( $10^{-7.2}$ M or pH 7.2)	$4 \times 10^{-5}$ ( $10^{-7.4}$ M or pH 7.4)
<b>Anions</b>		
Cl <sup>-</sup>	5–15	110

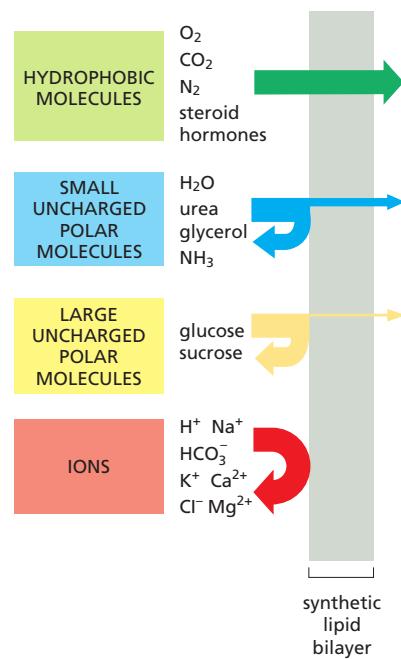
\*The cell must contain equal quantities of positive and negative charges (that is, it must be electrically neutral). Thus, in addition to Cl<sup>-</sup>, the cell contains many other anions not listed in this table; in fact, most cell constituents are negatively charged (HCO<sub>3</sub><sup>-</sup>, PO<sub>4</sub><sup>3-</sup>, nucleic acids, metabolites carrying phosphate and carboxyl groups, etc.). The concentrations of Ca<sup>2+</sup> and Mg<sup>2+</sup> given are for the free ions: although there is a total of about 20 mM Mg<sup>2+</sup> and 1–2 mM Ca<sup>2+</sup> in cells, both ions are mostly bound to other substances (such as proteins, free nucleotides, RNA, etc.) and, for Ca<sup>2+</sup>, stored within various organelles.

## Protein-Free Lipid Bilayers Are Impermeable to Ions

Given enough time, virtually any molecule will diffuse across a protein-free lipid bilayer down its concentration gradient. The rate of diffusion, however, varies enormously, depending partly on the size of the molecule but mostly on its relative hydrophobicity (solubility in oil). In general, the smaller the molecule and the more hydrophobic, or nonpolar, it is, the more easily it will diffuse across a lipid bilayer. Small nonpolar molecules, such as O<sub>2</sub> and CO<sub>2</sub>, readily dissolve in lipid bilayers and therefore diffuse rapidly across them. Small uncharged polar molecules, such as water or urea, also diffuse across a bilayer, albeit much more slowly (Figure 11–1 and see Movie 10.3). By contrast, lipid bilayers are essentially impermeable to charged molecules (ions), no matter how small: the charge and high degree of hydration of such molecules prevents them from entering the hydrocarbon phase of the bilayer (Figure 11–2).

## There Are Two Main Classes of Membrane Transport Proteins: Transporters and Channels

Like synthetic lipid bilayers, cell membranes allow small nonpolar molecules to permeate by diffusion. Cell membranes, however, also have to allow the passage of various polar molecules, such as ions, sugars, amino acids, nucleotides, water, and many cell metabolites that cross synthetic lipid bilayers only very slowly. Special **membrane transport proteins** transfer such solutes across cell membranes. These proteins occur in many forms and in all types of biological membranes. Each protein often transports only a specific molecular species or sometimes a class of molecules (such as ions, sugars, or amino acids). Studies in the 1950s found that bacteria with a single-gene mutation were unable to transport sugars across their plasma membrane, thereby demonstrating the specificity of membrane transport proteins. We now know that humans with similar mutations suffer from various inherited diseases that hinder the transport of a specific solute or solute class in the kidney, intestine, or other cell type. Individuals with the inherited disease *cystinuria*, for example, cannot transport certain amino acids (including cystine, the disulfide-linked dimer of cysteine) from either the urine or the intestine into the



**Figure 11–1** The relative permeability of a synthetic lipid bilayer to different classes of molecules. The smaller the molecule and, more importantly, the less strongly it associates with water, the more rapidly the molecule diffuses across the bilayer.

**Figure 11–2 Permeability coefficients for the passage of various molecules through synthetic lipid bilayers.** The rate of flow of a solute across the bilayer is directly proportional to the difference in its concentration on the two sides of the membrane. Multiplying this concentration difference (in mol/cm<sup>3</sup>) by the permeability coefficient (in cm/sec) gives the flow of solute in moles per second per square centimeter of bilayer. A concentration difference of tryptophan of 10<sup>-4</sup> mol/cm<sup>3</sup> (10<sup>-4</sup> mol / 10<sup>-3</sup> L = 0.1 M), for example, would cause a flow of 10<sup>-4</sup> mol/cm<sup>3</sup> × 10<sup>-7</sup> cm/sec = 10<sup>-11</sup> mol/sec through 1 cm<sup>2</sup> of bilayer, or 6 × 10<sup>4</sup> molecules/sec through 1 μm<sup>2</sup> of bilayer.

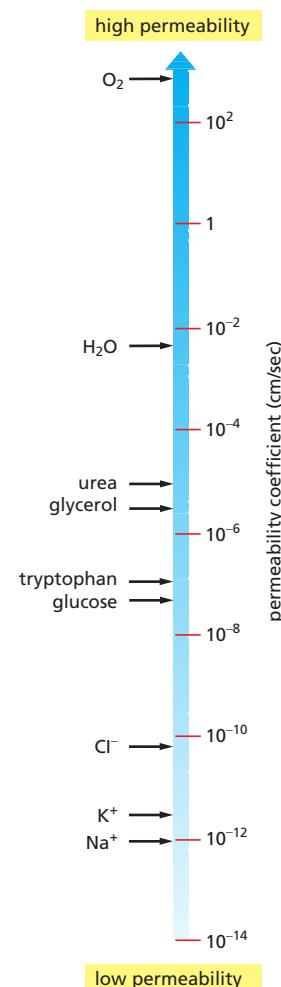
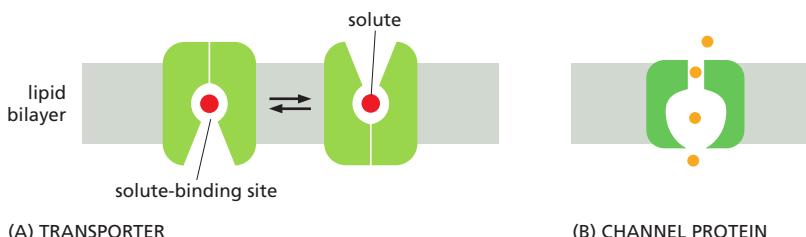
blood; the resulting accumulation of cystine in the urine leads to the formation of cystine stones in the kidneys.

All membrane transport proteins that have been studied in detail are multi-pass transmembrane proteins—that is, their polypeptide chains traverse the lipid bilayer multiple times. By forming a protein-lined pathway across the membrane, these proteins enable specific hydrophilic solutes to cross the membrane without coming into direct contact with the hydrophobic interior of the lipid bilayer.

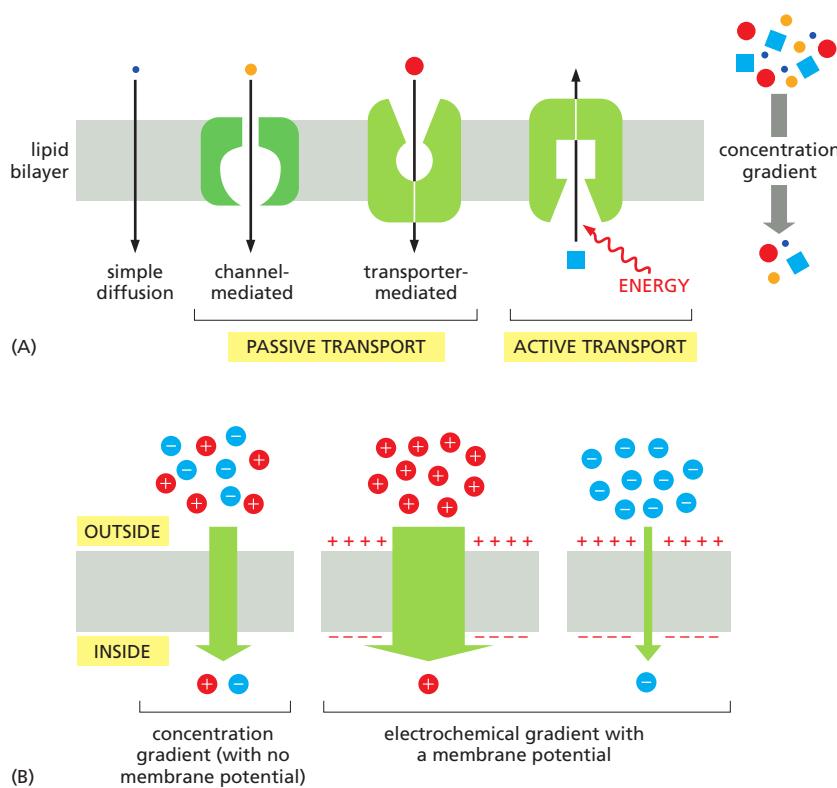
Transporters and channels are the two major classes of membrane transport proteins (Figure 11–3). **Transporters** (also called *carriers*, or *permeases*) bind the specific solute to be transported and undergo a series of conformational changes that alternately expose solute-binding sites on one side of the membrane and then on the other to transfer the solute across it. **Channels**, by contrast, interact with the solute to be transported much more weakly. They form continuous pores that extend across the lipid bilayer. When open, these pores allow specific solutes (such as inorganic ions of appropriate size and charge and in some cases small molecules, including water, glycerol, and ammonia) to pass through them and thereby cross the membrane. Not surprisingly, transport through channels occurs at a much faster rate than transport mediated by transporters. Although water can slowly diffuse across synthetic lipid bilayers, cells use dedicated channel proteins (called *water channels*, or *aquaporins*) that greatly increase the permeability of their membranes to water, as we discuss later.

### Active Transport Is Mediated by Transporters Coupled to an Energy Source

All channels and many transporters allow solutes to cross the membrane only passively (“downhill”), a process called **passive transport**. In the case of transport of a single uncharged molecule, the difference in the concentration on the two sides of the membrane—its *concentration gradient*—drives passive transport and determines its direction (Figure 11–4A). If the solute carries a net charge, however, both its concentration gradient and the electrical potential difference across the membrane, the *membrane potential*, influence its transport. The concentration gradient and the electrical gradient combine to form a net driving force, the **electrochemical gradient**, for each charged solute (Figure 11–4B). We discuss electrochemical gradients in more detail later and in Chapter 14. In fact, almost all plasma membranes have an electrical potential (i.e., a voltage) across them, with the inside usually negative with respect to the outside. This potential favors the entry of positively charged ions into the cell but opposes the entry of negatively charged ions (see Figure 11–4B); it also opposes the efflux of positively charged ions.



**Figure 11–3 Transporters and channel proteins.** (A) A transporter alternates between two conformations, so that the solute-binding site is sequentially accessible on one side of the bilayer and then on the other. (B) In contrast, a channel protein forms a pore across the bilayer through which specific solutes can passively diffuse.



**Figure 11-4** Different forms of membrane transport and the influence of the membrane. Passive transport down a concentration gradient (or an electrochemical gradient—see B below) occurs spontaneously, by diffusion, either through the lipid bilayer directly or through channels or passive transporters. By contrast, active transport requires an input of metabolic energy and is always mediated by transporters that pump the solute against its concentration or electrochemical gradient. (B) The electrochemical gradient of a charged solute (an ion) affects its transport. This gradient combines the membrane potential and the concentration gradient of the solute. The electrical and chemical gradients can work additively to increase the driving force on an ion across the membrane (middle) or can work against each other (right).

As shown in Figure 11-4A, in addition to passive transport, cells need to be able to actively pump certain solutes across the membrane “uphill,” against their electrochemical gradients. Such **active transport** is mediated by transporters whose pumping activity is directional because it is tightly coupled to a source of metabolic energy, such as an ion gradient or ATP hydrolysis, as discussed later. Transmembrane movement of small molecules mediated by transporters can be either active or passive, whereas that mediated by channels is always passive (see Figure 11-4A).

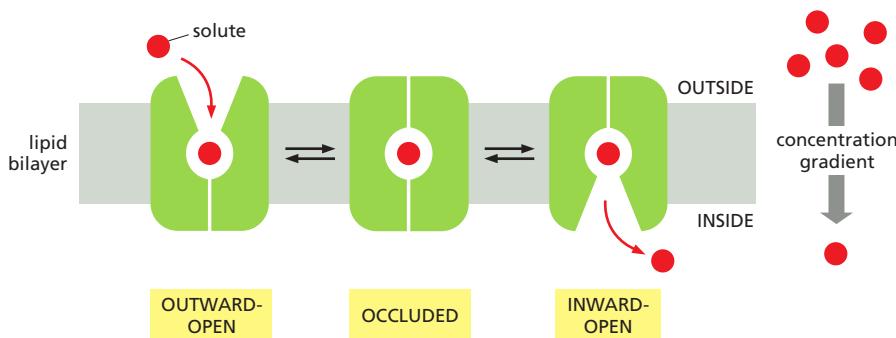
## Summary

Lipid bilayers are virtually impermeable to most polar molecules. To transport small water-soluble molecules into or out of cells or intracellular membrane-enclosed compartments, cell membranes contain various membrane transport proteins, each of which is responsible for transferring a particular solute or class of solutes across the membrane. There are two classes of membrane transport proteins—transporters and channels. Both form protein pathways across the lipid bilayer. Whereas transmembrane movement mediated by transporters can be either active or passive, solute flow through channel proteins is always passive. Both active and passive ion transport is influenced by the ion's concentration gradient and the membrane potential—that is, its electrochemical gradient.

## TRANSPORTERS AND ACTIVE MEMBRANE TRANSPORT

The process by which a transporter transfers a solute molecule across the lipid bilayer resembles an enzyme-substrate reaction, and in many ways transporters behave like enzymes. By contrast to ordinary enzyme-substrate reactions, however, the transporter does not modify the transported solute but instead delivers it unchanged to the other side of the membrane.

Each type of transporter has one or more specific binding sites for its solute (substrate). It transfers the solute across the lipid bilayer by undergoing reversible



conformational changes that alternately expose the solute-binding site first on one side of the membrane and then on the other—but never on both sides at the same time. The transition occurs through an intermediate state in which the solute is inaccessible, or occluded, from either side of the membrane (Figure 11–5). When the transporter is saturated (that is, when all solute-binding sites are occupied), the rate of transport is maximal. This rate, referred to as  $V_{\max}$  ( $V$  for velocity), is characteristic of the specific carrier.  $V_{\max}$  measures the rate at which the carrier can flip between its conformational states. In addition, each transporter has a characteristic affinity for its solute, reflected in the  $K_m$  of the reaction, which is equal to the concentration of solute when the transport rate is half its maximum value (Figure 11–6). As with enzymes, the binding of solute can be blocked by either competitive inhibitors (which compete for the same binding site and may or may not be transported) or noncompetitive inhibitors (which bind elsewhere and alter the structure of the transporter).

As we discuss shortly, it requires only a relatively minor modification of the model shown in Figure 11–5 to link a transporter to a source of energy in order to pump a solute uphill against its electrochemical gradient. Cells carry out such active transport in three main ways (Figure 11–7):

1. *Coupled transporters* harness the energy stored in concentration gradients to couple the uphill transport of one solute across the membrane to the downhill transport of another.
2. *ATP-driven pumps* couple uphill transport to the hydrolysis of ATP.
3. *Light- or redox-driven pumps*, which are known in bacteria, archaea, mitochondria, and chloroplasts, couple uphill transport to an input of energy from light, as with bacteriorhodopsin (discussed in Chapter 10), or from a redox reaction, as with cytochrome  $c$  oxidase (discussed in Chapter 14).

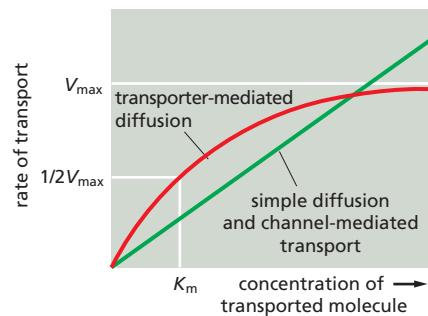
Amino acid sequence and three-dimensional structure comparisons suggest that, in many cases, there are strong similarities in structure between transporters that mediate active transport and those that mediate passive transport. Some bacterial transporters, for example, that use the energy stored in the  $H^+$  gradient across the plasma membrane to drive the active uptake of various sugars are structurally similar to the transporters that mediate passive glucose transport into most animal cells. This suggests an evolutionary relationship between various transporters. Given the importance of small metabolites and sugars as energy sources, it is not surprising that the superfamily of transporters is an ancient one.

We begin our discussion of active membrane transport by considering a class of coupled transporters that are driven by ion concentration gradients. These proteins have a crucial role in the transport of small metabolites across membranes in all cells. We then discuss ATP-driven pumps, including the  $Na^+-K^+$  pump that is found in the plasma membrane of most animal cells. Examples of the third class of active transport—light- or redox-driven pumps—are discussed in Chapter 14.

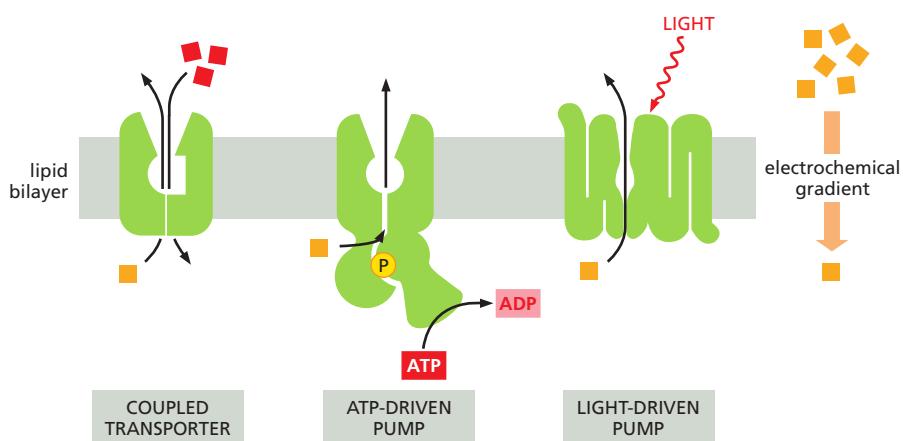
### Active Transport Can Be Driven by Ion-Concentration Gradients

Some transporters simply passively mediate the movement of a single solute from one side of the membrane to the other at a rate determined by their  $V_{\max}$  and

**Figure 11–5** A model of how a conformational change in a transporter mediates the passive movement of a solute. The transporter is shown in three conformational states: in the outward-open state, the binding sites for solute are exposed on the outside; in the occluded state, the same sites are not accessible from either side; and in the inward-open state, the sites are exposed on the inside. The transitions between the states occur randomly. They are completely reversible and do not depend on whether the solute-binding site is occupied. Therefore, if the solute concentration is higher on the outside of the bilayer, more solute binds to the transporter in the outward-open conformation than in the inward-open conformation, and there is a net transport of solute down its concentration gradient (or, if the solute is an ion, down its electrochemical gradient).



**Figure 11–6** The kinetics of simple diffusion compared with transporter-mediated diffusion. Whereas the rate of diffusion and channel-mediated transport is directly proportional to the solute concentration (within the physical limits imposed by total surface area or total channels available), the rate of transporter-mediated diffusion reaches a maximum ( $V_{\max}$ ) when the transporter is saturated. The solute concentration when the transport rate is at half its maximal value approximates the binding constant ( $K_m$ ) of the transporter for the solute and is analogous to the  $K_m$  of an enzyme for its substrate. The graph applies to a transporter moving a single solute; the kinetics of coupled transport of two or more solutes is more complex and exhibits cooperative behavior.

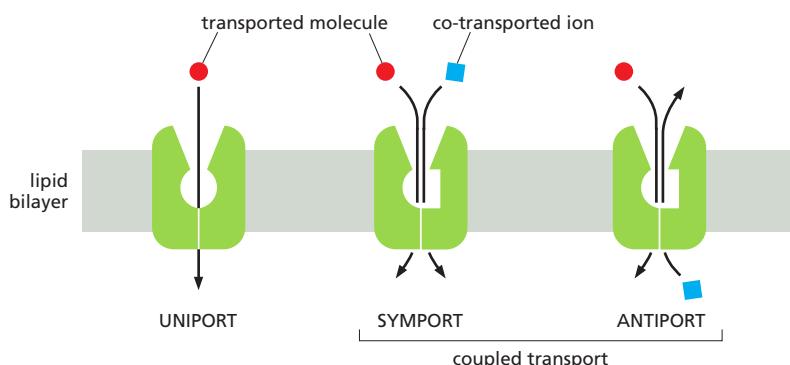


**Figure 11-7** Three ways of driving active transport. The actively transported molecule is shown in orange, and the energy source is shown in red. Redox driven active transport is discussed in Chapter 14 (see Figures 14–18 and 14–19).

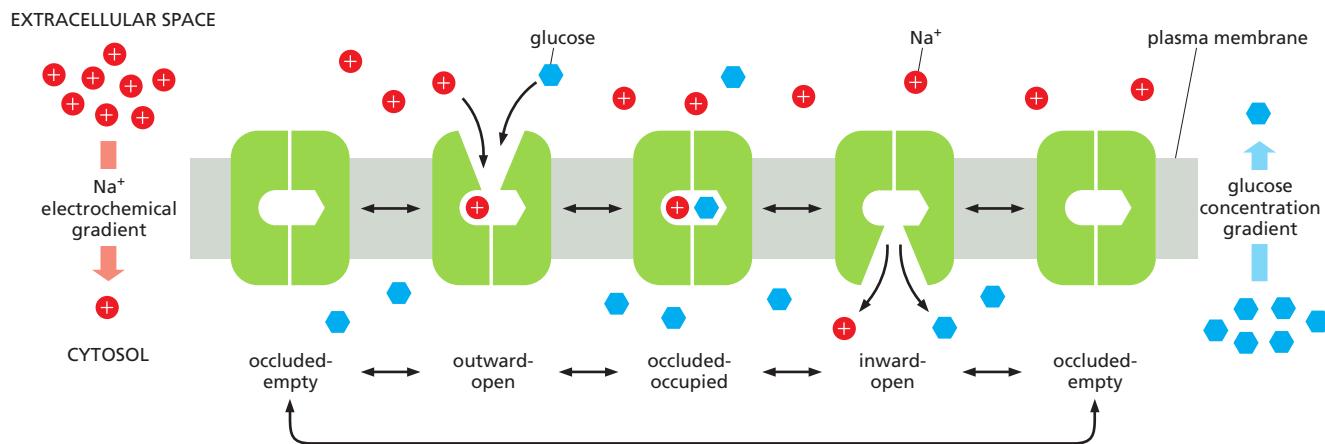
$K_m$ ; they are called **uniporters**. Others function as *coupled transporters*, in which the transfer of one solute strictly depends on the transport of a second. Coupled transport involves either the simultaneous transfer of a second solute in the same direction, performed by **symporters** (also called *co-transporters*), or the transfer of a second solute in the opposite direction, performed by **antiporters** (also called *exchangers*) (Figure 11–8).

The tight coupling between the transfer of two solutes allows the coupled transporters to harvest the energy stored in the electrochemical gradient of one solute, typically an inorganic ion, to transport the other. In this way, the free energy released during the movement of an inorganic ion down an electrochemical gradient is used as the driving force to pump other solutes uphill, against their electrochemical gradient. This strategy can work in either direction; some coupled transporters function as symporters, others as antiporters. In the plasma membrane of animal cells,  $\text{Na}^+$  is the usual co-transported ion because its electrochemical gradient provides a large driving force for the active transport of a second molecule. The  $\text{Na}^+$  that enters the cell during coupled transport is subsequently pumped out by an ATP-driven  $\text{Na}^+/\text{K}^+$  pump in the plasma membrane (as we discuss later), which, by maintaining the  $\text{Na}^+$  gradient, indirectly drives the coupled transport. Such ion-driven coupled transporters are just described are said to mediate *secondary active transport*. In contrast, ATP-driven pumps are said to mediate *primary active transport* because in these the free energy of ATP hydrolysis is used to directly drive the transport of a solute against its concentration gradient.

Intestinal and kidney epithelial cells contain a variety of symporters that are driven by the  $\text{Na}^+$  gradient across the plasma membrane. Each  $\text{Na}^+$ -driven symporter is specific for importing a small group of related sugars or amino acids into the cell. Because the  $\text{Na}^+$  tends to move into the cell down its electrochemical gradient, the sugar or amino acid is, in a sense, “dragged” into the cell with it. The greater the electrochemical gradient for  $\text{Na}^+$ , the more solute is pumped



**Figure 11–8** This schematic diagram shows transporters functioning as uniporters, symporters, and antiporters (Movie 11.1).

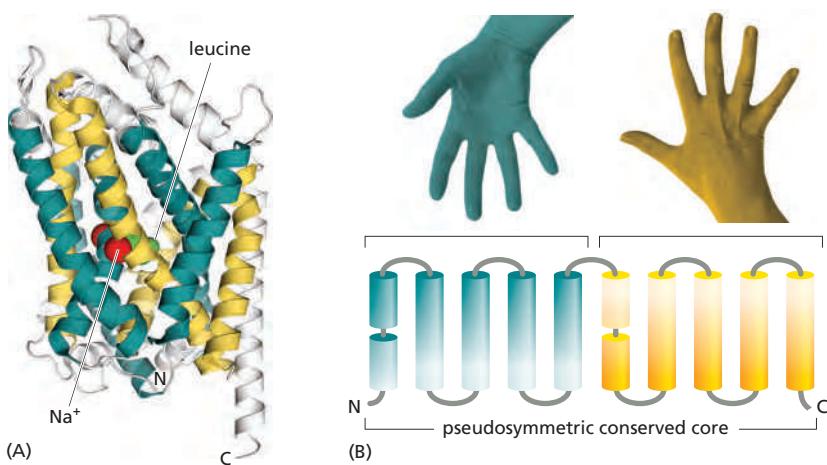


**Figure 11–9 Mechanism of glucose transport fueled by a Na<sup>+</sup> gradient.** As in the model shown in Figure 11–5, the transporter alternates between inward-open and outward-open states via an occluded intermediate state. Binding of Na<sup>+</sup> and glucose is cooperative—that is, the binding of either solute increases the protein's affinity for the other. Since the Na<sup>+</sup> concentration is much higher in the extracellular space than in the cytosol, glucose is more likely to bind to the transporter in the outward-facing state. The transition to the occluded state occurs only when both Na<sup>+</sup> and glucose are bound; their precise interactions in the solute-binding sites slightly stabilize the occluded state and thereby make this transition energetically favorable. Stochastic fluctuations caused by thermal energy drive the transporter randomly into the inward-open or outward-open conformation. If it opens outwardly, nothing is achieved, and the process starts all over. However, whenever it opens inwardly, Na<sup>+</sup> dissociates quickly in the low-Na<sup>+</sup>-concentration environment of the cytosol. Glucose dissociation is likewise enhanced when Na<sup>+</sup> is lost, because of cooperativity in binding of the two solutes. The overall result is the net transport of both Na<sup>+</sup> and glucose into the cell. Because the occluded state is not formed when only one of the solutes is bound, the transporter switches conformation only when it is fully occupied or fully empty, thereby assuring strict coupling of the transport of Na<sup>+</sup> and glucose.

into the cell (Figure 11–9). Neurotransmitters (released by nerve cells to signal at synapses—as we discuss later) are taken up again by Na<sup>+</sup> symporters after their release. These neurotransmitter transporters are important drug targets: stimulants, such as cocaine and antidepressants, inhibit them and thereby prolong signaling by the neurotransmitters, which are not cleared efficiently.

Despite their great variety, transporters share structural features that can explain how they function and how they evolved. Transporters are typically built from bundles of 10 or more  $\alpha$  helices that span the membrane. Solute- and ion-binding sites are located midway through the membrane, where some helices are broken or distorted and amino acid side chains and polypeptide backbone atoms form ion- and solute-binding sites. In the inward-open and outward-open conformations, these binding sites are accessible by passageways from one side of the membrane but not the other. In switching between the two conformations, the transporter protein transiently adopts an occluded conformation, in which both passageways are closed; this prevents the driving ion and the transported solute from crossing the membrane unaccompanied, which would deplete the cell's energy store to no purpose. Because only transporters with both types of binding sites appropriately filled change their conformation, tight coupling between ion and solute transport is assured.

Like enzymes, transporters can work in the reverse direction if ion and solute gradients are appropriately adjusted experimentally. This chemical symmetry is mirrored in their physical structure. Crystallographic analyses have revealed that transporters are built from *inverted repeats*: the packing of the transmembrane  $\alpha$  helices in one half of the helix bundle is structurally similar to the packing in the other half, but the two halves are inverted in the membrane relative to each other. Transporters are therefore said to be pseudosymmetric, and the passageways that open and close on either side of the membrane have closely similar geometries, allowing alternating access to the ion- and solute-binding sites in the center (Figure 11–10). It is thought that the two halves evolved by gene duplication of a smaller ancestor protein.



**Figure 11–10** Transporters are built from inverted repeats. (A) LeuT, a bacterial leucine/Na<sup>+</sup> symporter related to human neurotransmitter transporters, such as the serotonin transporter, is shown. The core of the transporter is built from two bundles, each composed of five  $\alpha$  helices (blue and yellow). The helices shown in gray differ among members of this transporter family and are thought to play regulatory roles, which are specific to a particular transporter. (B) Both core helix bundles are packed in a similar arrangement (shown as a hand, with the broken helix as the thumb), but the second bundle is inverted with respect to the first. The transporter's structural pseudosymmetry reflects its functional symmetry: the transporter can work in either direction, depending on the direction of the ion gradient. (Adapted from K.R. Vinothkumar and R. Henderson, *Q. Rev. Biophys.* 43:65–158, 2010. With permission from Cambridge University Press. PDB code: 3F3E.)

Some other types of important membrane transport proteins are also built from inverted repeats. Examples even include channel proteins such as the aquaporin water channel (discussed later) and the Sec61 channel through which nascent polypeptides move into the endoplasmic reticulum (discussed in Chapter 12). It is thought that these channels evolved from coupled transporters, in which the gating functions were lost, allowing them to open toward both sides of the membrane simultaneously to provide a continuous path across the membrane.

In bacteria, yeasts, and plants, as well as in many membrane-enclosed organelles of animal cells, most ion-driven active transport systems depend on H<sup>+</sup> rather than Na<sup>+</sup> gradients, reflecting the predominance of H<sup>+</sup> pumps in these membranes. An electrochemical H<sup>+</sup> gradient across the bacterial plasma membrane, for example, drives the inward active transport of many sugars and amino acids.

### Transporters in the Plasma Membrane Regulate Cytosolic pH

Most proteins operate optimally at a particular pH. Lysosomal enzymes, for example, function best at the low pH (~5) found in lysosomes, whereas cytosolic enzymes function best at the close-to-neutral pH (~7.2) found in the cytosol. It is therefore crucial that cells control the pH of their intracellular compartments.

Most cells have one or more types of Na<sup>+</sup>-driven antiporters in their plasma membrane that help to maintain the cytosolic pH at about 7.2. These transporters use the energy stored in the Na<sup>+</sup> gradient to pump out excess H<sup>+</sup>, which either leaks in or is produced in the cell by acid-forming reactions. Two mechanisms are used: either H<sup>+</sup> is directly transported out of the cell or HCO<sub>3</sub><sup>−</sup> is brought into the cell to neutralize H<sup>+</sup> in the cytosol (according to the reaction HCO<sub>3</sub><sup>−</sup> + H<sup>+</sup> → H<sub>2</sub>O + CO<sub>2</sub>). One of the antiporters that uses the first mechanism is a Na<sup>+</sup>-H<sup>+</sup> exchanger, which couples an influx of Na<sup>+</sup> to an efflux of H<sup>+</sup>. Another, which uses a combination of the two mechanisms, is a Na<sup>+</sup>-driven Cl<sup>−</sup>-HCO<sub>3</sub><sup>−</sup> exchanger that couples an influx of Na<sup>+</sup> and HCO<sub>3</sub><sup>−</sup> to an efflux of Cl<sup>−</sup> and H<sup>+</sup> (so that NaHCO<sub>3</sub> comes in and HCl goes out). The Na<sup>+</sup>-driven Cl<sup>−</sup>-HCO<sub>3</sub><sup>−</sup> exchanger is twice as effective as the Na<sup>+</sup>-H<sup>+</sup> exchanger: it pumps out one H<sup>+</sup> and neutralizes another for each Na<sup>+</sup> that enters the cell. If HCO<sub>3</sub><sup>−</sup> is available, as is usually the case, this antiporter is the most important transporter regulating the cytosolic pH. The pH inside the cell regulates both exchangers; when the pH in the cytosol falls, both exchangers increase their activity.

A Na<sup>+</sup>-independent Cl<sup>−</sup>-HCO<sub>3</sub><sup>−</sup> exchanger adjusts the cytosolic pH in the reverse direction. Like the Na<sup>+</sup>-dependent transporters, pH regulates the Na<sup>+</sup>-independent Cl<sup>−</sup>-HCO<sub>3</sub><sup>−</sup> exchanger, but the exchanger's activity increases as the cytosol becomes too alkaline. The movement of HCO<sub>3</sub><sup>−</sup> in this case is normally out of the cell, down its electrochemical gradient, which decreases the pH of the

cytosol. A  $\text{Na}^+$ -independent  $\text{Cl}^-$ – $\text{HCO}_3^-$  exchanger in the membrane of red blood cells (called band 3 protein—see Figure 10–38) facilitates the quick discharge of  $\text{CO}_2$  (as  $\text{HCO}_3^-$ ) as the cells pass through capillaries in the lung.

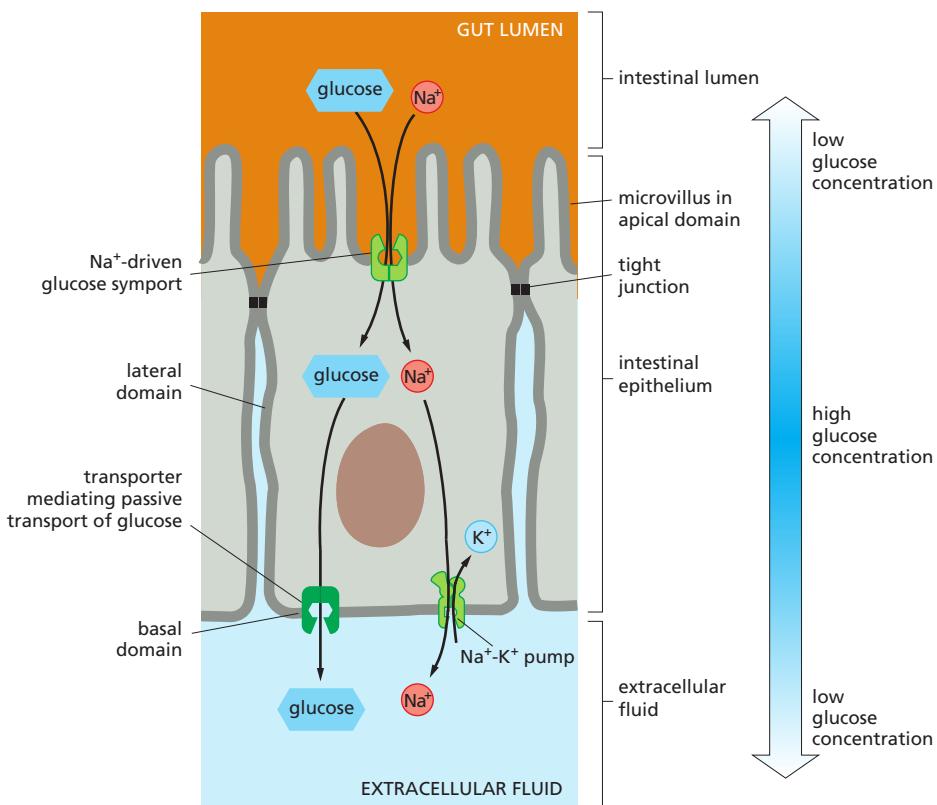
The intracellular pH is not entirely regulated by transporters in the plasma membrane: ATP-driven  $\text{H}^+$  pumps are used to control the pH of many intracellular compartments. As discussed in Chapter 13,  $\text{H}^+$  pumps maintain the low pH in lysosomes, as well as in endosomes and secretory vesicles. These  $\text{H}^+$  pumps use the energy of ATP hydrolysis to pump  $\text{H}^+$  into these organelles from the cytosol.

### An Asymmetric Distribution of Transporters in Epithelial Cells Underlies the Transcellular Transport of Solutes

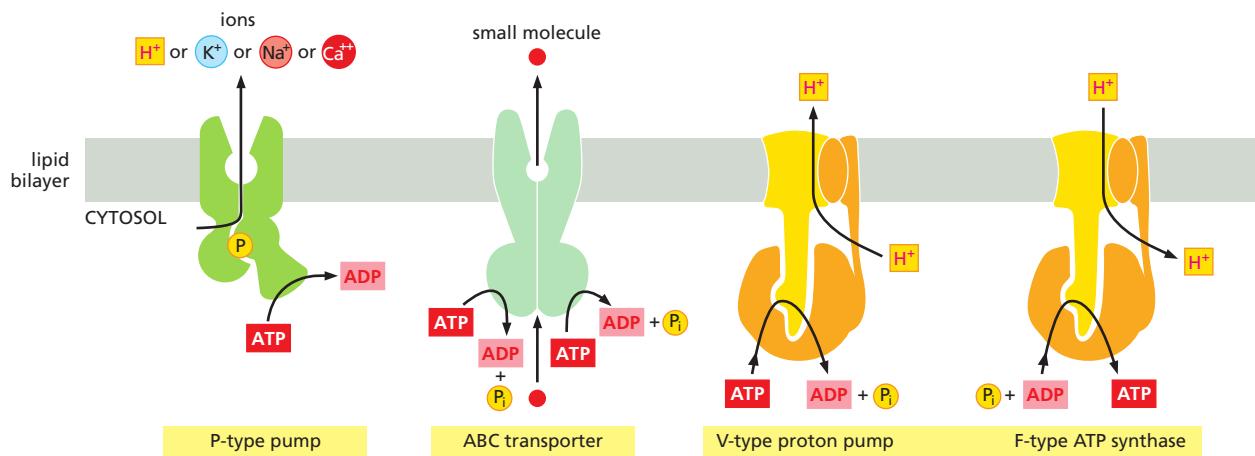
In epithelial cells, such as those that absorb nutrients from the gut, transporters are distributed nonuniformly in the plasma membrane and thereby contribute to the **transcellular transport** of absorbed solutes. By the actions of the transporters in these cells, solutes are moved across the epithelial cell layer into the extracellular fluid from where they pass into the blood. As shown in **Figure 11–11**,  $\text{Na}^+$ -linked symporters located in the apical (absorptive) domain of the plasma membrane actively transport nutrients into the cell, building up substantial concentration gradients for these solutes across the plasma membrane. Uniporters in the basal and lateral (basolateral) domains allow the nutrients to leave the cell passively down these concentration gradients.

In many of these epithelial cells, the plasma membrane area is greatly increased by the formation of thousands of microvilli, which extend as thin, fingerlike projections from the apical surface of each cell. Such microvilli can increase the total absorptive area of a cell as much as 25-fold, thereby enhancing its transport capabilities.

As we have seen, ion gradients have a crucial role in driving many essential transport processes in cells. Ion pumps that use the energy of ATP hydrolysis establish and maintain these gradients, as we discuss next.



**Figure 11–11** Transcellular transport. The transcellular transport of glucose across an intestinal epithelial cell depends on the nonuniform distribution of transporters in the cell's plasma membrane. The process shown here results in the transport of glucose from the intestinal lumen to the extracellular fluid (from where it passes into the blood). Glucose is pumped into the cell through the apical domain of the membrane by a  $\text{Na}^+$ -powered glucose symporter. Glucose passes out of the cell (down its concentration gradient) by passive movement through a glucose uniporter in the basal and lateral membrane domains. The  $\text{Na}^+$  gradient driving the glucose symport is maintained by the  $\text{Na}^+-\text{K}^+$  pump in the basal and lateral plasma membrane domains, which keeps the internal concentration of  $\text{Na}^+$  low (**Movie 11.2**). Adjacent cells are connected by impermeable tight junctions, which have a dual function in the transport process illustrated: they prevent solutes from crossing the epithelium between cells, allowing a concentration gradient of glucose to be maintained across the cell sheet (see Figure 19–18). They also serve as diffusion barriers (fences) within the plasma membrane, which help confine the various transporters to their respective membrane domains (see Figure 10–34).



### There Are Three Classes of ATP-Driven Pumps

ATP-driven pumps are often called *transport ATPases* because they hydrolyze ATP to ADP and phosphate and use the energy released to pump ions or other solutes across a membrane. There are three principal classes of ATP-driven pumps (Figure 11-12), and representatives of each are found in all prokaryotic and eukaryotic cells.

1. **P-type pumps** are structurally and functionally related multipass transmembrane proteins. They are called “P-type” because they phosphorylate themselves during the pumping cycle. This class includes many of the ion pumps that are responsible for setting up and maintaining gradients of  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{H}^+$ , and  $\text{Ca}^{2+}$  across cell membranes.
2. **ABC transporters** (ATP-Binding Cassette transporters) differ structurally from P-type ATPases and primarily pump small molecules across cell membranes.
3. **V-type pumps** are turbine-like protein machines, constructed from multiple different subunits. The V-type proton pump transfers  $\text{H}^+$  into organelles such as lysosomes, synaptic vesicles, and plant or yeast vacuoles ( $\text{V} = \text{vacuolar}$ ), to acidify the interior of these organelles (see Figure 13-37).

Structurally related to the V-type pumps is a distinct family of *F-type ATPases*, more commonly called *ATP synthases* because they normally work in reverse: instead of using ATP hydrolysis to drive  $\text{H}^+$  transport, they use the  $\text{H}^+$  gradient across the membrane to drive the synthesis of ATP from ADP and phosphate (see Figure 14-30). ATP synthases are found in the plasma membrane of bacteria, the inner membrane of mitochondria, and the thylakoid membrane of chloroplasts. The  $\text{H}^+$  gradient is generated either during the electron-transport steps of oxidative phosphorylation (in aerobic bacteria and mitochondria), during photosynthesis (in chloroplasts), or by the light-driven  $\text{H}^+$  pump (bacteriorhodopsin) in *Halobacterium*. We discuss some of these proteins in detail in Chapter 14.

For the remainder of this section, we focus on P-type pumps and ABC transporters.

### A P-type ATPase Pumps $\text{Ca}^{2+}$ into the Sarcoplasmic Reticulum in Muscle Cells

Eukaryotic cells maintain very low concentrations of free  $\text{Ca}^{2+}$  in their cytosol ( $\sim 10^{-7} \text{ M}$ ) in the face of a very much higher extracellular  $\text{Ca}^{2+}$  concentration ( $\sim 10^{-3} \text{ M}$ ). Therefore, even a small influx of  $\text{Ca}^{2+}$  significantly increases the concentration of free  $\text{Ca}^{2+}$  in the cytosol, and the flow of  $\text{Ca}^{2+}$  down its steep concentration gradient in response to extracellular signals is one means of transmitting these signals rapidly across the plasma membrane (discussed in Chapter 15). It is thus

**Figure 11-12** Three types of ATP-driven pumps. Like any enzyme, all ATP-driven pumps can work in either direction, depending on the electrochemical gradients of their solutes and the ATP/ADP ratio. When the ATP/ADP ratio is high, they hydrolyze ATP; when the ATP/ADP ratio is low, they can synthesize ATP. The F-type ATPase in mitochondria normally works in this “reverse” mode to make most of the cell’s ATP.

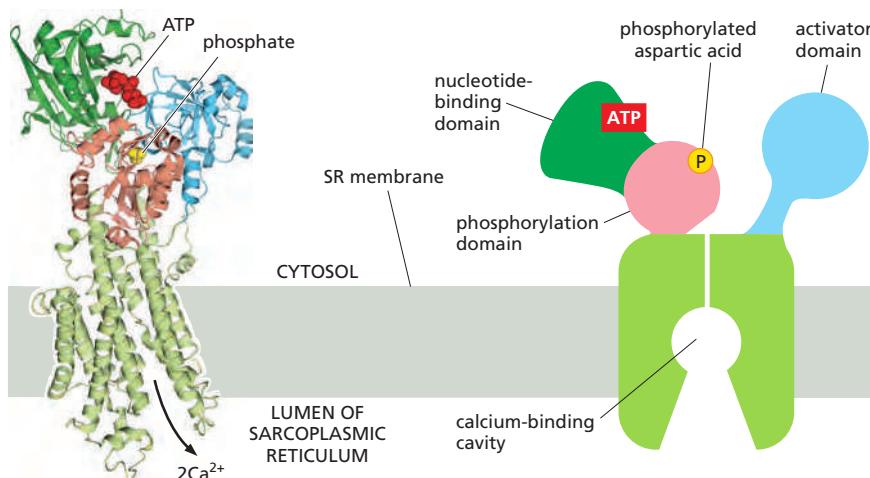
important that the cell maintains a steep  $\text{Ca}^{2+}$  gradient across its plasma membrane.  $\text{Ca}^{2+}$  transporters that actively pump  $\text{Ca}^{2+}$  out of the cell help maintain the gradient. One of these is a P-type  $\text{Ca}^{2+}$  ATPase; the other is an antiporter (called a  $\text{Na}^+-\text{Ca}^{2+}$  exchanger) that is driven by the  $\text{Na}^+$  electrochemical gradient (discussed in Chapter 15).

The  $\text{Ca}^{2+}$  pump, or  $\text{Ca}^{2+}$  ATPase, in the *sarcoplasmic reticulum* (SR) membrane of skeletal muscle cells is a well-understood P-type transport ATPase. The SR is a specialized type of endoplasmic reticulum that forms a network of tubular sacs in the muscle cell cytoplasm, and it serves as an intracellular store of  $\text{Ca}^{2+}$ . When an action potential depolarizes the muscle cell plasma membrane,  $\text{Ca}^{2+}$  is released into the cytosol from the SR through  $\text{Ca}^{2+}$ -release channels, stimulating the muscle to contract (discussed in Chapters 15 and 16). The  $\text{Ca}^{2+}$  pump, which accounts for about 90% of the membrane protein of the SR, moves  $\text{Ca}^{2+}$  from the cytosol back into the SR. The endoplasmic reticulum of nonmuscle cells contains a similar  $\text{Ca}^{2+}$  pump, but in smaller quantities.

Enzymatic studies and analyses of the three-dimensional structures of transport intermediates of the SR  $\text{Ca}^{2+}$  pump and related pumps have revealed the molecular mechanism of P-type transport ATPases in great detail. They all have similar structures, containing 10 transmembrane  $\alpha$  helices connected to three cytosolic domains (Figure 11–13). In the  $\text{Ca}^{2+}$  pump, amino acid side chains protruding from the transmembrane helices form two centrally positioned binding sites for  $\text{Ca}^{2+}$ . As shown in Figure 11–14, in the pump's ATP-bound nonphosphorylated state, these binding sites are accessible only from the cytosolic side of the SR membrane.  $\text{Ca}^{2+}$  binding triggers a series of conformational changes that close the passageway to the cytosol and activate a phosphotransfer reaction in which the terminal phosphate of the ATP is transferred to an aspartate that is highly conserved among all P-type ATPases. The ADP then dissociates and is replaced with a fresh ATP, causing another conformational change that opens a passageway to the SR lumen through which the two  $\text{Ca}^{2+}$  ions exit. They are replaced by two  $\text{H}^+$  ions and a water molecule that stabilize the empty  $\text{Ca}^{2+}$ -binding sites and close the passageway to the SR lumen. Hydrolysis of the labile phosphoryl-aspartate bond returns the pump to the initial conformation, and the cycle starts again. The transient self-phosphorylation of the pump during its cycle is an essential characteristic of all P-type pumps.

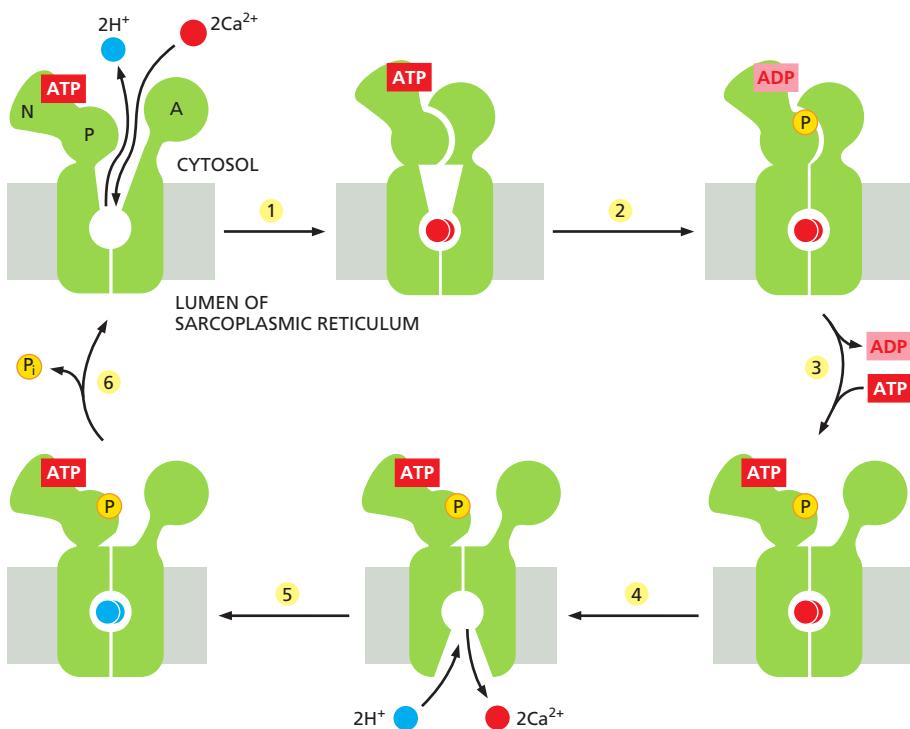
### The Plasma Membrane $\text{Na}^+-\text{K}^+$ Pump Establishes $\text{Na}^+$ and $\text{K}^+$ Gradients Across the Plasma Membrane

The concentration of  $\text{K}^+$  is typically 10–30 times higher inside cells than outside, whereas the reverse is true of  $\text{Na}^+$  (see Table 11–1, p. 598). A  $\text{Na}^+-\text{K}^+$  pump, or  $\text{Na}^+-\text{K}^+$  ATPase, found in the plasma membrane of virtually all animal cells maintains



**Figure 11–13** The structure of the sarcoplasmic reticulum  $\text{Ca}^{2+}$  pump.

The ribbon model (left), derived from x-ray crystallographic analyses, shows the pump in its phosphorylated, ATP-bound state. The three globular cytosolic domains of the pump—the nucleotide-binding domain (dark green), the activator domain (blue), and the phosphorylation domain (red), also shown schematically on the right—change conformation dramatically during the pumping cycle. These changes in turn alter the arrangement of the transmembrane helices, which allows the  $\text{Ca}^{2+}$  to be released from its binding cavity into the SR lumen (Movie 11.3). (PDB code: 3B9B.)

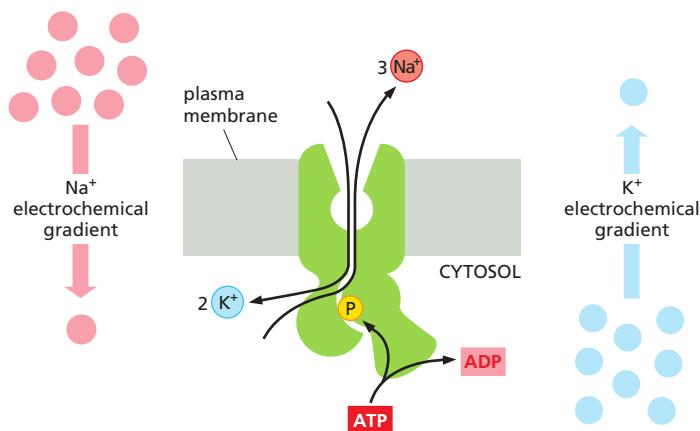


**Figure 11–14** The pumping cycle of the sarcoplasmic reticulum  $\text{Ca}^{2+}$  pump. Ion pumping proceeds by a series of stepwise conformational changes in which movements of the pump's three cytosolic domains [the nucleotide-binding domain (N), the phosphorylation domain (P), and the activator domain (A)] are mechanically coupled to movements of the transmembrane  $\alpha$  helices. Helix movement opens and closes passageways through which  $\text{Ca}^{2+}$  enters from the cytosol and binds to the two centrally located  $\text{Ca}^{2+}$  binding sites. The two  $\text{Ca}^{2+}$  then exit into the SR lumen and are replaced by two  $\text{H}^+$ , which are transported in the opposite direction. The  $\text{Ca}^{2+}$ -dependent phosphorylation and  $\text{H}^+$ -dependent dephosphorylation of aspartic acid are universally conserved steps in the reaction cycle of all P-type pumps: they cause the conformational transitions to occur in an orderly manner, enabling the proteins to do useful work. (Adapted from C. Toyoshima et al., *Nature* 432:361–368, 2004 and J.V. Möller et al., *Q. Rev. Biophys.* 43:501–566, 2010.)

these concentration differences. Like the  $\text{Ca}^{2+}$  pump, the  $\text{Na}^+-\text{K}^+$  pump belongs to the family of P-type ATPases and operates as an ATP-driven antiporter, actively pumping  $\text{Na}^+$  out of the cell against its steep electrochemical gradient and pumping  $\text{K}^+$  in (Figure 11–15).

We mentioned earlier that the  $\text{Na}^+$  gradient produced by the  $\text{Na}^+-\text{K}^+$  pump drives the transport of most nutrients into animal cells and also has a crucial role in regulating cytosolic pH. A typical animal cell devotes almost one-third of its energy to fueling this pump, and the pump consumes even more energy in nerve cells and in cells that are dedicated to transport processes, such as those forming kidney tubules.

Since the  $\text{Na}^+-\text{K}^+$  pump drives three positively charged ions out of the cell for every two it pumps in, it is *electrogenic*: it drives a net electric current across the membrane, tending to create an electrical potential, with the cell's inside being negative relative to the outside. This electrogenic effect of the pump, however, seldom directly contributes more than 10% to the membrane potential. The remaining 90%, as we discuss later, depends only indirectly on the  $\text{Na}^+-\text{K}^+$  pump.



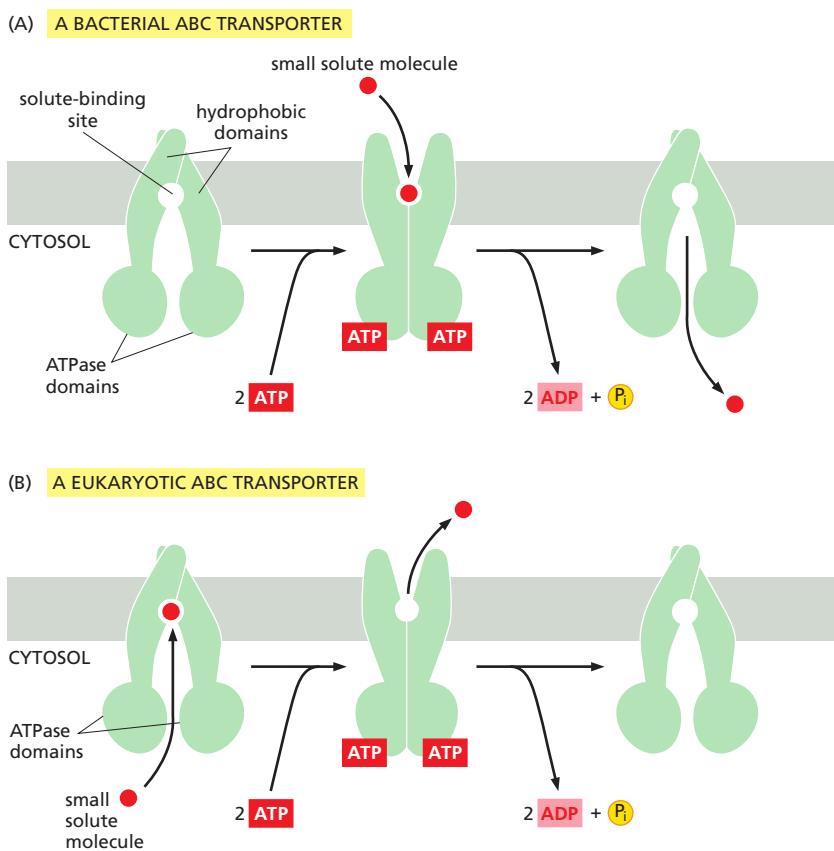
**Figure 11–15** The function of the  $\text{Na}^+-\text{K}^+$  pump. This P-type ATPase actively pumps  $\text{Na}^+$  out of and  $\text{K}^+$  into a cell against their electrochemical gradients. It is structurally closely related to the  $\text{Ca}^{2+}$  ATPase but differs in its selectivity for ions: for every molecule of ATP hydrolyzed by the pump, three  $\text{Na}^+$  are pumped out and two  $\text{K}^+$  are pumped in. As in the  $\text{Ca}^{2+}$  pump, an aspartate is phosphorylated and dephosphorylated during the pumping cycle (Movie 11.4).

## ABC Transporters Constitute the Largest Family of Membrane Transport Proteins

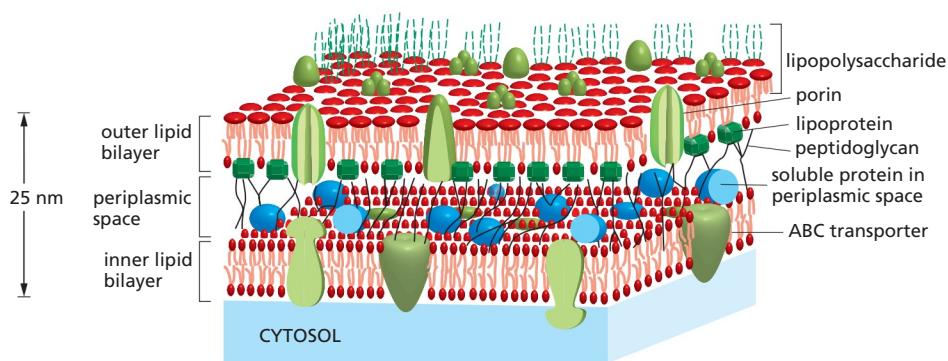
The last type of transport ATPase that we discuss is the family of the ABC transporters, so named because each member contains two highly conserved ATPase domains, or ATP-Binding “Cassettes,” on the cytosolic side of the membrane. ATP binding brings together the two ATPase domains, and ATP hydrolysis leads to their dissociation (Figure 11–16). These movements of the cytosolic domains are transmitted to the transmembrane segments, driving cycles of conformational changes that alternately expose solute-binding sites on one side of the membrane and then on the other, as we have seen for other transporters. In this way, ABC transporters harvest the energy released upon ATP binding and hydrolysis to drive transport of solutes across the bilayer. The transport is directional toward inside or toward outside, depending on the particular conformational change in the solute binding site that is linked to ATP hydrolysis (see Figure 11–16).

ABC transporters constitute the largest family of membrane transport proteins and are of great clinical importance. The first of these proteins to be characterized was found in bacteria. We have already mentioned that the plasma membranes of all bacteria contain transporters that use the H<sup>+</sup> gradient across the membrane to actively transport a variety of nutrients into the cell. In addition, bacteria use ABC transporters to import certain small molecules. In bacteria such as *E. coli* that have double membranes (Figure 11–17), the ABC transporters are located in the inner membrane, and an auxiliary mechanism operates to capture the nutrients and deliver them to the transporters (Figure 11–18).

In *E. coli*, 78 genes (an amazing 5% of the bacterium’s genes) encode ABC transporters, and animal genomes encode an even larger number. Although each transporter is thought to be specific for a particular molecule or class of molecules, the variety of substrates transported by this superfamily is great and includes inorganic ions, amino acids, mono- and polysaccharides, peptides, lipids, drugs, and, in some cases, even proteins that can be larger than the transporter itself.



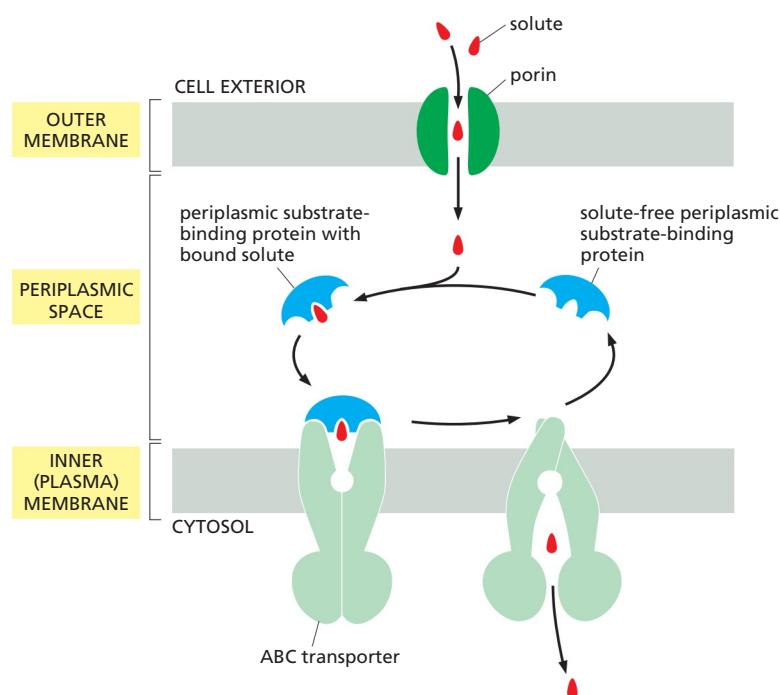
**Figure 11–16** Small-molecule transport by typical ABC transporters. ABC transporters consist of multiple domains. Typically, two hydrophobic domains, each built of six membrane-spanning  $\alpha$  helices, together form the translocation pathway and provide substrate specificity. Two ATPase domains protrude into the cytosol. In some cases, the two halves of the transporter are formed by a single polypeptide, whereas in other cases they are formed by two or more separate polypeptides that assemble into a similar structure. Without ATP bound, the transporter exposes a substrate-binding site on one side of the membrane. ATP binding induces a conformational change that exposes the substrate-binding site on the opposite side; ATP hydrolysis followed by ADP dissociation returns the transporter to its original conformation. Most individual ABC transporters are unidirectional. (A) Both importing and exporting ABC transporters are found in bacteria; an ABC importer is shown in this cartoon. The crystal structure of a bacterial ABC transporter is shown in Figure 3–76. (B) In eukaryotes, most ABC transporters export substances—either from the cytosol to the extracellular space or from the cytosol to a membrane-bound intracellular compartment such as the endoplasmic reticulum—or from the mitochondrial matrix to the cytosol.



**Figure 11–17** A small section of the double membrane of an *E. coli* bacterium. The inner membrane is the cell's plasma membrane. Between the inner and outer membranes is a highly porous, rigid peptidoglycan layer, composed of protein and polysaccharide that constitute the bacterial cell wall. It is attached to lipoprotein molecules in the outer membrane and fills the *periplasmic space* (only a little of the peptidoglycan layer is shown). This space also contains a variety of soluble protein molecules. The dashed threads (shown in green) at the top represent the polysaccharide chains of the special lipopolysaccharide molecules that form the external monolayer of the outer membrane; for clarity, only a few of these chains are shown. Bacteria with double membranes are called *Gram-negative* because they do not retain the dark blue dye used in Gram staining. Bacteria with single membranes (but thicker peptidoglycan cell walls), such as staphylococci and streptococci, retain the blue dye and are therefore called *Gram-positive*; their single membrane is analogous to the inner (plasma) membrane of *Gram-negative* bacteria.

The first eukaryotic ABC transporters identified were discovered because of their ability to pump hydrophobic drugs out of the cytosol. One of these transporters is the **multidrug resistance (MDR) protein**, also called P-glycoprotein. It is present at elevated levels in many human cancer cells and makes the cells simultaneously resistant to a variety of chemically unrelated cytotoxic drugs that are widely used in cancer chemotherapy. Treatment with any one of these drugs can result in the selective survival and overgrowth of those cancer cells that express an especially large amount of the MDR transporter. These cells pump drugs out of the cell very efficiently and are therefore relatively resistant to the drugs' toxic effects ([Movie 11.5](#)). Selection for cancer cells with resistance to one drug can thereby lead to resistance to a wide variety of anticancer drugs. Some studies indicate that up to 40% of human cancers develop multidrug resistance, making it a major hurdle in the battle against cancer.

A related and equally sinister phenomenon occurs in the protist *Plasmodium falciparum*, which causes malaria. More than 200 million people are infected worldwide with this parasite, which remains a major cause of human death, killing almost a million people every year. The development of resistance to the antimalarial drug *chloroquine* has hampered the control of malaria. The resistant *P. falciparum* have amplified a gene encoding an ABC transporter that pumps out the chloroquine.



**Figure 11–18** The auxiliary transport system associated with transport ATPases in bacteria with double membranes. The solute diffuses through channel proteins (porins) in the outer membrane and binds to a *periplasmic substrate-binding protein* that delivers it to the ABC transporter, which pumps it across the plasma membrane. The peptidoglycan is omitted for simplicity; its porous structure allows the substrate-binding proteins and water-soluble solutes to move through it by diffusion.

In most vertebrate cells, an ABC transporter in the endoplasmic reticulum (ER) membrane (named *transporter associated with antigen processing*, or *TAP transporter*) actively pumps a wide variety of peptides from the cytosol into the ER lumen. These peptides are produced by protein degradation in proteasomes (discussed in Chapter 6). They are carried from the ER to the cell surface, where they are displayed for scrutiny by cytotoxic T lymphocytes, which kill the cell if the peptides are derived from a virus or other microorganism lurking in the cytosol of an infected cell (discussed in Chapter 24).

Yet another member of the ABC transporter family is the *cystic fibrosis transmembrane conductance regulator* protein (CFTR), which was discovered through studies of the common genetic disease *cystic fibrosis*. This disease is caused by a mutation in the gene encoding CFTR, a Cl<sup>-</sup> transport protein in the plasma membrane of epithelial cells. CFTR regulates ion concentrations in the extracellular fluid, especially in the lung. One in 27 Caucasians carries a gene encoding a mutant form of this protein; in 1 in 2900, both copies of the gene are mutated, causing the disease. In contrast to other ABC transporters, ATP binding and hydrolysis in the CFTR protein do not drive the transport process. Instead, they control the opening and closing of a continuous channel, which provides a passive conduit for Cl<sup>-</sup> to move down its electrochemical gradient. Thus, some ABC proteins can function as transporters and others as gated channels.

### Summary

*Transporters bind specific solutes and transfer them across the lipid bilayer by undergoing conformational changes that alternately expose the solute-binding site on one side of the membrane and then on the other. Some transporters move a single solute “downhill,” whereas others can act as pumps to move a solute “uphill” against its electrochemical gradient, using energy provided by ATP hydrolysis, by a downhill flow of another solute (such as Na<sup>+</sup> or H<sup>+</sup>), or by light to drive the requisite series of conformational changes in an orderly manner. Transporters belong to a small number of protein families. Each family evolved from a common ancestral protein, and its members all operate by a similar mechanism. The family of P-type transport ATPases, which includes Ca<sup>2+</sup> and Na<sup>+</sup>-K<sup>+</sup> pumps, is an important example; each of these ATPases sequentially phosphorylates and dephosphorylates itself during the pumping cycle. The superfamily of ABC transporters is the largest family of membrane transport proteins and is especially important clinically. It includes proteins that are responsible for cystic fibrosis, for drug resistance in both cancer cells and malaria-causing parasites, and for pumping pathogen-derived peptides into the ER for cytotoxic lymphocytes to reorganize on the surface of infected cells.*

## CHANNELS AND THE ELECTRICAL PROPERTIES OF MEMBRANES

Unlike transporters, channels form pores across membranes. One class of channel proteins found in virtually all animals forms *gap junctions* between adjacent cells; each plasma membrane contributes equally to the formation of the channel, which connects the cytoplasm of the two cells. These channels are discussed in Chapter 19 and will not be considered further here. Both gap junctions and *porins*, the channels in the outer membranes of bacteria, mitochondria, and chloroplasts (discussed in Chapter 10), have relatively large and permissive pores, and it would be disastrous if they directly connected the inside of a cell to an extracellular space. Indeed, many bacterial toxins do exactly that to kill other cells (discussed in Chapter 24).

In contrast, most channels in the plasma membrane of animal and plant cells that connect the cytosol to the cell exterior necessarily have narrow, highly selective pores that can open and close rapidly. Because these proteins are concerned specifically with inorganic ion transport, they are referred to as **ion channels**. For transport efficiency, ion channels have an advantage over transporters, in that

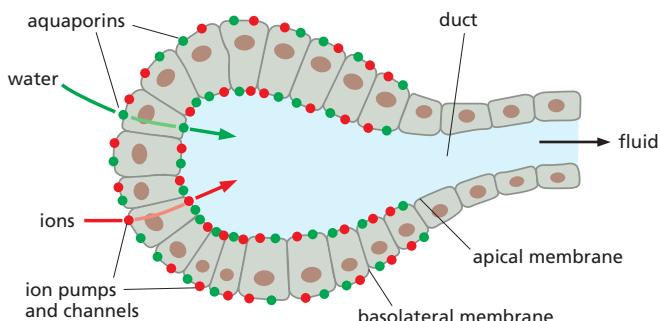
they can pass up to 100 million ions through one open channel each second—a rate  $10^5$  times greater than the fastest rate of transport mediated by any known transporter. As discussed earlier, however, channels cannot be coupled to an energy source to perform active transport, so the transport they mediate is always passive (downhill). Thus, the function of ion channels is to allow specific inorganic ions—primarily  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ , or  $\text{Cl}^-$ —to diffuse rapidly down their electrochemical gradients across the lipid bilayer. In this section, we will see that the ability to control ion fluxes through these channels is essential for many cell functions. Nerve cells (neurons), in particular, have made a specialty of using ion channels, and we will consider how they use many different ion channels to receive, conduct, and transmit signals. Before we discuss ion channels, however, we briefly consider the aquaporin water channels that we mentioned earlier.

### Aquaporins Are Permeable to Water But Impermeable to Ions

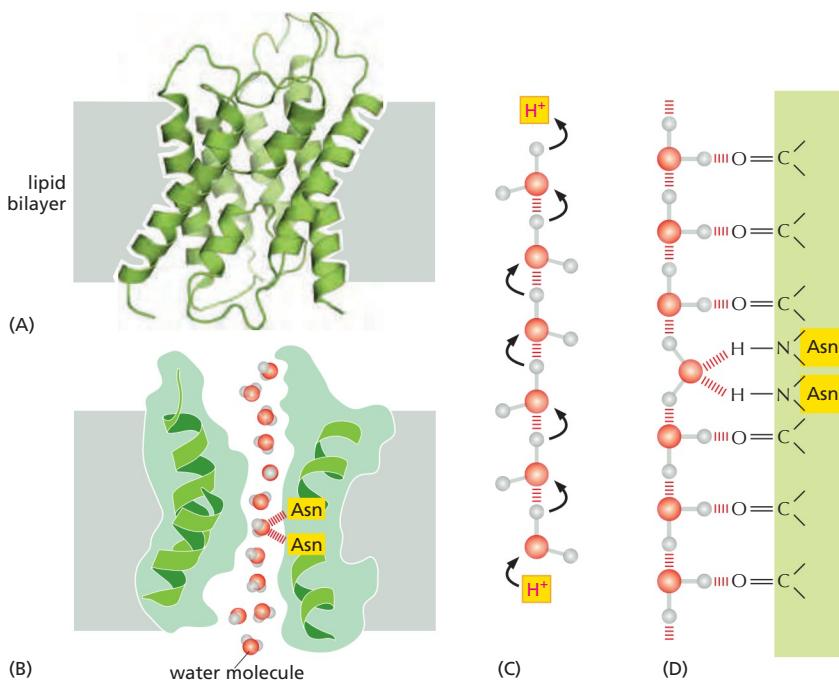
Because cells are mostly water (typically ~70% by weight), water movement across cell membranes is fundamentally important for life. Cells also contain a high concentration of solutes, including numerous negatively charged organic molecules that are confined inside the cell (the so-called *fixed anions*) and their accompanying cations that are required for charge balance. This creates an osmotic gradient, which mostly is balanced by an opposite osmotic gradient due to a high concentration of inorganic ions—chiefly  $\text{Na}^+$  and  $\text{Cl}^-$ —in the extracellular fluid. The small remaining osmotic force tends to “pull” water into the cell, causing it to swell until the forces are balanced. Because all biological membranes are moderately permeable to water (see Figure 11–2), cell volume equilibrates in minutes or less in response to an osmotic gradient. For most animal cells, however, osmosis has only a minor role in regulating cell volume. This is because most of the cytoplasm is in a gel-like state and resists large changes in its volume in response to changes in osmolarity.

In addition to the direct diffusion of water across the lipid bilayer, some prokaryotic and eukaryotic cells have **water channels**, or **aquaporins**, embedded in their plasma membrane to allow water to move more rapidly. Aquaporins are particularly abundant in animal cells that must transport water at high rates, such as the epithelial cells of the kidney or exocrine cells that must transport or secrete large volumes of fluids, respectively (Figure 11–19).

Aquaporins must solve a problem that is opposite to that facing ion channels. To avoid disrupting ion gradients across membranes, they have to allow the rapid passage of water molecules while completely blocking the passage of ions. The three-dimensional structure of an aquaporin reveals how it achieves this remarkable selectivity. The channels have a narrow pore that allows water molecules to traverse the membrane in single file, following the path of carbonyl oxygens that line one side of the pore (Figure 11–20A and B). Hydrophobic amino acids line the other side of the pore. The pore is too narrow for any hydrated ion to enter, and the energy cost of dehydrating an ion would be enormous because the hydrophobic wall of the pore cannot interact with a dehydrated ion to compensate for the loss of water. This design readily explains why the aquaporins cannot conduct  $\text{K}^+$ ,



**Figure 11–19** The role of aquaporins in fluid secretion. Cells lining the ducts of exocrine glands (as found, for example, in the pancreas and liver, and in mammary, sweat, and salivary glands) secrete large volumes of body fluids. These cells are organized into epithelial sheets in which their apical plasma membrane faces the lumen of the duct. Ion pumps and channels situated in the basolateral and apical plasma membrane move ions (mostly  $\text{Na}^+$  and  $\text{Cl}^-$ ) into the ductal lumen, creating an osmotic gradient between the surrounding tissue and the duct. Water molecules rapidly follow the osmotic gradient through aquaporins that are present in high concentrations in both the apical and basolateral membranes.

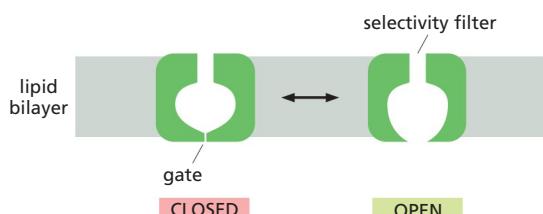


$\text{Na}^+$ ,  $\text{Ca}^{2+}$ , or  $\text{Cl}^-$  ions. These channels are also impermeable to  $\text{H}^+$ , which is mainly present in cells as  $\text{H}_3\text{O}^+$ . These hydronium ions diffuse through water extremely rapidly, using a molecular relay mechanism that requires the making and breaking of hydrogen bonds between adjacent water molecules (Figure 11–20C). Aquaporins contain two strategically placed asparagines, which bind to the oxygen atom of the central water molecule in the line of water molecules traversing the pore, imposing a bipolarity on the entire column of water molecules (Figure 11–20C and D). This makes it impossible for the “making and breaking” sequence of hydrogen bonds (shown in Figure 11–20C) to get past the central asparagine-bonded water molecule. Because both valences of this central oxygen are unavailable for hydrogen-bonding, the central water molecule cannot participate in an  $\text{H}^+$  relay, and the pore is therefore impermeable to  $\text{H}^+$ .

We now turn to ion channels, the subject of the rest of the chapter.

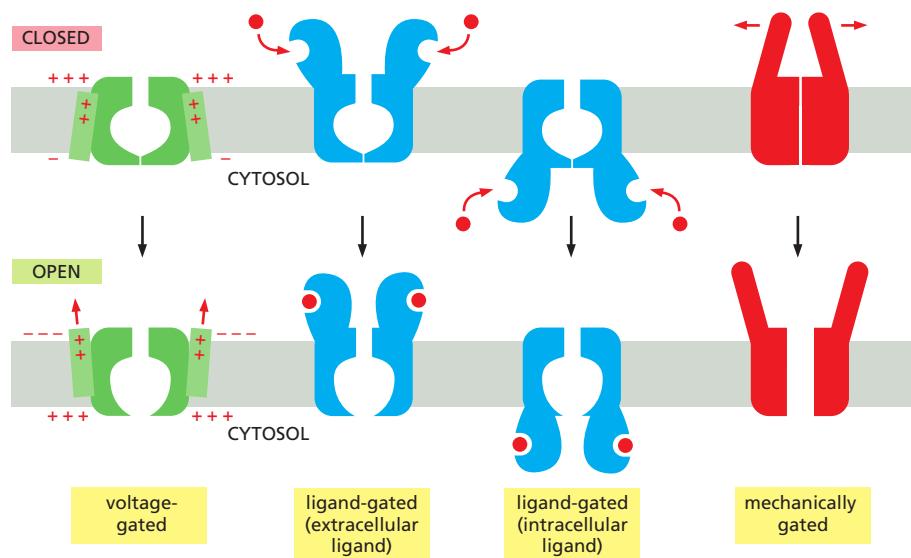
### Ion Channels Are Ion-Selective and Fluctuate Between Open and Closed States

Two important properties distinguish ion channels from aqueous pores. First, they show *ion selectivity*, permitting some inorganic ions to pass, but not others. This suggests that their pores must be narrow enough in places to force permeating ions into intimate contact with the walls of the channel so that only ions of appropriate size and charge can pass. The permeating ions have to shed most or all of their associated water molecules to pass, often in single file, through the narrowest part of the channel, which is called the *selectivity filter*; this limits their rate of passage (Figure 11–21). Thus, as the ion concentration increases, the flux of the ion through a channel increases proportionally but then levels off (saturates) at a maximum rate.



**Figure 11–20 The structure of aquaporins.** (A) A ribbon diagram of an aquaporin monomer. In the membrane, aquaporins form tetramers, with each monomer containing an aqueous pore in its center (not shown). Each individual aquaporin channel passes about  $10^9$  water molecules per second. (B) A longitudinal cross section through one aquaporin monomer, in the plane of the central pore. One face of the pore is lined with hydrophilic amino acids, which provide transient hydrogen bonds to water molecules; these bonds help line up the transiting water molecules in a single row and orient them as they traverse the pore. (C and D) A model explaining why aquaporins are impermeable to  $\text{H}^+$ . (C) In water,  $\text{H}^+$  diffuses extremely rapidly by being relayed from one water molecule to the next. (D) Carbonyl groups ( $\text{C}=\text{O}$ ) lining the hydrophilic face of the pore align water molecules, and two strategically placed asparagines in the center help tether a central water molecule such that both valences on its oxygen are occupied. This arrangement bipolarizes the entire line of water molecules, with each water molecule acting as a hydrogen-bond acceptor from its inner neighbor (Movie 11.6). (A and B, adapted from R.M. Stroud et al., *Curr. Opin. Struct. Biol.* 13:424–431, 2003. With permission from Elsevier.)

**Figure 11–21 A typical ion channel, which fluctuates between closed and open conformations.** The ion channel shown here in cross section forms a pore across the lipid bilayer only in the “open” conformational state. The pore narrows to atomic dimensions in one region (the selectivity filter), where the ion selectivity of the channel is largely determined. Another region of the channel forms the gate.



**Figure 11–22** The gating of ion channels. This schematic drawing shows several kinds of stimuli that open ion channels. Mechanically gated channels often have cytoplasmic extensions (not shown) that link the channel to the cytoskeleton.

The second important distinction between ion channels and aqueous pores is that ion channels are not continuously open. Instead, they are *gated*, which allows them to open briefly and then close again. Moreover, with prolonged (chemical or electrical) stimulation, most ion channels go into a closed “desensitized,” or “inactivated,” state, in which they are refractory to further opening until the stimulus has been removed, as we discuss later. In most cases, the gate opens in response to a specific stimulus. As shown in Figure 11–22, the main types of stimuli that are known to cause ion channels to open are a change in the voltage across the membrane (*voltage-gated channels*), a mechanical stress (*mechanically gated channels*), or the binding of a ligand (*ligand-gated channels*). The ligand can be either an extracellular mediator—specifically, a neurotransmitter (*transmitter-gated channels*)—or an intracellular mediator such as an ion (*ion-gated channels*) or a nucleotide (*nucleotide-gated channels*). In addition, protein phosphorylation and dephosphorylation regulates the activity of many ion channels; this type of channel regulation is discussed, together with nucleotide-gated ion channels, in Chapter 15.

More than 100 types of ion channels have been identified thus far, and new ones are still being discovered, each characterized by the ions it conducts, the mechanism by which it is gated, and its abundance and localization in the cell and in specific cells. Ion channels are responsible for the electrical excitability of muscle cells, and they mediate most forms of electrical signaling in the nervous system. A single neuron typically contains 10 or more kinds of ion channels, located in different domains of its plasma membrane. But ion channels are not restricted to electrically excitable cells. They are present in all animal cells and are found in plant cells and microorganisms: they propagate the leaf-closing response of the mimosa plant, for example (Movie 11.7), and allow the single-celled *Paramecium* to reverse direction after a collision.

Ion channels that are permeable mainly to  $K^+$  are found in the plasma membrane of almost all cells. An important subset of  $K^+$  channels opens even in an unstimulated or “resting” cell, and hence these are called  **$K^+$  leak channels**. Although this term applies to many different  $K^+$  channels, depending on the cell type, they serve a common purpose: by making the plasma membrane much more permeable to  $K^+$  than to other ions, they have a crucial role in maintaining the membrane potential across all plasma membranes, as we discuss next.

## The Membrane Potential in Animal Cells Depends Mainly on K<sup>+</sup> Leak Channels and the K<sup>+</sup> Gradient Across the Plasma Membrane

A **membrane potential** arises when there is a difference in the electrical charge on the two sides of a membrane, due to a slight excess of positive ions over negative ones on one side and a slight deficit on the other. Such charge differences can result both from active electrogenic pumping (see p. 608) and from passive ion diffusion. As we discuss in Chapter 14, electrogenic H<sup>+</sup> pumps in the mitochondrial inner membrane generate most of the membrane potential across this membrane. Electrogenic pumps also generate most of the electrical potential across the plasma membrane in plants and fungi. In typical animal cells, however, passive ion movements make the largest contribution to the electrical potential across the plasma membrane.

As explained earlier, due to the action of the Na<sup>+</sup>-K<sup>+</sup> pump, there is little Na<sup>+</sup> inside the cell, and other intracellular inorganic cations have to be plentiful enough to balance the charge carried by the cell's fixed anions—the negatively charged organic molecules that are confined inside the cell. This balancing role is performed largely by K<sup>+</sup>, which is actively pumped into the cell by the Na<sup>+</sup>-K<sup>+</sup> pump and can also move freely in or out through the *K<sup>+</sup> leak channels* in the plasma membrane. Because of the presence of these channels, K<sup>+</sup> comes almost to equilibrium, where an electrical force exerted by an excess of negative charges attracting K<sup>+</sup> into the cell balances the tendency of K<sup>+</sup> to leak out down its concentration gradient. The membrane potential (of the plasma membrane) is the manifestation of this electrical force, and we can calculate its equilibrium value from the steepness of the K<sup>+</sup> concentration gradient. The following argument may help to make this clear.

Suppose that initially there is no voltage gradient across the plasma membrane (the membrane potential is zero) but the concentration of K<sup>+</sup> is high inside the cell and low outside. K<sup>+</sup> will tend to leave the cell through the K<sup>+</sup> leak channels, driven by its concentration gradient. As K<sup>+</sup> begins to move out, each ion leaves behind an unbalanced negative charge, thereby creating an electrical field, or membrane potential, which will tend to oppose the further efflux of K<sup>+</sup>. The net efflux of K<sup>+</sup> halts when the membrane potential reaches a value at which this electrical driving force on K<sup>+</sup> exactly balances the effect of its concentration gradient—that is, when the electrochemical gradient for K<sup>+</sup> is zero. Although Cl<sup>-</sup> ions also equilibrate across the membrane, the membrane potential keeps most of these ions out of the cell because their charge is negative.

The equilibrium condition, in which there is no net flow of ions across the plasma membrane, defines the **resting membrane potential** for this idealized cell. A simple but very important formula, the **Nernst equation**, quantifies the equilibrium condition and, as explained in **Panel 11-1**, makes it possible to calculate the theoretical resting membrane potential if we know the ratio of internal and external ion concentrations. As the plasma membrane of a real cell is not exclusively permeable to K<sup>+</sup> and Cl<sup>-</sup>, however, the actual resting membrane potential is usually not exactly equal to that predicted by the Nernst equation for K<sup>+</sup> or Cl<sup>-</sup>.

## The Resting Potential Decays Only Slowly When the Na<sup>+</sup>-K<sup>+</sup> Pump Is Stopped

Movement of only a minute number of inorganic ions across the plasma membrane through ion channels suffices to set up the membrane potential. Thus, we can think of the membrane potential as arising from movements of charge that leave ion *concentrations* practically unaffected and result in only a very slight discrepancy in the number of positive and negative ions on the two sides of the membrane (**Figure 11-23**). Moreover, these movements of charge are generally rapid, taking only a few milliseconds or less.

Consider the change in the membrane potential in a real cell after the sudden inactivation of the Na<sup>+</sup>-K<sup>+</sup> pump. A slight drop in the membrane potential occurs immediately. This is because the pump is electrogenic and, when active, makes a

## THE NERNST EQUATION AND ION FLOW

The flow of any inorganic ion through a membrane channel is driven by the **electrochemical gradient** for that ion. This gradient represents the combination of two influences: the voltage gradient and the concentration gradient of the ion across the membrane. When these two influences just balance each other, the electrochemical gradient for the ion is zero, and there is no *net* flow of the ion through the channel. The voltage gradient (membrane potential) at which this equilibrium is reached is called the **equilibrium potential** for the ion. It can be calculated from an equation that will be derived below, called the **Nernst equation**.

The **Nernst equation** is

$$V = \frac{RT}{zF} \ln \frac{C_o}{C_i}$$

where

**V** = the equilibrium potential in volts (internal potential minus external potential)

**C<sub>o</sub>** and **C<sub>i</sub>** = outside and inside concentrations of the ion, respectively

**R** = the gas constant ( $8.3 \text{ J mol}^{-1} \text{ K}^{-1}$ )

**T** = the absolute temperature (K)

**F** = Faraday's constant ( $9.6 \times 10^4 \text{ J V}^{-1} \text{ mol}^{-1}$ )

**z** = the valence (charge) of the ion

**In** = logarithm to the base e

The Nernst equation is derived as follows:

A molecule in solution (a solute) tends to move from a region of high concentration to a region of low concentration simply due to the random movement of molecules, which results in their equilibrium.

Consequently, movement down a concentration gradient is accompanied by a favorable free-energy change ( $\Delta G < 0$ ), whereas movement up a concentration gradient is accompanied by an unfavorable free-energy change ( $\Delta G > 0$ ). (Free energy is introduced in Chapter 2 and discussed in the context of redox reactions in Panel 14-1, p. 765.)

The free-energy change per mole of solute moved across the plasma membrane ( $\Delta G_{\text{conc}}$ ) is equal to  $-RT \ln C_o / C_i$ .

If the solute is an ion, moving it into a cell across a membrane whose inside is at a voltage  $V$  relative to the outside will cause an additional free-energy change (per mole of solute moved) of  $\Delta G_{\text{volt}} = zFV$ .

At the point where the concentration and voltage gradients just balance,

$$\Delta G_{\text{conc}} + \Delta G_{\text{volt}} = 0$$

and the ion distribution is at equilibrium across the membrane.

Thus,

$$zFV - RT \ln \frac{C_o}{C_i} = 0$$

and, therefore,

$$V = \frac{RT}{zF} \ln \frac{C_o}{C_i}$$

or, using the constant that converts natural logarithms to base 10,

$$V = 2.3 \frac{RT}{zF} \log_{10} \frac{C_o}{C_i}$$

For a univalent cation,

$$2.3 \frac{RT}{F} = 58 \text{ mV at } 20^\circ\text{C} \text{ and } 61.5 \text{ mV at } 37^\circ\text{C}.$$

Thus, for such an ion at  $37^\circ\text{C}$ ,

$$V = +61.5 \text{ mV for } C_o / C_i = 10,$$

whereas

$$V = 0 \text{ for } C_o / C_i = 1.$$

The  $\text{K}^+$  equilibrium potential ( $V_K$ ), for example, is

$$61.5 \log_{10}([K^+]_o / [K^+]_i) \text{ millivolts}$$

( $-89 \text{ mV}$  for a typical cell, where  $[K^+]_o = 5 \text{ mM}$  and  $[K^+]_i = 140 \text{ mM}$ ).

At  $V_K$ , there is no net flow of  $\text{K}^+$  across the membrane.

Similarly, when the membrane potential has a value of

$$61.5 \log_{10}([Na^+]_o / [Na^+]_i),$$

the  $\text{Na}^+$  equilibrium potential ( $V_{\text{Na}}$ ),

there is no net flow of  $\text{Na}^+$ .

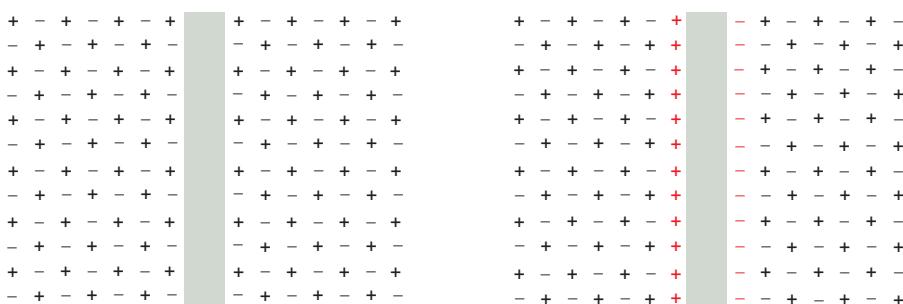
For any particular membrane potential,  $V_M$ , the net force tending to drive a particular type of ion out of the cell, is proportional to the difference between  $V_M$  and the equilibrium potential for the ion: hence,

for  $\text{K}^+$  it is  $V_M - V_K$

and for  $\text{Na}^+$  it is  $V_M - V_{\text{Na}}$ .

When there is a voltage gradient across the membrane, the ions responsible for it—the positive ions on one side and the negative ions on the other—are concentrated in thin layers on either side of the membrane because of the attraction between positive and negative electric charges. The number of ions that go to form the layer of charge adjacent to the membrane is minute compared with the total number inside the cell. For example, the movement of 6000  $\text{Na}^+$  ions across  $1 \mu\text{m}^2$  of membrane will carry sufficient charge to shift the membrane potential by about 100 mV.

Because there are about  $3 \times 10^7 \text{ Na}^+$  ions in a typical cell ( $1 \mu\text{m}^3$  of bulk cytoplasm), such a movement of charge will generally have a negligible effect on the ion concentration gradients across the membrane.



exact balance of charges on each side of the membrane; membrane potential = 0

a few of the positive ions (red) cross the membrane from right to left, leaving their negative counterions (red) behind; this sets up a nonzero membrane potential

small direct contribution to the membrane potential by pumping out three  $\text{Na}^+$  for every two  $\text{K}^+$  that it pumps in (see Figure 11–15). However, switching off the pump does not abolish the major component of the resting potential, which is generated by the  $\text{K}^+$  equilibrium mechanism just described. This component of the membrane potential persists as long as the  $\text{Na}^+$  concentration inside the cell stays low and the  $\text{K}^+$  ion concentration high—typically for many minutes. But the plasma membrane is somewhat permeable to all small ions, including  $\text{Na}^+$ . Therefore, without the  $\text{Na}^+-\text{K}^+$  pump, the ion gradients set up by the pump will eventually run down, and the membrane potential established by diffusion through the  $\text{K}^+$  leak channels will fall as well. As  $\text{Na}^+$  enters, the cell eventually comes to a new resting state where  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cl}^-$  are all at equilibrium across the membrane. The membrane potential in this state is much less than it was in the normal cell with an active  $\text{Na}^+-\text{K}^+$  pump.

The resting potential of an animal cell varies between  $-20 \text{ mV}$  and  $-120 \text{ mV}$ , depending on the organism and cell type. Although the  $\text{K}^+$  gradient always has a major influence on this potential, the gradients of other ions (and the disequilibrating effects of ion pumps) also have a significant effect: the more permeable the membrane for a given ion, the more strongly the membrane potential tends to be driven toward the equilibrium value for that ion. Consequently, changes in a membrane's permeability to ions can cause significant changes in the membrane potential. This is one of the key principles relating the electrical excitability of cells to the activities of ion channels.

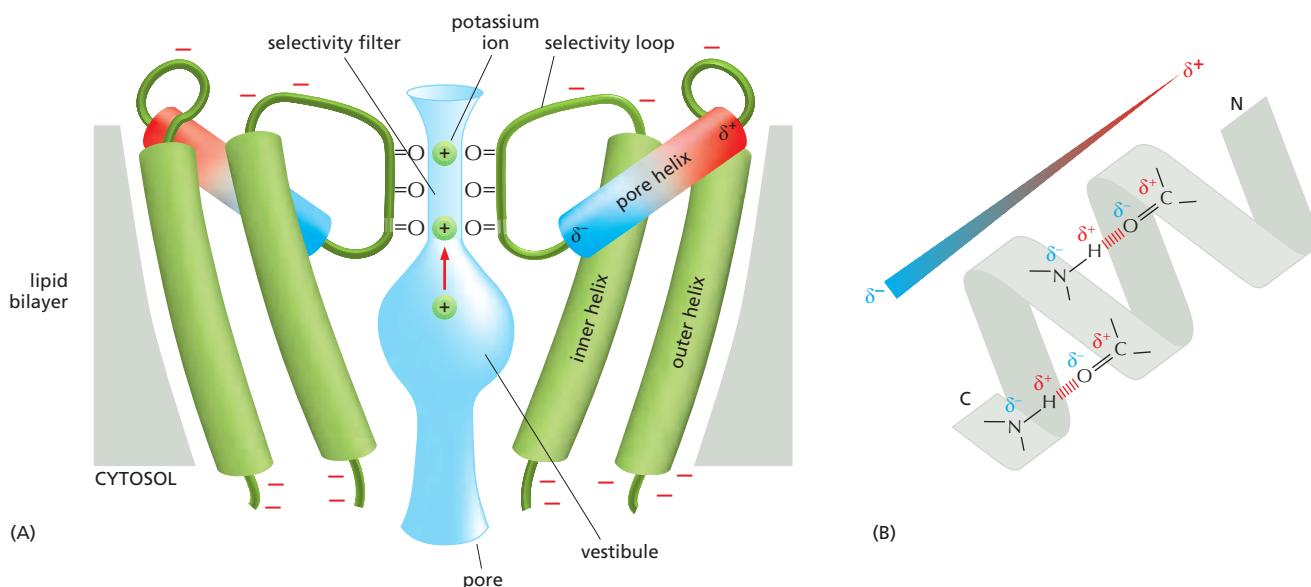
To understand how ion channels select their ions and how they open and close, we need to know their atomic structure. The first ion channel to be crystallized and studied by x-ray diffraction was a bacterial  $\text{K}^+$  channel. The details of its structure revolutionized our understanding of ion channels.

### The Three-Dimensional Structure of a Bacterial $\text{K}^+$ Channel Shows How an Ion Channel Can Work

Scientists were puzzled by the remarkable ability of ion channels to combine exquisite ion selectivity with a high conductance.  $\text{K}^+$  leak channels, for example, conduct  $\text{K}^+$  10,000-fold faster than  $\text{Na}^+$ , yet the two ions are both featureless spheres and have similar diameters ( $0.133 \text{ nm}$  and  $0.095 \text{ nm}$ , respectively). A single amino acid substitution in the pore of an animal cell  $\text{K}^+$  channel can result in a loss of ion selectivity and cell death. We cannot explain the normal  $\text{K}^+$  selectivity by pore size, because  $\text{Na}^+$  is smaller than  $\text{K}^+$ . Moreover, the high conductance rate is incompatible with the channel's having selective, high-affinity  $\text{K}^+$ -binding sites, as the binding of  $\text{K}^+$  ions to such sites would greatly slow their passage.

The puzzle was solved when the structure of a *bacterial  $\text{K}^+$  channel* was determined by x-ray crystallography. The channel is made from four identical transmembrane subunits, which together form a central pore through the membrane. Each subunit contributes two transmembrane  $\alpha$  helices, which are tilted outward in the membrane and together form a cone, with its wide end facing the outside of

**Figure 11–23** The ionic basis of a membrane potential. A small flow of inorganic ions through an ion channel carries sufficient charge to cause a large change in the membrane potential. The ions that give rise to the membrane potential lie in a thin ( $< 1 \text{ nm}$ ) surface layer close to the membrane, held there by their electrical attraction to their oppositely charged counterparts (counterions) on the other side of the membrane. For a typical cell, 1 microcoulomb of charge ( $6 \times 10^{12}$  monovalent ions) per square centimeter of membrane, transferred from one side of the membrane to the other, changes the membrane potential by roughly 1 V. This means, for example, that in a spherical cell of diameter  $10 \mu\text{m}$ , the number of  $\text{K}^+$  ions that have to flow out to alter the membrane potential by  $100 \text{ mV}$  is only about  $1/100,000$  of the total number of  $\text{K}^+$  ions in the cytosol. This amount is so minute that the intracellular  $\text{K}^+$  concentration remains virtually unchanged.



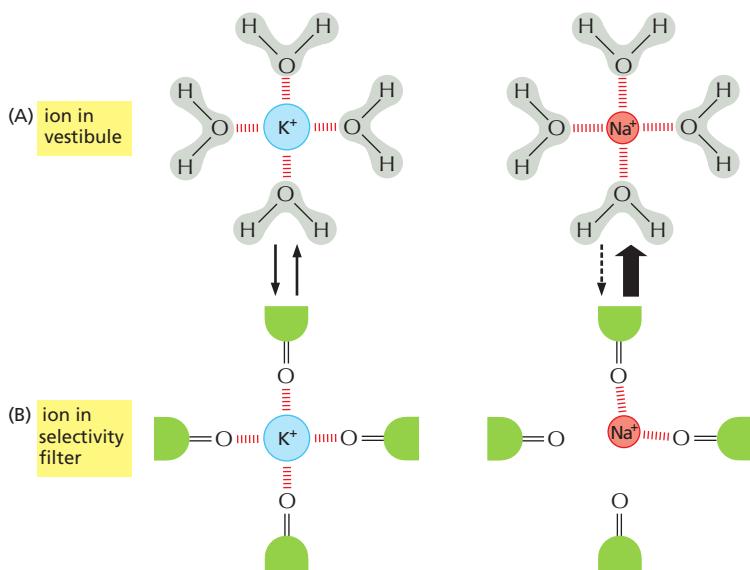
**Figure 11-24** The structure of a bacterial K<sup>+</sup> channel. (A) The transmembrane  $\alpha$  helices from only two of the four identical subunits are shown. From the cytosolic side, the pore (schematically shaded in blue) opens up into a vestibule in the middle of the membrane. The pore vestibule facilitates transport by allowing the K<sup>+</sup> ions to remain hydrated even though they are more than halfway across the membrane. The narrow selectivity filter of the pore links the vestibule to the outside of the cell. Carbonyl oxygens line the walls of the selectivity filter and form transient binding sites for dehydrated K<sup>+</sup> ions. Two K<sup>+</sup> ions occupy different sites in the selectivity filter, while a third K<sup>+</sup> ion is located in the center of the vestibule, where it is stabilized by electrical interactions with the more negatively charged ends of the pore helices. The ends of the four short “pore helices” (only two of which are shown) point precisely toward the center of the vestibule, thereby guiding K<sup>+</sup> ions into the selectivity filter (Movie 11.8). (B) Peptide bonds have an electric dipole, with more negative charge accumulated at the oxygen of the C=O bond and at the nitrogen of the N-H bond. In an  $\alpha$  helix, hydrogen bonds (red) align the dipoles. As a consequence, every  $\alpha$  helix has an electric dipole along its axis, resulting from summation of the dipoles of the individual peptide bonds, with a more negatively charged C-terminal end ( $\delta^-$ ) and a more positively charged N-terminal end ( $\delta^+$ ). (A, adapted from D.A. Doyle et al., *Science* 280:69–77, 1998.)

the cell where K<sup>+</sup> ions exit from the channel (Figure 11-24). The polypeptide chain that connects the two transmembrane helices forms a short  $\alpha$  helix (the *pore helix*) and a crucial loop that protrudes into the wide section of the cone to form the **selectivity filter**. The selectivity loops from the four subunits form a short, rigid, narrow pore, which is lined by the carbonyl oxygen atoms of their polypeptide backbones. Because the selectivity loops of all known K<sup>+</sup> channels have similar amino acid sequences, it is likely that they form a closely similar structure.

The structure of the selectivity filter explains the ion selectivity of the channel. A K<sup>+</sup> ion must lose almost all of its bound water molecules to enter the filter, where it interacts instead with the carbonyl oxygens lining the filter; the oxygens are rigidly spaced at the exact distance to accommodate a K<sup>+</sup> ion. A Na<sup>+</sup> ion, in contrast, cannot enter the filter because the carbonyl oxygens are too far away from the smaller Na<sup>+</sup> ion to compensate for the energy expense associated with the loss of water molecules required for entry (Figure 11-25).

Structural studies of K<sup>+</sup> channels and other ion channels have also indicated some general principles of how these channels open and close. The gating involves movement of the helices in the membrane so that they either obstruct or open the path for ion movement. Depending on the particular type of channel, helices tilt, rotate, or bend during gating. The structure of a closed K<sup>+</sup> channel shows that by tilting the inner helices, the pore constricts like a diaphragm at its cytosolic end (Figure 11-26). Bulky hydrophobic amino acid side chains block the small opening that remains, preventing the entry of ions.

Many other ion channels operate on similar principles: the channel's gating helices are allosterically coupled to domains that form the ion-conducting pathway; and a conformational change in the gate—in response, say, to ligand binding or altered membrane potential—brings about conformational change in the conducting pathway, either opening it or blocking it off.

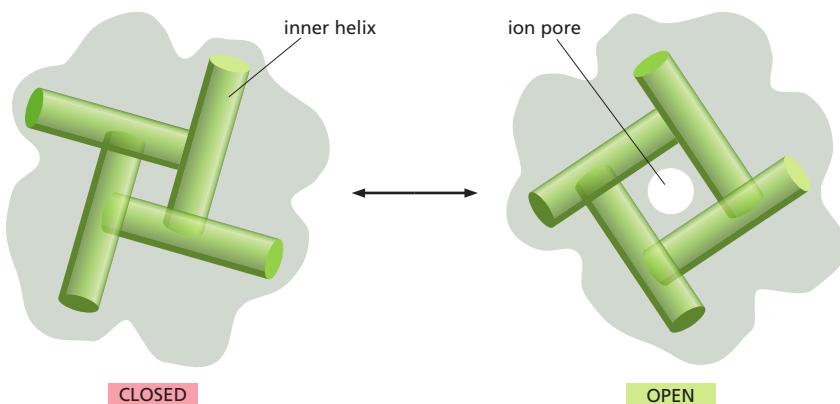


**Figure 11–25**  $K^+$  specificity of the selectivity filter in a  $K^+$  channel. The drawings show  $K^+$  and  $Na^+$  ions (A) in the vestibule and (B) in the selectivity filter of the pore, viewed in cross section. In the vestibule, the ions are hydrated. In the selectivity filter, they have lost their water, and the carbonyl oxygens are placed to accommodate a dehydrated  $K^+$  ion. The dehydration of the  $K^+$  ion requires energy, which is precisely balanced by the energy regained by the interaction of the ion with all of the carbonyl oxygens that serve as surrogate water molecules. Because the  $Na^+$  ion is too small to interact with the oxygens, it can enter the selectivity filter only at a great energetic expense. The filter therefore selects  $K^+$  ions with high specificity. (A, adapted from Y. Zhou et al., *Nature* 414:43–48, 2001. With permission from Macmillan Publishers Ltd.)

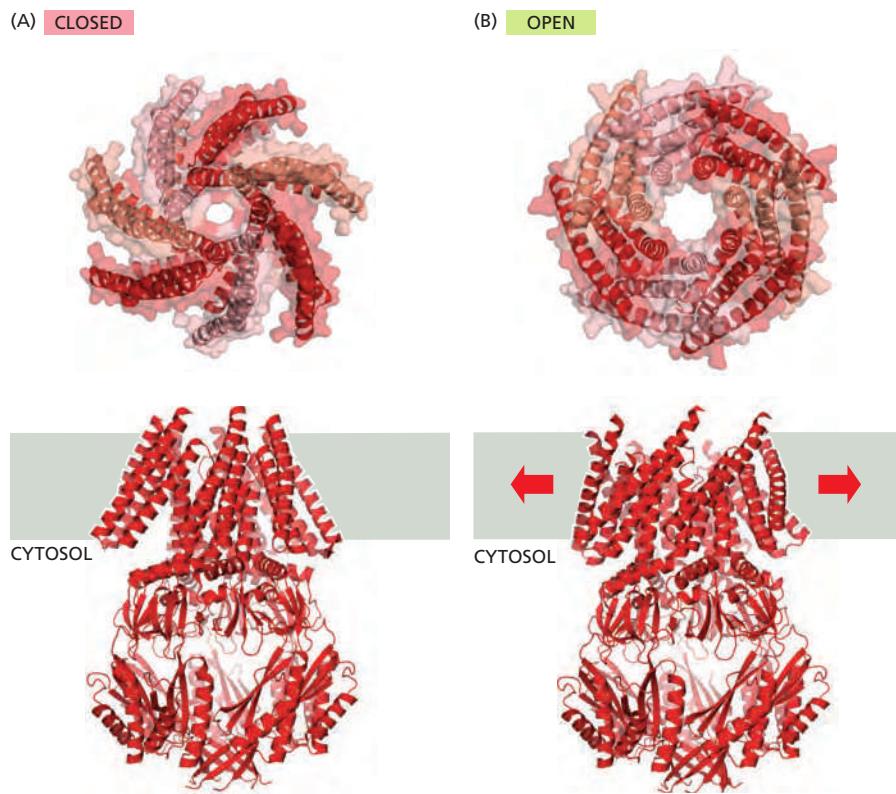
### Mechanosensitive Channels Protect Bacterial Cells Against Extreme Osmotic Pressures

All organisms, from single-cell bacteria to multicellular animals and plants, must sense and respond to mechanical forces in their external environment (such as sound, touch, pressure, shear forces, and gravity) and in their internal environment (such as osmotic pressure and membrane bending). Numerous proteins are known to be capable of responding to such mechanical forces, and a large subset of those proteins has been identified as possible **mechanosensitive channels**, but very few of the candidate proteins have been shown directly to be mechanically activated ion channels. One reason for this dearth in our knowledge is that most such channels are extremely rare. Auditory hair cells in the human cochlea, for example, contain extraordinarily sensitive mechanically gated ion channels, but each of the approximately 15,000 individual hair cells is thought to have a total of only 50–100 of them (Movie 11.9). Additional difficulties arise because the gating mechanisms of many mechanosensitive channel types require the channels to be embedded in complex architectures that require attachment to the extracellular matrix or to the cytoskeleton and are difficult to reconstitute in the test tube. The study of mechanosensitive receptors is a field of active investigation.

A well-studied class of mechanosensitive channels is found in the bacterial plasma membrane. These channels open in response to mechanical stretching of the lipid bilayer in which they are embedded. When a bacterium experiences a low-ionic-strength external environment (hypotonic conditions), such as



**Figure 11–26** A model for the gating of a bacterial  $K^+$  channel. The channel is viewed in cross section. To adopt the closed conformation, the four inner transmembrane helices that line the pore on the cytosolic side of the selectivity filter (see Figure 11–24) rearrange to close the cytosolic entrance to the channel. (Adapted from E. Perozo et al., *Science* 285:73–78, 1999.)



**Figure 11–27** The structure of mechanosensitive channels. The crystal structures of MscS in its (A) closed and (B) open conformation are shown. The side views (lower panels) show the entire protein, including the large intracellular domain. The face views (upper panels) show the transmembrane domains only. The open structure occupies more area in the lipid bilayer and is energetically favored when a membrane is stretched. This may explain why the MscS channel opens as pressure builds up inside the cell. (PDB codes: 2OAU, 2VV5.)

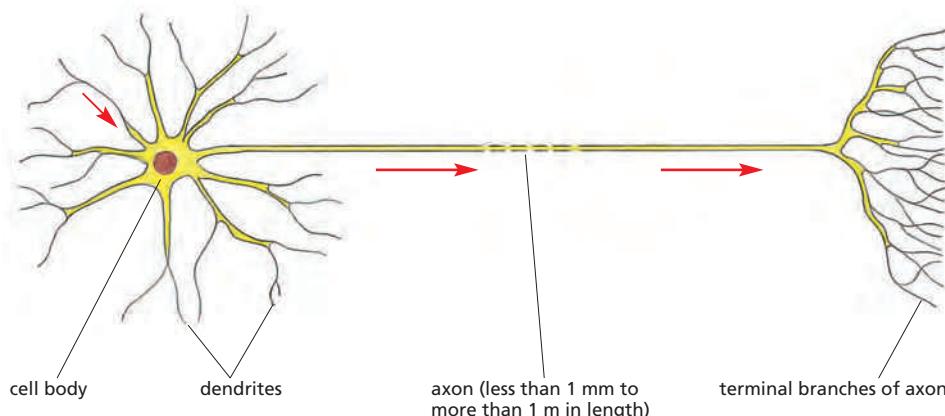
rainwater, the cell swells as water seeps in due to an increase in the osmotic pressure. If the pressure rises to dangerous levels, the cell opens mechanosensitive channels that allow small molecules to leak out. Bacteria that are experimentally placed in fresh water can rapidly lose more than 95% of their small molecules in this manner, including amino acids, sugars, and potassium ions. However, they keep their macromolecules safely inside and thus can recover quickly after environmental conditions return to normal.

Mechanical gating has been demonstrated using biophysical techniques in which force is exerted on pure lipid bilayers containing the bacterial mechanosensitive channels; for example, by applying suction with a micropipette. Such measurements demonstrate that the cell has several different channels that open at different levels of pressure. The mechanosensitive channel of small conductance, called the MscS channel, opens at low and moderate pressures (Figure 11–27). It is composed of seven identical subunits, which in the open state form a pore about 1.3 nm in diameter—just big enough to pass ions and small molecules. Large cytoplasmic domains limit the size of molecules that can reach the pore. The mechanosensitive channel of large conductance, called the MscL channel, opens to over 3 nm in diameter when the pressure gets so high that the cell might burst.

### The Function of a Neuron Depends on Its Elongated Structure

The cells that make most sophisticated use of channels are neurons. Before discussing how they do so, we digress briefly to describe how a typical neuron is organized.

The fundamental task of a **neuron**, or **nerve cell**, is to receive, conduct, and transmit signals. To perform these functions, neurons are often extremely elongated. In humans, for example, a single neuron extending from the spinal cord to a muscle in the foot may be as long as 1 meter. Every neuron consists of a cell body (containing the nucleus) with a number of thin processes radiating outward from it. Usually one long **axon** conducts signals away from the cell body toward distant



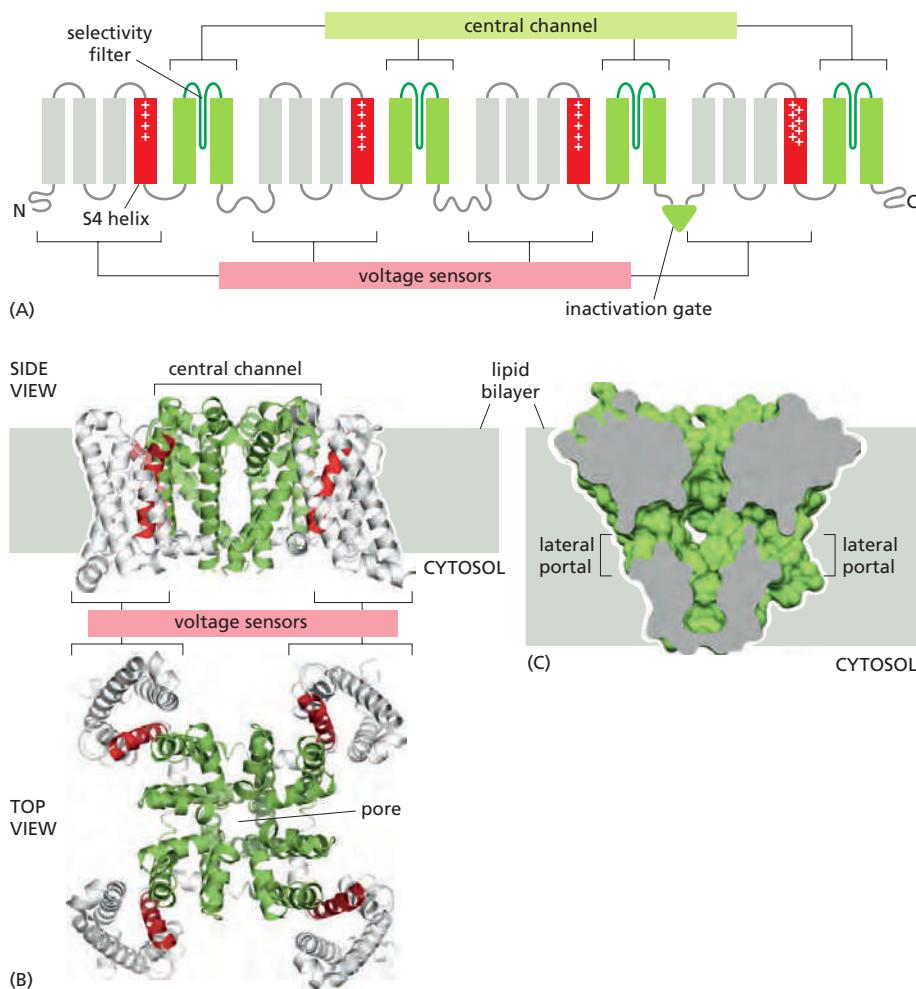
**Figure 11–28 A typical vertebrate neuron.** The arrows indicate the direction in which signals are conveyed. The single axon conducts signals away from the cell body, while the multiple dendrites (and the cell body) receive signals from the axons of other neurons. The axon terminals end on the dendrites or cell body of other neurons or on other cell types, such as muscle or gland cells.

targets, and several shorter, branching **dendrites** extend from the cell body like antennae, providing an enlarged surface area to receive signals from the axons of other neurons (Figure 11–28), although the cell body itself also receives such signals. A typical axon divides at its far end into many branches, passing on its message to many target cells simultaneously. Likewise, the extent of branching of the dendrites can be very great—in some cases sufficient to receive as many as 100,000 inputs on a single neuron.

Despite the varied significance of the signals carried by different classes of neurons, the form of the signal is always the same, consisting of changes in the electrical potential across the neuron's plasma membrane. The signal spreads because an electrical disturbance produced in one part of the membrane spreads to other parts, although the disturbance becomes weaker with increasing distance from its source, unless the neuron expends energy to amplify it as it travels. Over short distances, this attenuation is unimportant; in fact, many small neurons conduct their signals passively, without amplification. For long-distance communication, however, such passive spread is inadequate. Thus, larger neurons employ an active signaling mechanism, which is one of their most striking features. An electrical stimulus that exceeds a certain threshold strength triggers an explosion of electrical activity that propagates rapidly along the neuron's plasma membrane and is sustained by automatic amplification all along the way. This traveling wave of electrical excitation, known as an **action potential**, or *nerve impulse*, can carry a message without attenuation from one end of a neuron to the other at speeds of 100 meters per second or more. Action potentials are the direct consequence of the properties of voltage-gated cation channels, as we now discuss.

### Voltage-Gated Cation Channels Generate Action Potentials in Electrically Excitable Cells

The plasma membrane of all electrically excitable cells—not only neurons, but also muscle, endocrine, and egg cells—contains **voltage-gated cation channels**, which are responsible for generating the action potentials. An action potential is triggered by a **depolarization** of the plasma membrane—that is, by a shift in the membrane potential to a less negative value inside. (We shall see later how the action of a neurotransmitter causes depolarization.) In nerve and skeletal muscle cells, a stimulus that causes sufficient depolarization promptly opens the **voltage-gated Na<sup>+</sup> channels**, allowing a small amount of Na<sup>+</sup> to enter the cell down its electrochemical gradient. The influx of positive charge depolarizes the membrane further, thereby opening more Na<sup>+</sup> channels, which admit more Na<sup>+</sup> ions, causing still further depolarization. This self-amplification process (an example of *positive feedback*, discussed in Chapters 8 and 15) continues until, within a fraction of a millisecond, the electrical potential in the local region of membrane has shifted from its resting value of about -70 mV (in squid giant axon; about -40 mV in human) to almost as far as the Na<sup>+</sup> equilibrium potential of about +50 mV (see



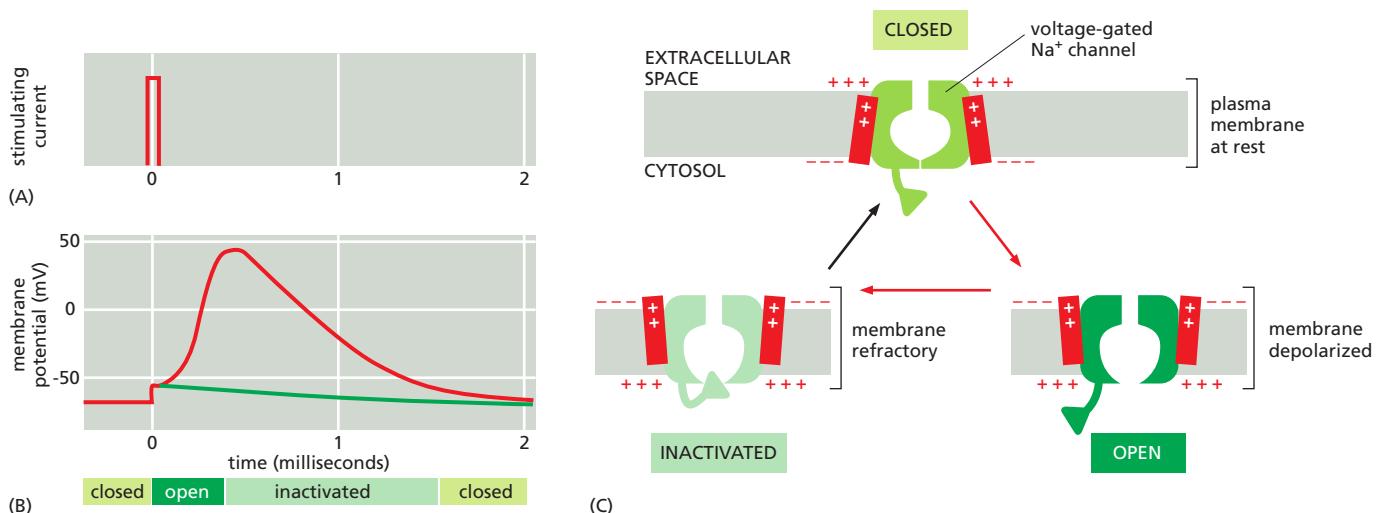
**Figure 11-29** Structural models of voltage-gated  $\text{Na}^+$  channels. (A) The channel in animal cells is built from a single polypeptide chain that contains four homologous domains. Each domain contains two transmembrane  $\alpha$  helices (green) that surround the central ion-conducting pore. They are separated by sequences (blue) that form the selectivity filter. Four  $\alpha$  additional helices (gray and red) in each domain constitute the voltage sensor. The S4 helices (red) are unique in that they contain an abundance of positively charged arginines. An inactivation gate that is part of a flexible loop connecting the third and fourth domains acts as a plug that obstructs the pore in the channel's inactivated state, as shown in Figure 11-30. (B) Side and top views of a homologous bacterial channel protein showing its arrangement within the membrane. (C) A cross section of the pore domain of the channel shown in (B) shows lateral portals, through which the central cavity is accessible from the hydrophobic core of the lipid bilayer. In the crystals, lipid acyl chains were found to intrude into the pore. These lateral portals are large enough to allow entry of small, hydrophobic, pore-blocking drugs that are commonly used as anesthetics and block ion conductance. (PDB code: 3RVZ.)

Panel 11-1, p. 616). At this point, when the net electrochemical driving force for the flow of  $\text{Na}^+$  is almost zero, the cell would come to a new resting state, with all of its  $\text{Na}^+$  channels permanently open, if the open conformation of the channel were stable. Two mechanisms act in concert to save the cell from such a permanent electrical spasm: the  $\text{Na}^+$  channels automatically inactivate and **voltage-gated  $\text{K}^+$  channels** open to restore the membrane potential to its initial negative value.

The  $\text{Na}^+$  channel is built from a single polypeptide chain that contains four structurally very similar domains. It is thought that these domains evolved by gene duplication followed by fusion into a single large gene (Figure 11-29A). In bacteria, in fact, the  $\text{Na}^+$  channel is a tetramer of four identical polypeptide chains, supporting this evolutionary idea.

Each domain contributes to the central channel, which is very similar to the  $\text{K}^+$  channel. Each domain also contains a *voltage sensor* that is characterized by an unusual transmembrane helix, S4, that contains many positively charged amino acids. As the membrane depolarizes, the S4 helices experience an electrostatic pulling force that attracts them to the now negatively charged extracellular side of the plasma membrane. The resulting conformational change opens the channel. The structure of a bacterial voltage-gated  $\text{Na}^+$  channel provides insights how the structural elements are arranged in the membrane (Figure 11-29B and C).

The  $\text{Na}^+$  channels also have an automatic inactivating mechanism, which causes the channels to reclose rapidly even though the membrane is still depolarized (see Figure 11-30). The  $\text{Na}^+$  channels remain in this *inactivated* state, unable to reopen, until after the membrane potential has returned to its initial negative value. The time necessary for a sufficient number of  $\text{Na}^+$  channels to recover from inactivation to support a new action potential, termed the *refractory period*, limits



**Figure 11-30**  $\text{Na}^+$  channels and an action potential. (A) An action potential is triggered by a brief pulse of current, which (B) partially depolarizes the membrane, as shown in the plot of membrane potential versus time. The green curve shows how the membrane potential would have simply relaxed back to the resting value after the initial depolarizing stimulus if there had been no voltage-gated  $\text{Na}^+$  channels in the membrane. The red curve shows the course of the action potential that is caused by the opening and subsequent inactivation of voltage-gated  $\text{Na}^+$  channels. The states of the  $\text{Na}^+$  channels are indicated in (B). The membrane cannot fire a second action potential until the  $\text{Na}^+$  channels have returned from the inactivated to the closed conformation; until then, the membrane is refractory to stimulation. (C) The three states of the  $\text{Na}^+$  channel. When the membrane is at rest (highly polarized), the closed conformation of the channel has the lowest free energy and is therefore most stable; when the membrane is depolarized, the energy of the *open* conformation is lower, so the channel has a high probability of opening. But the free energy of the *inactivated* conformation is lower still; therefore, after a randomly variable period spent in the open state, the channel becomes inactivated. Thus, the open conformation corresponds to a metastable state that can exist only transiently when the membrane depolarizes ([Movie 11.10](#)).

the repetitive firing rate of a neuron. The cycle from initial stimulus to the return to the original resting state takes a few milliseconds or less. The  $\text{Na}^+$  channel can therefore exist in three distinct states—closed, open, and inactivated—which contribute to the rise and fall of the action potential ([Figure 11-30](#)).

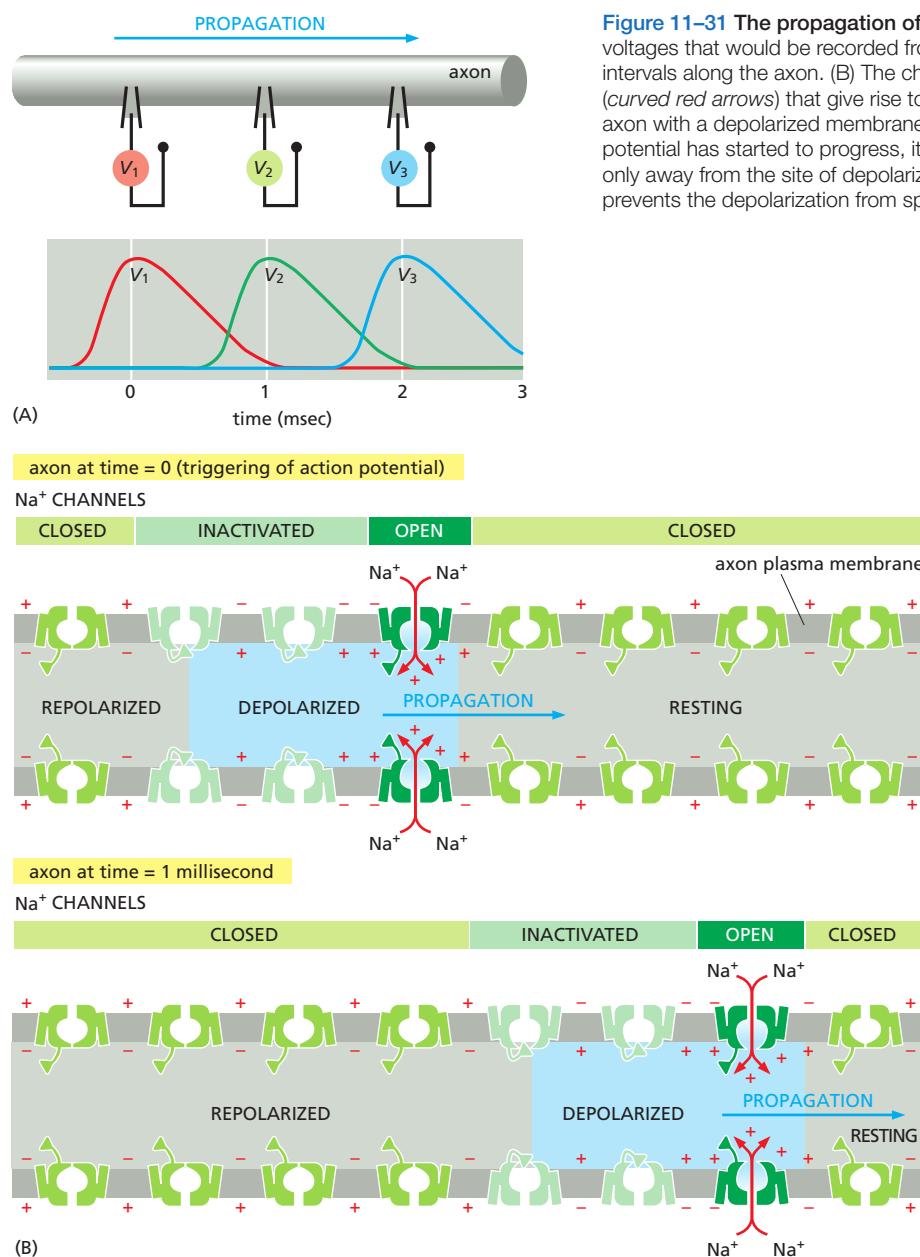
This description of an action potential applies only to a small patch of plasma membrane. The self-amplifying depolarization of the patch, however, is sufficient to depolarize neighboring regions of membrane, which then go through the same cycle. In this way, the action potential sweeps like a wave from the initial site of depolarization over the entire plasma membrane, as shown in [Figure 11-31](#).

### The Use of Channelrhodopsins Has Revolutionized the Study of Neural Circuits

**Channelrhodopsins** are photosensitive ion channels that open in response to light. They evolved as sensory receptors in photosynthetic green algae to allow the algae to swim toward light. The structure of channelrhodopsin closely resembles that of bacteriorhodopsin (see [Figure 10-31](#)). It contains a covalently bound retinal group that absorbs light and undergoes an isomerization reaction, which triggers a conformational change in the protein, opening an ion channel in the plasma membrane. In contrast to bacteriorhodopsin, which is a light-driven proton pump, channelrhodopsin is a light-driven cation channel.

Using genetic engineering techniques, channelrhodopsin can be expressed in virtually any cell type in vertebrates and invertebrates. Researchers first introduced the gene into cultured neurons and showed that flashing light could now activate the channelrhodopsin and induce the neurons to fire action potentials. Because the frequency of the light flashes determined the frequency of the action potentials, one can control the frequency of neuronal firing with millisecond precision.

Next, neurobiologists used the approach to activate specific neurons in the brain of experimental animals. Using a tiny fiber optic cable implanted near the



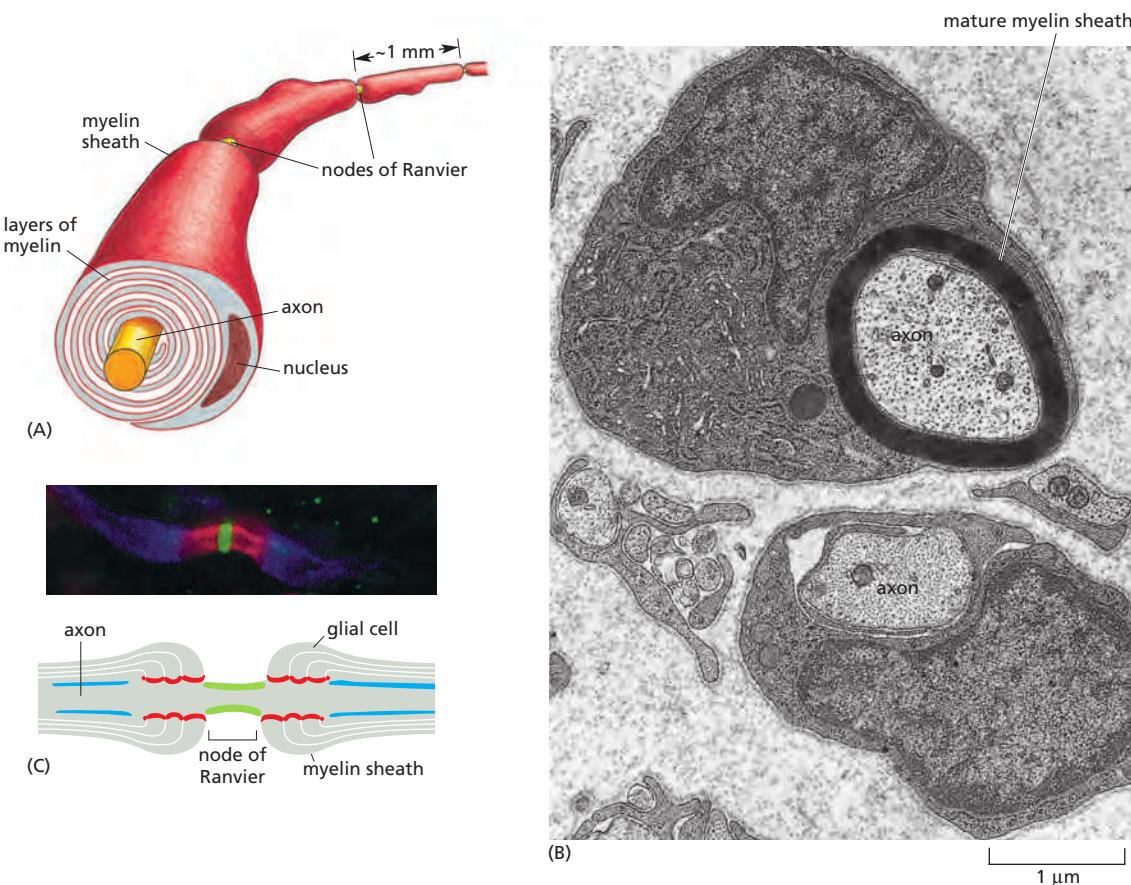
**Figure 11–31** The propagation of an action potential along an axon. (A) The voltages that would be recorded from a set of intracellular electrodes placed at intervals along the axon. (B) The changes in the  $\text{Na}^+$  channels and the current flows (curved red arrows) that give rise to a traveling action potential. The region of the axon with a depolarized membrane is shaded in blue. Note that once an action potential has started to progress, it has to continue in the same direction, traveling only away from the site of depolarization, because  $\text{Na}^+$ -channel inactivation prevents the depolarization from spreading backward.

relevant brain region, they could flash light to specifically activate the channelrhodopsin-containing neurons to fire action potentials. One group of researchers expressed channelrhodopsin in a subset of mouse neurons thought to be involved in aggression: when these cells were activated by light, the mouse immediately attacked anything in its environment—including other mice or even an inflated rubber glove (Figure 11–32); when the light was switched off, the neurons fell silent and the mouse's behavior returned to normal.

Since these pioneering studies, researchers have engineered additional light-responsive ion channels and transporters, including some that can rapidly



**Figure 11–32** Optogenetic control of aggression neurons in a living mouse. A gene encoding channelrhodopsin was introduced into a subpopulation of neurons in the hypothalamus of a mouse. When the neurons were exposed to flashing blue light using a tiny, implanted fiber optic cable, the channelrhodopsin channels opened, depolarizing and activating the cells. When the light was switched on, the mouse immediately became aggressive and attacked the inflated rubber glove; when the light was switched off, its behavior immediately returned to normal (Movie 11.11). (From D. Lin et al., *Nature* 470:221–226, 2011. With permission from Macmillan Publishers Ltd.)



inactivate specific neurons. It is therefore now possible to transiently activate or inhibit specific neurons in the brains of awake animals with remarkable spatial and temporal precision. In this way, the rapidly expanding new field of **optogenetics** is revolutionizing neurobiology, allowing neuroscientists to analyze the neurons and circuits underlying even the most complex behaviors in experimental animals, including nonhuman primates.

### Myelination Increases the Speed and Efficiency of Action Potential Propagation in Nerve Cells

The axons of many vertebrate neurons are insulated by a **myelin sheath**, which greatly increases the rate at which an axon can conduct an action potential. The importance of myelination is dramatically demonstrated by the demyelinating disease *multiple sclerosis*, in which the immune system destroys myelin sheaths in some regions of the central nervous system; in the affected regions, nerve impulse propagation greatly slows or even fails, often with devastating neurological consequences.

Myelin is formed by specialized non-neuronal supporting cells called **glial cells**. **Schwann cells** are the glial cells that myelinate axons in peripheral nerves, and **oligodendrocytes** do so in the central nervous system. These myelinating glial cells wrap layer upon layer of their own plasma membrane in a tight spiral around the axon (Figure 11–33A and B), thereby insulating the axonal membrane so that little current can leak across it. The myelin sheath is interrupted at regularly spaced *nodes of Ranvier*, where almost all the  $\text{Na}^+$  channels in the axon are concentrated (Figure 11–33C). This arrangement allows an action potential to propagate along a myelinated axon by jumping from node to node, a process called *saltatory conduction*. This type of conduction has two main advantages: action potentials travel very much faster, and metabolic energy is conserved because the active excitation is confined to the small regions of axonal plasma membrane at nodes of Ranvier.

**Figure 11–33 Myelination.**

(A) A myelinated axon from a peripheral nerve. Each Schwann cell wraps its plasma membrane concentrically around the axon to form a segment of myelin sheath about 1 mm long. For clarity, the membrane layers of the myelin are shown less compacted than they are in reality (see part B). (B) An electron micrograph of a nerve in the leg of a young rat. Two Schwann cells can be seen: one near the bottom is just beginning to myelinate its axon; the one above it has formed an almost mature myelin sheath. (C) Fluorescence micrograph and diagram of individual myelinated axons teased apart in a rat optic nerve, showing the confinement of the voltage-gated  $\text{Na}^+$  channels (green) in the axonal membrane at the node of Ranvier. A protein called Caspr (red) marks the junctions where the myelinating glial cell plasma membrane tightly abuts the axon on either side of the node. Voltage-gated  $\text{K}^+$  channels (blue) localize to regions in the axon plasma membrane well away from the node. (B, from Cedric S. Raine, in *Myelin* [P. Morell, ed.]. New York: Plenum, 1976; C, from M.N. Rasband and P. Shrager, *J. Physiol.* 525:63–73, 2000. With permission from Blackwell Publishing.)

## Patch-Clamp Recording Indicates That Individual Ion Channels Open in an All-or-Nothing Fashion

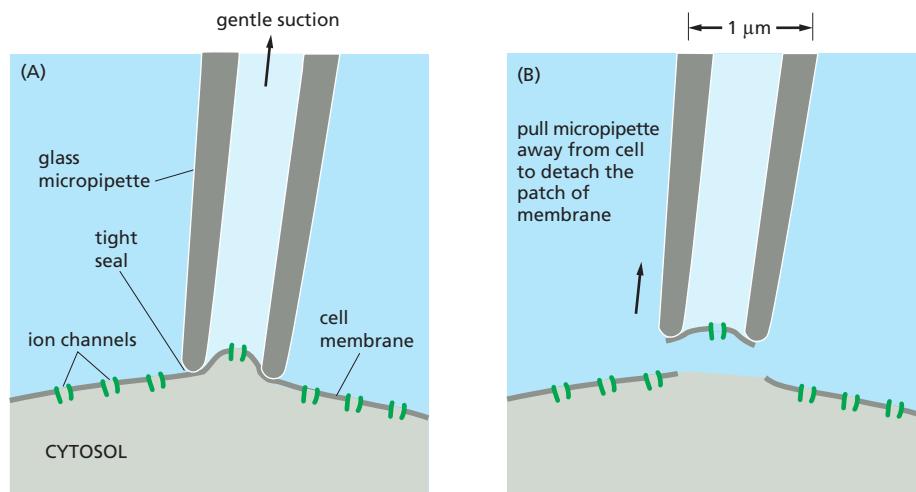
Neuron and skeletal muscle cell plasma membranes contain many thousands of voltage-gated  $\text{Na}^+$  channels, and the current crossing the membrane is the sum of the currents flowing through all of these. An intracellular microelectrode can record this aggregate current, as shown in Figure 11–31A. Remarkably, however, it is also possible to record current flowing through individual channels. **Patch-clamp recording**, developed in the 1970s and 1980s, revolutionized the study of ion channels and made it possible to examine transport through a single channel in a small patch of membrane covering the mouth of a micropipette (Figure 11–34). With this simple but powerful technique, one can study the detailed properties of ion channels in all sorts of cell types. This work led to the discovery that even cells that are not electrically excitable usually have a variety of ion channels in their plasma membrane. Many of these cells, such as yeasts, are too small to be investigated by the traditional electrophysiologist's method of impalement with an intracellular microelectrode.

Patch-clamp recording indicates that individual ion channels open in an all-or-nothing fashion. For example, a voltage-gated  $\text{Na}^+$  channel opens and closes at random, but when open, the channel always has the same large conductance, allowing more than 1000 ions to pass per millisecond (Figure 11–35). Therefore, the aggregate current crossing the membrane of an entire cell does not indicate the *degree* to which a typical individual channel is open but rather the *total number* of channels in its membrane that are open at any one time.

Some simple physical principles allow us to refine our understanding of voltage-gating from the perspective of a single  $\text{Na}^+$  channel. The interior of the resting neuron or muscle cell is at an electrical potential about 40–100 mV more negative than the external medium. Although this potential difference seems small, it exists across a plasma membrane only about 5 nm thick, so that the resulting voltage gradient is about 100,000 V/cm. Charged proteins in the membrane such as  $\text{Na}^+$  channels are thus subjected to a very large electrical field that can profoundly affect their conformation. Each conformation can “flip” to another conformation if given a sufficient jolt by the random thermal movements of the surroundings, and it is the relative stability of the closed, open, and inactivated conformations against flipping that is altered by changes in the membrane potential (see Figure 11–30C).

## Voltage-Gated Cation Channels Are Evolutionarily and Structurally Related

$\text{Na}^+$  channels are not the only kind of voltage-gated cation channel that can generate an action potential. The action potentials in some muscle, egg, and endocrine cells, for example, depend on *voltage-gated  $\text{Ca}^{2+}$  channels* rather than on  $\text{Na}^+$  channels.



**Figure 11–34** The technique of patch-clamp recording. Because of the extremely tight seal between the micropipette and the membrane, current can enter or leave the micropipette only by passing through the ion channels in the patch of membrane covering its tip. The term *clamp* is used because an electronic device is employed to maintain, or “clamp,” the membrane potential at a set value while recording the ionic current through individual channels. The current through these channels can be recorded with the patch still attached to the rest of the cell, as in (A), or detached, as in (B). The advantage of the detached patch is that it is easy to alter the composition of the solution on either side of the membrane to test the effect of various solutes on channel behavior. A detached patch can also be produced with the opposite orientation, so that the cytoplasmic surface of the membrane faces the inside of the pipette.

There is a surprising amount of structural and functional diversity within each of the different classes of voltage-gated cation channels, generated both by multiple genes and by the alternative splicing of RNA transcripts produced from the same gene. Nonetheless, the amino acid sequences of the known voltage-gated  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{2+}$  channels show striking similarities, demonstrating that they all belong to a large superfamily of evolutionarily and structurally related proteins and share many of the design principles. Whereas the single-celled yeast *S. cerevisiae* contains a single gene that codes for a voltage-gated  $\text{K}^+$  channel, the genome of the worm *C. elegans* contains 68 genes that encode different but related  $\text{K}^+$  channels. This complexity indicates that even a simple nervous system made up of only 302 neurons uses a large number of different ion channels to compute its responses.

Humans who inherit mutant genes encoding ion channels can suffer from a variety of nerve, muscle, brain, or heart diseases, depending in which cells the channel encoded by the mutant gene normally functions. Mutations in genes that encode voltage-gated  $\text{Na}^+$  channels in skeletal muscle cells, for example, can cause *myotonia*, a condition in which there is a delay in muscle relaxation after voluntary contraction, causing painful muscle spasms. In some cases, this occurs because the abnormal channels fail to inactivate normally; as a result,  $\text{Na}^+$  entry persists after an action potential finishes and repeatedly reinitiates membrane depolarization and muscle contraction. Similarly, mutations that affect  $\text{Na}^+$  or  $\text{K}^+$  channels in the brain can cause *epilepsy*, in which excessive synchronized firing of large groups of neurons causes epileptic seizures (convulsions, or fits).

The particular combination of ion channels conducting  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{2+}$  that are expressed in a neuron largely determines how the cell fires repetitive sequences of action potentials. Some nerve cells can repeat action potentials up to 300 times per second; other neurons fire short bursts of action potentials separated by periods of silence; while others rarely fire more than one action potential at a time. There is a remarkable diversity of neurons in the brain.

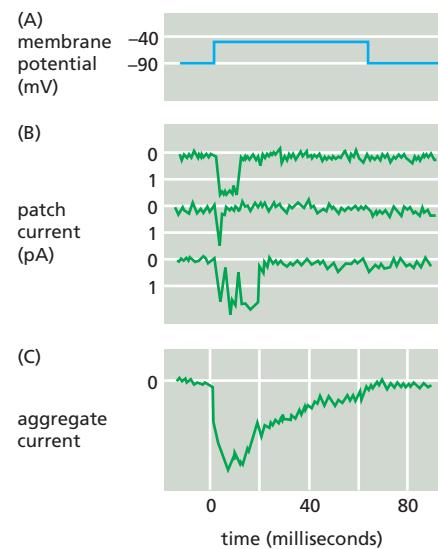
### Different Neuron Types Display Characteristic Stable Firing Properties

It is estimated that the human brain contains about  $10^{11}$  neurons and  $10^{14}$  synaptic connections. To make matters more complex, neural circuitry is continuously sculpted in response to experience, modified as we learn and store memories, and irreversibly altered by the gradual loss of neurons and their connections as we age. How can a system so complex be subject to such change and yet continue to function stably? One emerging theory suggests that individual neurons are self-tuning devices, constantly adjusting the expression of ion channels and neurotransmitter receptors in order to maintain a stable function. How might this work?

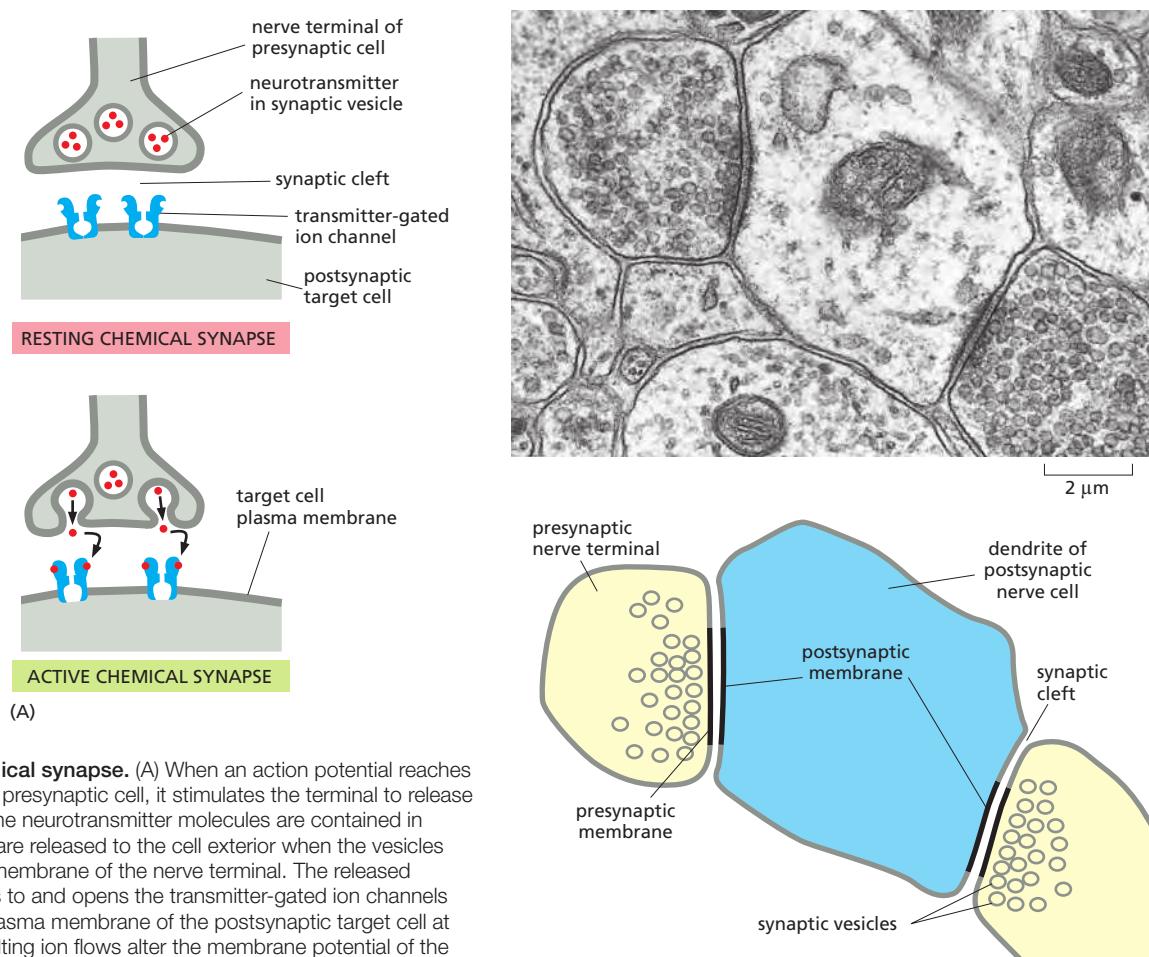
Neurons can be categorized into functionally different types, based in part on their propensity to fire action potentials and their pattern of firing. For example, some neurons fire action potentials at high frequencies, while others fire rarely. The firing properties of each neuron type are determined to a large extent by the ion channels that the cell expresses. The number of ion channels in a neuron's membrane is not fixed: as conditions change, a neuron can modify the numbers of depolarizing ( $\text{Na}^+$  and  $\text{Ca}^{2+}$ ) and hyperpolarizing ( $\text{K}^+$ ) channels and keep their proportions adjusted so as to maintain its characteristic firing behavior—a remarkable example of homeostatic control. The molecular mechanisms involved remain an important mystery.

### Transmitter-Gated Ion Channels Convert Chemical Signals into Electrical Ones at Chemical Synapses

Neuronal signals are transmitted from cell to cell at specialized sites of contact known as **synapses**. The usual mechanism of transmission is indirect. The cells are



**Figure 11–35** Patch-clamp measurements for a single voltage-gated  $\text{Na}^+$  channel. A tiny patch of plasma membrane was detached from an embryonic rat muscle cell, as in Figure 11–34. (A) The membrane was depolarized by an abrupt shift of potential from  $-90$  to about  $-40$  mV. (B) Three current records from three experiments performed on the same patch of membrane. Each major current step in (B) represents the opening and closing of a single channel. A comparison of the three records shows that, whereas the durations of channel opening and closing vary greatly, the rate at which current flows through an open channel (its conductance) is practically constant. The minor fluctuations in the current records arise largely from electrical noise in the recording apparatus. Current flowing into the cell, measured in picoamperes (pA), is shown as a downward deflection of the curve. By convention, the electrical potential on the outside of the cell is defined as zero. (C) The sum of the currents measured in 144 repetitions of the same experiment. This aggregate current is equivalent to the usual  $\text{Na}^+$  current that would be observed flowing through a relatively large region of membrane containing 144 channels. A comparison of (B) and (C) reveals that the time course of the aggregate current reflects the probability that any individual channel will be in the open state; this probability decreases with time as the channels in the depolarized membrane adopt their inactivated conformation. (Data from J. Patlak and R. Horn, *J. Gen. Physiol.* 79:333–351, 1982. With permission from The Rockefeller University Press.)



**Figure 11-36 A chemical synapse.** (A) When an action potential reaches the nerve terminal in a presynaptic cell, it stimulates the terminal to release its neurotransmitter. The neurotransmitter molecules are contained in synaptic vesicles and are released to the cell exterior when the vesicles fuse with the plasma membrane of the nerve terminal. The released neurotransmitter binds to and opens the transmitter-gated ion channels concentrated in the plasma membrane of the postsynaptic target cell at the synapse. The resulting ion flows alter the membrane potential of the postsynaptic membrane, thereby transmitting a signal from the excited nerve (B) (Movie 11.12). (B) A thin-section electron micrograph of two nerve terminal synapses on a dendrite of a postsynaptic cell. (B, courtesy of Cedric Raine.)

electrically isolated from one another, the *presynaptic cell* being separated from the *postsynaptic cell* by a narrow *synaptic cleft*. When an action potential arrives at the presynaptic site, the depolarization of the membrane opens voltage-gated  $\text{Ca}^{2+}$  channels that are clustered in the presynaptic membrane.  $\text{Ca}^{2+}$  influx triggers the release into the cleft of small signal molecules known as **neurotransmitters**, which are stored in membrane-enclosed *synaptic vesicles* and released by exocytosis (discussed in Chapter 13). The neurotransmitter diffuses rapidly across the synaptic cleft and provokes an electrical change in the postsynaptic cell by binding to and opening *transmitter-gated ion channels* (Figure 11-36). After the neurotransmitter has been secreted, it is rapidly removed: it is either destroyed by specific enzymes in the synaptic cleft or taken up by the presynaptic nerve terminal or by surrounding glial cells. Reuptake is mediated by a variety of  $\text{Na}^+$ -dependent neurotransmitter symporters (see Figure 11-8); in this way, neurotransmitters are recycled, allowing cells to keep up with high rates of release. Rapid removal ensures both spatial and temporal precision of signaling at a synapse. It decreases the chances that the neurotransmitter will influence neighboring cells, and it clears the synaptic cleft before the next pulse of neurotransmitter is released, so that the timing of repeated, rapid signaling events can be accurately communicated to the postsynaptic cell. As we shall see, signaling via such *chemical synapses* is far more versatile and adaptable than direct electrical coupling via gap junctions at *electrical synapses* (discussed in Chapter 19), which are also used by neurons but to a much smaller extent.

**Transmitter-gated ion channels**, also called **ionotropic receptors**, are built for rapidly converting extracellular chemical signals into electrical signals at

chemical synapses. The channels are concentrated in a specialized region of the postsynaptic plasma membrane at the synapse and open transiently in response to the binding of neurotransmitter molecules, thereby producing a brief permeability change in the membrane (see Figure 11–36A). Unlike the voltage-gated channels responsible for action potentials, transmitter-gated channels are relatively insensitive to the membrane potential and therefore cannot by themselves produce a self-amplifying excitation. Instead, they produce local permeability increases, and hence changes of membrane potential, that are graded according to the amount of neurotransmitter released at the synapse and how long it persists there. Only if the summation of small depolarizations at this site opens sufficient numbers of nearby voltage-gated cation channels can an action potential be triggered. This may require the opening of transmitter-gated ion channels at numerous synapses in close proximity on the target nerve cell.

### Chemical Synapses Can Be Excitatory or Inhibitory

Transmitter-gated ion channels differ from one another in several important ways. First, as receptors, they have highly selective binding sites for the neurotransmitter that is released from the presynaptic nerve terminal. Second, as channels, they are selective in the type of ions that they let pass across the plasma membrane; this determines the nature of the postsynaptic response. **Excitatory neurotransmitters** open cation channels, causing an influx of  $\text{Na}^+$ , and in many cases  $\text{Ca}^{2+}$ , that depolarizes the postsynaptic membrane toward the threshold potential for firing an action potential. **Inhibitory neurotransmitters**, by contrast, open either  $\text{Cl}^-$  channels or  $\text{K}^+$  channels, and this suppresses firing by making it harder for excitatory neurotransmitters to depolarize the postsynaptic membrane. Many transmitters can be either excitatory or inhibitory, depending on where they are released, what receptors they bind to, and the ionic conditions that they encounter. *Acetylcholine*, for example, can either excite or inhibit, depending on the type of acetylcholine receptors it binds to. Usually, however, acetylcholine, *glutamate*, and *serotonin* are used as excitatory transmitters, and  *$\gamma$ -aminobutyric acid (GABA)* and *glycine* are used as inhibitory transmitters. Glutamate, for instance, mediates most of the excitatory signaling in the vertebrate brain.

We have already discussed how the opening of  $\text{Na}^+$  or  $\text{Ca}^{2+}$  channels depolarizes a membrane. The opening of  $\text{K}^+$  channels has the opposite effect because the  $\text{K}^+$  concentration gradient is in the opposite direction—high concentration inside the cell, low outside. Opening  $\text{K}^+$  channels tends to keep the cell close to the equilibrium potential for  $\text{K}^+$ , which, as we discussed earlier, is normally close to the resting membrane potential because at rest  $\text{K}^+$  channels are the main type of channel that is open. When additional  $\text{K}^+$  channels open, it becomes harder to drive the cell away from the resting state. We can understand the effect of opening  $\text{Cl}^-$  channels similarly. The concentration of  $\text{Cl}^-$  is much higher outside the cell than inside (see Table 11–1, p. 598), but the membrane potential opposes its influx. In fact, for many neurons, the equilibrium potential for  $\text{Cl}^-$  is close to the resting potential—or even more negative. For this reason, opening  $\text{Cl}^-$  channels tends to buffer the membrane potential; as the membrane starts to depolarize, more negatively charged  $\text{Cl}^-$  ions enter the cell and counteract the depolarization. Thus, the opening of  $\text{Cl}^-$  channels makes it more difficult to depolarize the membrane and hence to excite the cell. Some powerful toxins act by blocking the action of inhibitory neurotransmitters: strychnine, for example, binds to glycine receptors and prevents their inhibitory action, causing muscle spasms, convulsions, and death.

However, not all chemical signaling in the nervous system operates through these ionotropic ligand-gated ion channels. In fact, most neurotransmitter molecules that are secreted by nerve terminals, including a large variety of neuropeptides, bind to **metabotropic receptors**, which regulate ion channels only indirectly through the action of small intracellular signal molecules (discussed in Chapter 15). All neurotransmitter receptors fall into one or other of these two

major classes—ionotropic or metabotropic—on the basis of their signaling mechanisms:

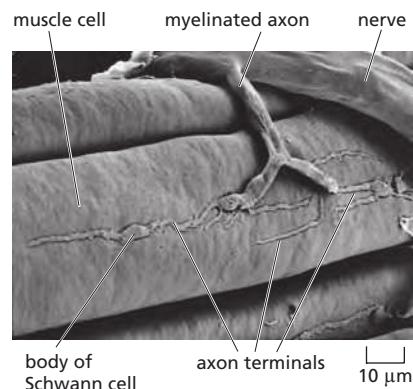
1. Ionotropic receptors are ion channels and feature fast chemical synapses. Acetylcholine, glycine, glutamate, and GABA all act on transmitter-gated ion channels, mediating excitatory or inhibitory signaling that is generally immediate, simple, and brief.
2. Metabotropic receptors are *G-protein-coupled receptors* (discussed in Chapter 15) that bind to all other neurotransmitters (and, confusingly, also acetylcholine, glutamate, and GABA). Signaling mediated by ligand-binding to metabotropic receptors tends to be far slower and more complex than that at ionotropic receptors, and longer-lasting in its consequences.

### The Acetylcholine Receptors at the Neuromuscular Junction Are Excitatory Transmitter-Gated Cation Channels

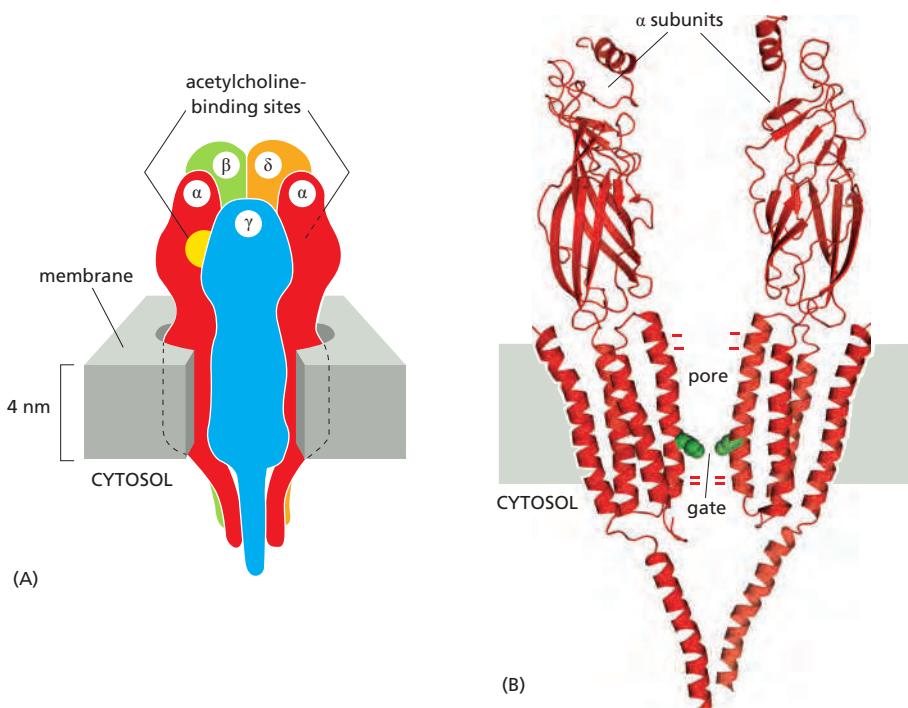
A well-studied example of a transmitter-gated ion channel is the **acetylcholine receptor** of skeletal muscle cells. This channel is opened transiently by acetylcholine released from the nerve terminal at a **neuromuscular junction**—the specialized chemical synapse between a motor neuron and a skeletal muscle cell (Figure 11–37). This synapse has been intensively investigated because it is readily accessible to electrophysiological study, unlike most of the synapses in the central nervous system, that is, the brain and spinal cord in vertebrates. Moreover, the acetylcholine receptors are densely packed in the muscle cell plasma membrane at a neuromuscular junction (about 20,000 such receptors per  $\mu\text{m}^2$ ), with relatively few receptors elsewhere in the same membrane.

The receptors are composed of five transmembrane polypeptides, two of one kind and three others, encoded by four separate genes (Figure 11–38A). The four genes are strikingly similar in sequence, implying that they evolved from a single ancestral gene. The two identical polypeptides in the pentamer each contribute one acetylcholine-binding site. When two acetylcholine molecules bind to the pentameric complex, they induce a conformational change that opens the channel. With ligand bound, the channel still flickers between open and closed states, but now it has a 90% probability of being open. This state continues—with acetylcholine binding and unbinding—until hydrolysis of the free acetylcholine by the enzyme *acetylcholinesterase* lowers its concentration at the neuromuscular junction sufficiently. Once freed of its bound neurotransmitter, the acetylcholine receptor reverts to its initial resting state. If the presence of acetylcholine persists for a prolonged time as a result of excessive nerve stimulation, the channel inactivates. Normally, the acetylcholine is rapidly hydrolyzed and the channel closes within about 1 millisecond, well before significant desensitization occurs. Desensitization would occur after about 20 milliseconds in the continued presence of acetylcholine.

The five subunits of the acetylcholine receptor are arranged in a ring, forming a water-filled transmembrane channel that consists of a narrow pore through the lipid bilayer, which widens into vestibules at both ends. Acetylcholine binding opens the channel by causing the helices that line the pore to rotate outward, thus disrupting a ring of hydrophobic amino acids that blocks ion flow in the closed state. Clusters of negatively charged amino acids at either end of the pore help to exclude negative ions and encourage any positive ion of diameter less than 0.65 nm to pass through (Figure 11–38B). The normal through-traffic consists chiefly of  $\text{Na}^+$  and  $\text{K}^+$ , together with some  $\text{Ca}^{2+}$ . Thus, unlike voltage-gated cation channels, such as the  $\text{K}^+$  channel discussed earlier, there is little selectivity among cations, and the relative contributions of the different cations to the current through the channel depend chiefly on their concentrations and on the electrochemical driving forces. When the muscle cell membrane is at its resting potential, the net driving force for  $\text{K}^+$  is near zero, since the voltage gradient nearly balances the  $\text{K}^+$  concentration gradient across the membrane (see Panel 11–1, p. 616). For  $\text{Na}^+$ , in contrast, the voltage gradient and the concentration gradient both act in the same direction to drive the ion into the cell. (The same is true for  $\text{Ca}^{2+}$ , but the



**Figure 11–37** A low-magnification scanning electron micrograph of a neuromuscular junction in a frog. The termination of a single axon on a skeletal muscle cell is shown. (From J. Desaki and Y. Uehara, *J. Neurocytol.* 10:101–110, 1981. With permission from Kluwer Academic Publishers.)



**Figure 11–38** A model for the structure of the skeletal muscle acetylcholine receptor. (A) Five homologous subunits ( $\alpha$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) combine to form a transmembrane pore. Both of the  $\alpha$  subunits contribute an acetylcholine-binding site nestled between adjoining subunits. (B) The pore is lined by a ring of five transmembrane  $\alpha$  helices, one contributed by each subunit (just the two  $\alpha$  subunits are shown). In its closed conformation, the pore is occluded by the hydrophobic side chains of five leucines (green), one from each  $\alpha$  helix, which form a gate near the middle of the lipid bilayer. When acetylcholine binds to both  $\alpha$  subunits, the channel undergoes a conformational change that opens the gate by an outward rotation of the helices containing the occluding leucines. Negatively charged side chains (indicated by the “ $-$ ” signs) at either end of the pore ensure that only positively charged ions pass through the channel. (PDB code: 2BG9.)

extracellular concentration of  $\text{Ca}^{2+}$  is so much lower than that of  $\text{Na}^+$  that  $\text{Ca}^{2+}$  makes only a small contribution to the total inward current.) Therefore, the opening of the acetylcholine-receptor channels leads to a large net influx of  $\text{Na}^+$  (a peak rate of about 30,000 ions per channel each millisecond). This influx causes a membrane depolarization that signals the muscle to contract, as discussed below.

### Neurons Contain Many Types of Transmitter-Gated Channels

The ion channels that open directly in response to the neurotransmitters acetylcholine, serotonin, GABA, and glycine contain subunits that are structurally similar and probably form transmembrane pores in the same way as the ionotropic acetylcholine receptor, even though they have distinct neurotransmitter-binding specificities and ion selectivities. These channels are all built from homologous polypeptide subunits, which assemble as a pentamer. Glutamate-gated ion channels are an exception, in that they are constructed from a distinct family of subunits and form tetramers resembling the  $\text{K}^+$  channels discussed earlier (see Figure 11–24A).

For each class of transmitter-gated ion channel, there are alternative forms of each type of subunit, which may be encoded by distinct genes or else generated by alternative RNA splicing of a single gene product. The subunits assemble in different combinations to form an extremely diverse set of distinct channel subtypes, with different ligand affinities, different channel conductances, different rates of opening and closing, and different sensitivities to drugs and toxins. Some vertebrate neurons, for example, have acetylcholine-gated ion channels that differ from those of muscle cells in that they are formed from two subunits of one type and three of another; but there are at least nine genes coding for different versions of the first type of subunit and at least three coding for different versions of the second. Subsets of such neurons performing different functions in the brain express different combinations of the genes for these subunits. In principle, and already to some extent in practice, it is possible to design drugs targeted against these narrowly defined subsets, thereby specifically influencing particular brain functions.

### Many Psychoactive Drugs Act at Synapses

Transmitter-gated ion channels have for a long time been important drug targets. A surgeon, for example, can relax muscles for the duration of an operation

by blocking the acetylcholine receptors on skeletal muscle cells with *curare*, a plant-derived drug that was originally used by South American Indians to make poison arrows. Most drugs used to treat insomnia, anxiety, depression, and schizophrenia exert their effects at chemical synapses, and many of these act by binding to transmitter-gated channels. Barbiturates, tranquilizers such as Valium, and sleeping pills such as Ambien, for example, bind to GABA receptors, potentiating the inhibitory action of GABA by allowing lower concentrations of this neurotransmitter to open  $\text{Cl}^-$  channels. Our increasing understanding of the molecular biology of ion channels should allow us to design a new generation of psychoactive drugs that will act still more selectively to alleviate the miseries of mental illness.

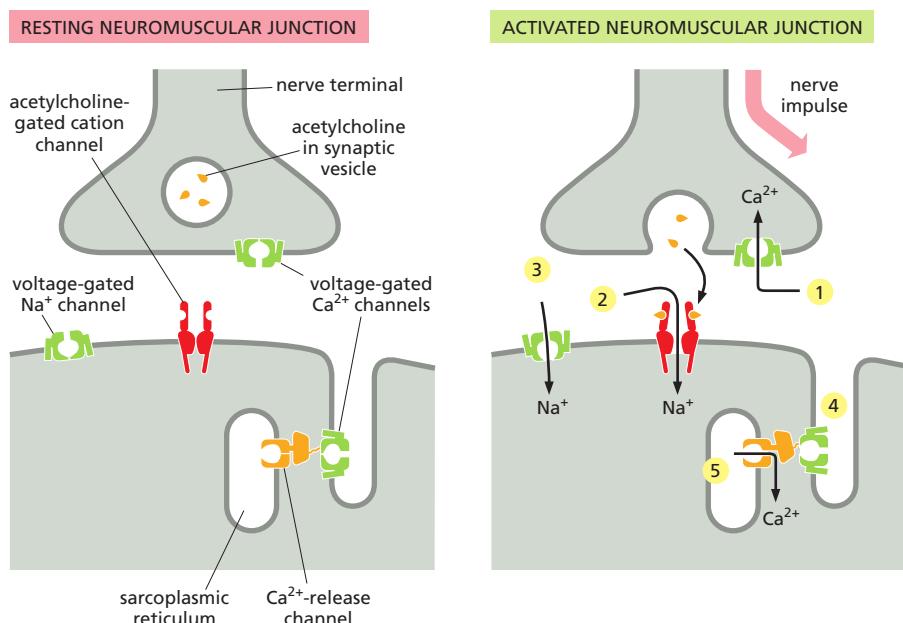
In addition to ion channels, many other components of the synaptic signaling machinery are potential targets for psychoactive drugs. As mentioned earlier, after release into the synaptic cleft, many neurotransmitters are cleared by reuptake mechanisms mediated by  $\text{Na}^+$ -driven symports. Inhibiting such transporters prolongs the effect of the neurotransmitter, thereby strengthening synaptic transmission. Many antidepressant drugs, including Prozac, inhibit the reuptake of serotonin; others inhibit the reuptake of both serotonin and norepinephrine.

Ion channels are the basic molecular units from which neuronal devices for signaling and computation are built. To provide a glimpse of how sophisticated these devices can be, we consider several examples that demonstrate how the coordinated activities of groups of ion channels allow you to move, feel, and remember.

### Neuromuscular Transmission Involves the Sequential Activation of Five Different Sets of Ion Channels

The following process, in which a nerve impulse stimulates a muscle cell to contract, illustrates the importance of ion channels to electrically excitable cells. This apparently simple response requires the sequential activation of at least five different sets of ion channels, all within a few milliseconds (**Figure 11–39**).

1. The process is initiated when a nerve impulse reaches the nerve terminal and depolarizes the plasma membrane of the terminal. The depolarization transiently opens voltage-gated  $\text{Ca}^{2+}$  channels in this presynaptic membrane. As the  $\text{Ca}^{2+}$  concentration outside cells is more than 1000 times



**Figure 11–39** The system of ion channels at a neuromuscular junction. These gated ion channels are essential for the stimulation of muscle contraction by a nerve impulse. The various channels are numbered in the sequence in which they are activated, as described in the text.

greater than the free  $\text{Ca}^{2+}$  concentration inside,  $\text{Ca}^{2+}$  flows into the nerve terminal. The increase in  $\text{Ca}^{2+}$  concentration in the cytosol of the nerve terminal triggers the local release of acetylcholine by exocytosis into the synaptic cleft.

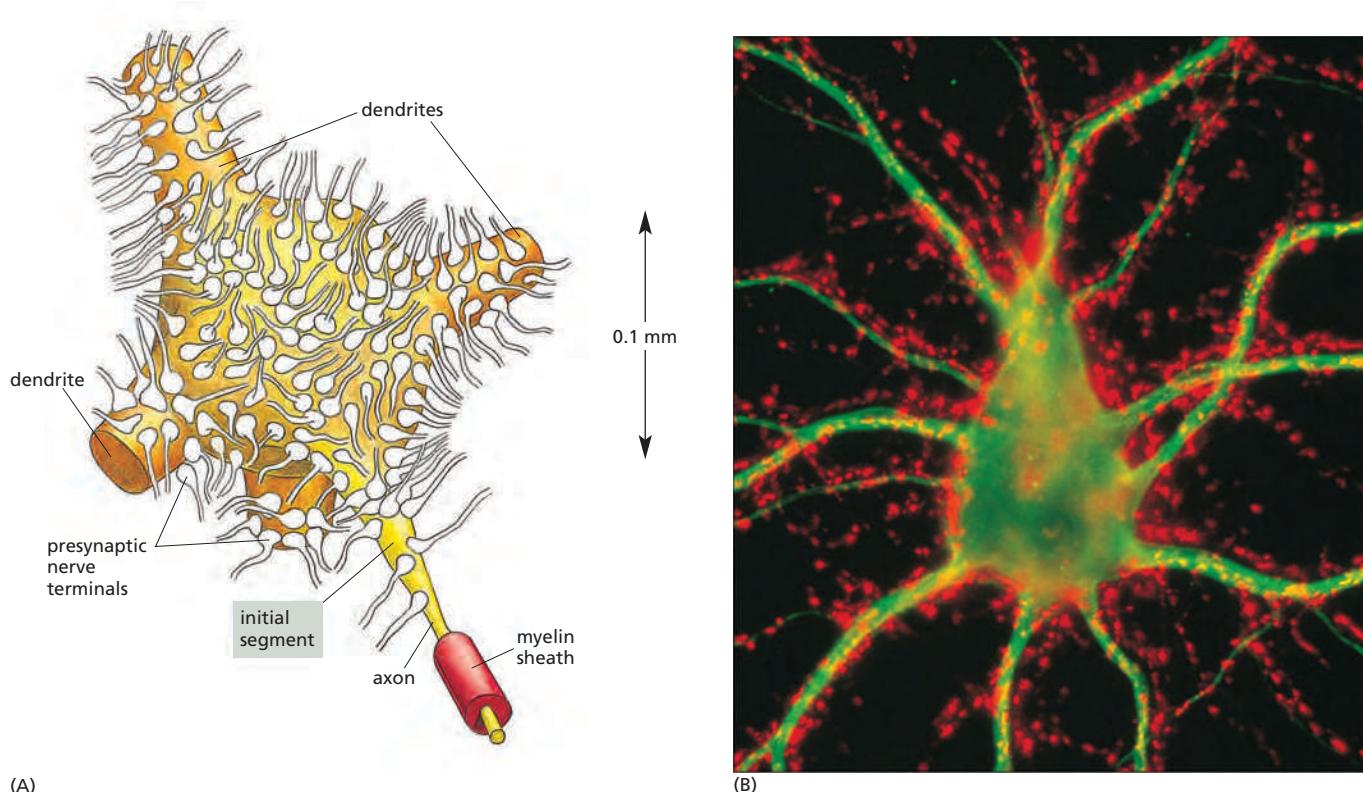
2. The released acetylcholine binds to acetylcholine receptors in the muscle cell plasma membrane, transiently opening the cation channels associated with them. The resulting influx of  $\text{Na}^+$  causes a local membrane depolarization.
3. The local depolarization opens voltage-gated  $\text{Na}^+$  channels in this membrane, allowing more  $\text{Na}^+$  to enter, which further depolarizes the membrane. This, in turn, opens neighboring voltage-gated  $\text{Na}^+$  channels and results in a self-propagating depolarization (an action potential) that spreads to involve the entire plasma membrane (see Figure 11–31).
4. The generalized depolarization of the muscle cell plasma membrane activates voltage-gated  $\text{Ca}^{2+}$  channels in the transverse tubules (T tubules—discussed in Chapter 16) of this membrane.
5. This in turn causes  $\text{Ca}^{2+}$ -release channels in an adjacent region of the sarcoplasmic reticulum (SR) membrane to open transiently and release  $\text{Ca}^{2+}$  stored in the SR into the cytosol. The T-tubule and SR membranes are closely apposed with the two types of channel joined together in a specialized structure, in which activation of the voltage-sensitive  $\text{Ca}^{2+}$  channel in the T-tubule plasma membrane causes a channel conformational change that is mechanically transmitted to the  $\text{Ca}^{2+}$ -release channel in the SR membrane, opening it and allowing  $\text{Ca}^{2+}$  to flow from the SR lumen into the cytoplasm (see Figure 16–35). The sudden increase in the cytosolic  $\text{Ca}^{2+}$  concentration causes the myofibrils in the muscle cell to contract.

Whereas the initiation of muscle contraction by a motor neuron is complex, an even more sophisticated interplay of ion channels is required for a neuron to integrate a large number of input signals at its synapses and compute an appropriate output, as we now discuss.

### Single Neurons Are Complex Computation Devices

In the central nervous system, a single neuron can receive inputs from thousands of other neurons, and it can in turn form synapses with many thousands of other cells. Several thousand nerve terminals, for example, make synapses on an average motor neuron in the spinal cord, almost completely covering its cell body and dendrites ([Figure 11–40](#)). Some of these synapses transmit signals from the brain or spinal cord; others bring sensory information from muscles or from the skin. The motor neuron must combine the information received from all these sources and react, either by firing action potentials along its axon or by remaining quiet.

Of the many synapses on a neuron, some tend to excite it, while others inhibit it. Neurotransmitter released at an excitatory synapse causes a small depolarization in the postsynaptic membrane called an *excitatory postsynaptic potential (excitatory PSP)*, whereas neurotransmitter released at an inhibitory synapse generally causes a small hyperpolarization called an *inhibitory PSP*. The plasma membrane of the dendrites and cell body of most neurons contains a relatively low density of voltage-gated  $\text{Na}^+$  channels, and so an individual excitatory PSP is generally too small to trigger an action potential. Instead, each incoming signal initiates a local PSP, which decreases with distance from the site of the synapse. If signals arrive simultaneously at several synapses in the same region of the dendritic tree, the total PSP in that neighborhood will be roughly the sum of the individual PSPs, with inhibitory PSPs making a negative contribution to the total. The PSPs from each neighborhood spread passively and converge on the cell body. For long-distance transmission, the combined magnitude of the PSP is then translated, or *encoded*, into the *frequency* of firing of action potentials: the greater the stimulation (depolarization), the higher the frequency of action potentials.



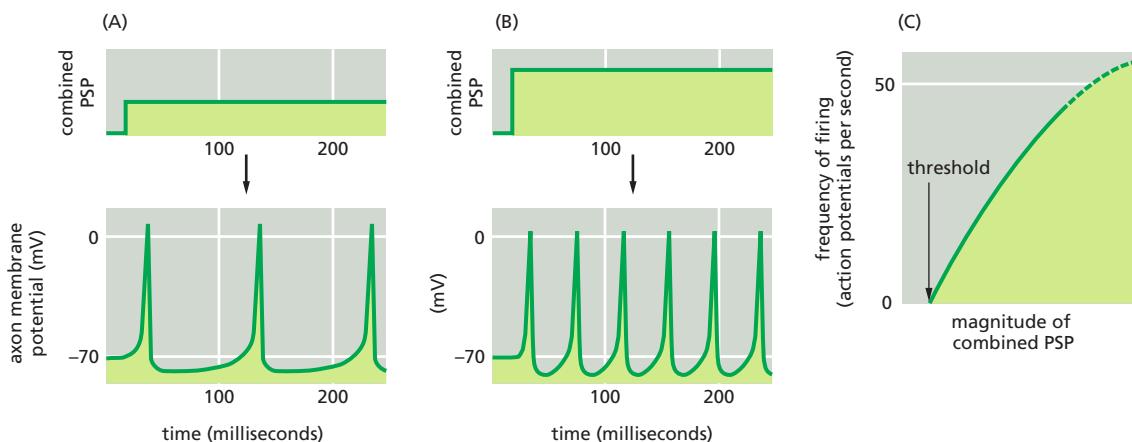
### Neuronal Computation Requires a Combination of at Least Three Kinds of $K^+$ Channels

The intensity of stimulation that a neuron receives is encoded by that neuron into action potential frequency for long-distance transmission. The encoding takes place at a specialized region of the axonal membrane known as the **initial segment**, or *axon hillock*, at the junction of the axon and the cell body (see Figure 11–40). This membrane is rich in voltage-gated  $Na^+$  channels; but it also contains at least four other classes of ion channels—three selective for  $K^+$  and one selective for  $Ca^{2+}$ —all of which contribute to the axon hillock's encoding function. The three varieties of  $K^+$  channels have different properties; we shall refer to them as *delayed*, *rapidly inactivating*, and  *$Ca^{2+}$ -activated  $K^+$  channels*.

To understand the need for multiple types of channels, consider first what would happen if the only voltage-gated ion channels present in the nerve cell were the  $Na^+$  channels. Below a certain threshold level of synaptic stimulation, the depolarization of the initial-segment membrane would be insufficient to trigger an action potential. With gradually increasing stimulation, the threshold would be crossed, the  $Na^+$  channels would open, and an action potential would fire. The action potential would be terminated by inactivation of the  $Na^+$  channels. Before another action potential could fire, these channels would have to recover from their inactivation. But that would require a return of the membrane voltage to a very negative value, which would not occur as long as the strong depolarizing stimulus (from PSPs) was maintained. An additional channel type is needed, therefore, to repolarize the membrane after each action potential to prepare the cell to fire again.

The **delayed  $K^+$  channels** perform this task, as discussed previously in relation to the propagation of the action potential (see Figure 11–31). They are voltage-gated, but because of their slower kinetics they open only during the falling phase of the action potential, when the  $Na^+$  channels are inactive. Their opening permits an efflux of  $K^+$  that drives the membrane back toward the  $K^+$  equilibrium potential, which is so negative that the  $Na^+$  channels rapidly recover from their inactivated state. Repolarization of the membrane also closes the delayed  $K^+$  channels. The initial segment is now reset so that the depolarizing stimulus from

**Figure 11–40** A motor neuron in the spinal cord. (A) Many thousands of nerve terminals synapse on the cell body and dendrites. These deliver signals from other parts of the organism to control the firing of action potentials along the single axon of this large cell. (B) Fluorescence micrograph showing a nerve cell body and its dendrites stained with a fluorescent antibody that recognizes a cytoskeletal protein (green) that is not present in axons. Thousands of axon terminals (red) from other nerve cells (not visible) make synapses on the cell body and dendrites; the terminals are stained with a fluorescent antibody that recognizes a protein in synaptic vesicles. (B, courtesy of Olaf Mundigl and Pietro de Camilli.)



synaptic inputs can fire another action potential. In this way, sustained stimulation of the dendrites and cell body leads to repetitive firing of the axon.

Repetitive firing in itself, however, is not enough. The frequency of firing has to reflect the intensity of stimulation, and a simple system of  $\text{Na}^+$  channels and delayed  $\text{K}^+$  channels is inadequate for this purpose. Below a certain threshold level of steady stimulation, the cell will not fire at all; above that threshold level, it will abruptly begin to fire at a relatively rapid rate. The **rapidly inactivating  $\text{K}^+$  channels** solve the problem. These, too, are voltage-gated and open when the membrane is depolarized, but their specific voltage sensitivity and kinetics of inactivation are such that they act to reduce the rate of firing at levels of stimulation that are only just above the threshold required for firing. Thus, they remove the discontinuity in the relationship between the firing rate and the intensity of stimulation. The result is a firing rate that is proportional to the strength of the depolarizing stimulus over a very broad range (Figure 11-41).

The process of encoding is usually further modulated by the two other types of ion channels in the initial segment that were mentioned earlier—voltage-gated  $\text{Ca}^{2+}$  channels and  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  channels. They act together to decrease the response of the cell to an unchanging, prolonged stimulation—a process called **adaptation**. These  $\text{Ca}^{2+}$  channels are similar to the  $\text{Ca}^{2+}$  channels that mediate the release of neurotransmitter from presynaptic axon terminals; they open when an action potential fires, transiently allowing  $\text{Ca}^{2+}$  into the axon cytosol at the initial segment.

The  **$\text{Ca}^{2+}$ -activated  $\text{K}^+$  channel** opens in response to a raised concentration of  $\text{Ca}^{2+}$  at the channel's cytoplasmic face (Figure 11-42). Prolonged, strong depolarizing stimuli will trigger a long train of action potentials, each of which permits a brief influx of  $\text{Ca}^{2+}$  through the voltage-gated  $\text{Ca}^{2+}$  channels, so that local cytosolic  $\text{Ca}^{2+}$  concentration gradually builds up to a level high enough to open the  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  channels. Because the resulting increased permeability of the membrane to  $\text{K}^+$  makes the membrane harder to depolarize, the delay between one action potential and the next is increased. In this way, a neuron that is stimulated continuously for a prolonged period becomes gradually less responsive to the constant stimulus.

Such adaptation, which can also occur by other mechanisms, allows a neuron—indeed, the nervous system generally—to react sensitively to *change*, even against a high background level of steady stimulation. It is one of the computational strategies that help us, for example, to feel a light touch on the shoulder and yet ignore the constant pressure of our clothing. We discuss adaptation as a general feature in cell signaling processes in more detail in Chapter 15.

Other neurons do different computations, reacting to their synaptic inputs in myriad ways, reflecting the different assortments of ion channels in their membrane. There are several hundred genes that code for ion channels in the human genome, with over 150 encoding voltage-gated channels alone. Further complexity is introduced by alternative splicing of RNA transcripts and assembling channel subunits in different combinations. Moreover, ion channels are selectively

**Figure 11-41** The magnitude of the combined postsynaptic potential (PSP) is reflected in the frequency of firing of action potentials. The mix of excitatory and inhibitory PSPs produces a *combined PSP* at the initial segment. A comparison of (A) and (B) shows how the firing frequency of an axon increases with an increase in the combined PSP, while (C) summarizes the general relationship.

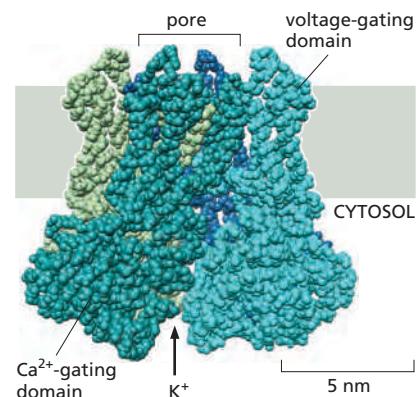
localized to different sites in the plasma membrane of a neuron. Some  $K^+$  and  $Ca^{2+}$  channels are concentrated in the dendrites and participate in processing the input that a neuron receives. As we have seen, other ion channels are located at the axon's initial segment, where they control action potential firing; and some ligand-gated channels are distributed over the cell body and, depending on their ligand occupancy, modulate the cell's general sensitivity to synaptic inputs. The multiplicity of ion channels and their locations evidently allows each of the many types of neurons to tune the electrical behavior to the particular tasks they perform.

One of the crucial properties of the nervous system is its ability to learn and remember. This property depends in part on the ability of individual synapses to strengthen or weaken depending on their use—a process called **synaptic plasticity**. We end this chapter by considering a remarkable type of ion channel that has a special role in some forms of synaptic plasticity. It is located at many excitatory synapses in the central nervous system, where it is gated by both voltage and the excitatory neurotransmitter glutamate. It is also the site of action of the psychoactive drug phencyclidine, or angel dust.

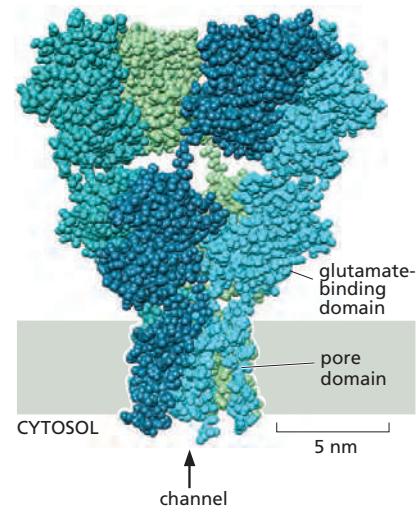
### Long-Term Potentiation (LTP) in the Mammalian Hippocampus Depends on $Ca^{2+}$ Entry Through NMDA-Receptor Channels

Practically all animals can learn, but mammals seem to learn exceptionally well (or so we like to think). In a mammal's brain, the region called the *hippocampus* has a special role in learning. When it is destroyed on both sides of the brain, the ability to form new memories is largely lost, although previous long-established memories remain. Some synapses in the hippocampus show a striking form of synaptic plasticity with repeated use: whereas occasional single action potentials in the presynaptic cells leave no lasting trace, a short burst of repetitive firing causes **long-term potentiation (LTP)**, such that subsequent single action potentials in the presynaptic cells evoke a greatly enhanced response in the postsynaptic cells. The effect lasts hours, days, or weeks, according to the number and intensity of the bursts of repetitive firing. Only the synapses that were activated exhibit LTP; synapses that have remained quiet on the same postsynaptic cell are not affected. However, while the cell is receiving a burst of repetitive stimulation via one set of synapses, if a single action potential is delivered at *another* synapse on its surface, that latter synapse also will undergo LTP, even though a single action potential delivered there at another time would leave no such lasting trace.

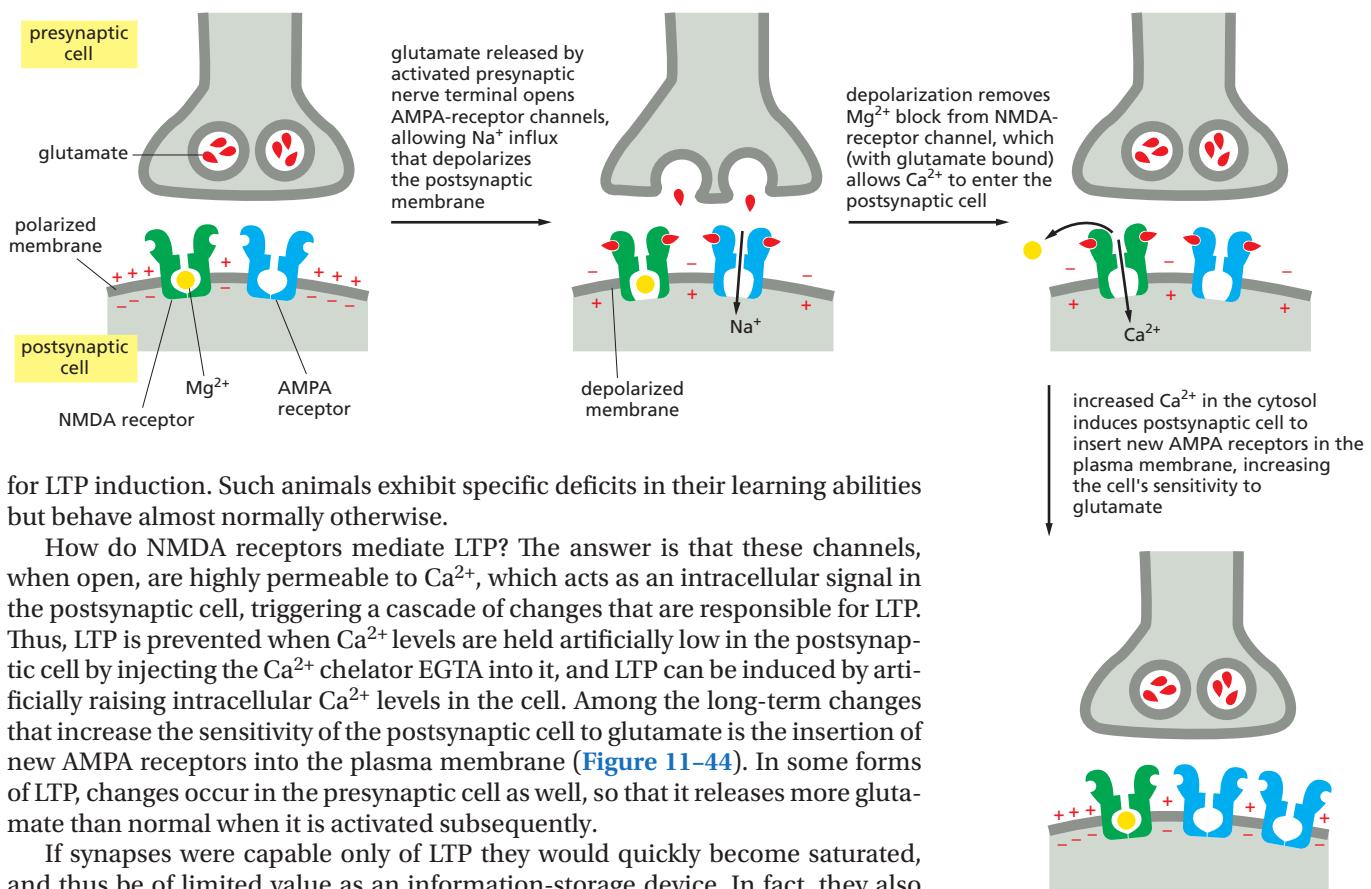
The underlying rule in such events seems to be that *LTP occurs on any occasion when a presynaptic cell fires (once or more) at a time when the postsynaptic membrane is strongly depolarized* (either through recent repetitive firing of the same presynaptic cell or by other means). This rule reflects the behavior of a particular class of ion channels in the postsynaptic membrane. Glutamate is the main excitatory neurotransmitter in the mammalian central nervous system, and glutamate-gated ion channels are the most common of all transmitter-gated channels in the brain. In the hippocampus, as elsewhere, most of the depolarizing current responsible for excitatory PSPs is carried by glutamate-gated ion channels called **AMPA receptors**, which operate in the standard way (Figure 11–43). But the current has, in addition, a second and more intriguing component, which is mediated by a separate subclass of glutamate-gated ion channels known as **NMDA receptors**, so named because they are selectively activated by the artificial glutamate analog *N*-methyl-D-aspartate. The NMDA-receptor channels are doubly gated, opening only when two conditions are satisfied simultaneously: glutamate must be bound to the receptor, and the membrane must be strongly depolarized. The second condition is required for releasing the  $Mg^{2+}$  that normally blocks the resting channel. This means that NMDA receptors are normally activated only when AMPA receptors are activated as well and depolarize the membrane. The NMDA receptors are critical for LTP. When they are selectively blocked with a specific inhibitor or inactivated genetically, LTP does not occur, even though ordinary synaptic transmission continues, indicating the importance of NMDA receptors.



**Figure 11–42** Structure of a  $Ca^{2+}$ -activated  $K^+$  channel. The channel contains four identical subunits (which are shown in different colors for clarity). It is both voltage- and  $Ca^{2+}$ -gated. The structure shown is a composite of the cytosolic and membrane portions of the channel that were separately crystallized. (PDB codes: 2R99, 1LNQ.)



**Figure 11–43** The structure of the AMPA receptor. This ionotropic glutamate receptor (named after the glutamate analog  $\alpha$ -Amino 3-hydroxy 5-Methyl 4-isoxazole Propionic Acid) is the most common mediator of fast, excitatory synaptic transmission in the central nervous system (CNS). (PDB code: 3KG2.)



for LTP induction. Such animals exhibit specific deficits in their learning abilities but behave almost normally otherwise.

How do NMDA receptors mediate LTP? The answer is that these channels, when open, are highly permeable to Ca<sup>2+</sup>, which acts as an intracellular signal in the postsynaptic cell, triggering a cascade of changes that are responsible for LTP. Thus, LTP is prevented when Ca<sup>2+</sup> levels are held artificially low in the postsynaptic cell by injecting the Ca<sup>2+</sup> chelator EGTA into it, and LTP can be induced by artificially raising intracellular Ca<sup>2+</sup> levels in the cell. Among the long-term changes that increase the sensitivity of the postsynaptic cell to glutamate is the insertion of new AMPA receptors into the plasma membrane (Figure 11–44). In some forms of LTP, changes occur in the presynaptic cell as well, so that it releases more glutamate than normal when it is activated subsequently.

If synapses were capable only of LTP they would quickly become saturated, and thus be of limited value as an information-storage device. In fact, they also exhibit **long-term depression (LTD)**, with the long-term effect of reducing the number of AMPA receptors in the post-synaptic membrane. This feat is accomplished by degrading AMPA receptors after their selective endocytosis. Surprisingly, LTD also requires NMDA receptor activation and a rise in Ca<sup>2+</sup>. How does Ca<sup>2+</sup> trigger opposite effects at the same synapse? It turns out that this bidirectional control of synaptic strength depends on the magnitude of the rise in Ca<sup>2+</sup>: high Ca<sup>2+</sup> levels activate protein kinases and LTP, whereas modest Ca<sup>2+</sup> levels activate protein phosphatases and LTD.

There is evidence that NMDA receptors have an important role in synaptic plasticity and learning in other parts of the brain, as well as in the hippocampus. Moreover, they have a crucial role in adjusting the anatomical pattern of synaptic connections in the light of experience during the development of the nervous system.

Thus, neurotransmitters released at synapses, besides relaying transient electrical signals, can also alter concentrations of intracellular mediators that bring about lasting changes in the efficacy of synaptic transmission. However, it is still uncertain how these changes endure for weeks, months, or a lifetime in the face of the normal turnover of cell constituents.

## Summary

*Ion channels form aqueous pores across the lipid bilayer and allow inorganic ions of appropriate size and charge to cross the membrane down their electrochemical gradients at rates about 1000 times greater than those achieved by any known transporter. The channels are “gated” and usually open transiently in response to a specific perturbation in the membrane, such as a change in membrane potential (voltage-gated channels), or the binding of a neurotransmitter to the channel (transmitter-gated channels).*

*K<sup>+</sup>-selective leak channels have an important role in determining the resting membrane potential across the plasma membrane in most animal cells.*

**Figure 11–44** The signaling events in long-term potentiation. Although not shown, transmission-enhancing changes can also occur in the presynaptic nerve terminals in LTP, which may be induced by retrograde signals from the postsynaptic cell.

Voltage-gated cation channels are responsible for the amplification and propagation of action potentials in electrically excitable cells, such as neurons and skeletal muscle cells. Transmitter-gated ion channels convert chemical signals to electrical signals at chemical synapses. Excitatory neurotransmitters, such as acetylcholine and glutamate, open transmitter-gated cation channels and thereby depolarize the postsynaptic membrane toward the threshold level for firing an action potential. Inhibitory neurotransmitters, such as GABA and glycine, open transmitter-gated  $\text{Cl}^-$  or  $\text{K}^+$  channels and thereby suppress firing by keeping the postsynaptic membrane polarized. A subclass of glutamate-gated ion channels, called NMDA-receptor channels, is highly permeable to  $\text{Ca}^{2+}$ , which can trigger the long-term changes in synapse efficacy (synaptic plasticity) such as LTP and LTD that are thought to be involved in some forms of learning and memory.

Ion channels work together in complex ways to control the behavior of electrically excitable cells. A typical neuron, for example, receives thousands of excitatory and inhibitory inputs, which combine by spatial and temporal summation to produce a combined postsynaptic potential (PSP) at the initial segment of its axon. The magnitude of the PSP is translated into the rate of firing of action potentials by a mixture of cation channels in the initial segment membrane.

## WHAT WE DON'T KNOW

- How do individual neurons establish and maintain their characteristic intrinsic firing properties?
- Even organisms with very simple nervous systems have dozens of different  $\text{K}^+$  channels. Why is it important to have so many?
- Why do cells that are not electrically active contain voltage-gated ion channels?
- How are memories stored for so many years in the human brain?

## PROBLEMS

Which statements are true? Explain why or why not.

**11–1** Transport by transporters can be either active or passive, whereas transport by channels is always passive.

**11–2** Transporters saturate at high concentrations of the transported molecule when all their binding sites are occupied; channels, on the other hand, do not bind the ions they transport and thus the flux of ions through a channel does not saturate.

**11–3** The membrane potential arises from movements of charge that leave ion concentrations practically unaffected, causing only a very slight discrepancy in the number of positive and negative ions on the two sides of the membrane.

Discuss the following problems.

**11–4** Order  $\text{Ca}^{2+}$ ,  $\text{CO}_2$ , ethanol, glucose, RNA, and  $\text{H}_2\text{O}$  according to their ability to diffuse through a lipid bilayer, beginning with the one that crosses the bilayer most readily. Explain your order.

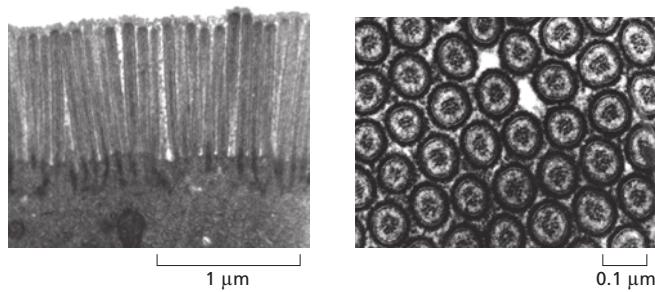
**11–5** How is it possible for some molecules to be at equilibrium across a biological membrane and yet not be at the same concentration on both sides?

**11–6** Ion transporters are “linked” together—not physically, but as a consequence of their actions. For example, cells can raise their intracellular pH, when it becomes too acidic, by exchanging external  $\text{Na}^+$  for internal  $\text{H}^+$ , using a  $\text{Na}^+-\text{H}^+$  antiporter. The change in internal  $\text{Na}^+$  is then redressed using the  $\text{Na}^+-\text{K}^+$  pump.

**A.** Can these two transporters, operating together, normalize both the  $\text{H}^+$  and the  $\text{Na}^+$  concentrations inside the cell?

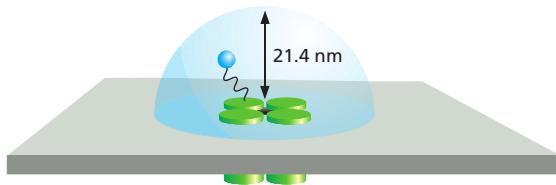
**B.** Does the linked action of these two pumps cause imbalances in either the  $\text{K}^+$  concentration or the membrane potential? Why or why not?

**11–7** Microvilli increase the surface area of intestinal cells, providing more efficient absorption of nutrients. Microvilli are shown in profile and cross section in **Figure Q11–1**. From the dimensions given in the figure, estimate the increase in surface area that microvilli provide (for the portion of the plasma membrane in contact with the lumen of the gut) relative to the corresponding surface of a cell with a “flat” plasma membrane.



**Figure Q11–1** Microvilli of intestinal epithelial cells in profile and cross section (Problem 11–7). (Left panel, from Rippel Electron Microscope Facility, Dartmouth College; right panel, from David Burgess.)

**11–8** According to Newton's laws of motion, an ion exposed to an electric field in a vacuum would experience a constant acceleration from the electric driving force, just as a falling body in a vacuum constantly accelerates due to gravity. In water, however, an ion moves at constant velocity in an electric field. Why do you suppose that is?



**Figure Q11-2** A “ball” tethered by a “chain” to a voltage-gated K<sup>+</sup> channel (Problem 11–9).

**11–9** In a subset of voltage-gated K<sup>+</sup> channels, the N-terminus of each subunit acts like a tethered ball that occludes the cytoplasmic end of the pore soon after it opens, thereby inactivating the channel. This “ball-and-chain” model for the rapid inactivation of voltage-gated K<sup>+</sup> channels has been elegantly supported for the *shaker* K<sup>+</sup> channel from *Drosophila melanogaster*. (The *shaker* K<sup>+</sup> channel in *Drosophila* is named after a mutant form that causes excitable behavior—even anesthetized flies keep twitching.) Deletion of the N-terminal amino acids from the normal *shaker* channel gives rise to a channel that opens in response to membrane depolarization, but stays open instead of rapidly closing as the normal channel does. A peptide (MAAVAGLYGLGEDRQHRKKQ) that corresponds to the deleted N-terminus can inactivate the open channel at 100 μM.

Is the concentration of free peptide (100 μM) that is required to inactivate the defective K<sup>+</sup> channel anywhere near the local concentration of the tethered ball on a normal channel? Assume that the tethered ball can explore a hemisphere [volume =  $(2/3)\pi r^3$ ] with a radius of 21.4 nm, which is the length of the polypeptide “chain” (Figure Q11-2). Calculate the concentration for one ball in this hemisphere. How does that value compare with the concentration of free peptide needed to inactivate the channel?

**11–10** The giant axon of the squid (Figure Q11-3) occupies a unique position in the history of our understanding of cell membrane potentials and nerve action. When an electrode is stuck into an intact giant axon, the membrane potential registers -70 mV. When the axon, suspended in a bath of seawater, is stimulated to conduct a nerve impulse, the membrane potential changes transiently from -70 mV to +40 mV.



**Figure Q11-3** The squid *Loligo* (Problem 11–10). This squid is about 15 cm in length.

**TABLE Q11-1** Ionic composition of seawater and of the cytosol in the squid giant axon (Problem 11–10).

Ion	Cytosol	Seawater
Na <sup>+</sup>	65 mM	430 mM
K <sup>+</sup>	344 mM	9 mM

For univalent ions and at 20°C (293 K), the Nernst equation reduces to

$$V = 58 \text{ mV} \times \log(C_0/C_i)$$

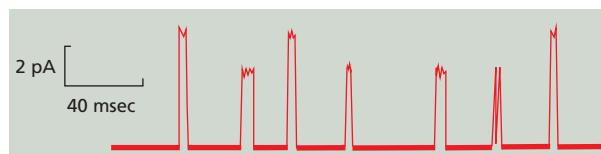
where C<sub>0</sub> and C<sub>i</sub> are the concentrations outside and inside, respectively.

Using this equation, calculate the potential across the resting membrane (1) assuming that it is due solely to K<sup>+</sup> and (2) assuming that it is due solely to Na<sup>+</sup>. (The Na<sup>+</sup> and K<sup>+</sup> concentrations in the axon cytosol and in seawater are given in Table Q11-1.) Which calculation is closer to the measured resting potential? Which calculation is closer to the measured action potential? Explain why these assumptions approximate the measured resting and action potentials.

**11–11** Acetylcholine-gated cation channels at the neuromuscular junction open in response to acetylcholine released by the nerve terminal and allow Na<sup>+</sup> ions to enter the muscle cell, which causes membrane depolarization and ultimately leads to muscle contraction.

**A.** Patch-clamp measurements show that young rat muscles have cation channels that respond to acetylcholine (Figure Q11-4). How many kinds of channel are there? How can you tell?

**B.** For each kind of channel, calculate the number of ions that enter in one millisecond. (One ampere is a current of one coulomb per second; one pA equals 10<sup>-12</sup> ampere. An ion with a single charge such as Na<sup>+</sup> carries a charge of 1.6 × 10<sup>-19</sup> coulomb.)



**Figure Q11-4** Patch-clamp measurements of acetylcholine-gated cation channels in young rat muscle (Problem 11–11).

## REFERENCES

### General

- Engel A & Gaub HE (2008) Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* 77, 127–148.
- Hille B (2001) Ionic Channels of Excitable Membranes, 3rd ed. Sunderland, MA: Sinauer.
- Stein WD (2014) Channels, Carriers, and Pumps: An Introduction to Membrane Transport, 2nd ed. San Diego, CA: Academic Press.
- Vinothkumar KR & Henderson R (2010) Structures of membrane proteins. *Q. Rev. Biophys.* 43, 65–158.

### Principles of Membrane Transport

- Al-Awqati Q (1999) One hundred years of membrane permeability: does Overton still rule? *Nat. Cell Biol.* 1, E201–E202.
- Forrest LR & Sansom MS (2000) Membrane simulations: bigger and better? *Curr. Opin. Struct. Biol.* 10, 174–181.
- Gouaux E & MacKinnon R (2005) Principles of selective ion transport in channels and pumps. *Science* 310, 1461–1465.
- Mitchell P (1977) Vectorial chemiosmotic processes. *Annu. Rev. Biochem.* 46, 996–1005.
- Tanford C (1983) Mechanism of free energy coupling in active transport. *Annu. Rev. Biochem.* 52, 379–409.

### Transporters and Active Membrane Transport

- Almers W & Stirling C (1984) Distribution of transport proteins over animal cell membranes. *J. Membr. Biol.* 77, 169–186.
- Baldwin SA & Henderson PJ (1989) Homologies between sugar transporters from eukaryotes and prokaryotes. *Annu. Rev. Physiol.* 51, 459–471.
- Doige CA & Ames GF (1993) ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. *Annu. Rev. Microbiol.* 47, 291–319.
- Forrest LR & Rudnick G (2009) The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology* 24, 377–386.
- Gadsby DC (2009) Ion channels versus ion pumps: the principal difference, in principle. *Nat. Rev. Mol. Cell Biol.* 10, 344–352.
- Higgins CF (2007) Multiple molecular mechanisms for multidrug resistance transporters. *Nature* 446, 749–757.
- Kaback HR, Sahin-Tóth M & Weinglass AB (2001) The kamikaze approach to membrane transport. *Nat. Rev. Mol. Cell Biol.* 2, 610–620.
- Kühlbrandt W (2004) Biology, structure and mechanism of P-type ATPases. *Nat. Rev. Mol. Cell Biol.* 5, 282–295.
- Lodish HF (1986) Anion-exchange and glucose transport proteins: structure, function, and distribution. *Harvey Lect.* 82, 19–46.
- Møller JV, Olesen C, Winther AML & Nissen P (2010) The sarcoplasmic  $\text{Ca}^{2+}$ -ATPase: design of a perfect chemi-osmotic pump. *Q. Rev. Biophys.* 43, 501–566.
- Perez C, Koshy C, Yıldız O & Ziegler C (2012) Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. *Nature* 490, 126–130.
- Rees D, Johnson E & Lewinson O (2009) ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* 10, 218–227.
- Romero MF & Boron WF (1999) Electrogenic  $\text{Na}^+/\text{HCO}_3^-$  cotransporters: cloning and physiology. *Annu. Rev. Physiol.* 61, 699–723.
- Rudnick G (2011) Cytoplasmic permeation pathway of neurotransmitter transporters. *Biochemistry* 50, 7462–7475.
- Saier MH Jr (2000) Vectorial metabolism and the evolution of transport systems. *J. Bacteriol.* 182, 5029–5035.
- Stein WD (2002) Cell volume homeostasis: ionic and nonionic mechanisms. The sodium pump in the emergence of animal cells. *Int. Rev. Cytol.* 215, 231–258.

Toyoshima C (2009) How  $\text{Ca}^{2+}$ -ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochim. Biophys. Acta* 1793, 941–946.

Yamashita A, Singh SK, Kawate T et al. (2005) Crystal structure of a bacterial homologue of  $\text{Na}^+/\text{Cl}^-$ -dependent neurotransmitter transporters. *Nature* 437, 215–223.

### Channels and the Electrical Properties of Membranes

- Armstrong C (1998) The vision of the pore. *Science* 280, 56–57.
- Arnadóttir J & Chalfie M (2010) Eukaryotic mechanosensitive channels. *Annu. Rev. Biophys.* 39, 111–137.
- Bezanilla F (2008) How membrane proteins sense voltage. *Nat. Rev. Mol. Cell Biol.* 9, 323–332.
- Catterall WA (2010) Ion channel voltage sensors: structure, function, and pathophysiology. *Neuron* 67, 915–928.
- Davis GW (2006) Homeostatic control of neural activity: from phenomenology to molecular design. *Annu. Rev. Neurosci.* 29, 307–323.
- Greengard P (2001) The neurobiology of slow synaptic transmission. *Science* 294, 1024–1030.
- Hodgkin AL & Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.
- Hodgkin AL & Huxley AF (1952) Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *J. Physiol.* 116, 449–472.
- Jessell TM & Kandel ER (1993) Synaptic transmission: a bidirectional and self-modifiable form of cell–cell communication. *Cell* 72(Suppl), 1–30.
- Julius D (2013) TRP channels and pain. *Annu. Rev. Cell Dev. Biol.* 29, 355–384.
- Katz B (1966) Nerve, Muscle and Synapse. New York: McGraw-Hill.
- King LS, Kozono D & Agre P (2004) From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.* 5, 687–698.
- Liao M, Cao E, Julius D & Cheng Y (2014) Single particle electron cryo-microscopy of a mammalian ion channel. *Curr. Opin. Struct. Biol.* 27, 1–7.
- MacKinnon R (2003) Potassium channels. *FEBS Lett.* 555, 62–65.
- Miesenböck G (2011) Optogenetic control of cells and circuits. *Annu. Rev. Cell Dev. Biol.* 27, 731–758.
- Moss SJ & Smart TG (2001) Constructing inhibitory synapses. *Nat. Rev. Neurosci.* 2, 240–250.
- Neher E & Sakmann B (1992) The patch clamp technique. *Sci. Am.* 266, 44–51.
- Nicholls JG, Fuchs PA, Martin AR & Wallace BG (2000) From Neuron to Brain, 4th ed. Sunderland, MA: Sinauer.
- Numa S (1987) A molecular view of neurotransmitter receptors and ionic channels. *Harvey Lect.* 83, 121–165.
- Payandeh J, Scheuer T, Zheng N & Catterall WA (2011) The crystal structure of a voltage-gated sodium channel. *Nature* 475, 353–358.
- Scannevin RH & Huganir RL (2000) Postsynaptic organization and regulation of excitatory synapses. *Nat. Rev. Neurosci.* 1, 133–141.
- Snyder SH (1996) Drugs and the Brain. New York: WH Freeman/Scientific American Books.
- Sobolevsky AI, Rosconi MP & Gouaux E (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462, 745–756.
- Stevens CF (2004) Presynaptic function. *Curr. Opin. Neurobiol.* 14, 341–345.
- Verkman AS (2013) Aquaporins. *Curr. Biol.* 23, R52–R55.

# Intracellular Compartments and Protein Sorting

CHAPTER  
**12**

Unlike a bacterium, which generally consists of a single intracellular compartment surrounded by a plasma membrane, a eukaryotic cell is elaborately subdivided into functionally distinct, membrane-enclosed compartments. Each compartment, or **organelle**, contains its own characteristic set of enzymes and other specialized molecules, and complex distribution systems transport specific products from one compartment to another. To understand the eukaryotic cell, it is essential to know how the cell creates and maintains these compartments, what occurs in each of them, and how molecules move between them.

Proteins confer upon each compartment its characteristic structural and functional properties. They catalyze the reactions that occur there and selectively transport small molecules into and out of the compartment. For membrane-enclosed organelles in the cytoplasm, proteins also serve as organelle-specific surface markers that direct new deliveries of proteins and lipids to the appropriate organelle.

An animal cell contains about 10 billion ( $10^{10}$ ) protein molecules of perhaps 10,000 kinds, and the synthesis of almost all of them begins in the **cytosol**, the space of the cytoplasm outside the membrane-enclosed organelles. Each newly synthesized protein is then delivered specifically to the organelle that requires it. The intracellular transport of proteins is the central theme of both this chapter and the next. By tracing the protein traffic from one compartment to another, one can begin to make sense of the otherwise bewildering maze of intracellular membranes.

## THE COMPARTMENTALIZATION OF CELLS

In this brief overview of the compartments of the cell and the relationships between them, we organize the organelles conceptually into a small number of discrete families, discuss how proteins are directed to specific organelles, and explain how proteins cross organelle membranes.

### All Eukaryotic Cells Have the Same Basic Set of Membrane-enclosed Organelles

Many vital biochemical processes take place in membranes or on their surfaces. Membrane-bound enzymes, for example, catalyze lipid metabolism; and oxidative phosphorylation and photosynthesis both require a membrane to couple the transport of  $H^+$  to the synthesis of ATP. In addition to providing increased membrane area to host biochemical reactions, intracellular membrane systems form enclosed compartments that are separate from the cytosol, thus creating functionally specialized aqueous spaces within the cell. In these spaces, subsets of molecules (proteins, reactants, ions) are concentrated to optimize the biochemical reactions in which they participate. Because the lipid bilayer of cell membranes is impermeable to most hydrophilic molecules, the membrane of an organelle must contain membrane transport proteins to import and export specific metabolites. Each organelle membrane must also have a mechanism for importing, and incorporating into the organelle, the specific proteins that make the organelle unique.

### IN THIS CHAPTER

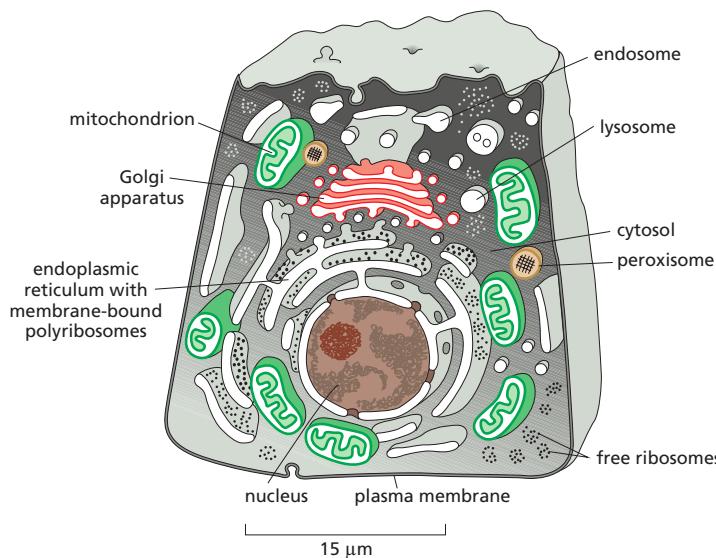
THE COMPARTMENTALIZATION OF CELLS

THE TRANSPORT OF MOLECULES BETWEEN THE NUCLEUS AND THE CYTOSOL

THE TRANSPORT OF PROTEINS INTO MITOCHONDRIA AND CHLOROPLASTS

PEROXISOMES

THE ENDOPLASMIC RETICULUM



**Figure 12–1** The major intracellular compartments of an animal cell. The cytosol (gray), endoplasmic reticulum, Golgi apparatus, nucleus, mitochondrion, endosome, lysosome, and peroxisome are distinct compartments isolated from the rest of the cell by at least one selectively permeable membrane (see Movie 9.2).

**Figure 12–1** illustrates the major intracellular compartments common to eukaryotic cells. The *nucleus* contains the genome (aside from mitochondrial and chloroplast DNA), and it is the principal site of DNA and RNA synthesis. The surrounding *cytoplasm* consists of the cytosol and the cytoplasmic organelles suspended in it. The cytosol constitutes a little more than half the total volume of the cell, and it is the main site of protein synthesis and degradation. It also performs most of the cell's *intermediary metabolism*—that is, the many reactions that degrade some small molecules and synthesize others to provide the building blocks for macromolecules (discussed in Chapter 2).

About half the total area of membrane in a eukaryotic cell encloses the labyrinthine spaces of the *endoplasmic reticulum (ER)*. The *rough ER* has many ribosomes bound to its cytosolic surface. Ribosomes are organelles that are not membrane-enclosed; they synthesize both soluble and integral membrane proteins, most of which are destined either for secretion to the cell exterior or for other organelles. We shall see that, whereas proteins are transported into other membrane-enclosed organelles only after their synthesis is complete, they are transported into the ER as they are synthesized. This explains why the ER membrane is unique in having ribosomes tethered to it. The ER also produces most of the lipid for the rest of the cell and functions as a store for  $\text{Ca}^{2+}$  ions. Regions of the ER that lack bound ribosomes are called *smooth ER*. The ER sends many of its proteins and lipids to the *Golgi apparatus*, which often consists of organized stacks of disc-like compartments called *Golgi cisternae*. The Golgi apparatus receives lipids and proteins from the ER and dispatches them to various destinations, usually covalently modifying them *en route*.

*Mitochondria* and *chloroplasts* generate most of the ATP that cells use to drive reactions requiring an input of free energy; chloroplasts are a specialized version of *plastids* (present in plants, algae, and some protozoa), which can also have other functions, such as the storage of food or pigment molecules. *Lysosomes* contain digestive enzymes that degrade defunct intracellular organelles, as well as macromolecules and particles taken in from outside the cell by endocytosis. On the way to lysosomes, endocytosed material must first pass through a series of organelles called *endosomes*. Finally, *peroxisomes* are small vesicular compartments that contain enzymes used in various oxidative reactions.

In general, each membrane-enclosed organelle performs the same set of basic functions in all cell types. But to serve the specialized functions of cells, these organelles vary in abundance and can have additional properties that differ from cell type to cell type.

On average, the membrane-enclosed compartments together occupy nearly half the volume of a cell (**Table 12–1**), and a large amount of intracellular membrane is required to make them. In liver and pancreatic cells, for example, the

endoplasmic reticulum has a total membrane surface area that is, respectively, 25 times and 12 times that of the plasma membrane (**Table 12–2**). The membrane-enclosed organelles are packed tightly in the cytoplasm, and, in terms of area and mass, the plasma membrane is only a minor membrane in most eukaryotic cells (**Figure 12–2**).

The abundance and shape of membrane-enclosed organelles are regulated to meet the needs of the cell. This is particularly apparent in cells that are highly specialized and therefore disproportionately rely on specific organelles. Plasma cells, for example, which secrete their own weight every day in antibody molecules into the bloodstream, contain vastly amplified amounts of rough ER, which is found in large, flat sheets. Cells that specialize in lipid synthesis also expand their ER, but in this case the organelle forms a network of convoluted tubules. Moreover, membrane-enclosed organelles are often found in characteristic positions in the cytoplasm. In most cells, for example, the Golgi apparatus is located close to the nucleus, whereas the network of ER tubules extends from the nucleus throughout the entire cytosol. These characteristic distributions depend on interactions of the organelles with the cytoskeleton. The localization of both the ER and the Golgi apparatus, for instance, depends on an intact microtubule array; if the microtubules are experimentally depolymerized with a drug, the Golgi apparatus fragments and disperses throughout the cell, and the ER network collapses toward the cell center (discussed in Chapter 16). The size, shape, composition, and location are all important and regulated features of these organelles that ultimately contribute to the organelle's function.

**TABLE 12–1 Relative Volumes Occupied by the Major Intracellular Compartments in a Liver Cell (Hepatocyte)**

Intracellular compartment	Percentage of total cell volume
Cytosol	54
Mitochondria	22
Rough ER cisternae	9
Smooth ER cisternae plus Golgi cisternae	6
Nucleus	6
Peroxisomes	1
Lysosomes	1
Endosomes	1

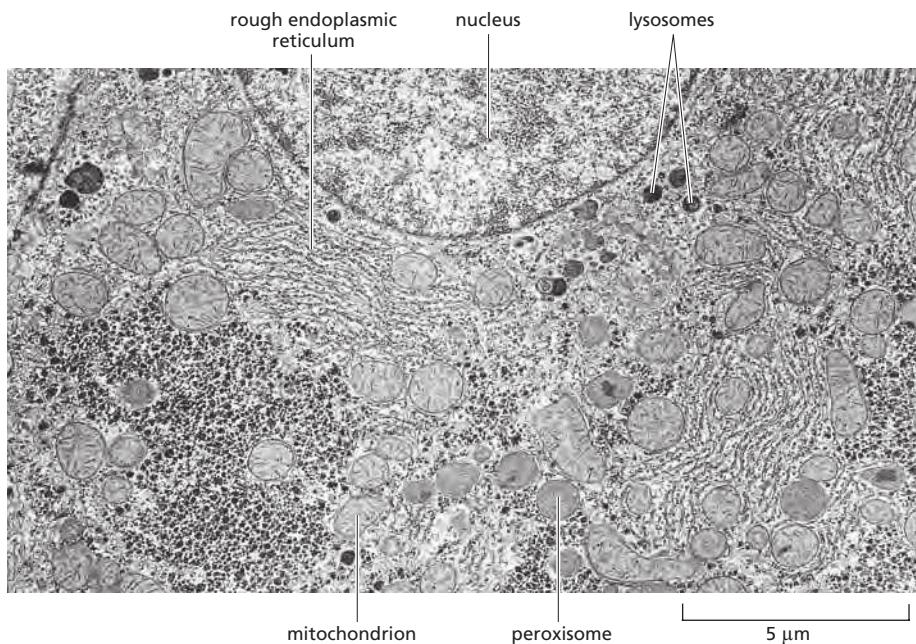
### Evolutionary Origins May Help Explain the Topological Relationships of Organelles

To understand the relationships between the compartments of the cell, it is helpful to consider how they might have evolved. The precursors of the first eukaryotic cells are thought to have been relatively simple cells that—like most bacterial and

**TABLE 12–2 Relative Amounts of Membrane Types in Two Kinds of Eukaryotic Cells**

Membrane Type	Percentage of total cell membrane	
	Liver hepatocyte*	Pancreatic exocrine cell*
Plasma membrane	2	5
Rough ER membrane	35	60
Smooth ER membrane	16	<1
Golgi apparatus membrane	7	10
Mitochondria		
Outer membrane	7	4
Inner membrane	32	17
Nucleus		
Inner membrane	0.2	0.7
Secretory vesicle membrane	Not determined	3
Lysosome membrane	0.4	Not determined
Peroxisome membrane	0.4	Not determined
Endosome membrane	0.4	Not determined

\*These two cells are of very different sizes: the average hepatocyte has a volume of about  $5000 \mu\text{m}^3$  compared with  $1000 \mu\text{m}^3$  for the pancreatic exocrine cell. Total cell membrane areas are estimated at about  $110,000 \mu\text{m}^2$  and  $13,000 \mu\text{m}^2$ , respectively.

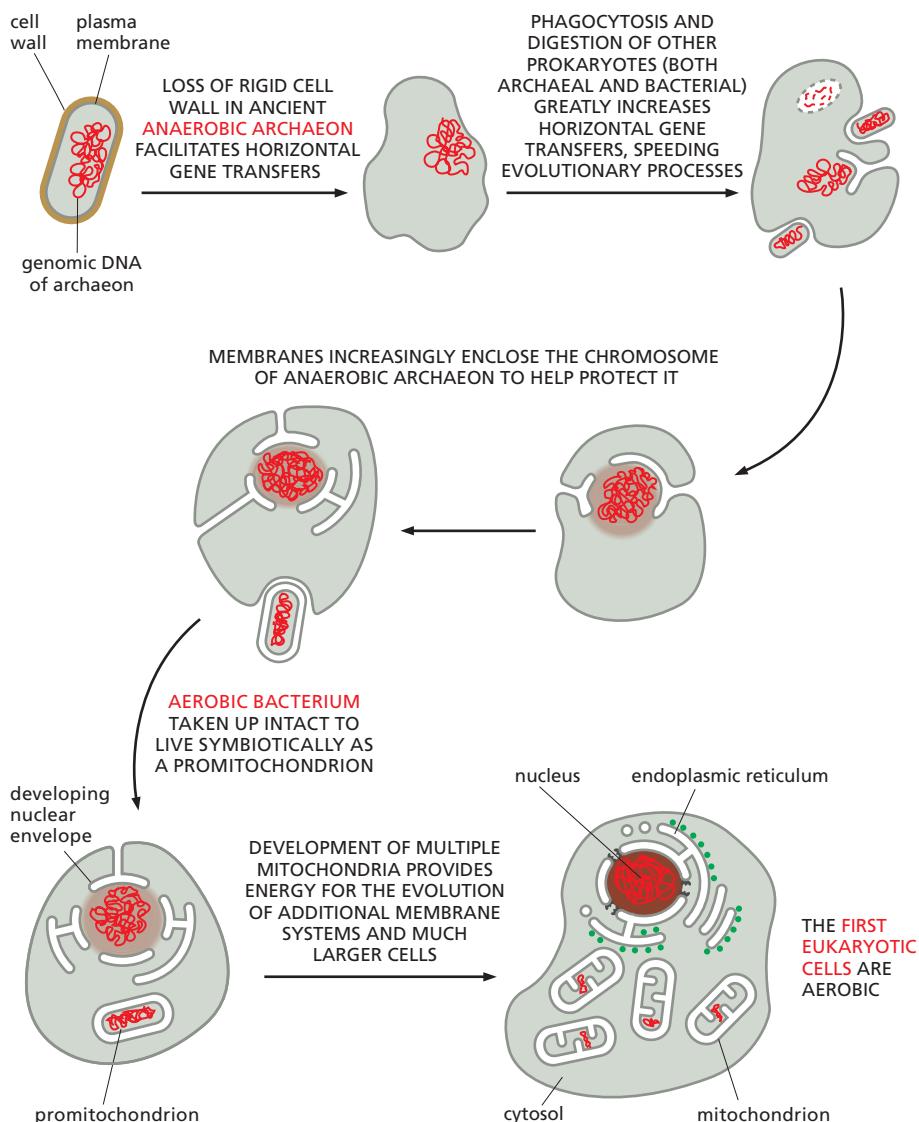


**Figure 12–2** An electron micrograph of part of a liver cell seen in cross section. Examples of most of the major intracellular organelles are indicated. (Courtesy of Daniel S. Friend.)

archaeal cells—have a plasma membrane but no internal membranes. The plasma membrane in such cells provides all membrane-dependent functions, including the pumping of ions, ATP synthesis, protein secretion, and lipid synthesis. Typical present-day eukaryotic cells are 10–30 times larger in linear dimension and 1000–10,000 times greater in volume than a typical bacterium such as *E. coli*. The profusion of internal membranes can be regarded, in part, as an adaptation to this increase in size: the eukaryotic cell has a much smaller ratio of surface area to volume, and its plasma membrane therefore presumably has too small an area to sustain the many vital functions that membranes perform. The extensive internal membrane systems of a eukaryotic cell alleviate this problem.

The evolution of internal membranes evidently went hand-in-hand with the specialization of membrane function. A hypothetical scheme for how the first eukaryotic cells, with a nucleus and ER, might have evolved by the invagination and pinching off of the plasma membrane of an ancestral cell is illustrated in Figure 12–3. This process would create membrane-enclosed organelles with an interior or **lumen** that is topologically equivalent to the exterior of the cell. We shall see that this topological relationship holds for all of the organelles involved in the secretory and endocytic pathways, including the ER, Golgi apparatus, endosomes, lysosomes, and peroxisomes. We can therefore think of all of these organelles as members of the same topologically equivalent compartment. As we discuss in detail in the next chapter, their interiors communicate extensively with one another and with the outside of the cell via *transport vesicles*, which bud off from one organelle and fuse with another (Figure 12–4).

As described in Chapter 14, mitochondria and plastids differ from the other membrane-enclosed organelles because they contain their own genomes. The nature of these genomes, and the close resemblance of the proteins in these organelles to those in some present-day bacteria, strongly suggest that mitochondria and plastids evolved from bacteria that were engulfed by other cells with which they initially lived in symbiosis (see Figures 1–29 and 1–31): the inner membrane of mitochondria and plastids presumably corresponds to the original plasma membrane of the bacterium, while the lumen of these organelles evolved from the bacterial cytosol. Like the bacteria from which they were derived, both mitochondria and plastids are enclosed by a double membrane and they remain isolated from the extensive vesicular traffic that connects the interiors of most of the other membrane-enclosed organelles to each other and to the outside of the cell.

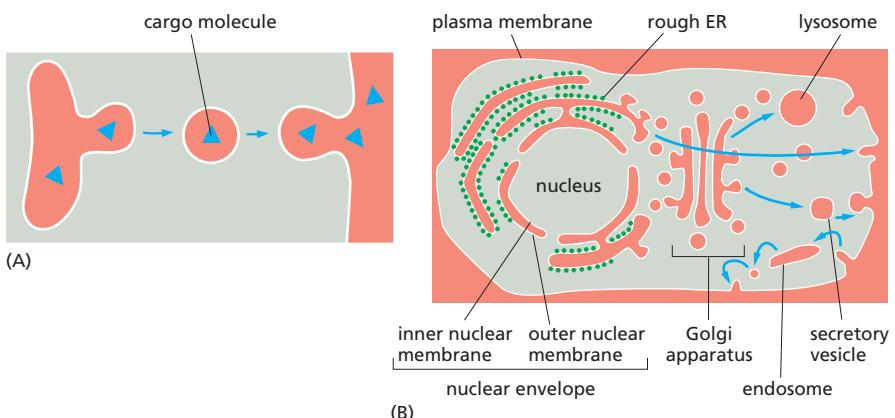


**Figure 12–3 One suggested pathway for the evolution of the eukaryotic cell and its internal membranes** As discussed in Chapter 1, there is evidence that the nuclear genome of a eukaryotic cell evolved from an ancient archaeon. For example, clear homologs of actin, tubulin, histones, and the nuclear DNA replication system are found in archaea, but not in bacteria. Thus, it is now thought that the first eukaryotic cells arose when an ancient anaerobic archaeon joined forces with an aerobic bacterium roughly 1.6 billion years ago. As indicated, the nuclear envelope may have originated from an invagination of the plasma membrane of this ancient archaeon—an invagination that protected its chromosome while still allowing access of the DNA to the cytosol (as required for DNA to direct protein synthesis). This envelope may have later pinched off completely from the plasma membrane, so as to produce a separate nuclear compartment surrounded by a double membrane. Because this double membrane is penetrated by nuclear pore complexes, the nuclear compartment is topologically equivalent to the cytosol. In contrast, the lumen of the ER is continuous with the space between the inner and outer nuclear membranes, and it is topologically equivalent to the extracellular space (see Figure 12–4). (Adapted from J. Martijn and T.J.G. Ettema, *Biochem. Soc. Trans.* 41: 451–457, 2013.)

The evolutionary schemes just described group the intracellular compartments in eukaryotic cells into four distinct families: (1) the nucleus and the cytosol, which communicate with each other through *nuclear pore complexes* and are thus topologically continuous (although functionally distinct); (2) all organelles that function in the secretory and endocytic pathways—including the ER, Golgi apparatus, endosomes, and lysosomes, the numerous classes of transport intermediates such as transport vesicles that move between them, and peroxisomes; (3) the mitochondria; and (4) the plastids (in plants only).

### Proteins Can Move Between Compartments in Different Ways

The synthesis of all proteins begins on ribosomes in the cytosol, except for the few that are synthesized on the ribosomes of mitochondria and plastids. Their subsequent fate depends on their amino acid sequence, which can contain **sorting signals** that direct their delivery to locations outside the cytosol or to organelle surfaces. Some proteins do not have a sorting signal and consequently remain in the cytosol as permanent residents. Many others, however, have specific sorting signals that direct their transport from the cytosol into the nucleus, the ER, mitochondria, plastids, or peroxisomes; sorting signals can also direct the transport of proteins from the ER to other destinations in the cell.



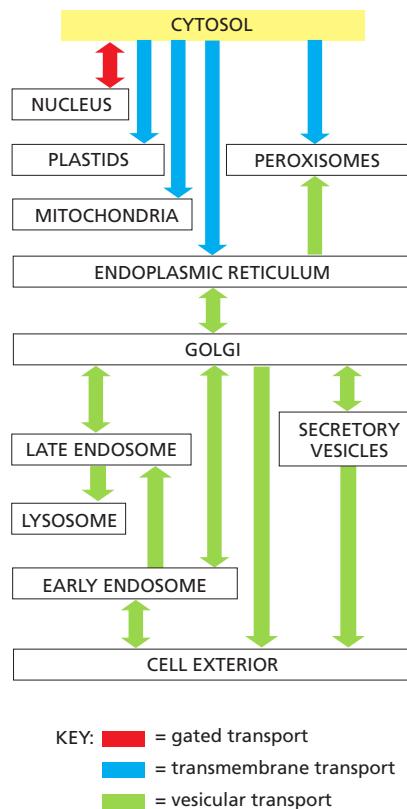
To understand the general principles by which sorting signals operate, it is important to distinguish three fundamentally different ways by which proteins move from one compartment to another. These three mechanisms are described below, and the transport steps at which they operate are outlined in **Figure 12–5**. We discuss the first two mechanisms (gated transport and transmembrane transport) in this chapter, and the third (vesicular transport, *green arrows* in Figure 12–5) in Chapter 13.

1. In **gated transport**, proteins and RNA molecules move between the cytosol and the nucleus through nuclear pore complexes in the nuclear envelope. The nuclear pore complexes function as selective gates that support the active transport of specific macromolecules and macromolecular assemblies between the two topologically equivalent spaces, although they also allow free diffusion of smaller molecules.
2. In **protein translocation**, transmembrane *protein translocators* directly transport specific proteins across a membrane from the cytosol into a space that is topologically distinct. The transported protein molecule usually must unfold to snake through the translocator. The initial transport of selected proteins from the cytosol into the ER lumen or mitochondria, for example, occurs in this way. Integral membrane proteins often use the same translocators but translocate only partially across the membrane, so that the protein becomes embedded in the lipid bilayer.
3. In **vesicular transport**, membrane-enclosed transport intermediates—which may be small, spherical transport vesicles or larger, irregularly shaped organelle fragments—ferry proteins from one topologically equivalent compartment to another. The transport vesicles and fragments become loaded with a cargo of molecules derived from the lumen of one compartment as they bud and pinch off from its membrane; they discharge their cargo into a second compartment by fusing with the membrane enclosing that compartment (**Figure 12–6**). The transfer of soluble proteins from the ER to the Golgi apparatus, for example, occurs in this way. Because the

**Figure 12–5 A simplified “roadmap” of protein traffic within a eukaryotic cell.** Proteins can move from one compartment to another by gated transport (red), protein translocation (blue), or vesicular transport (green). The sorting signals that direct a given protein's movement through the system, and thereby determine its eventual location in the cell, are contained in each protein's amino acid sequence. The journey begins with the synthesis of a protein on a ribosome in the cytosol and, for many proteins, terminates when the protein reaches its final destination. Other proteins shuttle back and forth between the nucleus and cytosol. At each intermediate station (boxes), a decision is made as to whether the protein is to be retained in that compartment or transported further. A sorting signal may direct either retention in or exit from a compartment.

We shall refer to this figure often as a guide in this chapter and the next, highlighting in color the particular pathway being discussed.

**Figure 12–4 Topologically equivalent compartments in the secretory and endocytic pathways in a eukaryotic cell.** Compartments are said to be *topologically equivalent* if they can communicate with one another, in the sense that molecules can get from one to the other without having to cross a membrane. Topologically equivalent spaces are shown in red. (A) Molecules can be carried from one compartment to another topologically equivalent compartment by vesicles that bud from one and fuse with the other. (B) In principle, cycles of membrane budding and fusion permit the lumen of any of the organelles shown to communicate with any other and with the cell exterior by means of transport vesicles. Blue arrows indicate the extensive outbound and inbound vesicular traffic (discussed in Chapter 13). Some organelles, most notably mitochondria and (in plant cells) plastids, do not take part in this communication and are isolated from the vesicular traffic between organelles shown here.



**Figure 12–6 Vesicle budding and fusion during vesicular transport.**

Transport vesicles bud from one compartment (donor) and fuse with another topologically equivalent (target) compartment. In the process, soluble components (red dots) are transferred from lumen to lumen. Note that membrane is also transferred and that the original orientation of both proteins and lipids in the donor compartment membrane is preserved in the target compartment membrane. Thus, membrane proteins retain their asymmetric orientation, with the same domains always facing the cytosol.

transported proteins do not cross a membrane, vesicular transport can move proteins only between compartments that are topologically equivalent (see Figure 12–4).

Each mode of protein transfer is usually guided by sorting signals in the transported protein, which are recognized by complementary *sorting receptors*. If a large protein is to be imported into the nucleus, for example, it must possess a sorting signal that receptor proteins recognize to guide it through the nuclear pore complex. If a protein is to be transferred directly across a membrane, it must possess a sorting signal that the translocator recognizes. Likewise, if a protein is to be loaded into a certain type of vesicle or retained in certain organelles, a complementary receptor in the appropriate membrane must recognize its sorting signal.

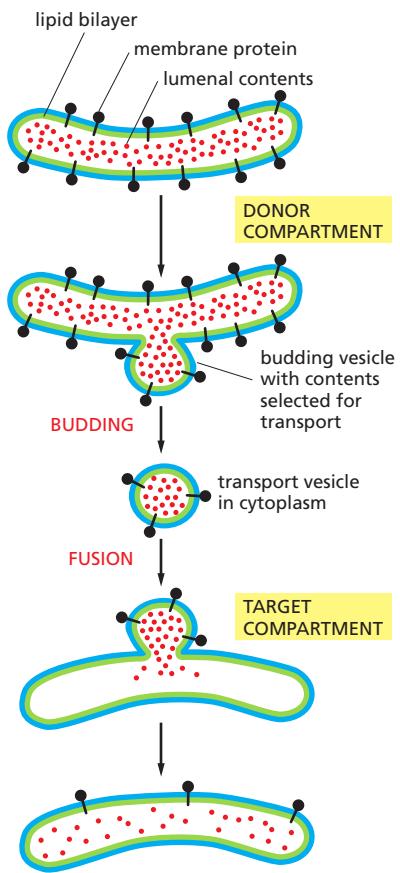
### Signal Sequences and Sorting Receptors Direct Proteins to the Correct Cell Address

Most protein sorting signals involved in transmembrane transport reside in a stretch of amino acid sequence, typically 15–60 residues long. Such **signal sequences** are often found at the N-terminus of the polypeptide chain, and in many cases specialized **signal peptidases** remove the signal sequence from the finished protein once the sorting process is complete. Signal sequences can also be internal stretches of amino acids, which remain part of the protein. Such signals are used in gated transport into the nucleus. Sorting signals can also be composed of multiple internal amino acid sequences that form a specific three-dimensional arrangement of atoms on the protein's surface; such **signal patches** are sometimes used for nuclear import and in vesicular transport.

Each signal sequence specifies a particular destination in the cell. Proteins destined for initial transfer to the ER usually have a signal sequence at their N-terminus that characteristically includes a sequence composed of about 5–10 hydrophobic amino acids. Many of these proteins will in turn pass from the ER to the Golgi apparatus, but those with a specific signal sequence of four amino acids at their C-terminus are recognized as ER residents and are returned to the ER. Proteins destined for mitochondria have signal sequences of yet another type, in which positively charged amino acids alternate with hydrophobic ones. Finally, many proteins destined for peroxisomes have a signal sequence of three characteristic amino acids at their C-terminus.

**Table 12–3** presents some specific signal sequences. Experiments in which the peptide is transferred from one protein to another by genetic engineering techniques have demonstrated the importance of each of these signal sequences for protein targeting. Placing the N-terminal ER signal sequence at the beginning of a cytosolic protein, for example, redirects the protein to the ER; removing or mutating the signal sequence of an ER protein causes its retention in the cytosol. Signal sequences are therefore both necessary and sufficient for protein targeting. Even though their amino acid sequences can vary greatly, the signal sequences of proteins having the same destination are functionally interchangeable; physical properties, such as hydrophobicity, often seem to be more important in the signal-recognition process than the exact amino acid sequence.

Signal sequences are recognized by complementary sorting receptors that guide proteins to their appropriate destination, where the receptors unload their cargo. The receptors function catalytically: after completing one round of targeting, they return to their point of origin to be reused. Most sorting receptors



**TABLE 12–3 Some Typical Signal Sequences**

Function of signal sequence	Example of signal sequence
Import into nucleus	-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-
Export from nucleus	-Met-Glu-Glu-Leu-Ser-Gln-Ala-Leu-Ala-Ser-Ser-Phe-
Import into mitochondria	<sup>+</sup> H <sub>3</sub> N-Met-Leu-Ser-Leu-Arg-Gln-Ser-Ile-Arg-Phe-Phe-Lys-Pro-Ala-Thr-Arg-Thr-Leu-Cys-Ser-Ser-Arg-Tyr-Leu-Leu-
Import into plastid	<sup>+</sup> H <sub>3</sub> N-Met-Val-Ala-Met-Ala-Ser-Leu-Gln-Ser-Ser-Met-Ser-Ser-Leu-Ser-Ser-Asn-Ser-Phe-Leu-Gly-Gln-Pro-Leu-Ser-Pro-Ile-Thr-Leu-Ser-Pro-Phe-Leu-Gln-Gly-
Import into peroxisomes	-Ser-Lys-Leu-COO-
Import into ER	<sup>+</sup> H <sub>3</sub> N-Met-Met-Ser-Phe-Val-Ser-Leu-Leu-Leu-Val-Gly-Ile-Leu-Phe-Trp-Ala-Thr-Glu-Ala-Glu-Gln-Leu-Thr-Lys-Cys-Glu-Val-Phe-Gln-
Return to ER	-Lys-Asp-Glu-Leu-COO-

Some characteristic features of the different classes of signal sequences are highlighted in color. Where they are known to be important for the function of the signal sequence, positively charged amino acids are shown in red and negatively charged amino acids are shown in green. Similarly, important hydrophobic amino acids are shown in orange and important hydroxylated amino acids are shown in blue. <sup>+</sup>H<sub>3</sub>N indicates the N-terminus of a protein; COO<sup>−</sup> indicates the C-terminus.

recognize classes of proteins rather than an individual protein species. They can therefore be viewed as public transportation systems, dedicated to delivering numerous different components to their correct location in the cell.

### Most Organelles Cannot Be Constructed *De Novo*: They Require Information in the Organelle Itself

When a cell reproduces by division, it has to duplicate its organelles, in addition to its chromosomes. In general, cells do this by incorporating new molecules into the existing organelles, thereby enlarging them; the enlarged organelles then divide and are distributed to the two daughter cells. Thus, each daughter cell inherits a complete set of specialized cell membranes from its mother. This inheritance is essential because a cell could not make such membranes from scratch. If the ER were completely removed from a cell, for example, how could the cell reconstruct it? As we discuss later, the membrane proteins that define the ER and perform many of its functions are themselves products of the ER. A new ER could not be made without an existing ER or, at least, a membrane that specifically contains the protein translocators required to import selected proteins into the ER from the cytosol (including the ER-specific translocators themselves). The same is true for mitochondria and plastids.

Thus, it seems that the information required to construct an organelle does not reside exclusively in the DNA that specifies the organelle's proteins. Information in the form of at least one distinct protein that preexists in the organelle membrane is also required, and this information is passed from parent cell to daughter cells in the form of the organelle itself. Presumably, such information is essential for the propagation of the cell's compartmental organization, just as the information in DNA is essential for the propagation of the cell's nucleotide and amino acid sequences.

As we discuss in more detail in Chapter 13, however, the ER buds off a constant stream of transport vesicles that incorporate only a subset of ER proteins and therefore have a composition different from the ER itself. Similarly, the plasma membrane constantly buds off various types of specialized endocytic vesicles. Thus, some organelles can form from other organelles and do not have to be inherited at cell division.

## Summary

Eukaryotic cells contain intracellular membrane-enclosed organelles that make up nearly half the cell's total volume. The main ones present in all eukaryotic cells are the endoplasmic reticulum, Golgi apparatus, nucleus, mitochondria, lysosomes, endosomes, and peroxisomes; plant cells also contain plastids such as chloroplasts. These organelles contain distinct sets of proteins, which mediate each organelle's unique function.

Each newly synthesized organelle protein must find its way from a ribosome in the cytosol, where the protein is made, to the organelle where it functions. It does so by following a specific pathway, guided by sorting signals in its amino acid sequence that function as either signal sequences or signal patches. Sorting signals are recognized by complementary sorting receptors, which deliver the protein to the appropriate target organelle. Proteins that function in the cytosol do not contain sorting signals and therefore remain there after they are synthesized.

During cell division, organelles such as the ER and mitochondria are distributed to each daughter cell. These organelles contain information that is required for their construction, and so they cannot be made de novo.

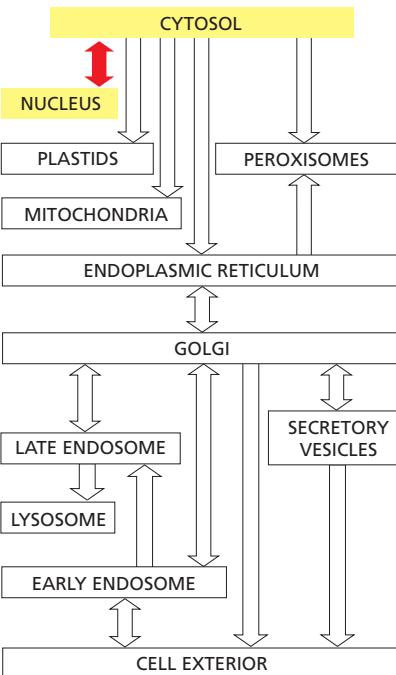
## THE TRANSPORT OF MOLECULES BETWEEN THE NUCLEUS AND THE CYTOSOL

The **nuclear envelope** encloses the DNA and defines the *nuclear compartment*. This envelope consists of two concentric membranes, which are penetrated by nuclear pore complexes (Figure 12–7). Although the inner and outer nuclear membranes are continuous, they maintain distinct protein compositions. The **inner nuclear membrane** contains proteins that act as binding sites for chromosomes and for the *nuclear lamina*, a protein meshwork that provides structural support for the nuclear envelope; the lamina also acts as an anchoring site for chromosomes and the cytoplasmic cytoskeleton (via protein complexes that span the nuclear envelope). The inner membrane is surrounded by the **outer nuclear membrane**, which is continuous with the membrane of the ER. Like the ER membrane (discussed later), the outer nuclear membrane is studded with ribosomes engaged in protein synthesis. The proteins made on these ribosomes are transported into the space between the inner and outer nuclear membranes (the *peri-nuclear space*), which is continuous with the ER lumen (see Figure 12–7).

Bidirectional traffic occurs continuously between the cytosol and the nucleus. The many proteins that function in the nucleus—including histones, DNA polymerases, RNA polymerases, transcriptional regulators, and RNA-processing proteins—are selectively imported into the nuclear compartment from the cytosol, where they are made. At the same time, almost all RNAs—including mRNAs, rRNAs, tRNAs, miRNAs, and snRNAs—are synthesized in the nuclear compartment and then exported to the cytosol. Like the import process, the export process is selective; mRNAs, for example, are exported only after they have been properly modified by RNA-processing reactions in the nucleus. In some cases, the transport process is complex. Ribosomal proteins, for instance, are made in the cytosol and imported into the nucleus, where they assemble with newly made ribosomal RNA into particles. The particles are then exported to the cytosol, where they assemble into ribosomes. Each of these steps requires selective transport across the nuclear envelope.

### Nuclear Pore Complexes Perforate the Nuclear Envelope

Large and elaborate **nuclear pore complexes** (NPCs) perforate the nuclear envelope in all eukaryotes. Each NPC is composed of a set of approximately 30 different proteins, or **nucleoporins**. Reflecting the high degree of internal symmetry, each nucleoporin is present in multiple copies, resulting in 500–1000 protein molecules in the fully assembled NPC, with an estimated mass of 66 million daltons in yeast and 125 million daltons in vertebrates (Figure 12–8). Most nucleoporins are



composed of repetitive protein domains of only a few different types, which have evolved through extensive gene duplication. Some of the scaffold nucleoporins (see Figure 12–8) are structurally related to vesicle coat protein complexes, such as clathrin and COPII coatomer (discussed in Chapter 13), which shape transport vesicles; and one protein is used as a common building block in both NPCs and vesicle coats. These similarities suggest a common evolutionary origin for NPCs and vesicle coats: they may derive from an early membrane-bending protein module that helped shape the elaborate membrane systems of eukaryotic cells, and in present-day cells stabilize the sharp membrane bends required to form a nuclear pore.

The nuclear envelope of a typical mammalian cell contains 3000–4000 NPCs, although that number varies widely, from a few hundred in glial cells to almost 20,000 in Purkinje neurons. The total traffic that passes through each NPC is enormous: each NPC can transport up to 1000 macromolecules per second and can transport in both directions at the same time. How it coordinates the bidirectional flow of macromolecules to avoid congestion and head-on collisions is not known.

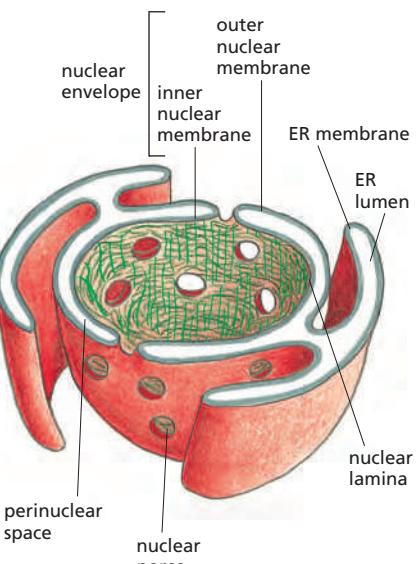
Each NPC contains aqueous passages, through which small water-soluble molecules can diffuse passively. Researchers have determined the effective size of these passages by injecting labeled water-soluble molecules of different sizes into the cytosol and then measuring their rate of diffusion into the nucleus. Small molecules (5000 daltons or less) diffuse in so fast that we can consider the nuclear envelope freely permeable to them. Large proteins, however, diffuse in much more slowly, and the larger a protein, the more slowly it passes through the NPC. Proteins larger than 60,000 daltons cannot enter by passive diffusion. This size cut-off to free diffusion is thought to result from the NPC structure (see Figure 12–8). The channel nucleoporins with extensive unstructured regions form a disordered tangle (much like a kelp bed in the ocean) that restricts the diffusion of large macromolecules while allowing smaller molecules to pass.

Because many cell proteins are too large to diffuse passively through the NPCs, the nuclear compartment and the cytosol can maintain different protein compositions. Mature cytosolic ribosomes, for example, are about 30 nm in diameter and thus cannot diffuse through the NPC, confining protein synthesis to the cytosol. But how does the nucleus export newly made ribosomal subunits or import large molecules, such as DNA polymerases and RNA polymerases, which have subunit molecular masses of 100,000–200,000 daltons? As we discuss next, these and most other transported protein and RNA molecules bind to specific receptor proteins that actively ferry large molecules through NPCs. Even small proteins like histones frequently use receptor-mediated mechanisms to cross the NPC, thereby increasing transport efficiency.

## Nuclear Localization Signals Direct Nuclear Proteins to the Nucleus

When proteins are experimentally extracted from the nucleus and reintroduced into the cytosol, even the very large ones reaccumulate efficiently in the nucleus. Sorting signals called **nuclear localization signals (NLSs)** are responsible for the selectivity of this active nuclear import process. The signals have been precisely defined by using recombinant DNA technology for numerous nuclear proteins, as well as for proteins that enter the nucleus only transiently (Figure 12–9). In many nuclear proteins, the signals consist of one or two short sequences that are rich in the positively charged amino acids lysine and arginine (see Table 12–3, p. 648), with the precise sequence varying for different proteins. Other nuclear proteins contain different signals, some of which are not yet characterized.

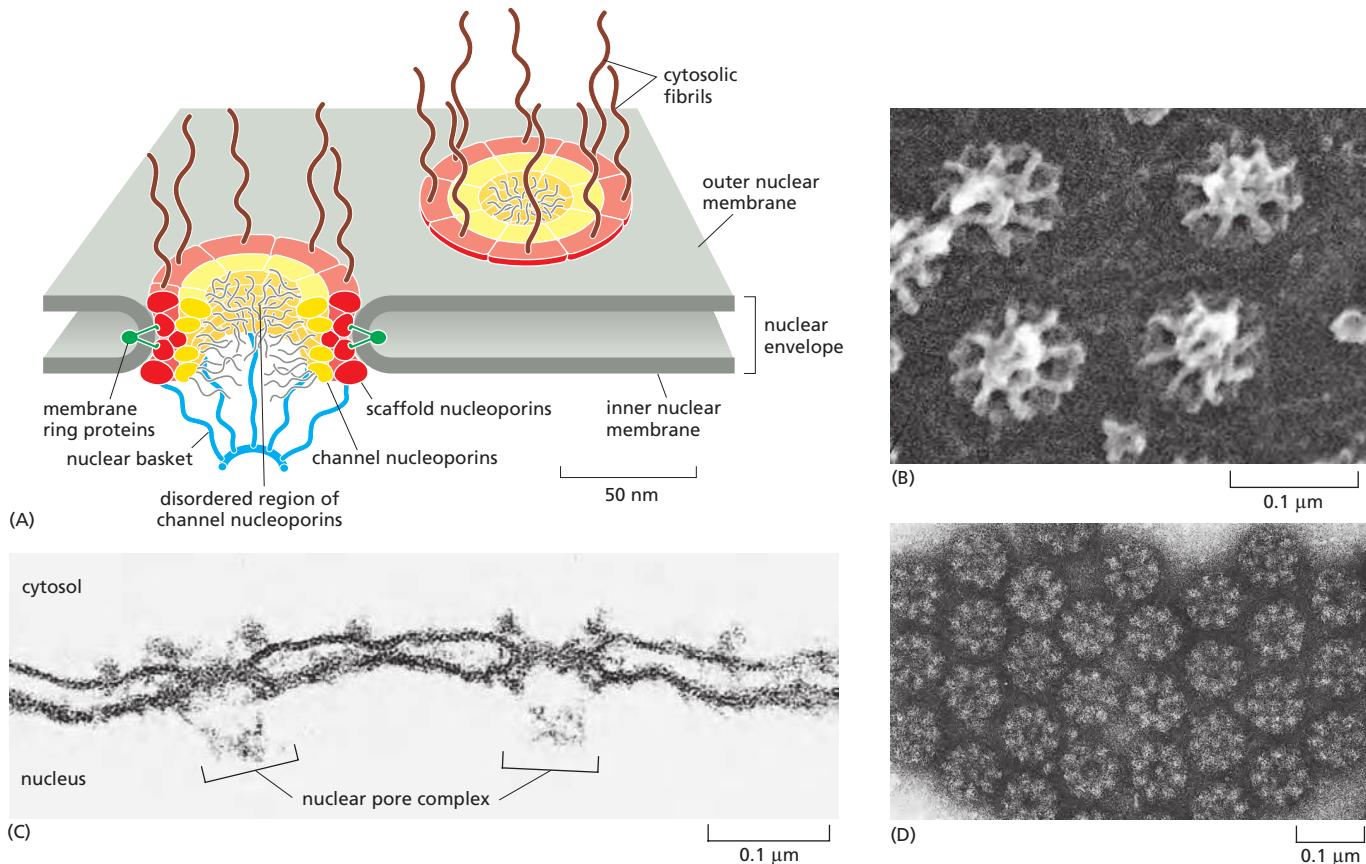
Nuclear localization signals can be located almost anywhere in the amino acid sequence and are thought to form loops or patches on the protein surface. Many function even when linked as short peptides to lysine side chains on the surface of a cytosolic protein, suggesting that the precise location of the signal within the amino acid sequence of a nuclear protein is not important. Moreover, as long as one of the protein subunits of a multicomponent complex displays a nuclear localization signal, the entire complex will be imported into the nucleus.



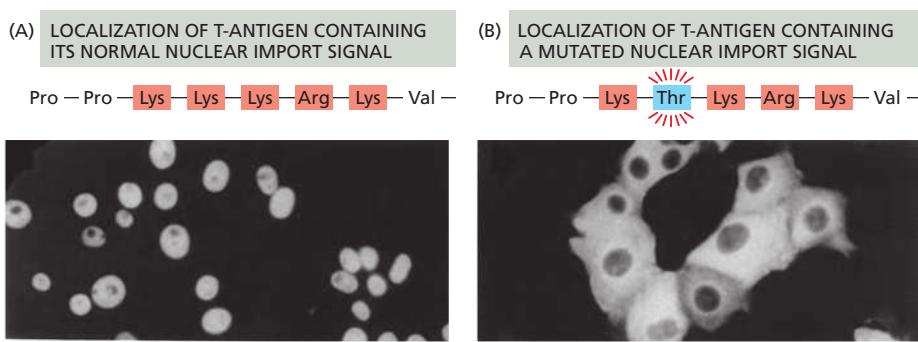
**Figure 12–7 The nuclear envelope.** The double-membrane envelope is penetrated by pores in which nuclear pore complexes (not shown) are positioned. The outer nuclear membrane is continuous with the endoplasmic reticulum (ER). The ribosomes that are normally bound to the cytosolic surface of the ER membrane and outer nuclear membrane are not shown. The nuclear lamina is a fibrous protein meshwork underlying the inner membrane.

One can visualize the transport of nuclear proteins through NPCs by coating gold particles with a nuclear localization signal, injecting the particles into the cytosol, and then following their fate by electron microscopy (Figure 12–10). The particles bind to the tentaclelike fibrils that extend from the scaffold nucleoporins at the rim of the NPC into the cytosol, and then proceed through the center of the NPC. Presumably, the unstructured regions of the nucleoporins that form a diffusion barrier for large molecules (mentioned earlier) are pushed away to allow the coated gold particles to squeeze through.

Macromolecular transport across NPCs differs fundamentally from the transport of proteins across the membranes of other organelles, in that it occurs



**Figure 12–8 The arrangement of NPCs in the nuclear envelope.** (A) In a vertebrate NPC, nucleoporins are arranged with striking eightfold rotational symmetry. In addition, immunoelectron microscopic studies show that the proteins that make up the central portion of the NPC are oriented symmetrically across the nuclear envelope, so that the nuclear and cytosolic sides look identical. The eightfold rotational and twofold transverse symmetry explains how such a huge structure can be formed from only about 30 different proteins: many of the nucleoporins are present in 8, 16, or 32 copies. Based on their approximate localization in the central portion of the NPC, nucleoporins can be classified into (1) transmembrane ring proteins that span the nuclear envelope and anchor the NPC to the envelope; (2) scaffold nucleoporins that form layered ring structures. Some scaffold nucleoporins are membrane-bending proteins that stabilize the sharp membrane curvature where the nuclear envelope is penetrated; and (3) channel nucleoporins that line a central pore. In addition to folded domains that anchor the proteins in specific places, many channel nucleoporins contain extensive unstructured regions, where the polypeptide chains are intrinsically disordered. The central pore is filled with a tangled mesh of these disordered domains that blocks the passive diffusion of large macromolecules. The disordered regions contain a large number of phenylalanine-glycine (FG) repeats. Fibrils protrude from both the cytosolic and the nuclear sides of the NPC. By contrast to the twofold transverse symmetry of the NPC core, the fibrils facing the cytosol and nucleus are different: on the nuclear side, the fibrils converge at their distal end to form a basketlike structure. The precise arrangement of individual nucleoporins in the assembled NPC is still a matter of intense debate, because atomic resolution analyses have been hindered by the sheer size and flexible nature of the NPC, and by difficulties in purifying sufficient amounts of homogeneous material. A combination of electron microscopy, computational analyses, and crystal structures of nucleoporin subcomplexes has been used to develop the current models of the NPC architecture. (B) A scanning electron micrograph of the nuclear side of the nuclear envelope of an oocyte (see also Figure 9–52). (C) An electron micrograph showing a side view of two NPCs (brackets); note that the inner and outer nuclear membranes are continuous at the edges of the pore. (D) An electron micrograph showing face-on views of negatively stained NPCs. The membrane has been removed by detergent extraction. Note that some of the NPCs contain material in their center, which is thought to be trapped macromolecules in transit through these NPCs. (A, adapted from A. Hoelz, E.W. Debler and G. Blobel, *Annu. Rev. Biochem.* 80:613–643, 2011. With permission from Annual Reviews; B, from M.W. Goldberg and T.D. Allen, *J. Cell Biol.* 119:1429–1440, 1992. With permission from The Rockefeller University Press; C, courtesy of Werner Franke and Ulrich Scheer; D, courtesy of Ron Milligan.)



**Figure 12-9 The function of a nuclear localization signal.** Immunofluorescence micrographs showing the cell location of SV40 virus T-antigen containing or lacking a short sequence that serves as a nuclear localization signal. (A) The normal T-antigen protein contains the lysine-rich sequence indicated and is imported to its site of action in the nucleus, as indicated by immunofluorescence staining with antibodies against the T-antigen. (B) T-antigen with an altered nuclear localization signal (a threonine replacing a lysine) remains in the cytosol. (From D. Kalderon, B. Roberts, W. Richardson and A. Smith, *Cell* 39:499–509, 1984. With permission from Elsevier.)

through a large, expandable, aqueous pore, rather than through a protein transporter spanning one or more lipid bilayers. For this reason, fully folded nuclear proteins can be transported into the nucleus through an NPC, and newly formed ribosomal subunits are transported out of the nucleus as an assembled particle. By contrast, proteins have to be extensively unfolded to be transported into most other organelles, as we discuss later.

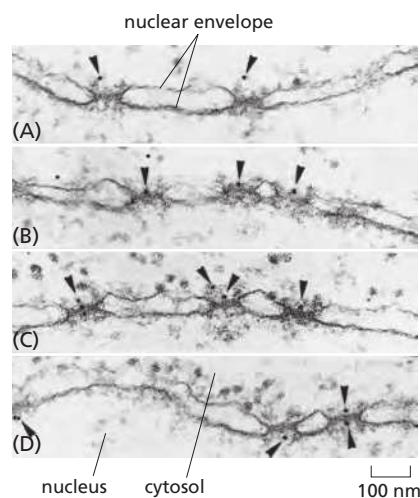
### Nuclear Import Receptors Bind to Both Nuclear Localization Signals and NPC Proteins

To initiate nuclear import, most nuclear localization signals must be recognized by **nuclear import receptors**, sometimes called *importins*, most of which are encoded by a family of related genes. Each family member encodes a receptor protein that can bind and transport the subset of cargo proteins containing the appropriate nuclear localization signal (Figure 12-11A). Nuclear import receptors do not always bind to nuclear proteins directly. Additional adaptor proteins can form a bridge between the import receptors and the nuclear localization signals on the proteins to be transported (Figure 12-11B). Some adaptor proteins are structurally related to nuclear import receptors, suggesting a common evolutionary origin. By using a variety of import receptors and adaptors, cells are able to recognize the broad repertoire of nuclear localization signals that are displayed on nuclear proteins.

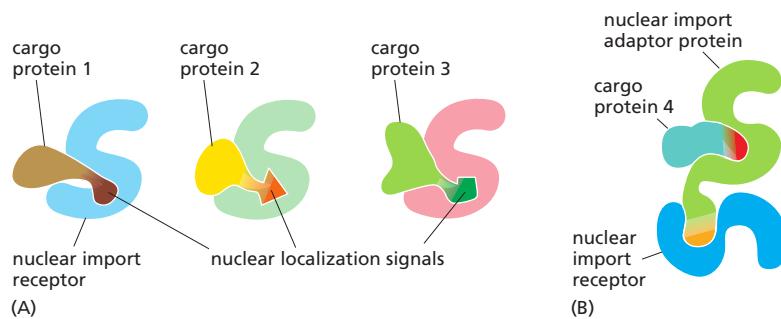
The import receptors are soluble cytosolic proteins that bind both to the nuclear localization signal on the cargo protein and to the phenylalanine-glycine (FG) repeats in the unstructured domains of the channel nucleoporins that line the central pore. FG-repeats are also found in the cytoplasmic and nuclear fibrils. FG-repeats in the unstructured tangle of the pore are thought to do double duty. They interact weakly, which gives the protein tangle gel-like properties that impose a permeability barrier to large macromolecules, and they serve as docking sites for nuclear import receptors. FG-repeats line the path through the NPCs taken by the import receptors and their bound cargo proteins. According to one model of nuclear transport, the receptor-cargo complexes move along the transport path by repeatedly binding, dissociating, and then re-binding to adjacent FG-repeat sequences. In this way, the complexes may hop from one nucleoporin to another to traverse the tangled interior of the NPC in a random walk. As import receptors bind to FG-repeats during this journey, they would disrupt interaction between the repeats and locally dissolve the gel phase of the protein tangle that fills the pore, allowing the passage of the receptor-cargo complex. Once inside the nucleus, the import receptors dissociate from their cargo and return to the cytosol. As we will see, this dissociation only occurs on the nuclear side of the NPC and thereby confers directionality to the import process.

### Nuclear Export Works Like Nuclear Import, But in Reverse

The nuclear export of large molecules, such as new ribosomal subunits and RNA molecules, occurs through NPCs and also depends on a selective transport



**Figure 12-10 Visualizing active import through NPCs.** This series of electron micrographs shows colloidal gold spheres (arrowheads) coated with peptides containing nuclear localization signals entering the nucleus through NPCs. The gold particles were injected into the cytosol of living cells, which then were fixed and prepared for electron microscopy at various times after injection. (A) Gold particles are first seen in proximity to the cytosolic fibrils of the NPCs. (B, C) They are then seen at the center of the NPCs, exclusively on the cytosolic face. (D) They then appear on the nuclear face. These gold particles have much larger diameters than the diffusion channels in the NPC and are imported by active transport. (From N. Panté and U. Aebi, *Science* 273:1729–1732, 1996. With permission from AAAS.)



**Figure 12-11 Nuclear import receptors (importins).** (A) Different nuclear import receptors bind different nuclear localization signals and thereby different cargo proteins. (B) Cargo protein 4 requires an adaptor protein to bind to its nuclear import receptor. The adaptors are structurally related to nuclear import receptors and recognize nuclear localization signals on cargo proteins. They also contain a nuclear localization signal that binds them to an import receptor, but this signal only becomes exposed when they are loaded with a cargo protein.

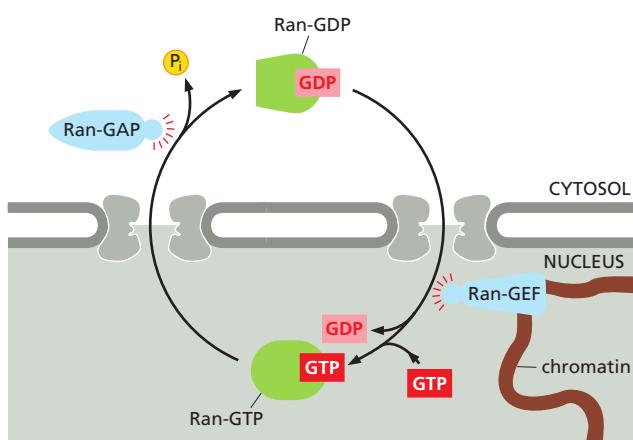
system. The transport system relies on **nuclear export signals** on the macromolecules to be exported, as well as on complementary **nuclear export receptors**, or *exportins*. These receptors bind to both the export signal and NPC proteins to guide their cargo through the NPC to the cytosol.

Many nuclear export receptors are structurally related to nuclear import receptors, and they are encoded by the same gene family of **nuclear transport receptors**, or *karyopherins*. In yeast, there are 14 genes encoding karyopherins; in animal cells, the number is significantly larger. It is often not possible to tell from their amino acid sequence alone whether a particular family member works as a nuclear import or nuclear export receptor. As might be expected, therefore, the import and export transport systems work in similar ways but in opposite directions: the import receptors bind their cargo molecules in the cytosol, release them in the nucleus, and are then exported to the cytosol for reuse, while the export receptors function in the opposite fashion.

### The Ran GTPase Imposes Directionality on Transport Through NPCs

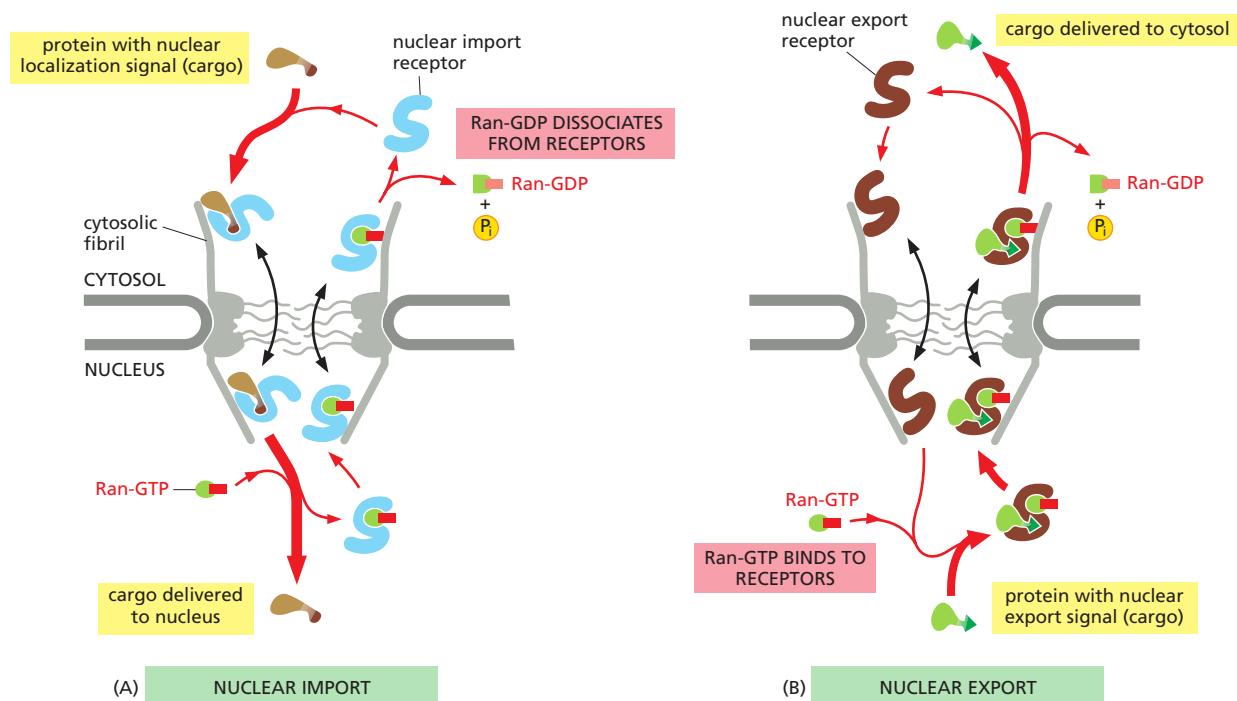
The import of nuclear proteins through NPCs concentrates specific proteins in the nucleus and thereby increases order in the cell. The cell fuels this ordering process by harnessing energy stored in concentration gradients of the GTP-bound form of the monomeric GTPase **Ran**, which is required for both nuclear import and export.

Like other GTPases, Ran is a molecular switch that can exist in two conformational states, depending on whether GDP or GTP is bound (discussed in Chapter 3). Two Ran-specific regulatory proteins trigger the conversion between the two states: a cytosolic *GTPase-activating protein* (*GAP*) triggers GTP hydrolysis and thus converts Ran-GTP to Ran-GDP, and a nuclear *guanine exchange factor* (*GEF*) promotes the exchange of GDP for GTP and thus converts Ran-GDP to Ran-GTP. Because *Ran-GAP* is located in the cytosol and *Ran-GEF* is located in the nucleus where it is anchored to chromatin, the cytosol contains mainly Ran-GDP, and the nucleus contains mainly Ran-GTP (Figure 12-12).



**Figure 12-12 The compartmentalization of Ran-GDP and Ran-GTP.** Localization of Ran-GDP in the cytosol and Ran-GTP in the nucleus results from the localization of two Ran regulatory proteins: Ran GTPase-activating protein (Ran-GAP) is located in the cytosol, and Ran guanine nucleotide exchange factor (Ran-GEF) binds to chromatin and is therefore located in the nucleus.

Ran-GDP is imported into the nucleus by its own import receptor, which is specific for the GDP-bound conformation of Ran. The Ran-GDP receptor is structurally unrelated to the main family of nuclear transport receptors. However, it also binds to FG-repeats in NPC channel nucleoporins.



This gradient of the two conformational forms of Ran drives nuclear transport in the appropriate direction. Docking of nuclear import receptors to FG-repeats on the cytosolic side of the NPC, for example, occurs whether or not these receptors are loaded with appropriate cargo. Import receptors, facilitated by FG-repeat binding, then enter the channel. If they reach the nuclear side of the pore complex, Ran-GTP binds to them, and, if the receptors arrive loaded with cargo molecules, the Ran-GTP binding causes the receptors to release their cargo (Figure 12–13A). Because the Ran-GDP in the cytosol does not bind to import (or export) receptors, unloading occurs only on the nuclear side of the NPC. In this way, the nuclear localization of Ran-GTP creates the directionality of the import process.

Having discharged its cargo in the nucleus, the empty import receptor with Ran-GTP bound is transported back through the pore complex to the cytosol. There, Ran-GAP triggers Ran-GTP to hydrolyze its bound GTP, thereby converting it to Ran-GDP, which dissociates from the receptor. The receptor is then ready for another cycle of nuclear import.

Nuclear export occurs by a similar mechanism, except that Ran-GTP in the nucleus promotes cargo binding to the export receptor, rather than promoting cargo dissociation. Once the export receptor moves through the pore to the cytosol, it encounters Ran-GAP, which induces the receptor to hydrolyze its GTP to GDP. As a result, the export receptor releases both its cargo and Ran-GDP in the cytosol. Free export receptors are then returned to the nucleus to complete the cycle (Figure 12–13B).

**Figure 12–13** How GTP hydrolysis by Ran in the cytosol provides directionality to nuclear transport.

Movement through the NPC of loaded nuclear transport receptors occurs along the FG-repeats displayed by certain NPC proteins. The differential localization of Ran-GTP in the nucleus and Ran-GDP in the cytosol provides directionality (red arrows) to both nuclear import (A) and nuclear export (B). Ran-GAP stimulates the hydrolysis of GTP to produce Ran-GDP on the cytosolic side of the NPC (see Figure 12–12).

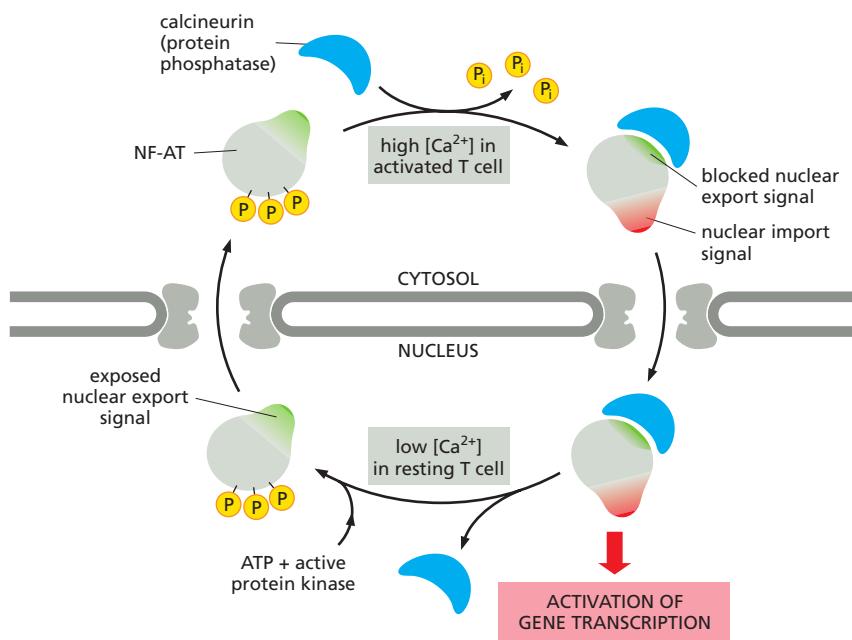
### Transport Through NPCs Can Be Regulated by Controlling Access to the Transport Machinery

Some proteins contain both nuclear localization signals and nuclear export signals. These proteins continually shuttle back and forth between the nucleus and the cytosol. The relative rates of their import and export determine the steady-state localization of such *shuttling proteins*: if the rate of import exceeds the rate of export, a protein will be located mainly in the nucleus; conversely, if the rate of export exceeds the rate of import, a protein will be located mainly in the cytosol. Thus, changing the rate of import, export, or both, can change the location of a protein.

Some shuttling proteins move continuously into and out of the nucleus. In other cases, however, the transport is stringently controlled. As discussed in Chapter 7, cells control the activity of some transcription regulators by keeping them out of the nucleus until they are needed there (Figure 12–14). In many cases, cells control transport by regulating nuclear localization and export signals—turning them on or off, often by phosphorylation of amino acids close to the signal sequences (Figure 12–15).

Other transcription regulators are bound to inhibitory cytosolic proteins that either anchor them in the cytosol (through interactions with the cytoskeleton or specific organelles) or mask their nuclear localization signals so that they cannot interact with nuclear import receptors. An appropriate stimulus releases the gene regulatory protein from its cytosolic anchor or mask, and it is then transported into the nucleus. One important example is the latent gene regulatory protein that controls the expression of proteins involved in cholesterol metabolism. The protein is made and stored in an inactive form as a transmembrane protein in the ER. When a cell is deprived of cholesterol, the protein is transported from the ER to the Golgi apparatus where it encounters specific proteases that cleave off the cytosolic domain, releasing it into the cytosol. This domain is then imported into the nucleus, where it activates the transcription of genes required for both cholesterol uptake and synthesis (Figure 12–16).

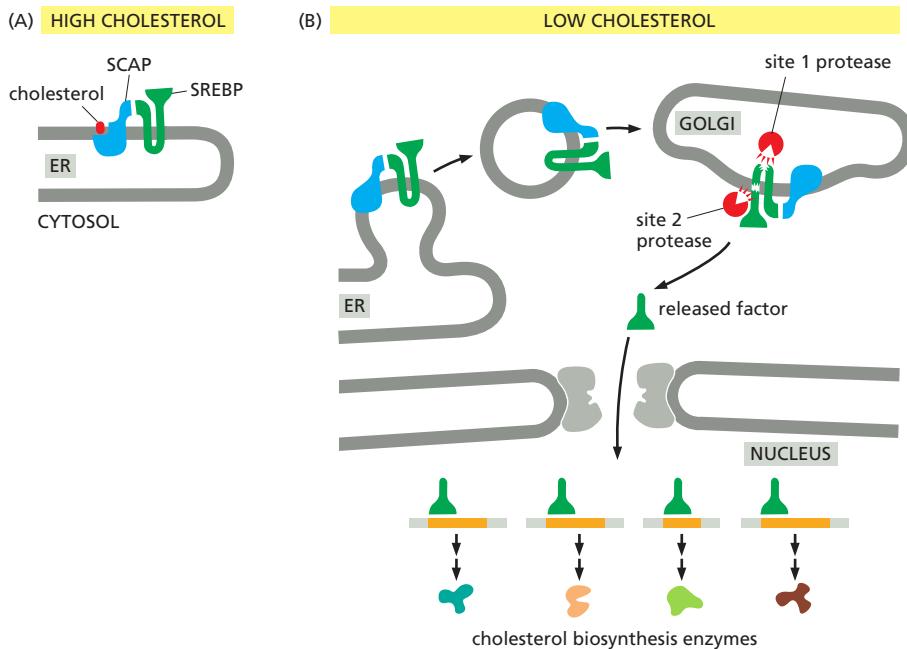
As we discuss in detail in Chapter 6, cells control the export of RNAs from the nucleus in a similar way. snRNAs, miRNAs, and tRNAs bind to the same family of nuclear export receptors just discussed, and they use the same Ran-GTP gradient to fuel the transport process. By contrast, the export of mRNAs out of the nucleus uses a different mechanism. mRNAs are exported as large assemblies, which can be as large as 100 million daltons (see Figure 6–37) and can contain hundreds of proteins of a few dozen different types. These mRNA ribonucleoprotein complexes (mRNPs) first dock at the nuclear side of the NPC, where they are extensively remodeled. Although Ran-GTP is indirectly involved in the export (because it imports the proteins that bind to the mRNA molecules), the translocation across the NPC is thought to be driven by ATP hydrolysis. How export directionality is assured is unclear. It is likely that the many accessory proteins tethered to the NPC's nuclear and cytoplasmic fibrils have important roles in remodeling the mRNPs as they pass through the pores, in particular stripping away nuclear proteins as the mRNPs exit on the cytosolic side of the NPC, thereby ensuring that transport is unidirectional. Upon entry into the cytosol, these nuclear mRNP proteins are rapidly returned to the nucleus.



**Figure 12–14 The control of nuclear transport in the early *Drosophila* embryo.** The embryo at this stage is a syncytium, shown here in cross section, with many nuclei in a common cytoplasm, arranged around the periphery, just beneath the plasma membrane. The transcription regulatory protein Dorsal is produced uniformly throughout the peripheral cytoplasm, but it can act only when inside the nuclei. The Dorsal protein has been stained with an enzyme-coupled antibody that yields a brown product, revealing that Dorsal is excluded from the nuclei at the dorsal side (top) of the embryo but is concentrated in the nuclei toward the ventral side (bottom) of the embryo. The regulated traffic of Dorsal into the nuclei controls the differential development between the back and belly of the animal. (Courtesy of Siegfried Roth.)

**Figure 12–15 The control of nuclear import during T cell activation.** The nuclear factor of activated T cells (NF-AT) is a transcription regulatory protein that, in the resting T cell, is found in the cytosol in a phosphorylated state. When T cells are activated by foreign antigen (discussed in Chapter 24), the intracellular  $\text{Ca}^{2+}$  concentration increases. In high  $\text{Ca}^{2+}$ , the protein phosphatase calcineurin binds to NF-AT and dephosphorylates it. The dephosphorylation exposes nuclear import signals and blocks a nuclear export signal. The complex of NF-AT and calcineurin is therefore imported into the nucleus, where NF-AT activates the transcription of numerous genes required for T cell activation.

The response shuts off when  $\text{Ca}^{2+}$  levels decrease, releasing NF-AT from calcineurin. Rephosphorylation of NF-AT inactivates the nuclear import signals and re-exposes the nuclear export signal, causing NF-AT to relocate to the cytosol. Some of the most potent immunosuppressive drugs, including cyclosporin A and FK506, inhibit the ability of calcineurin to dephosphorylate NF-AT and thereby block the nuclear accumulation of NF-AT and T cell activation (Movie 12.1).



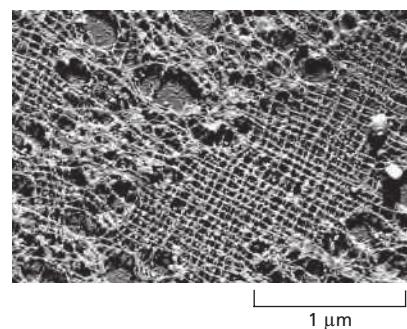
**Figure 12–16 Feedback regulation of cholesterol biosynthesis.** SREBP (sterol response element binding protein), a latent transcription regulator that controls expression of cholesterol biosynthetic enzymes, is initially synthesized as an ER membrane protein. It is anchored in the ER if there is sufficient cholesterol in the membrane by interaction with another ER membrane protein, called SCAP (SREBP cleavage activation protein), which binds cholesterol. If the cholesterol binding site on SCAP is empty (at low cholesterol concentrations), SCAP changes conformation and is packaged together with SREBP into transport vesicles, which deliver their cargo to the Golgi apparatus, where two Golgi-resident proteases cleave SREBP to free its cytosolic domain from the membrane. The cytosolic domain then moves into the nucleus, where it binds to the promoters of genes that encode proteins involved in cholesterol biosynthesis and activates their transcription. In this way, more cholesterol is made when its concentration falls below a threshold.

### During Mitosis the Nuclear Envelope Disassembles

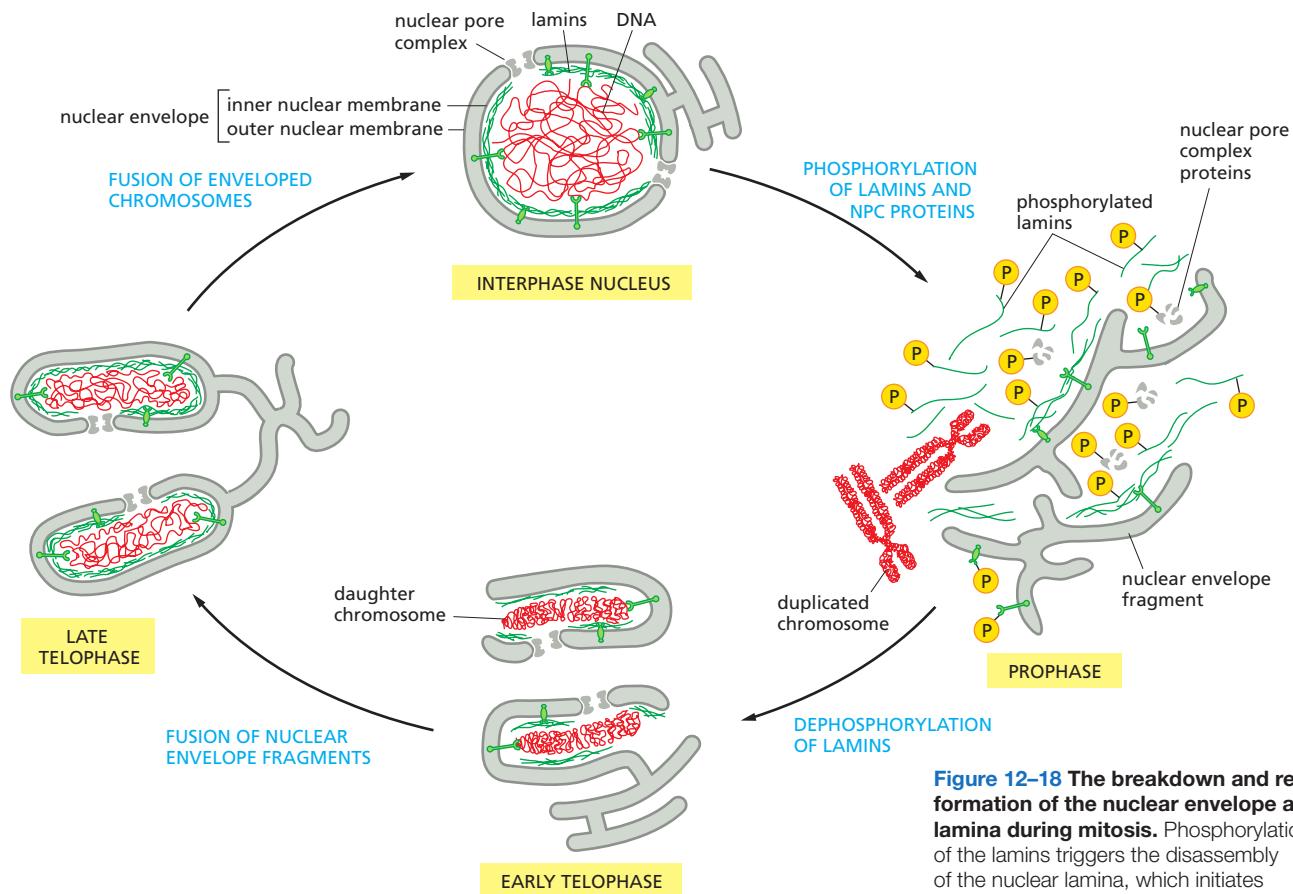
The **nuclear lamina**, located on the nuclear side of the inner nuclear membrane, is a meshwork of interconnected protein subunits called **nuclear lamins**. The lamins are a special class of intermediate filament proteins (discussed in Chapter 16) that polymerize into a two-dimensional lattice (Figure 12–17). The nuclear lamina gives shape and stability to the nuclear envelope, to which it is anchored by attachment to both the NPCs and transmembrane proteins of the inner nuclear membrane. The lamina also interacts directly with chromatin, which itself interacts with transmembrane proteins of the inner nuclear membrane. Together with the lamina, these inner membrane proteins provide structural links between the DNA and the nuclear envelope.

When a nucleus is dismantled during mitosis, the NPCs and nuclear lamina disassemble and the nuclear envelope fragments. The dismantling process is at least partly a consequence of direct phosphorylation of nucleoporins and lamins by the cyclin-dependent protein kinase (Cdk) that is activated at the onset of mitosis (discussed in Chapter 17). During this process, some NPC proteins become bound to nuclear import receptors, which play an important part in the reassembly of NPCs at the end of mitosis. Nuclear envelope membrane proteins—no longer tethered to the pore complexes, lamina, or chromatin—disperse throughout the ER membrane. The dynein motor protein, which moves along microtubules (discussed in Chapter 16), actively participates in tearing the nuclear envelope off the chromatin. Together, these processes break down the barriers that normally separate the nucleus and cytosol, and the nuclear proteins that are not bound to membranes or chromosomes intermix completely with the proteins of the cytosol (Figure 12–18).

Later in mitosis, the nuclear envelope reassembles on the surface of the daughter chromosomes. In addition to its crucial role in nuclear transport, the Ran GTPase also acts as a positional marker for chromatin during cell division, when the nuclear and cytosolic components intermix. Because Ran-GEF remains bound to chromatin when the nuclear envelope breaks down, Ran molecules close to chromatin are mainly in their GTP-bound conformation. By contrast, Ran molecules further away have a high likelihood of encountering Ran-GAP, which is distributed throughout the cytosol; these Ran molecules are mainly in their GDP-bound conformation. As a result, the chromosomes in mitotic cells are surrounded by a cloud of Ran-GTP. Ran-GTP releases the NPC proteins in proximity to the chromosomes from nuclear import receptors. The free NPC proteins attach



**Figure 12–17 The nuclear lamina.** An electron micrograph of a portion of the nuclear lamina in a *Xenopus* oocyte prepared by freeze-drying and metal shadowing. The lamina is formed by a regular lattice of specialized intermediate filaments. Lamins are only present in metazoan cells. Other, yet-unknown proteins may serve similar functions in species that lack lamins. (Courtesy of Ueli Aebi.)



**Figure 12–18** The breakdown and reformation of the nuclear envelope and lamina during mitosis. Phosphorylation of the lamins triggers the disassembly of the nuclear lamina, which initiates the nuclear envelope to break up. Dephosphorylation of the lamins reverses the process. An analogous phosphorylation and dephosphorylation cycle occurs for some nucleoporins and proteins of the inner nuclear membrane, and some of these dephosphorylations are also involved in the reassembly process. As indicated, the nuclear envelope initially re-forms around individual decondensing daughter chromosomes. Eventually, as decondensation progresses, these structures fuse to form a single complete nucleus.

Mitotic breakdown of the nuclear envelope occurs in all metazoan cells. However, in many other species, such as yeasts, the nuclear envelope remains intact during mitosis, and the nucleus divides by fission.

to the chromosome surface, where they assemble into new NPCs. At the same time, inner nuclear membrane proteins and dephosphorylated lamins bind again to chromatin. ER membranes wrap around groups of chromosomes until they form a sealed nuclear envelope (**Movie 12.2**). During this process, the NPCs start actively re-importing proteins that contain nuclear localization signals. Because the nuclear envelope is initially closely applied to the surface of the chromosomes, the newly formed nucleus excludes all proteins except those initially bound to the mitotic chromosomes and those that are selectively imported through NPCs. In this way, all other large proteins, including ribosomes, are kept out of the newly assembled nucleus.

As we discuss in Chapter 17, the cloud of Ran-GTP surrounding chromatin is also important in assembling the mitotic spindle in a dividing cell.

## Summary

The nuclear envelope consists of an inner and an outer nuclear membrane that are continuous with each other and with the ER membrane, and the space between the inner and outer nuclear membrane is continuous with the ER lumen. RNA molecules, which are made in the nucleus, and ribosomal subunits, which are assembled there, are exported to the cytosol; in contrast, all the proteins that function in the nucleus are synthesized in the cytosol and are then imported. The extensive traffic of materials between the nucleus and cytosol occurs through nuclear pore complexes (NPCs), which provide a direct passageway across the nuclear envelope. Small molecules diffuse passively through the NPCs, but large macromolecules have to be actively transported.

Proteins containing nuclear localization signals are actively transported into the nucleus through NPCs, while proteins containing nuclear export signals are transported out of the nucleus to the cytosol. Some proteins, including the nuclear

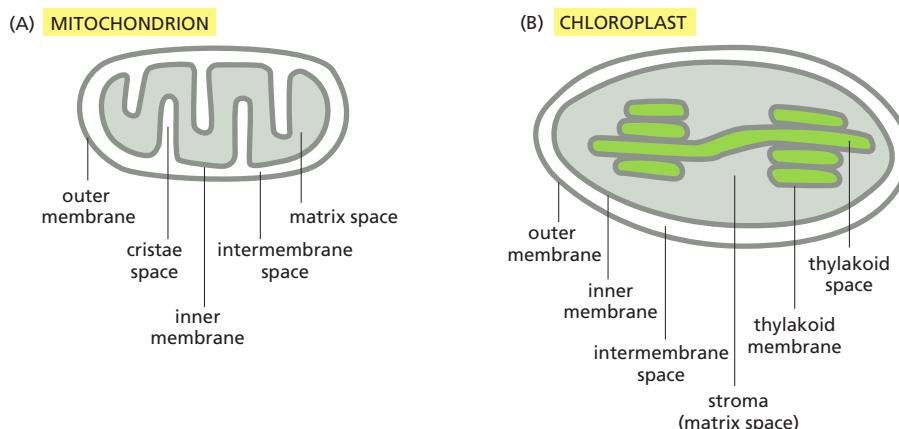
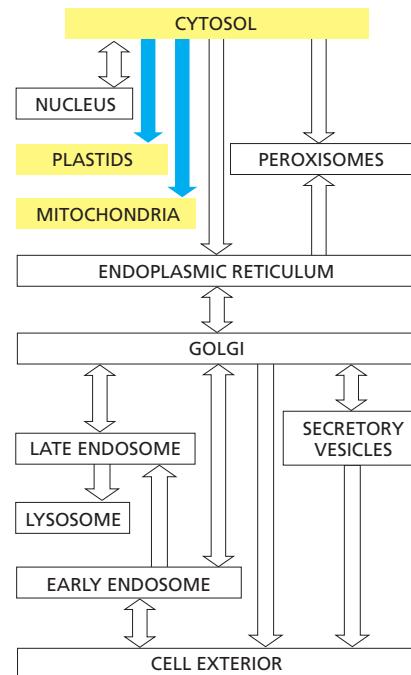
*import and export receptors, continually shuttle between the cytosol and nucleus.* The monomeric GTPase Ran provides both the free energy and the directionality for nuclear transport. Cells regulate the transport of nuclear proteins and RNA molecules through the NPCs by controlling the access of these molecules to the transport machinery. Newly transcribed messenger RNA and ribosomal RNA are exported from the nucleus as parts of large ribonucleoprotein complexes. Because nuclear localization signals are not removed, nuclear proteins can be imported repeatedly, as is required each time that the nucleus reassembles after mitosis.

## THE TRANSPORT OF PROTEINS INTO MITOCHONDRIA AND CHLOROPLASTS

Mitochondria and chloroplasts (a specialized form of plastids in green algae and plant cells) are double-membrane-enclosed organelles. They specialize in ATP synthesis, using energy derived from electron transport and oxidative phosphorylation in mitochondria and from photosynthesis in chloroplasts (discussed in Chapter 14). Although both organelles contain their own DNA, ribosomes, and other components required for protein synthesis, most of their proteins are encoded in the cell nucleus and imported from the cytosol. Each imported protein must reach the particular organelle subcompartment in which it functions.

There are different subcompartments in mitochondria (Figure 12–19A): the internal matrix space and the intermembrane space, which is continuous with the cristae space. These compartments are formed by the two concentric mitochondrial membranes: the **inner membrane**, which encloses the matrix space and forms extensive invaginations called *cristae*, and the **outer membrane**, which is in contact with the cytosol. Protein complexes provide boundaries at the junctions where the cristae invaginate and divide the inner membrane into two domains: one inner membrane domain surrounds the cristae space, and the other domain abuts the outer membrane. Chloroplasts also have an outer and inner membrane, which enclose an intermembrane space, and the stroma, which is the chloroplast equivalent of the mitochondrial matrix space (Figure 12–19B). They have an additional subcompartment, the *thylakoid space*, which is surrounded by the *thylakoid membrane*. The thylakoid membrane derives from the inner membrane during plastid development and is pinched off to become discontinuous with it. Each of the subcompartments in mitochondria and chloroplasts contains a distinct set of proteins.

New mitochondria and chloroplasts are produced by the growth of preexisting organelles, followed by fission (discussed in Chapter 14). The growth depends mainly on the import of proteins from the cytosol. The imported proteins must be transported across a number of membranes in succession and end up in the appropriate place. The process of protein movement across membranes is called *protein translocation*. This section explains how it occurs.



**Figure 12–19 The subcompartments of mitochondria and chloroplasts.** In contrast to the cristae of mitochondria (A), the thylakoids of chloroplasts (B) are not connected to the inner membrane and therefore form a sealed compartment with a separate internal space.

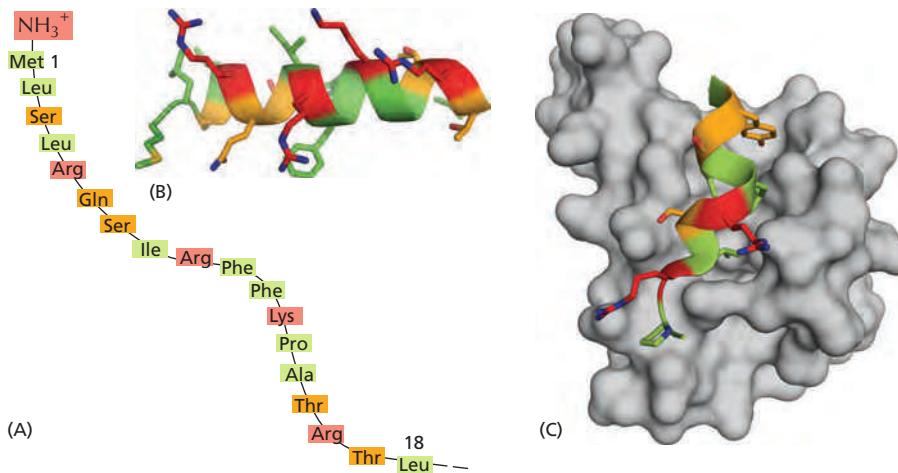
## Translocation into Mitochondria Depends on Signal Sequences and Protein Translocators

Proteins imported into **mitochondria** are usually taken up from the cytosol within seconds or minutes of their release from ribosomes. Thus, in contrast to protein translocation into the ER, which often takes place simultaneously with translation by a ribosome docked on the rough ER membrane (described later), mitochondrial proteins are first fully synthesized as **mitochondrial precursor proteins** in the cytosol and then translocated into mitochondria by a *post-translational* mechanism. One or more signal sequences direct all mitochondrial precursor proteins to their appropriate mitochondrial subcompartment. Many proteins entering the matrix space contain a signal sequence at their N-terminus that a signal peptidase rapidly removes after import. Other imported proteins, including all outer membrane and many inner membrane and intermembrane space proteins, have internal signal sequences that are not removed. The signal sequences are both necessary and sufficient for the import and correct localization of the proteins: when genetic engineering techniques are used to link these signals to a cytosolic protein, the signals direct the protein to the correct mitochondrial subcompartment.

The signal sequences that direct precursor proteins into the mitochondrial matrix space are best understood. They all form an amphiphilic  $\alpha$  helix, in which positively charged residues cluster on one side of the helix, while uncharged hydrophobic residues cluster on the opposite side. Specific receptor proteins that initiate protein translocation recognize this configuration rather than the precise amino acid sequence of the signal sequence (Figure 12-20).

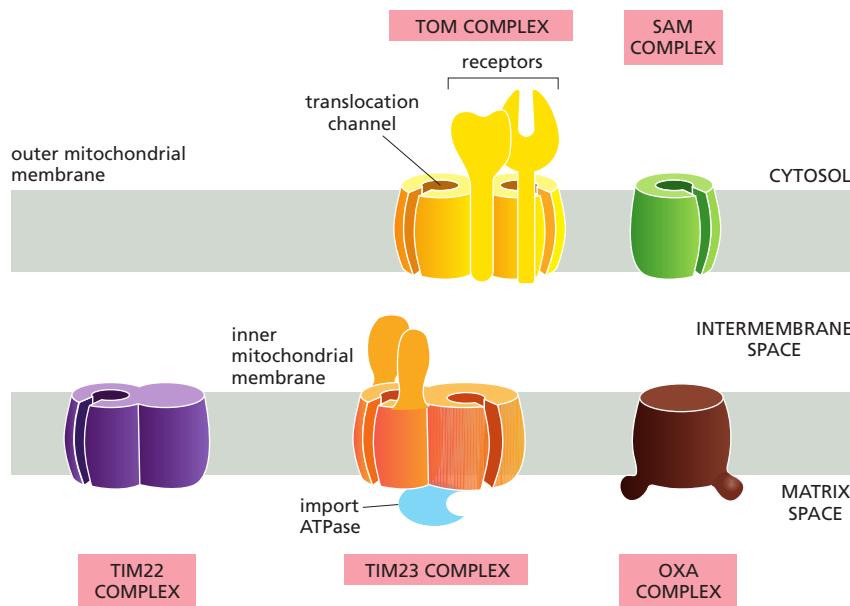
Multisubunit protein complexes that function as **protein translocators** mediate protein movement across mitochondrial membranes. The **TOM complex** transfers proteins across the outer membrane, and two **TIM complexes** (TIM23 and TIM22) transfer proteins across the inner membrane (Figure 12-21). These complexes contain some components that act as receptors for mitochondrial precursor proteins, and other components that form the translocation channels.

The TOM complex is required for the import of all nucleus-encoded mitochondrial proteins. It initially transports their signal sequences into the intermembrane space and helps to insert transmembrane proteins into the outer membrane.  $\beta$ -barrel proteins, which are particularly abundant in the outer membrane, are then passed on to an additional translocator, the **SAM complex**, which helps them to fold properly in the outer membrane. The TIM23 complex transports some soluble proteins into the matrix space and helps to insert transmembrane proteins into the inner membrane. The TIM22 complex mediates the insertion of a subclass of inner membrane proteins, including the transporter that moves ADP, ATP, and phosphate in and out of mitochondria. Yet another protein translocator in the inner mitochondrial membrane, the **OXA complex**, mediates the insertion of



**Figure 12–20** A signal sequence for mitochondrial protein import.

Cytochrome oxidase is a large multiprotein complex located in the inner mitochondrial membrane, where it functions as the terminal enzyme in the electron-transport chain (discussed in Chapter 14). (A) The first 18 amino acids of the precursor to subunit IV of this enzyme serve as a signal sequence for import of the subunit into the mitochondrion. (B) When the signal sequence is folded as an  $\alpha$  helix, the positively charged amino acids (red) are clustered on one face of the helix, while the nonpolar ones (green) are clustered primarily on the opposite face. Uncharged polar amino acids are shaded orange; nitrogen atoms on the side chains of Arg and Gln are colored blue. Signal sequences that direct proteins into the matrix space always have the potential to form such an amphiphilic  $\alpha$  helix, which is recognized by specific receptor proteins on the mitochondrial surface. (C) The structure of a signal sequence (of alcohol dehydrogenase, another mitochondrial matrix enzyme), bound to an import receptor (gray), as determined by nuclear magnetic resonance. The amphiphilic  $\alpha$  helix binds with its hydrophobic face to a hydrophobic groove in the receptor (PDB code: 1OM2).



**Figure 12–21 The protein translocators in the mitochondrial membranes.** The TOM, TIM, SAM, and OXA complexes are multimeric membrane protein assemblies that catalyze protein translocation across mitochondrial membranes. The protein components of the TIM22 and TIM23 complexes that line the import channel are structurally related, suggesting a common evolutionary origin of both TIM complexes. On the matrix side, the TIM23 complex is bound to a multimeric protein complex containing mitochondrial hsp70, which acts as an import ATPase, using ATP hydrolysis to pull proteins through the pore. In animal cells, subtle variations exist in the subunit composition of the translocator complexes to adapt the mitochondrial import machinery to the particular needs of specialized cell types. SAM = Sorting and Assembly Machinery; OXA = cytochrome OXidase Activity; TIM = Translocator of the Inner Mitochondrial membrane; TOM = Translocator of the Outer Membrane.

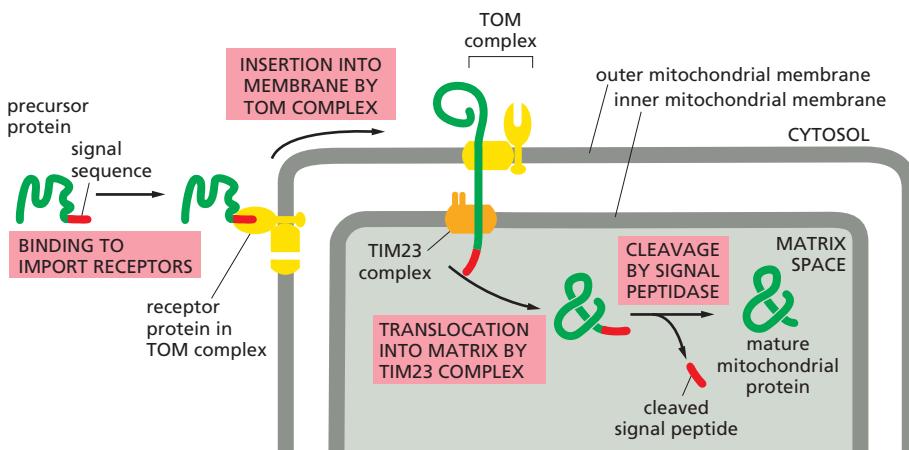
those inner membrane proteins that are synthesized within mitochondria. It also helps to insert some imported inner membrane proteins that are initially transported into the matrix space by the other complexes.

### Mitochondrial Precursor Proteins Are Imported as Unfolded Polypeptide Chains

We have learned almost everything we know about the molecular mechanism of protein import into mitochondria from analyses of cell-free, reconstituted translocation systems, in which purified mitochondria in a test tube import radiolabeled mitochondrial precursor proteins. By changing the conditions in the test tube, it is possible to establish the biochemical requirements for the import.

Mitochondrial precursor proteins do not fold into their native structures after they are synthesized; instead, they remain unfolded in the cytosol through interactions with other proteins. Some of these interacting proteins are general *chaperone proteins* of the *hsp70 family* (discussed in Chapter 6), whereas others are dedicated to mitochondrial precursor proteins and bind directly to their signal sequences. All the interacting proteins help to prevent the precursor proteins from aggregating or folding up spontaneously before they engage with the TOM complex in the outer mitochondrial membrane. As a first step in the import process, the import receptors of the TOM complex bind the signal sequence of the mitochondrial precursor protein. The interacting proteins are then stripped off, and the unfolded polypeptide chain is fed—signal sequence first—into the translocation channel.

In principle, a protein could reach the mitochondrial matrix space by either crossing the two membranes all at once or crossing one at a time. One can distinguish between these possibilities by cooling a cell-free mitochondrial import system to arrest the proteins at an intermediate step in the translocation process. The result is that the arrested proteins no longer contain their N-terminal signal sequence, indicating that the N-terminus must be in the matrix space where the signal peptidase is located, but the bulk of the protein can still be attacked from outside the mitochondria by externally added proteolytic enzymes. Clearly, the precursor proteins can pass through both mitochondrial membranes at once to enter the matrix space (Figure 12–22). The TOM complex first transports the signal sequence across the outer membrane to the intermembrane space, where it binds to a TIM complex, opening the channel in the complex. The polypeptide chain is then either translocated into the matrix space or inserted into the inner membrane.



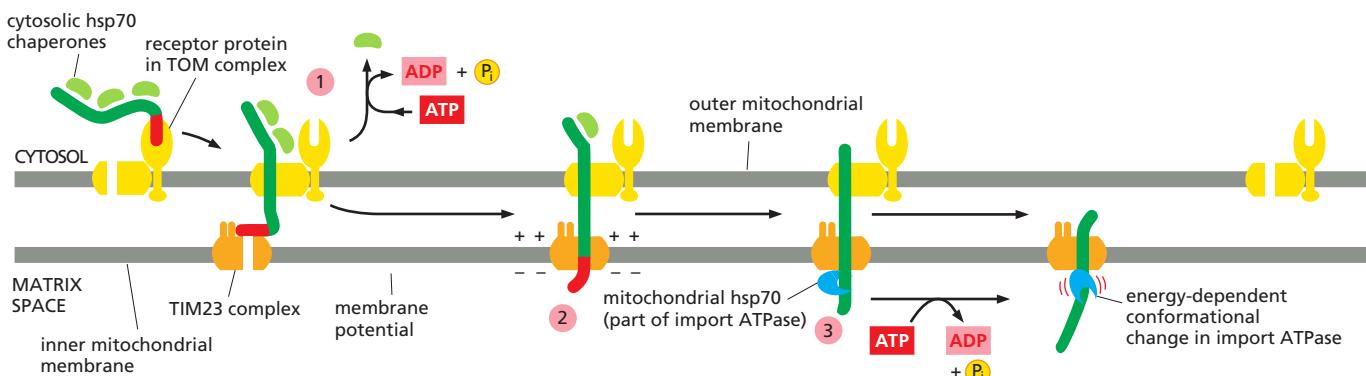
**Figure 12–22 Protein import by mitochondria.** The N-terminal signal sequence of the mitochondrial precursor protein is recognized by receptors of the TOM complex. The protein is then translocated through the TIM23 complex so that it transiently spans both mitochondrial membranes (Movie 12.3). The signal sequence is cleaved off by a signal peptidase in the matrix space to form the mature protein. The free signal sequence is then rapidly degraded (not shown).

Although the TOM and TIM complexes usually work together to translocate precursor proteins across both membranes at the same time, they can work independently. In isolated outer membranes, for example, the TOM complex can translocate the signal sequence of precursor proteins across the membrane. Similarly, if the outer membrane is experimentally disrupted in isolated mitochondria, the exposed TIM23 complex can efficiently import precursor proteins into the matrix space.

### ATP Hydrolysis and a Membrane Potential Drive Protein Import Into the Matrix Space

Directional transport requires energy, which in most biological systems is supplied by ATP hydrolysis. ATP hydrolysis fuels mitochondrial protein import at two discrete sites, one outside the mitochondria and one in the matrix space. In addition, protein import requires another energy source, which is the membrane potential across the inner mitochondrial membrane (Figure 12–23).

The first requirement for energy occurs at the initial stage of the translocation process, when the unfolded precursor protein, associated with chaperone proteins, interacts with the import receptors of the TOM complex. As discussed in Chapter 6, the binding and release of newly synthesized polypeptides from the chaperone proteins requires ATP hydrolysis.



**Figure 12–23 The role of energy in protein import into the mitochondrial matrix space.** (1) Bound cytosolic hsp70 chaperone is released from the precursor protein in a step that depends on ATP hydrolysis. After initial insertion of the signal sequence and of adjacent portions of the polypeptide chain into the TOM complex translocation channel, the signal sequence interacts with a TIM complex. (2) The signal sequence is then translocated into the matrix space in a process that requires the energy in the membrane potential across the inner membrane. (3) Mitochondrial hsp70, which is part of an import ATPase complex, binds to regions of the polypeptide chain as they become exposed in the matrix space, pulling the protein through the translocation channel, using the energy of ATP hydrolysis.

Once the signal sequence has passed through the TOM complex and is bound to a TIM complex, further translocation through the TIM translocation channel requires the membrane potential, which is the electrical component of the electrochemical H<sup>+</sup> gradient across the inner membrane (see Figure 11–4). Pumping of H<sup>+</sup> from the matrix space to the intermembrane space, driven by electron transport processes in the inner membrane (discussed in Chapter 14), maintains the electrochemical H<sup>+</sup> gradient across the inner membrane therefore not only powers most of the cell's ATP synthesis, but it also drives the translocation of the positively charged signal sequences through the TIM complexes by electrophoresis.

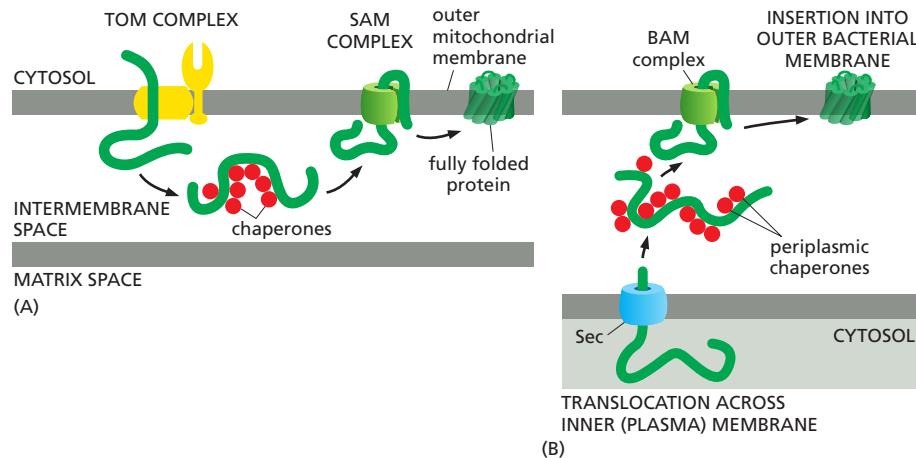
**Mitochondrial hsp70** also plays a crucial part in the import process. Mitochondria containing mutant forms of the protein fail to import precursor proteins. The mitochondrial hsp70 is part of a multisubunit protein assembly that is bound to the matrix side of the TIM23 complex and acts as a motor to pull the precursor protein into the matrix space. Like its cytosolic cousin, mitochondrial hsp70 has a high affinity for unfolded polypeptide chains, and it binds tightly to an imported protein chain as soon as the chain emerges from the TIM translocator in the matrix space. The hsp70 then undergoes a conformational change and releases the protein chain in an ATP-dependent step, exerting a ratcheting/pulling force on the protein being imported. This energy-driven cycle of binding and subsequent release provides the final driving force needed to complete protein import after a protein has initially inserted into the TIM23 complex (see Figure 12–23).

After the initial interaction with mitochondrial hsp70, many imported matrix proteins are passed on to another chaperone protein, *mitochondrial hsp60*. As discussed in Chapter 6, hsp60 helps the unfolded polypeptide chain to fold by binding and releasing it through cycles of ATP hydrolysis.

### Bacteria and Mitochondria Use Similar Mechanisms to Insert Porins into their Outer Membrane

The outer mitochondrial membrane, like the outer membrane of Gram-negative bacteria (see Figure 11–17), contains abundant pore-forming β-barrel proteins called **porins**, and it is thus freely permeable to inorganic ions and metabolites (but not to most proteins). In contrast to other outer membrane proteins, which are anchored in the membrane through transmembrane α-helical regions, the TOM complex cannot integrate porins into the lipid bilayer. Instead, porins are first transported unfolded into the intermembrane space, where they transiently bind specialized chaperone proteins, which keep the porins from aggregating (Figure 12–24A). They then bind to the SAM complex in the outer membrane, which both inserts them into the outer membrane and helps them fold properly.

One of the central subunits of the SAM complex is homologous to a bacterial outer membrane protein that helps insert β-barrel proteins into the bacterial outer

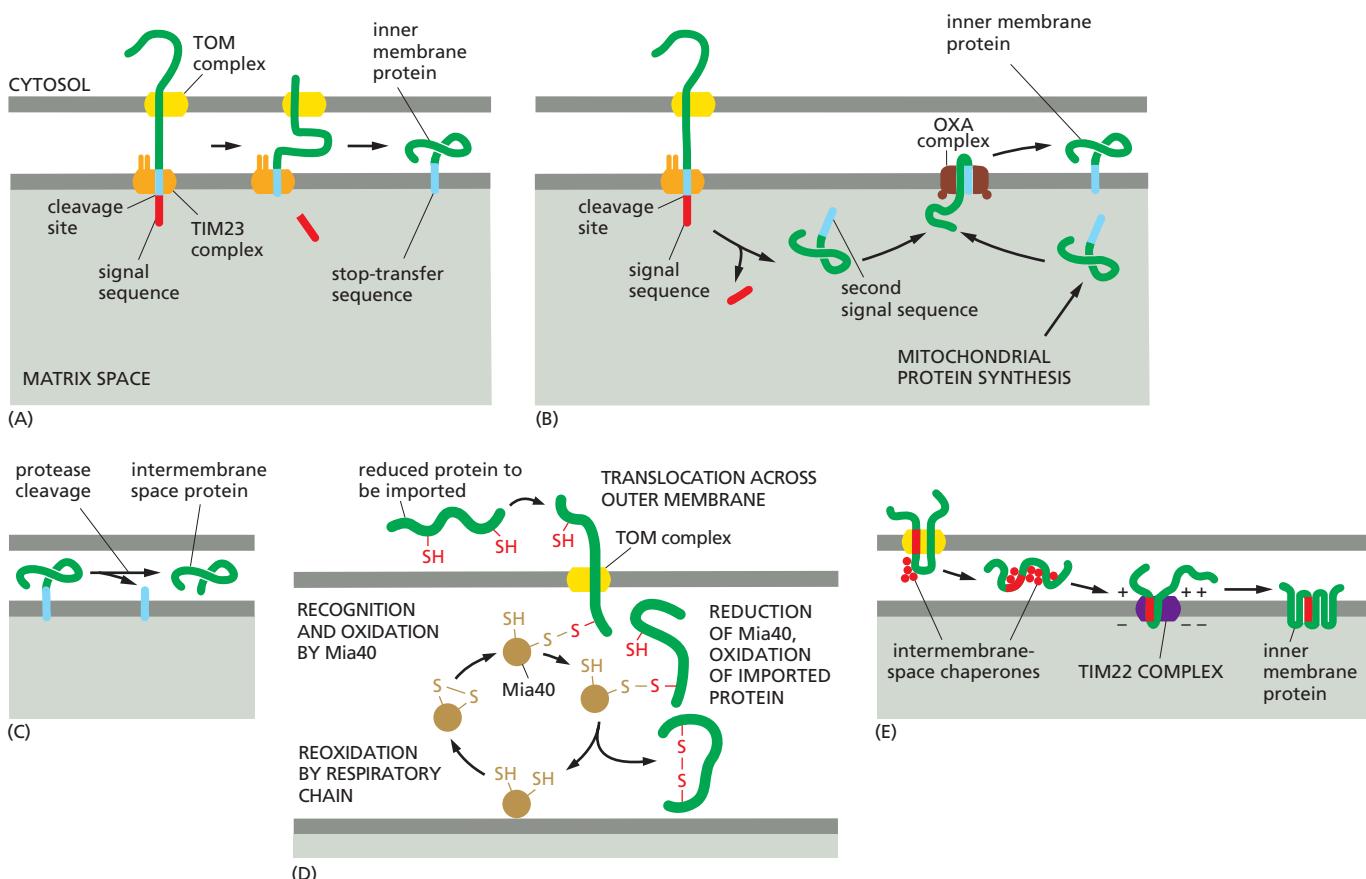


**Figure 12–24** Integration of porins into the outer mitochondrial and bacterial membranes. (A) After translocation through the TOM complex in the outer mitochondrial membrane, β-barrel proteins bind to chaperones in the intermembrane space. The SAM complex then inserts the unfolded polypeptide chain into the outer membrane and helps the chain fold. (B) A structurally related BAM complex in the outer membrane of Gram-negative bacteria catalyzes β-barrel protein insertion and folding (see Figure 11–17).

membrane from the periplasmic space (the equivalent of the intermembrane space in mitochondria) (Figure 12–24B). This conserved pathway for inserting  $\beta$ -barrel proteins further underscores the endosymbiotic origin of mitochondria.

### Transport Into the Inner Mitochondrial Membrane and Intermembrane Space Occurs Via Several Routes

The same mechanism that transports proteins into the matrix space using the TOM and TIM23 translocators (see Figure 12–22) also mediates the initial translocation of many proteins that are destined for the inner mitochondrial membrane or the intermembrane space. In the most common translocation route, only the N-terminal signal sequence of the transported protein actually enters the matrix space (Figure 12–25A). A hydrophobic amino acid sequence, strategically placed after the N-terminal signal sequence, acts as a *stop-transfer sequence*, preventing



**Figure 12–25 Protein import from the cytosol into the inner mitochondrial membrane and intermembrane space.** (A) The N-terminal signal sequence (red) initiates import into the matrix space (see Figure 12–22). A hydrophobic sequence (blue) that follows the matrix-targeting signal sequence binds to the TIM23 translocator (orange) in the inner membrane and stops translocation. The remainder of the protein is then pulled into the intermembrane space through the TOM translocator in the outer membrane, and the hydrophobic sequence is released into the inner membrane anchoring the protein there. (B) A second route for protein integration into the inner membrane first delivers the protein completely into the matrix space. Cleavage of the signal sequence (red) used for the initial translocation unmasks an adjacent hydrophobic signal sequence (blue) at the new N-terminus. This signal then directs the protein into the inner membrane, using the same OXA-dependent pathway that inserts proteins that are encoded by the mitochondrial genome and translated in the matrix space. (C) Some soluble proteins of the intermembrane space also use the pathways shown in (A) and (B) before they are released into the intermembrane space by a second signal peptidase, which has its active site in the intermembrane space and removes the hydrophobic signal sequence. (D) Some soluble intermembrane-space proteins become oxidized by the Mia40 protein (Mia = mitochondrial intermembrane space assembly) during import. Mia40 forms a covalent intermediate through an intermolecular disulfide bond, which helps pull the transported protein through the TOM complex. Mia40 becomes reduced in the process, and then is reoxidized by the electron transport chain, so that it can catalyze the next round of import. (E) Multipass inner membrane proteins that function as metabolite transporters contain internal signal sequences and snake through the TOM complex as a loop. They then bind to the chaperones in the intermembrane space, which guide the proteins to the TIM22 complex. The TIM22 complex is specialized for the insertion of multipass inner membrane proteins.

further translocation across the inner membrane. The remainder of the protein then crosses the outer membrane through the TOM complex into the intermembrane space; the signal sequence is cleaved off in the matrix, and the hydrophobic sequence, released from TIM23, remains anchored in the inner membrane.

In another transport route to the inner membrane or intermembrane space, the TIM23 complex initially translocates the entire protein into the matrix space (Figure 12–25B). A matrix signal peptidase then removes the N-terminal signal sequence, exposing a hydrophobic sequence at the new N-terminus. This signal sequence guides the protein to the OXA complex, which inserts the protein into the inner membrane. As mentioned earlier, the OXA complex is primarily used to insert proteins that are encoded and translated in the mitochondrion into the inner membrane, and only a few imported proteins use this pathway. Translocators that are closely related to the OXA complex are found in the plasma membrane of bacteria and in the thylakoid membrane of chloroplasts, where they insert membrane proteins by a similar mechanism.

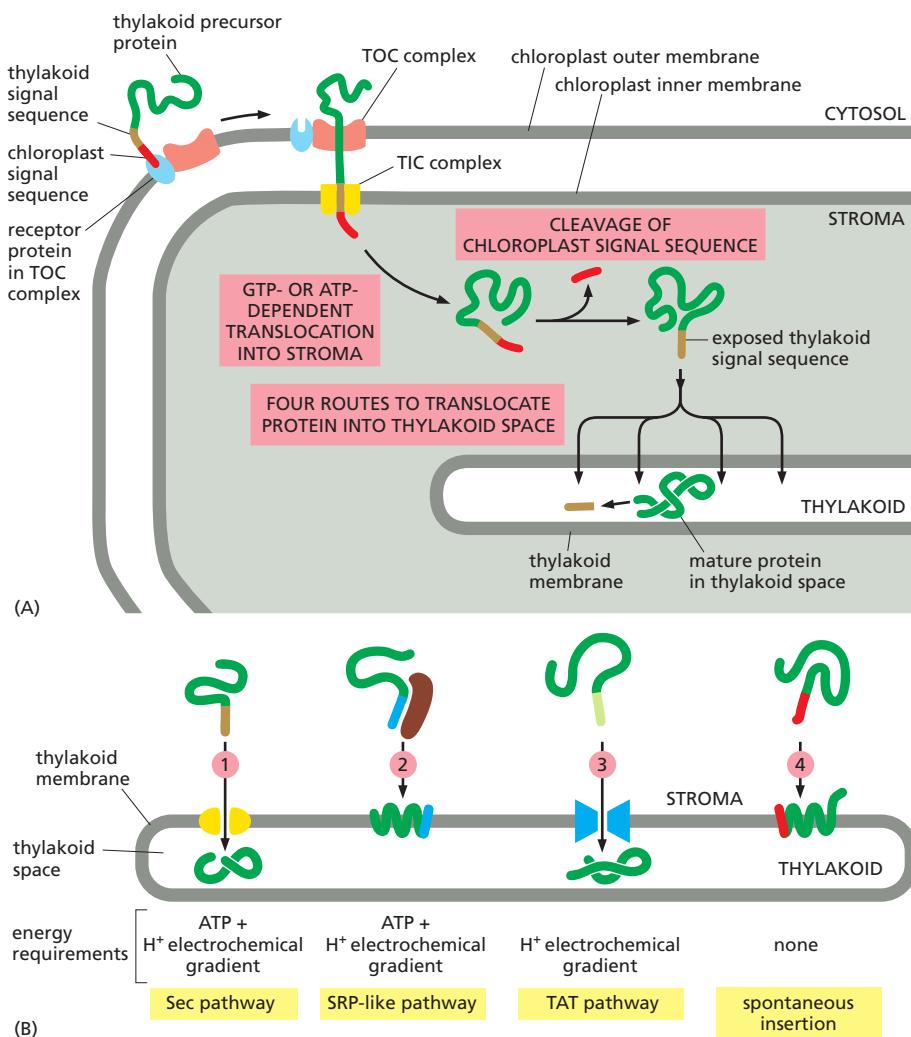
Many proteins that use these pathways to the inner membrane remain anchored there through their hydrophobic signal sequence (see Figure 12–25A,B). Others, however, are released into the intermembrane space by a protease that removes the membrane anchor (Figure 12–25C). Many of these cleaved proteins remain attached to the outer surface of the inner membrane as peripheral subunits of protein complexes that also contain transmembrane proteins.

Certain intermembrane-space proteins that contain cysteine motifs are imported by a yet different route. These proteins form a transient covalent disulfide bond to the Mia40 protein (Figure 12–25D). The imported proteins are then released in an oxidized form containing intrachain disulfide bonds. Mia40 becomes reduced in the process, and is then reoxidized by passing electrons to the electron transport chain in the inner mitochondrial membrane. In this way, the energy stored in the redox potential in the mitochondrial electron transport chain is tapped to drive protein import.

Mitochondria are the principal sites of ATP synthesis in the cell, but they also contain many metabolic enzymes, such as those of the citric acid cycle. Thus, in addition to proteins, mitochondria must also transport small metabolites across their membranes. While the outer membrane contains porins, which make the membrane freely permeable to such small molecules, the inner membrane does not. Instead, a family of metabolite-specific transporters transfers a vast number of small molecules across the inner membrane. In yeast cells, these transporters comprise a family of 35 different proteins, the most abundant of which transport ATP, ADP, and phosphate. These are multipass transmembrane proteins, which do not have cleavable signal sequences at their N-termini but instead contain internal signal sequences. They cross the TOM complex in the outer membrane, and intermembrane-space chaperones guide them to the TIM22 complex, which inserts them into the inner membrane by a process that requires the membrane potential, but not mitochondrial hsp70 or ATP (Figure 12–25E). An energetically favorable partitioning of the hydrophobic transmembrane regions into the inner membrane is likely to drive this process.

### Two Signal Sequences Direct Proteins to the Thylakoid Membrane in Chloroplasts

Protein transport into **chloroplasts** resembles transport into mitochondria. Both processes occur post-translationally, use separate translocation complexes in each membrane, require energy, and use amphiphilic N-terminal signal sequences that are removed after use. With the exception of some of the chaperone molecules, however, the protein components that form the translocation complexes differ. Moreover, whereas mitochondria harness the electrochemical H<sup>+</sup> gradient across their inner membrane to drive transport, chloroplasts, which have an electrochemical H<sup>+</sup> gradient across their thylakoid membrane but not their inner membrane, use GTP and ATP hydrolysis to power import across their



**Figure 12-26** Translocation of chloroplast precursor proteins into the thylakoid space. (A) The precursor protein contains an N-terminal chloroplast signal sequence (red), followed immediately by a thylakoid signal sequence (brown).

The chloroplast signal sequence initiates translocation into the stroma by a mechanism similar to that used for the translocation of mitochondrial precursor proteins into the matrix space, although the translocator complexes, TOC and TIC, are different. The signal sequence is then cleaved off, unmasking the thylakoid signal sequence, which initiates translocation across the thylakoid membrane.

(B) Translocation into the thylakoid space or thylakoid membrane can occur by any one of at least four routes: (1) a Sec pathway, so called because it uses components that are homologs of Sec proteins, which mediate protein translocation across the bacterial plasma membrane (discussed later); (2) an SRP-like pathway, so called because it uses a chloroplast homolog of the signal-recognition particle, or SRP (discussed later); (3) a TAT (twin arginine translocation) pathway, so called because two arginines are critical in the signal sequences that direct proteins into this pathway, which depends on the H<sup>+</sup> gradient across the thylakoid membrane; and (4) a spontaneous insertion pathway that seems not to require any protein translocator.

double membrane. The functional similarities may thus result from convergent evolution, reflecting the common requirements for translocation across a double membrane.

Although the signal sequences for import into chloroplasts superficially resemble those for import into mitochondria, the same plant cells have both mitochondria and chloroplasts, so proteins must partition appropriately between the two organelles. In plants, for example, a bacterial enzyme can be directed specifically to mitochondria if it is experimentally joined to an N-terminal signal sequence of a mitochondrial protein; the same enzyme joined to an N-terminal signal sequence of a chloroplast protein ends up in chloroplasts. Thus, the import receptors on each organelle distinguish between the different signal sequences.

Chloroplasts have an extra membrane-enclosed compartment, the **thylakoid**. Many chloroplast proteins, including the protein subunits of the photosynthetic system and of the ATP synthase (discussed in Chapter 14), are located in the thylakoid membrane. Like the precursors of some mitochondrial proteins, the precursors of these proteins are translocated from the cytosol to their final destination in two steps. First, they pass across the double membrane into the matrix space (called the **stroma** in chloroplasts), and then they either integrate into the thylakoid membrane or translocate into the thylakoid space (Figure 12-26A). The precursors of these proteins have a hydrophobic thylakoid signal sequence following the N-terminal chloroplast signal sequence. After the N-terminal signal sequence has been used to import the protein into the stroma, a stromal signal peptidase removes it, unmasking the thylakoid signal sequence that initiates transport

across the thylakoid membrane. There are at least four routes by which proteins cross or become integrated into the thylakoid membrane, distinguished by their need for different stromal chaperones and energy sources (Figure 12–26B).

## Summary

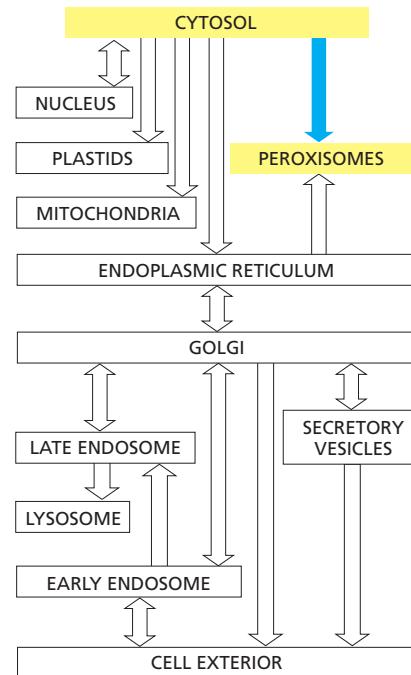
*Although mitochondria and chloroplasts have their own genetic systems, they produce only a small proportion of their own proteins. Instead, the two organelles import most of their proteins from the cytosol, using similar mechanisms. In both cases, proteins are transported in an unfolded state across both outer and inner membranes simultaneously into the matrix space or stroma. Both ATP hydrolysis and a membrane potential across the inner membrane drive translocation into mitochondria, whereas GTP and ATP hydrolysis drive translocation into chloroplasts. Chaperone proteins of the cytosolic hsp70 family maintain the precursor proteins in an unfolded state, and a second set of hsp70 proteins in the matrix space or stroma pulls the polypeptide chain into the organelle. Only proteins that contain a specific signal sequence are translocated. The signal sequence can either be located at the N-terminus and cleaved off after import or be internal and retained. Transport into the inner membrane sometimes uses a second, hydrophobic signal sequence that is unmasked when the first signal sequence is removed. In chloroplasts, import from the stroma into the thylakoid can occur by several routes, distinguished by the chaperones and energy source used.*

## PEROXISOMES

**Peroxisomes** differ from mitochondria and chloroplasts in many ways. Most notably, they are surrounded by only a single membrane, and they do not contain DNA or ribosomes. Thus, because peroxisomes lack a genome, all of their proteins are encoded in the nucleus. Peroxisomes acquire most of these proteins by selective import from the cytosol, although some of them enter the peroxisome membrane via the ER.

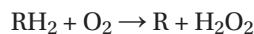
Because we do not discuss peroxisomes elsewhere, we shall digress to consider some of the functions of this diverse family of organelles, before discussing their biosynthesis. Virtually all eukaryotic cells have peroxisomes. They contain oxidative enzymes, such as *catalase* and *urate oxidase*, at such high concentrations that, in some cells, the peroxisomes stand out in electron micrographs because of the presence of a crystalloid protein core (Figure 12–27).

Like mitochondria, peroxisomes are major sites of oxygen utilization. One hypothesis is that peroxisomes are a vestige of an ancient organelle that performed all the oxygen metabolism in the primitive ancestors of eukaryotic cells. When the oxygen produced by photosynthetic bacteria first accumulated in the atmosphere, it would have been highly toxic to most cells. Peroxisomes might have lowered the intracellular concentration of oxygen, while also exploiting its chemical reactivity to perform useful oxidation reactions. According to this view, the later development of mitochondria rendered peroxisomes largely obsolete because many of the same biochemical reactions—which had formerly been carried out in peroxisomes without producing energy—were now coupled to ATP formation by means of oxidative phosphorylation. The oxidation reactions performed by peroxisomes in present-day cells could therefore partly be those whose functions were not taken over by mitochondria.

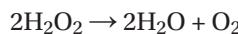


## Peroxisomes Use Molecular Oxygen and Hydrogen Peroxide to Perform Oxidation Reactions

Peroxisomes are so named because they usually contain one or more enzymes that use molecular oxygen to remove hydrogen atoms from specific organic substrates (designated here as R) in an oxidation reaction that produces *hydrogen peroxide* ( $H_2O_2$ ):



*Catalase* uses the H<sub>2</sub>O<sub>2</sub> generated by other enzymes in the organelle to oxidize a variety of other substrates—including formic acid, formaldehyde, and alcohol—by the “peroxidation” reaction: H<sub>2</sub>O<sub>2</sub> + R'H<sub>2</sub> → R' + 2H<sub>2</sub>O. This type of oxidation reaction is particularly important in liver and kidney cells, where the peroxisomes detoxify various harmful molecules that enter the bloodstream. About 25% of the ethanol we drink is oxidized to acetaldehyde in this way. In addition, when excess H<sub>2</sub>O<sub>2</sub> accumulates in the cell, catalase converts it to H<sub>2</sub>O through the reaction



A major function of the oxidation reactions performed in peroxisomes is the breakdown of fatty acid molecules. The process, called  $\beta$  *oxidation*, shortens the alkyl chains of fatty acids sequentially in blocks of two carbon atoms at a time, thereby converting the fatty acids to acetyl CoA. The peroxisomes then export the acetyl CoA to the cytosol for use in biosynthetic reactions. In mammalian cells,  $\beta$  oxidation occurs in both mitochondria and peroxisomes; in yeast and plant cells, however, this essential reaction occurs exclusively in peroxisomes.

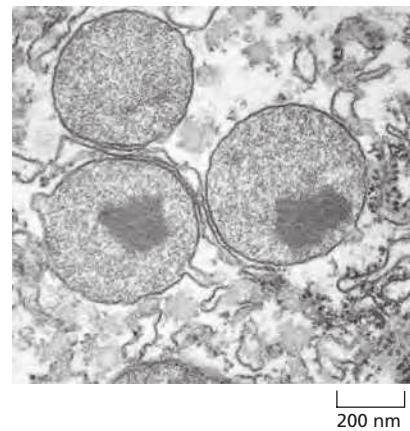
An essential biosynthetic function of animal peroxisomes is to catalyze the first reactions in the formation of *plasmalogens*, which are the most abundant class of phospholipids in myelin (Figure 12–28). Plasmalogen deficiencies cause profound abnormalities in the myelination of nerve-cell axons, which is one reason why many peroxisomal disorders lead to neurological disease.

Peroxisomes are unusually diverse organelles, and even in the various cell types of a single organism they may contain different sets of enzymes. They also adapt remarkably to changing conditions. Yeasts grown on sugar, for example, have few small peroxisomes. But when some yeasts are grown on methanol, numerous large peroxisomes are formed that oxidize methanol; and when grown on fatty acids, they develop numerous large peroxisomes that break down fatty acids to acetyl CoA by  $\beta$  oxidation.

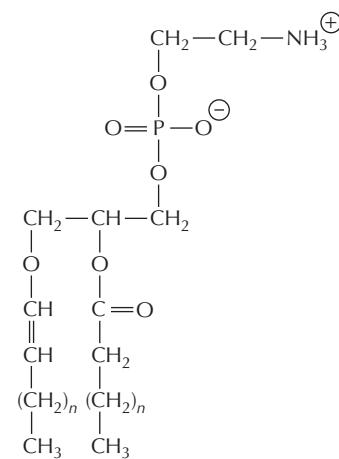
Peroxisomes are also important in plants. Two types of plant peroxisomes have been studied extensively. One is present in leaves, where it participates in *photorespiration* (discussed in Chapter 14) (Figure 12–29A). The other type of peroxisome is present in germinating seeds, where it converts the fatty acids stored in seed lipids into the sugars needed for the growth of the young plant. Because this conversion of fats to sugars is accomplished by a series of reactions known as the *glyoxylate cycle*, these peroxisomes are also called *glyoxysomes* (Figure 12–29B). In the glyoxylate cycle, two molecules of acetyl CoA produced by fatty acid breakdown in the peroxisome are used to make succinic acid, which then leaves the peroxisome and is converted into glucose in the cytosol. The glyoxylate cycle does not occur in animal cells, and animals are therefore unable to convert the fatty acids in fats into carbohydrates.

### A Short Signal Sequence Directs the Import of Proteins into Peroxisomes

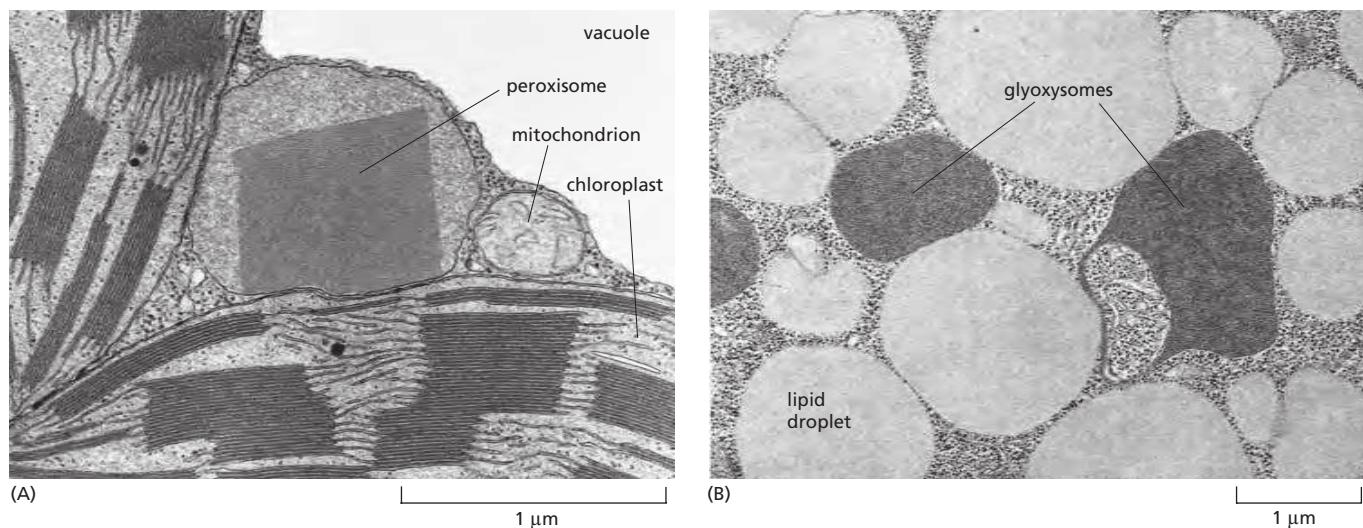
A specific sequence of three amino acids (Ser-Lys-Leu) located at the C-terminus of many peroxisomal proteins functions as an import signal (see Table 12–3, p. 648). Other peroxisomal proteins contain a signal sequence near the N-terminus. If either sequence is attached to a cytosolic protein, the protein is imported into peroxisomes. The import signals are first recognized by soluble receptor proteins in the cytosol. Numerous distinct proteins, called *peroxins*, participate in the import process, which is driven by ATP hydrolysis. A complex of at least six different peroxins forms a protein translocator in the peroxisome membrane. Even oligomeric proteins do not have to unfold to be imported. To allow the passage of such compactly folded cargo molecules, the pore formed by the transporter is thought to be dynamic in its dimensions, adapting in size to the particular cargo molecules to be transported. In this respect, the mechanism differs from that used by mitochondria and chloroplasts. One soluble import receptor, the peroxin Pex5 recognizes the C-terminal peroxisomal import signal. It accompanies its cargo all the way into peroxisomes and, after cargo release, cycles back to the cytosol. After



**Figure 12–27** An electron micrograph of three peroxisomes in a rat liver cell. The paracrystalline, electron-dense inclusions are composed primarily of the enzyme urate oxidase. (Courtesy of Daniel S. Friend.)



**Figure 12–28** The structure of a plasmalogen. Plasmalogens are very abundant in the myelin sheaths that insulate the axons of nerve cells. They make up some 80–90% of the myelin membrane phospholipids. In addition to an ethanolamine head group and a long-chain fatty acid attached to the same glycerol phosphate backbone used for phospholipids, plasmalogens contain an unusual fatty alcohol that is attached through an ether linkage (bottom left).

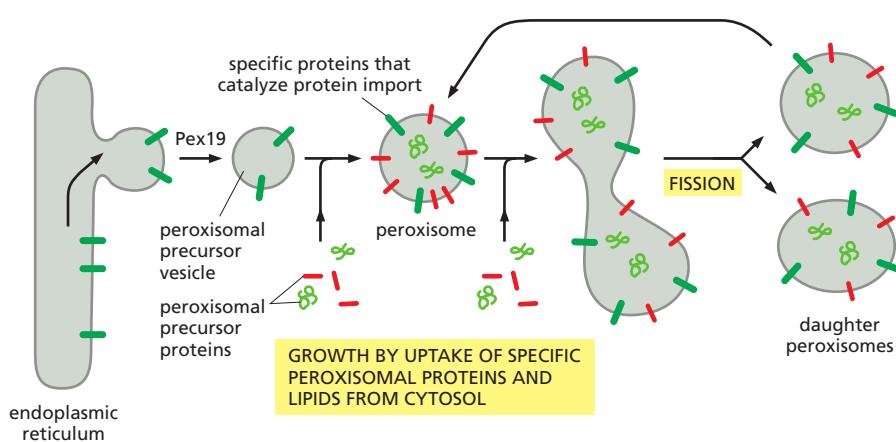


**Figure 12–29** Electron micrographs of two types of peroxisomes found in plant cells. (A) A peroxisome with a paracrystalline core in a tobacco leaf mesophyll cell. Its close association with chloroplasts is thought to facilitate the exchange of materials between these organelles during photorespiration. The vacuole in plant cells is equivalent to the lysosome in animal cells. (B) Peroxisomes in a fat-storing cotyledon cell of a tomato seed 4 days after germination. Here the peroxisomes (glyoxysomes) are associated with the lipid droplets that store fat, reflecting their central role in fat mobilization and gluconeogenesis during seed germination. (A, from S.E. Frederick and E.H. Newcomb, *J. Cell Biol.* 43:343–353, 1969. With permission from The Rockefeller Press; B, from W.P. Wergin, P.J. Gruber and E.H. Newcomb, *J. Ultrastruct. Res.* 30:533–557, 1970. With permission from Academic Press.)

delivering its cargo to the peroxisome lumen, Pex5 undergoes ubiquitylation. This modification is required to release Pex5 back into the cytosol, where the ubiquitin is removed. An ATPase composed of Pex1 and Pex6 harnesses the energy of ATP hydrolysis to help release Pex5 from peroxisomes.

The importance of this import process and of peroxisomes is demonstrated by the inherited human disease *Zellweger syndrome*, in which a defect in importing proteins into peroxisomes leads to a profound peroxisomal deficiency. These individuals, whose cells contain “empty” peroxisomes, have severe abnormalities in their brain, liver, and kidneys, and they die soon after birth. A mutation in the gene encoding peroxin Pex5 causes one form of the disease. A defect in Pex7, the receptor for the N-terminal import signal, causes a milder peroxisomal disease.

It has long been debated whether new peroxisomes arise from preexisting ones by organelle growth and fission—as mentioned earlier for mitochondria and plastids—or whether they derive as a specialized compartment from the endoplasmic reticulum (ER). Aspects of both views are true (Figure 12–30). Most peroxisomal membrane proteins are made in the cytosol and insert into the membrane of



**Figure 12–30** A model that explains how peroxisomes proliferate and how new peroxisomes arise. Peroxisomal precursor vesicles bud from the ER. At least two peroxisomal membrane proteins, Pex3 and Pex15, follow this route. The machinery that drives the budding reaction and that selects only peroxisomal proteins for packaging into these vesicles depends on Pex19 and other cytosolic proteins that are still unknown. Peroxisomal precursor vesicles may then fuse with one another or with preexisting peroxisomes. The peroxisome membrane contains import receptors and protein translocators that are required for the import of peroxisomal proteins made on cytosolic ribosomes, including new copies of the import receptors and translocator components. Presumably, the lipids required for growth are also imported, although some may derive directly from the ER in the membrane of peroxisomal precursor vesicles.

preexisting peroxisomes, but others are first integrated into the ER membrane, where they are packaged into specialized peroxisomal precursor vesicles. New precursor vesicles may then fuse with one another and begin importing additional peroxisomal proteins, using their own protein import machinery to grow into mature peroxisomes, which can undergo cycles of growth and fission.

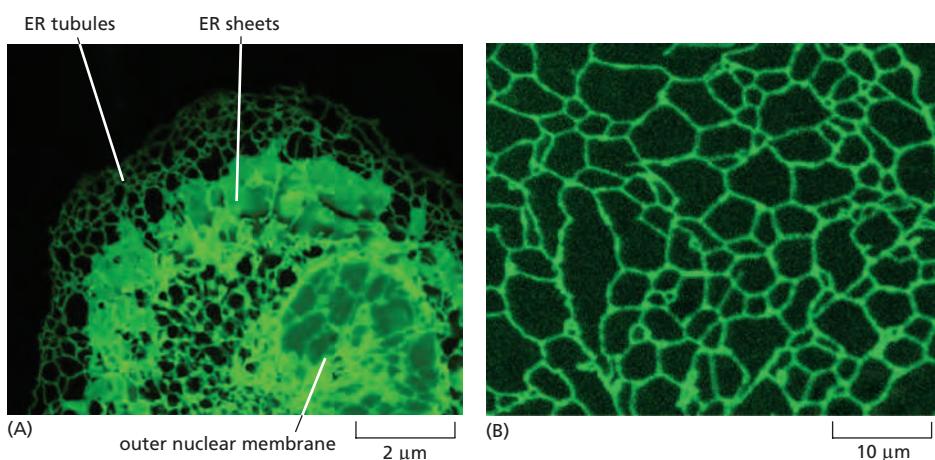
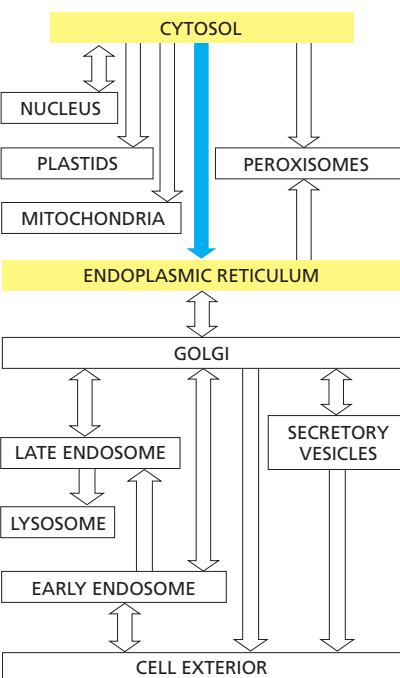
## Summary

*Peroxisomes are specialized for carrying out oxidation reactions using molecular oxygen. They generate hydrogen peroxide, which they employ for oxidative purposes—and contain catalase to destroy the excess. Like mitochondria and plastids, peroxisomes are self-replicating organelles. Because they do not contain DNA or ribosomes, however, all of their proteins are encoded in the cell nucleus. Some of these proteins are conveyed to peroxisomes via peroxisomal precursor vesicles that bud from the ER, but most are synthesized in the cytosol and directly imported. A specific sequence of three amino acids near the C-terminus of many of the latter proteins functions as a peroxisomal import signal. The mechanism of protein import differs from that of mitochondria and chloroplasts, in that even oligomeric proteins are imported from the cytosol without unfolding.*

## THE ENDOPLASMIC RETICULUM

All eukaryotic cells have an **endoplasmic reticulum (ER)**. Its membrane typically constitutes more than half of the total membrane of an average animal cell (see Table 12–2, p. 643). The ER is organized into a netlike labyrinth of branching tubules and flattened sacs that extends throughout the cytosol (Figure 12–31 and Movie 12.4). The tubules and sacs interconnect, and their membrane is continuous with the outer nuclear membrane; the compartment that they enclose therefore is also continuous with the space between the inner and outer nuclear membranes. Thus, the ER and nuclear membranes form a continuous sheet enclosing a single internal space, called the **ER lumen** or the **ER cisternal space**, which often occupies more than 10% of the total cell volume (see Table 12–1, p. 643).

As mentioned at the beginning of this chapter, the ER has a central role in both lipid and protein biosynthesis, and it also serves as an intracellular  $\text{Ca}^{2+}$  store that is used in many cell signaling responses (discussed in Chapter 15). The ER membrane is the site of production of all the transmembrane proteins and lipids for most of the cell's organelles, including the ER itself, the Golgi apparatus, lysosomes, endosomes, secretory vesicles, and the plasma membrane. The ER membrane is also the site at which most of the lipids for mitochondrial and peroxisomal membranes are made. In addition, almost all of the proteins that will be secreted to the cell exterior—plus those destined for the lumen of the ER, Golgi apparatus, or lysosomes—are initially delivered to the ER lumen.



**Figure 12–31 Fluorescent micrographs of the endoplasmic reticulum.** (A) An animal cell in tissue culture that was genetically engineered to express an ER membrane protein fused to a fluorescent protein. The ER extends as a network of tubules and sheets throughout the entire cytosol, so that all regions of the cytosol are close to some portion of the ER membrane. The outer nuclear membrane, which is continuous with the ER, is also stained. (B) Part of an ER network in a living plant cell that was genetically engineered to express a fluorescent protein in the ER. (A, courtesy of Patrick Chitwood and Gia Voeltz; B, courtesy of Petra Boevink and Chris Hawes.)

## The ER Is Structurally and Functionally Diverse

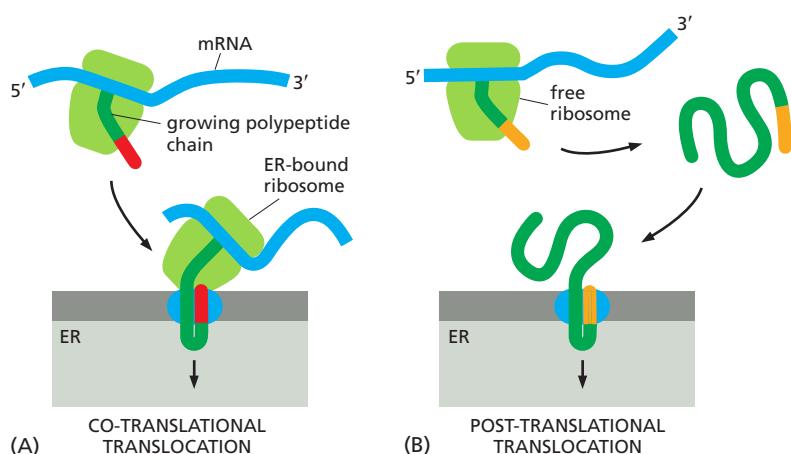
While the various functions of the ER are essential to every cell, their relative importance varies greatly between individual cell types. To meet different functional demands, distinct regions of the ER become highly specialized. We observe such functional specialization as dramatic changes in ER structure, and different cell types can therefore possess characteristically different types of ER membrane. One of the most remarkable ER specializations is the **rough ER**.

Mammalian cells begin to import most proteins into the ER before complete synthesis of the polypeptide chain—that is, import is a **co-translational** process (Figure 12–32A). In contrast, the import of proteins into mitochondria, chloroplasts, nuclei, and peroxisomes is a **post-translational** process (Figure 12–32B). In co-translational transport, the ribosome that is synthesizing the protein is attached directly to the ER membrane, enabling one end of the protein to be translocated into the ER while the rest of the polypeptide chain is being synthesized. These membrane-bound ribosomes coat the surface of the ER, creating regions termed **rough endoplasmic reticulum**, or **rough ER**; regions of ER that lack bound ribosomes are called **smooth endoplasmic reticulum**, or **smooth ER** (Figure 12–33).

Most cells have scanty regions of smooth ER, and the ER is often partly smooth and partly rough. Areas of smooth ER from which transport vesicles carrying newly synthesized proteins and lipids bud off for transport to the Golgi apparatus are called **transitional ER**. In certain specialized cells, the smooth ER is abundant and has additional functions. It is prominent, for example, in cells that specialize in lipid metabolism, such as cells that synthesize steroid hormones from cholesterol; the expanded smooth ER accommodates the enzymes that make cholesterol and modify it to form the hormones (see Figure 12–33B).

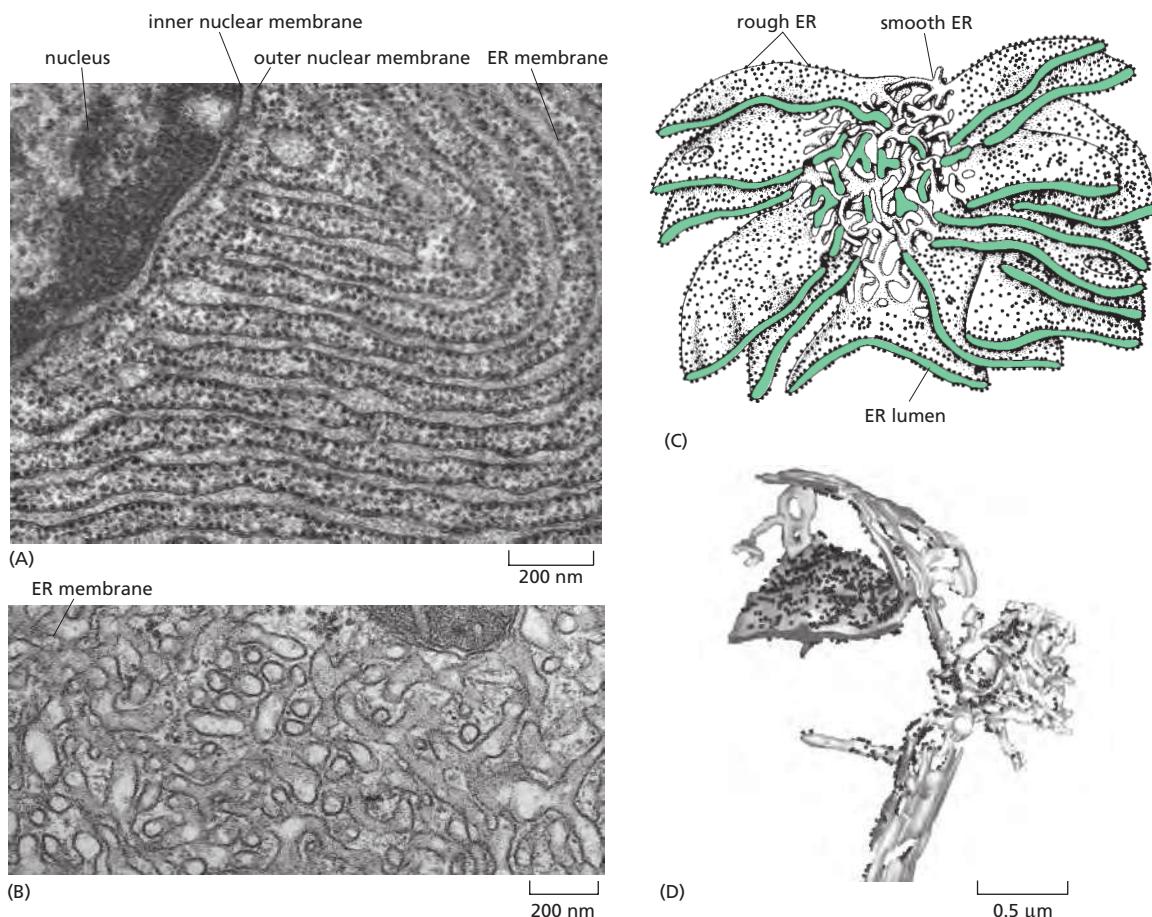
The main cell type in the liver, the *hepatocyte*, also has a substantial amount of smooth ER. It is the principal site of production of *lipoprotein particles*, which carry lipids via the bloodstream to other parts of the body. The enzymes that synthesize the lipid components of the particles are located in the membrane of the smooth ER, which also contains enzymes that catalyze a series of reactions to detoxify both lipid-soluble drugs and various harmful compounds produced by metabolism. The most extensively studied of these *detoxification reactions* are carried out by the *cytochrome P450* family of enzymes, which catalyze a series of reactions in which water-insoluble drugs or metabolites that would otherwise accumulate to toxic levels in cell membranes are rendered sufficiently water-soluble to leave the cell and be excreted in the urine. Because the rough ER alone cannot house enough of these and other necessary enzymes, a substantial portion of the membrane in a hepatocyte normally consists of smooth ER (see Table 12–2).

Another crucially important function of the ER in most eukaryotic cells is to sequester  $\text{Ca}^{2+}$  from the cytosol. The release of  $\text{Ca}^{2+}$  into the cytosol from the ER, and its subsequent reuptake, occurs in many rapid responses to extracellular



**Figure 12–32 Co-translational and post-translational protein translocation.**

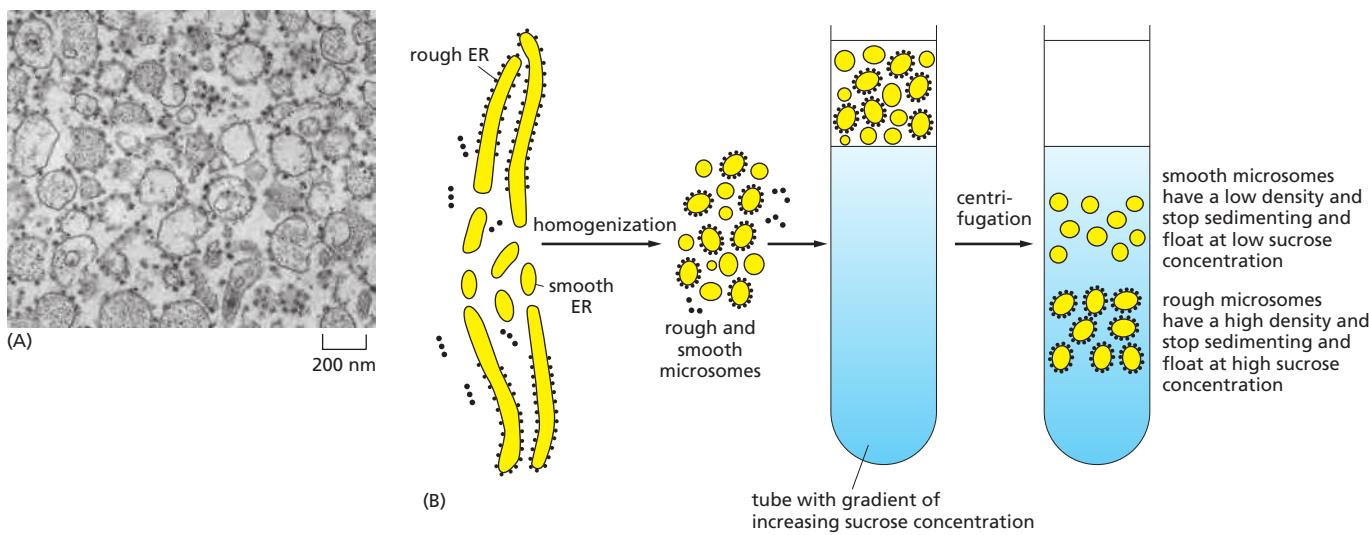
(A) Ribosomes bind to the ER membrane during co-translational translocation. (B) By contrast, cytosolic ribosomes complete the synthesis of a protein and release it prior to post-translational translocation. In both cases, the protein is directed to the ER by an ER signal sequence (red and orange).



**Figure 12-33 The rough and smooth ER.** (A) An electron micrograph of the rough ER in a pancreatic exocrine cell that makes and secretes large amounts of digestive enzymes every day. The cytosol is filled with closely packed sheets of ER membrane that is studded with ribosomes. At the top left is a portion of the nucleus and its nuclear envelope; note that the outer nuclear membrane, which is continuous with the ER, is also studded with ribosomes. (B) Abundant smooth ER in a steroid-hormone-secreting cell. This electron micrograph is of a testosterone-secreting Leydig cell in the human testis. (C) A three-dimensional reconstruction of a region of smooth ER and rough ER in a liver cell. The rough ER forms oriented stacks of flattened cisternae, each having a luminal space 20–30 nm wide. The smooth ER membrane is connected to these cisternae and forms a fine network of tubules 30–60 nm in diameter. The ER lumen is colored green. (D) A tomographic reconstruction of a portion of the ER network in a yeast cell. Membrane-bound ribosomes (tiny dark spheres) are seen in both flat sheets and tubular regions of irregular diameter, demonstrating that the ribosomes bind to ER membranes of different curvature in these cells. (A, courtesy of Lelio Orci; B, courtesy of Daniel S. Friend; C, after R.V. Krstić, Ultrastructure of the Mammalian Cell. New York: Springer-Verlag, 1979; D, from M. West et al., J. Cell Biol. 193:333–346, 2011. With permission from Rockefeller University Press.)

signals, as discussed in Chapter 15. A  $\text{Ca}^{2+}$  pump transports  $\text{Ca}^{2+}$  from the cytosol into the ER lumen. A high concentration of  $\text{Ca}^{2+}$ -binding proteins in the ER facilitates  $\text{Ca}^{2+}$  storage. In some cell types, and perhaps in most, specific regions of the ER are specialized for  $\text{Ca}^{2+}$  storage. Muscle cells have an abundant, modified smooth ER called the *sarcoplasmic reticulum*. The release and reuptake of  $\text{Ca}^{2+}$  by the sarcoplasmic reticulum trigger myofibril contraction and relaxation, respectively, during each round of muscle contraction (discussed in Chapter 16).

To study the functions and biochemistry of the ER, it is necessary to isolate it. This may seem to be a hopeless task because the ER is intricately interleaved with other components of the cytoplasm. Fortunately, when tissues or cells are disrupted by homogenization, the ER breaks into fragments, which reseal to form small (~100–200 nm in diameter) closed vesicles called **microsomes**. Microsomes are relatively easy to purify. To the biochemist, microsomes represent small authentic versions of the ER, still capable of protein translocation, protein glycosylation (discussed later),  $\text{Ca}^{2+}$  uptake and release, and lipid synthesis. Microsomes derived from rough ER are studded with ribosomes and are called **rough**



**microsomes.** The ribosomes are always found on the outside surface, so the interior of the microsome is biochemically equivalent to the lumen of the ER (Figure 12–34A).

Many vesicles similar in size to rough microsomes, but lacking attached ribosomes, are also found in cell homogenates. Such *smooth microsomes* are derived in part from smooth portions of the ER and in part from vesiculated fragments of the plasma membrane, Golgi apparatus, endosomes, and mitochondria (the ratio depending on the tissue). Thus, whereas rough microsomes are clearly derived from rough portions of ER, it is not easy to separate smooth microsomes derived from different organelles. The smooth microsomes prepared from liver or muscle cells are an exception. Because of the unusually large quantities of smooth ER or sarcoplasmic reticulum, respectively, most of the smooth microsomes in homogenates of these tissues are derived from the smooth ER or sarcoplasmic reticulum. The ribosomes attached to rough microsomes make them more dense than smooth microsomes. As a result, we can use equilibrium centrifugation to separate the rough and smooth microsomes (Figure 12–34B). Microsomes have been invaluable in elucidating the molecular aspects of ER function, as we discuss next.

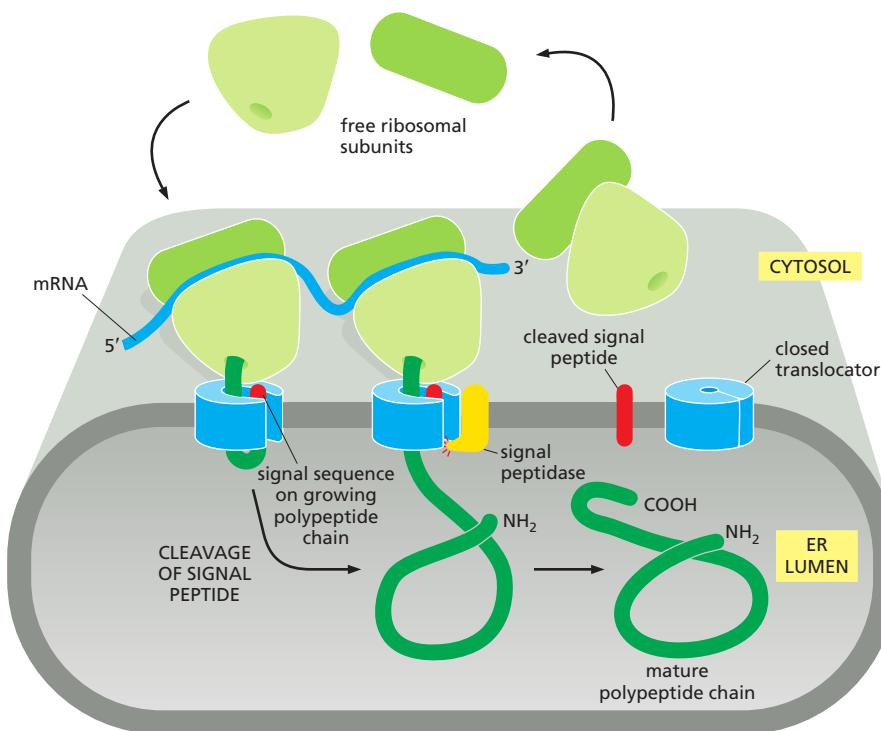
### Signal Sequences Were First Discovered in Proteins Imported into the Rough ER

The ER captures selected proteins from the cytosol as they are being synthesized. These proteins are of two types: *transmembrane proteins*, which are only partly translocated across the ER membrane and become embedded in it, and *water-soluble proteins*, which are fully translocated across the ER membrane and are released into the ER lumen. Some of the transmembrane proteins function in the ER, but many are destined to reside in the plasma membrane or the membrane of another organelle. The water-soluble proteins are destined either for secretion or for residence in the lumen of the ER or of another organelle. All of these proteins, regardless of their subsequent fate, are directed to the ER membrane by an **ER signal sequence**, which initiates their translocation by a common mechanism.

Signal sequences (and the signal sequence strategy of protein sorting) were first discovered in the early 1970s in secreted proteins that are translocated across the ER membrane as a first step toward their eventual discharge from the cell. In the key experiment, the mRNA encoding a secreted protein was translated by ribosomes *in vitro*. When microsomes were omitted from this cell-free system, the protein synthesized was slightly larger than the normal secreted protein. In the presence of microsomes derived from the rough ER, however, a protein of the correct size was produced. According to the *signal hypothesis*, the size difference reflects the initial presence of a signal sequence that directs the secreted protein

**Figure 12–34 The isolation of purified rough and smooth microsomes from the ER.** (A) A thin section electron micrograph of the purified rough ER fraction shows an abundance of ribosome-studded vesicles.

(B) When sedimented to equilibrium through a gradient of sucrose, the two types of microsomes separate from each other on the basis of their different densities. Note that the smooth fraction will also contain non-ER-derived material. (A, courtesy of George Palade.)

**Figure 12–35** The signal hypothesis.

A simplified view of protein translocation across the ER membrane, as originally proposed. When the ER signal sequence emerges from the ribosome, it directs the ribosome to a translocator on the ER membrane that forms a pore in the membrane through which the polypeptide is translocated. A signal peptidase is closely associated with the translocator and clips off the signal sequence during translation, and the mature protein is released into the lumen of the ER immediately after its synthesis is completed. The translocator is closed until the ribosome has bound, so that the permeability barrier of the ER membrane is maintained at all times.

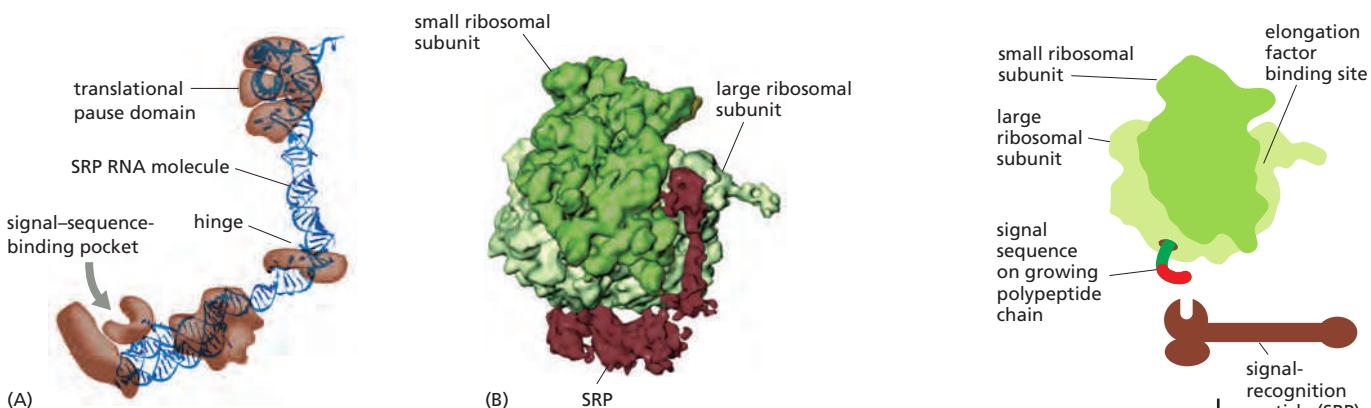
to the ER membrane and is then cleaved off by a *signal peptidase* in the ER membrane before the polypeptide chain has been completed (Figure 12–35). Cell-free systems in which proteins are imported into microsomes have provided powerful procedures for identifying, purifying, and studying the various components of the molecular machinery responsible for the ER import process.

### A Signal-Recognition Particle (SRP) Directs the ER Signal Sequence to a Specific Receptor in the Rough ER Membrane

The ER signal sequence is guided to the ER membrane by at least two components: a **signal-recognition particle (SRP)**, which cycles between the ER membrane and the cytosol and binds to the signal sequence, and an **SRP receptor** in the ER membrane. The SRP is a large complex; in animal cells, it consists of six different polypeptide chains bound to a single small RNA molecule. While the SRP and SRP receptor have fewer subunits in bacteria, homologs are present in all cells, indicating that this protein-targeting mechanism arose early in evolution and has been conserved.

ER signal sequences vary greatly in amino acid sequence, but each has eight or more nonpolar amino acids at its center (see Table 12–3, p. 648). How can the SRP bind specifically to so many different sequences? The answer has come from the crystal structure of the SRP protein, which shows that the signal-sequence-binding site is a large hydrophobic pocket lined by methionines. Because methionines have unbranched, flexible side chains, the pocket is sufficiently plastic to accommodate hydrophobic signal sequences of different sequences, sizes, and shapes.

The SRP is a rodlike structure, which wraps around the large ribosomal subunit, with one end binding to the ER signal sequence as it emerges from the ribosome as part of the newly made polypeptide chain; the other end blocks the elongation factor binding site at the interface between the large and small ribosomal subunits (Figure 12–36). This block halts protein synthesis as soon as the signal peptide has emerged from the ribosome. The transient pause presumably gives the ribosome enough time to bind to the ER membrane before completion of the polypeptide chain, thereby ensuring that the protein is not released into the cytosol. This safety

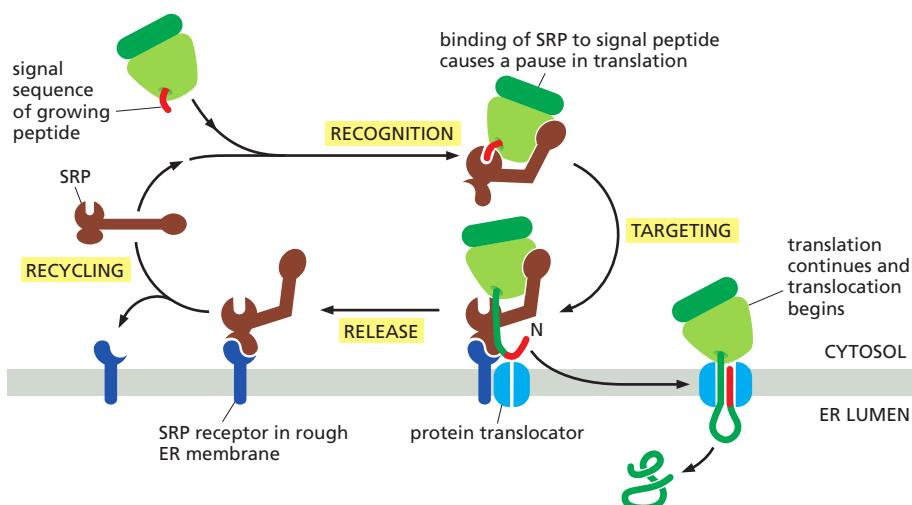


**Figure 12-36 The signal-recognition particle (SRP).** (A) A mammalian SRP is a rodlike ribonucleoprotein complex containing six protein subunits (brown) and one RNA molecule (blue). The SRP RNA forms a backbone that links the protein domain containing the signal-sequence-binding pocket to the domain responsible for pausing translation. Crystal structures of various SRP pieces from different species are assembled here into a composite model to approximate the structure of a complete SRP. (B) The three-dimensional outline of the SRP bound to a ribosome was determined by cryoelectron microscopy. SRP binds to the large ribosomal subunit so that its signal-sequence-binding pocket is positioned near the growing polypeptide chain exit site, and its translational pause domain is positioned at the interface between the ribosomal subunits, where it interferes with elongation factor binding. (C) As a signal sequence emerges from the ribosome and binds to the SRP, a conformational change in the SRP exposes a binding site for the SRP receptor. (B, adapted from M. Halic et al., *Nature* 427:808–814, 2004. With permission from Macmillan Publishers Ltd.)

device may be especially important for secreted and lysosomal hydrolases, which could wreak havoc in the cytosol; cells that secrete large amounts of hydrolases, however, take the added precaution of having high concentrations of hydrolase inhibitors in their cytosol. The pause also ensures that large portions of a protein that could fold into a compact structure are not made before reaching the translocator in the ER membrane. Thus, in contrast to the post-translational import of proteins into mitochondria and chloroplasts, chaperone proteins are not required to keep the protein unfolded.

When a signal sequence binds, SRP exposes a binding site for the SRP receptor (see Figure 12-36B,C), which is a transmembrane protein complex in the rough ER membrane. The binding of the SRP to its receptor brings the SRP-ribosome complex to an unoccupied protein translocator in the same membrane. The SRP and SRP receptor are then released, and the translocator transfers the growing polypeptide chain across the membrane (Figure 12-37).

This co-translational transfer process creates two spatially separate populations of ribosomes in the cytosol. **Membrane-bound ribosomes**, attached to the



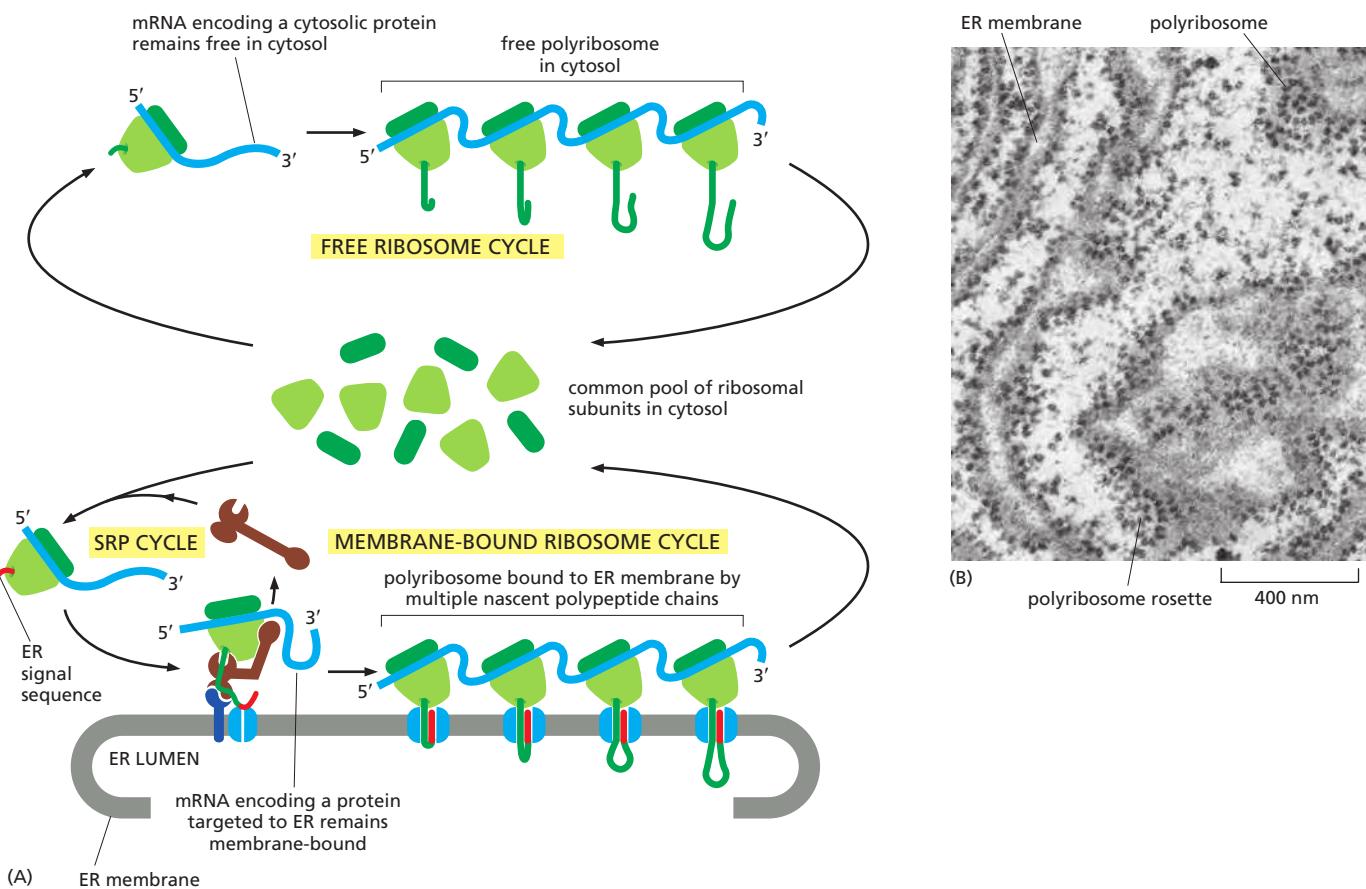
**Figure 12-37 How ER signal sequences and SRP direct ribosomes to the ER membrane.** The SRP and its receptor act in concert. The SRP binds to both the exposed ER signal sequence and the ribosome, thereby inducing a pause in translation. The SRP receptor in the ER membrane, which in animal cells is composed of two different polypeptide chains, binds the SRP-ribosome complex and directs it to the translocator. In a poorly understood reaction, the SRP and SRP receptor are then released, leaving the ribosome bound to the translocator in the ER membrane. The translocator then inserts the polypeptide chain into the membrane and transfers it across the lipid bilayer. Because one of the SRP proteins and both chains of the SRP receptor contain GTP-binding domains, it is thought that conformational changes that occur during cycles of GTP binding and hydrolysis (discussed in Chapter 15) ensure that SRP release occurs only after the ribosome has become properly engaged with the translocator in the ER membrane. The translocator is closed until the ribosome has bound, so that the permeability barrier of the ER membrane is maintained at all times.

cytosolic side of the ER membrane, are engaged in the synthesis of proteins that are being concurrently translocated into the ER. **Free ribosomes**, unattached to any membrane, synthesize all other proteins encoded by the nuclear genome. Membrane-bound and free ribosomes are structurally and functionally identical. They differ only in the proteins they are making at any given time.

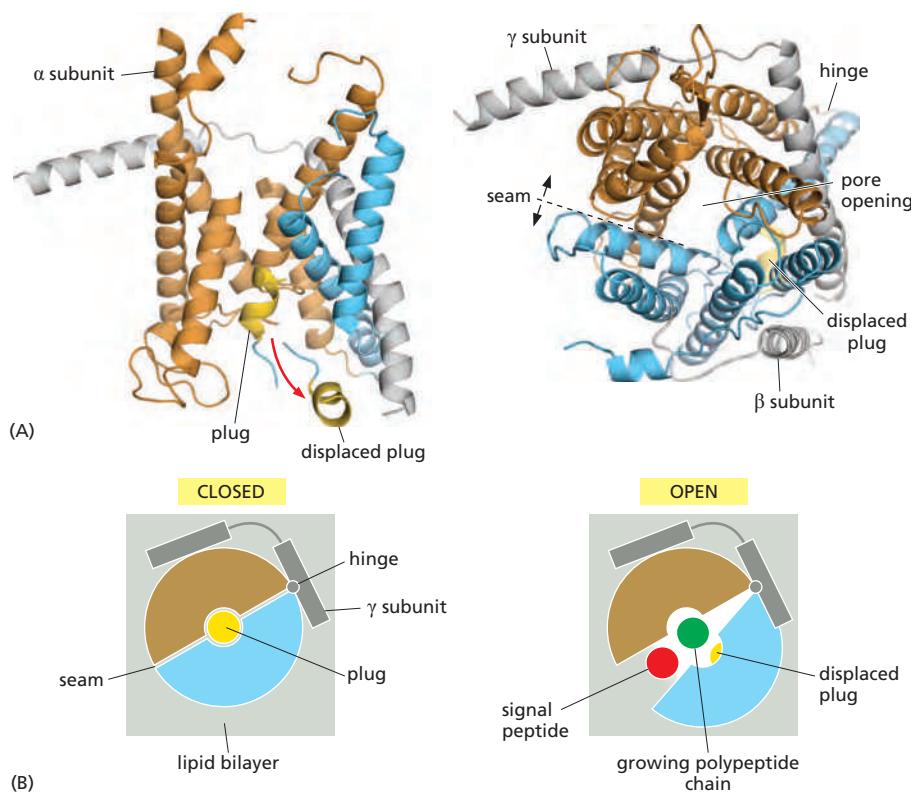
Since many ribosomes can bind to a single mRNA molecule, a **polyribosome** is usually formed. If the mRNA encodes a protein with an ER signal sequence, the polyribosome becomes attached to the ER membrane, directed there by the signal sequences on multiple growing polypeptide chains. The individual ribosomes associated with such an mRNA molecule can return to the cytosol when they finish translation and intermix with the pool of free ribosomes. The mRNA itself, however, remains attached to the ER membrane by a changing population of ribosomes, each transiently held at the membrane by the translocator (Figure 12–38).

### The Polypeptide Chain Passes Through an Aqueous Channel in the Translocator

It had long been debated whether polypeptide chains are transferred across the ER membrane in direct contact with the lipid bilayer or through a channel in a protein translocator. The debate ended with the identification of the translocator, which was shown to form a water-filled channel in the membrane through



**Figure 12–38 Free and membrane-bound polyribosomes.** (A) A common pool of ribosomes synthesizes the proteins that stay in the cytosol and those that are transported into the ER. The ER signal sequence on a newly formed polypeptide chain binds to SRP, which directs the translating ribosome to the ER membrane. The mRNA molecule remains permanently bound to the ER as part of a polyribosome, while the ribosomes that move along it are recycled; at the end of each round of protein synthesis, the ribosomal subunits are released and rejoin the common pool in the cytosol. (B) A thin section electron micrograph of polyribosomes attached to the ER membrane. The plane of section in some places cuts through the ER roughly parallel to the membrane, giving a face-on view of the rosettelike pattern of the polyribosomes. (B, courtesy of George Palade.)



**Figure 12-39 Structure of the Sec61 complex.** (A) A side view (left) and a top view (right, seen from the cytosol) of the structure of the Sec61 complex of the archaeon *Methanococcus jannaschii*. The Sec61 $\alpha$  subunit has an inverted repeat structure (see Figure 11–10) and is shown in blue and beige to indicate this pseudo-symmetry; the two smaller  $\beta$  and  $\gamma$  subunits are shown in gray. In the side view, some helices in front have been omitted to make the inside of the pore visible. The yellow short helix is thought to form a plug that seals the pore when the translocator is closed. To open, the complex rearranges itself to move the plug helix out of the way, as indicated by the red arrow. A ring of hydrophobic amino acid side chains is thought to form a tight-fitting diaphragm around translocating polypeptide chain to prevent leaks of other molecules across the membrane. The pore of the Sec61 complex can also open sideways at a lateral seam. (B) Models of the closed and open states of the translocator are shown in top view, illustrating how a signal sequence (or a transmembrane segment) could be released into the lipid bilayer after opening of the seam. (PDB codes: 1RH5 and 1RHZ.)

which the polypeptide chain passes. The core of the translocator, called the **Sec61 complex**, is built from three subunits that are highly conserved from bacteria to eukaryotic cells. The structure of the Sec61 complex suggests that  $\alpha$  helices contributed by the largest subunit surround a central channel through which the polypeptide chain traverses the membrane (Figure 12–39). The channel is gated by a short  $\alpha$  helix that is thought to keep the translocator closed when it is idle and to move aside when it is engaged in passing a polypeptide chain. According to this view, the pore is a dynamic gated channel that opens only transiently when a polypeptide chain traverses the membrane. In an idle translocator, it is important to keep the channel closed, so that the membrane remains impermeable to ions, such as  $\text{Ca}^{2+}$ , which otherwise would leak out of the ER. As a polypeptide chain is translocating, a ring of hydrophobic amino acid side chains is thought to provide a flexible seal to prevent ion leaks.

The structure of the Sec61 complex suggests that the pore can also open along a seam on its side. Indeed, some structures of the translocator show it locked in an open-seam conformation. This opening allows a translocating peptide chain lateral access into the hydrophobic core of the membrane, a process that is important both for the release of a cleaved signal peptide into the membrane (see Figure 12–35) and for the integration of transmembrane proteins into the bilayer, as we discuss later.

**Figure 12–40 A ribosome (green) bound to the ER protein translocator (blue).** (A) A side-view reconstruction of the complex from electron microscopic images. (B) A view of the translocator seen from the ER lumen. The translocator contains Sec61, accessory proteins, and detergent used in the preparation. Domains of accessory proteins extend across the membrane and form the luminal bulge. (C) A schematic drawing of a membrane-bound ribosome attached to the translocator, indicating the location of the tunnel in the large ribosomal subunit through which the growing polypeptide chain exits from the ribosome. The mRNA (not shown) would be located between the small and large ribosomal subunits. (Adapted from J.F. Ménéret et al., *J. Mol. Biol.* 348:445–457, 2005. With permission from Academic Press.)

In eukaryotic cells, four Sec61 complexes form a large translocator assembly that can be visualized on ER-bound ribosomes after detergent solubilization of the ER membrane (Figure 12–40). It is likely that this assembly includes other membrane complexes that associate with the translocator, such as enzymes that modify the growing polypeptide chain, including oligosaccharide transferase and the signal peptidase. The assembly of a translocator with these accessory components is called the **translocon**.

### Translocation Across the ER Membrane Does Not Always Require Ongoing Polypeptide Chain Elongation

As we have seen, translocation of proteins into mitochondria, chloroplasts, and peroxisomes occurs post-translationally, after the protein has been made and released into the cytosol, whereas translocation across the ER membrane usually occurs during translation (co-translationally). This explains why ribosomes are bound to the ER but not to other organelles.

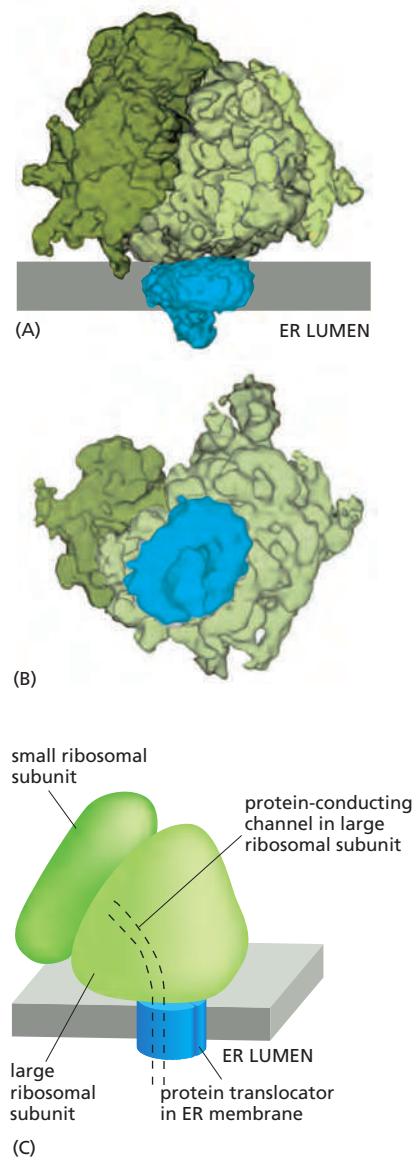
Some completely synthesized proteins, however, are imported into the ER, demonstrating that translocation does not always require ongoing translation. Post-translational protein translocation is especially common across the yeast ER membrane and the bacterial plasma membrane (which is thought to be evolutionarily related to the ER). To function in post-translational translocation, the ER translocator needs accessory proteins that feed the polypeptide chain into the pore and drive translocation (Figure 12–41). In bacteria, a translocation motor protein, the *SecA ATPase*, attaches to the cytosolic side of the translocator, where it undergoes cyclic conformational changes driven by ATP hydrolysis. Each time an ATP is hydrolyzed, a portion of the SecA protein inserts into the pore of the translocator, pushing a short segment of the passenger protein with it. As a result of this ratchet mechanism, the SecA ATPase progressively pushes the polypeptide chain of the transported protein across the membrane.

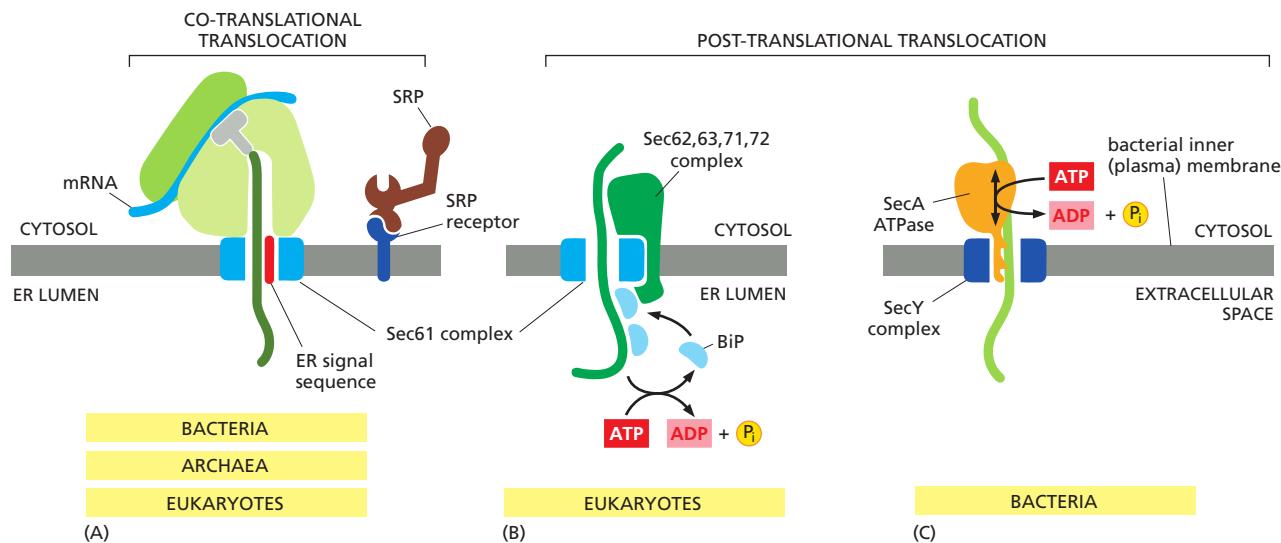
Eukaryotic cells use a different set of accessory proteins that associate with the Sec61 complex. These proteins span the ER membrane and use a small domain on the luminal side of the ER membrane to deposit an hsp70-like chaperone protein (called *BiP*, for *binding protein*) onto the polypeptide chain as it emerges from the pore into the ER lumen. ATP-dependent cycles of BiP binding and release drive unidirectional translocation, as described earlier for the mitochondrial hsp70 proteins that pull proteins across mitochondrial membranes.

Proteins that are transported into the ER by a post-translational mechanism are first released into the cytosol, where they bind to chaperone proteins to prevent folding, as discussed earlier for proteins destined for mitochondria and chloroplasts.

### In Single-Pass Transmembrane Proteins, a Single Internal ER Signal Sequence Remains in the Lipid Bilayer as a Membrane-spanning $\alpha$ Helix

The ER signal sequence in the growing polypeptide chain is thought to trigger the opening of the pore in the Sec61 protein translocator: after the signal sequence is released from the SRP and the growing chain has reached a sufficient length, the





**Figure 12–41** Three ways in which protein translocation can be driven through structurally similar translocators.

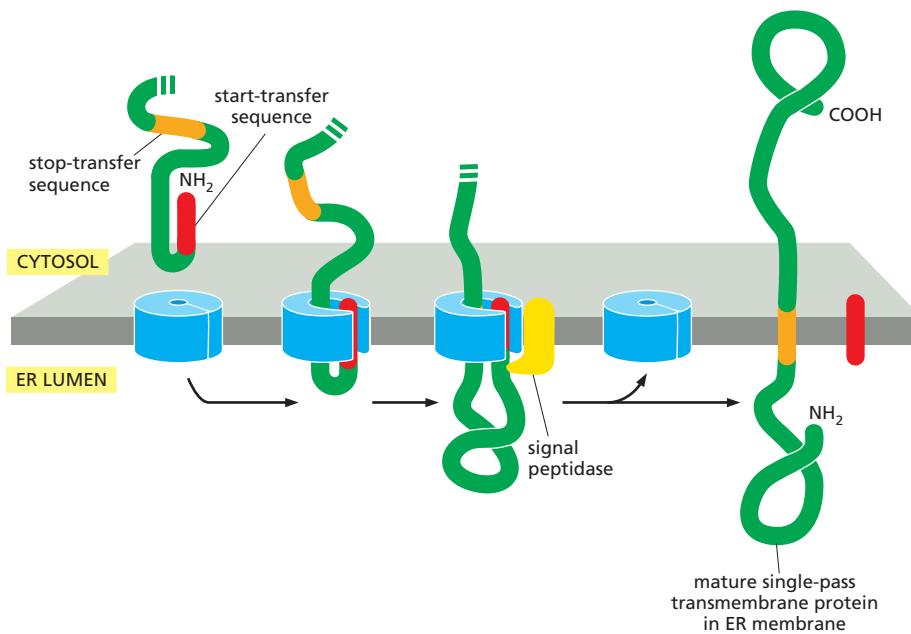
(A) Co-translational translocation. The ribosome is brought to the membrane by the SRP and SRP receptor and then engages with the Sec61 protein translocator. The growing polypeptide chain is threaded across the membrane as it is made. No additional energy is needed, as the only path available to the growing chain is to cross the membrane. (B) Post-translational translocation in eukaryotic cells requires an additional complex composed of Sec62, Sec63, Sec71, and Sec72 proteins, which is attached to the Sec61 translocator and deposits BiP molecules onto the translocating chain as it emerges from the translocator in the lumen of the ER. ATP-driven cycles of BiP binding and release pull the protein into the lumen, a mechanism that closely resembles the mechanism of mitochondrial import in Figure 12–23. (C) Post-translational translocation in bacteria. The completed polypeptide chain is fed from the cytosolic side into the bacterial homolog of the Sec61 complex (called the SecY complex in bacteria) in the plasma membrane by the SecA ATPase. ATP hydrolysis-driven conformational changes drive a pistonlike motion in SecA, each cycle pushing about 20 amino acids of the protein chain through the pore of the translocator. The Sec pathway used for protein translocation across the thylakoid membrane in chloroplasts uses a similar mechanism (see Figure 12–26B).

Whereas the Sec61 translocator, SRP, and SRP receptor are found in all organisms, SecA is found exclusively in bacteria, and the Sec62, 63, 71, 72 complex is found exclusively in eukaryotic cells. (Adapted from P. Walter and A.E. Johnson, *Annu. Rev. Cell Biol.* 10:87–119, 1994. With permission from Annual Reviews.)

signal sequence binds to a specific site inside the pore itself, thereby opening the pore. An ER signal sequence is therefore recognized twice: first by an SRP in the cytosol and then by a binding site in the pore of the protein translocator, where it serves as a **start-transfer signal** (or start-transfer peptide) that opens the pore (for example, see Figure 12–35 for how this works for a soluble protein). Dual recognition may help ensure that only appropriate proteins enter the lumen of the ER.

While bound in the translocation pore, a signal sequence is in contact not only with the Sec61 complex, which forms the walls of the pore, but also, along the lateral seam, with the hydrophobic core of the lipid bilayer. This was shown in chemical cross-linking experiments in which the signal sequence and the hydrocarbon chains of lipids were covalently linked together. When the nascent polypeptide chain grew long enough, the ER signal peptidase cleaved off the signal sequence and released it from the pore into the membrane, where it was rapidly degraded to amino acids by other proteases in the ER membrane. To release the signal sequence into the membrane, the translocator opens laterally along the seam (see Figures 12–35 and 12–39). The translocator is therefore gated in two directions: it opens to form a pore across the membrane to let the hydrophilic portions of proteins cross the lipid bilayer, and it opens laterally within the membrane to let hydrophobic portions of proteins partition into the lipid bilayer. Lateral gating of the pore is an essential step during the integration of transmembrane proteins.

The integration of membrane proteins requires that some parts of the polypeptide chain be translocated across the lipid bilayer whereas others are not. Despite this additional complexity, all modes of insertion of membrane proteins are simply variants of the sequence of events just described for transferring a soluble protein into the lumen of the ER. We begin by describing the three ways in



**Figure 12–42** How a single-pass transmembrane protein with a cleaved ER signal sequence is integrated into the ER membrane. In this protein, the co-translational translocation process is initiated by an N-terminal ER signal sequence (red) that functions as a start-transfer signal, opening the translocator as in Figure 12–35. In addition to this start-transfer sequence, however, the protein also contains a stop-transfer sequence (orange); when this sequence enters the translocator and interacts with a binding site within the pore, the translocator opens at the seam and discharges the protein laterally into the lipid bilayer, where the stop-transfer sequence remains to anchor the protein in the membrane. (In this figure and the two figures that follow, the ribosomes have been omitted for clarity.)

which **single-pass transmembrane proteins** (see Figure 10–17) become inserted into the ER membrane.

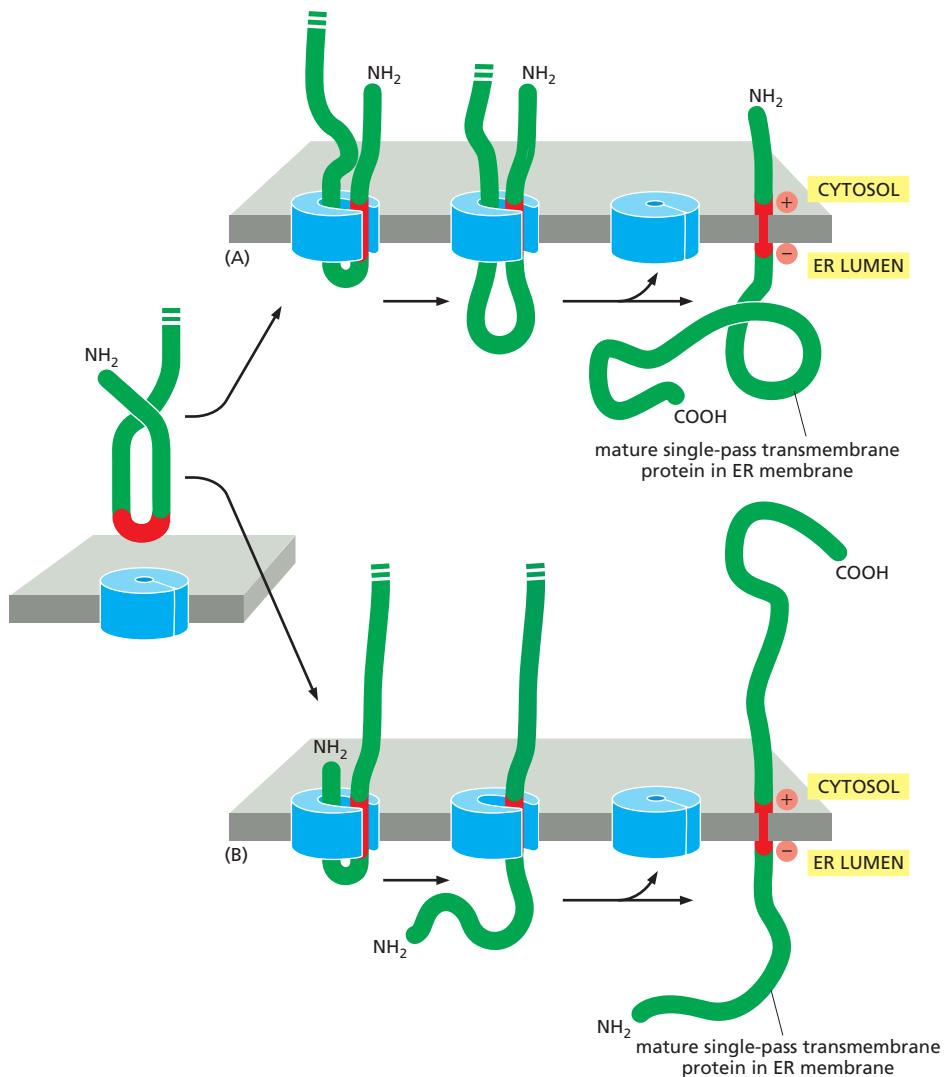
In the simplest case, an N-terminal signal sequence initiates translocation, just as for a soluble protein, but an additional hydrophobic segment in the polypeptide chain stops the transfer process before the entire polypeptide chain is translocated. This **stop-transfer signal** anchors the protein in the membrane after the ER signal sequence (the start-transfer signal) has been cleaved off and released from the translocator (Figure 12–42). The lateral gating mechanism transfers the stop-transfer sequence into the bilayer, where it remains as a single  $\alpha$ -helical membrane-spanning segment, with the N-terminus of the protein on the luminal side of the membrane and the C-terminus on the cytosolic side.

In the other two cases, the signal sequence is internal, rather than at the N-terminal end of the protein. As for an N-terminal ER signal sequence, the SRP binds to an internal signal sequence by recognizing its hydrophobic  $\alpha$ -helical features. The SRP brings the ribosome making the protein to the ER membrane, and the ER signal sequence then serves as a start-transfer signal that initiates the protein's translocation. After release from the translocator, the internal start-transfer sequence remains in the lipid bilayer as a single membrane-spanning  $\alpha$  helix.

Internal start-transfer sequences can bind to the translocation apparatus in either of two orientations; this in turn determines which protein segment (the one preceding or the one following the start-transfer sequence) is moved across the membrane into the ER lumen. In one case, the resulting membrane protein has its C-terminus on the luminal side (pathway A in Figure 12–43), while in the other, it has its N-terminus on the luminal side (pathway B in Figure 12–43). The orientation of the start-transfer sequence depends on the distribution of nearby charged amino acids, as described in the figure legend.

### Combinations of Start-Transfer and Stop-Transfer Signals Determine the Topology of Multipass Transmembrane Proteins

In **multipass transmembrane proteins**, the polypeptide chain passes back and forth repeatedly across the lipid bilayer as hydrophobic  $\alpha$  helices (see Figure 10–17). It is thought that an internal signal sequence serves as a start-transfer signal in these proteins to initiate translocation, which continues until the translocator encounters a stop-transfer sequence; in double-pass transmembrane proteins, for example, the polypeptide can then be released into the bilayer

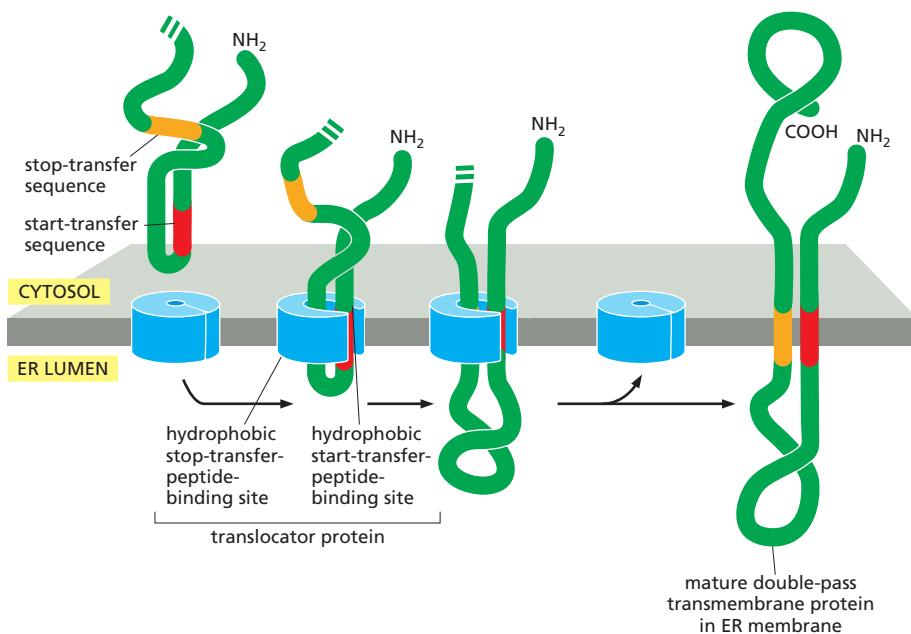


**Figure 12–43 Integration of a single-pass transmembrane protein with an internal signal sequence into the ER membrane.** An internal ER signal sequence that functions as a start-transfer signal can bind to the translocator in one of two ways, leading to a membrane protein that has either its C-terminus (pathway A) or its N-terminus (pathway B) in the ER lumen. Proteins are directed into either pathway by features in the polypeptide chain flanking the internal start-transfer sequence: if there are more positively charged amino acids immediately preceding the hydrophobic core of the start-transfer sequence than there are following it, the membrane protein is inserted into the translocator in the orientation shown in pathway A, whereas if there are more positively charged amino acids immediately following the hydrophobic core of the start-transfer sequence than there are preceding it, the membrane protein is inserted into the translocator in the orientation shown in pathway B. Because translocation cannot start before a start-transfer sequence appears outside the ribosome, translocation of the N-terminal portion of the protein shown in (B) can occur only after this portion has been fully synthesized.

Note that there are two ways to insert a single-pass membrane-spanning protein whose N-terminus is located in the ER lumen: that shown in Figure 12–42 and that shown here in (B).

**(Figure 12–44).** In more complex multipass proteins, in which many hydrophobic  $\alpha$  helices span the bilayer, a second start-transfer sequence reinitiates translocation further down the polypeptide chain until the next stop-transfer sequence causes polypeptide release, and so on for subsequent start-transfer and stop-transfer sequences (Figure 12–45 and Movie 12.5).

Hydrophobic start-transfer and stop-transfer signal sequences both act to fix the topology of the protein in the membrane by locking themselves into the membrane as membrane-spanning  $\alpha$  helices; and they can do this in either orientation. Whether a given hydrophobic signal sequence functions as a start-transfer or stop-transfer sequence must depend on its location in a polypeptide chain, since its function can be switched by changing its location in the protein by using recombinant DNA techniques. Thus, the distinction between start-transfer and stop-transfer sequences results mostly from their relative order in the growing polypeptide chain. It seems that the SRP begins scanning an unfolded polypeptide chain for hydrophobic segments at its N-terminus and proceeds toward the C-terminus, in the direction that the protein is synthesized. By recognizing the first appropriate hydrophobic segment to emerge from the ribosome, the SRP sets the “reading frame” for membrane integration: after the SRP initiates translocation, the translocator recognizes the next appropriate hydrophobic segment in the direction of transfer as a stop-transfer sequence, causing the region of the polypeptide chain in between to be threaded across the membrane. A similar

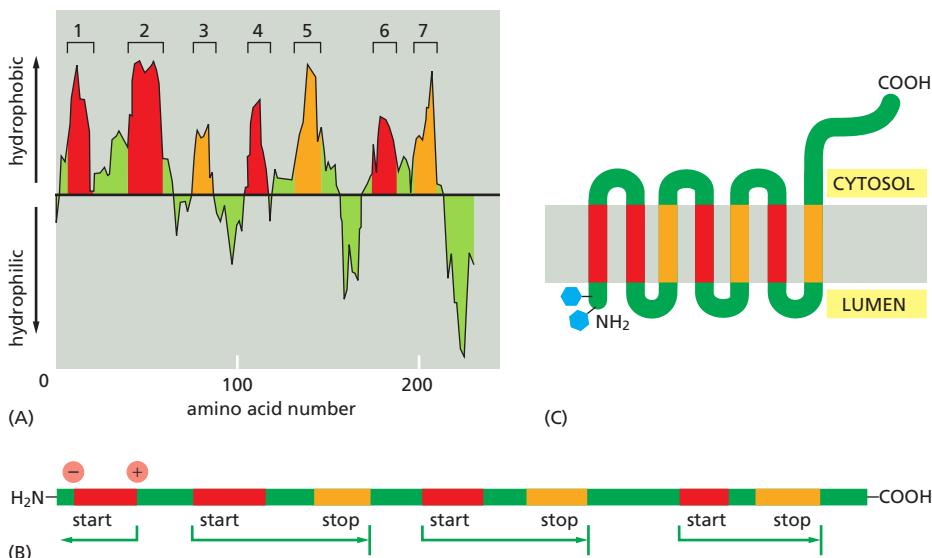


**Figure 12–44 Integration of a double-pass transmembrane protein with an internal signal sequence into the ER membrane.** In this protein, an internal ER signal sequence acts as a start-transfer signal (as in Figure 12–43) and initiates the transfer of the C-terminal part of the protein. At some point after a stop-transfer sequence has entered the translocator, the translocator discharges the sequence laterally into the membrane.

scanning process continues until all of the hydrophobic regions in the protein have been inserted into the membrane as transmembrane  $\alpha$  helices.

Because membrane proteins are always inserted from the cytosolic side of the ER in this programmed manner, all copies of the same polypeptide chain will have the same orientation in the lipid bilayer. This generates an asymmetrical ER membrane in which the protein domains exposed on one side are different from those exposed on the other side. This asymmetry is maintained during the many membrane budding and fusion events that transport the proteins made in the ER to other cell membranes (discussed in Chapter 13). Thus, the way in which a newly synthesized protein is inserted into the ER membrane determines the orientation of the protein in all of the other membranes as well.

When proteins are extracted with detergent from a membrane and then reconstituted into artificial lipid vesicles, a random mixture of right-side-out and inside-out protein orientations usually results. Thus, the protein asymmetry observed in cell membranes seems not to be an inherent property of the proteins, but instead results solely from the process by which proteins are inserted into the ER membrane from the cytosol.



**Figure 12–45 The insertion of the multipass membrane protein rhodopsin into the ER membrane.** Rhodopsin is the light-sensitive protein in rod photoreceptor cells in the mammalian retina (discussed in Chapter 15). (A) A hydrophathy plot (see Figure 10–20) identifies seven short hydrophobic regions in rhodopsin. (B) The hydrophobic region nearest the N-terminus serves as a start-transfer sequence that causes the preceding N-terminal portion of the protein to pass across the ER membrane. Subsequent hydrophobic sequences function in alternation as start-transfer and stop-transfer sequences. The green arrows indicate the paired start and stop signals inserted into the translocator. (C) The final integrated rhodopsin has its N-terminus located in the ER lumen and its C-terminus located in the cytosol. The blue hexagons represent covalently attached oligosaccharides.

## ER Tail-anchored Proteins Are Integrated into the ER Membrane by a Special Mechanism

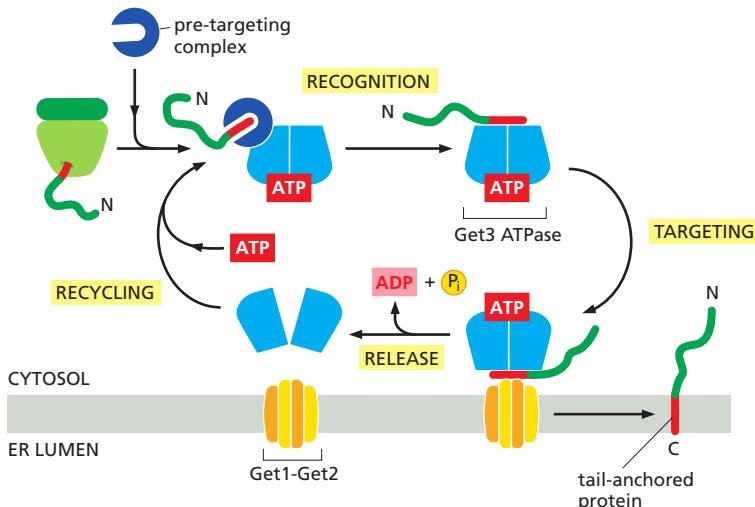
Many important membrane proteins are anchored in the membrane by a C-terminal transmembrane, hydrophobic  $\alpha$  helix. These **ER tail-anchored proteins** include a large number of SNARE protein subunits that guide vesicular traffic (discussed in Chapter 13). When such a tail-anchored protein inserts into the ER membrane from the cytosol, only a few amino acids that follow the transmembrane  $\alpha$  helix on its C-terminal side are translocated into the ER lumen, while most of the protein remains in the cytosol. Because of the unique position of the transmembrane  $\alpha$  helix in the protein sequence, translation terminates while the C-terminal amino acids that will form the transmembrane  $\alpha$  helix have not yet emerged from the ribosome exit tunnel. Recognition by SRP is therefore not possible. It was long thought that these proteins are released from the ribosome and the hydrophobic C-terminal tail spontaneously partitions into the ER membrane. Such a mechanism could not explain, however, why ER tail-anchored proteins insert into the ER membrane selectively and not also into all other membranes in the cell. It is now clear that a specialized targeting machinery is involved that is fueled by ATP hydrolysis (Figure 12–46). Although the components and details differ, this post-translational targeting mechanism is conceptually similar to SRP-dependent protein targeting (see Figure 12–37).

Not all tail-anchored proteins are inserted into the ER, however. Some proteins contain a C-terminal membrane anchor that contains additional sorting information that directs the protein to mitochondria or peroxisomes. How these proteins are sorted there remains unknown.

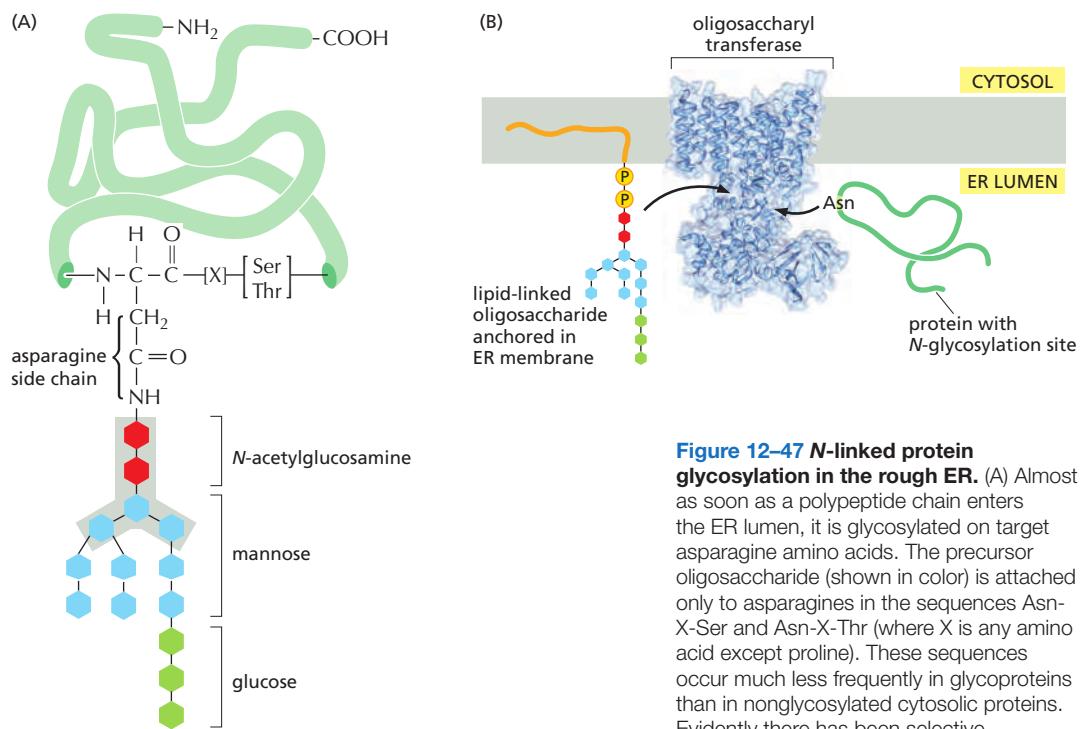
## Translocated Polypeptide Chains Fold and Assemble in the Lumen of the Rough ER

Many of the proteins in the lumen of the ER are in transit, *en route* to other destinations; others, however, normally reside there and are present at high concentrations. These **ER resident proteins** contain an **ER retention signal** of four amino acids at their C-terminus that is responsible for retaining the protein in the ER (see Table 12–3, p. 648; discussed in Chapter 13). Some of these proteins function as catalysts that help the many proteins that are translocated into the ER lumen to fold and assemble correctly.

One important ER resident protein is *protein disulfide isomerase (PDI)*, which catalyzes the oxidation of free sulphydryl (SH) groups on cysteines to form disulfide (S-S) bonds. Almost all cysteines in protein domains exposed to either the extracellular space or the lumen of organelles in the secretory and endocytic pathways are disulfide-bonded. By contrast, disulfide bonds form only very rarely in domains exposed to the cytosol, because of the reducing environment there.



**Figure 12–46** The insertion mechanism for tail-anchored proteins. In this post-translational pathway for the insertion of tail-anchored ER membrane proteins, a soluble pre-targeting complex captures the hydrophobic C-terminal  $\alpha$  helix after it emerges from the ribosomal exit tunnel and loads it onto the Get3 ATPase. The resulting complex is targeted to the ER membrane by interaction with the Get1–Get2 receptor complex that functions as a membrane protein insertion machine. After Get3 hydrolyzes its bound ATP, the tail-anchored protein is released from the receptor and inserted into the ER membrane. ADP release and renewed ATP binding recycles Get3 back to the cytosol.



**Figure 12–47** *N*-linked protein glycosylation in the rough ER. (A) Almost as soon as a polypeptide chain enters the ER lumen, it is glycosylated on target asparagine amino acids. The precursor oligosaccharide (shown in color) is attached only to asparagines in the sequences Asn-X-Ser and Asn-X-Thr (where X is any amino acid except proline). These sequences occur much less frequently in glycoproteins than in nonglycosylated cytosolic proteins. Evidently there has been selective pressure against these sequences during protein evolution, presumably because glycosylation at too many sites would interfere with protein folding. The five sugars in the gray box form the core region of this oligosaccharide. For many glycoproteins, only the core sugars survive the extensive oligosaccharide trimming that takes place in the Golgi apparatus. (B) The precursor oligosaccharide is transferred from a dolichol lipid anchor to the asparagine as an intact unit in a reaction catalyzed by a transmembrane oligosaccharyl transferase enzyme. One copy of this enzyme is associated with each protein translocator in the ER membrane. (The translocator is not shown.) Oligosaccharyl transferase contains 13 transmembrane  $\alpha$  helices and a large ER luminal domain that contains its substrate-binding sites. The asparagine binds a tunnel that penetrates the enzyme interior. There, the amino group of the asparagine is twisted out of the plane that stabilizes the otherwise poorly reactive amide bond, activating it for reaction with the dolichol-oligosaccharide. The structure shown is of a prokaryotic homolog that closely resembles the catalytic subunit of the eukaryotic oligosaccharyl transferase. (PDB code: 3RCE.)

Another ER resident protein is the chaperone protein BiP. We have already discussed how BiP pulls proteins post-translationally into the ER through the Sec61 ER translocator. Like other chaperones (discussed in Chapter 13), BiP recognizes incorrectly folded proteins, as well as protein subunits that have not yet assembled into their final oligomeric complexes. It does so by binding to exposed amino acid sequences that would normally be buried in the interior of correctly folded or assembled polypeptide chains. An example of a BiP-binding site is a stretch of alternating hydrophobic and hydrophilic amino acids that would normally be buried in a  $\beta$  sheet with its hydrophobic side oriented towards the hydrophobic core of the folded protein. The bound BiP both prevents the protein from aggregating and helps keep it in the ER (and thus out of the Golgi apparatus and later parts of the secretory pathway). Like some other members of the hsp70 family of chaperone proteins, which bind unfolded proteins and facilitate their import into mitochondria and chloroplasts, BiP hydrolyzes ATP to shuttle between high- and low-affinity binding states, which allow it to hold on to and let go of its substrate proteins in a dynamic cycle.

### Most Proteins Synthesized in the Rough ER Are Glycosylated by the Addition of a Common *N*-Linked Oligosaccharide

The covalent addition of oligosaccharides to proteins is one of the major biosynthetic functions of the ER. About half of the soluble and membrane-bound proteins that are processed in the ER—including those destined for transport to the Golgi apparatus, lysosomes, plasma membrane, or extracellular space—are **glycoproteins** that are modified in this way. Many proteins in the cytosol and nucleus are also glycosylated, but not with oligosaccharides: they carry a much simpler sugar modification, in which a single *N*-acetylglucosamine group is added to a serine or threonine of the protein.

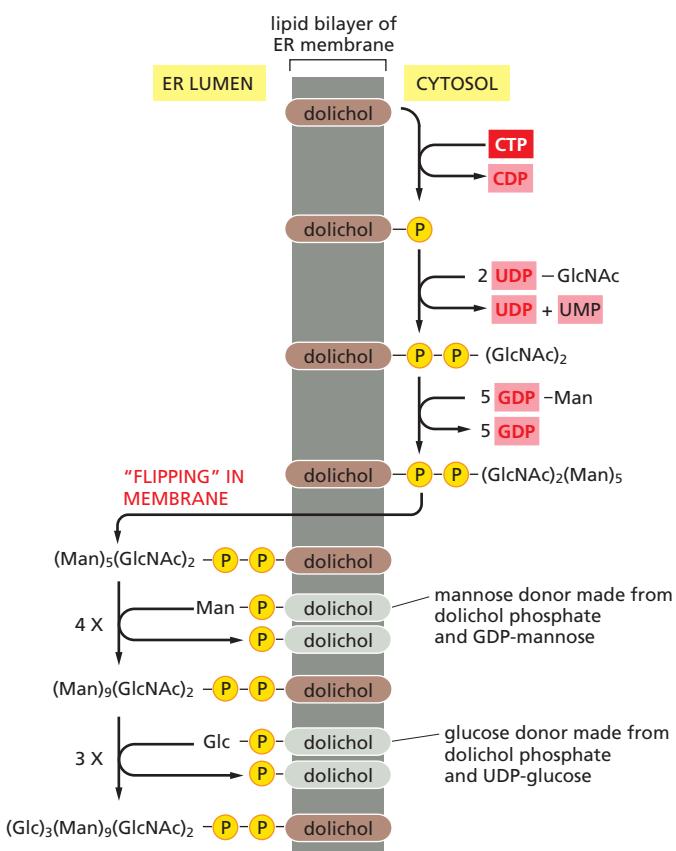
During the most common form of **protein glycosylation** in the ER, a pre-formed **precursor oligosaccharide** (composed of *N*-acetylglucosamine, mannose, and glucose, and containing a total of 14 sugars) is transferred *en bloc* to proteins. Because this oligosaccharide is transferred to the side-chain NH<sub>2</sub> group of an asparagine in the protein, it is said to be ***N*-linked** or **asparagine-linked** (Figure 12–47A). The transfer is catalyzed by a membrane-bound enzyme complex, an

*oligosaccharyl transferase*, which has its active site exposed on the luminal side of the ER membrane; this explains why cytosolic proteins are not glycosylated in this way. A special lipid molecule called **dolichol** anchors the precursor oligosaccharide in the ER membrane. The precursor oligosaccharide is transferred to the target asparagine in a single enzymatic step immediately after that amino acid has reached the ER lumen during protein translocation. The precursor oligosaccharide is linked to the dolichol lipid by a high-energy pyrophosphate bond, which provides the activation energy that drives the glycosylation reaction (Figure 12–47B). One copy of oligosaccharyl transferase is associated with each protein translocator, allowing it to scan and glycosylate the incoming polypeptide chains efficiently.

The precursor oligosaccharide is built up sugar by sugar on the membrane-bound dolichol lipid and is then transferred to a protein. The sugars are first activated in the cytosol by the formation of *nucleotide (UDP or GDP)-sugar intermediates*, which then donate their sugar (directly or indirectly) to the lipid in an orderly sequence. Part way through this process, the lipid-linked oligosaccharide is flipped, with the help of a transporter, from the cytosolic to the luminal side of the ER membrane (Figure 12–48).

All of the diversity of the *N*-linked oligosaccharide structures on mature glycoproteins results from the later modification of the original precursor oligosaccharide. While still in the ER, three glucoses (see Figure 12–47) and one mannose are quickly removed from the oligosaccharides of most glycoproteins. We shall return to the importance of glucose trimming shortly. This oligosaccharide “trimming,” or “processing,” continues in the Golgi apparatus, as we discuss in Chapter 13.

The *N*-linked oligosaccharides are by far the most common oligosaccharides, being found on 90% of all glycoproteins. Less frequently, oligosaccharides are linked to the hydroxyl group on the side chain of a serine, threonine, or hydroxylysine amino acid. A first sugar of these *O*-linked oligosaccharides is added in the ER and the oligosaccharide is then further extended in the Golgi apparatus (see Figure 13–32).



**Figure 12–48 Synthesis of the lipid-linked precursor oligosaccharide in the rough ER membrane.** The oligosaccharide is assembled sugar by sugar onto the carrier lipid dolichol (a polyisoprenoid; see Panel 2–5, pp. 98–99). Dolichol is long and very hydrophobic: its 22 five-carbon units can span the thickness of a lipid bilayer more than three times, so that the attached oligosaccharide is firmly anchored in the membrane. The first sugar is linked to dolichol by a pyrophosphate bridge. This high-energy bond activates the oligosaccharide for its eventual transfer from the lipid to an asparagine side chain of a growing polypeptide on the luminal side of the rough ER. As indicated, the synthesis of the oligosaccharide starts on the cytosolic side of the ER membrane and continues on the luminal face after the  $(\text{Man})_5(\text{GlcNAc})_2$  lipid intermediate is flipped across the bilayer by a transporter (which is not shown). All the subsequent glycosyl transfer reactions on the luminal side of the ER involve transfers from dolichol-P-glucose and dolichol-P-mannose; these activated, lipid-linked monosaccharides are synthesized from dolichol phosphate and UDP-glucose or GDP-mannose (as appropriate) on the cytosolic side of the ER and are then flipped across the ER membrane. GlcNAc = *N*-acetylglucosamine; Man = mannose; Glc = glucose.

## Oligosaccharides Are Used as Tags to Mark the State of Protein Folding

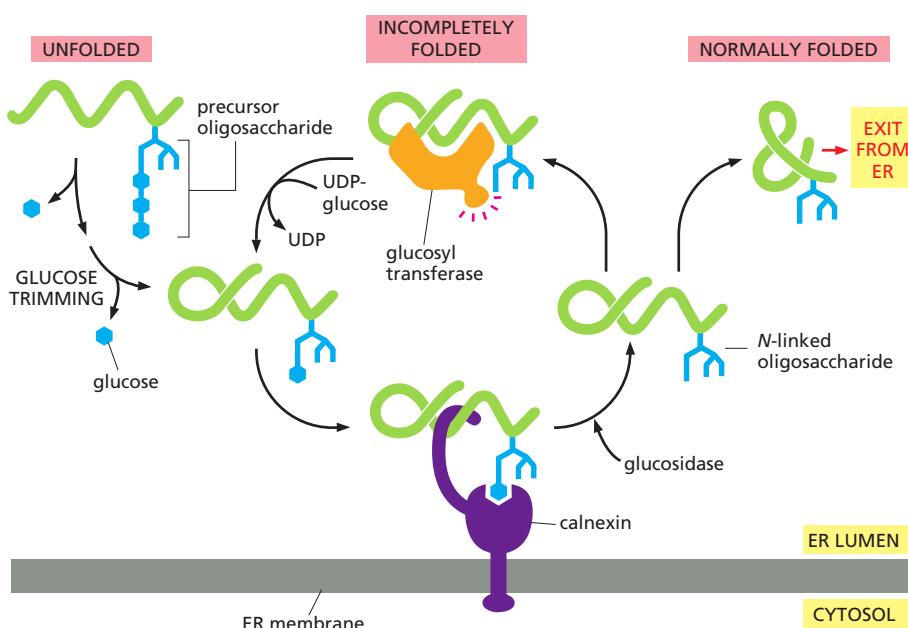
It has long been debated why glycosylation is such a common modification of proteins that enter the ER. One particularly puzzling observation has been that some proteins require *N*-linked glycosylation for proper folding in the ER, yet the precise location of the oligosaccharides attached to the protein's surface does not seem to matter. A clue to the role of glycosylation in protein folding came from studies of two ER chaperone proteins, which are called **calnexin** and **calreticulin** because they require  $\text{Ca}^{2+}$  for their activities. These chaperones are carbohydrate-binding proteins, or *lectins*, which bind to oligosaccharides on incompletely folded proteins and retain them in the ER. Like other chaperones, they prevent incompletely folded proteins from irreversibly aggregating. Both calnexin and calreticulin also promote the association of incompletely folded proteins with another ER chaperone, which binds to cysteines that have not yet formed disulfide bonds.

Calnexin and calreticulin recognize *N*-linked oligosaccharides that contain a single terminal glucose, and they therefore bind proteins only after two of the three glucoses on the precursor oligosaccharide have been removed during glucose trimming by ER glucosidases. When the third glucose has been removed, the glycoprotein dissociates from its chaperone and can leave the ER.

How, then, do calnexin and calreticulin distinguish properly folded from incompletely folded proteins? The answer lies in yet another ER enzyme, a glucosyl transferase that keeps adding a glucose to those oligosaccharides that have lost their last glucose. It adds the glucose, however, only to oligosaccharides that are attached to unfolded proteins. Thus, an unfolded protein undergoes continuous cycles of glucose trimming (by glucosidase) and glucose addition (by glucosyl transferase), maintaining an affinity for calnexin and calreticulin until it has achieved its fully folded state (**Figure 12–49**).

## Improperly Folded Proteins Are Exported from the ER and Degraded in the Cytosol

Despite all the help from chaperones, many protein molecules (more than 80% for some proteins) translocated into the ER fail to achieve their properly folded or oligomeric state. Such proteins are exported from the ER back into the cytosol, where they are degraded in proteasomes (discussed in Chapter 6). In many ways, the mechanism of retrotranslocation is similar to other post-translational



**Figure 12–49** The role of *N*-linked glycosylation in ER protein folding.

The ER-membrane-bound chaperone protein calnexin binds to incompletely folded proteins containing one terminal glucose on *N*-linked oligosaccharides, trapping the protein in the ER. Removal of the terminal glucose by a glucosidase releases the protein from calnexin. A glucosyl transferase is the crucial enzyme that determines whether the protein is folded properly or not: if the protein is still incompletely folded, the enzyme transfers a new glucose from UDP-glucose to the *N*-linked oligosaccharide, renewing the protein's affinity for calnexin and retaining it in the ER. The cycle repeats until the protein has folded completely. Calreticulin functions similarly, except that it is a soluble ER resident protein. Another ER chaperone, ERp57 (not shown), collaborates with calnexin and calreticulin in retaining an incompletely folded protein in the ER. ERp57 recognizes free sulfhydryl groups, which are a sign of incomplete disulfide bond formation.

modes of translocation. For example, like translocation into mitochondria or chloroplasts, chaperone proteins are necessary to keep the polypeptide chain in an unfolded state prior to and during translocation. Similarly, a source of energy is required to provide directionality to the transport and to pull the protein into the cytosol. Finally, a translocator is necessary.

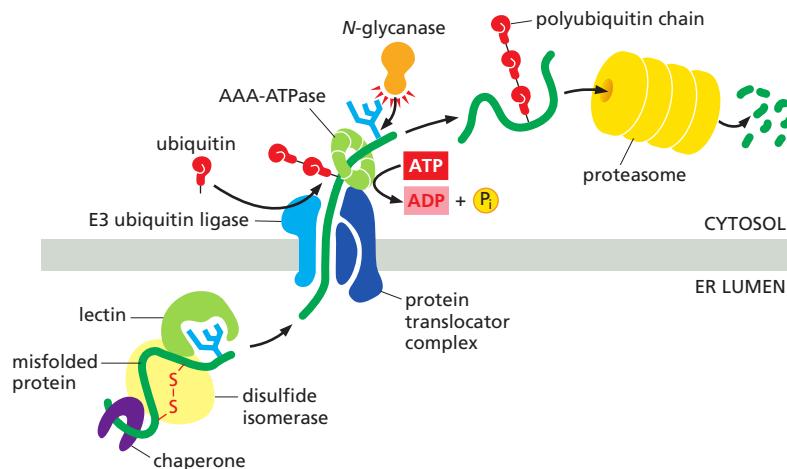
Selecting proteins from the ER for degradation is a challenging process: misfolded proteins or unassembled protein subunits should be degraded, but folding intermediates of newly made proteins should not. Help in making this distinction comes from the *N*-linked oligosaccharides, which serve as timers that measure how long a protein has spent in the ER. The slow trimming of a particular mannose on the core oligosaccharide tree by an enzyme (a mannosidase) in the ER creates a new oligosaccharide structure that ER-luminal lectins of the retrotranslocation apparatus recognize. Proteins that fold and exit from the ER faster than the mannosidase can remove its target mannose therefore escape degradation.

In addition to the lectins in the ER that recognize the oligosaccharides, chaperones and protein disulfide isomerases (enzymes mentioned earlier that catalyze the formation and breakage of S-S bonds) associate with the proteins that must be degraded. The chaperones prevent the unfolded proteins from aggregating, and the disulfide isomerases break disulfide bonds that may have formed incorrectly, so that a linear polypeptide chain can be translocated back into the cytosol.

Multiple translocator complexes move different proteins from the ER membrane or lumen into the cytosol. A common feature is that they each contain an E3 ubiquitin ligase enzyme, which attaches polyubiquitin tags to the unfolded proteins as they emerge into the cytosol, marking them for destruction. Fueled by the energy derived from ATP hydrolysis, a hexameric AAA-ATPase of the family of AAA-ATPases (see Figure 6-85) pulls the unfolded protein through the translocator into the cytosol. An *N*-glycanase removes its oligosaccharide chains *en bloc*. Guided by its ubiquitin tag, the deglycosylated polypeptide is rapidly fed into proteasomes, where it is degraded (Figure 12-50).

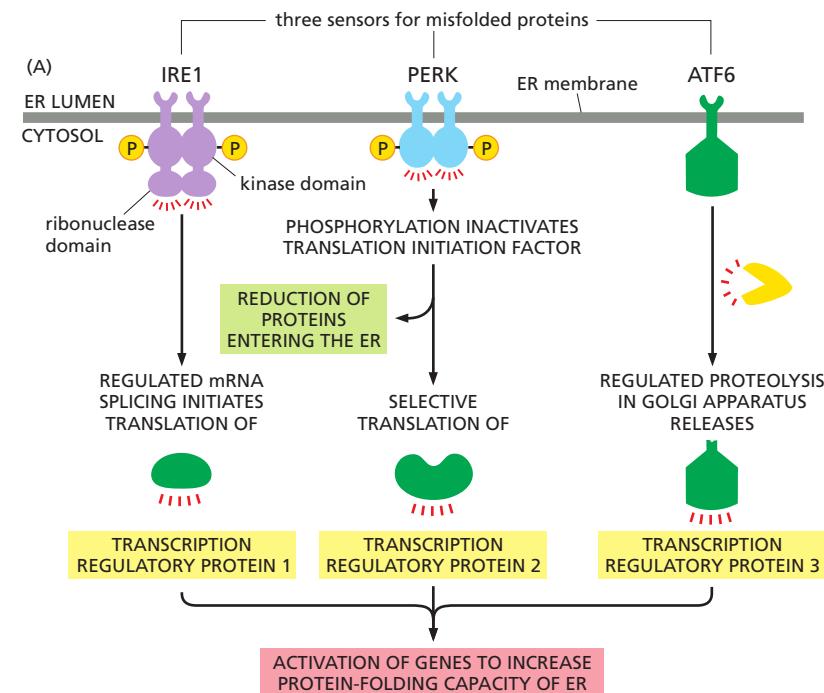
### Misfolded Proteins in the ER Activate an Unfolded Protein Response

Cells carefully monitor the amount of misfolded protein in various compartments. An accumulation of misfolded proteins in the cytosol, for example, triggers a *heat-shock response* (discussed in Chapter 6), which stimulates the transcription of genes encoding cytosolic chaperones that help to refold the proteins. Similarly, an accumulation of misfolded proteins in the ER triggers an **unfolded protein response**, which includes an increased transcription of genes encoding proteins involved in retrotranslocation and protein degradation in the cytosol, ER chaperones, and many other proteins that help to increase the protein-folding capacity of the ER.

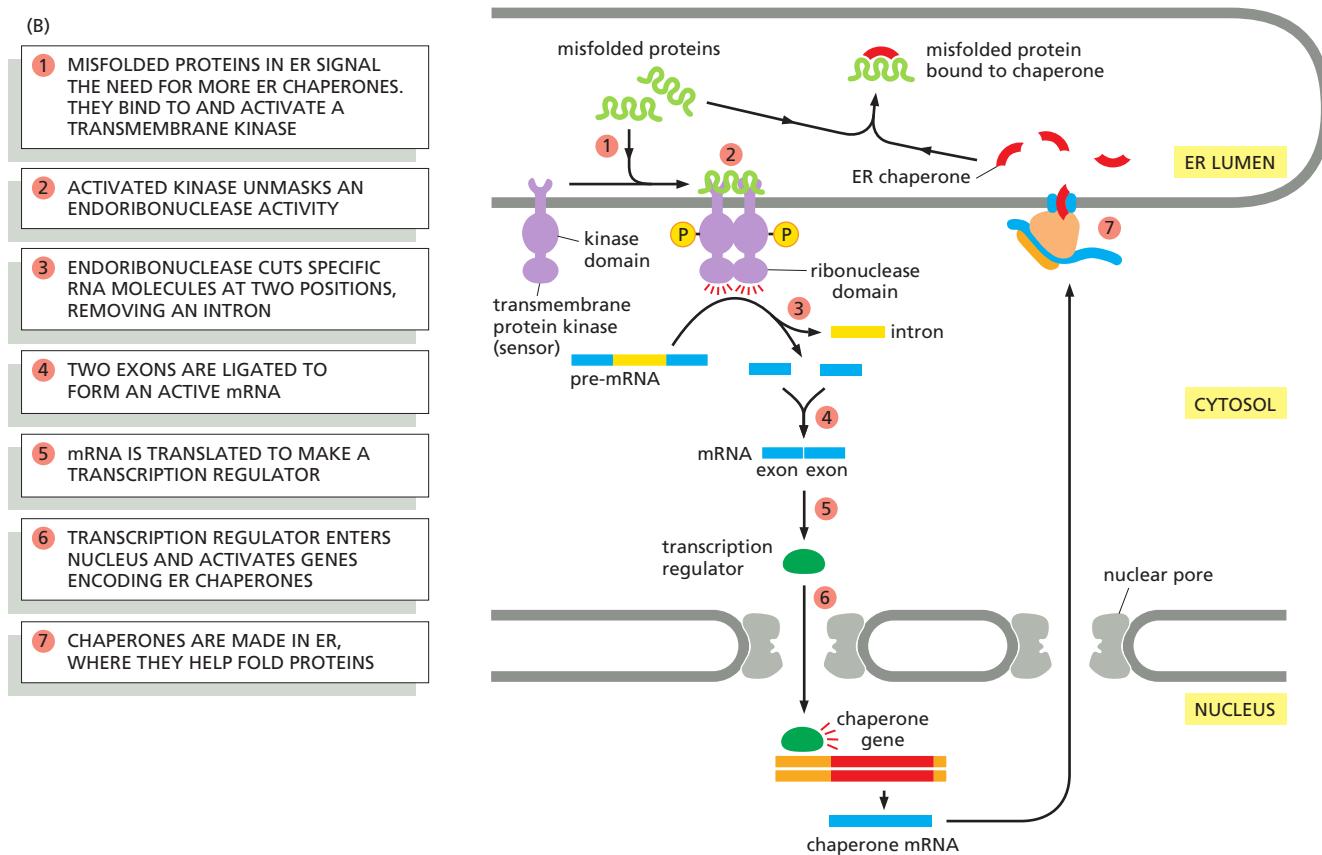


**Figure 12–50** The export and degradation of misfolded ER proteins. Misfolded soluble proteins in the ER lumen are recognized and targeted to a translocator complex in the ER membrane. They first interact in the ER lumen with chaperones, disulfide isomerases, and lectins. They are then exported into the cytosol through the translocator. In the cytosol, they are ubiquitylated, deglycosylated, and degraded in proteasomes. Misfolded membrane proteins follow a similar pathway but use a different translocator.

How do misfolded proteins in the ER signal to the nucleus? There are three parallel pathways that execute the unfolded protein response (Figure 12–51A). The first pathway, which was initially discovered in yeast cells, is particularly remarkable. Misfolded proteins in the ER activate a transmembrane protein kinase in the ER, called IRE1, which causes the kinase to oligomerize and phosphorylate itself. (Some cell-surface receptor kinases in the plasma membrane are activated in a



**Figure 12–51 The unfolded protein response.** (A) By three parallel intracellular signaling pathways, the accumulation of misfolded proteins in the ER lumen signals to the nucleus to activate the transcription of genes that encode proteins that help the cell cope with misfolded proteins in the ER. (B) Regulated RNA splicing is a key regulatory switch in pathway 1 of the unfolded protein response (Movie 12.6).



similar way, as discussed in Chapter 15.) The oligomerization and autophosphorylation of IRE1 activates an endoribonuclease domain in the cytosolic portion of the same molecule, which cleaves a specific cytosolic mRNA molecule at two positions, excising an intron. (This is a unique exception to the rule that introns are spliced out while the RNA is still in the nucleus.) The separated exons are then joined by an RNA ligase, generating a spliced mRNA, which is translated to produce an active transcription regulatory protein. This protein activates the transcription of genes encoding the proteins that help mediate the unfolded protein response (Figure 12–51B).

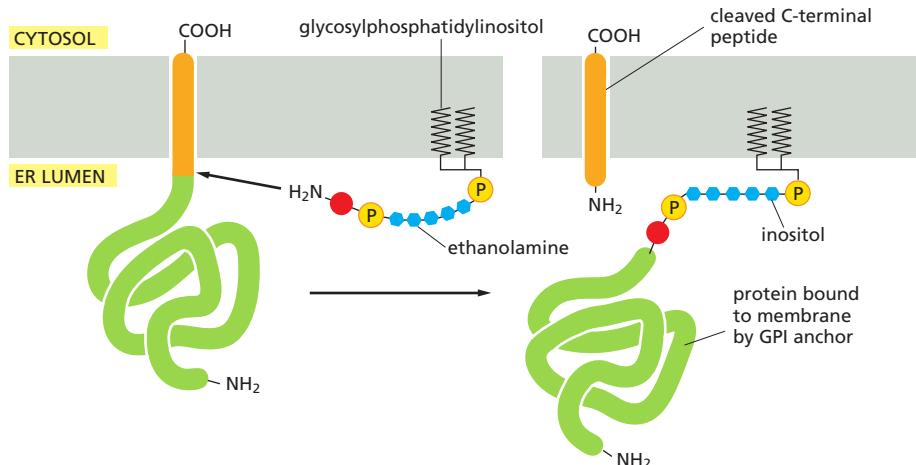
Misfolded proteins also activate a second transmembrane kinase in the ER, PERK, which inhibits a translation initiation factor by phosphorylating it, thereby reducing the production of new proteins throughout the cell. One consequence of the reduction in protein synthesis is to reduce the flux of proteins into the ER, thereby reducing the load of proteins that need to be folded there. Some proteins, however, are preferentially translated when translation initiation factors are scarce (discussed in Chapter 7, p. 424), and one of these is a transcription regulator that helps activate the transcription of the genes encoding proteins active in the unfolded protein response.

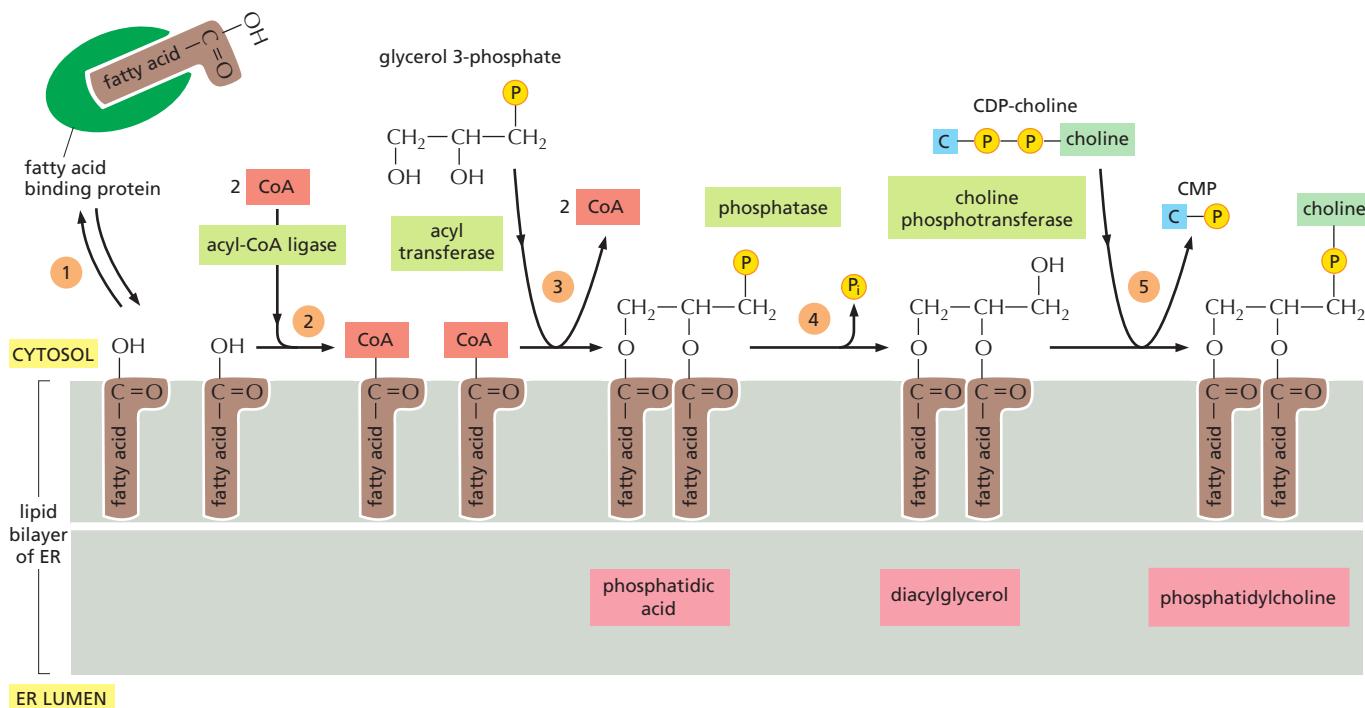
Finally, a third transcription regulator, ATF6, is initially synthesized as a transmembrane ER protein. Because it is embedded in the ER membrane, it cannot activate the transcription of genes in the nucleus. When misfolded proteins accumulate in the ER, however, the ATF6 protein is transported to the Golgi apparatus, where it encounters proteases that cleave off its cytosolic domain, which can now migrate to the nucleus and help activate the transcription of genes encoding proteins involved in the unfolded protein response. (This mechanism is similar to that described in Figure 12–16 for activation of the transcription regulator that controls cholesterol biosynthesis.) The relative importance of each of these three pathways in the unfolded protein response differs in different cell types, enabling each cell type to tailor the unfolded protein response to its particular needs.

### Some Membrane Proteins Acquire a Covalently Attached Glycosylphosphatidylinositol (GPI) Anchor

As discussed in Chapter 10, several cytosolic enzymes catalyze the covalent addition of a single fatty acid chain or prenyl group to selected proteins. The attached lipids help direct and attach these proteins to cell membranes. A related process is catalyzed by ER enzymes that covalently attach a **glycosylphosphatidylinositol (GPI) anchor** to the C-terminus of some membrane proteins destined for the plasma membrane. This linkage forms in the lumen of the ER, where, at the same time, the transmembrane segment of the protein is cleaved off (Figure 12–52). A large number of plasma membrane proteins are modified in this way. Since they are attached to the exterior of the plasma membrane only by their GPI anchors,

**Figure 12–52 The attachment of a GPI anchor to a protein in the ER.** GPI-anchored proteins are targeted to the ER membrane by an N-terminal signal sequence (not shown), which is removed (see Figure 12–42). Immediately after the completion of protein synthesis, the precursor protein remains anchored in the ER membrane by a hydrophobic C-terminal sequence of 15–20 amino acids; the rest of the protein is in the ER lumen. Within less than a minute, an enzyme in the ER cuts the protein free from its membrane-bound C-terminus and simultaneously attaches the new C-terminus to an amino group on a preassembled GPI intermediate. The sugar chain contains an inositol attached to the lipid from which the GPI anchor derives its name. It is followed by a glucosamine and three mannoses. The terminal mannose links to a phosphoethanolamine that provides the amino group to attach the protein. The signal that specifies this modification is contained within the hydrophobic C-terminal sequence and a few amino acids adjacent to it on the luminal side of the ER membrane; if this signal is added to other proteins, they too become modified in this way. Because of the covalently linked lipid anchor, the protein remains membrane-bound, with all of its amino acids exposed initially on the luminal side of the ER and eventually on the exterior of the plasma membrane.





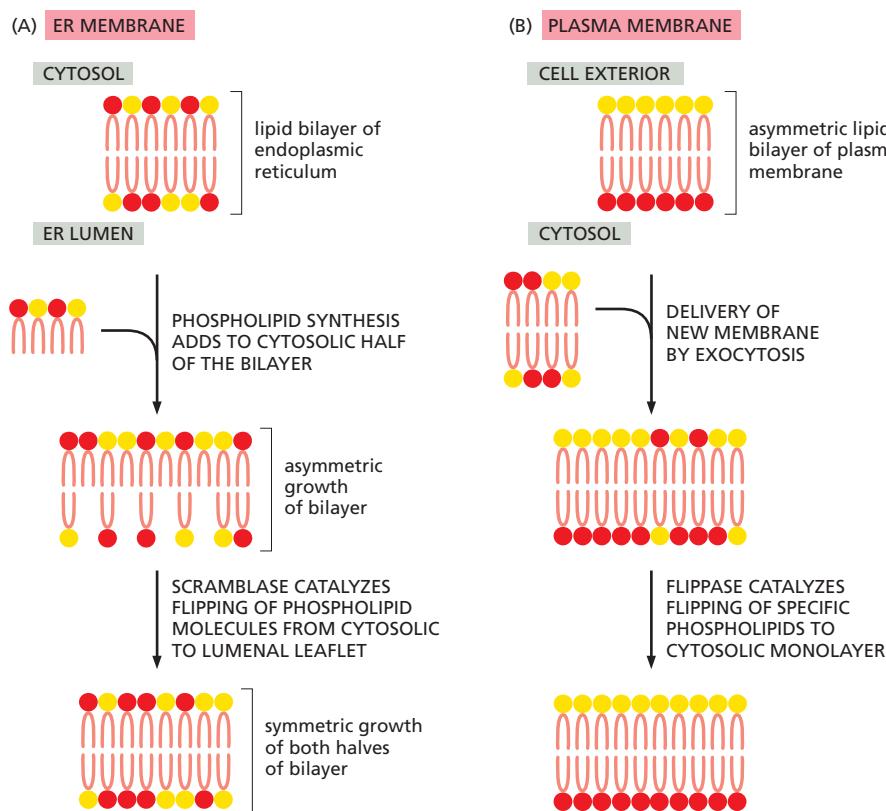
they can in principle be released from cells in soluble form in response to signals that activate a specific phospholipase in the plasma membrane. Trypanosome parasites, for example, use this mechanism to shed their coat of GPI-anchored surface proteins when attacked by the immune system. GPI anchors may also be used to direct plasma membrane proteins into *lipid rafts* and thus segregate the proteins from other membrane proteins (see Figure 10–13).

**Figure 12–53 The synthesis of phosphatidylcholine.** As illustrated, this phospholipid is synthesized from glycerol 3-phosphate, cytidine-diphosphocholine (CDP-choline), and fatty acids delivered to the ER by a cytosolic fatty acid binding protein.

### The ER Assembles Most Lipid Bilayers

The ER membrane is the site of synthesis of nearly all of the cell's major classes of lipids, including both phospholipids and cholesterol, required for the production of new cell membranes. The major phospholipid made is *phosphatidylcholine*, which can be formed in three steps from choline, two fatty acids, and glycerol phosphate (Figure 12–53). Each step is catalyzed by enzymes in the ER membrane, which have their active sites facing the cytosol, where all of the required metabolites are found. Thus, phospholipid synthesis occurs exclusively in the cytosolic leaflet of the ER membrane. Because fatty acids are not soluble in water, they are shepherded from their sites of synthesis to the ER by a fatty acid binding protein in the cytosol. After arrival in the ER membrane and activation with CoA, acyl transferases successively add two fatty acids to glycerol phosphate to produce phosphatidic acid. Phosphatidic acid is sufficiently water-insoluble to remain in the lipid bilayer; it cannot be extracted from the bilayer by the fatty acid binding protein. It is therefore this first step that enlarges the ER lipid bilayer. The later steps determine the head group of a newly formed lipid molecule and therefore the chemical nature of the bilayer, but they do not result in net membrane growth. The two other major membrane phospholipids—phosphatidylethanolamine and phosphatidylserine (see Figure 10–3)—as well as the minor phospholipid phosphatidylinositol (PI), are all synthesized in this way.

Because phospholipid synthesis takes place in the cytosolic leaflet of the ER lipid bilayer, there needs to be a mechanism that transfers some of the newly formed phospholipid molecules to the luminal leaflet of the bilayer. In synthetic lipid bilayers, lipids do not “flip-flop” in this way (see Figure 10–10). In the ER, however, phospholipids equilibrate across the membrane within minutes, which is almost 100,000 times faster than can be accounted for by spontaneous “flip-flop.” This rapid trans-bilayer movement is mediated by a poorly characterized



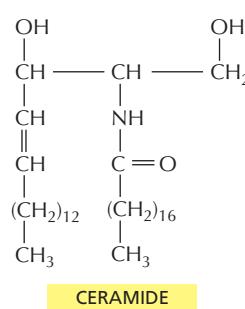
**Figure 12–54** The role of phospholipid translocators in lipid bilayer synthesis.

(A) Because new lipid molecules are added only to the cytosolic half of the ER membrane bilayer and lipid molecules do not flip spontaneously from one monolayer to the other, a transmembrane phospholipid translocator (called a scramblase) is required to transfer lipid molecules from the cytosolic half to the luminal half so that the membrane grows as a bilayer. The scramblase is not specific for particular phospholipid head groups and therefore equilibrates the different phospholipids between the two monolayers. (B) Fueled by ATP hydrolysis, a head-group-specific flipase in the plasma membrane actively flips phosphatidylserine and phosphatidylethanolamine directionally from the extracellular to the cytosolic leaflet, creating the characteristically asymmetric lipid bilayer of the plasma membrane of animal cells (see Figure 10–15).

phospholipid translocator called a *scramblase*, which nonselectively equilibrates phospholipids between the two leaflets of the lipid bilayer (Figure 12–54). Thus, the different types of phospholipids are thought to be equally distributed between the two leaflets of the ER membrane.

The plasma membrane contains a different type of phospholipid translocator that belongs to the family of P-type pumps (discussed in Chapter 11). These *flipases* specifically recognize those phospholipids that contain free amino groups in their head groups (phosphatidylserine and phosphatidylethanolamine—see Figure 10–3) and transfers them from the extracellular to the cytosolic leaflet, using the energy of ATP hydrolysis. The plasma membrane therefore has a highly asymmetric phospholipid composition, which is actively maintained by the flipases (see Figure 10–15). The plasma membrane also contains a scramblase but, in contrast to the ER scramblase, which is always active, the plasma membrane enzyme is regulated and only activated in some situations, such as in apoptosis and in activated platelets, where it acts to abolish the lipid bilayer asymmetry; the resulting exposure of phosphatidylserine on the surface of apoptotic cells serves as a signal for phagocytic cells to ingest and degrade the dead cell.

The ER also produces cholesterol and ceramide (Figure 12–55). Ceramide is made by condensing the amino acid serine with a fatty acid to form the amino alcohol sphingosine (see Figure 10–3); a second fatty acid is then covalently added to form ceramide. The ceramide is exported to the Golgi apparatus, where it serves as a precursor for the synthesis of two types of lipids: oligosaccharide chains are added to form *glycosphingolipids* (glycolipids; see Figure 10–16), and phosphocholine head groups are transferred from phosphatidylcholine to other ceramide molecules to form *sphingomyelin* (discussed in Chapter 10). Thus, both glycolipids and sphingomyelin are produced relatively late in the process of membrane synthesis. Because they are produced by enzymes that have their active sites exposed to the Golgi lumen, they are found exclusively in the noncytosolic leaflet of the lipid bilayers that contain them.



**Figure 12–55** The structure of ceramide.

As discussed in Chapter 13, the plasma membrane and the membranes of the Golgi apparatus, lysosomes, and endosomes all form part of a membrane system that communicates with the ER by means of transport vesicles, which transfer both proteins and lipids. Mitochondria and plastids, however, do not belong to this system, and they therefore require different mechanisms to import proteins and lipids for growth. We have already seen that they import most of their proteins from the cytosol. Although mitochondria modify some of the lipids they import, they do not synthesize lipids *de novo*; instead, their lipids have to be imported from the ER, either directly or indirectly by way of other cell membranes. In either case, special mechanisms are required for the transfer.

The details of how lipid distribution between different membranes is catalyzed and regulated are not known. Water-soluble carrier proteins called *phospholipid exchange proteins* (or *phospholipid transfer proteins*) are thought to transfer individual phospholipid molecules between membranes, functioning much like fatty acid binding proteins that shepherd fatty acids through the cytosol (see Figure 12-54). In addition, mitochondria are often seen in close juxtaposition to ER membranes in electron micrographs, and specific junction complexes have been identified that hold the ER and outer mitochondrial membrane in close proximity. It is thought that these junction complexes provide specific contact-dependent lipid transfer mechanisms that operate between these adjacent membranes.

## Summary

*The extensive ER network serves as a factory for the production of almost all of the cell's lipids. In addition, a major portion of the cell's protein synthesis occurs on the cytosolic surface of the rough ER: virtually all proteins destined for secretion or for the ER itself, the Golgi apparatus, the lysosomes, the endosomes, and the plasma membrane are first imported into the ER from the cytosol. In the ER lumen, the proteins fold and oligomerize, disulfide bonds are formed, and N-linked oligosaccharides are added. The pattern of N-linked glycosylation is used to indicate the extent of protein folding, so that proteins leave the ER only when they are properly folded. Proteins that do not fold or oligomerize correctly are translocated back into the cytosol, where they are deglycosylated, polyubiquitylated, and degraded in proteasomes. If misfolded proteins accumulate in excess in the ER, they trigger an unfolded protein response, which activates appropriate genes in the nucleus to help the ER cope.*

*Only proteins that carry a special ER signal sequence are imported into the ER. The signal sequence is recognized by a signal-recognition particle (SRP), which binds both the growing polypeptide chain and the ribosome and directs them to a receptor protein on the cytosolic surface of the rough ER membrane. This binding to the ER membrane initiates the translocation process that threads a loop of polypeptide chain across the ER membrane through the hydrophilic pore of a protein translocator.*

*Soluble proteins—destined for the ER lumen, for secretion, or for transfer to the lumen of other organelles—pass completely into the ER lumen. Transmembrane proteins destined for the ER or for other cell membranes are translocated part way across the ER membrane and remain anchored there by one or more membrane-spanning  $\alpha$ -helical segments in their polypeptide chains. These hydrophobic portions of the protein can act either as start-transfer or stop-transfer signals during the translocation process. When a polypeptide contains multiple, alternating start-transfer and stop-transfer signals, it will pass back and forth across the bilayer multiple times as a multipass transmembrane protein.*

*The asymmetry of protein insertion and glycosylation in the ER establishes the sidedness of the membranes of all the other organelles that the ER supplies with membrane proteins.*

## WHAT WE DON'T KNOW

- How do nuclear import receptors negotiate the tangled gel-like interior of a nuclear pore complex so efficiently?
- Is the nuclear pore complex a rigid structure or can it expand and contract, depending on the cargo transported?
- Sequence comparisons show that signal sequences for an individual protein such as insulin are quite conserved across species, much more so than would be expected from our current understanding that all that matters for their function are general structural features such as hydrophobicity. What other functions might signal sequences have that could account for their evolutionary sequence conservation?
- How are polyribosomes on the endoplasmic reticulum membrane arranged so that the next initiating ribosome will find an unoccupied translocator?
- Why does the signal-recognition particle have an indispensable RNA subunit?

## PROBLEMS

Which statements are true? Explain why or why not.

**12–1** Like the lumen of the ER, the interior of the nucleus is topologically equivalent to the outside of the cell.

**12–2** ER-bound and free ribosomes, which are structurally and functionally identical, differ only in the proteins they happen to be making at a particular time.

**12–3** To avoid the inevitable collisions that would occur if two-way traffic through a single pore were allowed, nuclear pore complexes are specialized so that some mediate import while others mediate export.

**12–4** Peroxisomes are found in only a few specialized types of eukaryotic cell.

Discuss the following problems.

**12–5** What is the fate of a protein with no sorting signal?

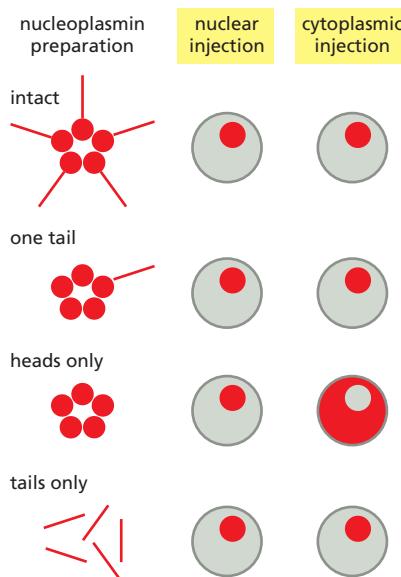
**12–6** The rough ER is the site of synthesis of many classes of membrane proteins. Some of these proteins remain in the ER, whereas others are sorted to compartments such as the Golgi apparatus, lysosomes, and the plasma membrane. One measure of the difficulty of the sorting problem is the degree of “purification” that must be achieved during transport from the ER. Are proteins bound for the plasma membrane common or rare among all ER membrane proteins?

A few simple considerations allow one to answer this question. In a typical growing cell that is dividing once every 24 hours, the equivalent of one new plasma membrane must transit the ER every day. If the ER membrane is 20 times the area of a plasma membrane, what is the ratio of plasma membrane proteins to other membrane proteins in the ER? (Assume that all proteins on their way to the plasma membrane remain in the ER for 30 minutes on average before exiting, and that the ratio of proteins to lipids in the ER and plasma membranes is the same.)

**12–7** Before nuclear pore complexes were well understood, it was unclear whether nuclear proteins diffused passively into the nucleus and accumulated there by binding to residents of the nucleus such as chromosomes, or whether they were actively imported and accumulated regardless of their affinity for nuclear components.

A classic experiment that addressed this problem used several forms of radioactive nucleoplasmin, which is a large pentameric protein involved in chromatin assembly. In this experiment, either the intact protein or the nucleoplasmin heads, tails, or heads with a single tail were injected into the cytoplasm of a frog oocyte or into the nucleus (Figure Q12–1). All forms of nucleoplasmin, except heads, accumulated in the nucleus when injected into the cytoplasm, and all forms were retained in the nucleus when injected there.

**A.** What portion of the nucleoplasmin molecule is responsible for localization in the nucleus?



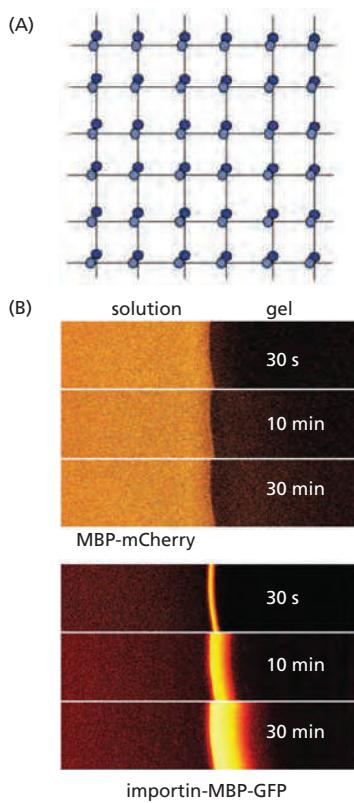
**Figure Q12–1** Cellular location of injected nucleoplasmin and nucleoplasmin components (Problem 12–7). Schematic diagrams of autoradiographs show the cytoplasm and nucleus with the location of nucleoplasmin indicated by the red areas.

**B.** How do these experiments distinguish between active transport, in which a nuclear localization signal triggers transport by the nuclear pore complex, and passive diffusion, in which a binding site for a nuclear component allows accumulation in the nucleus?

**12–8** Assuming that 32 million histone octamers are required to package the human genome, how many histone molecules must be transported per second per nuclear pore complex in cells whose nuclei contain 3000 nuclear pores and are dividing once per day?

**12–9** The nuclear pore complex (NPC) creates a barrier to the free exchange of molecules between the nucleus and cytosol, but in a way that remains mysterious. In yeast, for example, the central pore of the NPC has a diameter of 35 nm and is 30 nm long, which is somewhat smaller than its vertebrate counterpart. Even so, it is large enough to accommodate virtually all components of the cytosol. Yet the pore allows passive diffusion of molecules only up to about 40 kd; entry of anything larger requires help from a nuclear import receptor. Selective permeability is controlled by protein components of the NPC that have unstructured, polar tails extending into the central pore. These tails are characterized by periodic repeats of the hydrophobic amino acids phenylalanine (F) and glycine (G).

At high enough concentration (~50 mM), the FG-repeat domains of these proteins can form a gel, with a meshwork of interactions between the hydrophobic FG repeats (Figure Q12–2A). These gels allow passive diffusion of small molecules, but they prevent entry of larger proteins such as the fluorescent protein mCherry fused to maltose binding protein (MBP) (Figure Q12–2B). (The fusion to MBP makes mCherry too large to enter the nucleus by passive diffusion.) However, if the nuclear import receptor, importin, is fused to a similar protein, MBP-GFP, the importin-MBP-GFP fusion readily enters the gel (Figure Q12–2B).



**Figure Q12-2** FG-repeat gel and influx of proteins into the nucleus (Problem 12–9). (A) Cartoon of the meshwork (gel) formed by pairwise interactions between hydrophobic FG repeats. For FG-repeats separated by 17 amino acids, as is typical, the mesh formed by extended amino acid side chains would correspond to about 4 nm on a side, which would be large enough to account for the characteristic passive diffusion of proteins through nuclear pores. (B) Diffusion of MBP-mCherry and importin-MBP-GFP into a gel of FG-repeats. In each group, the solution is shown at left and the gel at right. The bright areas indicate regions that contain the fluorescent proteins.

**A.** FG-repeats only form gels *in vitro* at relatively high concentration (50 mM). Is this concentration reasonable for FG repeats in the NPC core? In yeast, there are about 5000 FG-repeats in each NPC. Given the dimensions of the yeast nuclear pore (35 nm diameter and 30 nm length), calculate the concentration of FG-repeats in the cylindrical volume of the pore. Is this concentration comparable to the one used *in vitro*?

**B.** A second question is whether the diffusion of importin-MBP-GFP through the FG-repeat gel is fast enough to account for the efficient flow of materials between the nucleus and cytosol. From experiments of the type shown in Figure Q12-2B, the diffusion coefficient ( $D$ ) of importin-MBP-GFP through the FG-repeat gel was determined to be about  $0.1 \mu\text{m}^2/\text{s}$ . The equation for diffusion is  $t = x^2/2D$ , where  $t$  is time and  $x$  is distance. Calculate the time it would take importin-MBP-GFP to diffuse through a yeast nuclear pore (30 nm) if the pore consisted of a gel of FG-repeats. Does this time seem fast enough for the needs of a eukaryotic cell?

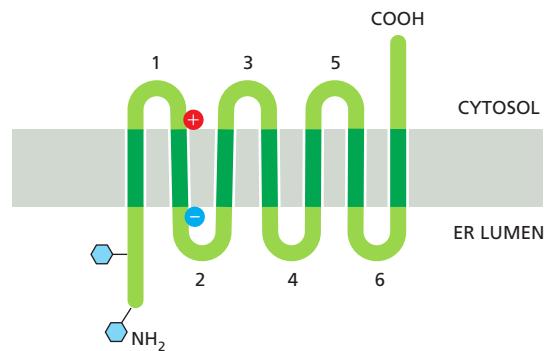
**12-10** Components of the TIM complexes, the multi-subunit protein translocators in the mitochondrial inner membrane, are much less abundant than those of the TOM

complex. They were initially identified using a genetic trick. The yeast *Ura3* gene, whose product is an enzyme that is normally located in the cytosol where it is essential for synthesis of uracil, was modified so that the protein carried an import signal for the mitochondrial matrix. A population of cells carrying the modified *Ura3* gene in place of the normal gene was then grown in the absence of uracil. Most cells died, but the rare cells that grew were shown to be defective for mitochondrial import. Explain how this selection identifies cells with defects in components required for import into the mitochondrial matrix. Why don't normal cells with the modified *Ura3* gene grow in the absence of uracil? Why do cells that are defective for mitochondrial import grow in the absence of uracil?

**12-11** If the enzyme dihydrofolate reductase (DHFR), which is normally located in the cytosol, is engineered to carry a mitochondrial targeting sequence at its N-terminus, it is efficiently imported into mitochondria. If the modified DHFR is first incubated with methotrexate, which binds tightly to the active site, the enzyme remains in the cytosol. How do you suppose that the binding of methotrexate interferes with mitochondrial import?

**12-12** Why do mitochondria need a special translocator to import proteins across the outer membrane, when the membrane already has large pores formed by porins?

**12-13** Examine the multipass transmembrane protein shown in Figure Q12-3. What would you predict would be the effect of converting the first hydrophobic transmembrane segment to a hydrophilic segment? Sketch the arrangement of the modified protein in the ER membrane.



**Figure Q12-3** Arrangement of a multipass transmembrane protein in the ER membrane (Problem 12–13). Blue hexagons represent covalently attached oligosaccharides. The positions of positively and negatively charged amino acids flanking the second transmembrane segment are shown.

**12-14** All new phospholipids are added to the cytosolic leaflet of the ER membrane, yet the ER membrane has a symmetrical distribution of different phospholipids in its two leaflets. By contrast, the plasma membrane, which receives all its membrane components ultimately from the ER, has a very asymmetrical distribution of phospholipids in the two leaflets of its lipid bilayer. How is the symmetry generated in the ER membrane, and how is the asymmetry generated and maintained in the plasma membrane?

## REFERENCES

### General

Palade G (1975) Intracellular aspects of the process of protein synthesis. *Science* 189, 347–358.

### The Compartmentalization of Cells

B Blobel G (1980) Intracellular protein topogenesis. *Proc. Natl Acad. Sci. USA* 77, 1496–1500.  
DeVos DP, Gráf R & Field MC (2014) Evolution of the nucleus. *Curr. Opin. Cell Biol.* 28, 8–15.  
Warren G & Wickner W (1996) Organelle inheritance. *Cell* 84, 395–400.

### The Transport of Molecules Between the Nucleus and the Cytosol

Adam SA & Gerace L (1991) Cytosolic proteins that specifically bind nuclear location signals are receptors for nuclear import. *Cell* 66, 837–847.  
Burke B & Stewart CL (2013) The nuclear lamins: flexibility in function. *Nat. Rev. Mol. Cell Biol.* 14, 13–24.  
Cole CN & Scarcelli JJ (2006) Transport of messenger RNA from the nucleus to the cytoplasm. *Curr. Opin. Cell Biol.* 18, 299–306.  
Güttlinger S, Laurell E & Kutay U (2009) Orchestrating nuclear envelope disassembly and reassembly during mitosis. *Nat. Rev. Mol. Cell Biol.* 10, 178–191.  
Hetz MW & Wente SR (2009) Border control at the nucleus: biogenesis and organization of the nuclear membrane and pore complexes. *Dev. Cell* 17, 606–616.  
Hoelz A, Debler EW & Blobel G (2011) The structure of the nuclear pore complex. *Annu. Rev. Biochem.* 80, 613–643.  
Hülsmann BB, Labokha AA & Görlich D (2012) The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* 150, 738–751.  
Köhler A & Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. *Nat. Rev. Mol. Cell Biol.* 8, 761–773.  
Rothbäumer A & Kutay U (2013) Poring over pores: nuclear pore complex insertion into the nuclear envelope. *Trends Biochem. Sci.* 38, 292–301.  
Strambio-De-Castilla C, Niepel M & Rout MP (2010) The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat. Rev. Mol. Cell Biol.* 11, 490–501.  
Tran EJ & Wente SR (2006) Dynamic nuclear pore complexes: life on the edge. *Cell* 125, 1041–1053.

### The Transport of Proteins Into Mitochondria and Chloroplasts

Chacinska A, Koehler CM, Milenkovic D et al. (2009) Importing mitochondrial proteins: machineries and mechanisms. *Cell* 138, 628–644.  
Jarvis P & Robinson C (2004) Mechanisms of protein import and routing in chloroplasts. *Curr. Biol.* 14, R1064–R1077.  
Kessler F & Schnell DJ (2009) Chloroplast biogenesis: diversity and regulation of the protein import apparatus. *Curr. Opin. Cell Biol.* 21, 494–500.  
Prakash S & Matouschek A (2004) Protein unfolding in the cell. *Trends Biochem. Sci.* 29, 593–600.  
Schleiff E & Becker T (2011) Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat. Rev. Mol. Cell Biol.* 12, 48–59.

### Peroxisomes

Dimitrov L, Lam SK & Schekman R (2013) The role of the endoplasmic reticulum in peroxisome biogenesis. *Cold Spring Harb. Perspect. Biol.* 5, a013243.  
Fujiki Y, Yagita Y & Matsuzaki T (2012) Peroxisome biogenesis disorders. *Biochim. Biophys. Acta* 1822, 1337–1342.  
Schliebs W, Girzalsky W & Erdmann R (2010) Peroxisomal protein import and ERAD: variations on a common theme. *Nat. Rev. Mol. Cell Biol.* 11, 885–890.

Tabak HF, Braakman I & van der Zand A (2013) Peroxisome formation and maintenance are dependent on the endoplasmic reticulum. *Annu. Rev. Biochem.* 82, 723–744.

### The Endoplasmic Reticulum

Akopian D, Shen K, Zhang X & Shan SO (2013) Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.* 82, 693–721.  
Blobel G & Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* 67, 835–851.  
Borgese N, Mok W, Kreibich G & Sabatini DD (1974) Ribosomal-membrane interaction: *in vitro* binding of ribosomes to microsomal membranes. *J. Mol. Biol.* 88, 559–580.  
Braakman I & Bulleid NJ (2011) Protein folding and modification in the mammalian endoplasmic reticulum. *Annu. Rev. Biochem.* 80, 71–99.  
Brodsky JL & Skach WR (2011) Protein folding and quality control in the endoplasmic reticulum: recent lessons from yeast and mammalian cell systems. *Curr. Opin. Cell Biol.* 23, 464–475.  
Chen S, Novick P & Ferro-Novick S (2013) ER structure and function. *Curr. Opin. Cell Biol.* 25, 428–433.  
Clark MR (2011) Flippin' lipids. *Nat. Immunol.* 12, 373–375.  
Daleke DL (2003) Regulation of transbilayer plasma membrane phospholipid asymmetry. *J. Lipid Res.* 44, 233–242.  
Deshaies RJ, Sanders SL, Feldheim DA & Schekman R (1991) Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* 349, 806–808.  
Getting MJ (1999) Role and regulation of the ER chaperone BiP. *Semin. Cell Dev. Biol.* 10, 465–472.  
Görlich D, Prehn S, Hartmann E et al. (1992) A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* 71, 489–503.  
Hegde RS & Ploegh HL (2010) Quality and quantity control at the endoplasmic reticulum. *Curr. Opin. Cell Biol.* 22, 437–446.  
Hegde RS & Keenan RJ (2011) Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 12, 787–798.  
Levine T & Loewen C (2006) Inter-organelle membrane contact sites: through a glass, darkly. *Curr. Opin. Cell Biol.* 18, 371–378.  
López-Marqués RL, Holthuis JCM & Pomorski TG (2011) Pumping lipids with P4-ATPases. *Biol. Chem.* 392, 67–76.  
Mamathambika BS & Bardwell JC (2008) Disulfide-linked protein folding pathways. *Annu. Rev. Cell Dev. Biol.* 24, 211–235.  
Marciniak SJ & Ron D (2006) Endoplasmic reticulum stress signaling in disease. *Physiol. Rev.* 86, 1133–1149.  
Milstein C, Brownlee GG, Harrison TM & Mathews MB (1972) A possible precursor of immunoglobulin light chains. *Nat. New Biol.* 239, 117–120.  
Park E & Rapoport TA (2012) Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annu. Rev. Biophys.* 41, 21–40.  
Römisch K (2005) Endoplasmic reticulum-associated degradation. *Annu. Rev. Cell Dev. Biol.* 21, 435–456.  
Rowland AA & Voeltz GK (2012) Endoplasmic reticulum-mitochondria contacts: function of the junction. *Nat. Rev. Mol. Cell Biol.* 13, 607–625.  
Trombetta ES & Parodi AJ (2003) Quality control and protein folding in the secretory pathway. *Annu. Rev. Cell Dev. Biol.* 19, 649–676.  
Tsai B, Ye Y & Rapoport TA (2002) Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nat. Rev. Mol. Cell Biol.* 3, 246–255.  
Walter P & Ron D (2011) The unfolded protein response: from stress pathway to homeostatic regulation. *Science* 334, 1081–1086.  
von Heijne G (2011) Introduction to theme “membrane protein folding and insertion”. *Annu. Rev. Biochem.* 80, 157–160.

## CHAPTER

## 13

## Intracellular Membrane Traffic

Every cell must eat, communicate with the world around it, and quickly respond to changes in its environment. To help accomplish these tasks, cells continually adjust the composition of their plasma membrane and internal compartments in rapid response to need. They use an elaborate internal membrane system to add and remove cell-surface proteins, such as receptors, ion channels, and transporters (Figure 13–1). Through the process of *exocytosis*, the secretory pathway delivers newly synthesized proteins, carbohydrates, and lipids either to the plasma membrane or the extracellular space. By the converse process of *endocytosis*, cells remove plasma membrane components and deliver them to internal compartments called *endosomes*, from where they can be recycled to the same or different regions of the plasma membrane or be delivered to lysosomes for degradation. Cells also use endocytosis to capture important nutrients, such as vitamins, cholesterol, and iron; these are taken up together with the macromolecules to which they bind and are then moved on to endosomes and lysosomes, from where they can be transported into the cytosol for use in various biosynthetic processes.

The interior space, or *lumen*, of each membrane-enclosed compartment along the secretory and endocytic pathways is equivalent to the lumen of most other membrane-enclosed compartments and to the cell exterior, in the sense that proteins can travel in this space without having to cross a membrane as they are passed from one compartment to another by means of numerous membrane-enclosed transport containers. These containers are formed from the donor compartment and are either small, spherical *vesicles*, larger irregular vesicles, or tubules. We shall use the term **transport vesicle** to apply to all forms of these containers.

Within a eukaryotic cell, transport vesicles continually bud off from one membrane and fuse with another, carrying membrane components and soluble luminal molecules, which are referred to as **cargo** (Figure 13–2). This vesicular traffic flows along highly organized, directional routes, which allow the cell to secrete, eat, and remodel its plasma membrane and organelles. The *secretory pathway* leads outward from the endoplasmic reticulum (ER) toward the Golgi apparatus and cell surface, with a side route leading to lysosomes, while the *endocytic pathway* leads inward from the plasma membrane. In each case, *retrieval pathways*

## IN THIS CHAPTER

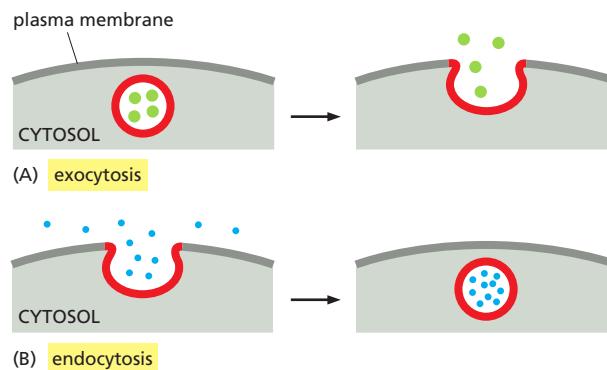
THE MOLECULAR MECHANISMS OF MEMBRANE TRANSPORT AND THE MAINTENANCE OF COMPARTMENTAL DIVERSITY

TRANSPORT FROM THE ER THROUGH THE GOLGI APPARATUS

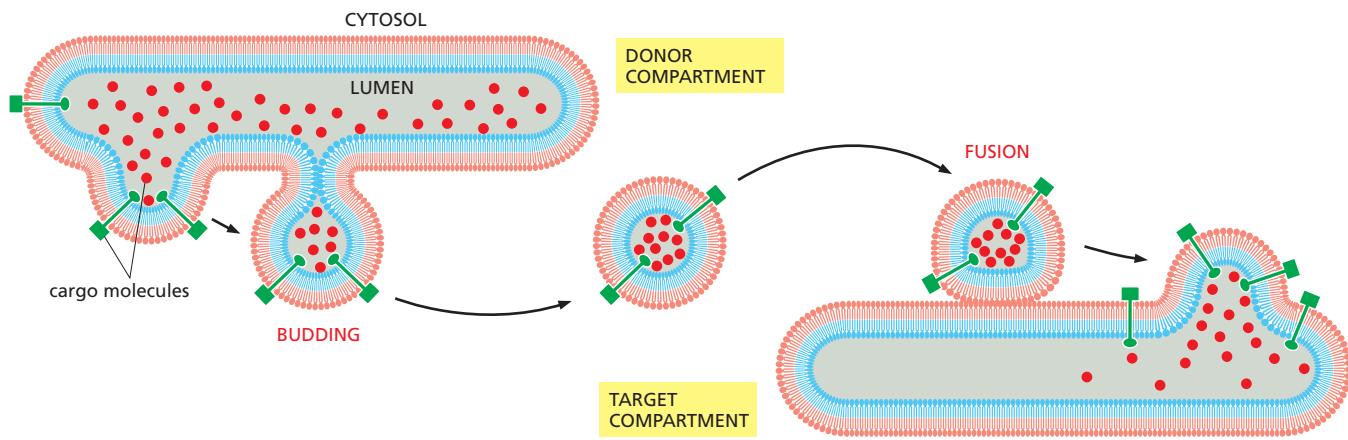
TRANSPORT FROM THE TRANS GOLGI NETWORK TO LYSOSOMES

TRANSPORT INTO THE CELL FROM THE PLASMA MEMBRANE: ENDOCYTOSIS

TRANSPORT FROM THE TRANS GOLGI NETWORK TO THE CELL EXTERIOR: EXOCYTOSIS



**Figure 13–1 Exocytosis and endocytosis.** (A) In exocytosis, a transport vesicle fuses with the plasma membrane. Its content is released into the extracellular space, while the vesicle membrane (red) becomes continuous with the plasma membrane. (B) In endocytosis, a plasma membrane patch (red) is internalized, forming a transport vesicle. Its content derives from the extracellular space.

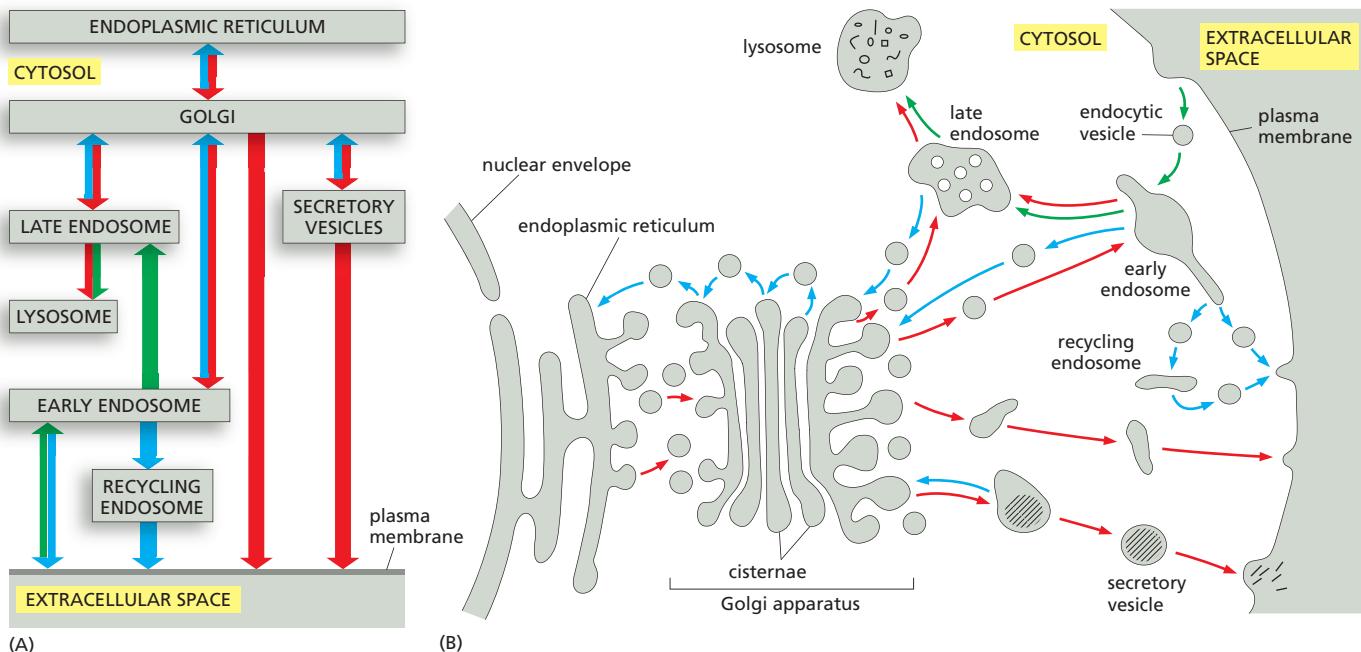


balance the flow of membrane between compartments in the opposite direction, bringing membrane and selected proteins back to the compartment of origin (**Figure 13-3**).

To perform its function, each transport vesicle that buds from a compartment must be selective. It must take up only the appropriate molecules and must fuse only with the appropriate target membrane. A vesicle carrying cargo from the ER to the Golgi apparatus, for example, must exclude most proteins that are to stay in the ER, and it must fuse only with the Golgi apparatus and not with any other organelle.

We begin this chapter by considering the molecular mechanisms of budding and fusion that underlie all vesicle transport. We then discuss the fundamental problem of how, in the face of this transport, the cell maintains the molecular and

**Figure 13-2 Vesicle transport.** Transport vesicles bud off from one compartment and fuse with another. As they do so, they carry material as cargo from the lumen (the space within a membrane-enclosed compartment) and membrane of the donor compartment to the lumen and membrane of the target compartment, as shown.



**Figure 13-3 A “road-map” of the secretory and endocytic pathways.** (A) In this schematic roadmap, which was introduced in Chapter 12, the endocytic and secretory pathways are illustrated with green and red arrows, respectively. In addition, blue arrows denote retrieval pathways for the backflow of selected components. (B) The compartments of the eukaryotic cell involved in vesicle transport. The lumen of each membrane-enclosed compartment is topologically equivalent to the outside of the cell. All compartments shown communicate with one another and the outside of the cell by means of transport vesicles. In the secretory pathway (red arrows), protein molecules are transported from the ER to the plasma membrane or (via endosomes) to lysosomes. In the endocytic pathway (green arrows), molecules are ingested in endocytic vesicles derived from the plasma membrane and delivered to early endosomes and then (via late endosomes) to lysosomes. Many endocytosed molecules are retrieved from early endosomes and returned (some via recycling endosomes) to the cell surface for reuse; similarly, some molecules are retrieved from the early and late endosomes and returned to the Golgi apparatus, and some are retrieved from the Golgi apparatus and returned to the ER. All of these retrieval pathways are shown with blue arrows, as in part (A).

functional differences between its compartments. Finally, we consider the function of the Golgi apparatus, lysosomes, secretory vesicles, and endosomes, as we trace the pathways that connect these organelles.

## THE MOLECULAR MECHANISMS OF MEMBRANE TRANSPORT AND THE MAINTENANCE OF COMPARTMENTAL DIVERSITY

Vesicle transport mediates a continuous exchange of components between the ten or more chemically distinct, membrane-enclosed compartments that collectively comprise the secretory and endocytic pathways. With this massive exchange, how can each compartment maintain its special identity? To answer this question, we must first consider what defines the character of a compartment. Above all, it is the composition of the enclosing membrane: molecular markers displayed on the cytosolic surface of the membrane serve as guidance cues for incoming traffic to ensure that transport vesicles fuse only with the correct compartment. Many of these membrane markers, however, are found on more than one compartment, and it is the specific combination of marker molecules that gives each compartment its molecular address.

How are these membrane markers kept at high concentration on one compartment and at low concentration on another? To answer this question, we need to consider how patches of membrane, enriched or depleted in specific membrane components, bud off from one compartment and transfer to another.

We begin by discussing how cells segregate proteins into separate membrane domains by assembling a special protein coat on the membrane's cytosolic face. We consider how coats form, what they are made of, and how they are used to extract specific cargo components from a membrane and compartment lumen for delivery to another compartment. Finally, we discuss how transport vesicles dock at the appropriate target membrane and then fuse with it to deliver their cargo.

### There Are Various Types of Coated Vesicles

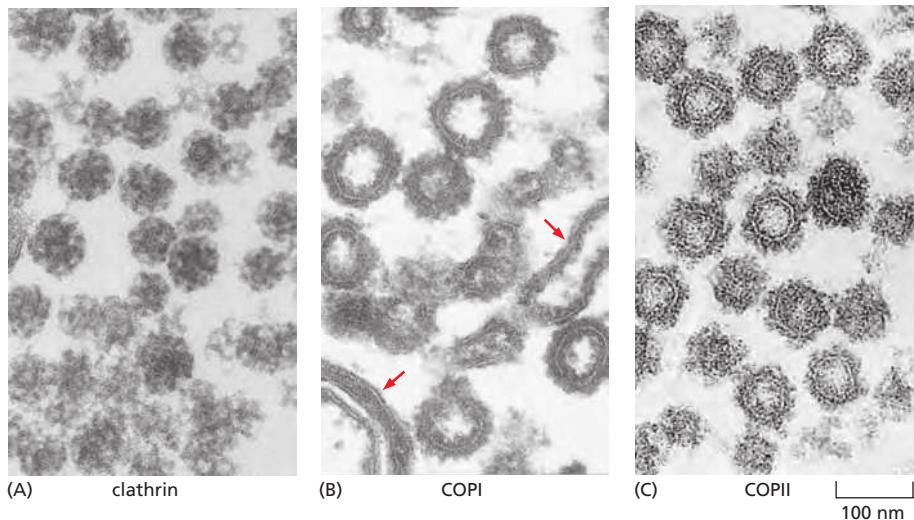
Most transport vesicles form from specialized, coated regions of membranes. They bud off as **coated vesicles**, which have a distinctive cage of proteins covering their cytosolic surface. Before the vesicles fuse with a target membrane, they discard their coat, as is required for the two cytosolic membrane surfaces to interact directly and fuse.

The coat performs two main functions that are reflected in a common two-layered structure. First, an inner coat layer concentrates specific membrane proteins in a specialized patch, which then gives rise to the vesicle membrane. In this way, the inner layer selects the appropriate membrane molecules for transport. Second, an outer coat layer assembles into a curved, basketlike lattice that deforms the membrane patch and thereby shapes the vesicle.

There are three well-characterized types of coated vesicles, distinguished by their major coat proteins: *clathrin-coated*, *COP-I-coated*, and *COP-II-coated* (Figure 13–4). Each type is used for different transport steps. Clathrin-coated vesicles, for example, mediate transport from the Golgi apparatus and from the plasma membrane, whereas COPI- and COPII-coated vesicles most commonly mediate transport from the ER and from the Golgi cisternae (Figure 13–5). There is, however, much more variety in coated vesicles and their functions than this short list suggests. As we discuss below, there are several types of clathrin-coated vesicles, each specialized for a different transport step, and the COPI- and COPII-coated vesicles may be similarly diverse.

### The Assembly of a Clathrin Coat Drives Vesicle Formation

**Clathrin-coated vesicles**, the first coated vesicles to be identified, transport material from the plasma membrane and between endosomal and Golgi compartments. **COPI-coated vesicles** and **COPII-coated vesicles** transport material early



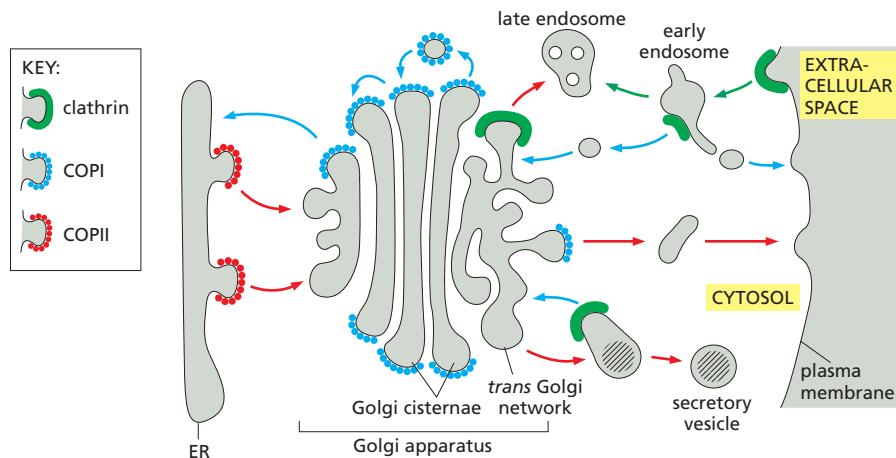
**Figure 13-4** Electron micrographs of clathrin-coated, COPI-coated, and COPII-coated vesicles. All are shown in electron micrographs at the same scale. (A) Clathrin-coated vesicles. (B) COPI-coated vesicles and Golgi cisternae (red arrows) from a cell-free system in which COPI-coated vesicles bud in the test tube. (C) COPII-coated vesicles. (A and B, from L. Orci, B. Glick and J. Rothman, *Cell* 46:171–184, 1986. With permission from Elsevier; C, courtesy of Charles Barlowe and Lelio Orci.)

in the secretory pathway: COPI-coated vesicles bud from Golgi compartments, and COPII-coated vesicles bud from the ER (see Figure 13-5). We discuss clathrin-coated vesicles first, as they provide a good example of how vesicles form.

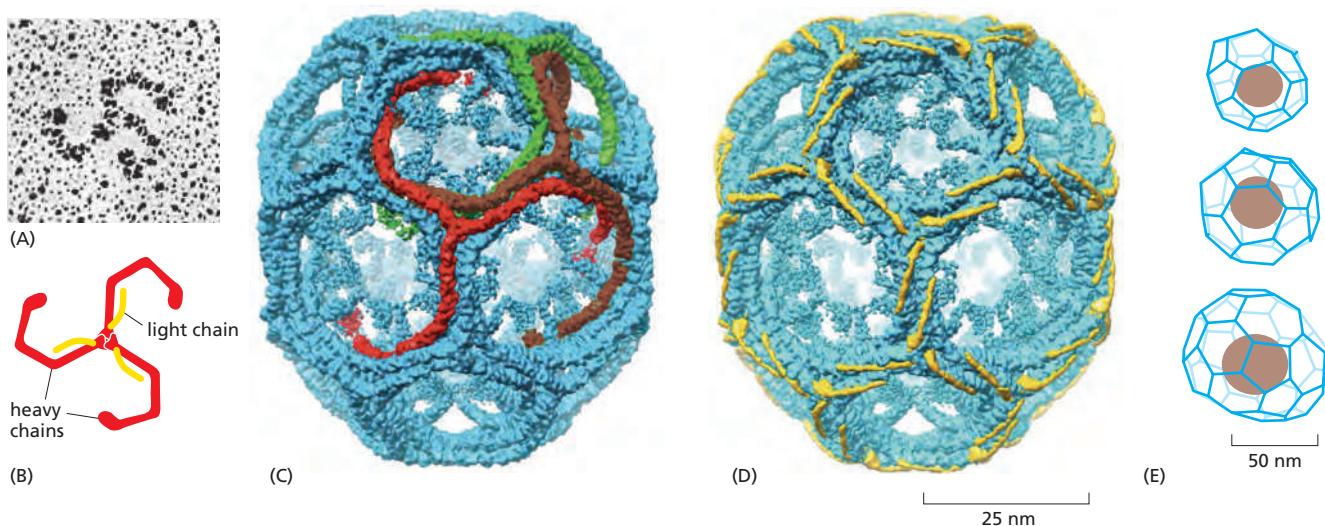
The major protein component of clathrin-coated vesicles is **clathrin** itself, which forms the outer layer of the coat. Each clathrin subunit consists of three large and three small polypeptide chains that together form a three-legged structure called a *triskelion* (Figure 13-6A,B). Clathrin triskelions assemble into a basketlike framework of hexagons and pentagons to form coated pits (buds) on the cytosolic surface of membranes (Figure 13-7). Under appropriate conditions, isolated triskelions spontaneously self-assemble into typical polyhedral cages in a test tube, even in the absence of the membrane vesicles that these baskets normally enclose (Figure 13-6C,D). Thus, the clathrin triskelions determine the geometry of the clathrin cage (Figure 13-6E).

### Adaptor Proteins Select Cargo into Clathrin-Coated Vesicles

**Adaptor proteins**, another major coat component in clathrin-coated vesicles, form a discrete inner layer of the coat, positioned between the clathrin cage and the membrane. They bind the clathrin coat to the membrane and trap various transmembrane proteins, including transmembrane receptors that capture soluble cargo molecules inside the vesicle—so-called *cargo receptors*. In this way, the adaptor proteins select a specific set of transmembrane proteins, together with the soluble proteins that interact with them, and package them into each newly formed clathrin-coated transport vesicle (Figure 13-8).



**Figure 13-5** Use of different coats for different steps in vesicle traffic. Different coat proteins select different cargo and shape the transport vesicles that mediate the various steps in the secretory and endocytic pathways. When the same coats function in different places in the cell, they usually incorporate different coat protein subunits that modify their properties (not shown). Many differentiated cells have additional pathways beside those shown here, including a sorting pathway from the *trans* Golgi network to the apical surface of epithelial cells and a specialized recycling pathway for proteins of synaptic vesicles in the nerve terminals of neurons (see Figure 11–36). The arrows are colored as in Figure 13–3.



**Figure 13–6** The structure of a clathrin coat. (A) Electron micrograph of a clathrin triskelion shadowed with platinum. (B) Each triskelion is composed of three clathrin heavy chains and three clathrin light chains, as shown in the diagram. (C and D) A cryoelectron micrograph taken of a clathrin coat composed of 36 triskelions organized in a network of 12 pentagons and 6 hexagons, with some heavy chains (C) and light chains (D) highlighted ([Movie 13.1](#)). The light chains link to the actin cytoskeleton, which helps generate force for membrane budding and vesicle movement, and their phosphorylation regulates clathrin coat assembly. The interwoven legs of the clathrin triskelions form an outer shell from which the N-terminal domains of the triskelions protrude inward. These domains bind to the adaptor proteins shown in Figure 13–8. The coat shown was assembled biochemically from pure clathrin triskelions and is too small to enclose a membrane vesicle. (E) Images of clathrin-coated vesicles isolated from bovine brain. The clathrin coats are constructed in a similar but less regular way, from pentagons, a larger number of hexagons, and sometime heptagons, resembling the architecture of deformed soccer balls. The structures were determined by cryoelectron microscopy and tomographic reconstruction. (A, from E. Ungewickell and D. Branton, *Nature* 289:420–422, 1981; C and D, from A. Fotin et al., *Nature* 432:573–579, 2004. All with permission from Macmillan Publishers Ltd; E, from Y. Cheng et al., *J. Mol. Biol.* 365:892–899, 2007. With permission from Elsevier.)

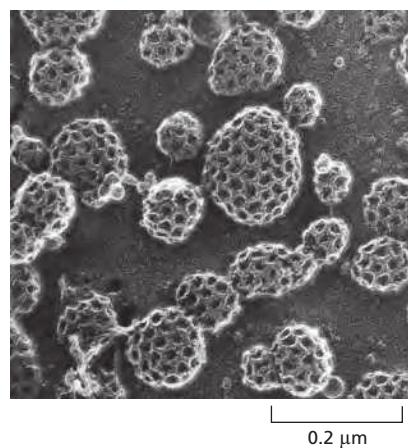
There are several types of adaptor proteins. The best characterized have four different protein subunits; others are single-chain proteins. Each type of adaptor protein is specific for a different set of cargo receptors. Clathrin-coated vesicles budding from different membranes use different adaptor proteins and thus package different receptors and cargo molecules.

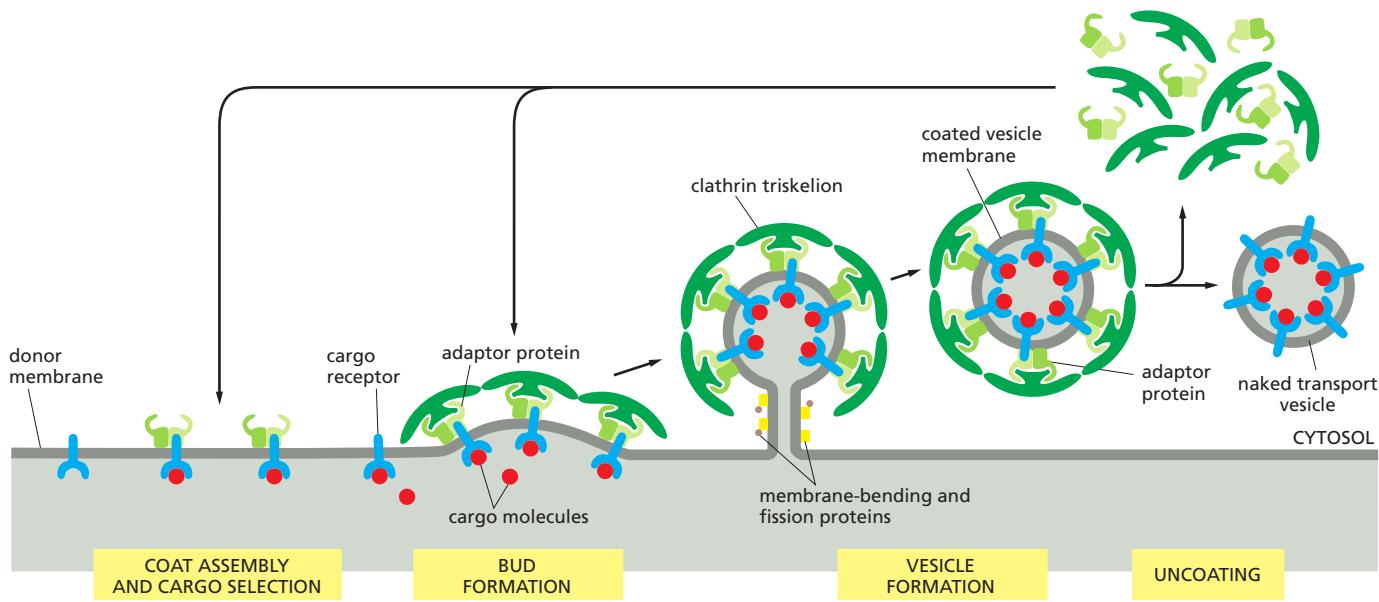
The assembly of adaptor proteins on the membrane is tightly controlled, in part by the cooperative interaction of the adaptor proteins with other components of the coat. The adaptor protein AP2 serves as a well-understood example. When it binds to a specific phosphorylated phosphatidylinositol lipid (a *phosphoinositide*), it alters its conformation, exposing binding sites for cargo receptors in the membrane. The simultaneous binding to the cargo receptors and lipid head groups greatly enhances the binding of AP2 to the membrane ([Figure 13–9](#)).

Because several requirements must be met simultaneously to stably bind AP2 proteins to a membrane, the proteins act as *coincidence detectors* that only assemble at the right time and place. Upon binding, they induce membrane curvature, which makes the binding of additional AP2 proteins in its proximity more likely. The cooperative assembly of the AP2 coat layer then is further amplified by clathrin binding, which leads to the formation and budding of a transport vesicle.

Adaptor proteins found in other coats also bind to phosphoinositides, which not only have a major role in directing when and where coats assemble in the cell, but also are used much more widely as molecular markers of compartment identity. This helps to control vesicular traffic, as we now discuss.

**Figure 13–7** Clathrin-coated pits and vesicles. This rapid-freeze, deep-etch electron micrograph shows numerous clathrin-coated pits and vesicles on the inner surface of the plasma membrane of cultured fibroblasts. The cells were rapidly frozen in liquid helium, fractured, and deep-etched to expose the cytoplasmic surface of the plasma membrane. (Courtesy of John Heuser.)



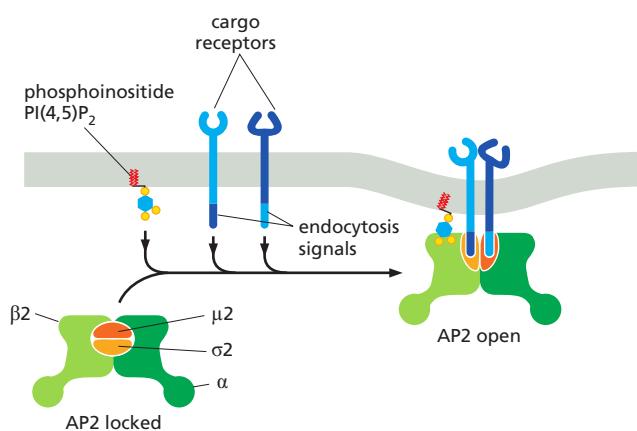


### Phosphoinositides Mark Organelles and Membrane Domains

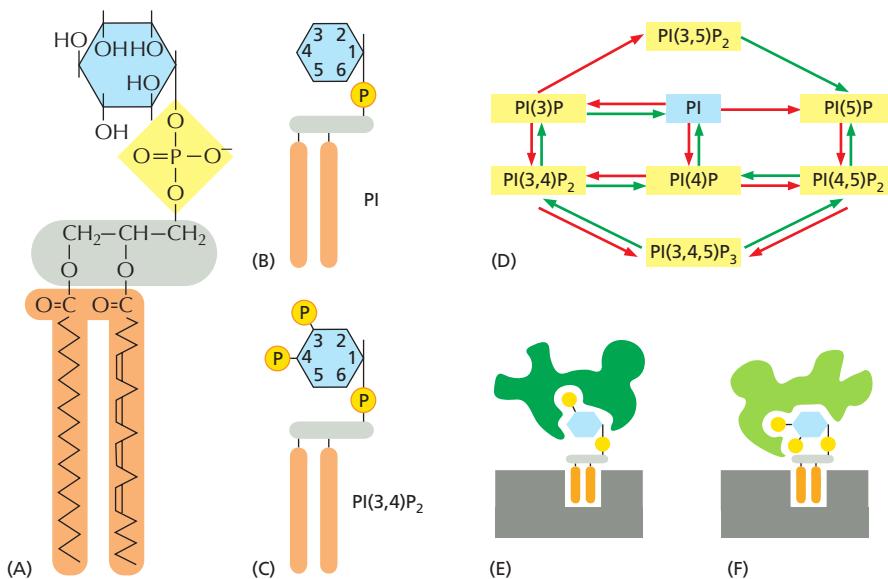
Although inositol phospholipids typically comprise less than 10% of the total phospholipids in a membrane, they have important regulatory functions. They can undergo rapid cycles of phosphorylation and dephosphorylation at the 3', 4', and 5' positions of their inositol sugar head groups to produce various types of **phosphoinositides (phosphatidylinositol phosphates, or PIPs)**. The interconversion of phosphatidylinositol (PI) and PIPs is highly compartmentalized: different organelles in the endocytic and secretory pathways have distinct sets of PI and PIP kinases and PIP phosphatases (Figure 13–10). The distribution, regulation, and local balance of these enzymes determine the steady-state distribution of each PIP species. As a consequence, the distribution of PIPs varies from organelle to organelle, and often within a continuous membrane from one region to another, thereby defining specialized membrane domains.

Many proteins involved at different steps in vesicle transport contain domains that bind with high specificity to the head groups of particular PIPs, distinguishing one phosphorylated form from another (see Figure 13–10 E and F). Local control of the PI and PIP kinases and PIP phosphatases can therefore be used to rapidly control the binding of proteins to a membrane or membrane domain. The production of a particular type of PIP recruits proteins containing matching PIP-binding domains. The PIP-binding proteins then help regulate vesicle formation and other steps in the control of vesicle traffic (Figure 13–11). The same strategy is widely used to recruit specific intracellular signaling proteins to the plasma membrane in response to extracellular signals (discussed in Chapter 15).

**Figure 13–8** The assembly and disassembly of a clathrin coat. The assembly of the coat introduces curvature into the membrane, which leads in turn to the formation of a coated bud (called a coated pit if it is in the plasma membrane). The adaptor proteins bind both clathrin triskelions and membrane-bound cargo receptors, thereby mediating the selective recruitment of both membrane and soluble cargo molecules into the vesicle. Other membrane-bending and fission proteins are recruited to the neck of the budding vesicle, where sharp membrane curvature is introduced. The coat is rapidly lost shortly after the vesicle buds off.



**Figure 13–9** Lipid-induced conformation switching of AP2. The AP2 adaptor protein complex has four subunits ( $\alpha$ ,  $\beta 2$ ,  $\mu 2$ , and  $\sigma 2$ ). Upon interaction with the phosphoinositide  $\text{PI}(4,5)\text{P}_2$  (see Figure 13–10) in the cytosolic leaflet of the plasma membrane, AP2 rearranges so that binding sites for cargo receptors become exposed. Each AP2 complex binds four  $\text{PI}(4,5)\text{P}_2$  molecules (for clarity, only one is shown). In the open AP2 complex, the  $\mu 2$  and  $\sigma 2$  subunits bind the cytosolic tails of cargo receptors that display the appropriate endocytosis signals. These signals consist of short amino acid sequence motifs. When AP2 binds tightly to the membrane, it induces curvature, which favors the binding of additional AP2 complexes in the vicinity.



### Membrane-Bending Proteins Help Deform the Membrane During Vesicle Formation

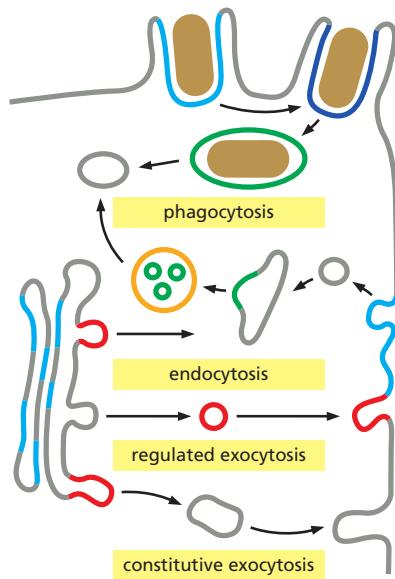
The forces generated by clathrin coat assembly alone are not sufficient to shape and pinch off a vesicle from the membrane. Other membrane-bending and force-generating proteins participate at every stage of the process. Membrane-bending proteins that contain crescent-shaped domains, called *BAR domains*, bind to and impose their shape on the underlying membrane via electrostatic interactions with the lipid head groups (Figure 13–12; also see Figure 10–40). Such BAR-domain proteins are thought to help AP2 nucleate clathrin-mediated endocytosis by shaping the plasma membrane to allow a clathrin-coated bud to form. Some of these proteins also contain amphiphilic helices that induce membrane curvature after being inserted as wedges into the cytoplasmic leaflet of the membrane. Other BAR-domain proteins are important in shaping the neck of a budding vesicle, where stabilization of sharp membrane bends is essential. Finally, the clathrin machinery nucleates the local assembly of actin filaments that introduce tension to help pinch off and propel the forming vesicle away from the membrane.

### Cytoplasmic Proteins Regulate the Pinching-Off and Uncoating of Coated Vesicles

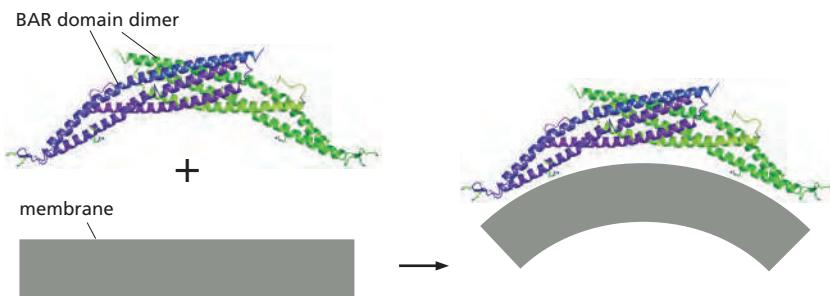
As a clathrin-coated bud grows, soluble cytoplasmic proteins, including **dynamin**, assemble at the neck of each bud (Figure 13–13). Dynamin contains a PI(4,5)P<sub>2</sub>-binding domain, which tethers the protein to the membrane, and a GTPase domain, which regulates the rate at which vesicles pinch off from the membrane.

**Figure 13–11** The intracellular location of phosphoinositides. Different types of PIPs are located in different membranes and membrane domains, where they are often associated with specific vesicle transport events. The membrane of secretory vesicles, for example, contains PI(4)P. When the vesicles fuse with the plasma membrane, a PI 5-kinase that is localized there converts the PI(4)P into PI(4,5)P<sub>2</sub>. The PI(4,5)P<sub>2</sub>, in turn, helps recruit adaptor proteins, which initiate the formation of a clathrin-coated pit, as the first step in clathrin-mediated endocytosis. Once the clathrin-coated vesicle buds off from the plasma membrane, a PI(5)P phosphatase hydrolyzes PI(4,5)P<sub>2</sub>, which weakens the binding of the adaptor proteins, promoting vesicle uncoating. We discuss phagocytosis and the distinction between regulated and constitutive exocytosis later in the chapter. (Modified from M.A. de Matteis and A. Godi, *Nat. Cell Biol.* 6:487–492, 2004. With permission from Macmillan Publishers Ltd.)

**Figure 13–10** Phosphatidylinositol (PI) and phosphoinositides (PIPs). (A, B) The structure of PI shows the free hydroxyl groups in the inositol sugar that can in principle be modified. (C) Phosphorylation of one, two, or three of the hydroxyl groups on PI by PI and PIP kinases produces a variety of PIP species. They are named according to the ring position (in parentheses) and the number of phosphate groups (subscript) added to PI. PI(3,4)P<sub>2</sub> is shown. (D) Animal cells have several PI and PIP kinases and a similar number of PIP phosphatases, which are localized to different organelles, where they are regulated to catalyze the production of particular PIPs. The red and green arrows show the kinase and phosphatase reactions, respectively. (E, F) Phosphoinositide head groups are recognized by protein domains that discriminate between the different forms. In this way, select groups of proteins containing such domains are recruited to regions of membrane in which these phosphoinositides are present. PI(3)P and PI(4,5)P<sub>2</sub> are shown. (D, modified from M.A. de Matteis and A. Godi, *Nat. Cell Biol.* 6:487–492, 2004. With permission from Macmillan Publishers Ltd.)



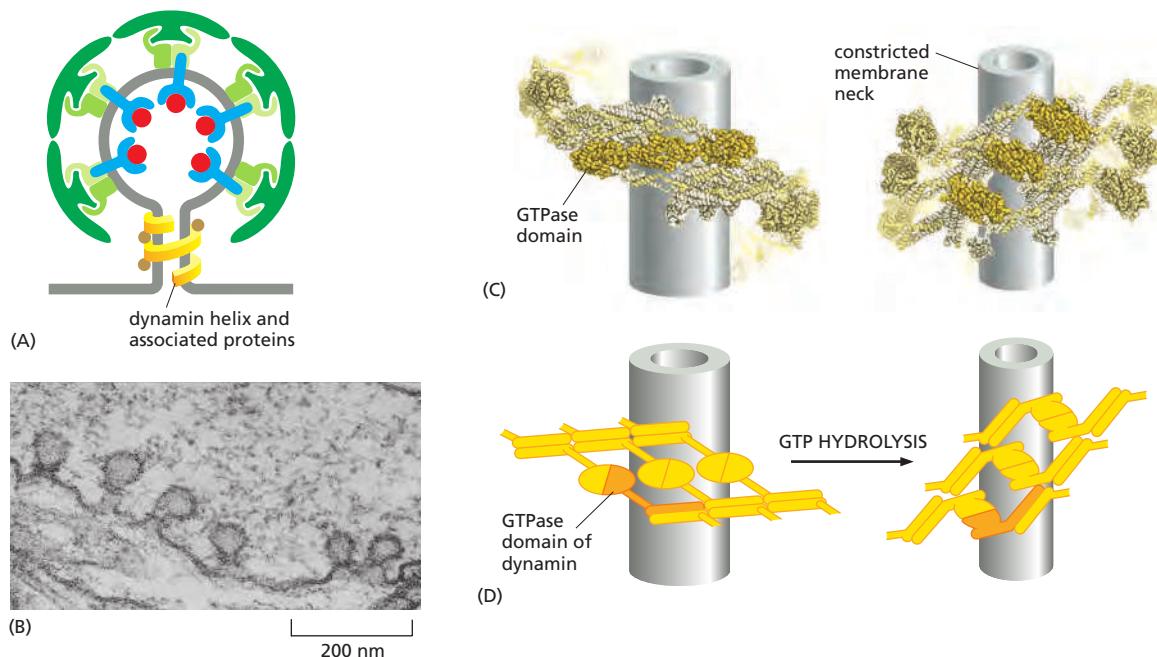
KEY: PI(3)P PI(4)P PI(4,5)P<sub>2</sub> PI(3,5)P<sub>2</sub> PI(3,4,5)P<sub>3</sub>



The pinching-off process brings the two noncytosolic leaflets of the membrane into close proximity and fuses them, sealing off the forming vesicle (see Figure 13–2). To perform this task, dynamin recruits other proteins to the neck of the bud. Together with dynamin, they help bend the patch of membrane—by directly distorting the bilayer structure, or by changing its lipid composition through the recruitment of lipid-modifying enzymes, or by both mechanisms.

Once released from the membrane, the vesicle rapidly loses its clathrin coat. A PIP phosphatase that is co-packaged into clathrin-coated vesicles depletes PI(4,5)P<sub>2</sub> from the membrane, which weakens the binding of the adaptor proteins. In addition, an hsp70 chaperone protein (see Figure 6–80) functions as an uncoating ATPase, using the energy of ATP hydrolysis to peel off the clathrin coat. *Auxilin*, another vesicle protein, is thought to activate the ATPase. The release of the coat, however, must not happen prematurely, so additional control mechanisms must somehow prevent the clathrin from being removed before it has formed a complete vesicle (discussed below).

**Figure 13–12 The structure of BAR domains.** BAR-domain proteins are diverse and enable many membrane-bending processes in the cell. BAR domains are built from coiled coils that dimerize into modules with a positively charged inner surface, which preferentially interacts with negatively charged lipid head groups to bend membranes. Local membrane deformations caused by BAR-domain proteins facilitate the binding of additional BAR-domain proteins, thereby generating a positive feedback cycle for curvature propagation. Individual BAR-domain proteins contain a distinctive curvature and often have additional features that adapt them to their specific tasks: some have short amphiphilic helices that cause further membrane deformation by wedge insertion; others are flanked by PIP-binding domains that direct them to membranes enriched in cognate phosphoinositides.



**Figure 13–13 The role of dynamin in pinching off clathrin-coated vesicles.** (A) Multiple dynamin molecules assemble into a spiral around the neck of the forming bud. The dynamin spiral is thought to recruit other proteins to the bud neck, which, together with dynamin, destabilize the interacting lipid bilayers so that the noncytoplasmic leaflets flow together. The newly formed vesicle then pinches off from the membrane. Specific mutations in dynamin can either enhance or block the pinching-off process. (B) Dynamin was discovered as the protein defective in the *shibire* mutant of *Drosophila*. These mutant flies become paralyzed because clathrin-mediated endocytosis stops, and the synaptic vesicle membrane fails to recycle, blocking neurotransmitter release. Deeply invaginated clathrin-coated pits form in the nerve endings of the fly's nerve cells, with a belt of mutant dynamin assembled around the neck, as shown in this thin-section electron micrograph. The pinching-off process fails because the required membrane fusion does not take place. (C, D) A model of how conformational changes in the GTPase domains of membrane-assembled dynamin may power a conformational change that constricts the neck of the bud. A single dynamin molecule is shown in orange in D. (B, from J.H. Koenig and K. Ikeda, *J. Neurosci.* 9:3844–3860, 1989. With permission from the Society of Neuroscience; C and D, adapted from M.G.J. Ford, S. Jenni and J. Nunnari, *Nature* 477:561–566, 2011. With permission from Macmillan Publishers.)

## Monomeric GTPases Control Coat Assembly

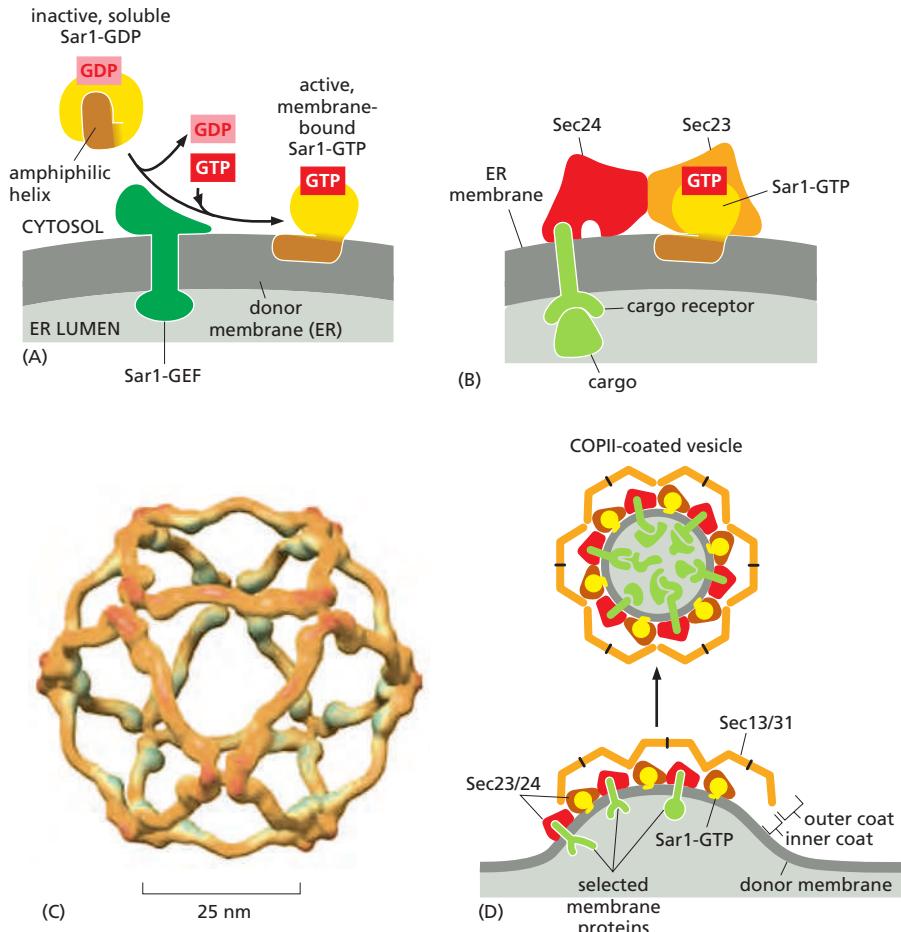
To balance the vesicle traffic to and from a compartment, coat proteins must assemble only when and where they are needed. While local production of PIPs plays a major part in regulating the assembly of clathrin coats on the plasma membrane and Golgi apparatus, cells superimpose additional ways of regulating coat formation. *Coat-recruitment GTPases*, for example, control the assembly of clathrin coats on endosomes and the COPI and COPII coats on Golgi and ER membranes.

Many steps in vesicle transport depend on a variety of GTP-binding proteins that control both the spatial and temporal aspects of vesicle formation and fusion. As discussed in Chapter 3, GTP-binding proteins regulate most processes in eukaryotic cells. They act as molecular switches, which flip between an active state with GTP bound and an inactive state with GDP bound. Two classes of proteins regulate the flipping: *guanine nucleotide exchange factors (GEFs)* activate the proteins by catalyzing the exchange of GDP for GTP, and *GTPase-activating proteins (GAPs)* inactivate the proteins by triggering the hydrolysis of the bound GTP to GDP (see Figures 3–68 and 15–7). Although both monomeric GTP-binding proteins (monomeric GTPases) and trimeric GTP-binding proteins (G proteins) have important roles in vesicle transport, the roles of the monomeric GTPases are better understood, and we focus on them here.

**Coat-recruitment GTPases** are members of a family of monomeric GTPases. They include the **ARF proteins**, which are responsible for the assembly of both COPI and clathrin coats assembly at Golgi membranes, and the **Sar1 protein**, which is responsible for the assembly of COPII coats at the ER membrane. Coat-recruitment GTPases are usually found in high concentration in the cytosol in an inactive, GDP-bound state. When a COPII-coated vesicle is to bud from the ER membrane, for example, a specific Sar1-GEF embedded in the ER membrane binds to cytosolic Sar1, causing the Sar1 to release its GDP and bind GTP in its place. (Recall that GTP is present in much higher concentration in the cytosol than GDP and therefore will spontaneously bind after GDP is released.) In its GTP-bound state, the Sar1 protein exposes an amphiphilic helix, which inserts into the cytoplasmic leaflet of the lipid bilayer of the ER membrane. The tightly bound Sar1 now recruits adaptor coat protein subunits to the ER membrane to initiate budding (Figure 13–14). Other GEFs and coat-recruitment GTPases operate in a similar way on other membranes.

The coat-recruitment GTPases also have a role in coat disassembly. The hydrolysis of bound GTP to GDP causes the GTPase to change its conformation so that its hydrophobic tail pops out of the membrane, causing the vesicle's coat to disassemble. Although it is not known what triggers the GTP hydrolysis, it has been proposed that the GTPases work like timers, which hydrolyze GTP at slow but predictable rates, to ensure that vesicle formation is synchronized with the requirements of the moment. COPII coats accelerate GTP hydrolysis by Sar1, and a fully formed vesicle will be produced only when bud formation occurs faster than the timed disassembly process; otherwise, disassembly will be triggered before a vesicle pinches off, and the process will have to start again, perhaps at a more appropriate time and place. Once a vesicle pinches off, GTP hydrolysis releases Sar1, but the sealed coat is sufficiently stabilized through many cooperative interactions, including binding to the cargo receptors in the membrane, that it may stay on the vesicle until the vesicle docks at a target membrane. There, a kinase phosphorylates the coat proteins, which completes coat disassembly and readies the vesicle for fusion.

Clathrin- and COPI-coated vesicles, by contrast, shed their coat soon after they pinch off. For COPI vesicles, the curvature of the vesicle membrane serves as a trigger to begin uncoating. An ARF-GAP is recruited to the COPI coat as it assembles. It interacts with the membrane, and senses the lipid packing density. It becomes activated when the curvature of the membrane approaches that of a transport vesicle. It then inactivates ARF, causing the coat to disassemble.

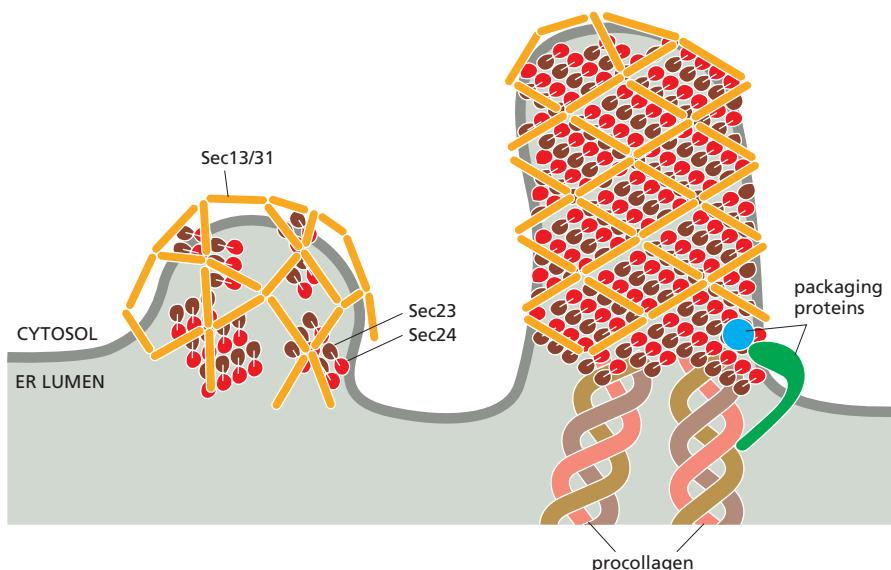


**Figure 13–14 Formation of a COPII-coated vesicle.** (A) Inactive, soluble Sar1-GDP binds to a Sar1-GEF in the ER membrane, causing the Sar1 to release its GDP and bind GTP. A GTP-triggered conformational change in Sar1 exposes an amphiphilic helix, which inserts into the cytoplasmic leaflet of the ER membrane, initiating membrane bending (which is not shown). (B) GTP-bound Sar1 binds to a complex of two COPII adaptor coat proteins, called Sec23 and Sec24, which form the inner coat. Sec24 has several different binding sites for the cytosolic tails of cargo receptors. The entire surface of the complex that attaches to the membrane is gently curved, matching the diameter of COPII-coated vesicles. (C) A complex of two additional COPII coat proteins, called Sec13 and Sec 31, forms the outer shell of the coat. Like clathrin, they can assemble on their own into symmetrical cages with appropriate dimensions to enclose a COPII-coated vesicle. (D) Membrane-bound, active Sar1-GTP recruits COPII adaptor proteins to the membrane. They select certain transmembrane proteins and cause the membrane to deform. The adaptor proteins then recruit the outer coat proteins which help form a bud. A subsequent membrane fusion event pinches off the coated vesicle. Other coated vesicles are thought to form in a similar way. (C, modified from S.M. Stagg et al., *Nature* 439:234–238, 2006. With permission from Macmillan Publishers Ltd.)

### Not All Transport Vesicles Are Spherical

Although vesicle-budding is similar at various locations in the cell, each cell membrane poses its own special challenges. The plasma membrane, for example, is comparatively flat and stiff, owing to its cholesterol-rich lipid composition and underlying actin-rich cortex. Thus, the coordinated action of clathrin coats and membrane-bending proteins has to produce sufficient force to introduce curvature, especially at the neck of the bud where sharp bends are required for the pinching-off processes. In contrast, vesicle-budding from many intracellular membranes occurs preferentially at regions where the membranes are already curved, such as the rims of the Golgi cisternae or ends of membrane tubules. In these places, the primary function of the coats is to capture the appropriate cargo proteins rather than to deform the membrane.

Transport vesicles also occur in various sizes and shapes. Diverse COPII vesicles are required for the transport of large cargo molecules. Collagen, for example, is assembled in the ER as 300-nm-long, stiff procollagen rods that then are secreted from the cell where they are cleaved by proteases to collagen, which is embedded into the extracellular matrix (discussed in Chapter 19). Procollagen rods do not fit into the 60–80 nm COPII vesicles normally observed. To circumvent this problem, the procollagen cargo molecules bind to transmembrane *packaging proteins* in the ER, which control the assembly of the COPII coat components (Figure 13–15). These events drive the local assembly of much larger COPII vesicles that accommodate the oversized cargo. Human mutations in genes encoding such packaging proteins result in collagen defects with severe consequences, such as skeletal abnormalities and other developmental defects. Similar mechanisms must regulate the sizes of vesicles required to secrete other large macromolecular complexes, including the lipoprotein particles that transport lipids out of cells.



**Figure 13–15** Packaging of procollagen into large tubular COPII-coated vesicles. The cartoons show models for two COPII coat assembly modes. The models are based on cryo-electron tomography images of reconstituted COPII vesicles. On a spherical membrane (left), the Sec23/24 inner coat proteins assemble in patches that anchor the Sec13/31 outer coat protein cage. The Sec13/31 rods assemble a cage of triangles, squares, and pentagons. When procollagen needs to be packaged (right), special packaging proteins sense the cargo and modify the coat assembly process. This interaction recruits the COPII inner coat protein Sec24 and locally enhances the rate with which Sar1 cycles on and off the membrane (not shown). In addition, a monoubiquitin is added to the Sec31 protein, changing the assembly properties of the outer cage. Sec23/24 proteins arrange in larger arrays and Sec13/31 arrange in a regular lattice of diamond shapes. As the result, a large tubular vesicle is formed that can accommodate the large cargo molecules. The packaging proteins are not part of the budding vesicle but remain in the ER. (Modified from G. Zanetti et al., eLife 2:e00951, 2013.)

Many other vesicle budding events likewise involve variations of common mechanisms. When living cells are genetically engineered to express fluorescent membrane components, the endosomes and *trans* Golgi network are seen in a fluorescence microscope to continually send out long tubules. Coat proteins assemble onto the membrane tubules and help recruit specific cargo. The tubules then either regress or pinch off with the help of dynamin-like proteins to form transport vesicles of different sizes and shapes.

Tubules have a higher surface-to-volume ratio than the larger organelles from which they form. They are therefore relatively enriched in membrane proteins compared with soluble cargo proteins. As we discuss later, this property of tubules is an important feature for sorting proteins in endosomes.

### Rab Proteins Guide Transport Vesicles to Their Target Membrane

To ensure an orderly flow of vesicle traffic, transport vesicles must be highly accurate in recognizing the correct target membrane with which to fuse. Because of the diversity and crowding of membrane systems in the cytoplasm, a vesicle is likely to encounter many potential target membranes before it finds the correct one. Specificity in targeting is ensured because all transport vesicles display surface markers that identify them according to their origin and type of cargo, and target membranes display complementary receptors that recognize the appropriate markers. This crucial process occurs in two steps. First, *Rab proteins* and *Rab effectors* direct the vesicle to specific spots on the correct target membrane. Second, *SNARE proteins* and *SNARE regulators* mediate the fusion of the lipid bilayers.

**Rab proteins** play a central part in the specificity of vesicle transport. Like the coat-recruitment GTPases discussed earlier (see Figure 13–14), Rab proteins are also monomeric GTPases. With over 60 known members, the Rab subfamily is the largest of the monomeric GTPase subfamilies. Each Rab protein is associated with one or more membrane-enclosed organelles of the secretory or endocytic pathways, and each of these organelles has at least one Rab protein on its cytosolic surface (Table 13–1). Their highly selective distribution on these membrane systems makes Rab proteins ideal molecular markers for identifying each membrane type and guiding vesicle traffic between them. Rab proteins can function on transport vesicles, on target membranes, or both.

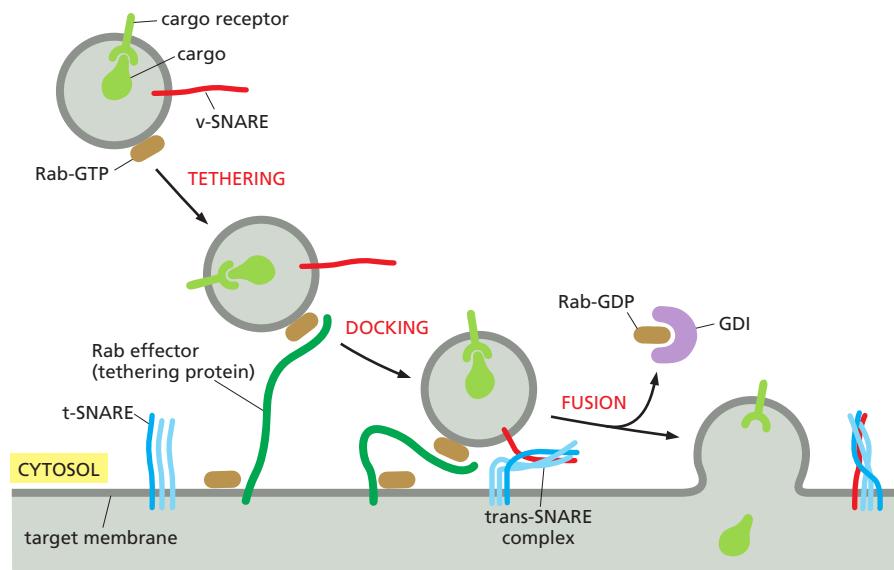
Like the coat-recruitment GTPases, Rab proteins cycle between a membrane and the cytosol and regulate the reversible assembly of protein complexes on the membrane. In their GDP-bound state, they are inactive and bound to another protein (*Rab-GDP dissociation inhibitor*, or *GDI*) that keeps them soluble in the

**TABLE 13–1 Subcellular Locations of Some Rab Proteins**

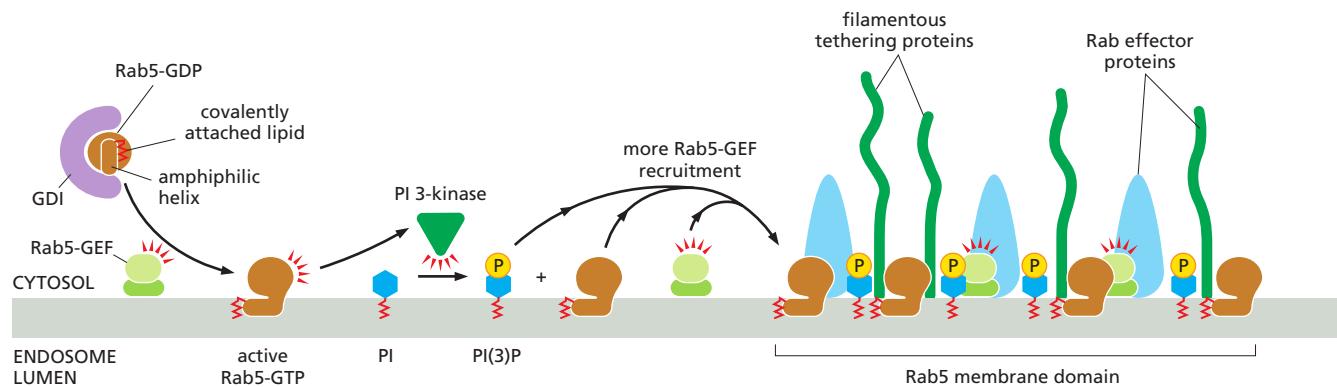
Protein	Organelle
Rab1	ER and Golgi complex
Rab2	<i>cis</i> Golgi network
Rab3A	Synaptic vesicles, secretory vesicles
Rab4/Rab11	Recycling endosomes
Rab5	Early endosomes, plasma membrane, clathrin-coated vesicles
Rab6	Medial and <i>trans</i> Golgi
Rab7	Late endosomes
Rab8	Cilia
Rab9	Late endosomes, <i>trans</i> Golgi

cytosol; in their GTP-bound state, they are active and tightly associated with the membrane of an organelle or transport vesicle. Membrane-bound Rab-GEFs activate Rab proteins on both transport vesicle and target membranes; for some membrane fusion events, activated Rab molecules are required on both sides of the reaction. Once in the GTP-bound state and membrane-bound through a now-exposed lipid anchor, Rab proteins bind to other proteins, called **Rab effectors**, which are the downstream mediators of vesicle transport, membrane tethering, and membrane fusion (Figure 13–16). The rate of GTP hydrolysis sets the concentration of active Rab and, consequently, the concentration of its effectors on the membrane.

In contrast to the highly conserved structure of Rab proteins, the structures and functions of Rab effectors vary greatly, and the same Rab proteins can often bind to many different effectors. Some Rab effectors are *motor proteins* that propel vesicles along actin filaments or microtubules to their target membrane. Others are *tethering proteins*, some of which have long, threadlike domains that serve as “fishing lines” that can extend to link two membranes more than 200 nm apart; other tethering proteins are large protein complexes that link two membranes that are closer together and interact with a wide variety of other proteins that facilitate the membrane fusion step. The tethering complex that docks COPII-coated



**Figure 13–16 Tethering of a transport vesicle to a target membrane.** Rab effector proteins interact with active Rab proteins (Rab-GTPs, yellow) located on the target membrane, vesicle membrane, or both, to establish the first connection between the two membranes that are going to fuse. In the example shown here, the Rab effector is a filamentous tethering protein (dark green). Next, SNARE proteins on the two membranes (red and blue) pair, docking the vesicle to the target membrane and catalyzing the fusion of the two apposed lipid bilayers. During docking and fusion, a Rab-GAP (not shown) induces the Rab protein to hydrolyze its bound GTP to GDP, causing the Rab to dissociate from the membrane and return to the cytosol as Rab-GDP, where it is bound by a GDI protein that keeps the Rab soluble and inactive.



vesicles, for example, contains a protein kinase that phosphorylates the coat proteins to complete the uncoating process. Coupling uncoating to vesicle delivery helps to ensure directionality of the transport process and fusion with the proper membrane. Rab effectors can also interact with SNAREs to couple membrane tethering to fusion (see Figure 13–16).

The assembly of Rab proteins and their effectors on a membrane is cooperative and results in the formation of large, specialized membrane patches. Rab5, for example, assembles on endosomes and mediates the capture of endocytic vesicles arriving from the plasma membrane. The experimental depletion of Rab5 causes disappearance of the entire endosomal and lysosomal membrane system, highlighting the crucial role of Rab proteins in organelle biogenesis and maintenance.

A Rab5 domain concentrates tethering proteins that catch incoming vesicles. Its assembly on endosomal membranes begins when a Rab5-GDP/GDI complex encounters a Rab-GEF. GDI is released and Rab5-GDP is converted to Rab5-GTP. Active Rab5-GTP becomes anchored to the membrane and recruits more Rab5-GEF to the endosome, thereby stimulating the recruitment of more Rab5 to the same site. In addition, active Rab5 activates a PI 3-kinase, which locally converts PI to PI(3)P, which in turn binds some of the Rab effectors including tethering proteins and stabilizes their local membrane attachment (Figure 13–17). This type of positive feedback greatly amplifies the assembly process and helps to establish functionally distinct membrane domains within a continuous membrane.

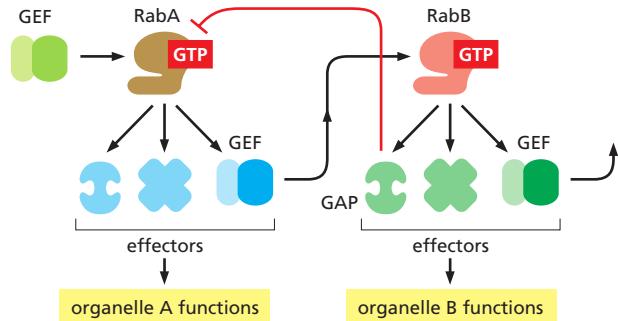
The endosomal membrane provides a striking example of how different Rab proteins and their effectors help to create multiple specialized membrane domains, each fulfilling a particular set of functions. Thus, while the Rab5 membrane domain receives incoming endocytic vesicles from the plasma membrane, distinct Rab11 and Rab4 domains in the same membrane organize the budding of recycling vesicles that return proteins from the endosome to the plasma membrane.

### Rab Cascades Can Change the Identity of an Organelle

A Rab domain can be disassembled and replaced by a different Rab domain, changing the identity of an organelle. Such ordered recruitment of sequentially acting Rab proteins is called a **Rab cascade**. Over time, for example, Rab5 domains are replaced by Rab7 domains on endosomal membranes. This converts an early endosome, marked by Rab5, into a late endosome, marked by Rab7. Because the set of Rab effectors recruited by Rab7 is different from that recruited by Rab5, this change reprograms the compartment: as we discuss later, it alters the membrane dynamics, including the incoming and outgoing traffic, and repositions the organelle away from the plasma membrane toward the cell interior. All of the cargo contained in the early endosome that has not been recycled to the plasma membrane is now part of a late endosome. This process is also referred to as *endosome maturation*. The self-amplifying nature of the Rab domains renders the process of endosome maturation unidirectional and irreversible (Figure 13–18).

**Figure 13–17** The formation of a Rab5 domain on the endosome membrane. A Rab5-GEF on the endosome membrane binds a Rab5 protein and induces it to exchange GDP for GTP. GDI is lost and GTP binding alters the conformation of the Rab protein, exposing an amphiphilic helix and a covalently attached lipid group, which together anchor the Rab5-GTP to the membrane. Active Rab5 activates PI 3-kinase, which converts PI into PI(3)P. PI(3)P and active Rab5 together bind a variety of Rab effector proteins that contain PI(3)P-binding sites, including filamentous tethering proteins that catch incoming clathrin-coated endocytic vesicles from the plasma membrane. With the help of another effector, active Rab5 also recruits more Rab5-GEF, further enhancing the assembly of the Rab5 domain on the membrane.

Controlled cycles of GTP hydrolysis and GDP–GTP exchange dynamically regulate the size and activity of such Rab domains. Unlike SNAREs, which are integral membrane proteins, the GDP/GTP cycle, coupled to the membrane/cytosol translocation cycle, endows the Rab machinery with the ability to undergo assembly and disassembly on the membrane. (Adapted from M. Zerial and H. McBride, *Nat. Rev. Mol. Cell Biol.* 2:107–117, 2001. With permission from Macmillan Publishers Ltd.)



**Figure 13–18 A model for a generic Rab cascade.** The local activation of a RabA-GEF leads to assembly of a RabA domain on the membrane. Active RabA recruits its effector proteins, one of which is a GEF for RabB. The RabB-GEF then recruits RabB to the membrane, which in turn begins to recruit its effectors, among them a GAP for RabA. The RabA-GAP activates RabA GTP hydrolysis leading to the inactivation of the RabA and the disassembly of the RabA domain as the RabB domain grows. In this way, the RabA domain is irreversibly replaced by the RabB domain. In principle, this sequence can be continued by the recruitment of a next GEF by RabB. (Adapted from A.H. Hutagalung and P.J. Novick, *Physiol. Rev.* 91:119–149, 2011. With permission from The American Physiological Society.)

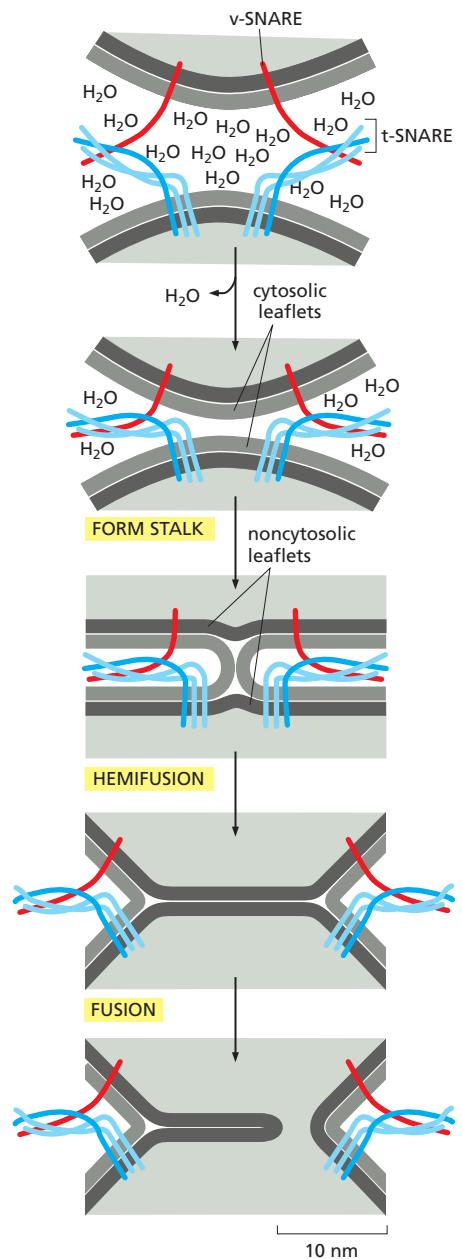
### SNAREs Mediate Membrane Fusion

Once a transport vesicle has been tethered to its target membrane, it unloads its cargo by membrane fusion. Membrane fusion requires bringing the lipid bilayers of two membranes to within 1.5 nm of each other so that they can merge. When the membranes are in such close apposition, lipids can flow from one bilayer to the other. For this close approach, water must be displaced from the hydrophilic surface of the membrane—a process that is highly energetically unfavorable and requires specialized *fusion proteins* that overcome this energy barrier. We have already discussed the role of dynamin in a related task during the pinching-off of clathrin-coated vesicles (see Figure 13–13).

The **SNARE proteins** (also called SNAREs, for short) catalyze the membrane fusion reactions in vesicle transport. There are at least 35 different SNAREs in an animal cell, each associated with a particular organelle in the secretory or endocytic pathways. These transmembrane proteins exist as complementary sets, with **v-SNAREs** usually found on vesicle membranes and **t-SNAREs** usually found on target membranes (see Figure 13–16). A v-SNARE is a single polypeptide chain, whereas a t-SNARE is usually composed of three proteins. The v-SNAREs and t-SNAREs have characteristic helical domains, and when a v-SNARE interacts with a t-SNARE, the helical domains of one wrap around the helical domains of the other to form a very stable four-helix bundle. The resulting *trans-SNARE complex* locks the two membranes together. Biochemical membrane fusion assays with all different SNARE combinations show that v- and t-SNARE pairing is highly specific. The SNAREs thus provide an additional layer of specificity in the transport process by helping to ensure that vesicles fuse only with the correct target membrane.

The trans-SNARE complexes catalyze membrane fusion by using the energy that is freed when the interacting helices wrap around each other to pull the membrane faces together, simultaneously squeezing out water molecules from the interface (Figure 13–19). When liposomes containing purified v-SNAREs are mixed with liposomes containing complementary t-SNAREs, their membranes fuse, albeit slowly. In the cell, other proteins recruited to the fusion site, presumably Rab effectors, cooperate with SNAREs to accelerate fusion. Fusion does not always follow immediately after v-SNAREs and t-SNAREs pair. As we discuss later, in the process of regulated exocytosis, fusion is delayed until secretion is triggered by a specific extracellular signal.

Rab proteins, which can regulate the availability of SNARE proteins, exert an additional layer of control. t-SNAREs in target membranes are often associated with inhibitory proteins that must be released before the t-SNARE can function. Rab proteins and their effectors trigger the release of such SNARE inhibitory



**Figure 13–19 A model for how SNARE proteins may catalyze membrane fusion.** Bilayer fusion occurs in multiple steps. A tight pairing between v- and t-SNAREs forces lipid bilayers into close apposition and expels water molecules from the interface. Lipid molecules in the two interacting (cytosolic) leaflets of the bilayers then flow between the membranes to form a connecting stalk. Lipids of the two noncytosolic leaflets then contact each other, forming a new bilayer, which widens the fusion zone (*hemifusion*, or half-fusion). Rupture of the new bilayer completes the fusion reaction.

proteins. In this way, SNARE proteins are concentrated and activated in the correct location on the membrane, where tethering proteins capture incoming vesicles. Rab proteins thus speed up the process by which appropriate SNARE proteins in two membranes find each other.

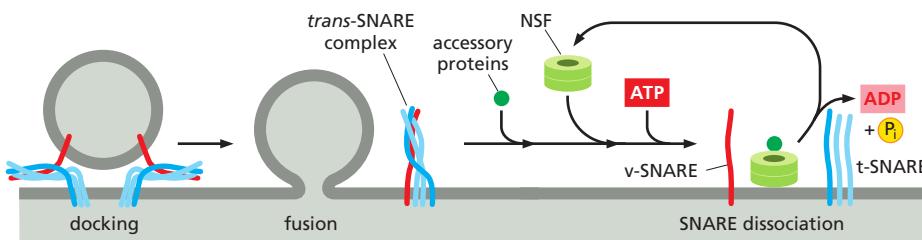
For vesicle transport to operate normally, transport vesicles must incorporate the appropriate SNARE and Rab proteins. Not surprisingly, therefore, many transport vesicles will form only if they incorporate the appropriate complement of SNARE and Rab proteins in their membrane. How this crucial control process operates during vesicle budding remains a mystery.

### Interacting SNAREs Need to Be Pried Apart Before They Can Function Again

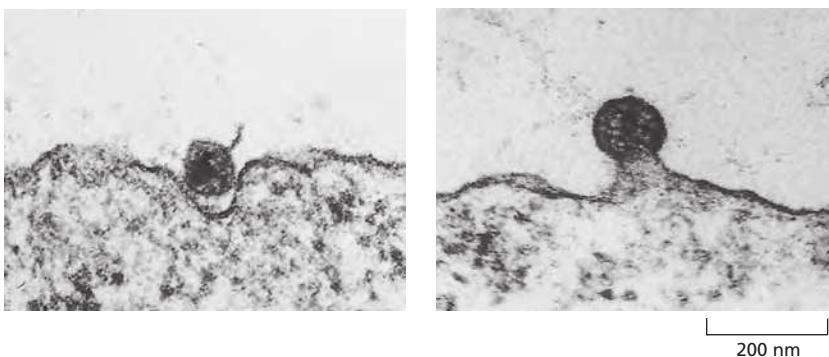
Most SNARE proteins in cells have already participated in multiple rounds of vesicle transport and are sometimes present in a membrane as stable complexes with partner SNAREs. The complexes have to disassemble before the SNAREs can mediate new rounds of transport. A crucial protein called NSF cycles between membranes and the cytosol and catalyzes the disassembly process. NSF is a hexameric ATPase of the family of AAA-ATPases (see Figure 6–85) that uses the energy of ATP hydrolysis to unravel the intimate interactions between the helical domains of paired SNARE proteins (Figure 13–20). The requirement for NSF-mediated reactivation of SNAREs by SNARE complex disassembly helps prevent membranes from fusing indiscriminately: if the t-SNAREs in a target membrane were always active, any membrane containing an appropriate v-SNARE might fuse whenever the two membranes made contact. It is not known how the activity of NSF is controlled so that the SNARE machinery is activated at the right time and place. It is also not known how v-SNAREs are selectively retrieved and returned to their compartment of origin so that they can be reused in newly formed transport vesicles.

Membrane fusion is important in other processes beside vesicle transport. The plasma membranes of a sperm and an egg fuse during fertilization, and myoblasts fuse with one another during the development of multinucleate muscle fibers (discussed in Chapter 22). Likewise, the ER network and mitochondria fuse and fragment in a dynamic way (discussed in Chapters 12 and 14). All cell membrane fusions require special proteins and are tightly regulated to ensure that only appropriate membranes fuse. The controls are crucial for maintaining both the identity of cells and the individuality of each type of intracellular compartment.

The membrane fusions catalyzed by viral fusion proteins are well understood. These proteins have a crucial role in permitting the entry of enveloped viruses (which have a lipid-bilayer-based membrane coat) into the cells that they infect (discussed in Chapters 5 and 23). For example, viruses such as the human immunodeficiency virus (HIV), which causes AIDS, bind to cell-surface receptors and then fuse with the plasma membrane of the target cell (Figure 13–21). This fusion event allows the viral nucleic acid inside the nucleocapsid to enter the cytosol, where it replicates. Other viruses, such as the influenza virus, first enter the cell by receptor-mediated endocytosis (discussed later) and are delivered to endosomes; the low pH in endosomes activates a fusion protein in the viral envelope that catalyzes the fusion of the viral and endosomal membranes, releasing the viral nucleic acid into the cytosol. Viral fusion proteins and SNAREs promote lipid bilayer fusion in similar ways.



**Figure 13–20** Dissociation of SNARE pairs by NSF after a membrane fusion cycle. After a v-SNARE and t-SNARE have mediated the fusion of a transport vesicle with a target membrane, NSF binds to the SNARE complex and, with the help of accessory proteins, hydrolyzes ATP to pry the SNAREs apart.



**Figure 13–21** The entry of enveloped viruses into cells. Electron micrographs showing how HIV enters a cell by fusing its membrane with the plasma membrane of the cell. (From B.S. Stein et al., *Cell* 49:659–668, 1987. With permission from Elsevier.)

## Summary

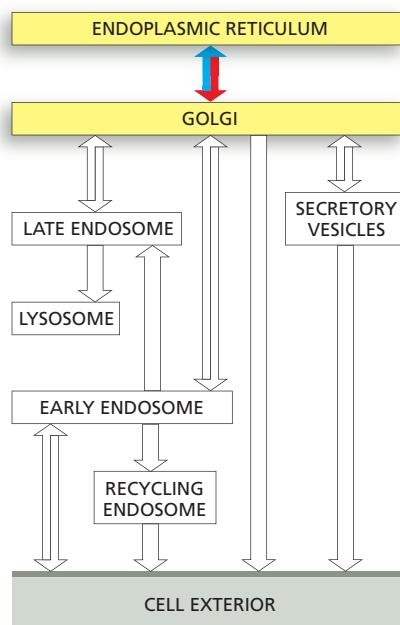
Directed and selective transport of particular membrane components from one membrane-enclosed compartment to another in a eukaryotic cell maintains the differences between those compartments. Transport vesicles, which can be spherical, tubular, or irregularly shaped, bud from specialized coated regions of the donor membrane. The assembly of the coat helps to collect specific membrane and soluble cargo molecules for transport and to drive the formation of the vesicle.

There are various types of coated vesicles. The best characterized are clathrin-coated vesicles, which mediate transport from the plasma membrane and the trans Golgi network, and COPI- and COPII-coated vesicles, which mediate transport between Golgi cisternae and between the ER and the Golgi apparatus, respectively. Coats have a common two-layered structure: an inner layer formed of adaptor proteins links the outer layer (or cage) to the vesicle membrane and also traps specific cargo molecules for packaging into the vesicle. The coat is shed before the vesicle fuses with its appropriate target membrane.

Local synthesis of specific phosphoinositides creates binding sites that trigger clathrin coat assembly and vesicle budding. In addition, monomeric GTPases help regulate various steps in vesicle transport, including both vesicle budding and docking. The coat-recruitment GTPases, including Sar1 and the ARF proteins, regulate coat assembly and disassembly. A large family of Rab proteins functions as vesicle-targeting GTPases. Rab proteins are recruited to both, forming transport vesicles and target membranes. The assembly and disassembly of Rab proteins and their effectors in specialized membrane domains are dynamically controlled by GTP binding and hydrolysis. Active Rab proteins recruit Rab effectors, such as motor proteins, which transport vesicles along actin filaments or microtubules, and filamentous tethering proteins, which help ensure that the vesicles deliver their contents only to the appropriate target membrane. Complementary v-SNARE proteins on transport vesicles and t-SNARE proteins on the target membrane form stable trans-SNARE complexes, which force the two membranes into close apposition so that their lipid bilayers can fuse.

## TRANSPORT FROM THE ER THROUGH THE GOLGI APPARATUS

As discussed in Chapter 12, newly synthesized proteins cross the ER membrane from the cytosol to enter the secretory pathway. During their subsequent transport, from the ER to the Golgi apparatus and from the Golgi apparatus to the cell surface and elsewhere, these proteins are successively modified as they pass through a series of compartments. Transfer from one compartment to the next involves a delicate balance between forward and backward (retrieval) transport pathways. Some transport vesicles select cargo molecules and move them to the next compartment in the pathway, while others retrieve escaped proteins and return them to a previous compartment where they normally function. Thus, the pathway from the ER to the cell surface consists of many sorting steps, which continuously select membrane and soluble luminal proteins for packaging and transport.



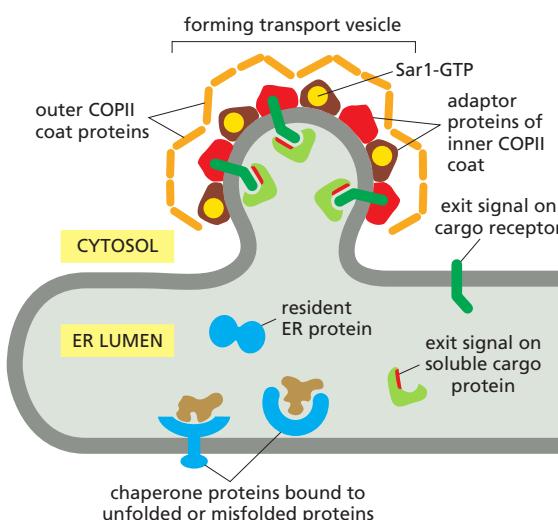
In this section, we focus mainly on the **Golgi apparatus** (also called the **Golgi complex**). It is a major site of carbohydrate synthesis, as well as a sorting and dispatching station for products of the ER. The cell makes many polysaccharides in the Golgi apparatus, including the pectin and hemicellulose of the cell wall in plants and most of the glycosaminoglycans of the extracellular matrix in animals (discussed in Chapter 19). The Golgi apparatus also lies on the exit route from the ER, and a large proportion of the carbohydrates that it makes are attached as oligosaccharide side chains to the many proteins and lipids that the ER sends to it. A subset of these oligosaccharide groups serve as tags to direct specific proteins into vesicles that then transport them to lysosomes. But most proteins and lipids, once they have acquired their appropriate oligosaccharides in the Golgi apparatus, are recognized in other ways for targeting into the transport vesicles going to other destinations.

### Proteins Leave the ER in COPII-Coated Transport Vesicles

To initiate their journey along the secretory pathway, proteins that have entered the ER and are destined for the Golgi apparatus or beyond are first packaged into COPII-coated transport vesicles. These vesicles bud from specialized regions of the ER called *ER exit sites*, whose membrane lacks bound ribosomes. Most animal cells have ER exit sites dispersed throughout the ER network.

Entry into vesicles that leave the ER can be a selective process or can happen by default. Many membrane proteins are actively recruited into such vesicles, where they become concentrated. These cargo membrane proteins display exit (transport) signals on their cytosolic surface that adaptor proteins of the inner COPII coat recognize (Figure 13–22); some of these components act as cargo receptors and are recycled back to the ER after they have delivered their cargo to the Golgi apparatus. Soluble cargo proteins in the ER lumen, by contrast, have exit signals that attach them to transmembrane cargo receptors. Proteins without exit signals can also enter transport vesicles, including protein molecules that normally function in the ER (so-called *ER resident proteins*), some of which slowly leak out of the ER and are delivered to the Golgi apparatus. Different cargo proteins enter the transport vesicles with substantially different rates and efficiencies, which may result from differences in their folding and oligomerization efficiencies and kinetics, as well as the factors already discussed. The exit step from the ER is a major checkpoint at which quality control is exerted on the proteins that a cell secretes or displays on its surface, as we discussed in Chapter 12.

The exit signals that direct soluble proteins out of the ER for transport to the Golgi apparatus and beyond are not well understood. Some transmembrane proteins that serve as cargo receptors for packaging some secretory proteins into COPII-coated vesicles are lectins that bind to oligosaccharides on the secreted



**Figure 13–22** The recruitment of membrane and soluble cargo molecules into ER transport vesicles. Membrane proteins are packaged into budding transport vesicles through interactions of exit signals on their cytosolic tails with adaptor proteins of the inner COPII coat. Some of these membrane proteins function as cargo receptors, binding soluble proteins in the ER lumen and helping to package them into vesicles. Other proteins may enter the vesicle by bulk flow. A typical 50 nm transport vesicle contains about 200 membrane proteins, which can be of many different types. As indicated, unfolded or incompletely assembled proteins are bound to chaperones and transiently retained in the ER compartment.

proteins. One such lectin, for example, binds to mannose on two secreted blood-clotting factors (Factor V and Factor VIII), thereby packaging the proteins into transport vesicles in the ER; its role in protein transport was identified because humans who lack it owing to an inherited mutation have lowered serum levels of Factors V and VIII, and they therefore bleed excessively.

### Only Proteins That Are Properly Folded and Assembled Can Leave the ER

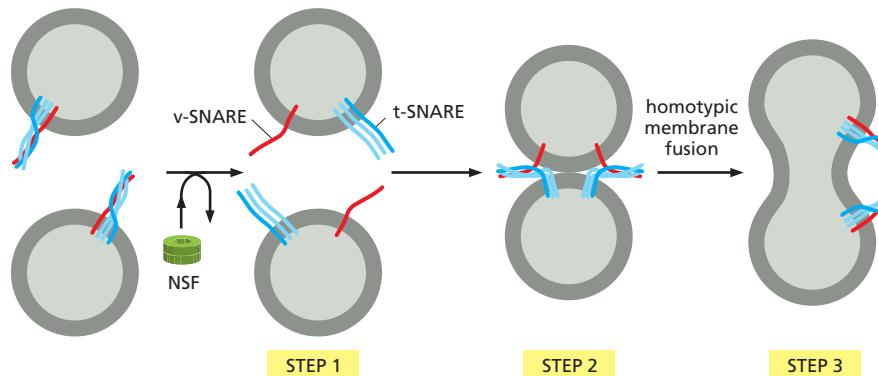
To exit from the ER, proteins must be properly folded and, if they are subunits of multiprotein complexes, they need to be completely assembled. Those that are misfolded or incompletely assembled transiently remain in the ER, where they are bound to chaperone proteins (discussed in Chapter 6) such as *BiP* or *calnexin*. The chaperones may cover up the exit signals or somehow anchor the proteins in the ER. Such failed proteins are eventually transported back into the cytosol, where they are degraded by proteasomes (discussed in Chapters 6 and 12). This quality-control step prevents the onward transport of misfolded or misassembled proteins that could potentially interfere with the functions of normal proteins. Such failures are surprisingly common. More than 90% of the newly synthesized subunits of the T cell receptor (discussed in Chapter 24) and of the acetylcholine receptor (discussed in Chapter 11), for example, are normally degraded without ever reaching the cell surface where they function. Thus, cells must make a large excess of some protein molecules to produce a select few that fold, assemble, and function properly.

Sometimes, however, there are drawbacks to the stringent quality-control mechanism. The predominant mutations that cause cystic fibrosis, a common inherited disease, result in the production of a slightly misfolded form of a plasma membrane protein important for  $\text{Cl}^-$  transport. Although the mutant protein would function normally if it reached the plasma membrane, it is retained in the ER and then is degraded by cytosolic proteasomes. This devastating disease thus results not because the mutation inactivates the protein but because the active protein is discarded before it reaches the plasma membrane.

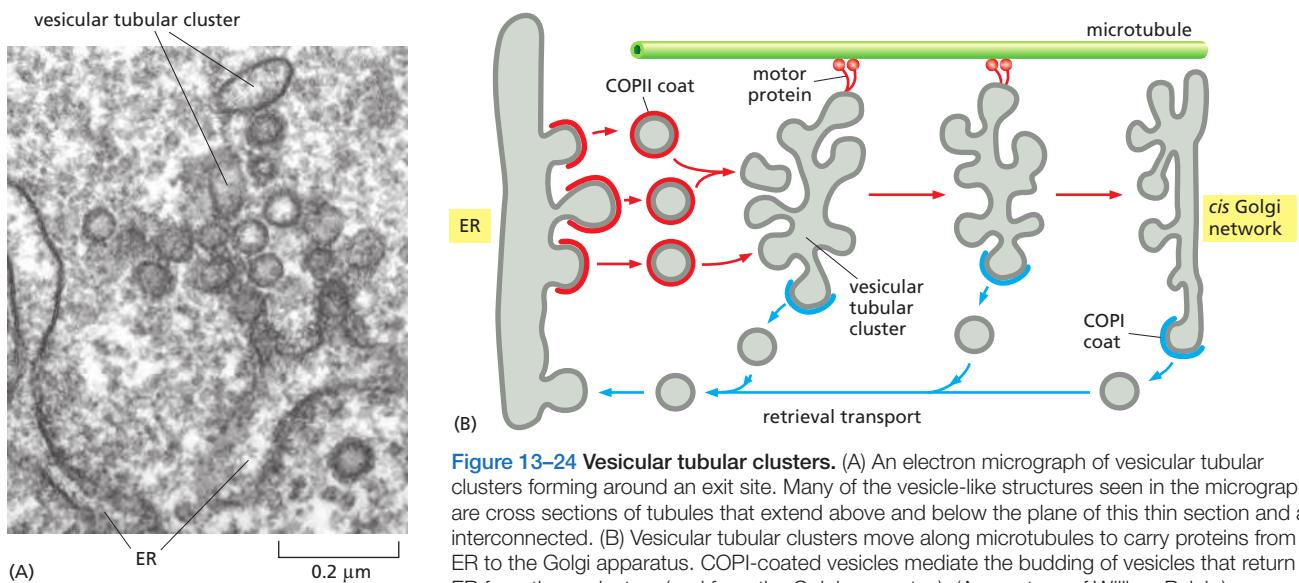
### Vesicular Tubular Clusters Mediate Transport from the ER to the Golgi Apparatus

After transport vesicles have budded from ER exit sites and have shed their coat, they begin to fuse with one another. The fusion of membranes from the same compartment is called *homotypic fusion*, to distinguish it from *heterotypic fusion*, in which a membrane from one compartment fuses with the membrane of a different compartment. As with heterotypic fusion, homotypic fusion requires a set of matching SNAREs. In this case, however, the interaction is symmetrical, with both membranes contributing v-SNAREs and t-SNAREs (Figure 13–23).

The structures formed when ER-derived vesicles fuse with one another are called *vesicular tubular clusters*, because they have a convoluted appearance in



**Figure 13–23 Homotypic membrane fusion.** In step 1, NSF pries apart identical pairs of v-SNAREs and t-SNAREs in both membranes (see Figure 13–20). In steps 2 and 3, the separated matching SNAREs on adjacent identical membranes interact, which leads to membrane fusion and the formation of one continuous compartment. Subsequently, the compartment grows by further homotypic fusion with vesicles from the same kind of membrane, displaying matching SNAREs. Homotypic fusion occurs when ER-derived transport vesicles fuse with one another, but also when endosomes fuse to generate larger endosomes. Rab proteins help regulate the extent of homotypic fusion and hence the size of a cell's compartments (not shown).



**Figure 13–24 Vesicular tubular clusters.** (A) An electron micrograph of vesicular tubular clusters forming around an exit site. Many of the vesicle-like structures seen in the micrograph are cross sections of tubules that extend above and below the plane of this thin section and are interconnected. (B) Vesicular tubular clusters move along microtubules to carry proteins from the ER to the Golgi apparatus. COPII-coated vesicles mediate the budding of vesicles that return to the ER from these clusters (and from the Golgi apparatus). (A, courtesy of William Balch.)

the electron microscope (Figure 13–24A). These clusters constitute a compartment that is separate from the ER and lacks many of the proteins that function in the ER. They are generated continually and function as transport containers that bring material from the ER to the Golgi apparatus. The clusters move quickly along microtubules to the Golgi apparatus with which they fuse (Figure 13–24B and Movie 13.2).

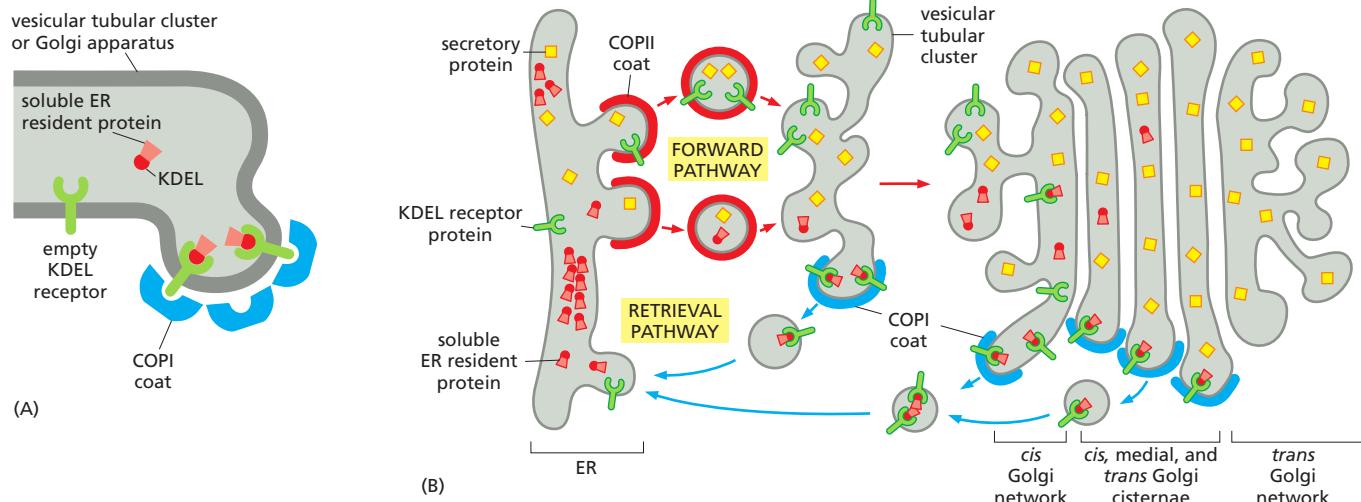
As soon as vesicular tubular clusters form, they begin to bud off transport vesicles of their own. Unlike the COPII-coated vesicles that bud from the ER, these vesicles are COPI-coated (see Figure 13–24A). COPI-coated vesicles are unique in that the components that make up the inner and outer coat layers are recruited as a preassembled complex, called *coatomer*. They function as a *retirement pathway*, carrying back ER resident proteins that have escaped, as well as proteins such as cargo receptors and SNAREs that participated in the ER budding and vesicle fusion reactions. This retrieval process demonstrates the exquisite control mechanisms that regulate coat assembly reactions. The COPI coat assembly begins only seconds after the COPII coats have been shed, and remains a mystery how this switch in coat assembly is controlled.

The retrieval (or retrograde) transport continues as the vesicular tubular clusters move toward the Golgi apparatus. Thus, the clusters continuously mature, gradually changing their composition as selected proteins are returned to the ER. The retrieval continues from the Golgi apparatus, after the vesicular tubular clusters have delivered their cargo.

### The Retrieval Pathway to the ER Uses Sorting Signals

The retrieval pathway for returning escaped proteins back to the ER depends on *ER retrieval signals*. Resident ER membrane proteins, for example, contain signals that bind directly to COPI coats and are thus packaged into COPI-coated transport vesicles for retrograde delivery to the ER. The best-characterized retrieval signal of this type consists of two lysines, followed by any two other amino acids, at the extreme C-terminal end of the ER membrane protein. It is called a *KKXX sequence*, based on the single-letter amino acid code.

Soluble ER resident proteins, such as BiP, also contain a short ER retrieval signal at their C-terminal end, but it is different: it consists of a Lys-Asp-Glu-Leu or a similar sequence. If this signal (called the *KDEL sequence*) is removed from BiP by genetic engineering, the protein is slowly secreted from the cell. If the signal is transferred to a protein that is normally secreted, the protein is now efficiently returned to the ER, where it accumulates.



Unlike the retrieval signals on ER membrane proteins, which can interact directly with the COPI coat, soluble ER resident proteins must bind to specialized receptor proteins such as the *KDEL receptor*—a multipass transmembrane protein that binds to the KDEL sequence and packages any protein displaying it into COPI-coated retrograde transport vesicles (Figure 13–25). To accomplish this task, the KDEL receptor itself must cycle between the ER and the Golgi apparatus, and its affinity for the KDEL sequence must differ in these two compartments. The receptor must have a high affinity for the KDEL sequence in vesicular tubular clusters and the Golgi apparatus, so as to capture escaped, soluble ER resident proteins that are present there at low concentration. It must have a low affinity for the KDEL sequence in the ER, however, to unload its cargo in spite of the very high concentration of KDEL-containing soluble resident proteins in the ER.

How does the affinity of the KDEL receptor change depending on the compartment in which it resides? The answer is likely related to the lower pH in the Golgi compartments, which is regulated by H<sup>+</sup> pumps. As we discuss later, pH-sensitive protein–protein interactions form the basis for many of the protein sorting steps in the cell.

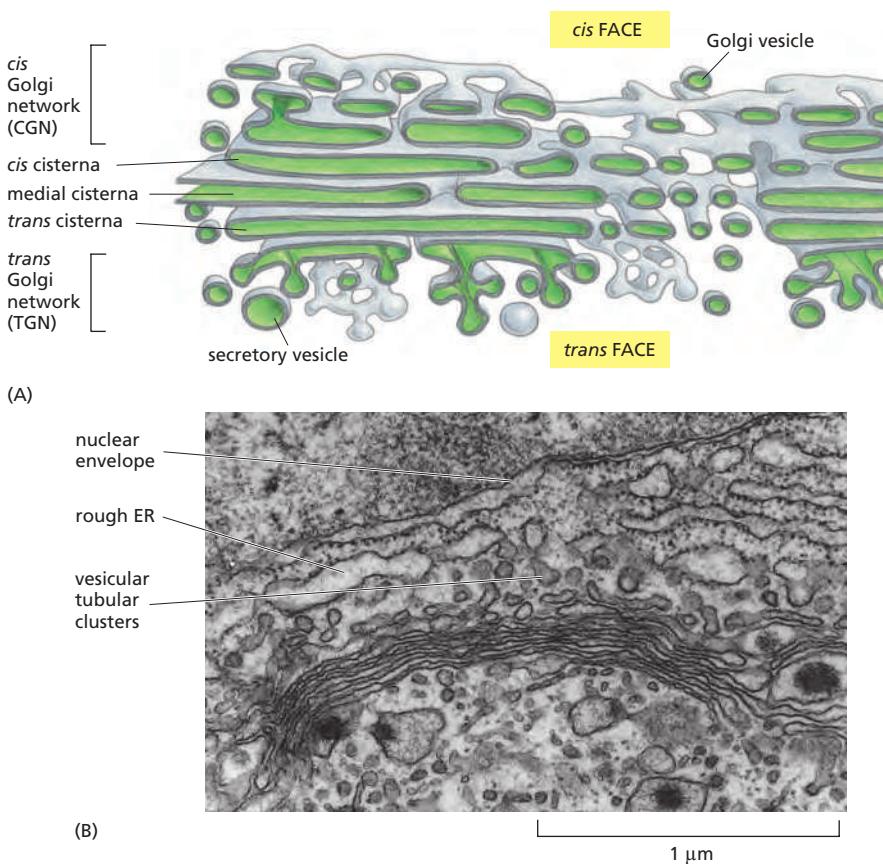
Most membrane proteins that function at the interface between the ER and Golgi apparatus, including v- and t-SNAREs and some cargo receptors, also enter the retrieval pathway back to the ER.

### Many Proteins Are Selectively Retained in the Compartments in Which They Function

The KDEL retrieval pathway only partly explains how ER resident proteins are maintained in the ER. As mentioned, cells that express genetically modified ER resident proteins, from which the KDEL sequence has been experimentally removed, secrete these proteins. But the rate of secretion is much slower than for a normal secretory protein. It seems that a mechanism that is independent of their KDEL signal normally retains ER resident proteins and that only those proteins that escape this retention mechanism are captured and returned via the KDEL receptor. A suggested retention mechanism is that ER resident proteins bind to one another, thus forming complexes that are too big to enter transport vesicles efficiently. Because ER resident proteins are present in the ER at very high concentrations (estimated to be millimolar), relatively low-affinity interactions would suffice to retain most of the proteins in such complexes.

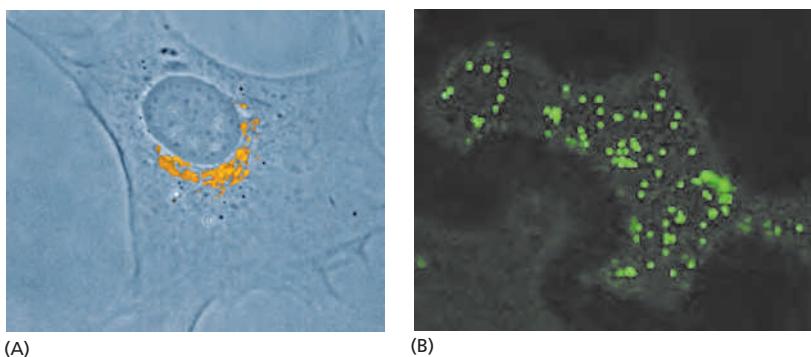
Aggregation of proteins that function in the same compartment is a general mechanism that compartments use to organize and retain their resident proteins. Golgi enzymes that function together, for example, also bind to each other and are thereby restrained from entering transport vesicles leaving the Golgi apparatus.

**Figure 13–25** Retrieval of soluble ER resident proteins. ER resident proteins that escape from the ER are returned by vesicle transport. (A) The KDEL receptor present in both vesicular tubular clusters and the Golgi apparatus captures the soluble ER resident proteins and carries them in COPI-coated transport vesicles back to the ER. (Recall that the COPI-coated vesicles shed their coats as soon as they are formed.) Upon binding its ligands in the tubular cluster or Golgi, the KDEL receptor may change conformation, so as to facilitate its recruitment into budding COPI-coated vesicles. (B) The retrieval of ER proteins begins in vesicular tubular clusters and continues from later parts of the Golgi apparatus. In the environment of the ER, the ER resident proteins dissociate from the KDEL receptor, which is then returned to the Golgi apparatus for reuse. We discuss the different compartments of the Golgi apparatus shortly.



### The Golgi Apparatus Consists of an Ordered Series of Compartments

Because it could be selectively visualized by silver stains, the Golgi apparatus was one of the first organelles described by early light microscopists. It consists of a collection of flattened, membrane-enclosed compartments called *cisternae*, that somewhat resemble a stack of pita breads. Each Golgi stack typically consists of four to six cisternae (Figure 13-26), although some unicellular flagellates can have more than 20. In animal cells, tubular connections between corresponding cisternae link many stacks, thus forming a single complex, which is usually located near the cell nucleus and close to the centrosome (Figure 13-27A). This localization depends on microtubules. If microtubules are experimentally depolymerized, the Golgi apparatus reorganizes into individual stacks that are found throughout the cytoplasm, adjacent to ER exit sites. Some cells, including most plant cells, have hundreds of individual Golgi stacks dispersed throughout the cytoplasm where they are typically found adjacent to ER exit sites (Figure 13-27B).



**Figure 13-26** The Golgi apparatus.

(A) Three-dimensional reconstruction from electron micrographs of the Golgi apparatus in a secretory animal cell. The *cis* face of the Golgi stack is that closest to the ER. (B) A thin-section electron micrograph of an animal cell. In plant cells, the Golgi apparatus is generally more distinct and more clearly separated from other intracellular membranes than in animal cells. (A, redrawn from A. Rambour and Y. Clermont, *Eur. J. Cell Biol.* 51:189–200, 1990. With permission from Wissenschaftliche Verlagsgesellschaft; B, courtesy of Brij J. Gupta.)

**Figure 13-27** Localization of the Golgi apparatus in animal and plant cells.

(A) The Golgi apparatus in a cultured fibroblast stained with a fluorescent antibody that recognizes a Golgi resident protein (bright orange). The Golgi apparatus is polarized, facing the direction in which the cell was crawling before fixation. (B) The Golgi apparatus in a plant cell that is expressing a fusion protein consisting of a resident Golgi enzyme fused to green fluorescent protein. (A, courtesy of John Henley and Mark McNiven; B, courtesy of Chris Hawes.)

During their passage through the Golgi apparatus, transported molecules undergo an ordered series of covalent modifications. Each Golgi stack has two distinct faces: a *cis* face (or entry face) and a *trans* face (or exit face). Both *cis* and *trans* faces are closely associated with special compartments, each composed of a network of interconnected tubular and cisternal structures: the *cis* Golgi network (CGN) and the *trans* Golgi network (TGN), respectively. The CGN is a collection of fused vesicular tubular clusters arriving from the ER. Proteins and lipids enter the *cis* Golgi network and exit from the *trans* Golgi network, bound for the cell surface or another compartment. Both networks are important for protein sorting: proteins entering the CGN can either move onward in the Golgi apparatus or be returned to the ER. Similarly, proteins exiting from the TGN move onward and are sorted according to their next destination: endosomes, secretory vesicles, or the cell surface. They also can be returned to an earlier compartment. Some membrane proteins are retained in the part of the Golgi apparatus where they function.

As described in Chapter 12, a single species of *N*-linked oligosaccharide is attached *en bloc* to many proteins in the ER and then trimmed while the protein is still in the ER. The oligosaccharide intermediates created by the trimming reactions serve to help proteins fold and to help transport misfolded proteins to the cytosol for degradation in proteasomes. Thus, they play an important role in controlling the quality of proteins exiting from the ER. Once these ER functions have been fulfilled, the cell reutilizes the oligosaccharides for new functions. This begins in the Golgi apparatus, which generates the heterogeneous oligosaccharide structures seen in mature proteins. After arrival in the CGN, proteins enter the first of the Golgi processing compartments (the *cis* Golgi cisternae). They then move to the next compartment (the medial cisternae) and finally to the *trans* cisternae, where glycosylation is completed. The lumen of the *trans* cisternae is thought to be continuous with the TGN, the place where proteins are segregated into different transport packages and dispatched to their final destinations.

The oligosaccharide processing steps occur in an organized sequence in the Golgi stack, with each cisterna containing a characteristic mixture of processing enzymes. Proteins are modified in successive stages as they move from cisterna to cisterna across the stack, so that the stack forms a multistage processing unit.

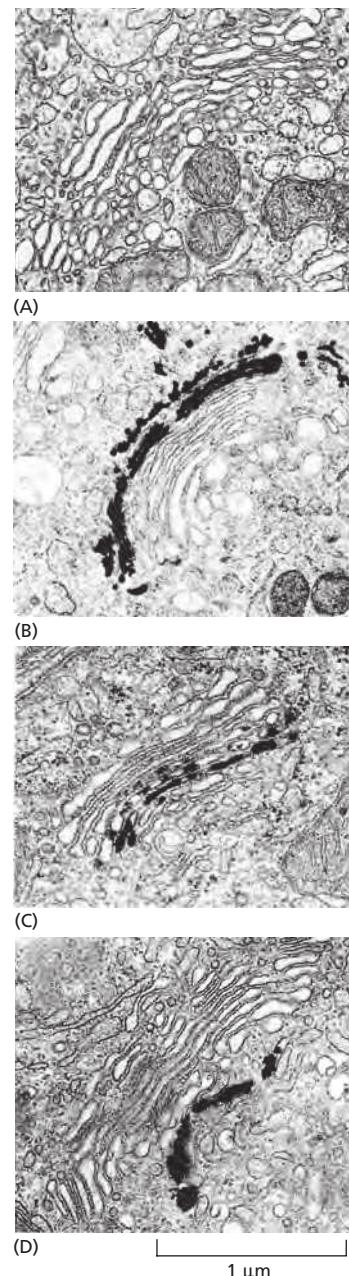
Investigators discovered the functional differences between the *cis*, medial, and *trans* subdivisions of the Golgi apparatus by localizing the enzymes involved in processing *N*-linked oligosaccharides in distinct regions of the organelle, both by physical fractionation of the organelle and by labeling the enzymes in electron microscope sections with antibodies (Figure 13–28). The removal of mannose and the addition of *N*-acetylglucosamine, for example, occur in the *cis* and medial cisternae, while the addition of galactose and sialic acid occurs in the *trans* cisterna and *trans* Golgi network. Figure 13–29 summarizes the functional compartmentalization of the Golgi apparatus.

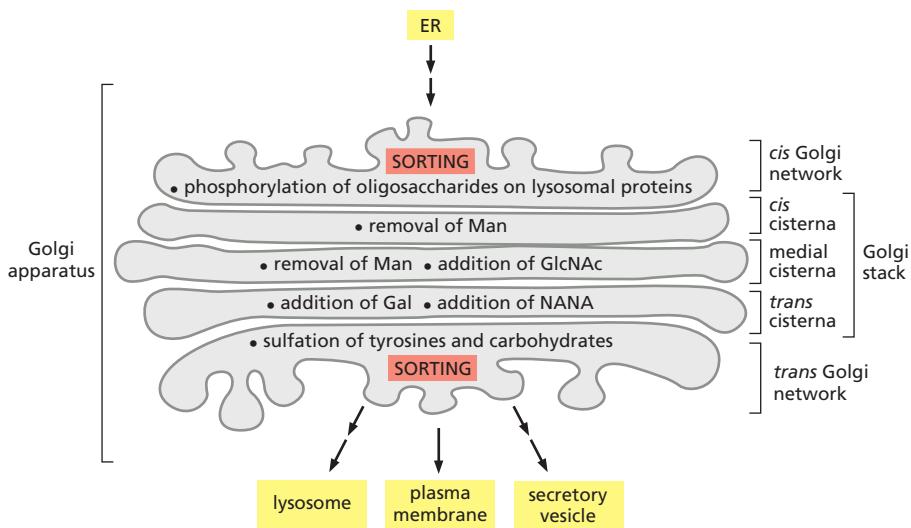
### Oligosaccharide Chains Are Processed in the Golgi Apparatus

Whereas the ER lumen is full of soluble luminal resident proteins and enzymes, the resident proteins in the Golgi apparatus are all membrane bound, as the enzymatic reactions apparently occur entirely on membrane surfaces. All of the Golgi glycosidases and glycosyl transferases, for example, are single-pass transmembrane proteins, many of which are organized in multienzyme complexes.

**Figure 13–28 Molecular compartmentalization of the Golgi apparatus.**

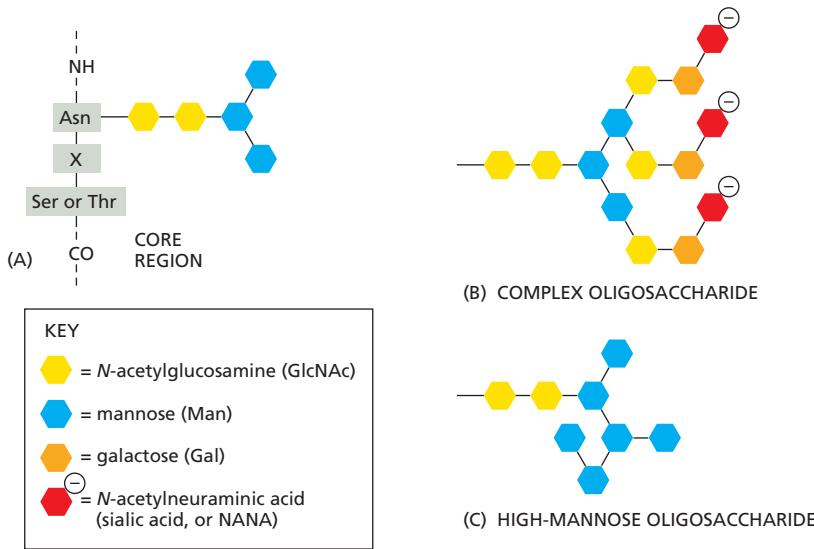
A series of electron micrographs shows the Golgi apparatus (A) unstained, (B) stained with osmium, which preferentially labels the cisternae of the *cis* compartment, and (C and D) stained to reveal the location of specific enzymes. Nucleoside diphosphatase is found in the *trans* Golgi cisternae (C), while acid phosphatase is found in the *trans* Golgi network (D). Note that usually more than one cisterna is stained. The enzymes are therefore thought to be highly enriched rather than precisely localized to a specific cisterna. (Courtesy of Daniel S. Friend.)





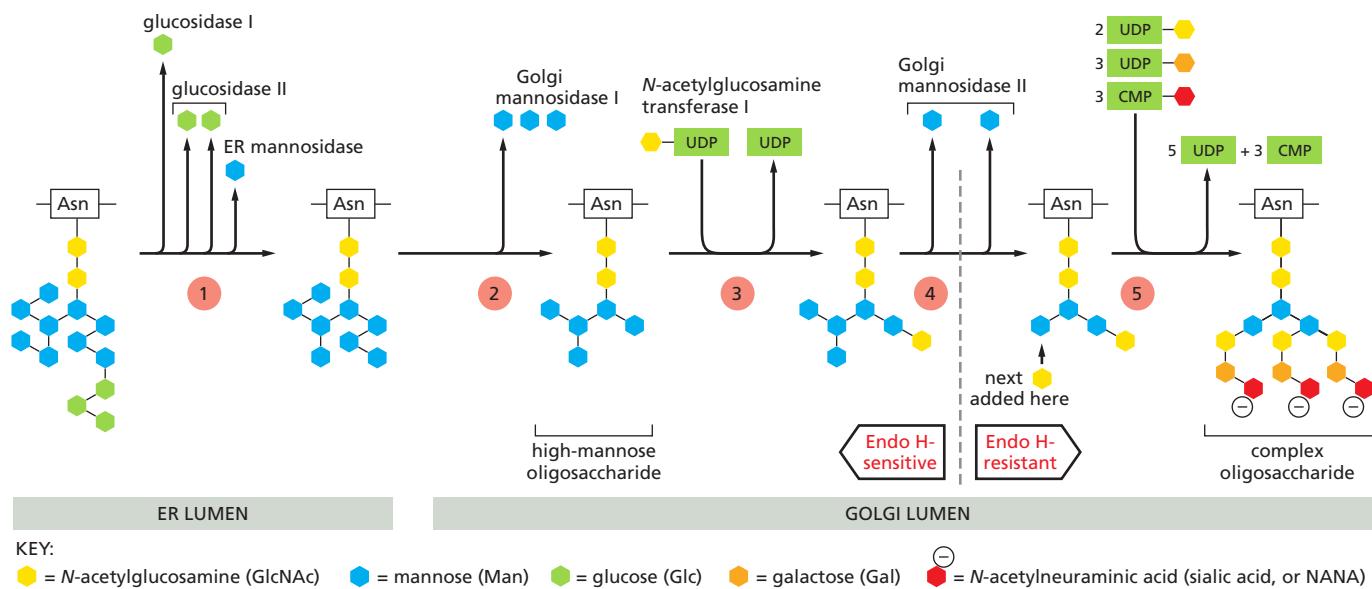
**Figure 13–29 Oligosaccharide processing in Golgi compartments.** The localization of each processing step shown was determined by a combination of techniques, including biochemical subfractionation of the Golgi apparatus membranes and electron microscopy after staining with antibodies specific for some of the processing enzymes. Processing enzymes are not restricted to a particular cisterna; instead, their distribution is graded across the stack, such that early-acting enzymes are present mostly in the *cis* Golgi cisternae and later-acting enzymes are mostly in the *trans* Golgi cisternae. Man, mannose; GlcNAc, *N*-acetylglucosamine; Gal, galactose; NANA, *N*-acetylneurameric acid (sialic acid).

Two broad classes of *N*-linked oligosaccharides, the **complex oligosaccharides** and the **high-mannose oligosaccharides**, are attached to mammalian glycoproteins. Sometimes, both types are attached (in different places) to the same polypeptide chain. Complex oligosaccharides are generated when the original *N*-linked oligosaccharide added in the ER is trimmed and further sugars are added; by contrast, high-mannose oligosaccharides are trimmed but have no new sugars added to them in the Golgi apparatus (Figure 13–30). The sialic acids in the



**Figure 13–30** The two main classes of asparagine-linked (*N*-linked) oligosaccharides found in mature mammalian glycoproteins. (A) Both complex oligosaccharides and high-mannose oligosaccharides share a common *core region* derived from the original *N*-linked oligosaccharide added in the ER (see Figure 12–50) and typically containing two *N*-acetylglucosamines (GlcNAc) and three mannoses (Man). (B) Each complex oligosaccharide consists of a *core region*, together with a *terminal region* that contains a variable number of copies of a special trisaccharide unit (*N*-acetylglucosamine–galactose–sialic acid) linked to the core mannoses. Frequently, the terminal region is truncated and contains only GlcNAc and galactose (Gal) or just GlcNAc. In addition, a fucose may be added, usually to the core GlcNAc attached to the asparagine (Asn). Thus, although the steps of processing and subsequent sugar addition are rigidly ordered, complex oligosaccharides can be heterogeneous. Moreover, although the complex oligosaccharide shown has three terminal branches, two and four branches are also common, depending on the glycoprotein and the cell in which it is made. (C) High-mannose oligosaccharides are not trimmed back all the way to the core region and contain additional mannoses. Hybrid oligosaccharides with one Man branch and one GlcNAc and Gal branch are also found (not shown).

The three amino acids indicated in (A) constitute the sequence recognized by the oligosaccharyl transferase enzyme that adds the initial oligosaccharide to the protein. Ser, serine; Thr, threonine; X, any amino acid, except proline.



**Figure 13–31 Oligosaccharide processing in the ER and the Golgi apparatus.** The processing pathway is highly ordered, so that each step shown depends on the previous one. Step 1: Processing begins in the ER with the removal of the glucoses from the oligosaccharide initially transferred to the protein. Then a mannosidase in the ER membrane removes a specific mannose. The remaining steps occur in the Golgi stack. Step 2: Golgi mannosidase I removes three more mannoses. Step 3: N-acetylglucosamine transferase I then adds an N-acetylglucosamine. Step 4: Mannosidase II then removes two additional mannoses. This yields the final core of three mannoses that is present in a complex oligosaccharide. At this stage, the bond between the two N-acetylglucosamines in the core becomes resistant to attack by a highly specific endoglycosidase (*Endo H*). Since all later structures in the pathway are also Endo H-resistant, treatment with this enzyme is widely used to distinguish complex from high-mannose oligosaccharides. Step 5: Finally, as shown in Figure 13–30, additional N-acetylglucosamines, galactoses, and sialic acids are added. These final steps in the synthesis of a complex oligosaccharide occur in the cisternal compartments of the Golgi apparatus: three types of glycosyl transferases act sequentially, using sugar substrates that have been activated by linkage to the indicated nucleotide; the membranes of the Golgi cisternae contain specific carrier proteins that allow each sugar nucleotide to enter in exchange for the nucleoside phosphates that are released after the sugar is attached to the protein on the luminal face.

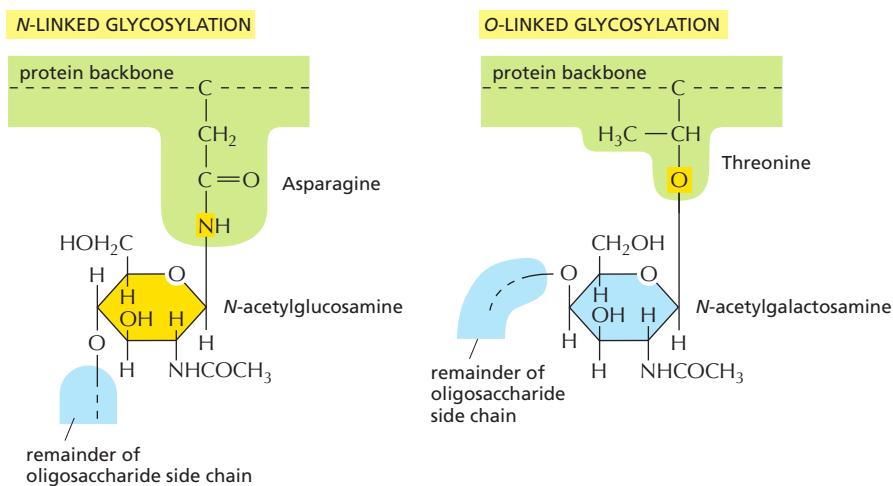
Note that, as a biosynthetic organelle, the Golgi apparatus differs from the ER: all sugars in the Golgi are assembled inside the lumen from sugar nucleotide, whereas in the ER, the *N*-linked precursor oligosaccharide is assembled partly in the cytosol and partly in the lumen, and most luminal reactions use dolichol-linked sugars as their substrates (see Figure 12–51).

complex oligosaccharides are of special importance because they bear a negative charge. Whether a given oligosaccharide remains high-mannose or is processed depends largely on its position in the protein. If the oligosaccharide is accessible to the processing enzymes in the Golgi apparatus, it is likely to be converted to a complex form; if it is inaccessible because its sugars are tightly held to the protein's surface, it is likely to remain in a high-mannose form. The processing that generates complex oligosaccharide chains follows the highly ordered pathway shown in **Figure 13–31**.

Beyond these commonalities in oligosaccharide processing that are shared among most cells, the products of the carbohydrate modifications carried out in the Golgi apparatus are highly complex and have given rise to a new field of study called glycobiology. The human genome, for example, encodes hundreds of different Golgi glycosyl transferases and many glycosidases, which are expressed differently from one cell type to another, resulting in a variety of glycosylated forms of a given protein or lipid in different cell types and at varying stages of differentiation, depending on the spectrum of enzymes expressed by the cell. The complexity of modifications is not limited to *N*-linked oligosaccharides but also occurs on *O*-linked sugars, as we discuss next.

### Proteoglycans Are Assembled in the Golgi Apparatus

In addition to the *N*-linked oligosaccharide alterations made to proteins as they pass through the Golgi cisternae *en route* from the ER to their final destinations, many proteins are also modified in the Golgi apparatus in other ways. Some proteins have sugars added to the hydroxyl groups of selected serines or threonines,



**Figure 13-32** *N*- and *O*-linked glycosylation. In each case, only the single sugar group that is directly attached to the protein chain is shown.

or, in some cases—such as collagens—to hydroxylated proline and lysine side chains. This **O-linked glycosylation** (Figure 13-32), like the extension of *N*-linked oligosaccharide chains, is catalyzed by a series of glycosyl transferase enzymes that use the sugar nucleotides in the lumen of the Golgi apparatus to add sugars to a protein one at a time. Usually, *N*-acetylgalactosamine is added first, followed by a variable number of additional sugars, ranging from just a few to 10 or more.

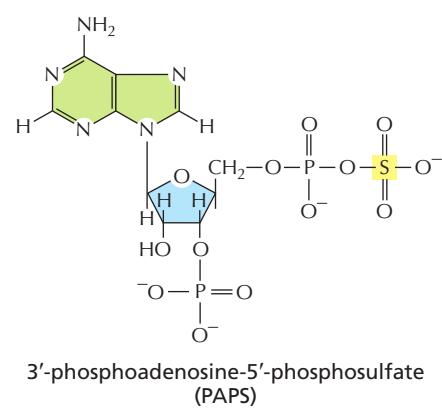
The Golgi apparatus confers the heaviest *O*-linked glycosylation of all on *mucins*, the glycoproteins in mucus secretions, and on *proteoglycan core proteins*, which it modifies to produce **proteoglycans**. As discussed in Chapter 19, this process involves the polymerization of one or more *glycosaminoglycan chains* (long, unbranched polymers composed of repeating disaccharide units; see Figure 19-35) onto serines on a core protein. Many proteoglycans are secreted and become components of the extracellular matrix, while others remain anchored to the extracellular face of the plasma membrane. Still others form a major component of slimy materials, such as the mucus that is secreted to form a protective coating on the surface of many epithelia.

The sugars incorporated into glycosaminoglycans are heavily sulfated in the Golgi apparatus immediately after these polymers are made, thus adding a significant portion of their characteristically large negative charge. Some tyrosines in proteins also become sulfated shortly before they exit from the Golgi apparatus. In both cases, the sulfation depends on the sulfate donor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) (Figure 13-33), which is transported from the cytosol into the lumen of the *trans* Golgi network.

### What Is the Purpose of Glycosylation?

There is an important difference between the construction of an oligosaccharide and the synthesis of other macromolecules such as DNA, RNA, and protein. Whereas nucleic acids and proteins are copied from a template in a repeated series of identical steps using the same enzyme or set of enzymes, complex carbohydrates require a different enzyme at each step, each product being recognized as the exclusive substrate for the next enzyme in the series. The vast abundance of glycoproteins and the complicated pathways that have evolved to synthesize them emphasize that the oligosaccharides on glycoproteins and glycosphingolipids have very important functions.

*N*-linked glycosylation, for example, is prevalent in all eukaryotes, including yeasts. *N*-linked oligosaccharides also occur in a very similar form in archaeal cell wall proteins, suggesting that the whole machinery required for their synthesis is evolutionarily ancient. *N*-linked glycosylation promotes protein folding in two ways. First, it has a direct role in making folding intermediates more soluble, thereby preventing their aggregation. Second, the sequential modifications of the *N*-linked oligosaccharide establish a “glyco-code” that marks the progression of



**Figure 13-33** The structure of PAPS.

protein folding and mediates the binding of the protein to chaperones (discussed in Chapter 12) and lectins—for example, in guiding ER-to-Golgi transport. As we discuss later, lectins also participate in protein sorting in the *trans* Golgi network.

Because chains of sugars have limited flexibility, even a small *N*-linked oligosaccharide protruding from the surface of a glycoprotein (Figure 13–34) can limit the approach of other macromolecules to the protein surface. In this way, for example, the presence of oligosaccharides tends to make a glycoprotein more resistant to digestion by proteolytic enzymes. It may be that the oligosaccharides on cell-surface proteins originally provided an ancestral cell with a protective coat; compared to the rigid bacterial cell wall, such a sugar coat has the advantage that it leaves the cell with the freedom to change shape and move.

The sugar chains have since become modified to serve other purposes as well. The mucus coat of lung and intestinal cells, for example, protects against many pathogens. The recognition of sugar chains by *lectins* in the extracellular space is important in many developmental processes and in cell-cell recognition: *selectins*, for example, are transmembrane lectins that function in cell-cell adhesion during blood cell migration, as discussed in Chapter 19. The presence of oligosaccharides may modify a protein's antigenic and functional properties, making glycosylation an important factor in the production of proteins for pharmaceutical purposes.

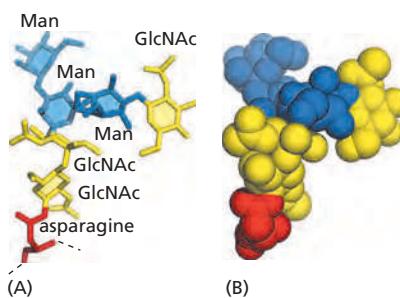
Glycosylation can also have important regulatory roles. Signaling through the cell-surface signaling receptor Notch, for example, is an important factor in determining the cell's fate in development (discussed in Chapter 21). Notch is a transmembrane protein that is *O*-glycosylated by addition of a single fucose to some serines, threonines, and hydroxylysines. Some cell types express an additional glycosyl transferase that adds an *N*-acetylglucosamine to each of these fucoses in the Golgi apparatus. This addition changes the specificity of Notch for the cell-surface signal proteins that activate it.

### Transport Through the Golgi Apparatus May Occur by Cisternal Maturation

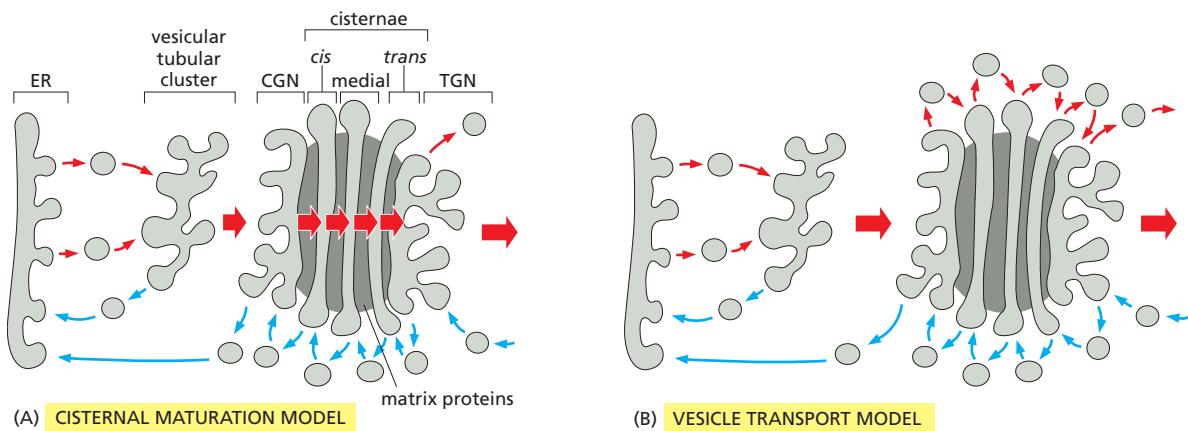
It is still uncertain how the Golgi apparatus achieves and maintains its polarized structure and how molecules move from one cisterna to another, and it is likely that more than one mechanism is involved in each case. One hypothesis, called the **cisternal maturation model**, views the Golgi cisternae as dynamic structures that mature from early to late by acquiring and then losing specific Golgi-resident proteins. According to this view, new *cis* cisternae continually form as vesicular tubular clusters arrive from the ER and progressively mature to become a medial cisterna and then a *trans* cisterna (Figure 13–35A). A cisterna therefore moves through the Golgi stack with cargo in its lumen. Retrograde transport of the Golgi enzymes by budding COPI-coated vesicles explains their characteristic distribution. As we discuss later, when a cisterna finally moves forward to become part of the *trans* Golgi network, various types of coated vesicles bud off it until this network disappears, to be replaced by a maturing cisterna just behind. At the same time, other transport vesicles are continually retrieving membrane from post-Golgi compartments and returning it to the *trans* Golgi network.

The cisternal maturation model is supported by studies using Golgi enzymes from different cisternae that were fluorescently labeled with different colors. Such studies performed in yeast cells where Golgi cisternae are not stacked reveal that individual Golgi cisternae change their color, thereby demonstrating that they change their complement of resident enzymes as they mature, even though they are not stacked. In further support of the model, electron microscopic observations found that large structures such as procollagen rods in fibroblasts and scales in certain algae move progressively through the Golgi stack.

An alternative view holds that Golgi cisternae are long-lived structures that retain their characteristic set of Golgi-resident proteins firmly in place, and cargo proteins are transported from one cisterna to the next by transport vesicles (Figure 13–35B). According to this **vesicle transport model**, retrograde flow of vesicles



**Figure 13–34** The three-dimensional structure of a small *N*-linked oligosaccharide. The structure was determined by x-ray crystallographic analysis of a glycoprotein. This oligosaccharide contains only 6 sugars, whereas there are 14 sugars in the *N*-linked oligosaccharide that is initially transferred to proteins in the ER (see Figure 12–47). (A) A backbone model showing all atoms except hydrogens; only the asparagine of the protein is shown. (B) A space-filling model, with the asparagine and sugars indicated using the same color scheme as in (A). (B, courtesy of Richard Feldmann.)



**Figure 13–35** Two possible models explaining the organization of the Golgi apparatus and how proteins move through it.

It is likely that the transport through the Golgi apparatus in the forward direction (red arrows) involves elements of both models. (A) According to the cisternal maturation model, each Golgi cisterna matures as it migrates outward through the stack. At each stage, the Golgi resident proteins that are carried forward in a maturing cisterna are moved backward to an earlier compartment in COPI-coated vesicles. When a newly formed cisterna moves to a medial position, for example, “leftover” *cis* Golgi enzymes would be extracted and transported retrogradely to a new *cis* cisterna behind. Likewise, the medial enzymes would be received by retrograde transport from the cisternae just ahead. In this way, a *cis* cisterna would mature to a medial and then *trans* cisterna as it moves outward. (B) In the vesicle transport model, Golgi cisternae are static compartments, which contain a characteristic complement of resident enzymes. The passing of molecules from *cis* to *trans* through the Golgi is accomplished by forward-moving transport vesicles, which bud from one cisterna and fuse with the next in a *cis*-to-*trans* direction.

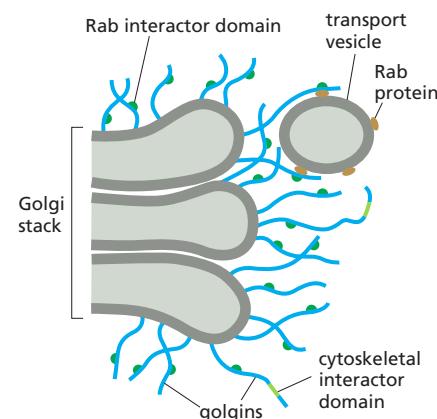
retrieves escaped ER and Golgi proteins and returns them to upstream compartments. Directional flow could be achieved because forward-moving cargo molecules are selectively packaged into forward-moving vesicles. Although both forward- and backward-moving vesicles would likely be COPI-coated, the coats may contain different adaptor proteins that confer selectivity on the packaging of cargo molecules. Alternatively, transport vesicles shuttling between Golgi cisternae might not be directional at all, transporting cargo randomly back and forth; directional flow would then occur because of the continual input to the *cis* cisterna and output from the *trans* cisterna.

The vesicle transport model is supported by experiments that show that cargo molecules are present in small COPI-coated vesicles and that these vesicles can deliver them to Golgi cisternae over large distances. In addition, when experimentally aggregated membrane proteins are introduced into Golgi cisternae, they can be observed staying in place, while soluble cargo, even if present as large aggregates, traverses the Golgi at normal rates.

It is likely that aspects of both models are true. A stable core of long-lasting cisternae might exist in the center of each Golgi cisterna, while regions at the rim may undergo continuous maturation, perhaps utilizing Rab cascades that change their identity. As matured pieces of the cisternae are formed, they might break off and fuse with downstream cisternae by homotypic fusion mechanisms, taking large cargo molecules with them. In addition, small COPI-coated vesicles might transport small cargo in the forward direction and retrieve escaped Golgi enzymes and return them to their appropriate upstream cisternae.

### Golgi Matrix Proteins Help Organize the Stack

The unique architecture of the Golgi apparatus depends on both the microtubule cytoskeleton, as already mentioned, and cytoplasmic Golgi matrix proteins, which form a scaffold between adjacent cisternae and give the Golgi stack its structural integrity. Some of the matrix proteins, called *golgins*, form long tethers composed of stiff coiled-coil domains with interspersed hinge regions. Golgins form a forest of tentacles that can extend 100–400 nm from the surface of the Golgi stack. They are thought to help retain Golgi transport vesicles close to the organelle through interactions with Rab proteins (Figure 13–36). When the cell prepares to divide, mitotic protein kinases phosphorylate the Golgi matrix proteins, causing the Golgi



**Figure 13–36** A model of golgin function.

Filamentous golgins anchored to Golgi membranes capture transport vesicles by binding to Rab proteins on the vesicle surface.

apparatus to fragment and disperse throughout the cytosol. The Golgi fragments are then distributed evenly to the two daughter cells, where the matrix proteins are dephosphorylated, leading to the reassembly of the Golgi stack. Similarly, during apoptosis, proteolytic cleavage of golgins by caspases ensues (discussed in Chapter 18), fragments the Golgi apparatus as the cell self-destructs.

## Summary

*Correctly folded and assembled proteins in the ER are packaged into COPII-coated transport vesicles that pinch off from the ER membrane. Shortly thereafter, the vesicles shed their coat and fuse with one another to form vesicular tubular clusters. In animal cells, the clusters then move on microtubule tracks to the Golgi apparatus, where they fuse with one another to form the cis Golgi network. Any resident ER proteins that escape from the ER are returned there from the vesicular tubular clusters and Golgi apparatus by retrograde transport in COPI-coated vesicles.*

The Golgi apparatus, unlike the ER, contains many sugar nucleotides, which glycosyl transferase enzymes use to glycosylate lipid and protein molecules as they pass through the Golgi apparatus. The mannoses on the N-linked oligosaccharides that are added to proteins in the ER are often initially removed, and further sugars are added. Moreover, the Golgi apparatus is the site where O-linked glycosylation occurs and where glycosaminoglycan chains are added to core proteins to form proteoglycans. Sulfation of the sugars in proteoglycans and of selected tyrosines on proteins also occurs in a late Golgi compartment.

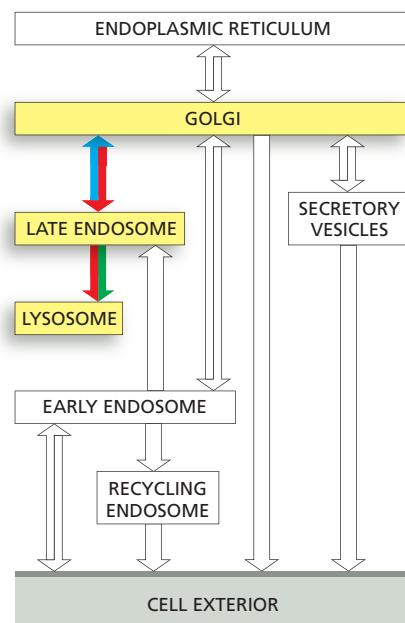
The Golgi apparatus modifies the many proteins and lipids that it receives from the ER and then distributes them to the plasma membrane, lysosomes, and secretory vesicles. The Golgi apparatus is a polarized organelle, consisting of one or more stacks of disc-shaped cisternae. Each stack is organized as a series of at least three functionally distinct compartments, termed *cis*, *medial*, and *trans* cisternae. The *cis* and *trans* cisternae are each connected to special sorting stations, called the *cis* Golgi network and the *trans* Golgi network, respectively. Proteins and lipids move through the Golgi stack in the *cis*-to-*trans* direction. This movement may occur by vesicle transport, by progressive maturation of the *cis* cisternae as they migrate continuously through the stack, or, most likely, by a combination of these two mechanisms. Continual retrograde vesicle transport from upstream to more downstream cisternae is thought to keep the enzymes concentrated in the cisternae where they are needed. The finished new proteins end up in the *trans* Golgi network, which packages them in transport vesicles and dispatches them to their specific destinations in the cell.

## TRANSPORT FROM THE TRANS GOLGI NETWORK TO LYSOSOMES

The *trans* Golgi network sorts all of the proteins that pass through the Golgi apparatus (except those that are retained there as permanent residents) according to their final destination. The sorting mechanism is especially well understood for those proteins destined for the lumen of lysosomes, and in this section we consider this selective transport process. We begin with a brief account of lysosome structure and function.

### Lysosomes Are the Principal Sites of Intracellular Digestion

Lysosomes are membrane-enclosed organelles filled with soluble hydrolytic enzymes that digest macromolecules. Lysosomes contain about 40 types of hydrolytic enzymes, including proteases, nucleases, glycosidases, lipases, phospholipases, phosphatases, and sulfatases. All are **acid hydrolases**; that is, hydrolases that work best at acidic pH. For optimal activity, they need to be activated by proteolytic cleavage, which also requires an acid environment. The lysosome provides this acidity, maintaining an interior pH of about 4.5–5.0. By this arrangement, the contents of the cytosol are doubly protected against attack by the cell's



own digestive system: the membrane of the lysosome keeps the digestive enzymes out of the cytosol, but, even if they leak out, they can do little damage at the cytosolic pH of about 7.2.

Like all other membrane-enclosed organelles, the lysosome not only contains a unique collection of enzymes, but also has a unique surrounding membrane. Most of the lysosome membrane proteins, for example, are highly glycosylated, which helps to protect them from the lysosome proteases in the lumen. Transport proteins in the lysosome membrane carry the final products of the digestion of macromolecules—such as amino acids, sugars, and nucleotides—to the cytosol, where the cell can either reuse or excrete them.

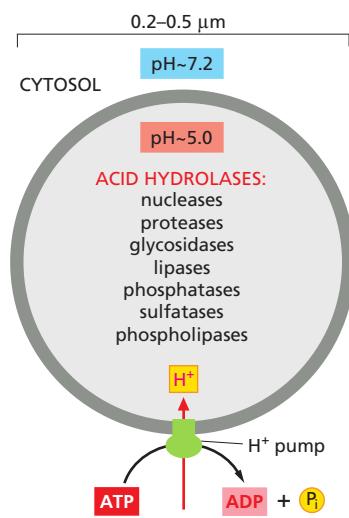
A *vacuolar H<sup>+</sup> ATPase* in the lysosome membrane uses the energy of ATP hydrolysis to pump H<sup>+</sup> into the lysosome, thereby maintaining the lumen at its acidic pH (Figure 13–37). The lysosome H<sup>+</sup> pump belongs to the family of V-type ATPases and has a similar architecture to the mitochondrial and chloroplast ATP synthases (F-type ATPases), which convert the energy stored in H<sup>+</sup> gradients into ATP (see Figure 11–12). By contrast to these enzymes, however, the vacuolar H<sup>+</sup> ATPase exclusively works in reverse, pumping H<sup>+</sup> into the organelle. Similar or identical V-type ATPases acidify all endocytic and exocytic organelles, including lysosomes, endosomes, some compartments of the Golgi apparatus, and many transport and secretory vesicles. In addition to providing a low-pH environment that is suitable for reactions occurring in the organelle lumen, the H<sup>+</sup> gradient provides a source of energy that drives the transport of small metabolites across the organelle membrane.

### Lysosomes Are Heterogeneous

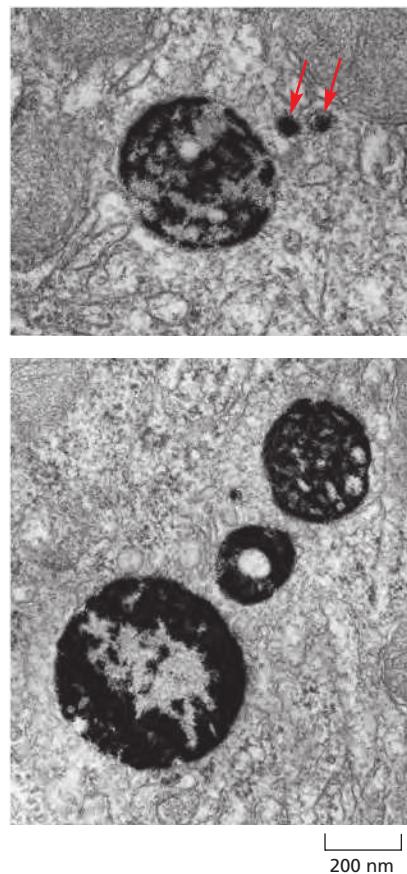
Lysosomes are found in all eukaryotic cells. They were initially discovered by the biochemical fractionation of cell extracts; only later were they seen clearly in the electron microscope. Although extraordinarily diverse in shape and size, staining them with specific antibodies shows they are members of a single family of organelles. They can also be identified by histochemical techniques that reveal which organelles contain acid hydrolase (Figure 13–38).

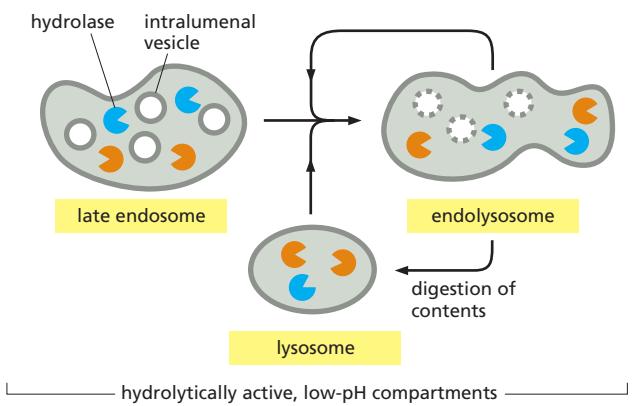
The heterogeneous morphology of lysosomes contrasts with the relatively uniform structures of many other cell organelles. The diversity reflects the wide variety of digestive functions that acid hydrolases mediate, including the breakdown of intra- and extracellular debris, the destruction of phagocytosed microorganisms, and the production of nutrients for the cell. Their morphological diversity, however, also reflects the way lysosomes form. Late endosomes containing material received from both the plasma membrane by endocytosis and newly synthesized lysosomal hydrolases fuse with preexisting lysosomes to form structures that are sometimes referred to as *endolysosomes*, which then fuse with one another (Figure 13–39). When the majority of the endocytosed material within an endolysosome has been digested so that only resistant or slowly digestible residues remain, these organelles become “classical” lysosomes. These are relatively dense, round, and small, but they can enter the cycle again by fusing with late endosomes or endolysosomes. Thus, there is no real distinction between endolysosomes and lysosomes: they are the same except that they are in different stages of a maturation cycle. For this reason, lysosomes are sometimes viewed as a heterogeneous collection of distinct organelles, the common feature of which is a high content of

**Figure 13–38 Histochemical visualization of lysosomes.** These electron micrographs show two sections of a cell stained to reveal the location of acid phosphatase, a marker enzyme for lysosomes. The larger membrane-enclosed organelles, containing dense precipitates of lead phosphate, are lysosomes. Their diverse morphology reflects variations in the amount and nature of the material they are digesting. The precipitates are produced when tissue fixed with glutaraldehyde (to fix the enzyme in place) is incubated with a phosphatase substrate in the presence of lead ions. Red arrows in the top panel indicate two small vesicles thought to be carrying acid hydrolases from the Golgi apparatus. (Courtesy of Daniel S. Friend.)



**Figure 13–37 Lysosomes.** The acid hydrolases are hydrolytic enzymes that are active under acidic conditions. An H<sup>+</sup> ATPase in the membrane pumps H<sup>+</sup> into the lysosome, maintaining its lumen at an acidic pH.





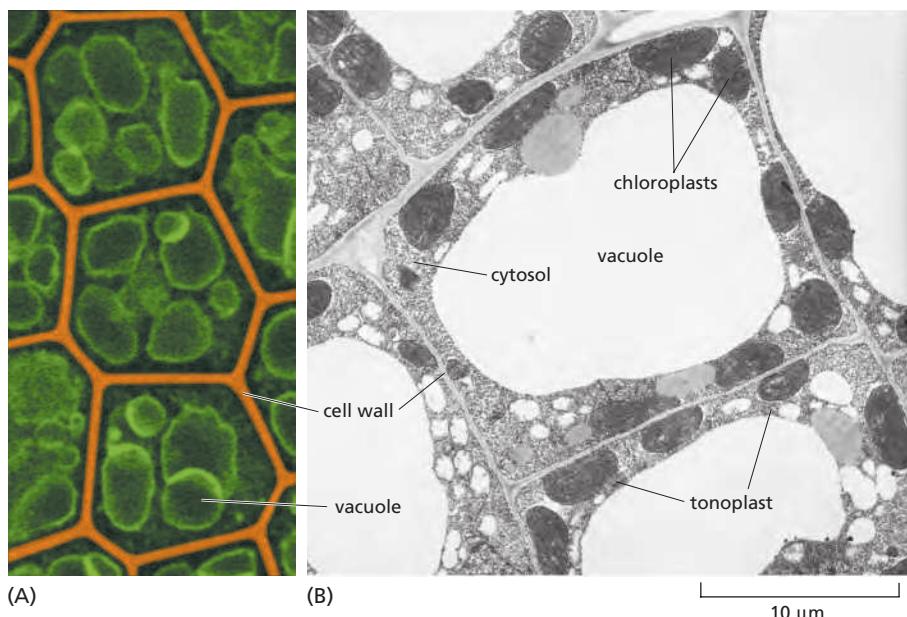
**Figure 13–39 A model for lysosome maturation.** Late endosomes fuse with preexisting lysosomes (bottom arrow) or preexisting endolysosomes (top arrow). Endolysosomes eventually mature into lysosomes as hydrolases complete the digestion of their contents, which can include intraluminal vesicles. Lysosomes also fuse with phagosomes, as we discuss later.

hydrolytic enzymes. It is especially hard to apply a narrower definition than this in plant cells, as we discuss next.

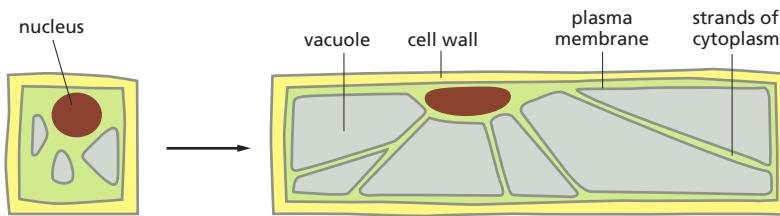
### Plant and Fungal Vacuoles Are Remarkably Versatile Lysosomes

Most plant and fungal cells (including yeasts) contain one or several very large, fluid-filled vesicles called **vacuoles**. They typically occupy more than 30% of the cell volume, and as much as 90% in some cell types (Figure 13–40). Vacuoles are related to animal cell lysosomes and contain a variety of hydrolytic enzymes, but their functions are remarkably diverse. The plant vacuole can act as a storage organelle for both nutrients and waste products, as a degradative compartment, as an economical way of increasing cell size, and as a controller of *turgor pressure* (the osmotic pressure that pushes outward on the cell wall and keeps the plant from wilting) (Figure 13–41). The same cell may have different vacuoles with distinct functions, such as digestion and storage.

The vacuole is important as a homeostatic device, enabling plant cells to withstand wide variations in their environment. When the pH in the environment drops, for example, the flux of H<sup>+</sup> into the cytosol is balanced, at least in part, by an increased transport of H<sup>+</sup> into the vacuole, which tends to keep the pH in the cytosol constant. Similarly, many plant cells maintain an almost constant turgor pressure despite large changes in the tonicity of the fluid in their immediate environment. They do so by changing the osmotic pressure of the cytosol and vacuole—in part by the controlled breakdown and resynthesis of polymers such as polyphosphate in the vacuole, and in part by altering the transport rates of sugars,



**Figure 13–40 The plant cell vacuole.** (A) A confocal image of cells from an *Arabidopsis* embryo that is expressing an aquaporin–YFP (yellow fluorescent protein) fusion protein in its tonoplast, or vacuole membrane (green); the cell walls have been false-colored orange. Each cell contains several large vacuoles. (B) This electron micrograph of cells in a young tobacco leaf shows the cytosol as a thin layer, containing chloroplasts, pressed against the cell wall by the enormous vacuole. (A, courtesy of C. Carroll and L. Frigerio, based on S. Gattolin et al., *Mol. Plant* 4:180–189, 2011. With permission from Oxford University Press; B, courtesy of J. Burgess.)



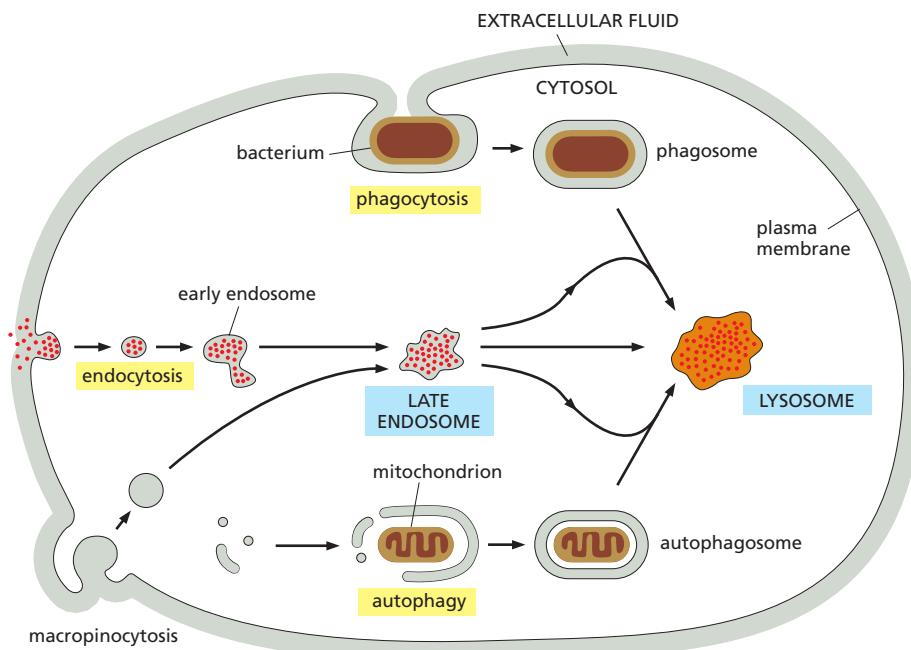
amino acids, and other metabolites across the plasma membrane and the vacuolar membrane. The turgor pressure regulates the activities of distinct transporters in each membrane to control these fluxes.

Humans often harvest substances stored in plant vacuoles—from rubber to opium to the flavoring of garlic. Many stored products have a metabolic function. Proteins, for example, can be preserved for years in the vacuoles of the storage cells of many seeds, such as those of peas and beans. When the seeds germinate, these proteins are hydrolyzed, and the resulting amino acids provide a food supply for the developing embryo. Anthocyanin pigments stored in vacuoles color the petals of many flowers so as to attract pollinating insects, while noxious molecules released from vacuoles when a plant is eaten or damaged provide a defense against predators.

### Multiple Pathways Deliver Materials to Lysosomes

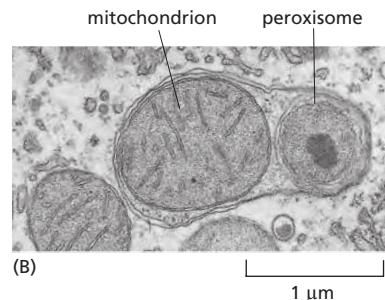
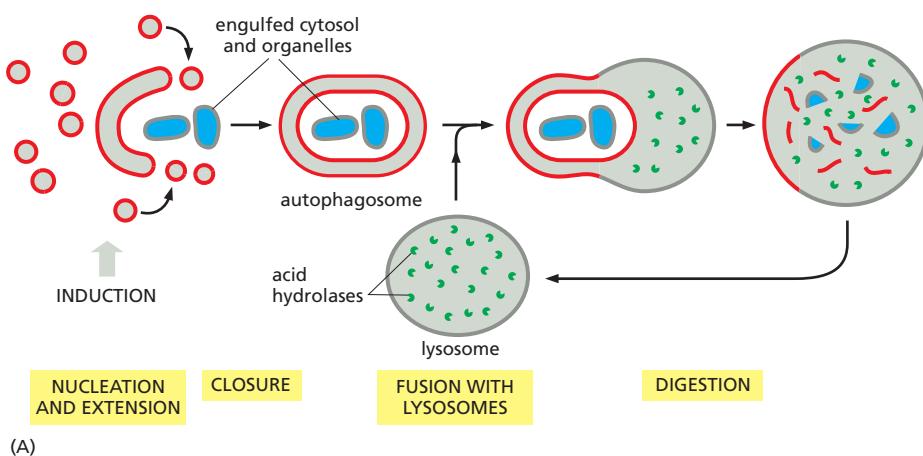
Lysosomes are meeting places where several streams of intracellular traffic converge. A route that leads outward from the ER via the Golgi apparatus delivers most of the lysosome's digestive enzymes, while at least four paths from different sources feed substances into lysosomes for digestion.

The best studied of these degradation pathways is the one followed by macromolecules taken up from extracellular fluid by *endocytosis*. A similar pathway found in phagocytic cells, such as macrophages and neutrophils in vertebrates, is dedicated to the engulfment, or *phagocytosis*, of large particles and microorganisms to form *phagosomes*. A third pathway called *macropinocytosis* specializes in the nonspecific uptake of fluids, membrane, and particles attached to the plasma membrane. We will return to discuss these pathways later in the chapter. A fourth pathway called *autophagy* originates in the cytoplasm of the cell itself and is used to digest cytosol and worn-out organelles, as we discuss next. The four paths to degradation in lysosomes are illustrated in Figure 13–42.



**Figure 13–41** The role of the vacuole in controlling the size of plant cells. A plant cell can achieve a large increase in volume without increasing the volume of the cytosol. Localized weakening of the cell wall orients a turgor-driven cell enlargement that accompanies the uptake of water into an expanding vacuole. The cytosol is eventually confined to a thin peripheral layer, which is connected to the nuclear region by strands of cytosol stabilized by bundles of actin filaments (not shown).

**Figure 13–42** Four pathways to degradation in lysosomes. Materials in each pathway are derived from a different source. Note that the autophagosome has a double membrane. In all cases, the final step is the fusion with lysosomes.



**Figure 13–43** A model of autophagy.

(A) Activation of a signaling pathway initiates a nucleation event in the cytoplasm. A crescent of autophagosomal membrane grows by fusion of vesicles of unknown origin and eventually fuses to form a double-membrane-enclosed autophagosome, which sequesters a portion of the cytoplasm. The autophagosome then fuses with lysosomes containing acid hydrolases that digest its content. During the formation of the autophagosome membrane, a ubiquitin-like protein becomes activated by covalent attachment of a phosphatidylethanolamine lipid anchor. These proteins then mediate vesicle tethering and fusion, leading to the formation of a crescent-shaped membrane structure that assembles around its target (not shown). (B) An electron micrograph of an autophagosome containing a mitochondrion and a peroxisome. (B, courtesy of Daniel S. Friend, from D.W. Fawcett, A Textbook of Histology, 12th ed. New York: Chapman and Hall, 1994. With permission from Kluwer.)

### Autophagy Degrades Unwanted Proteins and Organelles

All cell types dispose of obsolete parts by a lysosome-dependent process called **autophagy**, or “self-eating.” The degradation process is important during normal cell growth and in development, where it helps restructure differentiating cells, but also in adaptive responses to stresses such as starvation and infection. Autophagy can remove large objects—macromolecules, large protein aggregates, and even whole organelles—that other disposal mechanisms such as proteasomal degradation cannot handle. Defects in autophagy may prevent cells from clearing away invading microbes, unwanted protein aggregates and abnormal proteins, and thereby contribute to diseases ranging from infectious disorders to neurodegeneration and cancer.

In the initial stages of autophagy, cytoplasmic cargo becomes surrounded by a double membrane that assembles by the fusion of small vesicles of unknown origin, forming an **autophagosome** (Figure 13–43). A few tens of different proteins have been identified in yeast and animal cells that participate in the process, which must be tightly regulated: either too little or too much can be deleterious. The whole process occurs in the following sequence of steps:

1. Induction by activation of signaling molecules: Protein kinases (including the mTOR complex 1, discussed in Chapter 15) that relay information about the metabolic status of the cell, become activated and signal to the autophagic machinery.
2. Nucleation and extension of a delimiting membrane into a crescent-shaped cup: Membrane vesicles, characterized by the presence of ATG9, the only transmembrane protein involved in the process, are recruited to an assembly site, where they nucleate autophagosome formation. ATG9 is not incorporated into the autophagosome: a retrieval pathway must remove it from the assembling structure.
3. Closure of the membrane cup around the target to form a sealed double-membrane-enclosed autophagosome.
4. Fusion of the autophagosome with lysosomes, catalyzed by SNAREs.
5. Digestion of the inner membrane and the luminal contents of the autophagosome.

Autophagy can be either nonselective or selective. In **nonselective autophagy**, a bulk portion of cytoplasm is sequestered in autophagosomes. It might occur, for example, in starvation conditions: when external nutrients are limiting, metabolites derived from the digestion of the captured cytosol might help the cell survive. In **selective autophagy** specific cargo is packaged into autophagosomes that tend to contain little cytosol, and their shape reflects the shape of the cargo. Selective autophagy mediates the degradation of worn out, or otherwise unwanted, mitochondria, peroxisomes, ribosomes, and ER; it can also be used to destroy invading microbes.

The selective autophagy of worn out or damaged mitochondria is called *mitophagy*. As discussed in Chapters 12 and 14, when mitochondria function normally, the inner mitochondrial membrane is energized by an electrochemical H<sup>+</sup> gradient that drives ATP synthesis and the import of mitochondrial precursor proteins and metabolites. Damaged mitochondria cannot maintain the gradient, so protein import is blocked. As a consequence, a protein kinase called Pink1, which is normally imported into mitochondria, is instead retained on the mitochondrial surface where it recruits the ubiquitin ligase Parkin from the cytosol. Parkin ubiquitylates mitochondrial outer membrane proteins, which mark the organelle for selective destruction in autophagosomes. Mutations in Pink1 or Parkin cause a form of early-onset Parkinson's disease, a degenerative disorder of the central nervous system. It is not known why the neurons that die prematurely in this disease are particularly reliant on mitophagy.

### A Mannose 6-Phosphate Receptor Sorts Lysosomal Hydrolases in the *Trans* Golgi Network

We now consider the pathway that delivers lysosomal hydrolases from the TGN to lysosomes. The enzymes are first delivered to endosomes in transport vesicles that bud from the TGN, before they move on to endolysosomes and lysosomes (see Figure 13–39). The vesicles that leave the TGN incorporate the lysosomal proteins and exclude the many other proteins being packaged into different transport vesicles for delivery elsewhere.

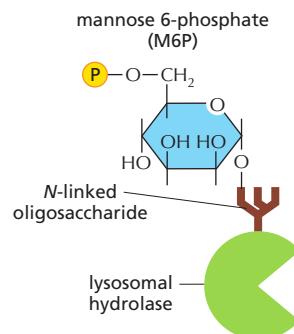
How are lysosomal hydrolases recognized and selected in the TGN with the required accuracy? In animal cells they carry a unique marker in the form of *mannose 6-phosphate* (M6P) groups, which are added exclusively to the N-linked oligosaccharides of these soluble lysosomal enzymes as they pass through the lumen of the *cis* Golgi network (Figure 13–44). Transmembrane **M6P receptor proteins**, which are present in the TGN, recognize the M6P groups and bind to the lysosomal hydrolases on the luminal side of the membrane and to adaptor proteins in assembling clathrin coats on the cytosolic side. In this way, the receptors help package the hydrolases into clathrin-coated vesicles that bud from the TGN and deliver their contents to early endosomes.

The M6P receptor protein binds to M6P at pH 6.5–6.7 in the TGN lumen and releases it at pH 6, which is the pH in the lumen of endosomes. Thus, after the receptor is delivered, the lysosomal hydrolases dissociate from the M6P receptors, which are retrieved into transport vesicles that bud from endosomes. These vesicles are coated with *retromer*, a coat protein complex specialized for endosome-to-TGN transport, which returns the receptors to the TGN for reuse (Figure 13–45).

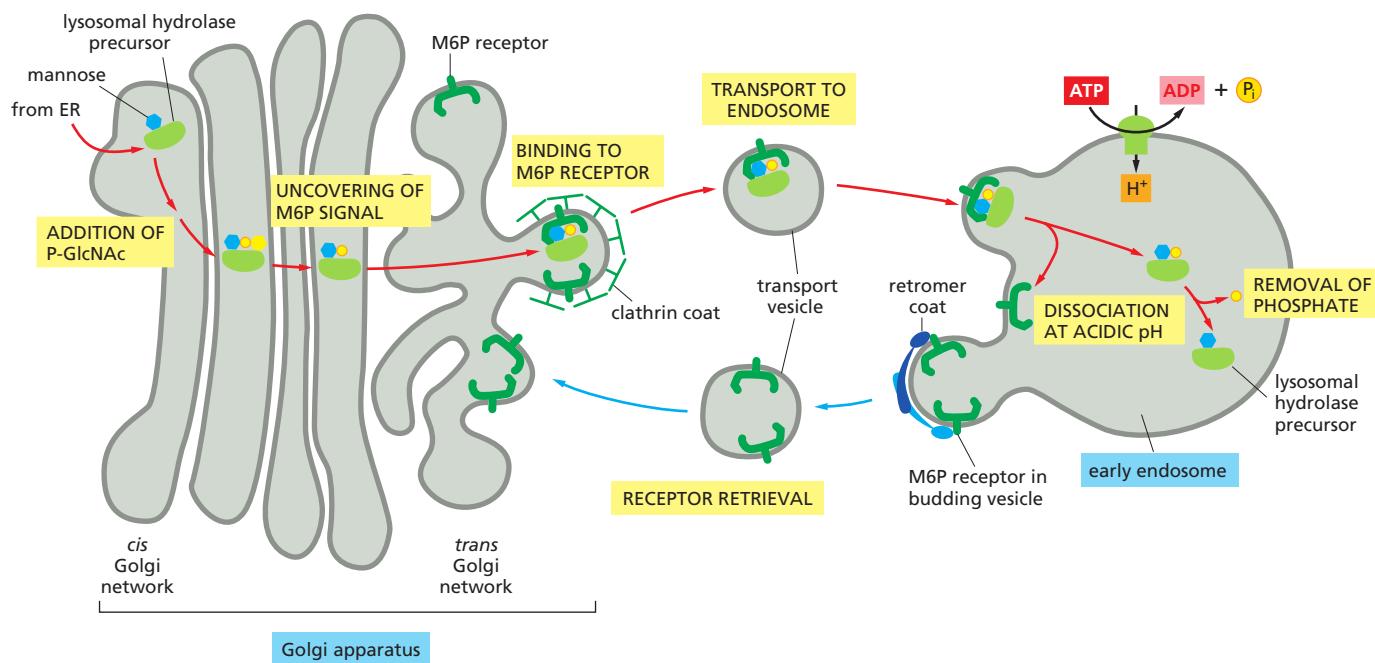
Transport in either direction requires signals in the cytoplasmic tail of the M6P receptor that direct this protein to the endosome or back to the TGN. These signals are recognized by the retromer complex that recruits M6P receptors into transport vesicles that bud from endosomes. The recycling of the M6P receptor resembles the recycling of the KDEL receptor discussed earlier, although it differs in the type of coated vesicles that mediate the transport.

Not all the hydrolase molecules that are tagged with M6P get to lysosomes. Some escape the normal packaging process in the *trans* Golgi network and are transported “by default” to the cell surface, where they are secreted into the extracellular fluid. Some M6P receptors, however, also take a detour to the plasma membrane, where they recapture the escaped lysosomal hydrolases and return them by *receptor-mediated endocytosis* (discussed later) to lysosomes via early and late endosomes. As lysosomal hydrolases require an acidic milieu to work, they can do little harm in the extracellular fluid, which usually has a neutral pH of 7.4.

For the sorting system that segregates lysosomal hydrolases and dispatches them to endosomes to work, the M6P groups must be added only to the appropriate glycoproteins in the Golgi apparatus. This requires specific recognition of the hydrolases by the Golgi enzymes responsible for adding M6P. Since all glycoproteins leave the ER with identical N-linked oligosaccharide chains, the signal for



**Figure 13–44** The structure of mannose 6-phosphate on a lysosomal hydrolase.



adding the M6P units to oligosaccharides must reside somewhere in the polypeptide chain of each hydrolase. Genetic engineering experiments have revealed that the recognition signal is a cluster of neighboring amino acids on each protein's surface, known as a *signal patch* (Figure 13–46). Since most lysosomal hydrolases contain multiple oligosaccharides, they acquire many M6P groups, providing a high-affinity signal for the M6P receptor.

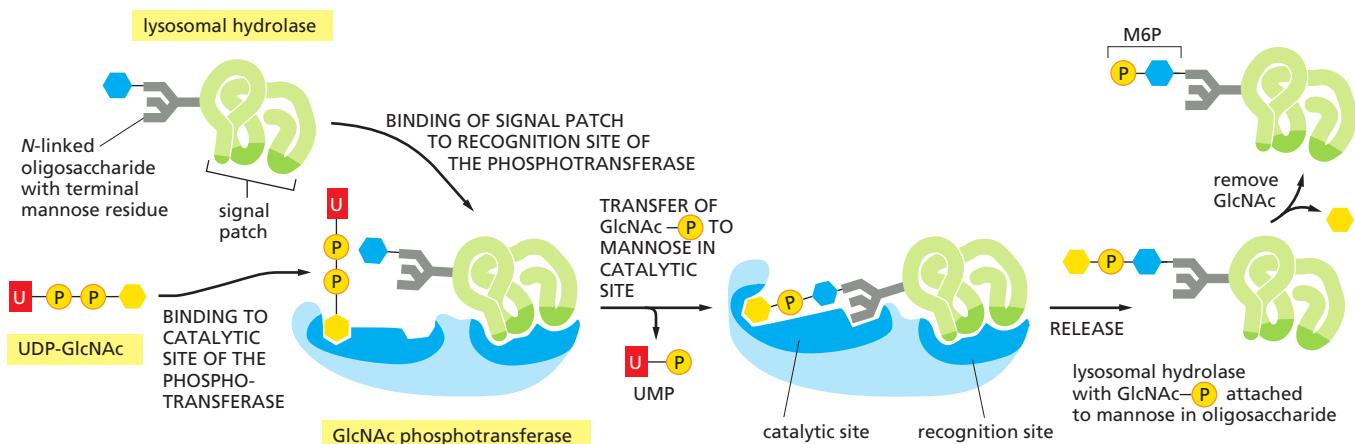
### Defects in the GlcNAc Phosphotransferase Cause a Lysosomal Storage Disease in Humans

Genetic defects that affect one or more of the lysosomal hydrolases cause a number of human **lysosomal storage diseases**. The defects result in an accumulation of undigested substrates in lysosomes, with severe pathological consequences, most often in the nervous system. In most cases, there is a mutation in a structural gene that codes for an individual lysosomal hydrolase. This occurs in *Hurler's disease*, for example, in which the enzyme required for the breakdown of certain types of glycosaminoglycan chains is defective or missing. The most severe form of lysosomal storage disease, however, is a very rare inherited metabolic disorder called *inclusion-cell disease* (*I-cell disease*). In this condition, almost all of the hydrolytic enzymes are missing from the lysosomes of many cell types, and their undigested substrates accumulate in these lysosomes, which consequently form large *inclusions* in the cells. The consequent pathology is complex, affecting all organ systems, skeletal integrity, and mental development; individuals rarely live beyond six or seven years.

*I-cell disease* is due to a single gene defect and, like most genetic enzyme deficiencies, it is recessive—that is, it occurs only in individuals having two copies of the defective gene. In patients with *I-cell disease*, all the hydrolases missing from lysosomes are found in the blood: because they fail to sort properly in the Golgi apparatus, they are secreted rather than transported to lysosomes. The mis-sorting has been traced to a defective or missing GlcNAc phosphotransferase. Because lysosomal enzymes are not phosphorylated in the *cis* Golgi network, the M6P receptors do not segregate them into the appropriate transport vesicles in the TGN. Instead, the lysosomal hydrolases are carried to the cell surface and secreted.

In *I-cell disease*, the lysosomes in some cell types, such as hepatocytes, contain a normal complement of lysosomal enzymes, implying that there is another

**Figure 13–45** The transport of newly synthesized lysosomal hydrolases to endosomes. The sequential action of two enzymes in the *cis* and *trans* Golgi network adds mannose 6-phosphate (M6P) groups to the precursors of lysosomal enzymes (see Figure 13–46). The M6P-tagged hydrolases then segregate from all other types of proteins in the TGN because adaptor proteins (not shown) in the clathrin coat bind the M6P receptors, which, in turn, bind the M6P-modified lysosomal hydrolases. The clathrin-coated vesicles bud off from the TGN, shed their coat, and fuse with early endosomes. At the lower pH of the endosome, the hydrolases dissociate from the M6P receptors, and the empty receptors are retrieved in retromer-coated vesicles to the TGN for further rounds of transport. In the endosomes, the phosphate is removed from the M6P attached to the hydrolases, which may further ensure that the hydrolases do not return to the TGN with the receptor.



pathway for directing hydrolases to lysosomes that is used by some cell types but not others. Alternative sorting receptors function in these M6P-independent pathways. Similarly, an M6P-independent pathway in all cells sorts the membrane proteins of lysosomes from the TGN for transport to late endosomes, and those proteins are therefore normal in I-cell disease.

### Some Lysosomes and Multivesicular Bodies Undergo Exocytosis

Targeting of material to lysosomes is not necessarily the end of the pathway. *Lysosomal secretion* of undigested content enables all cells to eliminate indigestible debris. For most cells, this seems to be a minor pathway, used only when the cells are stressed. Some cell types, however, contain specialized lysosomes that have acquired the necessary machinery for fusion with the plasma membrane. *Melanocytes* in the skin, for example, produce and store pigments in their lysosomes. These pigment-containing *melanosomes* release their pigment into the extracellular space of the epidermis by exocytosis. The pigment is then taken up by keratinocytes, leading to normal skin pigmentation. In some genetic disorders, defects in melanosome exocytosis block this transfer process, leading to forms of hypopigmentation (albinism). Under certain conditions, multivesicular bodies can also fuse with the plasma membrane. If that occurs, their intraluminal vesicles are released from cells. Circulating small vesicles, also called *exosomes*, have been observed in the blood and may be used to transport components between cells, although the importance of such a mechanism of potential communication between distant cells is unknown. Some exosomes may derive from direct vesicle budding events at the plasma membrane, which is a topologically equivalent process (see Figure 13-57).

### Summary

*Lysosomes are specialized for the intracellular digestion of macromolecules. They contain unique membrane proteins and a wide variety of soluble hydrolytic enzymes that operate best at pH 5, which is the internal pH of lysosomes. An ATP-driven H<sup>+</sup> pump in the lysosomal membrane maintains this low pH. Newly synthesized lysosomal proteins transported from the lumen of the ER, through the Golgi apparatus; they are then carried from the trans Golgi network to endosomes by means of clathrin-coated transport vesicles, before moving on to lysosomes.*

*The lysosomal hydrolases contain N-linked oligosaccharides that are covalently modified in a unique way in the cis Golgi so that their mannoses are phosphorylated. These mannose 6-phosphate (M6P) groups are recognized by an M6P receptor protein in the trans Golgi network that segregates the hydrolases and helps package them into budding transport vesicles that deliver their contents to endosomes. The M6P receptors shuttle back and forth between the trans Golgi network and the endosomes. The low pH in endosomes and the removal of the phosphate from the*

**Figure 13-46** The recognition of a lysosomal hydrolase. A GlcNAc phosphotransferase recognizes lysosomal hydrolases in the Golgi apparatus. The enzyme has separate catalytic and recognition sites. The catalytic site binds both high-mannose N-linked oligosaccharides and UDP-GlcNAc. The recognition site binds to a signal patch that is present only on the surface of lysosomal hydrolases. A second enzyme cleaves off the GlcNAc, leaving the mannose 6-phosphate exposed.

*M6P* group cause the lysosomal hydrolases to dissociate from these receptors, making the transport of the hydrolases unidirectional. A separate transport system uses clathrin-coated vesicles to deliver resident lysosomal membrane proteins from the trans Golgi network to endosomes.

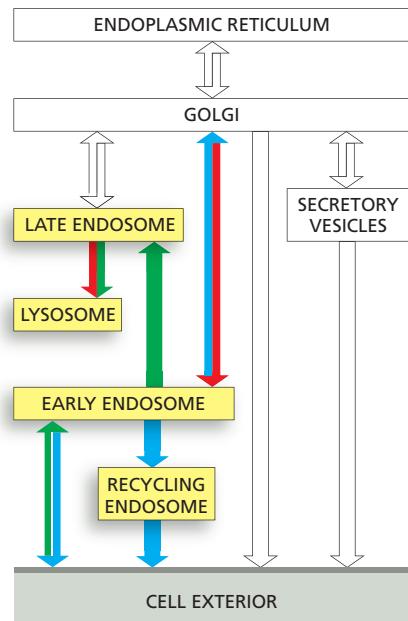
## TRANSPORT INTO THE CELL FROM THE PLASMA MEMBRANE: ENDOCYTOSIS

The routes that lead inward from the cell surface start with the process of **endocytosis**, by which cells take up plasma membrane components, fluid, solutes, macromolecules, and particulate substances. Endocytosed cargo includes receptor-ligand complexes, a spectrum of nutrients and their carriers, extracellular matrix components, cell debris, bacteria, viruses, and, in specialized cases, even other cells. Through endocytosis, the cell regulates the composition of its plasma membrane in response to changing extracellular conditions.

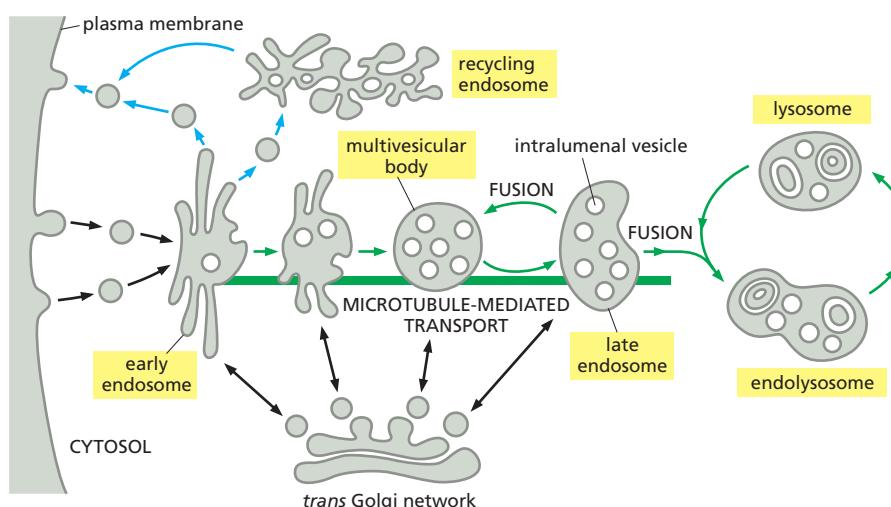
In endocytosis, the material to be ingested is progressively enclosed by a small portion of the plasma membrane, which first invaginates and then pinches off to form an **endocytic vesicle** containing the ingested substance or particle. Most eukaryotic cells constantly form endocytic vesicles, a process called *pinocytosis* ("cell drinking"); in addition, some specialized cells contain dedicated pathways that take up large particles on demand, a process called *phagocytosis* ("cell eating"). Endocytic vesicles form at the plasma membrane by multiple mechanisms that differ in both the molecular machinery used and how that machinery is regulated.

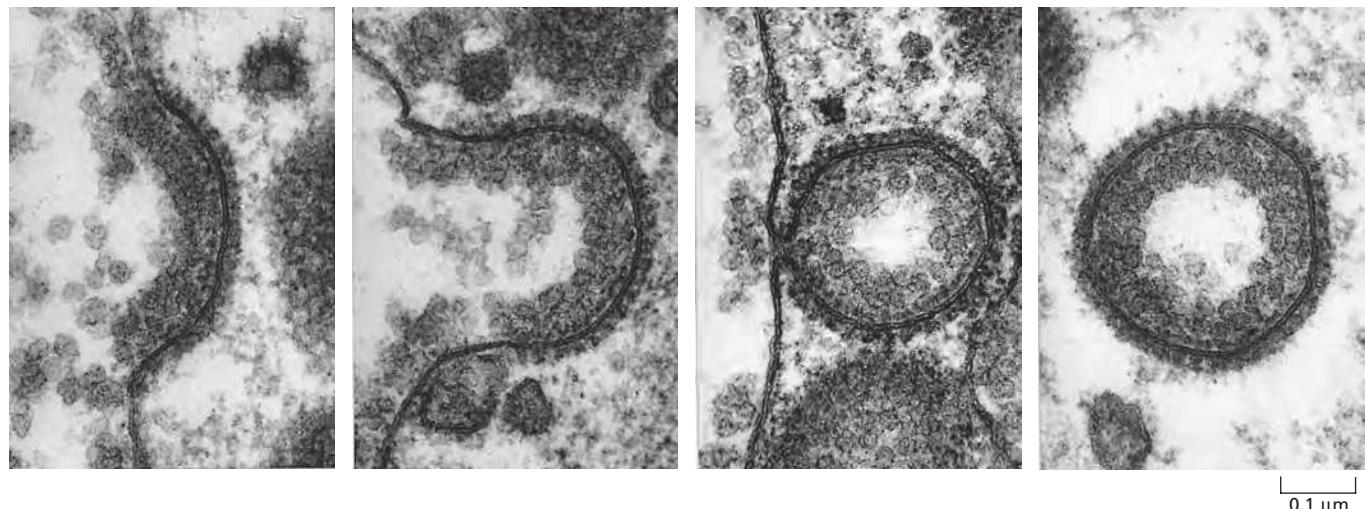
Once generated at the plasma membrane, most endocytic vesicles fuse with a common receiving compartment, the *early endosome*, where internalized cargo is sorted: some cargo molecules are returned to the plasma membrane, either directly or via a *recycling endosome*, and others are designated for degradation by inclusion in a *late endosome*. Late endosomes form from a bulbous, vacuolar portion of early endosomes by a process called *endosome maturation*. This conversion process changes the protein composition of the endosome membrane, patches of which invaginate and become incorporated within the organelles as *intraluminal vesicles*, while the endosome itself moves from the cell periphery to a location close to the nucleus. As an endosome matures, it ceases to recycle material to the plasma membrane and irreversibly commits its remaining contents to degradation: late endosomes fuse with one another and with lysosomes to form endolysosomes, which degrade their contents, as discussed earlier (Figure 13–47).

Each of the stages of endosome maturation—from the early endosome to the endolysosome—is connected through bidirectional vesicle transport pathways to



**Figure 13–47** Endosome maturation: the endocytic pathway from the plasma membrane to lysosomes. Endocytic vesicles fuse near the cell periphery with an early endosome, which is the primary sorting station. Tubular portions of the early endosome bud off vesicles that recycle endocytosed cargo back to the plasma membrane—either directly, or indirectly via recycling endosomes. Recycling endosomes can store proteins until they are needed. Conversion of early endosomes to late endosomes is accompanied by loss of the tubular projections. Membrane proteins destined for degradation are internalized in intraluminal vesicles. The developing late endosome, or multivesicular body, moves on microtubules to the cell interior. Fully matured late endosomes no longer send vesicles to the plasma membrane, and they fuse with one another and with endolysosomes and lysosomes to degrade their contents. Each stage of endosome maturation is connected via transport vesicles with the TGN, providing a continuous supply of newly synthesized lysosomal proteins.





the TGN. These pathways allow insertion of newly synthesized materials, such as lysosomal enzymes arriving from the ER, and the retrieval of components, such as the M6P receptor, back into the early parts of the secretory pathway. We next discuss how the cell uses and controls the various features of endocytic trafficking.

### Pinocytic Vesicles Form from Coated Pits in the Plasma Membrane

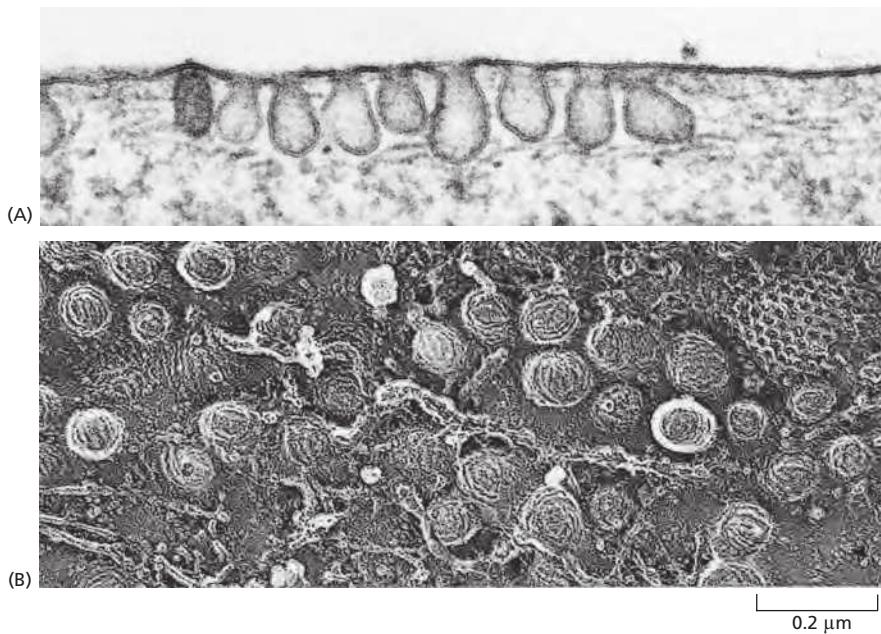
Virtually all eukaryotic cells continually ingest portions of their plasma membrane in the form of small pinocytic (endocytic) vesicles. The rate at which plasma membrane is internalized in this process of **pinocytosis** varies between cell types, but it is usually surprisingly high. A macrophage, for example, ingests 25% of its own volume of fluid each hour. This means that it must ingest 3% of its plasma membrane each minute, or 100% in about half an hour. Fibroblasts endocytose at a somewhat lower rate (1% of their plasma membrane per minute), whereas some amoebae ingest their plasma membrane even more rapidly. Since a cell's surface area and volume remain unchanged during this process, it is clear that the same amount of membrane being removed by endocytosis is being added to the cell surface by the converse process of *exocytosis*. In this sense, endocytosis and exocytosis are linked processes that can be considered to constitute an *endocytic-exocytic cycle*. The coupling between exocytosis and endocytosis is particularly strict in specialized structures characterized by high membrane turnover, such as a nerve terminal.

The endocytic part of the cycle often begins at **clathrin-coated pits**. These specialized regions typically occupy about 2% of the total plasma membrane area. The lifetime of a clathrin-coated pit is short: within a minute or so of being formed, it invaginates into the cell and pinches off to form a clathrin-coated vesicle (**Figure 13-48**). About 2500 clathrin-coated vesicles pinch off from the plasma membrane of a cultured fibroblast every minute. The coated vesicles are even more transient than the coated pits: within seconds of being formed, they shed their coat and fuse with early endosomes.

### Not All Pinocytic Vesicles Are Clathrin-Coated

In addition to clathrin-coated pits and vesicles, cells can form other types of pinocytic vesicles, such as **caveolae** (from the Latin for “little cavities”), originally recognized by their ability to transport molecules across endothelial cells that form the inner lining of blood vessels. Caveolae, sometimes seen in the electron microscope as deeply invaginated flasks, are present in the plasma membrane of most vertebrate cell types (**Figure 13-49**). They are thought to form from *lipid rafts* in the plasma membrane (discussed in Chapter 10), which are especially rich

**Figure 13-48** The formation of clathrin-coated vesicles from the plasma membrane. These electron micrographs illustrate the probable sequence of events in the formation of a clathrin-coated vesicle from a clathrin-coated pit. The clathrin-coated pits and vesicles shown are larger than those seen in normal-sized cells; they are from a very large hen oocyte and they take up lipoprotein particles to form yolk. The lipoprotein particles bound to their membrane-bound receptors appear as a dense, fuzzy layer on the extracellular surface of the plasma membrane—which is the inside surface of the coated pit and vesicle. (Courtesy of M.M. Perry and A.B. Gilbert, *J. Cell Sci.* 39:257–272, 1979. With permission from The Company of Biologists.)



**Figure 13–49** Caveolae in the plasma membrane of a fibroblast. (A) This electron micrograph shows a plasma membrane with a very high density of caveolae. (B) This rapid-freeze deep-etch image demonstrates the characteristic “cauliflower” texture of the cytosolic face of the caveolae membrane. The characteristic texture is thought to result from aggregates of caveolins and cavins. A clathrin-coated pit is also seen at the upper right. (Courtesy of R.G.W. Anderson, from K.G. Rothberg et al., *Cell* 68:673–682, 1992. With permission from Elsevier.)

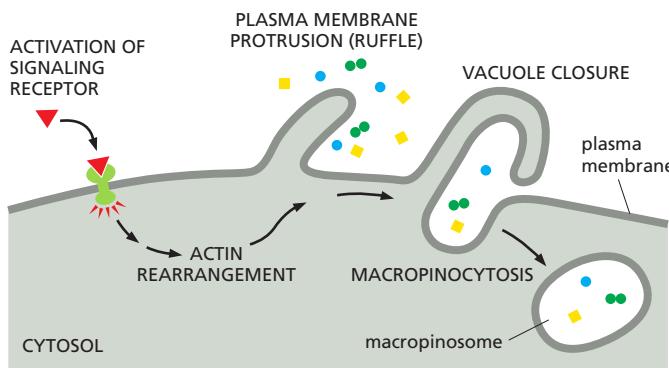
in cholesterol, glycosphingolipids, and glycosylphosphatidylinositol (GPI)-anchored membrane proteins (see Figure 10–13). The major structural proteins in caveolae are **caveolins**, a family of unusual integral membrane proteins that each insert a hydrophobic loop into the membrane from the cytosolic side but do not extend across the membrane. On their cytosolic side, caveolins are bound to large protein complexes of caving proteins, which are thought to stabilize the membrane curvature.

In contrast to clathrin-coated and COPI- or COPII-coated vesicles, caveolae are usually static structures. Nonetheless, they can be induced to pinch off and serve as endocytic vesicles to transport cargo to early endosomes or to the plasma membrane on the opposite side of a polarized cell (in a process called *transcytosis*, which we discuss later). Some animal viruses such as SV40 and papillomavirus (which causes warts) enter cells in vesicles derived from caveolae. The viruses are first delivered to early endosomes and move from there in transport vesicles to the lumen of the ER. The viral genome exits across the ER membrane into the cytosol, from where it is imported into the nucleus to start the infection cycle. Cholera toxin (discussed in Chapters 15 and 19) also enters the cell through caveolae and is transported to the ER before entering the cytosol.

**Macropinocytosis** is another clathrin-independent endocytic mechanism that can be activated in practically all animal cells. In most cell types, it does not operate continually but rather is induced for a limited time in response to cell-surface receptor activation by specific cargoes, including growth factors, integrin ligands, apoptotic cell remnants, and some viruses. These ligands activate a complex signaling pathway, resulting in a change in actin dynamics and the formation of cell-surface protrusions, called *ruffles* (discussed in Chapter 16). When ruffles collapse back onto the cell, large fluid-filled endocytic vesicles form, called *macropinosomes* (Figure 13–50), which transiently increase the bulk fluid uptake of a cell by up to tenfold. Macropinocytosis is a dedicated degradative pathway: macropinosomes acidify and then fuse with late endosomes or endolysosomes, without recycling their cargo back to the plasma membrane.

### Cells Use Receptor-Mediated Endocytosis to Import Selected Extracellular Macromolecules

In most animal cells, clathrin-coated pits and vesicles provide an efficient pathway for taking up specific macromolecules from the extracellular fluid. In this process, called **receptor-mediated endocytosis**, the macromolecules bind to



**Figure 13–50 Schematic representation of macropinocytosis.** Cell signaling events lead to a reprogramming of actin dynamics, which in turn triggers the formation of cell-surface ruffles. As the ruffles collapse back onto the cell surface, they nonspecifically trap extracellular fluid and macromolecules and particles contained in it, forming large vacuoles, or macropinosomes, as shown.

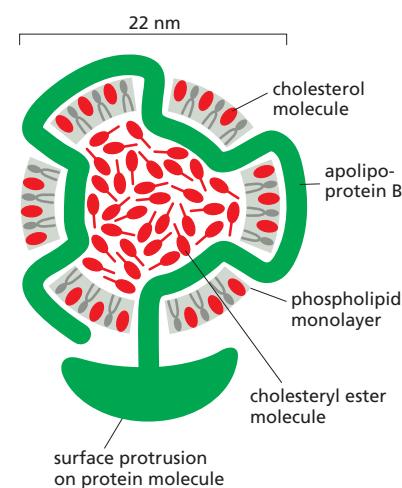
complementary transmembrane receptor proteins, which accumulate in coated pits, and then enter the cell as receptor-macromolecule complexes in clathrin-coated vesicles (see Figure 13–48). Because ligands are selectively captured by receptors, receptor-mediated endocytosis provides a selective concentrating mechanism that increases the efficiency of internalization of particular ligands more than a hundredfold. In this way, even minor components of the extracellular fluid can be efficiently taken up in large amounts. A particularly well-understood and physiologically important example is the process that mammalian cells use to import cholesterol.

Many animal cells take up cholesterol through receptor-mediated endocytosis and, in this way, acquire most of the cholesterol they require to make new membrane. If the uptake is blocked, cholesterol accumulates in the blood and can contribute to the formation in blood vessel (artery) walls of *atherosclerotic plaques*, deposits of lipid and fibrous tissue that can cause strokes and heart attacks by blocking arterial blood flow. In fact, it was a study of humans with a strong genetic predisposition for *atherosclerosis* that first revealed the mechanism of receptor-mediated endocytosis.

Most cholesterol is transported in the blood as cholesteryl esters in the form of lipid-protein particles known as **low-density lipoproteins (LDLs)** (Figure 13–51). When a cell needs cholesterol for membrane synthesis, it makes transmembrane receptor proteins for LDL and inserts them into its plasma membrane. Once in the plasma membrane, the *LDL receptors* diffuse until they associate with clathrin-coated pits that are in the process of forming. There, an endocytosis signal in the cytoplasmic tail of the LDL receptors binds the membrane-bound adaptor protein AP2 after its conformation has been locally unlocked by binding to PI(4,5)P<sub>2</sub> on the plasma membrane. Coincidence detection, as discussed earlier, thus imparts both efficiency and selectivity to the process (see Figure 13–9). AP2 then recruits clathrin to initiate endocytosis.

Since coated pits constantly pinch off to form coated vesicles, any LDL particles bound to LDL receptors in the coated pits are rapidly internalized in coated vesicles. After shedding their clathrin coats, the vesicles deliver their contents to early endosomes. Once the LDL and LDL receptors encounter the low pH in early endosomes, LDL is released from its receptor and is delivered via late endosomes to lysosomes. There, the cholesteryl esters in the LDL particles are hydrolyzed to free cholesterol, which is now available to the cell for new membrane synthesis (Movie 13.3). If too much free cholesterol accumulates in a cell, the cell shuts off both its own cholesterol synthesis and the synthesis of LDL receptors, so that it ceases both to make or to take up cholesterol.

This regulated pathway for cholesterol uptake is disrupted in individuals who inherit defective genes encoding LDL receptors. The resulting high levels of blood cholesterol predispose these individuals to develop atherosclerosis prematurely, and many would die at an early age of heart attacks resulting from coronary artery disease if they were not treated with drugs such as statins that lower the level of blood cholesterol. In some cases, the receptor is lacking altogether. In others, the receptors are defective—in either the extracellular binding site for LDL or the



**Figure 13–51 A low-density lipoprotein (LDL) particle.** Each spherical particle has a mass of  $3 \times 10^6$  daltons. It contains a core of about 1500 cholesterol molecules esterified to long-chain fatty acids. A lipid monolayer composed of about 800 phospholipid and 500 unesterified cholesterol molecules surrounds the core of cholesteryl esters. A single molecule of apolipoprotein B, a 500,000-dalton beltlike protein, organizes the particle and mediates the specific binding of LDL to cell-surface LDL receptors.

intracellular binding site that attaches the receptor AP2 adaptor protein in clathrin-coated pits. In the latter case, normal numbers of LDL receptors are present, but they fail to become localized in clathrin-coated pits. Although LDL binds to the surface of these mutant cells, it is not internalized, directly demonstrating the importance of clathrin-coated pits for the receptor-mediated endocytosis of cholesterol.

More than 25 distinct receptors are known to participate in receptor-mediated endocytosis of different types of molecules. They all apparently use clathrin-dependent internalization routes and are guided into clathrin-coated pits by signals in their cytoplasmic tails that bind to adaptor proteins in the clathrin coat. Many of these receptors, like the LDL receptor, enter coated pits irrespective of whether they have bound their specific ligands. Others enter preferentially when bound to a specific ligand, suggesting that a ligand-induced conformational change is required for them to activate the signal sequence that guides them into the pits. Since most plasma membrane proteins fail to become concentrated in clathrin-coated pits, the pits serve as molecular filters, preferentially collecting certain plasma membrane proteins (receptors) over others.

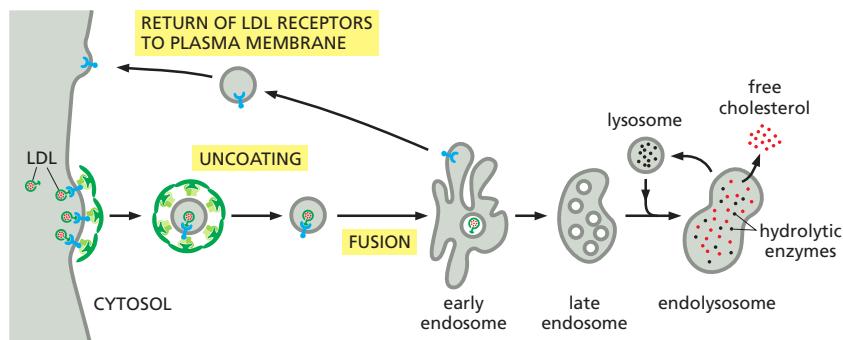
Electron-microscope studies of cultured cells exposed simultaneously to different labeled ligands demonstrate that many kinds of receptors can cluster in the same coated pit, whereas some other receptors cluster in different clathrin-coated pits. The plasma membrane of one clathrin-coated pit can accommodate more than 100 receptors of assorted varieties.

### Specific Proteins Are Retrieved from Early Endosomes and Returned to the Plasma Membrane

Early endosomes are the main sorting station in the endocytic pathway, just as the *cis* and *trans* Golgi networks serve this function in the secretory pathway. In the mildly acidic environment of the early endosome, many internalized receptor proteins change their conformation and release their ligand, as already discussed for the M6P receptors. Those endocytosed ligands that dissociate from their receptors in the early endosome are usually doomed to destruction in lysosomes (although cholesterol is an exception, as just discussed), along with the other soluble contents of the endosome. Some other endocytosed ligands, however, remain bound to their receptors, and thereby share the fate of the receptors.

In the early endosome, the LDL receptor dissociates from its ligand, LDL, and is recycled back to the plasma membrane for reuse, leaving the discharged LDL to be carried to lysosomes (Figure 13–52). The recycling transport vesicles bud from long, narrow tubules that extend from the early endosomes. It is likely that the geometry of these tubules helps the sorting process: because tubules have a large membrane area enclosing a small volume, membrane proteins become enriched over soluble proteins. The transport vesicles return the LDL receptor directly to the plasma membrane.

The **transferrin receptor** follows a similar recycling pathway as the LDL receptor, but unlike the LDL receptor it also recycles its ligand. Transferrin is a soluble



**Figure 13–52** The receptor-mediated endocytosis of LDL. Note that the LDL dissociates from its receptors in the acidic environment of the early endosome. After a number of steps, the LDL ends up in endolysosomes and lysosomes, where it is degraded to release free cholesterol. In contrast, the LDL receptors are returned to the plasma membrane via transport vesicles that bud off from the tubular region of the early endosome, as shown. For simplicity, only one LDL receptor is shown entering the cell and returning to the plasma membrane. Whether it is occupied or not, an LDL receptor typically makes one round trip into the cell and back to the plasma membrane every 10 minutes, making a total of several hundred trips in its 20-hour life-span.

protein that carries iron in the blood. Cell-surface transferrin receptors deliver transferrin with its bound iron to early endosomes by receptor-mediated endocytosis. The low pH in the endosome induces transferrin to release its bound iron, but the iron-free transferrin itself (called apotransferrin) remains bound to its receptor. The receptor-apotransferrin complex enters the tubular extensions of the early endosome and from there is recycled back to the plasma membrane. When the apotransferrin returns to the neutral pH of the extracellular fluid, it dissociates from the receptor and is thereby freed to pick up more iron and begin the cycle again. Thus, transferrin shuttles back and forth between the extracellular fluid and early endosomes, avoiding lysosomes and delivering iron to the cell interior, as needed for cells to grow and proliferate.

### Plasma Membrane Signaling Receptors are Down-Regulated by Degradation in Lysosomes

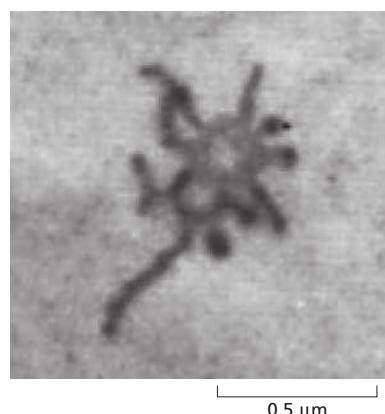
A second pathway that endocytosed receptors can follow from endosomes is taken by many signaling receptors, including opioid receptors and the receptor that binds *epidermal growth factor* (EGF). EGF is a small, extracellular signal protein that stimulates epidermal and various other cells to divide. Unlike LDL receptors, EGF receptors accumulate in clathrin-coated pits only after binding their ligand, and most do not recycle but are degraded in lysosomes, along with the ingested EGF. EGF binding therefore first activates intracellular signaling pathways and then leads to a decrease in the concentration of EGF receptors on the cell surface, a process called *receptor downregulation*, that reduces the cell's subsequent sensitivity to EGF (see Figure 15–20).

Receptor downregulation is highly regulated. The activated receptors are first covalently modified on the cytosolic face with the small protein ubiquitin. Unlike *polyubiquitylation*, which adds a chain of ubiquitins that typically targets a protein for degradation in proteasomes (discussed in Chapter 6), ubiquitin tagging for sorting into the clathrin-dependent endocytic pathway adds just one or a few single ubiquitin molecules to the protein—a process called *monoubiquitylation* or *multiubiquitylation*, respectively. Ubiquitin-binding proteins recognize the attached ubiquitin and help direct the modified receptors into clathrin-coated pits. After delivery to the early endosome, other ubiquitin-binding proteins that are part of *ESCRT complexes* (ESCRT = Endosome Sorting Complex Required for Transport) recognize and sort the ubiquitylated receptors into intraluminal vesicles, which are retained in the maturing late endosome (see Figure 13–47). In this way, addition of ubiquitin blocks receptor recycling to the plasma membrane and directs the receptors into the degradation pathway, as we discuss next.

### Early Endosomes Mature into Late Endosomes

The endosomal compartments can be made visible in the electron microscope by adding a readily detectable tracer molecule, such as the enzyme peroxidase, to the extracellular medium and allowing varying lengths of time for the cell to endocytose the tracer. The distribution of the molecule after its uptake reveals the sequence of events. Within a minute or so after adding the tracer, it starts to appear in **early endosomes**, just beneath the plasma membrane (Figure 13–53). By 5–15 minutes later, it has moved to **late endosomes**, close to the Golgi apparatus and near the nucleus.

How early endosomes arise is not entirely clear, but their membrane and volume are mainly derived from incoming endocytic vesicles that fuse with one another (Movie 13.4). Early endosomes are relatively small and patrol the cytoplasm underlying the plasma membrane in jerky back-and-forth movements along microtubules, capturing incoming vesicles. Typically, an early endosome receives incoming vesicles for about 10 minutes, during which time membrane and fluid is rapidly recycled to the plasma membrane. Some of the incoming cargo, however, accumulates over the lifetime of the early endosome, eventually to be included in the late endosome.



**Figure 13–53** Electron micrograph of an **early endosome**. The endosome is labeled with endocytosed horseradish peroxidase, a widely used enzyme marker, detected in this case by an electron-dense reaction product. Many tubular extensions protrude from the central vacuolar space of the early endosome, which will later mature to give rise to a late endosome. (From J. Tooze and M. Hollinshead, *J. Cell Biol.* 118:813–830, 1992.)

Early endosomes have tubular and vacuolar domains (see Figure 13–53). Most of the membrane surface is in the tubules and most of the volume is in the vacuolar domain. During **endosome maturation**, the two domains have different fates: the vacuolar portions of the early endosome are retained and transformed into late endosomes; the tubular portions shrink. Maturing endosomes, also called *multivesicular bodies*, migrate along microtubules toward the cell interior, shedding membrane tubules and vesicles that recycle material to the plasma membrane and TGN, and receiving newly synthesized lysosomal proteins. As they concentrate in a perinuclear region of the cell, the multivesicular bodies fuse with each other, and eventually with endolysosomes and lysosomes (see Figure 13–47).

Many changes occur during the maturation process. (1) The endosome changes shape and location, as the tubular domains are lost and the vacuolar domains are thoroughly modified. (2) Rab proteins, phosphoinositide lipids, fusion machinery (SNAREs and tethers), and microtubule motor proteins all participate in a molecular makeover of the cytosolic face of the endosome membrane, changing the functional characteristics of the organelle. (3) A V-type ATPase in the endosome membrane pumps H<sup>+</sup> from the cytosol into the endosome lumen and acidifies the organelle. Crucially, the increasing acidity that accompanies maturation renders lysosomal hydrolases increasingly more active, influencing many receptor-ligand interactions, thereby controlling receptor loading and unloading. (4) Intraluminal vesicles sequester endocytosed signaling receptors inside the endosome, thus halting the receptor signaling activity. (5) Lysosome proteins are delivered from the TGN to the maturing endosome. Most of these events occur gradually but eventually lead to a complete transformation of the endosome into an early endolysosome.

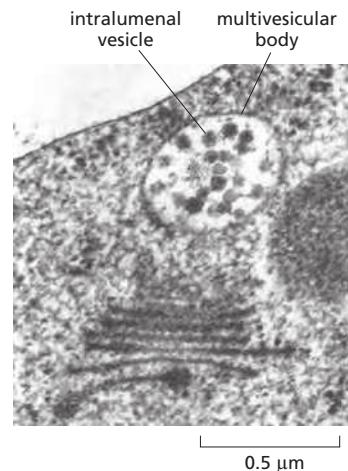
In addition to committing selected cargo for degradation, the maturation process is important for lysosome maintenance. The continual delivery of lysosome components from the TGN to maturing endosomes, ensures a steady supply of new lysosome proteins. The endocytosed materials mix in early endosomes with newly arrived acid hydrolases. Although mild digestion may start here, many hydrolases are synthesized and delivered as proenzymes, called *zymogens*, which contain extra inhibitory domains that keep the hydrolases inactive until these domains are proteolytically removed at later stages of endosome maturation. Moreover, the pH in early endosomes is not low enough to activate lysosomal hydrolases optimally. By these means, cells can retrieve membrane proteins intact from early endosomes and recycle them back to the plasma membrane.

### ESCRT Protein Complexes Mediate the Formation of Intraluminal Vesicles in Multivesicular Bodies

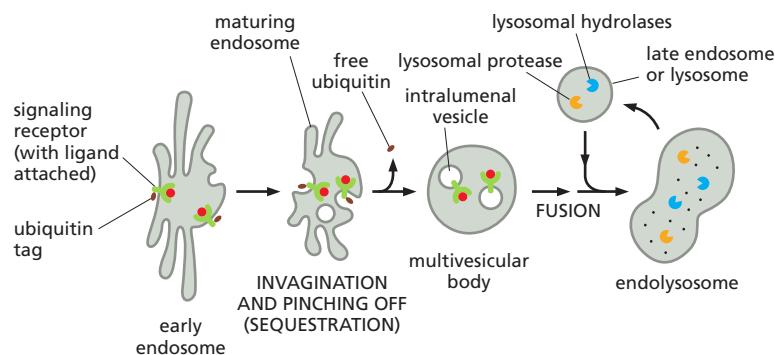
As endosomes mature, patches of their membrane invaginate into the endosome lumen and pinch off to form intraluminal vesicles. Because of their appearance in the electron microscope such maturing endosomes are also called **multivesicular bodies** (Figure 13–54).

The multivesicular bodies carry endocytosed membrane proteins that are to be degraded. As part of the protein-sorting process, receptors destined for degradation, such as the occupied EGF receptors described previously, selectively partition into the invaginating membrane of the multivesicular bodies. In this way, both the receptors and any signaling proteins strongly bound to them are sequestered away from the cytosol where they might otherwise continue signaling. They also are made fully accessible to the digestive enzymes that eventually will degrade them (Figure 13–55). In addition to endocytosed membrane proteins, multivesicular bodies include the soluble content of early endosomes destined for late endosomes and digestion in lysosomes.

As discussed earlier, sorting into intraluminal vesicles requires one or multiple ubiquitin tags, which are added to the cytosolic domains of membrane proteins. These tags initially help guide the proteins into clathrin-coated vesicles in the plasma membrane. Once delivered to the endosomal membrane, the ubiquitin tags are recognized again, this time by a series of cytosolic **ESCRT protein**



**Figure 13–54** Electron micrograph of a multivesicular body. The large amount of internal membrane will be delivered to the lysosome, for digestion. (Courtesy of Andrew Staehelin, from A. Driouch, A. Jauneau and L.A. Staehelin; *Plant Physiol.* 113:487–492, 1997. With permission from the American Society of Plant Biologists.)



**Figure 13–55** The sequestration of endocytosed proteins into intraluminal vesicles of multivesicular bodies.

Ubiquitylated membrane proteins are sorted into domains on the endosome membrane, which invaginate and pinch off to form intraluminal vesicles. The ubiquitin marker is removed and returned to the cytosol for reuse before the intraluminal vesicle closes. Eventually, proteases and lipases in lysosomes digest all of the internal membranes. The invagination processes are essential for complete digestion of endocytosed membrane proteins: because the outer membrane of the multivesicular body becomes continuous with the lysosomal membrane, which is resistant to lysosomal hydrolases; the hydrolases, for example, could not digest the cytosolic domains of endocytosed transmembrane proteins, such as the EGF receptor shown here, if the protein were not localized in intraluminal vesicles.

complexes, (*ESCRT-0*, *-I*, *-II*, and *-III*), which bind sequentially and ultimately mediate the sorting process into the intraluminal vesicles. Membrane invagination into multivesicular bodies also depends on a lipid kinase that phosphorylates phosphatidylinositol to produce PI(3)P, which serves as an additional docking site for the ESCRT complexes; these complexes require both PI(3)P and the presence of ubiquitylated cargo proteins to bind to the endosomal membrane. *ESCRT-III* forms large multimeric assemblies on the membrane that bend the membrane (Figure 13–56).

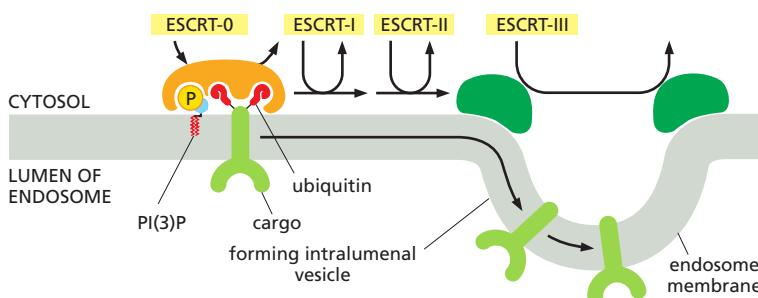
Mutant cells compromised in ESCRT function display signaling defects. In such cells, activated receptors cannot be down-regulated by endocytosis and packaging into multivesicular bodies. The still-active receptors therefore mediate prolonged signaling, which can lead to uncontrolled cell proliferation and cancer.

Processes that shape membranes often use similar machinery. Because of strong similarities in their protein sequences, researchers think that ESCRT complexes are evolutionarily related to components that mediate cell-membrane deformation in cytokinesis in archaea. Similarly, the ESCRT machinery that drives the internal budding from the endosome membrane to form intraluminal vesicles is also used in animal cell cytokinesis and virus budding, which are topologically equivalent, as both processes involve budding away from the cytosolic surface of the membrane (Figure 13–57).

### Recycling Endosomes Regulate Plasma Membrane Composition

The fates of endocytosed receptors—and of any ligands remaining bound to them—vary according to the specific type of receptor. As we discussed, most receptors are recycled and returned to the same plasma membrane domain from which they came; some proceed to a different domain of the plasma membrane, thereby mediating **transcytosis**; and some progress to lysosomes, where they are degraded.

Receptors on the surface of polarized epithelial cells can transfer specific macromolecules from one extracellular space to another by transcytosis. A newborn, for example, obtains antibodies from its mother's milk (which help protect it against infection) by transporting them across the epithelium of its gut. The lumen of the gut is acidic, and, at this low pH, the antibodies in the milk bind to specific receptors on the apical (absorptive) surface of the gut epithelial cells. The receptor-antibody complexes are internalized via clathrin-coated pits and



**Figure 13–56** Sorting of endocytosed membrane proteins into the intraluminal vesicles of a multivesicular body.

A series of complex binding events passes the ubiquitylated cargo proteins sequentially from one ESCRT complex (ESCRT-0) to the next, eventually concentrating them in membrane areas that bud into the lumen of the endosome to form intraluminal vesicles. ESCRT-III assembles into expansive multimeric structures and mediates invagination. The mechanisms of how cargo molecules are shepherded into the vesicles and how the vesicles are formed without including the ESCRT complexes themselves remain unknown. ESCRT complexes are soluble in the cytosol, are recruited to the membrane sequentially as needed, and are then released back into the cytosol as the vesicle pinches off.

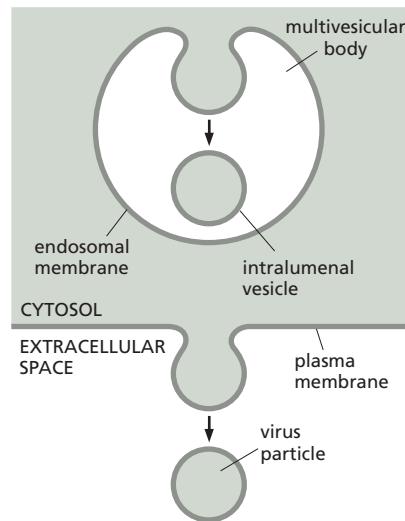
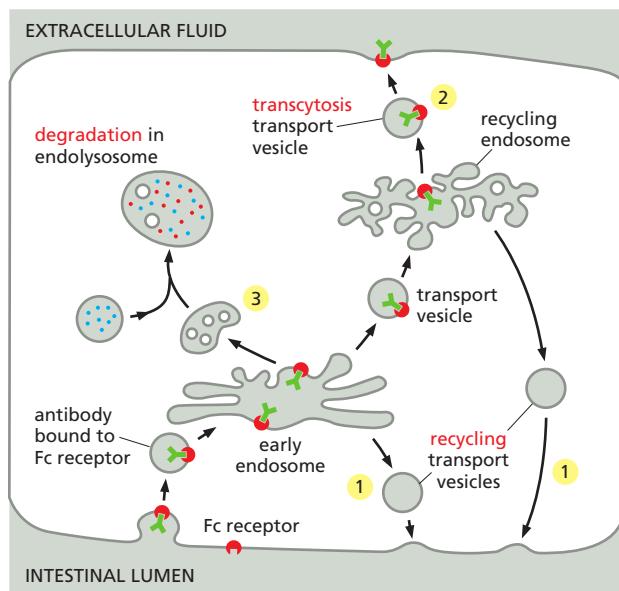
vesicles and are delivered to early endosomes. The complexes remain intact and are retrieved in transport vesicles that bud from the early endosome and subsequently fuse with the basolateral domain of the plasma membrane. On exposure to the neutral pH of the extracellular fluid that bathes the basolateral surface of the cells, the antibodies dissociate from their receptors and eventually enter the baby's bloodstream.

The transcytotic pathway from the early endosome back to the plasma membrane is not direct. The receptors first move from the early endosome to the **recycling endosome**. The variety of pathways that different receptors follow from early endosomes implies that, in addition to binding sites for their ligands and binding sites for coated pits, many receptors also possess sorting signals that guide them into the appropriate transport pathway (Figure 13–58).

Cells can regulate the release of membrane proteins from recycling endosomes, thus adjusting the flux of proteins through the transcytotic pathway according to need. This regulation, the mechanism of which is uncertain, allows recycling endosomes to play an important part in adjusting the concentration of specific plasma membrane proteins. Fat cells and muscle cells, for example, contain large intracellular pools of the glucose transporters that are responsible for the uptake of glucose across the plasma membrane. These membrane transport proteins are stored in specialized recycling endosomes until the hormone *insulin* stimulates the cell to increase its rate of glucose uptake. In response to the insulin signal, transport vesicles rapidly bud from the recycling endosome and deliver large numbers of glucose transporters to the plasma membrane, thereby greatly increasing the rate of glucose uptake into the cell (Figure 13–59). Similarly, kidney cells regulate the insertion of aquaporins and V-ATPase into the plasma membrane to increase water and acid excretion, respectively, both in response to hormones.

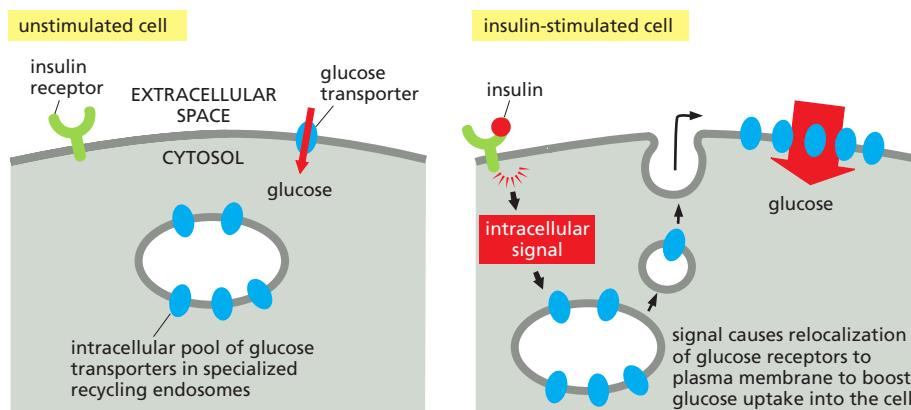
## Specialized Phagocytic Cells Can Ingest Large Particles

**Phagocytosis** is a special form of endocytosis in which a cell uses large endocytic vesicles called **phagosomes** to ingest large particles such as microorganisms and dead cells. Phagocytosis is distinct, both in purpose and mechanism, from macropinocytosis, which we discussed earlier. In protozoa, phagocytosis is a form of feeding; large particles taken up into phagosomes end up in lysosomes, and the products of the subsequent digestive processes pass into the cytosol to be used as food. However, few cells in multicellular organisms are able to ingest such large particles efficiently. In the gut of animals, for example, extracellular processes break down food particles, and cells import the small products of hydrolysis.



**Figure 13–57** Conserved mechanism in multivesicular body formation and **virus budding**. In the two topologically equivalent processes indicated by the arrows, ESCRT complexes (not shown) shape membranes into buds that bulge away from the cytosol.

**Figure 13–58** Possible fates for transmembrane receptor proteins that have been endocytosed. Three pathways from the early endosomal compartment in an epithelial cell are shown. Retrieved receptors are returned (1) to the same plasma membrane domain from which they came (**recycling**) or (2) via a recycling endosome to a different domain of the plasma membrane (**transcytosis**). (3) Receptors that are not specifically retrieved from early or recycling endosomes follow the pathway from the endosomal compartment to lysosomes, where they are degraded (**degradation**). If the ligand that is endocytosed with its receptor stays bound to the receptor in the acidic environment of the endosome, it shares the same fate as the receptor; otherwise, it is delivered to lysosomes. Recycling endosomes are a way-station on the transcytotic pathway. In the transcytosis example shown here, an antibody Fc receptor on a gut epithelial cell binds antibody and is endocytosed, eventually carrying the antibody to the basolateral plasma membrane. The receptor is called an Fc receptor because it binds the Fc part of the antibody (discussed in Chapter 24).



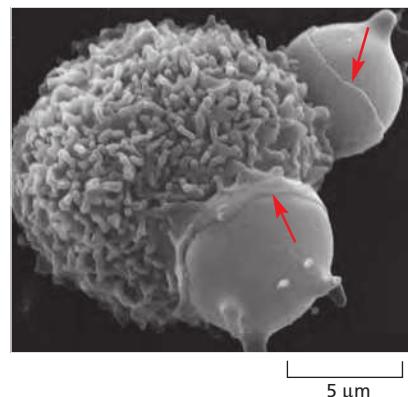
**Figure 13–59 Storage of plasma membrane proteins in recycling endosomes.** Recycling endosomes can serve as an intracellular storage site for specialized plasma membrane proteins that can be mobilized when needed. In the example shown, insulin binding to the insulin receptor triggers an intracellular signaling pathway that causes the rapid insertion of glucose transporters into the plasma membrane of a fat or muscle cell, greatly increasing its glucose intake.

Phagocytosis is important in most animals for purposes other than nutrition, and it is carried out mainly by specialized cells—so-called *professional phagocytes*. In mammals, two important classes of white blood cells that act as professional phagocytes are **macrophages** and **neutrophils** (Movie 13.5). These cells develop from hemopoietic stem cells (discussed in Chapter 22), and they ingest invading microorganisms to defend us against infection. Macrophages also have an important role in scavenging senescent cells and cells that have died by apoptosis (discussed in Chapter 18). In quantitative terms, the clearance of senescent and dead cells is by far the most important: our macrophages, for example, phagocytose more than  $10^{11}$  senescent red blood cells in each of us every day.

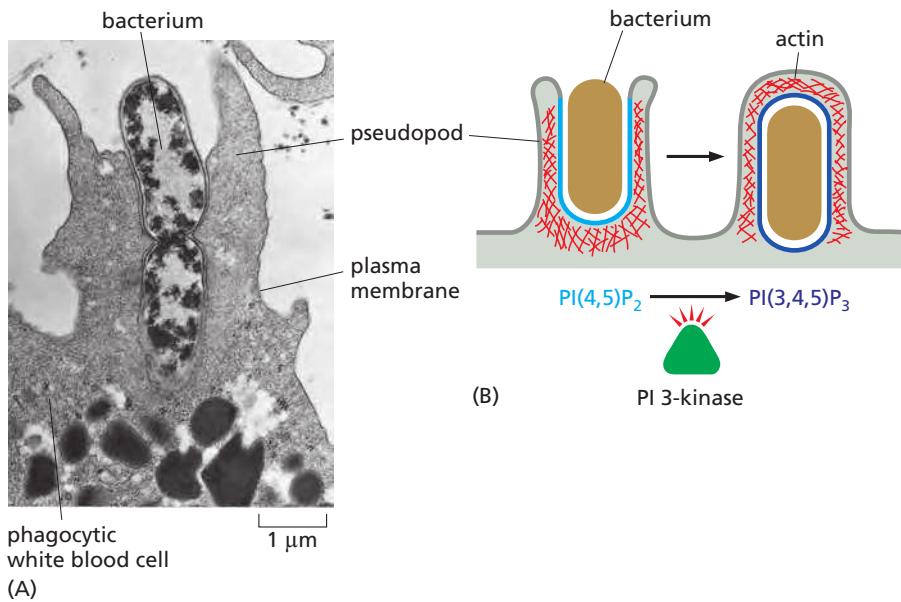
The diameter of a phagosome is determined by the size of its ingested particles, and those particles can be almost as large as the phagocytic cell itself (Figure 13–60). Phagosomes fuse with lysosomes, and the ingested material is then degraded. Indigestible substances remain in the lysosomes, forming *residual bodies* that can be excreted from cells by exocytosis, as mentioned earlier. Some of the internalized plasma membrane components never reach the lysosome, because they are retrieved from the phagosome in transport vesicles and returned to the plasma membrane.

Some pathogenic bacteria have evolved elaborate mechanisms to prevent phagosome–lysosome fusion. The bacterium *Legionella pneumophila*, for example, which causes Legionnaires' disease (discussed in Chapter 23), injects into its unfortunate host a Rab-modifying enzyme that causes certain Rab proteins to misdirect membrane traffic, thereby preventing phagosome–lysosome fusion. The bacterium, thus spared from lysosomal degradation, remains in the modified phagosome, growing and dividing as an intracellular pathogen, protected from the host's adaptive immune system.

Phagocytosis is a cargo-triggered process. That is, it requires the activation of cell-surface receptors that transmit signals to the cell interior. Thus, to be phagocytosed, particles must first bind to the surface of the phagocyte (although not all particles that bind are ingested). Phagocytes have a variety of cell surface receptors that are functionally linked to the phagocytic machinery of the cell. The best-characterized triggers of phagocytosis are antibodies, which protect us by binding to the surface of infectious microorganisms (pathogens) and initiating a series of events that culminate in the invader being phagocytosed. When antibodies initially attack a pathogen, they coat it with antibody molecules that bind to *Fc receptors* on the surface of macrophages and neutrophils, activating the receptors to induce the phagocytic cell to extend pseudopods, which engulf the particle and fuse at their tips to form a phagosome (Figure 13–61A). Localized actin polymerization, initiated by Rho family GTPases and their activating Rho-GEFs (discussed in Chapters 15 and 16), shapes the pseudopods. The activated Rho GTPases switch on the kinase activity of local PI kinases to produce PI(4,5)P<sub>2</sub> in the membrane (see Figure 13–11), which stimulates actin polymerization. To seal off the phagosome and complete the engulfment, actin is depolymerized by a PI 3-kinase that converts the PI(4,5)P<sub>2</sub> to PI(3,4,5)P<sub>3</sub>, which is required for closure



**Figure 13–60 Phagocytosis by a macrophage.** A scanning electron micrograph of a mouse macrophage phagocytosing two chemically altered red blood cells. The red arrows point to edges of thin processes (pseudopods) of the macrophage that are extending as collars to engulf the red cells. (Courtesy of Jean Paul Revel.)



**Figure 13–61** A neutrophil reshaping its plasma membrane during phagocytosis. (A) An electron micrograph of a neutrophil phagocytosing a bacterium, which is in the process of dividing. (B) Pseudopod extension and phagosome formation are driven by actin polymerization and reorganization, which respond to the accumulation of specific phosphoinositides in the membrane of the forming phagosome: PI(4,5)P<sub>2</sub> stimulates actin polymerization, which promotes pseudopod formation, and then PI(3,4,5)P<sub>3</sub> depolymerizes actin filaments at the base. (A, courtesy of Dorothy F. Bainton, Phagocytic Mechanisms in Health and Disease. New York: Intercontinental Medical Book Corporation, 1971.)

of the phagosome and may also contribute to reshaping the actin network to help drive the invagination of the forming phagosome (Figure 13–61B). In this way, the ordered generation and consumption of specific phosphoinositides guides sequential steps in phagosome formation.

Several other classes of receptors that promote phagocytosis have been characterized. Some recognize *complement* components, which collaborate with antibodies in targeting microbes for destruction (discussed in Chapter 24). Others directly recognize oligosaccharides on the surface of certain pathogens. Still others recognize cells that have died by apoptosis. Apoptotic cells lose the asymmetric distribution of phospholipids in their plasma membrane. As a consequence, negatively charged phosphatidylserine, which is normally confined to the cytosolic leaflet of the lipid bilayer, is now exposed on the outside of the cell, where it helps to trigger the phagocytosis of the dead cell.

Remarkably, macrophages will also phagocytose a variety of inanimate particles—such as glass or latex beads and asbestos fibers—yet they do not phagocytose live cells in their own body. The living cells display “don’t-eat-me” signals in the form of cell-surface proteins that bind to inhibiting receptors on the surface of macrophages. The inhibitory receptors recruit tyrosine phosphatases that antagonize the intracellular signaling events required to initiate phagocytosis, thereby locally inhibiting the phagocytic process. Thus phagocytosis, like many other cell processes, depends on a balance between positive signals that activate the process and negative signals that inhibit it. Apoptotic cells are thought both to gain “eat-me” signals (such as extracellularly exposed phosphatidylserine) and to lose their “don’t-eat-me” signals, causing them to be very rapidly phagocytosed by macrophages.

## Summary

Cells ingest fluid, molecules, and particles by endocytosis, in which localized regions of the plasma membrane invaginate and pinch off to form endocytic vesicles. In most cells, endocytosis internalizes a large fraction of the plasma membrane every hour. The cells remain the same size because most of the plasma membrane components (proteins and lipids) that are endocytosed are continually returned to the cell surface by exocytosis. This large-scale endocytic-exocytic cycle is mediated largely by clathrin-coated pits and vesicles but clathrin-independent endocytic pathways also contribute.

While many of the endocytosed molecules are quickly recycled to the plasma membrane, others eventually end up in lysosomes, where they are degraded. Most of the ligands that are endocytosed with their receptors dissociate from their receptors

in the acidic environment of the endosome and eventually end up in lysosomes, while most of the receptors are recycled via transport vesicles back to the cell surface for reuse. Many cell-surface signaling receptors become tagged with ubiquitin when activated by binding their extracellular ligands. Ubiquitylation guides activated receptors into clathrin-coated pits, they and their ligands are efficiently internalized and delivered to early endosomes.

Early endosomes, rapidly mature into late endosomes. During maturation, patches of the endosomal membrane containing ubiquitylated receptors invaginate and pinch off to form intraluminal vesicles. This process is mediated by ESCRT complexes and sequesters the receptors away from the cytosol, which terminates their signaling activity. Late endosomes migrate along microtubules toward the interior of the cell where they fuse with one another and with lysosomes to form endolysosomes, where degradation occurs.

In some cases, both receptor and ligand are transferred to a different plasma membrane domain, causing the ligand to be released at a different surface from where it originated, a process called transcytosis. In some cells, endocytosed plasma membrane proteins and lipids can be stored in recycling endosomes, for as long as necessary until they are needed.

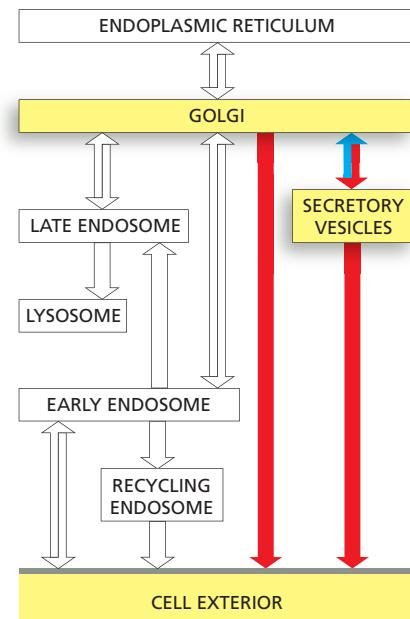
## TRANSPORT FROM THE TRANS GOLGI NETWORK TO THE CELL EXTERIOR: EXOCYTOSIS

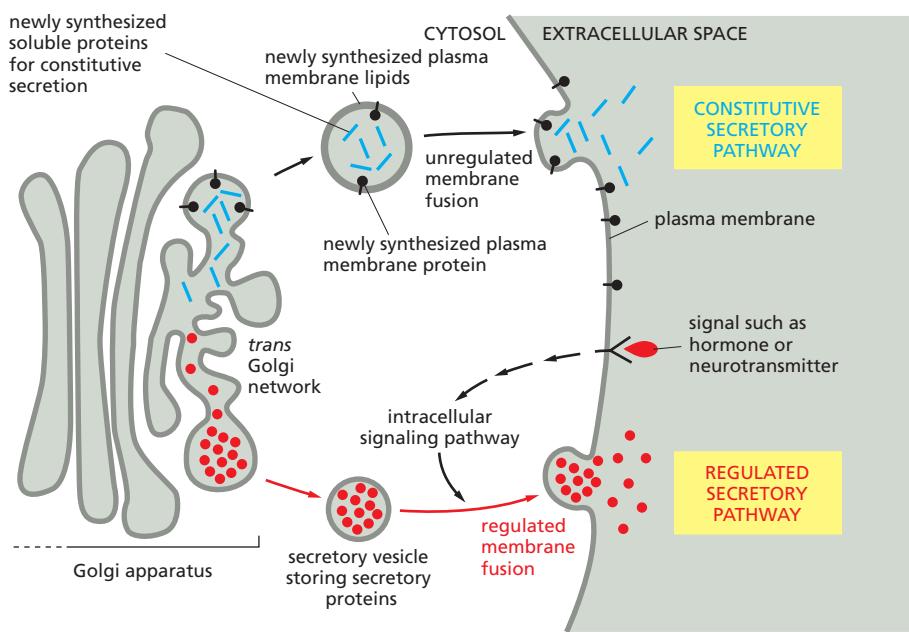
Having considered the cell's endocytic and internal digestive systems and the various types of incoming membrane traffic that converge on lysosomes, we now return to the Golgi apparatus and examine the secretory pathways that lead outward to the cell exterior. Transport vesicles destined for the plasma membrane normally leave the TGN in a steady stream as irregularly shaped tubules. The membrane proteins and the lipids in these vesicles provide new components for the cell's plasma membrane, while the soluble proteins inside the vesicles are secreted to the extracellular space. The fusion of the vesicles with the plasma membrane is called **exocytosis**. This is the route, for example, by which cells secrete most of the proteoglycans and glycoproteins of the *extracellular matrix*, as discussed in Chapter 19.

All cells require this **constitutive secretory pathway**, which operates continuously (Movie 13.6). Specialized secretory cells, however, have a second secretory pathway in which soluble proteins and other substances are initially stored in *secretory vesicles* for later release by exocytosis. This is the **regulated secretory pathway**, found mainly in cells specialized for secreting products rapidly on demand—such as hormones, neurotransmitters, or digestive enzymes (Figure 13–62). In this section, we consider the role of the Golgi apparatus in both of these pathways and compare the two mechanisms of secretion.

### Many Proteins and Lipids Are Carried Automatically from the Trans Golgi Network (TGN) to the Cell Surface

A cell capable of regulated secretion must separate at least three classes of proteins before they leave the TGN—those destined for lysosomes (via endosomes), those destined for secretory vesicles, and those destined for immediate delivery to the cell surface (Figure 13–63). We have already discussed how proteins destined for lysosomes are tagged with M6P for packaging into specific departing vesicles, and analogous signals are thought to direct *secretory proteins* into secretory vesicles. The nonselective constitutive secretory pathway transports most other proteins directly to the cell surface. Because entry into this pathway does not require a particular signal, it is also called the **default pathway**. Thus, in an unpolarized cell such as a white blood cell or a fibroblast, it seems that any protein in the lumen of the Golgi apparatus is automatically carried by the constitutive pathway to the cell surface unless it is specifically returned to the ER, retained as a resident protein in the Golgi apparatus itself, or selected for the pathways that lead to regulated secretion or to lysosomes. In polarized cells, where different products have to be





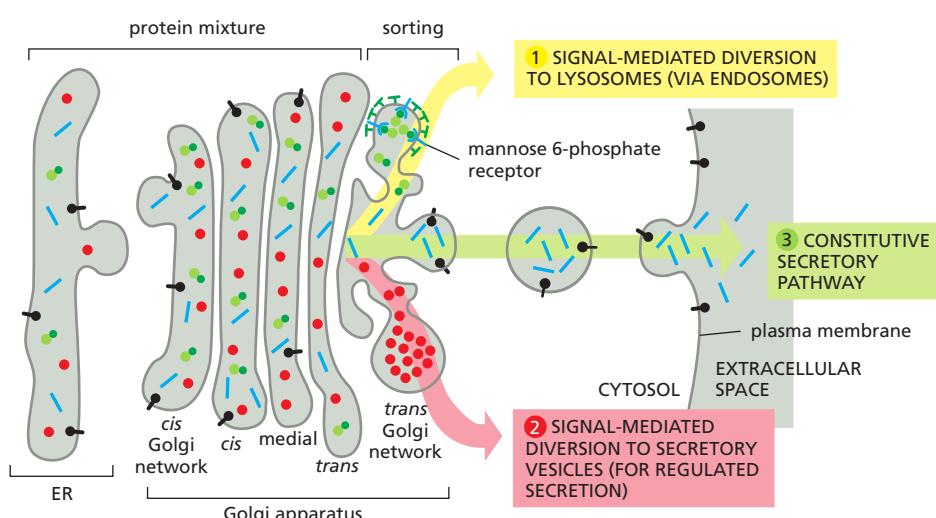
**Figure 13–62** The constitutive and regulated secretory pathways. The two pathways diverge in the TGN. The constitutive secretory pathway operates in all cells. Many soluble proteins are continually secreted from the cell by this pathway, which also supplies the plasma membrane with newly synthesized membrane lipids and proteins. Specialized secretory cells also have a regulated secretory pathway, by which selected proteins in the TGN are diverted into secretory vesicles, where the proteins are concentrated and stored until an extracellular signal stimulates their secretion. The regulated secretion of small molecules, such as histamine and neurotransmitters, occurs by a similar pathway; these molecules are actively transported from the cytosol into preformed secretory vesicles. There they are often bound to specific macromolecules (proteoglycans, for histamine) so that they can be stored at high concentration without generating an excessively high osmotic pressure.

delivered to different domains of the cell surface, we shall see that the options are more complex.

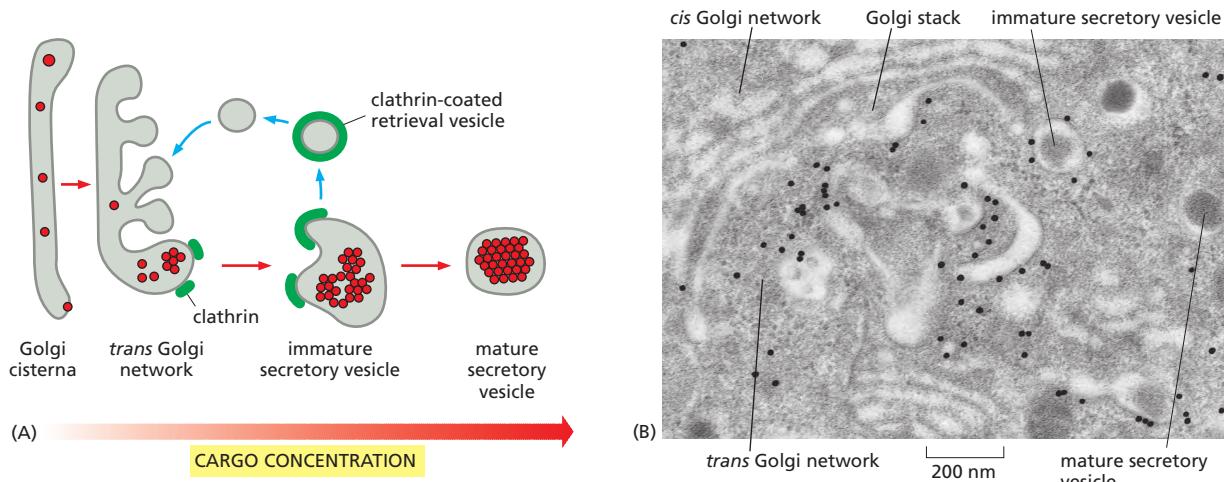
### Secretory Vesicles Bud from the Trans Golgi Network

Cells that are specialized for secreting some of their products rapidly on demand concentrate and store these products in **secretory vesicles** (often called *dense-core secretory granules* because they have dense cores when viewed in the electron microscope). Secretory vesicles form from the TGN, and they release their contents to the cell exterior by exocytosis in response to specific signals. The secreted product can be either a small molecule (such as histamine or a neuropeptide) or a protein (such as a hormone or digestive enzyme).

Proteins destined for secretory vesicles (called *secretory proteins*) are packaged into appropriate vesicles in the TGN by a mechanism that involves the selective aggregation of the secretory proteins. Clumps of aggregated, electron-dense material can be detected by electron microscopy in the lumen of the TGN. The signals that direct secretory proteins into such aggregates are not well defined and may be quite diverse. When a gene encoding a secretory protein is artificially expressed in a secretory cell that normally does not make the protein, the foreign protein is appropriately packaged into secretory vesicles. This observation



**Figure 13–63** The three best-understood pathways of protein sorting in the trans Golgi network. (1) Proteins with the mannose 6-phosphate (M6P) marker (see Figure 13–45) are diverted to lysosomes (via endosomes) in clathrin-coated transport vesicles. (2) Proteins with signals directing them to secretory vesicles are concentrated in such vesicles as part of a regulated secretory pathway that is present only in specialized secretory cells. (3) In unpolarized cells, a constitutive secretory pathway delivers proteins with no special features to the cell surface. In polarized cells, such as epithelial cells, however, secreted and plasma membrane proteins are selectively directed to either the apical or the basolateral plasma membrane domain, so a specific signal must mediate at least one of these two pathways, as we discuss later.



shows that, although the proteins that an individual cell expresses and packages in secretory vesicles differ, they contain common sorting signals, which function properly even when the proteins are expressed in cells that do not normally make them.

It is unclear how the aggregates of secretory proteins are segregated into secretory vesicles. Secretory vesicles have unique proteins in their membrane, some of which might serve as receptors for aggregated protein in the TGN. The aggregates are much too big, however, for each molecule of the secreted protein to be bound by its own cargo receptor, as occurs for transport of the lysosomal enzymes. The uptake of the aggregates into secretory vesicles may therefore more closely resemble the uptake of particles by phagocytosis at the cell surface, where the plasma membrane zippers up around large structures.

Initially, most of the membrane of the secretory vesicles that leave the TGN is only loosely wrapped around the clusters of aggregated secretory proteins. Morphologically, these *immature secretory vesicles* resemble dilated *trans* Golgi cisternae that have pinched off from the Golgi stack. As the vesicles mature, they fuse with one another and their contents become concentrated (Figure 13–64A), probably as the result of both the continuous retrieval of membrane that is recycled to the TGN, and the progressive acidification of the vesicle lumen that results from the increasing concentration of V-type ATPases in the vesicle membrane that acidify all endocytic and exocytic organelles (see Figure 13–37). The degree of concentration of proteins during the formation and maturation of secretory vesicles is only a small part of the total 200–400-fold concentration of these proteins that occurs after they leave the ER. Secretory and membrane proteins become concentrated as they move from the ER through the Golgi apparatus because of an extensive retrograde retrieval process mediated by COPI-coated transport vesicles that carry soluble ER resident proteins back to the ER, while excluding the secretory and membrane proteins (see Figure 13–25).

Membrane recycling is important for returning Golgi components to the Golgi apparatus, as well as for concentrating the contents of secretory vesicles. The vesicles that mediate this retrieval originate as clathrin-coated buds on the surface of immature secretory vesicles, often being seen even on budding secretory vesicles that have not yet pinched off from the Golgi stack (Figure 13–64B).

Because the final mature secretory vesicles are so densely filled with contents, the secretory cell can disgorge large amounts of material promptly by exocytosis when triggered to do so (Figure 13–65).

### Precursors of Secretory Proteins Are Proteolytically Processed During the Formation of Secretory Vesicles

Concentration is not the only process to which secretory proteins are subjected as the secretory vesicles mature. Many protein hormones and small neuropeptides,

**Figure 13–64** The formation of secretory vesicles. (A) Secretory proteins become segregated and highly concentrated in secretory vesicles by two mechanisms. First, they aggregate in the ionic environment of the TGN; often the aggregates become more condensed as secretory vesicles mature and their lumen becomes more acidic. Second, clathrin-coated vesicles retrieve excess membrane and luminal content present in immature secretory vesicles as the secretory vesicles mature. (B) This electron micrograph shows secretory vesicles forming from the TGN in an insulin-secreting  $\beta$  cell of the pancreas. Anti-clathrin antibodies conjugated to gold spheres (black dots) have been used to locate clathrin molecules. The immature secretory vesicles, which contain insulin precursor protein (proinsulin), contain clathrin patches, which are no longer seen on the mature secretory vesicle. (B, courtesy of Lelio Orci.)

as well as many secreted hydrolytic enzymes, are synthesized as inactive precursors. Proteolysis is necessary to liberate the active molecules from these precursor proteins. The cleavages occur in the secretory vesicles and sometimes in the extracellular fluid after secretion. Additionally, many of the precursor proteins have an N-terminal *pro-peptide* that is cleaved off to yield the mature protein. These proteins are synthesized as *pre-pro-proteins*, the *pre-peptide* consisting of the ER signal peptide that is cleaved off earlier in the rough ER (see Figure 12–36). In other cases, peptide signaling molecules are made as *polyproteins* that contain multiple copies of the same amino acid sequence. In still more complex cases, a variety of peptide signaling molecules are synthesized as parts of a single polyprotein that acts as a precursor for multiple end products, which are individually cleaved from the initial polypeptide chain. The same polyprotein may be processed in various ways to produce different peptides in different cell types (Figure 13–66).

Why is proteolytic processing so common in the secretory pathway? Some of the peptides produced in this way, such as the *enkephalins* (five-amino-acid neuropeptides with morphine-like activity), are undoubtedly too short in their mature forms to be co-translationally transported into the ER lumen or to include the necessary signal for packaging into secretory vesicles. In addition, for secreted hydrolytic enzymes—or any other protein whose activity could be harmful inside the cell that makes it—delaying activation of the protein until it reaches a secretory vesicle, or until after it has been secreted, has a clear advantage: the delay prevents the protein from acting prematurely inside the cell in which it is synthesized.

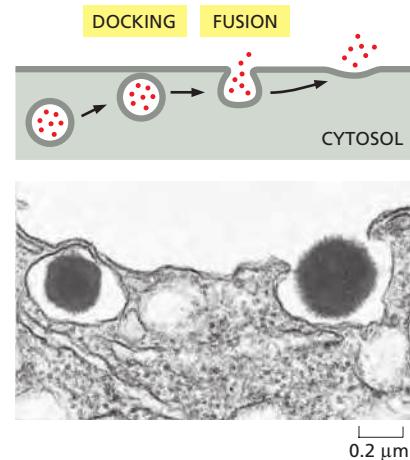
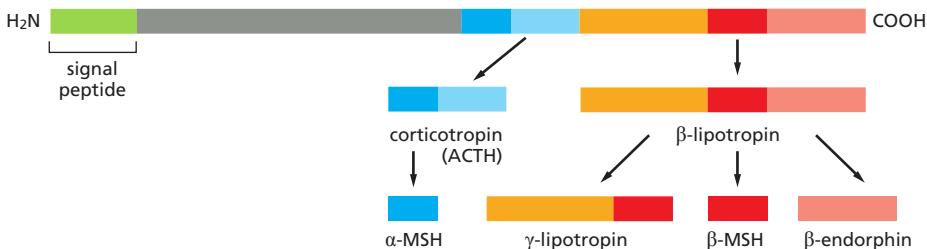
### Secretory Vesicles Wait Near the Plasma Membrane Until Signaled to Release Their Contents

Once loaded, a secretory vesicle has to reach the site of secretion, which in some cells is far away from the TGN. Nerve cells are the most extreme example. Secretory proteins, such as peptide neurotransmitters (neuropeptides), which will be released from nerve terminals at the end of the axon, are made and packaged into secretory vesicles in the cell body. They then travel along the axon to the nerve terminals, which can be a meter or more away. As discussed in Chapter 16, motor proteins propel the vesicles along axonal microtubules, whose uniform orientation guides the vesicles in the proper direction. Microtubules also guide transport vesicles to the cell surface for constitutive exocytosis.

Whereas transport vesicles containing materials for constitutive release fuse with the plasma membrane once they arrive there, secretory vesicles in the regulated pathway wait at the membrane until the cell receives a signal to secrete, and they then fuse. The signal can be an electrical nerve impulse (an action potential) or an extracellular signal molecule, such as a hormone: in either case, it leads to a transient increase in the concentration of free  $\text{Ca}^{2+}$  in the cytosol.

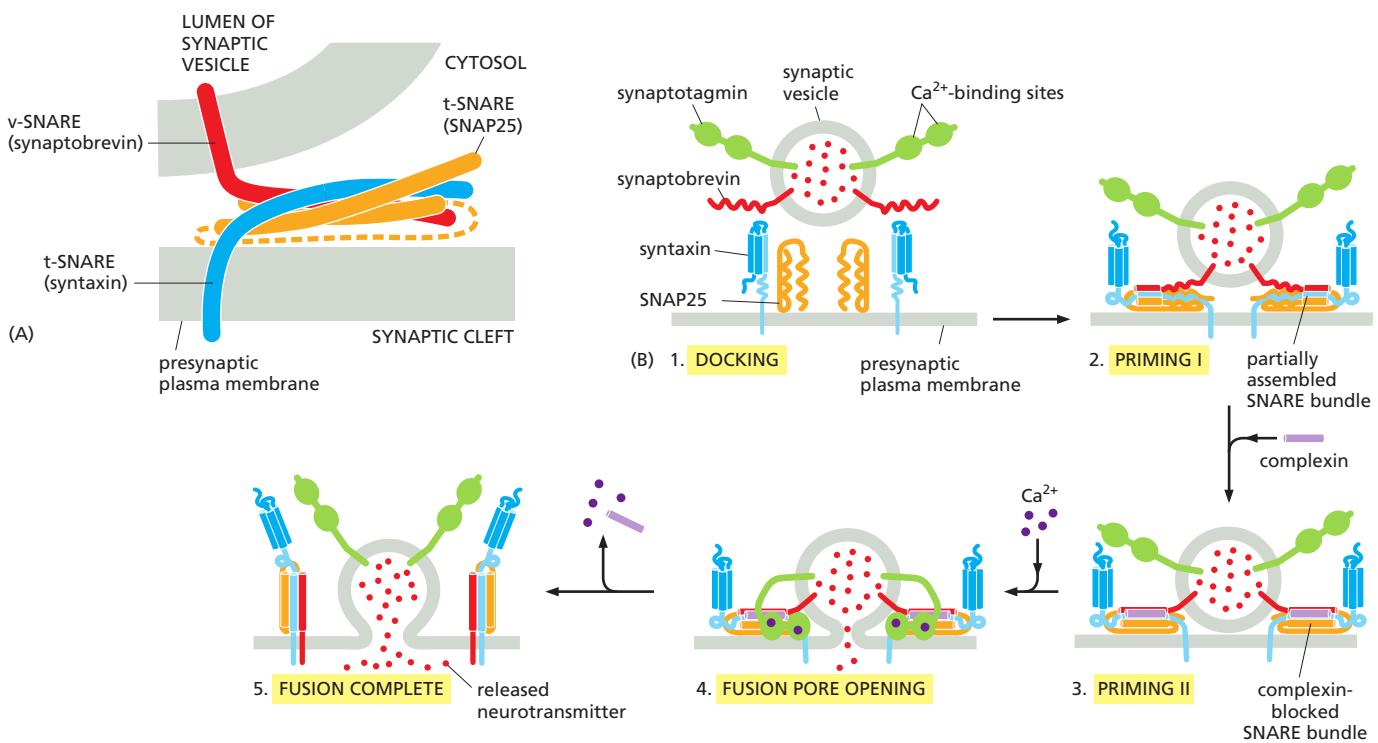
### For Rapid Exocytosis, Synaptic Vesicles Are Primed at the Presynaptic Plasma Membrane

Nerve cells (and some endocrine cells) contain two types of secretory vesicles. As for all secretory cells, these cells package proteins and neuropeptides in dense-cored secretory vesicles in the standard way for release by the regulated secretory pathway. In addition, however, they use another specialized class of tiny ( $\approx 50$  nm



**Figure 13–65** Exocytosis of secretory vesicles. The process is illustrated schematically (top) and in an electron micrograph that shows the release of insulin from a secretory vesicle of a pancreatic  $\beta$  cell. (Courtesy of Lelio Orci, from L. Orci, J.-D. Vassalli and A. Perrelet, *Sci. Am.* 259:85–94, 1988.)

**Figure 13–66** Alternative processing pathways for the prohormone proopiomelanocortin. The initial cleavages are made by proteases that cut next to pairs of positively charged amino acids (Lys-Arg, Lys-Lys, Arg-Lys, or Arg-Arg pairs). Trimming reactions then produce the final secreted products. Different cell types produce different concentrations of individual processing enzymes, so that the same prohormone precursor is cleaved to produce different peptide hormones. In the anterior lobe of the pituitary gland, for example, only corticotropin (ACTH) and  $\beta$ -lipotropin are produced from proopiomelanocortin, whereas in the intermediate lobe of the pituitary gland mainly  $\alpha$ -melanocyte stimulating hormone ( $\alpha$ -MSH),  $\gamma$ -lipotropin,  $\beta$ -MSH, and  $\beta$ -endorphin are produced— $\alpha$ -MSH from ACTH and the other three from  $\beta$ -lipotropin, as shown.



**Figure 13–67** Exocytosis of synaptic vesicles. For orientation at a synapse, see Figure 11–36. (A) The trans-SNARE complex responsible for docking synaptic vesicles at the plasma membrane of nerve terminals consists of three proteins. The v-SNARE *synaptobrevin* and the t-SNARE *syntaxin* are both transmembrane proteins, and each contributes one  $\alpha$  helix to the complex. By contrast to other SNAREs discussed earlier, the t-SNARE *SNAP25* is a peripheral membrane protein that contributes two  $\alpha$  helices to the four-helix bundle; the two helices are connected by a loop (dashed line) that lies parallel to the membrane and has fatty acyl chains (not shown) attached to anchor it there. The four  $\alpha$  helices are shown as rods for simplicity. (B) At the synapse, the basic SNARE machinery is modulated by the  $\text{Ca}^{2+}$  sensor *synaptotagmin* and an additional protein called *complexin*. Synaptic vesicles first dock at the membrane (Step 1), and the SNARE bundle partially assembles (Step 2), resulting in a “primed vesicle” that is already drawn close to the membrane. The SNARE bundle assembles further but the additional binding of complexin prevents fusion (Step 3). Upon arrival of an action potential,  $\text{Ca}^{2+}$  enters the cell and binds to synaptotagmin, which releases the block and opens a fusion pore (Step 4). Further rearrangements complete the fusion reaction (Step 5) and release the fusion machinery, which now can be reused. This elaborate arrangement allows the fusion machinery to respond on the millisecond time scale essential for rapid and repetitive synaptic signaling. (A, adapted from R.B. Sutton et al., *Nature* 395:347–353, 1998. With permission from Macmillan Publishers Ltd.; B, adapted from Z.P. Pang and T.C. Südhof, *Curr. Opin. Cell Biol.* 22:496–505, 2010. With permission from Elsevier.)

diameter) secretory vesicles called **synaptic vesicles**. These vesicles store small *neurotransmitter molecules*, such as acetylcholine, glutamate, glycine, and  $\gamma$ -aminobutyric acid (GABA), which mediate rapid signaling from nerve cell to its target cell at chemical synapses. When an action potential arrives at a nerve terminal, it causes an influx of  $\text{Ca}^{2+}$  through voltage-gated  $\text{Ca}^{2+}$  channels, which triggers the synaptic vesicles to fuse with the plasma membrane and release their contents to the extracellular space (see Figure 11–36). Some neurons fire more than 1000 times per second, releasing neurotransmitters each time.

The speed of transmitter release (taking only milliseconds) indicates that the proteins mediating the fusion reaction do not undergo complex, multistep rearrangements. Rather, after vesicles have been docked at the presynaptic plasma membrane, they undergo a priming step, which prepares them for rapid fusion. In the primed state, the SNAREs are partly paired, their helices are not fully wound into the final four-helix bundle required for fusion (Figure 13–67). Proteins called *complexins* freeze the SNARE complexes in this metastable state. The brake imposed by the complexins is released by another synaptic vesicle protein, *synaptotagmin*, which contains  $\text{Ca}^{2+}$ -binding domains. A rise in cytosolic  $\text{Ca}^{2+}$  triggers binding of synaptotagmin to phospholipids and to the SNAREs, displacing the complexins. As the SNARE bundle zippers up completely, a fusion pore

opens and the neurotransmitters are released. At a typical synapse, only a small number of the docked vesicles are primed and ready for exocytosis. The use of only a small number of vesicles at a time allows each synapse to fire over and over again in quick succession. With each firing, new synaptic vesicles dock and become primed to replace those that have fused and released their contents.

### Synaptic Vesicles Can Form Directly from Endocytic Vesicles

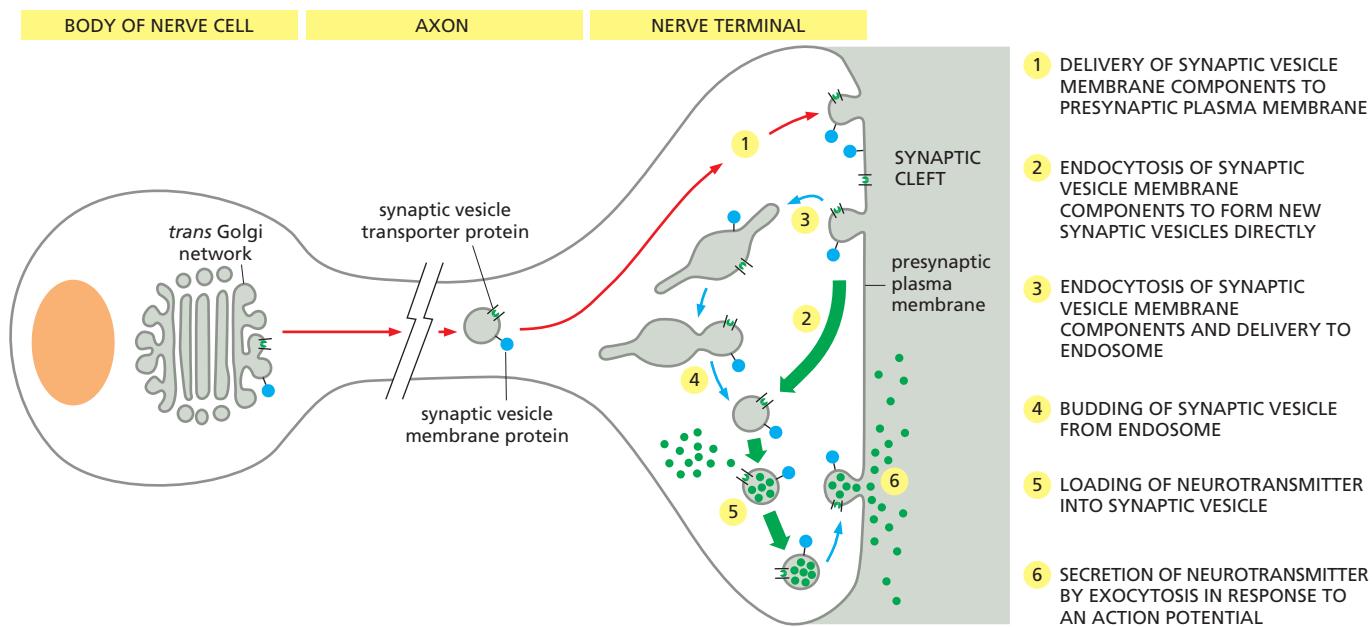
For the nerve terminal to respond rapidly and repeatedly, synaptic vesicles need to be replenished very quickly after they discharge. Thus, most synaptic vesicles are generated not from the Golgi membrane in the nerve cell body but by local recycling from the presynaptic plasma membrane in the nerve terminals (Figure 13–68). Similarly, newly made membrane components of the synaptic vesicles are initially delivered to the plasma membrane by the constitutive secretory pathway and then retrieved by endocytosis. But instead of fusing with endosomes, most of the endocytic vesicles immediately fill with neurotransmitter to become synaptic vesicles.

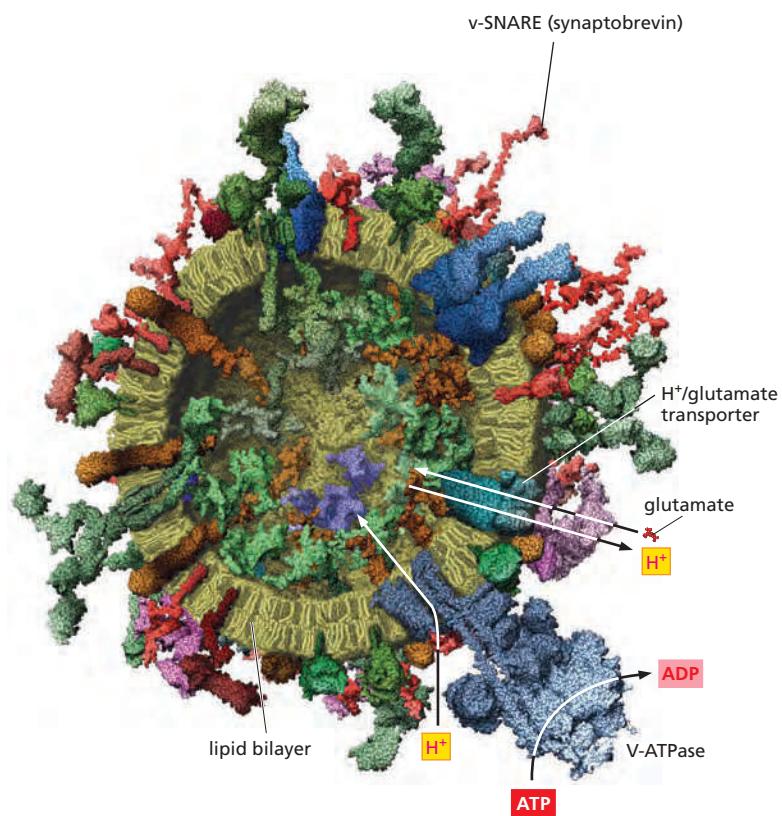
The membrane components of a synaptic vesicle include transporters specialized for the uptake of neurotransmitter from the cytosol, where the small-molecule neurotransmitters that mediate fast synaptic signaling are synthesized. Once filled with neurotransmitter, the synaptic vesicles can be used again (see Figure 13–68). Because synaptic vesicles are abundant and relatively uniform in size, they can be purified in large numbers and, consequently, are the best-characterized organelle of the cell, in that all of their membrane components have been identified by quantitative proteomic analyses (Figure 13–69). Extending this analysis to a complete presynaptic terminal, allows us to model the crowded environment in which these reactions occur.

### Secretory Vesicle Membrane Components Are Quickly Removed from the Plasma Membrane

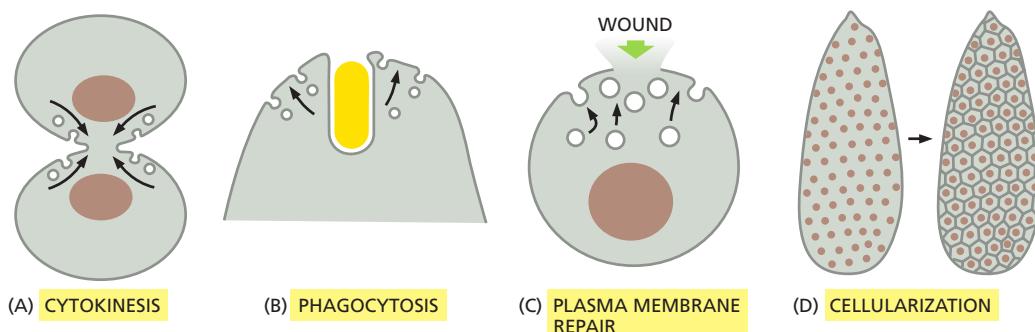
When a secretory vesicle fuses with the plasma membrane, its contents are discharged from the cell by exocytosis, and its membrane becomes part of the plasma membrane. Although this should greatly increase the surface area of the plasma membrane, it does so only transiently, because membrane components are removed from the surface by endocytosis almost as fast as they are added by exocytosis, a process reminiscent of the endocytic-exocytic cycle discussed earlier. After their removal from the plasma membrane, the proteins of the secretory

**Figure 13–68** The formation of synaptic vesicles in a nerve cell. These tiny uniform vesicles are found only in nerve cells and in some endocrine cells, where they store and secrete small-molecule neurotransmitters. The import of neurotransmitter directly into the small endocytic vesicles that form from the plasma membrane is mediated by membrane transporters that function as antiports and are driven by an H<sup>+</sup> gradient maintained by V-ATPase H<sup>+</sup> pumps in the vesicle membrane (discussed in Chapter 11).





**Figure 13–69 Scale models of a brain presynaptic terminal and a synaptic vesicle.** The illustrations show sections through a pre-synaptic terminal (A; enlarged in B) and a synaptic vesicle (C) in which proteins and lipids are drawn to scale based on their known stoichiometry and either known or approximated structures. The relative localization of protein molecules in different regions of the presynaptic terminal was inferred from super-resolution imaging and electron microscopy. The model in (A) contains 300,000 proteins of 60 different kinds that vary in abundance from 150 copies to 20,000 copies. In the model in (C), only 70% of the membrane proteins present in the membrane are shown; a complete model would therefore show a membrane that is even more crowded than this picture suggests ([Movie 13.7](#)). Each synaptic vesicle membrane contains 7000 phospholipid molecules and 5700 cholesterol molecules. Each also contains close to 50 different integral membrane protein molecules, which vary widely in their relative abundance and together contribute about 600 transmembrane  $\alpha$  helices. The transmembrane v-SNARE synaptobrevin is the most abundant protein in the vesicle (~70 copies per vesicle). By contrast, the V-ATPase, which uses ATP hydrolysis to pump H<sup>+</sup> into the vesicle lumen, is present in 1–2 copies per vesicle. The H<sup>+</sup> gradient provides the energy for neurotransmitter import by an H<sup>+</sup>/neurotransmitter antiport, which loads each vesicle with 1800 neurotransmitter molecules, such as glutamate, one of which is shown to scale. (A and B, from B.G. Wilhelm et al., *Science* 344:1023–1028, 2014. With permission from AAAS; C, adapted from S. Takamori et al., *Cell* 127:831–846, 2006. With permission from Elsevier.)



vesicle membrane are either recycled or shuttled to lysosomes for degradation. The amount of secretory vesicle membrane that is temporarily added to the plasma membrane can be enormous: in a pancreatic acinar cell discharging digestive enzymes for delivery to the gut lumen, about  $900 \mu\text{m}^2$  of vesicle membrane is inserted into the apical plasma membrane (whose area is only  $30 \mu\text{m}^2$ ) when the cell is stimulated to secrete.

Control of membrane traffic thus has a major role in maintaining the composition of the various membranes of the cell. To maintain each membrane-enclosed compartment in the secretory and endocytic pathways at a constant size, the balance between the outward and inward flows of membrane needs to be precisely regulated. For cells to grow, however, the forward flow needs to be greater than the retrograde flow, so that the membrane can increase in area. For cells to maintain a constant size, the forward and retrograde flows must be equal. We still know very little about the mechanisms that coordinate these flows.

### Some Regulated Exocytosis Events Serve to Enlarge the Plasma Membrane

An important task of regulated exocytosis is to deliver more membrane to enlarge the surface area of a cell's plasma membrane when such a need arises. A spectacular example is the plasma membrane expansion that occurs during the cellularization process in a fly embryo, which initially is a syncytium—a single cell containing about 6000 nuclei surrounded by a single plasma membrane (see Figure 21–15). Within tens of minutes, the embryo is converted into the same number of cells. This process of *cellularization* requires a vast amount of new plasma membrane, which is added by a carefully orchestrated fusion of cytoplasmic vesicles, eventually forming the plasma membranes that enclose the separate cells. Similar vesicle fusion events are required to enlarge the plasma membrane when other animal cells or plant cells divide during *cytokinesis* (discussed in Chapter 17).

Many animal cells, especially those subjected to mechanical stresses, frequently experience small ruptures in their plasma membrane. In a remarkable process thought to involve both homotypic vesicle–vesicle fusion and exocytosis, a temporary cell-surface patch is quickly fashioned from locally available internal-membrane sources, such as lysosomes. In addition to providing an emergency barrier against leaks, the patch reduces membrane tension over the wounded area, allowing the bilayer to flow back together to restore continuity and seal the puncture. This membrane repair process, the fusion and exocytosis of vesicles is triggered by the sudden increase of  $\text{Ca}^{2+}$ , which is abundant in the extracellular space and rushes into the cell as soon as the plasma membrane is punctured. **Figure 13–70** shows four examples in which regulated exocytosis leads to plasma membrane expansion.

### Polarized Cells Direct Proteins from the Trans Golgi Network to the Appropriate Domain of the Plasma Membrane

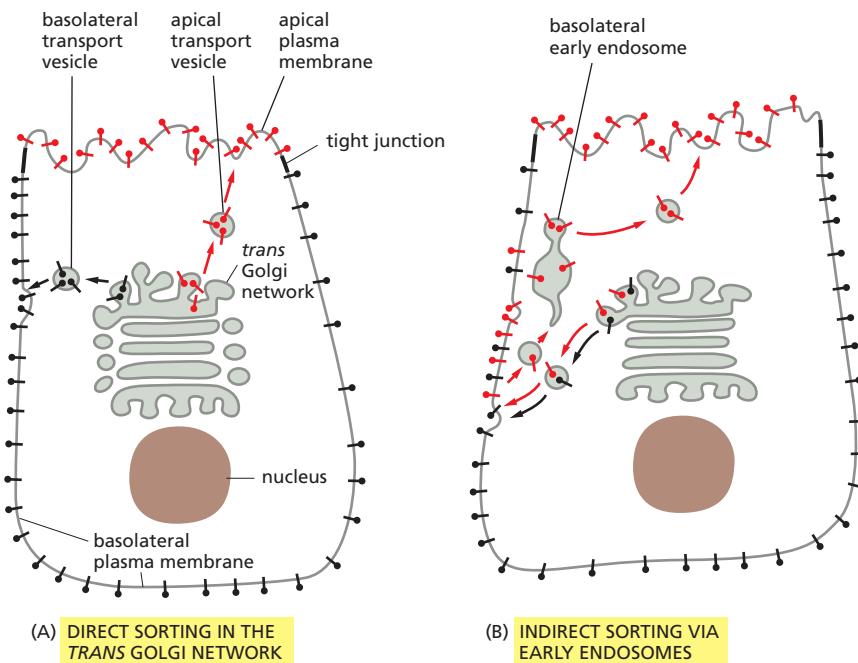
Most cells in tissues are *polarized*, with two or more molecularly and functionally distinct plasma membrane domains. This raises the general problem of how the

**Figure 13–70** Four examples of regulated exocytosis leading to plasma membrane enlargement. The vesicles fusing with the plasma membrane during cytokinesis (A) and phagocytosis (B) are thought to be derived from endosomes, whereas those involved in wound repair (C) may be derived from plasma membranes. The vast amount of new plasma membrane inserted during cellularization in a fly embryo occurs by the fusion of cytoplasmic vesicles (D).

delivery of membrane from the Golgi apparatus is organized so as to maintain the differences between one cell-surface domain and another. A typical epithelial cell, for example, has an *apical domain*, which faces either an internal cavity or the outside world and often has specialized features such as cilia or a brush border of microvilli. It also has a *basolateral domain*, which covers the rest of the cell. The two domains are separated by a ring of *tight junctions* (see Figure 19–21), which prevent proteins and lipids (in the outer leaflet of the lipid bilayer) from diffusing between the two domains, so that the differences between the two domains are maintained.

In principle, differences between plasma membrane domains need not depend on the targeted delivery of the appropriate membrane components. Instead, membrane components could be delivered to all regions of the cell surface indiscriminately but then be selectively stabilized in some locations and selectively eliminated in others. Although this strategy of random delivery followed by selective retention or removal seems to be used in certain cases, deliveries are often specifically directed to the appropriate membrane domain. Epithelial cells lining the gut, for example, secrete digestive enzymes and mucus at their apical surface and components of the basal lamina at their basolateral surface. Such cells must have ways of directing vesicles carrying different cargoes to different plasma membrane domains. Proteins from the ER destined for different domains travel together until they reach the TGN, where they are separated and dispatched in secretory or transport vesicles to the appropriate plasma membrane domain (Figure 13–71).

The apical plasma membrane of most epithelial cells is greatly enriched in glycosphingolipids, which help protect this exposed surface from damage—for example, from the digestive enzymes and low pH in sites such as the gut or stomach, respectively. Similarly, plasma membrane proteins that are linked to the lipid bilayer by a GPI anchor (see Figure 12–52) are found predominantly in the apical plasma membrane. If recombinant DNA techniques are used to attach a GPI anchor to a protein that would normally be delivered to the basolateral surface, the protein is usually delivered to the apical surface instead. GPI-anchored proteins are thought to be directed to the apical membrane because they associate with glycosphingolipids in lipid rafts that form in the membrane of the TGN. As discussed in Chapter 10, lipid rafts form in the TGN and plasma membrane when glycosphingolipids and cholesterol molecules self-associate (see Figure 10–13).



**Figure 13–71** Two ways of sorting plasma membrane proteins in a polarized epithelial cell. (A) In the direct pathway, proteins destined for different plasma membrane domains are sorted and packaged into different transport vesicles. The lipid-raft-dependent delivery system to the apical domain described in the text is an example of the direct pathway. (B) In the indirect pathway, a protein is retrieved from the inappropriate plasma membrane domain by endocytosis and then transported to the correct domain via early endosomes—that is, by transcytosis. The indirect pathway, for example, is used in liver hepatocytes to deliver proteins to the apical domain that lines bile ducts.

Having selected a unique set of cargo molecules, the rafts then bud from the TGN into transport vesicles destined for the apical plasma membrane. This process is similar to the selective partitioning of some membrane proteins into the specialized lipid domains in caveolae at the plasma membrane discussed earlier.

Membrane proteins destined for delivery to the basolateral membrane contain sorting signals in their cytosolic tail. When present in an appropriate structural context, these signals are recognized by coat proteins that package them into appropriate transport vesicles in the TGN. The same basolateral signals that are recognized in the TGN also function in early endosomes to redirect the proteins back to the basolateral plasma membrane after they have been endocytosed.

## Summary

*Cells can secrete molecules by exocytosis in either a constitutive or a regulated fashion. Whereas the regulated pathways operate only in specialized secretory cells, a constitutive secretory pathway operates in all eukaryotic cells, characterized by continual vesicle transport from the TGN to the plasma membrane. In the regulated pathways, the molecules are stored either in secretory vesicles or in synaptic vesicles, which do not fuse with the plasma membrane to release their contents until they receive an appropriate signal. Secretory vesicles containing proteins for secretion bud from the TGN. The secretory proteins become concentrated during the formation and maturation of the secretory vesicles. Synaptic vesicles, which are confined to nerve cells and some endocrine cells, form from both endocytic vesicles and from endosomes, and they mediate the regulated secretion of small-molecule neurotransmitters at the axon terminals of nerve cells.*

*Proteins are delivered from the TGN to the plasma membrane by the constitutive pathway unless they are diverted into other pathways or retained in the Golgi apparatus. In polarized cells, the transport pathways from the TGN to the plasma membrane operate selectively to ensure that different sets of membrane proteins, secreted proteins, and lipids are delivered to the different domains of the plasma membrane.*

## WHAT WE DON'T KNOW

- How are targeting and fusion proteins such as SNAREs regulated, so that they can be returned to their respective donor compartments in an inactive state?
- How does a cell balance exocytic and endocytic events to keep its plasma membrane a constant size?
- Can newly formed daughter cells generate a Golgi apparatus *de novo*, or do they have to inherit it?
- How do lysosomes avoid digesting their own membranes?
- How does a cell maintain the right amount of every component (organelles, molecules), and how does it change these amounts as needed (for example, to greatly expand the endoplasmic reticulum when the cell needs to produce large amounts of secreted proteins)?

## PROBLEMS

### Which statements are true? Explain why or why not.

**13–1** In all events involving fusion of a vesicle to a target membrane, the cytosolic leaflets of the vesicle and target bilayers always fuse together, as do the leaflets that are not in contact with the cytosol.

**13–2** There is one strict requirement for the exit of a protein from the ER: it must be correctly folded.

**13–3** All the glycoproteins and glycolipids in intracellular membranes have oligosaccharide chains facing the luminal side, and all those in the plasma membrane have oligosaccharide chains facing the outside of the cell.

### Discuss the following problems.

**13–4** In a nondividing cell such as a liver cell, why must the flow of membrane between compartments be balanced, so that the retrieval pathways match the outward flow? Would you expect the same balanced flow in a gut epithelial cell, which is actively dividing?

**13–5** Enveloped viruses, which have a membrane coat, gain access to the cytosol by fusing with a cell membrane. Why do you suppose that these viruses encode their own special fusion protein, rather than making use of a cell's SNAREs?

**13–6** For fusion of a vesicle with its target membrane to occur, the membranes have to be brought to within 1.5 nm so that the two bilayers can join (Figure Q13–1). Assuming that the relevant portions of the two membranes at the fusion site are circular regions 1.5 nm in diameter, calculate the number of water molecules that would remain between the membranes. (Water is 55.5 M and the volume of a cylinder is  $\pi r^2 h$ .) Given that an average phospholipid

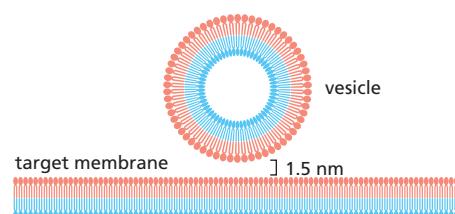


Figure Q13–1 Close approach of a vesicle and its target membrane in preparation for fusion (Problem 13–6).

occupies a membrane surface area of  $0.2 \text{ nm}^2$ , how many phospholipids would be present in each of the opposing monolayers at the fusion site? Are there sufficient water molecules to bind to the hydrophilic head groups of this number of phospholipids? (It is estimated that 10–12 water molecules are normally associated with each phospholipid head group at the exposed surface of a membrane.)

**13–7** SNAREs exist as complementary partners that carry out membrane fusions between appropriate vesicles and their target membranes. In this way, a vesicle with a particular variety of v-SNARE will fuse only with a membrane that carries the complementary t-SNARE. In some instances, however, fusions of identical membranes (homotypic fusions) are known to occur. For example, when a yeast cell forms a bud, vesicles derived from the mother cell's vacuole move into the bud where they fuse with one another to form a new vacuole. These vesicles carry both v-SNAREs and t-SNAREs. Are both types of SNAREs essential for this homotypic fusion event?

To test this point, you have developed an ingenious assay for fusion of vacuolar vesicles. You prepare vesicles from two different mutant strains of yeast: strain B has a defective gene for vacuolar alkaline phosphatase (Pase); strain A is defective for the protease that converts the precursor of alkaline phosphatase (pro-Pase) into its active form (Pase) (Figure Q13–2A). Neither strain has active alkaline phosphatase, but when extracts of the strains are mixed, vesicle fusion generates active alkaline phosphatase, which can be easily measured (Figure Q13–2).

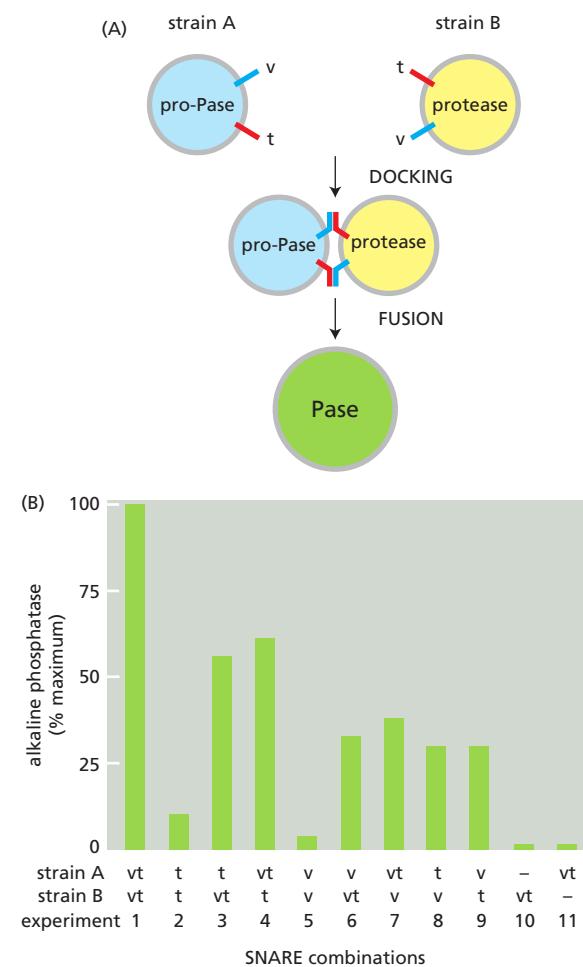
Now you delete the genes for the vacuolar v-SNARE, t-SNARE, or both in each of the two yeast strains. You prepare vacuolar vesicles from each and test them for their ability to fuse, as measured by the alkaline phosphatase assay (Figure Q13–2B).

What do these data say about the requirements for v-SNAREs and t-SNAREs in the fusion of vacuolar vesicles? Does it matter which kind of SNARE is on which vesicle?

**13–8** If you were to remove the ER retrieval signal from protein disulfide isomerase (PDI), which is normally a soluble resident of the ER lumen, where would you expect the modified PDI to be located?

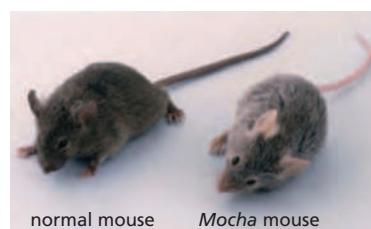
**13–9** The KDEL receptor must shuttle back and forth between the ER and the Golgi apparatus to accomplish its task of ensuring that soluble ER proteins are retained in the ER lumen. In which compartment does the KDEL receptor bind its ligands more tightly? In which compartment does it bind its ligands more weakly? What is thought to be the basis for its different binding affinities in the two compartments? If you were designing the system, in which compartment would you have the highest concentration of KDEL receptor? Would you predict that the KDEL receptor, which is a transmembrane protein, would itself possess an ER retrieval signal?

**13–10** How does the low pH of lysosomes protect the rest of the cell from lysosomal enzymes in case the lysosome breaks?



**Figure Q13–2** SNARE requirements for vesicle fusion (Problem 13–7). (A) Scheme for measuring the fusion of vacuolar vesicles. (B) Results of fusions of vesicles with different combinations of v-SNAREs and t-SNAREs. The SNAREs present on the vesicles of the two strains are indicated as v (v-SNARE) and t (t-SNARE).

**13–11** Melanosomes are specialized lysosomes that store pigments for eventual release by exocytosis. Various cells such as skin and hair cells then take up the pigment, which accounts for their characteristic pigmentation. Mouse mutants that have defective melanosomes often have pale or unusual coat colors. One such light-colored mouse, the *Mocha* mouse (Figure Q13–3), has a defect in the gene for one of the subunits of the adaptor protein complex AP3, which is associated with coated vesicles budding from the *trans* Golgi network. How might the loss of AP3 cause a defect in melanosomes?



**Figure Q13–3** A normal mouse and the *Mocha* mouse (Problem 13–11). In addition to its light coat color, the *Mocha* mouse has a poor sense of balance. (Courtesy of Margit Burmeister.)

## REFERENCES

### General

- Harrison SC & Kirchhausen T (2010) Structural biology: Conservation in vesicle coats. *Nature* 466, 1048–1049.
- Pfeffer SR (2013) A prize for membrane magic. *Cell* 155, 1203–1206.
- Thor F, Gautschi M, Geiger R & Helenius A (2009) Bulk flow revisited: transport of a soluble protein in the secretory pathway. *Traffic* 10, 1819–1830.

### The Molecular Mechanisms of Membrane Transport and the Maintenance of Compartmental Diversity

- Antony B (2011) Mechanisms of membrane curvature sensing. *Annu. Rev. Biochem.* 80, 101–123.
- Ferguson SM & De Camilli P (2012) Dynamin, a membrane-remodelling GTPase. *Nat. Rev. Mol. Cell Biol.* 13, 75–88.
- Frost A, Unger VM & De Camilli P (2009) The BAR domain superfamily: membrane-molding macromolecules. *Cell* 137, 191–196.
- Grosshans BL, Ortiz D & Novick P (2006) Rabs and their effectors: achieving specificity in membrane traffic. *Proc. Natl Acad. Sci. USA* 103, 11821–11827.
- Hughson FM (2010) Copy coats: COPI mimics clathrin and COPII. *Cell* 142, 19–21.
- Jackson LP, Kelly BT, McCoy AJ et al. (2010) A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell* 141, 1220–1229.
- Jahn R & Scheller RH (2006) SNAREs—engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* 7, 631–643.
- Jean S & Kiger AA (2012) Coordination between RAB GTPase and phosphoinositide regulation and functions. *Nat. Rev. Mol. Cell Biol.* 13, 463–470.
- Jin L, Pahuja KB, Wickliffe KE et al. (2012) Ubiquitin-dependent regulation of COPII coat size and function. *Nature* 482, 495–500.
- Martens S & McMahon HT (2008) Mechanisms of membrane fusion: disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* 9, 543–556.
- McNew JA, Parlati F, Fukuda R et al. (2000) Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* 407, 153–159.
- Miller EA & Schekman R (2013) COPII—a flexible vesicle formation system. *Curr. Opin. Cell Biol.* 25, 420–427.
- Pfeffer SR (2013) Rab GTPase regulation of membrane identity. *Curr. Opin. Cell Biol.* 25, 414–419.
- Saito K, Chen M, Bard F et al. (2009) TANGO1 facilitates cargo loading at endoplasmic reticulum exit sites. *Cell* 136, 891–902.
- Seaman MN (2005) Recycle your receptors with retromer. *Trends Cell Biol.* 15, 68–75.

### Transport from the ER Through the Golgi Apparatus

- Ellgaard L & Helenius A (2003) Quality control in the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.* 4, 181–191.
- Emr S, Glick BS, Linstedt AD et al. (2009) Journeys through the Golgi—taking stock in a new era. *J. Cell Biol.* 187, 449–453.
- Farquhar MG & Palade GE (1998) The Golgi apparatus: 100 years of progress and controversy. *Trends Cell Biol.* 8, 2–10.
- Ladinsky MS, Mastronarde DN, McIntosh JR et al. (1999) Golgi structure in three dimensions: functional insights from the normal rat kidney cell. *J. Cell Biol.* 144, 1135–1149.
- Munro S (2011) The golgin coiled-coil proteins of the Golgi apparatus. *Cold Spring Harb. Perspect. Biol.* 3, a005256.
- Pfeffer S (2010) How the Golgi works: a cisternal progenitor model. *Proc. Natl Acad. Sci. USA* 107, 19614–19618.
- Varki A (2011) Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb. Perspect. Biol.* 3, a005462.

### Transport from the Trans Golgi Network to Lysosomes

- Andrews NW (2000) Regulated secretion of conventional lysosomes. *Trends Cell Biol.* 10, 316–321.

- Bonifacino JS & Rojas R (2006) Retrograde transport from endosomes to the trans-Golgi network. *Nat. Rev. Mol. Cell Biol.* 7, 568–579.
- de Duve C (2005) The lysosome turns fifty. *Nat. Cell Biol.* 7, 847–849.
- Futerman AH & van Meer G (2004) The cell biology of lysosomal storage disorders. *Nat. Rev. Mol. Cell Biol.* 5, 554–565.
- Kraft C & Martens S (2012) Mechanisms and regulation of autophagosome formation. *Curr. Opin. Cell Biol.* 24, 496–501.
- Mizushima N, Yoshimori T & Ohsumi Y (2011) The role of Atg proteins in autophagosome formation. *Annu. Rev. Cell Dev. Biol.* 27, 107–132.
- Parzych KR & Klionsky DJ (2014) An overview of autophagy: morphology, mechanism, and regulation. *Antioxid. Redox Signal.* 20, 460–473.

### Transport into the Cell from the Plasma Membrane: Endocytosis

- Bonifacino JS & Traub LM (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* 72, 395–447.
- Brown MS & Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232, 34–47.
- Conner SD & Schmid SL (2003) Regulated portals of entry into the cell. *Nature* 422, 37–44.
- Doherty GJ & McMahon HT (2009) Mechanisms of endocytosis. *Annu. Rev. Biochem.* 78, 857–902.
- Howes MT, Mayor S & Parton RG (2010) Molecules, mechanisms, and cellular roles of clathrin-independent endocytosis. *Curr. Opin. Cell Biol.* 22, 519–527.
- Huotari J & Helenius A (2011) Endosome maturation. *EMBO J.* 30, 3481–3500.
- Hurley JH & Hanson PI (2010) Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nat. Rev. Mol. Cell Biol.* 11, 556–566.
- Kelly BT & Owen DJ (2011) Endocytic sorting of transmembrane protein cargo. *Curr. Opin. Cell Biol.* 23, 404–412.
- Maxfield FR & McGraw TE (2004) Endocytic recycling. *Nat. Rev. Mol. Cell Biol.* 5, 121–132.
- McMahon HT & Boucrot E (2011) Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* 12, 517–533.
- Mercer J & Helenius A (2012) Gulp rather than sipping: macropinocytosis as a way of virus entry. *Curr. Opin. Microbiol.* 15, 490–499.
- Sorkin A & von Zastrow M (2009) Endocytosis and signalling: intertwining molecular networks. *Nat. Rev. Mol. Cell Biol.* 10, 609–622.
- Tjelle TE, Lovdal T & Berg T (2000) Phagosome dynamics and function. *BioEssays* 22, 255–263.

### Transport from the Trans Golgi Network to the Cell Exterior: Exocytosis

- Burgess TL & Kelly RB (1987) Constitutive and regulated secretion of proteins. *Annu. Rev. Cell Biol.* 3, 243–293.
- Li F, Pincet F, Perez E et al. (2011) Complexin activates and clamps SNAREpins by a common mechanism involving an intermediate energetic state. *Nat. Struct. Mol. Biol.* 18, 941–946.
- Martin TF (1997) Stages of regulated exocytosis. *Trends Cell Biol.* 7, 271–276.
- Mellman I & Nelson WJ (2008) Coordinated protein sorting, targeting and distribution in polarized cells. *Nat. Rev. Mol. Cell Biol.* 9, 833–845.
- Mostov K, Su T & ter Beest M (2003) Polarized epithelial membrane traffic: conservation and plasticity. *Nat. Cell Biol.* 5, 287–293.
- Pang ZP & Südhof TC (2010) Cell biology of  $\text{Ca}^{2+}$ -triggered exocytosis. *Curr. Opin. Cell Biol.* 22, 496–505.
- Schuck S & Simons K (2004) Polarized sorting in epithelial cells: raft clustering and the biogenesis of the apical membrane. *J. Cell Sci.* 117, 5955–5964.

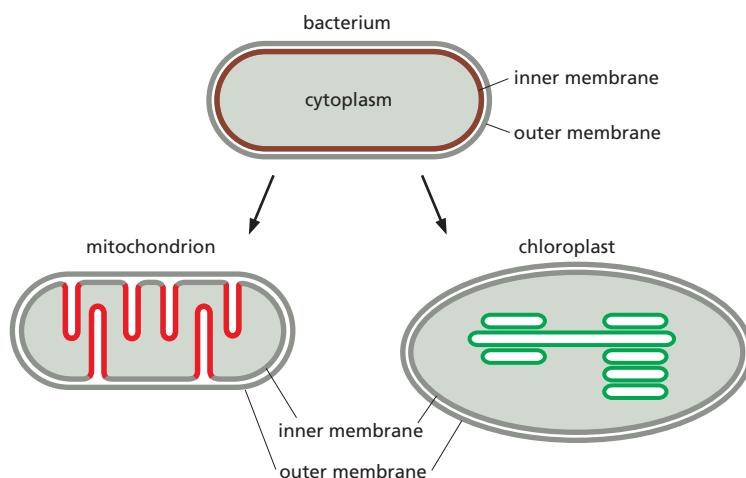
# Energy Conversion: Mitochondria and Chloroplasts

CHAPTER  
**14**

To maintain their high degree of organization in a universe that is constantly drifting toward chaos, cells have a constant need for a plentiful supply of ATP, as we have explained in Chapter 2. In eukaryotic cells, most of the ATP that powers life processes is produced by specialized, membrane-enclosed, *energy-converting organelles*. These are of two types. **Mitochondria**, which occur in virtually all cells of animals, plants, and fungi, burn food molecules to produce ATP by *oxidative phosphorylation*. **Chloroplasts**, which occur only in plants and green algae, harness solar energy to produce ATP by *photosynthesis*. In electron micrographs, the most striking features of both mitochondria and chloroplasts are their extensive internal membrane systems. These internal membranes contain sets of membrane protein complexes that work together to produce most of the cell's ATP. In bacteria, simpler versions of essentially the same protein complexes produce ATP, but they are located in the cell's plasma membrane (Figure 14-1).

Comparisons of DNA sequences indicate that the energy-converting organelles in present-day eukaryotes originated from prokaryotic cells that were endocytosed during the evolution of eukaryotes (discussed in Chapter 1). This explains why mitochondria and chloroplasts contain their own DNA, which still encodes a subset of their proteins. Over time, these organelles have lost most of their own genomes and become heavily dependent on proteins that are encoded by genes in the nucleus, synthesized in the cytosol, and then imported into the organelle. And the eukaryotic cells now rely on these organelles not only for the ATP they need for biosynthesis, solute transport, and movement, but also for many important biosynthetic reactions that occur inside each organelle.

The common evolutionary origin of the energy-converting machinery in mitochondria, chloroplasts, and prokaryotes (archaea and bacteria) is reflected in the fundamental mechanism that they share for harnessing energy. This is known as **chemiosmotic coupling**, signifying a link between the chemical bond-forming reactions that generate ATP ("chemi") and membrane transport processes



## IN THIS CHAPTER

### THE MITOCHONDRION

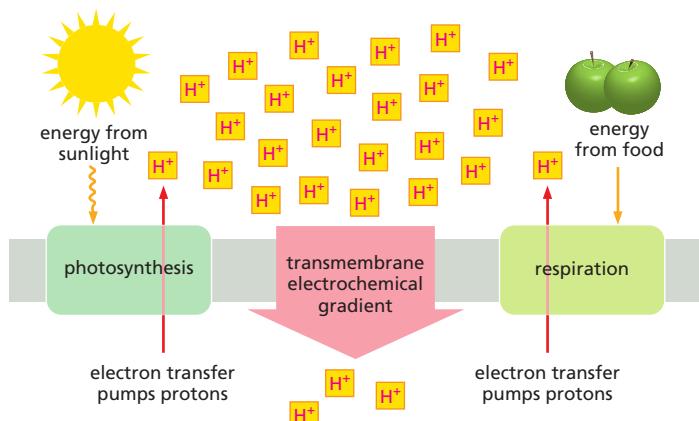
### THE PROTON PUMPS OF THE ELECTRON-TRANSPORT CHAIN

### ATP PRODUCTION IN MITOCHONDRIA

### CHLOROPLASTS AND PHOTOSYNTHESIS

### THE GENETIC SYSTEMS OF MITOCHONDRIA AND CHLOROPLASTS

**Figure 14-1** The membrane systems of bacteria, mitochondria, and chloroplasts are related. Mitochondria and chloroplasts are cell organelles that have originated from bacteria and have retained the bacterial energy-conversion mechanisms. Like their bacterial ancestors, mitochondria and chloroplasts have an outer and an inner membrane. Each of the membranes colored in this diagram contains energy-harvesting electron-transport chains. The deep invaginations of the mitochondrial inner membrane and the internal membrane system of the chloroplast harbor the machinery for cellular respiration and photosynthesis, respectively.



**Figure 14-2 Stage 1 of chemiosmotic coupling.** Energy from sunlight or the oxidation of food compounds is captured to generate an electrochemical proton gradient across a membrane. The electrochemical gradient serves as a versatile energy store that drives energy-requiring reactions in mitochondria, chloroplasts, and bacteria.

(“osmotic”). The chemiosmotic process occurs in two linked stages, both of which are performed by protein complexes in a membrane.

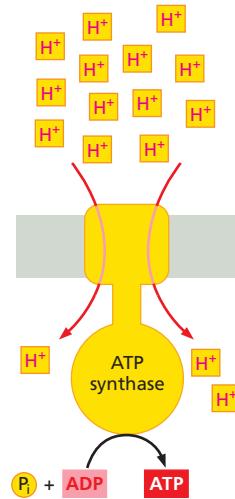
**Stage 1:** High-energy electrons (derived from the oxidation of food molecules, from pigments excited by sunlight, or from other sources described later) are transferred along a series of electron-transport protein complexes that form an *electron-transport chain* embedded in a membrane. Each electron transfer releases a small amount of energy that is used to pump protons ( $H^+$ ) and thereby generate a large *electrochemical gradient* across the membrane (Figure 14-2). As discussed in Chapter 11, such an electrochemical gradient provides a way of storing energy, and it can be harnessed to do useful work when ions flow back across the membrane.

**Stage 2:** The protons flow back down their electrochemical gradient through an elaborate membrane protein machine called *ATP synthase*, which catalyzes the production of ATP from ADP and inorganic phosphate ( $P_i$ ). This ubiquitous enzyme works like a turbine in the membrane, driven by protons, to synthesize ATP (Figure 14-3). In this way, the energy derived from food or sunlight in stage 1 is converted into the chemical energy of a phosphate bond in ATP.

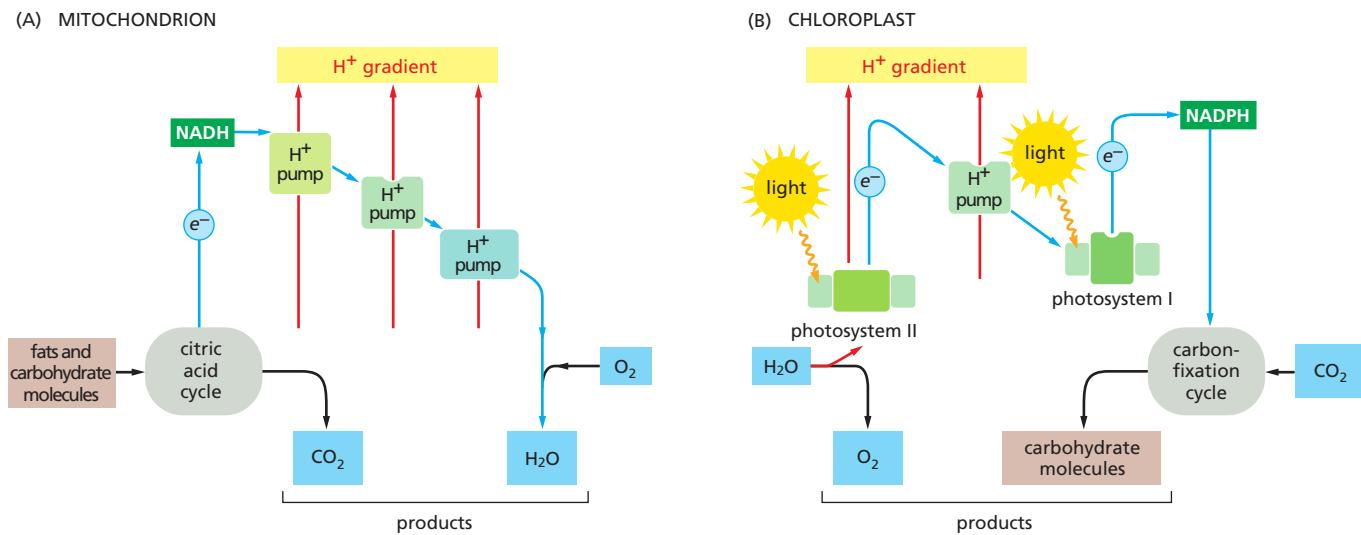
Electrons move through protein complexes in biological systems via tightly bound metal ions or other carriers that take up and release electrons easily, or by special small molecules that pick electrons up at one location and deliver them to another. For mitochondria, the first of these electron carriers is  $NAD^+$ , a water-soluble small molecule that takes up two electrons and one  $H^+$  derived from food molecules (fats and carbohydrates) to become NADH. NADH transfers these electrons from the sites where the food molecules are degraded to the inner mitochondrial membrane. There, the electrons from the energy-rich NADH are passed from one membrane protein complex to the next, passing to a lower-energy compound at each step, until they reach a final complex in which they combine with molecular oxygen ( $O_2$ ) to produce water. The energy released at each step as the electrons flow down this path from the energy-rich NADH to the low-energy water molecule drives  $H^+$  pumps in the inner mitochondrial membrane, utilizing three different membrane protein complexes. Together, these complexes generate the proton-motive force harnessed by ATP synthase to produce the ATP that serves as the universal energy currency throughout the cell (see Chapter 2).

Figure 14-4 compares the electron-transport processes in mitochondria, which harness energy from food molecules, with those in chloroplasts, which harness energy from sunlight. The energy-conversion systems of mitochondria and chloroplasts can be described in similar terms, and we shall see later in the chapter that two of their key components are closely related. One of these is the ATP synthase, and the other is a proton pump (colored green in Figure 14-4).

Among the crucial constituents that are unique to photosynthetic organisms are the two *photosystems*. These use the green pigment chlorophyll to capture



**Figure 14-3 Stage 2 of chemiosmotic coupling.** An ATP synthase (yellow) embedded in the lipid bilayer of a membrane harnesses the electrochemical proton gradient across the membrane, using it as a local energy store to drive ATP synthesis. The red arrows show the direction of proton movement through the ATP synthase.



**Figure 14-4 Electron-transport processes.** (A) The mitochondrion converts energy from chemical fuels. (B) The chloroplast converts energy from sunlight. In both cases, electron flow is indicated by blue arrows. Each of the protein complexes (green) is embedded in a membrane. In the mitochondrion, fats and carbohydrates from food molecules are fed into the citric acid cycle and provide electrons to generate the energy-rich compound NADH from  $NAD^+$ . These electrons then flow down an energy gradient as they pass from one complex to the next in the electron-transport chain, until they combine with molecular  $O_2$  in the last complex to produce water. The energy released at each stage is harnessed to pump  $H^+$  across the membrane. In the chloroplast, by contrast, electrons are extracted from water through the action of light in the photosystem II complex and molecular  $O_2$  is released. The electrons pass on to the next complex in the chain, which uses some of their energy to pump protons across the membrane, before passing to photosystem I, where sunlight generates high-energy electrons that combine with  $NADP^+$  to produce NADPH. NADPH then enters the carbon-fixation cycle along with  $CO_2$  to generate carbohydrates.

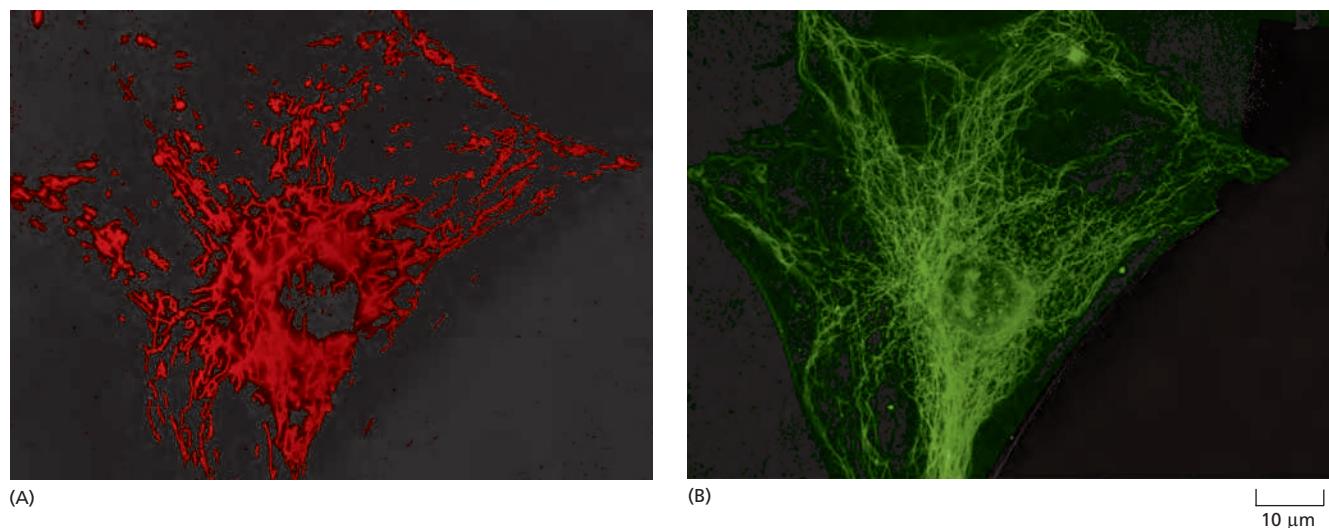
light energy and power the transfer of electrons, not unlike a photocell in a solar panel. The chloroplasts drive electron transfer in the direction opposite to that in mitochondria: electrons are taken from water to produce  $O_2$ , and these electrons are used (via NADPH, a molecule closely related to the NADH used in mitochondria) to synthesize carbohydrates from  $CO_2$  and water. These carbohydrates then serve as the source for all other compounds a plant cell needs.

Thus, both mitochondria and chloroplasts make use of an electron-transfer chain to produce an  $H^+$  gradient that powers reactions that are critical for the cell. However, chloroplasts generate  $O_2$  and take up  $CO_2$ , whereas mitochondria consume  $O_2$  and release  $CO_2$  (see Figure 14-4).

## THE MITOCHONDRION

Mitochondria occupy up to 20% of the cytoplasmic volume of a eukaryotic cell. Although they are often depicted as short, bacterium-like bodies with a diameter of 0.5–1  $\mu m$ , they are in fact remarkably dynamic and plastic, moving about the cell, constantly changing shape, dividing, and fusing (Movie 14.1). Mitochondria are often associated with the microtubular cytoskeleton (Figure 14-5), which determines their orientation and distribution in different cell types. Thus, in highly polarized cells such as neurons, mitochondria can move long distances (up to a meter or more in the extended axons of neurons), being propelled along the tracks of the microtubular cytoskeleton. In other cells, mitochondria remain fixed at points of high energy demand; for example, in skeletal or cardiac muscle cells, they pack between myofibrils, and in sperm cells they wrap tightly around the flagellum (Figure 14-6).

Mitochondria also interact with other membrane systems in the cell, most notably the endoplasmic reticulum (ER). Contacts between mitochondria and ER define specialized domains thought to facilitate the exchange of lipids between the two membrane systems. These contacts also appear to induce mitochondrial

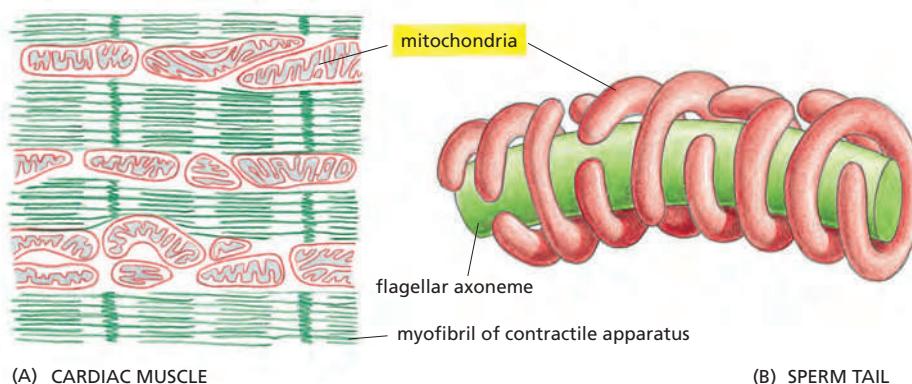


fission, which, as we discuss later, is involved in the distribution and partitioning of mitochondria within cells (**Figure 14–7**).

The acquisition of mitochondria was a prerequisite for the evolution of complex animals. Without mitochondria, present-day animal cells would have had to generate all of their ATP through anaerobic glycolysis. When glycolysis converts glucose to pyruvate, it releases only a small fraction of the total free energy that is potentially available from glucose oxidation (see Chapter 2). In mitochondria, the metabolism of sugars is complete: pyruvate is imported into the mitochondrion and ultimately oxidized by  $O_2$  to  $CO_2$  and  $H_2O$ , which allows 15 times more ATP to be made from a sugar than by glycolysis alone. As explained later, this became possible only when enough molecular oxygen accumulated in the Earth's atmosphere to allow organisms to take full advantage, via respiration, of the large amounts of energy potentially available from the oxidation of organic compounds.

Mitochondria are large enough to be seen in the light microscope, and they were first identified in the nineteenth century. Real progress in understanding their internal structure and function, however, depended on biochemical procedures developed in 1948 for isolating intact mitochondria, and on electron microscopy, which was first used to look at cells at about the same time.

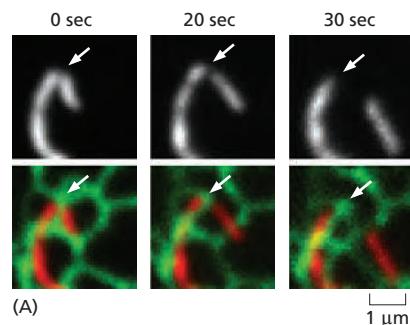
**Figure 14–5** The relationship between mitochondria and microtubules. (A) A light micrograph of chains of elongated mitochondria in a living mammalian cell in culture. The cell was stained with a fluorescent dye (rhodamine 123) that specifically labels mitochondria in living cells. (B) An immunofluorescence micrograph of the same cell stained (after fixation) with fluorescent antibodies that bind to microtubules. Note that the mitochondria tend to be aligned along microtubules. (Courtesy of Lan Bo Chen.)



**Figure 14–6** Localization of mitochondria near sites of high ATP demand in cardiac muscle and a sperm tail. (A) Cardiac muscle in the wall of the heart is the most heavily used muscle in the body, and its continual contractions require a reliable energy supply. It has limited built-in energy stores and has to depend on a steady supply of ATP from the copious mitochondria aligned close to the contractile myofibrils (see Figure 16–32). (B) During sperm development, microtubules wind helically around the flagellar axoneme, where they are thought to help localize the mitochondria in the tail to produce the structure shown.

**Figure 14–7 Interaction of mitochondria with the endoplasmic reticulum.**

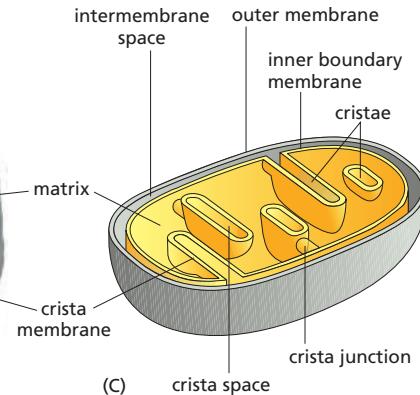
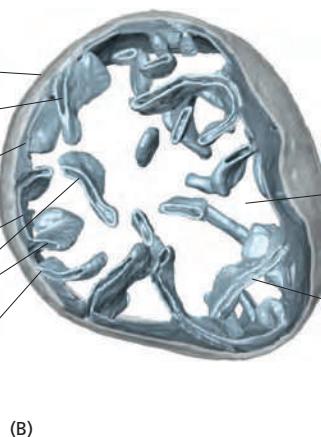
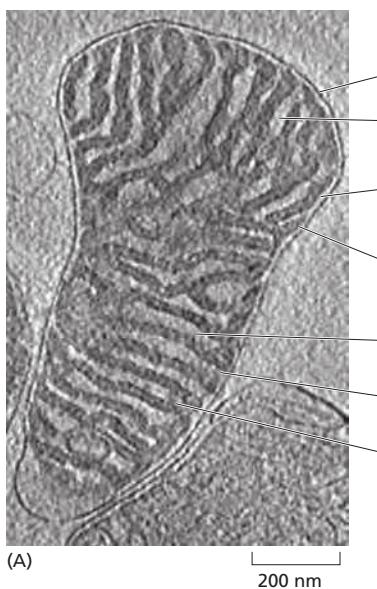
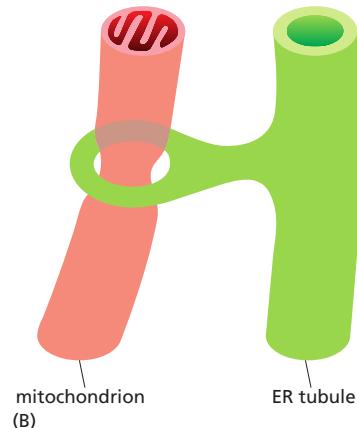
(A) Fluorescence light microscopy shows that tubules of the ER (green) wrap around parts of the mitochondrial network (red) in mammalian cells. The mitochondria then divide at the contact sites. After contact is established, fission occurs within less than a minute, as indicated by time-lapse microscopy. (B) Schematic drawing of an ER tubule wrapped around part of the mitochondrial reticulum. It is thought that ER-mitochondrial contacts also mediate the exchange of lipids between the two membrane systems. (A, adapted from J.R. Friedman et al., *Science* 334:358–362, 2011.)



## The Mitochondrion Has an Outer Membrane and an Inner Membrane

Like the bacteria from which they originated, mitochondria have an outer and an inner membrane. The two membranes have distinct functions and properties, and delineate separate compartments within the organelle. The inner membrane, which surrounds the internal **mitochondrial matrix** compartment (**Figure 14–8**), is highly folded to form invaginations known as **cristae** (the singular is **crista**), which contain in their membranes the proteins of the electron-transport chain. Where the inner membrane runs parallel to the outer membrane, between the cristae, it is known as the **inner boundary membrane**. The narrow (20–30 nm) gap between the inner boundary membrane and the outer membrane is known as the **intermembrane space**. The cristae are about 20 nm-wide membrane discs or tubules that protrude deeply into the matrix and enclose the **crista space**. The **crista membrane** is continuous with the inner boundary membrane, and where their membranes join, the membrane forms narrow membrane tubes or slits, known as **crista junctions**.

Like the bacterial outer membrane, the **outer mitochondrial membrane** is freely permeable to ions and to small molecules as large as 5000 daltons. This



**Figure 14–8 Structure of a mitochondrion.** (A) Tomographic slice through a three-dimensional map of a mouse heart mitochondrion determined by electron-microscope tomography. The outer membrane envelops the inner boundary membrane. The inner membrane is highly folded into tubular or lamellar cristae, which crisscross the matrix. The dense matrix, which contains most of the mitochondrial protein, appears dark in the electron microscope, whereas the intermembrane space and the crista space appear light due to their lower protein content. The inner boundary membrane follows the outer membrane closely at a distance of  $\approx$ 20 nm. The inner membrane turns sharply at the cristae junctions, where the cristae join the inner boundary membrane. (B) Tomographic surface-rendered portion of a yeast mitochondrion, showing how flattened cristae project into the matrix from the inner membrane (Movie 14.2). (C) Schematic drawing of a mitochondrion showing the outer membrane (gray), and the inner membrane (yellow). Note that the inner membrane is compartmentalized into the inner boundary membrane and the crista membrane. There are three distinct spaces: the inner membrane space, the crista space, and the matrix. (A, courtesy of Tobias Brandt; B, from K. Davies et al., *Proc. Natl. Acad. Sci. USA* 109:13602–13607, 2012. With permission from the National Academy of Sciences.)

**Figure 14–9 Biochemical fractionation of purified mitochondria into separate components.** Large numbers of mitochondria are isolated from homogenized tissue by centrifugation and then suspended in a medium of low osmotic strength. In such a medium, water flows into mitochondria and greatly expands the matrix space (yellow). While the cristae of the inner membrane unfold to accommodate the swelling, the outer membrane—which has no folds—breaks, releasing structures composed of the inner membrane surrounding the matrix. These techniques have made it possible to study the protein composition of the inner membrane (comprising a mixture of cristae, boundary membranes, and cristae junctions), the outer membrane, and the matrix.

is because it contains many porin molecules, a special class of  $\beta$ -barrel-type membrane protein that creates aqueous pores across the membrane (see Figure 10–23). As a consequence, the intermembrane space between the outer and inner membrane has the same pH and ionic composition as the cytoplasm, and there is no electrochemical gradient across the outer membrane.

If purified mitochondria are gently disrupted and then fractionated (Figure 14–9), the biochemical composition of membranes and mitochondrial compartments can be determined.

### The Inner Membrane Cristae Contain the Machinery for Electron Transport and ATP Synthesis

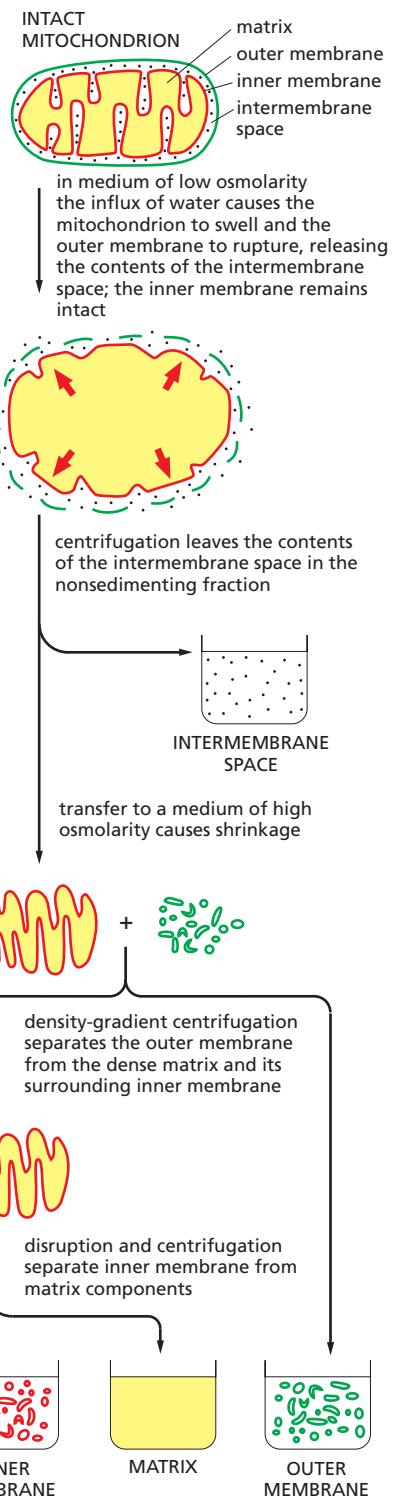
Unlike the outer mitochondrial membrane, the **inner mitochondrial membrane** is a diffusion barrier to ions and small molecules, just like the bacterial inner membrane. However, selected ions, most notably protons and phosphate, as well as essential metabolites such as ATP and ADP, can pass through it by means of special transport proteins.

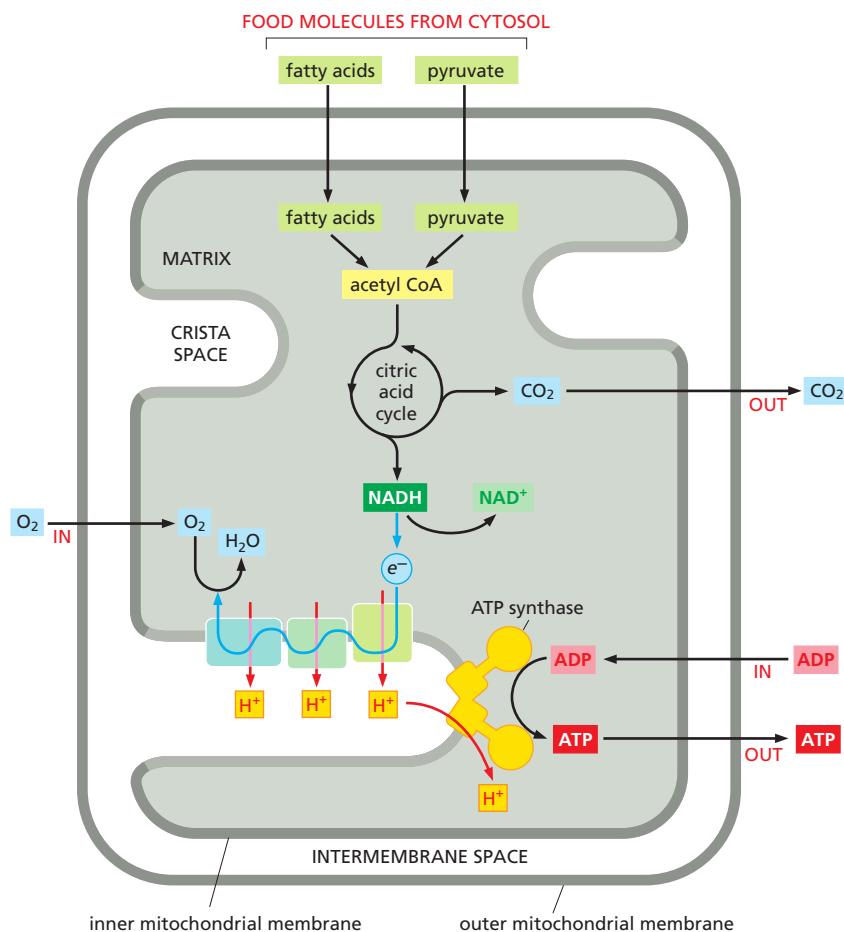
The inner mitochondrial membrane is highly differentiated into functionally distinct regions with different protein compositions. As discussed in Chapter 10, the lateral segregation of membrane regions with different protein and lipid compositions is a key feature of cells. In the inner mitochondrial membrane, the boundary membrane region is thought to contain the machinery for protein import, new membrane insertion, and assembly of the respiratory-chain complexes. The membranes of the cristae, which are continuous with the boundary membrane, contain the ATP synthase enzyme that produces most of the cell's ATP; they also contain the large protein complexes of the **respiratory chain**—the name given to the mitochondrion's electron-transport chain.

At the cristae junctions, where the membranes of the cristae join the boundary membrane, special protein complexes provide a diffusion barrier that segregates the membrane proteins in the two regions of the inner membrane; these complexes are also thought to anchor the cristae to the outer membrane, thus maintaining the highly folded topology of the inner membrane. Cristae membranes have one of the highest protein densities of all biological membranes, with a lipid content of 25% and a protein content of 75% by weight. The folding of the inner membrane into cristae greatly increases the membrane area available for oxidative phosphorylation. In highly active cardiac muscle cells, for example, the total area of cristae membranes can be up to 20 times larger than the area of the cell's plasma membrane. In total, the surface area of cristae membranes in each human body adds up to roughly the size of a football field.

### The Citric Acid Cycle in the Matrix Produces NADH

Together with the cristae that project into it, the matrix is the major working part of the mitochondrion. Mitochondria can use both pyruvate and fatty acids as fuel. Pyruvate is derived from glucose and other sugars, whereas fatty acids are derived from fats. Both of these fuel molecules are transported across the inner mitochondrial membrane by specialized transport proteins, and they are then converted to the crucial metabolic intermediate *acetyl CoA* by enzymes located in the mitochondrial matrix (see Chapter 2).





**Figure 14–10** A summary of the energy-converting metabolism in mitochondria.

Pyruvate and fatty acids enter the mitochondrion (*top of the figure*) and are broken down to acetyl CoA. The acetyl CoA is metabolized by the citric acid cycle, which reduces NAD<sup>+</sup> to NADH, which then passes its high-energy electrons to the first complex in the electron-transport chain. In the process of oxidative phosphorylation, these electrons pass along the electron-transport chain in the inner membrane cristae to oxygen (O<sub>2</sub>). This electron transport generates a proton gradient, which drives the production of ATP by the ATP synthase (see Figure 14–3). Electrons from the oxidation of succinate, a reaction intermediate in the citric acid cycle (see Panel 2–9, pp. 106–107), take a separate path to enter this electron-transport chain (not shown, see p. 772).

The membranes that comprise the mitochondrial inner membrane—the inner boundary membrane and the crista membrane—contain different mixtures of proteins and they are therefore shaded differently in this diagram.

The acetyl groups in acetyl CoA are oxidized in the matrix via the *citric acid cycle*, also called the Krebs cycle (see Figure 2–57 and Movie 2.6). The oxidation of these carbon atoms in acetyl CoA produces CO<sub>2</sub>, which diffuses out of the mitochondrion to be released to the environment as a waste product. More importantly, the citric acid cycle saves a great deal of the bond energy released by this oxidation in the form of electrons carried by NADH. This NADH transfers its electrons from the matrix to the electron-transport chain in the inner mitochondrial membrane, where—through the *chemiosmotic coupling* process described previously (see Figures 14–2 and 14–3)—the energy that was carried by NADH electrons is converted into phosphate-bond energy in ATP. **Figure 14–10** outlines this sequence of reactions schematically.

The matrix contains the genetic system of the mitochondrion, including the mitochondrial DNA and the ribosomes. The mitochondrial DNA (see section on genetic systems, p. 800) is organized into compact bodies—the nucleoids—by special scaffolding proteins that also function as transcription regulatory proteins. The large number of enzymes required for the maintenance of the mitochondrial genetic system, as well as for many other essential reactions to be outlined next, accounts for the very high protein concentration in the matrix; at more than 500 mg/mL, this concentration is close to that in a protein crystal.

### Mitochondria Have Many Essential Roles in Cellular Metabolism

Mitochondria not only generate most of the cell's ATP; they also provide many other essential resources for biosynthesis and cell growth. Before describing in detail the remarkable machinery of the respiratory chain, we diverge briefly to touch on some of these important roles.

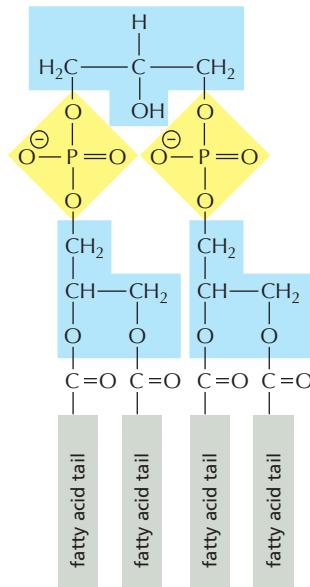
Mitochondria are critical for buffering the redox potential in the cytosol. Cells need a constant supply of the electron acceptor NAD<sup>+</sup> for the central reaction in glycolysis that converts glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate (see Figure 2–48). This NAD<sup>+</sup> is converted to NADH in the process, and the NAD<sup>+</sup> needs to be regenerated by transferring the high-energy NADH electrons somewhere. The NADH electrons will eventually be used to help drive oxidative phosphorylation inside the mitochondrion. But the inner mitochondrial membrane is impermeable to NADH. The electrons are therefore passed from the NADH to smaller molecules in the cytosol that can move through the inner mitochondrial membrane. Once in the matrix, these smaller molecules transfer their electrons to NAD<sup>+</sup> to form mitochondrial NADH, after which they are returned to the cytosol for recharging—creating a so-called *shuttle system* for the NADH electrons.

In addition to ATP, biosynthesis in the cytosol requires both a constant supply of reducing power in the form of NADPH and small carbon-rich molecules to serve as building blocks (discussed in Chapter 2). Descriptions of biosynthesis often state that the needed carbon skeletons come directly from the breakdown of sugars, whereas the NADPH is produced in the cytosol by a side pathway for the breakdown of sugars (the *pentose phosphate pathway*, an alternative to glycolysis). But under conditions where nutrients abound and plenty of ATP is available, mitochondria help to generate both the reducing power and the carbon-rich building blocks (the “carbon skeletons” in Panel 2–1, pp. 90–91) needed for cell growth. For this purpose, excess citrate produced in the mitochondrial matrix by the citric acid cycle (see Panel 2–9, pp. 106–107) is transported down its electrochemical gradient to the cytosol, where it is metabolized to produce essential components of the cell. Thus, for example, as part of a cell’s response to growth signals, large amounts of acetyl CoA are produced in the cytosol from citrate exported from mitochondria, accelerating the production of the fatty acids and sterols that build new membranes (described in Chapter 10). Cancer cells are frequently mutated in ways that enhance this pathway, as part of their program of abnormal growth (see Figure 20–26).

The urea cycle is a central metabolic pathway in mammals that converts the ammonia (NH<sub>4</sub><sup>+</sup>) produced by the breakdown of nitrogen-containing compounds (such as amino acids) to the urea excreted in urine. Two critical steps of the urea cycle are carried out in the mitochondria of liver cells, while the remaining steps occur in the cytosol. Mitochondria also play an essential part in the metabolic adaptation of cells to different nutritional conditions. For example, under conditions of starvation, proteins in our bodies are broken down to amino acids, and the amino acids are imported into mitochondria and oxidized to produce NADH for ATP production.

The biosynthesis of *heme groups*—which, as we shall see in the next section, play a central part in electron transfer—is another critical process that is shared between the mitochondrion and the cytoplasm. Iron-sulfur clusters, which are essential not only for electron transfer in the respiratory chain (see p. 766), but also for the maintenance and stability of the nuclear genome, are produced in mitochondria (and chloroplasts). Nuclear genome instability, a hallmark of cancer, can sometimes be linked to the decreased function of cellular proteins that contain iron-sulfur clusters.

Mitochondria also have a central role in membrane biosynthesis. Cardiolipin is a two-headed phospholipid (Figure 14–11) that is confined to the inner mitochondrial membrane, where it is also produced. But mitochondria are also a major source of phospholipids for the biogenesis of other cell membranes. Phosphatidylethanolamine, phosphatidylglycerol, and phosphatidic acid are synthesized in the mitochondrion, while phosphatidylinositol, phosphatidylcholine, and phosphatidylserine are primarily synthesized in the endoplasmic reticulum (ER). As described in Chapter 12, most of the cell’s membranes are assembled in the ER. The exchange of lipids between the ER and mitochondria is thought to occur at special sites of close contact (see Figure 14–7) by an as-yet unknown mechanism.



**Figure 14–11** The structure of cardiolipin. Cardiolipin consists of two covalently linked phospholipid units, with a total of four rather than the usual two fatty acid chains (see Figure 10–3). Cardiolipin is only produced in the mitochondrial inner membrane, where it interacts closely with membrane proteins involved in oxidative phosphorylation and ATP transport. In cristae, its two juxtaposed phosphate groups may act as a local proton trap on the membrane surface.

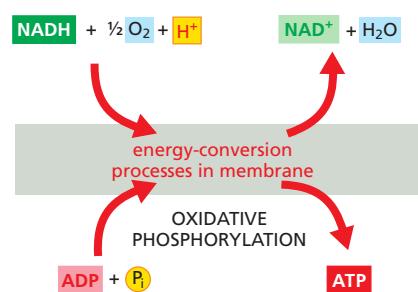
Finally, mitochondria are important calcium buffers, taking up calcium from the ER and sarcoplasmic reticulum at special membrane junctions. Cellular calcium levels control muscle contraction (see Chapter 16) and alterations are implicated in neurodegeneration and apoptosis. Clearly, cells and organisms depend on mitochondria in many different ways.

We now return to the central function of the mitochondrion in respiratory ATP generation.

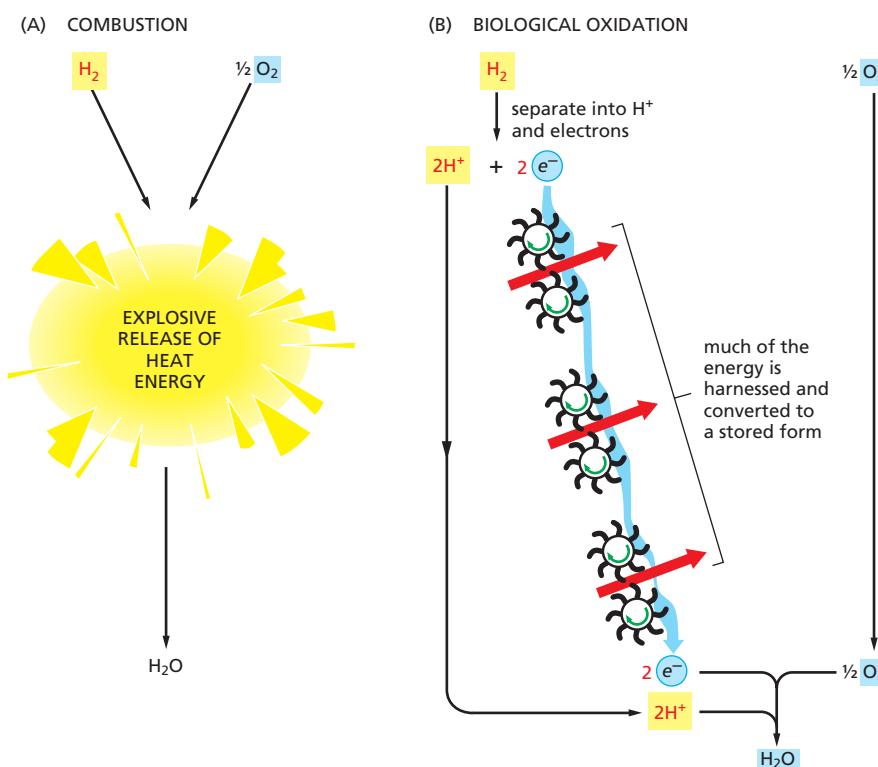
### A Chemiosmotic Process Couples Oxidation Energy to ATP Production

Although the citric acid cycle that takes place in the mitochondrial matrix is considered to be part of aerobic metabolism, it does not itself use oxygen. Only the final step of oxidative metabolism consumes molecular oxygen ( $O_2$ ) directly (see Figure 14–10). Nearly all the energy available from metabolizing carbohydrates, fats, and other foodstuffs in earlier stages is saved in the form of energy-rich compounds that feed electrons into the respiratory chain in the inner mitochondrial membrane. These electrons, most of which are carried by NADH, finally combine with  $O_2$  at the end of the respiratory chain to form water. The energy released during the complex series of electron transfers from NADH to  $O_2$  is harnessed in the inner membrane to generate an electrochemical gradient that drives the conversion of  $ADP + P_i$  to ATP. For this reason, the term **oxidative phosphorylation** is used to describe this final series of reactions (Figure 14–12).

The total amount of energy released by biological oxidation in the respiratory chain is equivalent to that released by the explosive combustion of hydrogen when it combines with oxygen in a single step to form water. But the combustion of hydrogen in a single-step chemical reaction, which has a strongly negative  $\Delta G$ , releases this large amount of energy unproductively as heat. In the respiratory chain, the same energetically favorable reaction  $H_2 + \frac{1}{2} O_2 \rightarrow H_2O$  is divided into small steps (Figure 14–13). This stepwise process allows the cell to store nearly half of the total energy that is released in a useful form. At each step, the electrons, which can be thought of as having been removed from a hydrogen molecule to



**Figure 14–12** The major net energy conversion catalyzed by the mitochondrion. In the process of oxidative phosphorylation, the mitochondrial inner membrane serves as a device that changes one form of chemical-bond energy to another, converting a major part of the energy of NADH oxidation into phosphate-bond energy in ATP.



**Figure 14–13** A comparison of biological oxidation with combustion. (A) If hydrogen were simply burned, nearly all of the energy would be released in the form of heat. (B) In biological oxidation reactions, about half of the released energy is stored in a form useful to the cell by means of the electron-transport chain (the respiratory chain) in the crista membrane of the mitochondrion. Only the rest of the energy is released as heat. In the cell, the protons and electrons shown here as being derived from  $H_2$  are removed from hydrogen atoms that are covalently linked to NADH molecules.

produce two protons, pass through a series of electron carriers in the inner mitochondrial membrane. At each of three distinct steps along the way (marked by the three electron-transport complexes of the respiratory chain, see below), much of the energy is utilized for pumping protons across the membrane. At the end of the electron-transport chain, the electrons and protons recombine with molecular oxygen into water.

Water is a very low-energy molecule and is thus very stable; it can serve as an electron donor only when a large amount of energy from an external source is spent on splitting it into protons, electrons, and molecular oxygen. This is exactly what happens in oxygenic photosynthesis, where the external energy source is the sun, as we shall see later in the section on chloroplasts (p. 782).

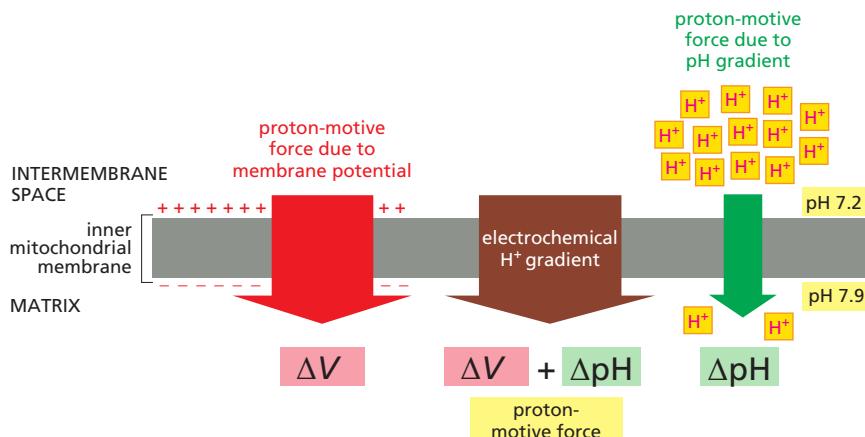
### The Energy Derived from Oxidation Is Stored as an Electrochemical Gradient

In mitochondria, the process of electron transport begins when two electrons and a proton are removed from NADH (to regenerate NAD<sup>+</sup>). These electrons are passed to the first of about 20 different electron carriers in the respiratory chain. The electrons start at a large negative redox potential (see Panel 14-1, p. 765)—that is, at a high energy level—which gradually drops as they pass along the chain. The proteins involved are grouped into three large *respiratory enzyme complexes*, each composed of protein subunits that sit in the inner mitochondrial membrane. Each complex in the chain has a higher affinity for electrons than its predecessor, and electrons pass sequentially from one complex to the next until they are finally transferred to molecular oxygen, which has the highest electron affinity of all.

The net result is the pumping of H<sup>+</sup> out of the matrix across the inner membrane, driven by the energetically favorable flow of electrons. This transmembrane movement of H<sup>+</sup> has two major consequences:

1. It generates a pH gradient across the inner mitochondrial membrane, with a high pH in the matrix (close to 8) and a lower pH in the intermembrane space. Since ions and small molecules equilibrate freely across the outer mitochondrial membrane, the pH in the intermembrane space is the same as in the cytosol (generally around pH 7.4).
2. It generates a voltage gradient across the inner mitochondrial membrane, creating a *membrane potential* with the matrix side negative and the crista space side positive.

The pH gradient ( $\Delta\text{pH}$ ) reinforces the effect of the membrane potential ( $\Delta V$ ), because the latter acts to attract any positive ion into the matrix and to push any negative ion out. Together,  $\Delta\text{pH}$  and  $\Delta V$  make up the **electrochemical gradient**, which is measured in units of millivolts (mV). This gradient exerts a **proton-motive force**, which tends to drive H<sup>+</sup> back into the matrix (Figure 14-14).



**Figure 14-14** The electrochemical proton gradient across the inner mitochondrial membrane. This gradient is composed of a large force due to the membrane potential ( $\Delta V$ ) and a smaller force due to the H<sup>+</sup> concentration gradient—that is, the pH gradient ( $\Delta\text{pH}$ ). Both forces combine to generate the proton-motive force, which pulls H<sup>+</sup> back into the mitochondrial matrix. The exact relationship between these forces is expressed by the Nernst equation (see Panel 11-1, p. 616).

The electrochemical gradient across the inner membrane of a respiring mitochondrion is typically about 180 mV (inside negative), and it consists of a membrane potential of about 150 mV and a pH gradient of about 0.5 to 0.6 pH units (each  $\Delta\text{pH}$  of 1 pH unit is equivalent to a membrane potential of about 60 mV). The electrochemical gradient drives not only ATP synthesis but also the transport of selected molecules across the inner mitochondrial membrane, including the import of selected proteins from the cytoplasm (discussed in Chapter 12).

### Summary

*The mitochondrion performs most cellular oxidations and produces the bulk of the animal cell's ATP. A mitochondrion has two separate membranes: the outer membrane and the inner membrane. The inner membrane surrounds the innermost space (the matrix) of the mitochondrion and forms the cristae, which project into the matrix. The matrix and the inner membrane cristae are the major working parts of the mitochondrion. The membranes that form cristae account for a major part of the membrane surface area in most cells, and they contain the mitochondrion's electron-transport chain (the respiratory chain).*

*The mitochondrial matrix contains a large variety of enzymes, including those that convert pyruvate and fatty acids to acetyl CoA and those that oxidize this acetyl CoA to  $\text{CO}_2$  through the citric acid cycle. These oxidation reactions produce large amounts of NADH, whose high-energy electrons are passed to the respiratory chain. The respiratory chain then uses the energy derived from transporting electrons from NADH to molecular oxygen to pump  $\text{H}^+$  out of the matrix. This produces a large electrochemical proton gradient across the inner mitochondrial membrane, composed of contributions from both a membrane potential and a pH difference. This electrochemical gradient exerts a force to drive  $\text{H}^+$  back into the matrix. This proton-motive force is harnessed both to produce ATP and for the selective transport of metabolites across the inner mitochondrial membrane.*

## THE PROTON PUMPS OF THE ELECTRON-TRANSPORT CHAIN

Having considered in general terms how a mitochondrion uses electron transport to generate a proton-motive force, we now turn to the molecular mechanisms that underlie this membrane-based energy-conversion process. In describing the respiratory chain of mitochondria, we accomplish the larger purpose of explaining how an electron-transport process can pump protons across a membrane. As stated at the beginning of this chapter, mitochondria, chloroplasts, archaea, and bacteria use very similar chemiosmotic mechanisms. In fact, these mechanisms underlie the function of all living organisms—including anaerobes that derive all their energy from electron transfers between two inorganic molecules, as we shall see later.

We start with some of the basic principles on which all of these processes depend.

### The Redox Potential Is a Measure of Electron Affinities

In chemical reactions, any electrons removed from one molecule are always passed to another, so that whenever one molecule is oxidized, another is reduced. As with any other chemical reaction, the tendency of such **redox reactions** to proceed spontaneously depends on the free-energy change ( $\Delta G$ ) for the electron transfer, which in turn depends on the relative affinities of the two molecules for electrons.

Because electron transfers provide most of the energy for life, it is worth taking the time to understand them. As discussed in Chapter 2, acids donate protons and bases accept them (see Panel 2-2, p. 93). Acids and bases exist in conjugate acid-base pairs, in which the acid is readily converted into the base by the loss of a

proton. For example, acetic acid ( $\text{CH}_3\text{COOH}$ ) is converted into its conjugate base, the acetate ion ( $\text{CH}_3\text{COO}^-$ ), in the reaction:



In an exactly analogous way, pairs of compounds such as NADH and  $\text{NAD}^+$  are called **redox pairs**, since NADH is converted to  $\text{NAD}^+$  by the loss of electrons in the reaction:



NADH is a strong electron donor: because two of its electrons are engaged in a covalent bond which releases energy when broken, the free-energy change for passing these electrons to many other molecules is favorable. Energy is required to form this bond from  $\text{NAD}^+$ , two electrons, and a proton (the same amount of energy that was released when the bond was broken). Therefore  $\text{NAD}^+$ , the redox partner of NADH, is of necessity a weak electron acceptor.

We can measure the tendency to transfer electrons from any redox pair experimentally. All that is required is the formation of an electrical circuit linking a 1:1 (equimolar) mixture of the redox pair to a second redox pair that has been arbitrarily selected as a reference standard, so that we can measure the voltage difference between them (Panel 14-1). This voltage difference is defined as the **redox potential**; electrons move spontaneously from a redox pair like NADH/ $\text{NAD}^+$  with a lower redox potential (a lower affinity for electrons) to a redox pair like  $\text{O}_2/\text{H}_2\text{O}$  with a higher redox potential (a higher affinity for electrons). Thus, NADH is a good molecule for donating electrons to the respiratory chain, while  $\text{O}_2$  is well suited to act as the “sink” for electrons at the end of the chain. As explained in Panel 14-1, the difference in redox potential,  $\Delta E'_0$ , is a direct measure of the standard free-energy change ( $\Delta G^\circ$ ) for the transfer of an electron from one molecule to another.

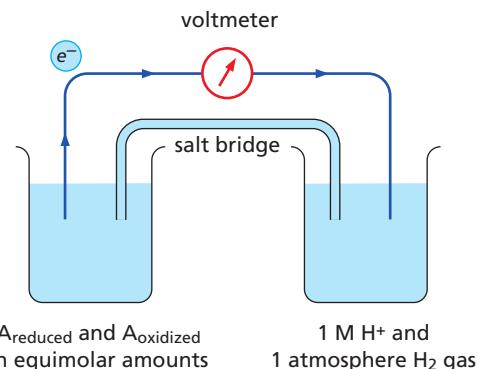
### Electron Transfers Release Large Amounts of Energy

As just discussed, those pairs of compounds that have the most negative redox potentials have the weakest affinity for electrons and therefore are useful as carriers with a strong tendency to donate electrons. Conversely, those pairs that have the most positive redox potentials have the greatest affinity for electrons and therefore are useful as carriers with a strong tendency to accept electrons. A 1:1 mixture of NADH and  $\text{NAD}^+$  has a redox potential of  $-320$  mV, indicating that NADH has a strong tendency to donate electrons; a 1:1 mixture of  $\text{H}_2\text{O}$  and  $\frac{1}{2}\text{O}_2$  has a redox potential of  $+820$  mV, indicating that  $\text{O}_2$  has a strong tendency to accept electrons. The difference in redox potential is  $1140$  mV, which means that the transfer of each electron from NADH to  $\text{O}_2$  under these standard conditions is enormously favorable, since  $\Delta G^\circ = -109$  kJ/mole, and twice this amount of energy is gained for the two electrons transferred per NADH molecule (see Panel 14-1). If we compare this free-energy change with that for the formation of the phospho-anhydride bonds in ATP, where  $\Delta G^\circ = 30.6$  kJ/mole (see Figure 2-50), we see that, under standard conditions, the oxidation of one NADH molecule releases more than enough energy to synthesize seven molecules of ATP from ADP and  $\text{P}_i$ . (In the cell, the number of ATP molecules generated will be lower because the standard conditions are far from the physiological ones; in addition, small amounts of energy are inevitably dissipated as heat along the way.)

### Transition Metal Ions and Quinones Accept and Release Electrons Readily

The electron-transport properties of the membrane protein complexes in the respiratory chain depend upon electron-carrying *cofactors*, most of which are *transition metals* such as Fe, Cu, Ni, and Mn, bound to proteins in the complexes. These metals have special properties that allow them to promote both enzyme catalysis and electron-transfer reactions. Most relevant here is the fact that their ions exist in several different oxidation states with closely spaced redox potentials, which enables them to accept or give up electrons readily; this property is

## HOW REDOX POTENTIALS ARE MEASURED

THE STANDARD REDOX POTENTIAL,  $E'_0$ 

The standard redox potential for a redox pair, defined as  $E'_0$ , is measured for a standard state where all of the reactants are at a concentration of 1 M, including  $\text{H}^+$ . Since biological reactions occur at pH 7, biologists instead define the standard state as  $\text{A}_{\text{reduced}} = \text{A}_{\text{oxidized}}$  and  $\text{H}^+ = 10^{-7}$  M. This standard redox potential is designated by the symbol  $E'_0$ , in place of  $E_0$ .

One beaker (left) contains substance A with an equimolar mixture of the reduced ( $\text{A}_{\text{reduced}}$ ) and oxidized ( $\text{A}_{\text{oxidized}}$ ) members of its redox pair. The other beaker contains the hydrogen reference standard ( $2\text{H}^+ + 2e^- \rightleftharpoons \text{H}_2$ ), whose redox potential is arbitrarily assigned as zero by international agreement. (A salt bridge formed from a concentrated KCl solution allows  $\text{K}^+$  and  $\text{Cl}^-$  to move between the beakers, as required to neutralize the charges when electrons flow between the beakers.) The metal wire (dark blue) provides a resistance-free path for electrons, and a voltmeter then measures the redox potential of substance A. If electrons flow from  $\text{A}_{\text{reduced}}$  to  $\text{H}^+$ , as indicated here, the redox pair formed by substance A is said to have a negative redox potential. If they instead flow from  $\text{H}_2$  to  $\text{A}_{\text{oxidized}}$ , the redox pair is said to have a positive redox potential.

examples of redox reactions	standard redox potential $E'_0$
$\text{NADH} \rightleftharpoons \text{NAD}^+ + \text{H}^+ + 2e^-$	-320 mV
reduced ubiquinone $\rightleftharpoons$ oxidized ubiquinone $+ 2\text{H}^+ + 2e^-$	+30 mV
reduced cytochrome c $\rightleftharpoons$ oxidized cytochrome c $+ e^-$	+230 mV
$\text{H}_2\text{O} \rightleftharpoons \frac{1}{2}\text{O}_2 + 2\text{H}^+ + 2e^-$	+820 mV

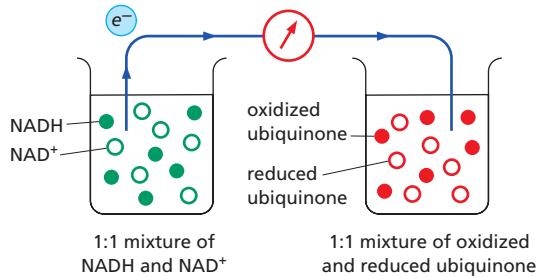
CALCULATION OF  $\Delta G^\circ$  FROM REDOX POTENTIALS

To determine the energy change for an electron transfer, the  $\Delta G^\circ$  of the reaction (kJ/mole) is calculated as follows:

$$\Delta G^\circ = -n(0.096) \Delta E'_0, \text{ where } n \text{ is the number of electrons transferred across a redox potential change of } \Delta E'_0 \text{ millivolts (mV), and}$$

$$\Delta E'_0 = E'_0(\text{acceptor}) - E'_0(\text{donor})$$

## EXAMPLE:



For the transfer of one electron from NADH to ubiquinone:

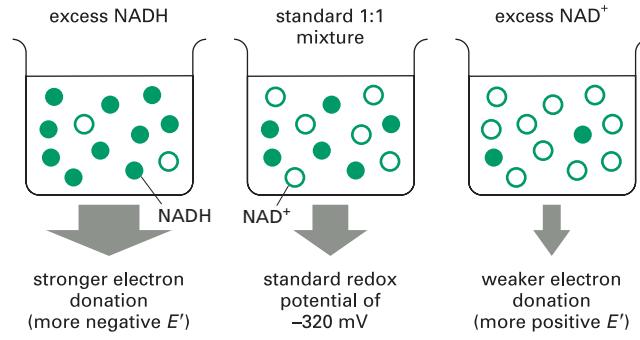
$$\Delta E'_0 = +30 - (-320) = +350 \text{ mV}$$

$$\Delta G^\circ = -n(0.096)\Delta E'_0 = -1(0.096)(350) = -34 \text{ kJ/mole}$$

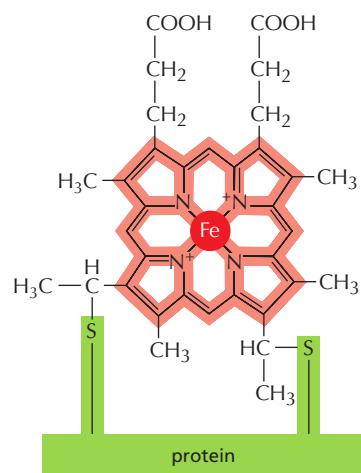
The same calculation reveals that the transfer of one electron from ubiquinone to oxygen has an even more favorable  $\Delta G^\circ$  of -76 kJ/mole. The  $\Delta G^\circ$  value for the transfer of one electron from NADH to oxygen is the sum of these two values, -110 kJ/mole.

## EFFECT OF CONCENTRATION CHANGES

As explained in Chapter 2 (see p. 60), the actual free-energy change for a reaction,  $\Delta G$ , depends on the concentration of the reactants and generally will be different from the standard free-energy change,  $\Delta G^\circ$ . The standard redox potentials are for a 1:1 mixture of the redox pair. For example, the standard redox potential of -320 mV is for a 1:1 mixture of NADH and  $\text{NAD}^+$ . But when there is an excess of NADH over  $\text{NAD}^+$ , electron transfer from NADH to an electron acceptor becomes more favorable. This is reflected by a more negative redox potential and a more negative  $\Delta G$  for electron transfer.



**Figure 14–15** The structure of the heme group attached covalently to cytochrome c. The porphyrin ring of the heme is shown in red. There are six different cytochromes in the respiratory chain. Because the hemes in different cytochromes have slightly different structures and are kept in different local environments by their respective proteins, each has a different affinity for an electron, and a slightly different spectroscopic signature.



exploited by the membrane protein complexes in the respiratory chain to move electrons both within and between complexes.

Unlike the colorless atoms H, C, N, and O that constitute the bulk of biological molecules, transition metal ions are often brightly colored, which makes the proteins that contain them easy to study by spectroscopic methods using visible light. One family of such colored proteins, the **cytochromes**, contains a bound *heme group*, in which an iron atom is tightly held by four nitrogen atoms at the corners of a square in a *porphyrin ring* (Figure 14–15). Similar porphyrin rings are responsible both for the red color of blood and for the green color of leaves, binding an iron in hemoglobin or a magnesium in chlorophyll, respectively.

Iron-sulfur proteins contain a second major family of electron-transfer cofactors. In this case, either two or four iron atoms are bound to an equal number of sulfur atoms and to cysteine side chains, forming **iron-sulfur clusters** in the protein (Figure 14–16). Like the cytochrome hemes, these clusters carry one electron at a time.

The simplest of the electron-transfer cofactors in the respiratory chain—and the only one that is not always bound to a protein—is a quinone (called *ubiquinone*, or *coenzyme Q*). A **quinone** (Q) is a small hydrophobic molecule that is freely mobile in the lipid bilayer. This *electron carrier* can accept or donate either one or two electrons. Upon reduction (note that reduced quinones are called quinols), it picks up a proton from water along with each electron (Figure 14–17).

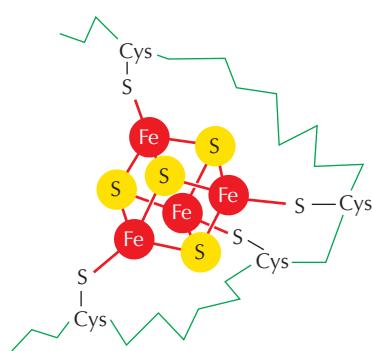
In the mitochondrial electron-transport chain, six different cytochrome hemes, eight iron-sulfur clusters, three copper atoms, a flavin mononucleotide (another electron-transfer cofactor), and ubiquinone work in a defined sequence to carry electrons from NADH to O<sub>2</sub>. In total, this pathway involves more than 60 different polypeptides arranged in three large membrane protein complexes, each of which binds several of the above electron-carrying cofactors.

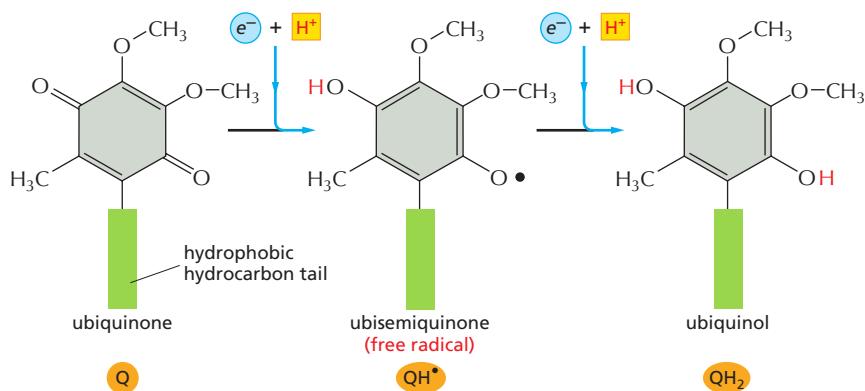
As we would expect, the electron-transfer cofactors have increasing affinities for electrons (higher redox potentials) as the electrons move along the respiratory chain. The redox potentials have been fine-tuned during evolution by the protein environment of each cofactor, which alters the cofactor's normal affinity for electrons. Because iron-sulfur clusters have a relatively low affinity for electrons, they predominate in the first half of the respiratory chain; in contrast, the heme cytochromes predominate further down the chain, where a higher electron affinity is required.

### NADH Transfers Its Electrons to Oxygen Through Three Large Enzyme Complexes Embedded in the Inner Membrane

Membrane proteins are difficult to purify because they are insoluble in aqueous solutions, and they are easily disrupted by the detergents that are required to solubilize them. But by using mild nonionic detergents, such as octylglucoside or dodecyl maltoside (see Figure 10–28), they can be solubilized and purified in their native form, and even crystallized for structure determination. Each of the three different detergent-solubilized respiratory-chain complexes can be re-inserted

**Figure 14–16** The structure of an iron-sulfur cluster. These dark brown clusters consist either of four iron and four sulfur atoms, as shown here, or of two irons and two sulfurs linked to cysteines in the polypeptide chain via covalent sulfur bridges, or to histidines. Although they contain several iron atoms, each iron-sulfur cluster can carry only one electron at a time. Nine different iron-sulfur clusters participate in electron transport in the respiratory chain.



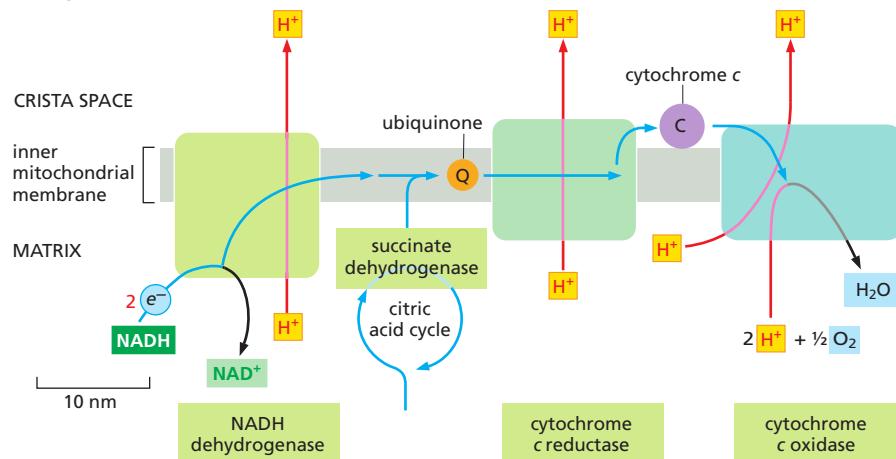


into artificial lipid bilayer vesicles and shown to pump protons across the membrane as electrons pass through them.

In the mitochondrion, the three complexes are linked in series, serving as electron-transport-driven H<sup>+</sup> pumps that pump protons out of the matrix to acidify the crista space (Figure 14-18):

1. The **NADH dehydrogenase complex** (often referred to as *Complex I*) is the largest of these respiratory enzyme complexes. It accepts electrons from NADH and passes them through a flavin mononucleotide and eight iron-sulfur clusters to the lipid-soluble electron carrier ubiquinone. The reduced ubiquinol then transfers its electrons to cytochrome *c* reductase.
2. The **cytochrome *c* reductase** (also called the *cytochrome b-c<sub>1</sub> complex*) is a large membrane protein assembly that functions as a dimer. Each monomer contains three cytochrome hemes and an iron-sulfur cluster. The complex accepts electrons from ubiquinol and passes them on to the small, soluble protein cytochrome *c*, which is located in the crista space and carries electrons one at a time to cytochrome *c* oxidase.
3. The **cytochrome *c* oxidase complex** contains two cytochrome hemes and three copper atoms. The complex accepts electrons one at a time from cytochrome *c* and passes them to molecular oxygen. In total, four electrons and four protons are needed to convert one molecule of oxygen to water.

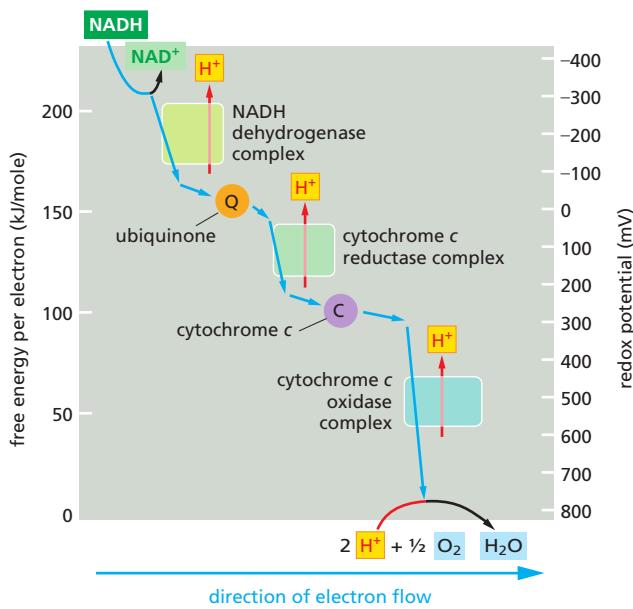
We have previously discussed how the redox potential reflects electron affinities. Figure 14-19 presents an outline of the redox potentials measured along the respiratory chain. These potentials change in three large steps, one across each proton-translocating respiratory complex. The change in redox potential between any two electron carriers is directly proportional to the free energy released when an electron transfers between them. Each complex acts as an energy-conversion device by harnessing some of this free-energy change to pump H<sup>+</sup> across the inner membrane, thereby creating an electrochemical proton gradient as electrons pass along the chain.



**Figure 14-17 Quinone electron carriers.** Ubiquinone in the lipid bilayer picks up one H<sup>+</sup> (red) from the aqueous environment for each electron (blue) it accepts, in two steps, from respiratory-chain complexes. The first step involves the acquisition of a proton and an electron and converts the ubiquinone into an unstable ubisemiquinone radical. In the second step, it becomes a fully reduced ubiquinone (called ubiquinol), which is freely mobile as an electron carrier in the lipid bilayer of the membrane. When the ubiquinol donates its electrons to the next complex in the chain, the two protons are released. The long hydrophobic tail (green) that confines ubiquinone to the membrane consists of 6–10 five-carbon isoprene units, depending on the organism. The corresponding electron carrier in the photosynthetic membranes of chloroplasts is plastoquinone, which has almost the same structure and works in the same way. For simplicity, we refer to both ubiquinone and plastoquinone in this chapter as quinone (abbreviated as Q).

**Figure 14-18 The path of electrons through the three respiratory-chain proton pumps.** (Movie 14.3) The approximate size and shape of each complex is shown. During the transfer of electrons from NADH to oxygen (blue arrows), ubiquinone and cytochrome *c* serve as mobile carriers that ferry electrons from one complex to the next. During the electron-transfer reactions, protons are pumped across the membrane by each of the respiratory enzyme complexes, as indicated (red arrows).

For historical reasons, the three proton pumps in the respiratory chain are sometimes denoted as Complex I, Complex III, and Complex IV, according to the order in which electrons pass through them from NADH. Electrons from the oxidation of succinate by succinate dehydrogenase (designated as Complex II) are fed into the electron-transport chain in the form of reduced ubiquinone. Although embedded in the crista membrane, succinate dehydrogenase does not pump protons and thus does not contribute to the proton-motive force; it is therefore not considered to be an integral part of the respiratory chain.



**Figure 14-19** Redox potential changes along the mitochondrial electron-transport chain. The redox potential (designated  $E'_0$ ) increases as electrons flow down the respiratory chain to oxygen. The standard free-energy change in kilojoules,  $\Delta G^\circ$ , for the transfer of each of the two electrons donated by an NADH molecule can be obtained from the left-hand ordinate [ $\Delta G^\circ = -n(0.096) \Delta E'_0$ , where  $n$  is the number of electrons transferred across a redox potential change of  $\Delta E'_0$  mV]. Electrons flow through a respiratory enzyme complex by passing in sequence through the multiple electron carriers in each complex (blue arrows). As indicated, part of the favorable free-energy change is harnessed by each enzyme complex to pump H<sup>+</sup> across the inner mitochondrial membrane (red arrows). The NADH dehydrogenase pumps up to four H<sup>+</sup> per electron, the cytochrome c reductase complex pumps two, whereas the cytochrome c oxidase complex pumps one per electron.

Note that NADH is not the only source of electrons for the respiratory chain. The flavin FADH<sub>2</sub>, which is generated by fatty acid oxidation (see Figure 2-56) and by succinate dehydrogenase in the citric acid cycle (see Figure 2-57), also contributes. Its two electrons are passed directly to ubiquinone, bypassing NADH dehydrogenase.

X-ray crystallography has elucidated the structure of each of the three respiratory-chain complexes in great detail, and we next examine each of them in turn to see how they work.

### The NADH Dehydrogenase Complex Contains Separate Modules for Electron Transport and Proton Pumping

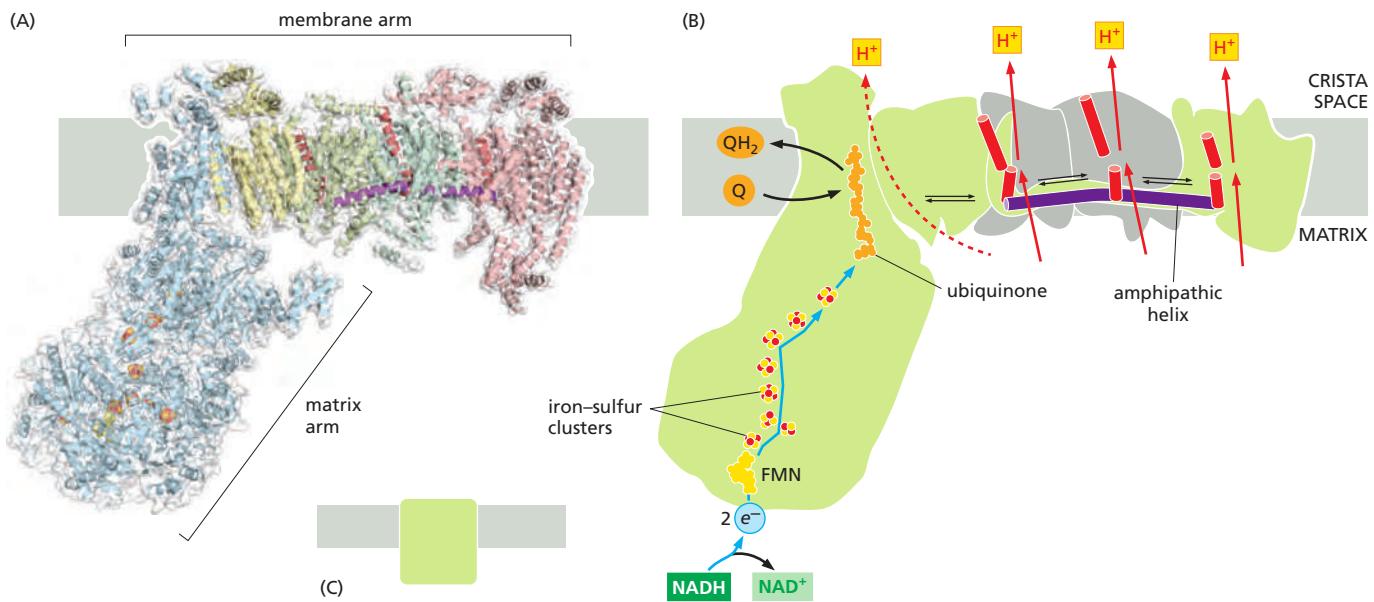
The NADH dehydrogenase complex is a massive assembly of membrane and nonmembrane proteins that receives electrons from NADH and passes them to ubiquinone. In animal mitochondria, it consists of more than 40 different protein subunits, with a molecular mass of nearly a million daltons. The x-ray structures of the NADH dehydrogenase complex from fungi and bacteria show that it is L-shaped, with both a hydrophobic membrane arm and a hydrophilic arm that projects into the mitochondrial matrix (Figure 14-20).

Electron transfer and proton pumping are physically separated in the NADH dehydrogenase complex, with electron transfer occurring in the matrix arm and proton pumping in the membrane arm. The NADH docks near the tip of the matrix arm, where it transfers its electrons via a bound flavin mononucleotide to a string of iron-sulfur clusters that runs down the arm, acting like a wire to carry electrons to a protein-bound molecule of ubiquinone. Electron transfer to the quinone is thought to trigger proton translocation in a set of proton pumps in the membrane arm, and for this to happen the two processes must be energetically and mechanically linked. A mechanical link is thought to be provided by a 6-nm long, amphipathic  $\alpha$  helix that runs parallel to the membrane surface on the matrix side of the membrane arm. This helix may act like the connecting rod in a steam engine to generate a mechanical, energy-transducing power stroke that links the quinone-binding site to the proton-translocating modules in the membrane (see Figure 14-20).

The reduction of each quinone by the transfer of two electrons can cause four protons to be pumped out of the matrix into the crista space. In this way, NADH dehydrogenase generates roughly half of the total proton-motive force in mitochondria.

### Cytochrome c Reductase Takes Up and Releases Protons on the Opposite Side of the Crista Membrane, Thereby Pumping Protons

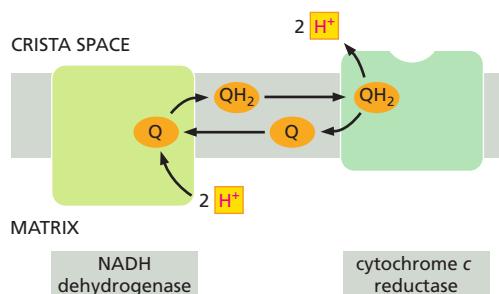
As described previously, when a quinone molecule (Q) accepts its two electrons, it also takes up two protons to form a quinol (QH<sub>2</sub>; see Figure 14-17). In



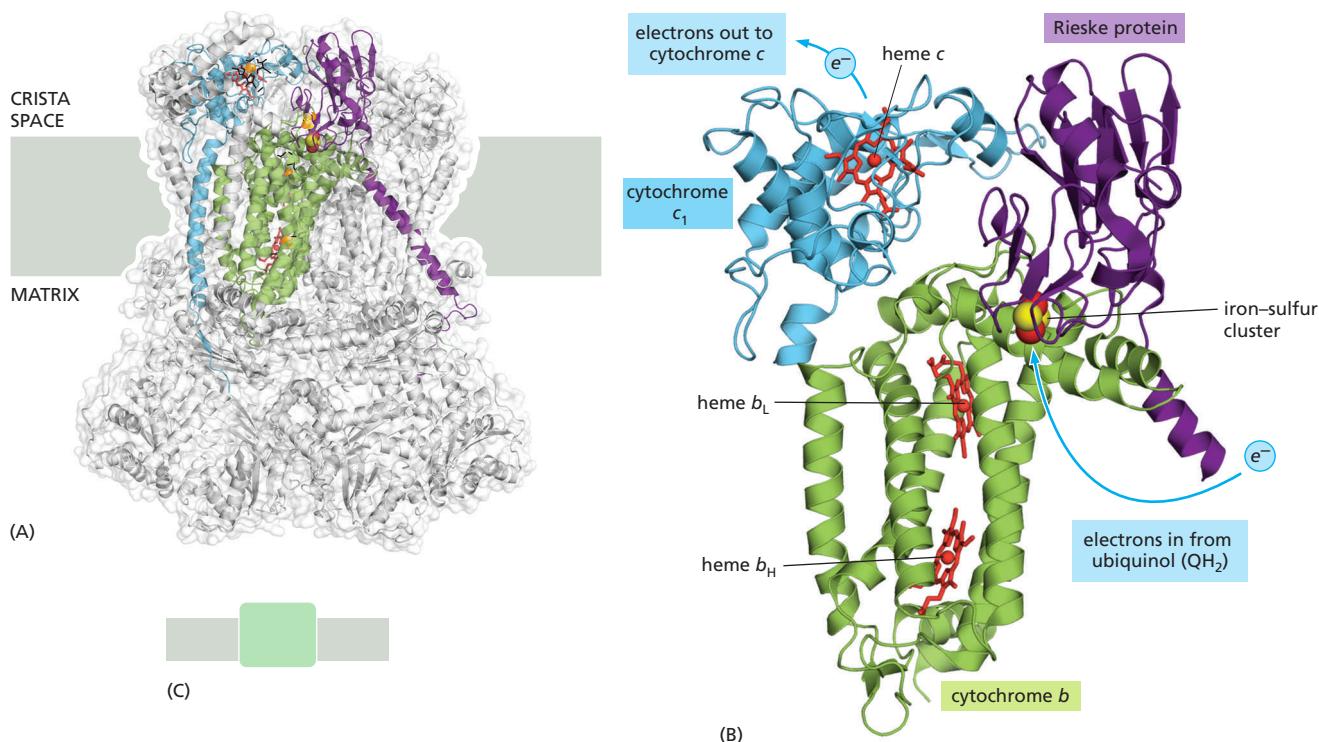
**Figure 14–20** The structure of NADH dehydrogenase. (A) The model of the mitochondrial complex shown here is based on the x-ray structure of the smaller bacterial complex, which works in the same way. The matrix arm of NADH dehydrogenase (also known as Complex I) contains eight iron–sulfur (FeS) clusters that appear to participate in electron transport. The membrane contains more than 70 transmembrane helices, forming three distinct proton-pumping modules, while the matrix arm contains the electron-transport cofactors. (B) NADH donates two electrons, via a bound flavin mononucleotide (FMN; yellow), to a chain of seven iron–sulfur clusters (red and yellow spheres). From the terminal iron–sulfur cluster, the electrons pass to ubiquinone (orange). Electron transfer results in conformational changes (black arrows) that are thought to be transmitted to a long amphipathic  $\alpha$  helix (purple) on the matrix side of the membrane arm, which pulls on discontinuous transmembrane helices (red) in three membrane subunits, each of which resembles an antiporter (see Chapter 11). This movement is thought to change the conformation of charged residues in the three proton channels, resulting in the translocation of three protons out of the matrix. A fourth proton may be translocated at the interface of the two arms (dotted line). (C) This shows the symbol for NADH dehydrogenase used throughout this chapter. (Adapted from R.G. Efremov, R. Baradaran and L.A. Sazanov, *Nature* 465:441–445, 2010. PDB code: 3M9S.)

the respiratory chain, ubiquinol transfers electrons from NADH dehydrogenase to cytochrome *c* reductase. Because the protons in this  $QH_2$  molecule are taken up from the matrix and released on the opposite side of the crista membrane, two protons are transferred from the matrix into the crista space per pair of electrons transferred (Figure 14–21). This vectorial transfer of protons supplements the electrochemical proton gradient that is created by the NADH dehydrogenase proton pumping just discussed.

Cytochrome *c* reductase is a large assembly of membrane protein subunits. Three subunits form a catalytic core that passes electrons from ubiquinol to cytochrome *c*, with a structure that has been highly conserved from bacterial ancestors (Figure 14–22). It pumps protons by a vectorial transfer of protons that involves a binding site for a second molecule of ubiquinone; the elaborate



**Figure 14–21** How a directional release and uptake of protons by a quinone pump pumps protons across a membrane. Two protons are picked up on the matrix side of the inner mitochondrial membrane when the reaction  $Q + 2e^- + 2H^+ \rightarrow QH_2$  is catalyzed by the NADH dehydrogenase complex. This molecule of ubiquinol ( $QH_2$ ) diffuses rapidly in the plane of the membrane, becoming bound to the crista side of cytochrome *c* reductase. When its oxidation by cytochrome *c* reductase generates two protons and two electrons (see Figure 14–17), the two protons are released into the crista space. The flow of electrons is not shown in this diagram.

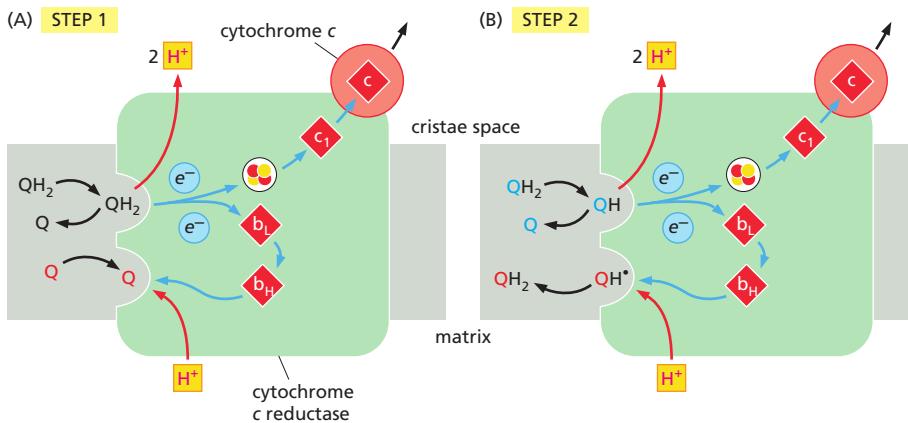


**Figure 14–22 The structure of cytochrome c reductase.** Cytochrome c reductase (also known as the cytochrome  $b-c_1$  complex) is a dimer of two identical 240,000-dalton halves, each composed of 11 different protein molecules in mammals. (A) A structure graphic of the entire dimer, showing in color the three proteins that form the functional core of the enzyme complex: cytochrome  $b$  (green) and cytochrome  $c_1$  (blue) are colored in one half, and the Rieske protein (purple) containing an  $\text{Fe}_2\text{S}_2$  iron–sulfur cluster (red and yellow) is colored in the other. These three protein subunits interact across the two halves. (B) Transfer of electrons through cytochrome c reductase to the small, soluble carrier protein cytochrome  $c$ . Electrons entering from ubiquinol near the matrix side of the membrane are captured by the iron–sulfur cluster of the Rieske protein, which moves its iron–sulfur group back and forth to transfer these electrons to heme  $c$  (red). Heme  $c$  then transfers them to the carrier molecule cytochrome  $c$ . (C) This shows the symbol for cytochrome c reductase used throughout this chapter. (PDB code: 1EZV.)

redox loop mechanism used is called the *Q cycle* because while one of the electrons received from each  $\text{QH}_2$  molecule is transferred from ubiquinone through the complex to the carrier protein cytochrome  $c$ , the other electron is recycled back into the quinone pool. Through the mechanism illustrated in Figure 14–23, the Q cycle increases the total amount of redox energy that can be stored in the electrochemical proton gradient. As a result, two protons are pumped across the crista membrane for every electron that is transferred from NADH dehydrogenase to cytochrome  $c$ .

### The Cytochrome $c$ Oxidase Complex Pumps Protons and Reduces $\text{O}_2$ Using a Catalytic Iron–Copper Center

The final link in the mitochondrial electron-transport chain is cytochrome  $c$  oxidase. The cytochrome  $c$  oxidase complex accepts electrons from the soluble electron carrier cytochrome  $c$ , and it uses yet a different, third mechanism to pump protons across the inner mitochondrial membrane. The structure of the mammalian complex is illustrated in Figure 14–24. The atomic-resolution structures, combined with studies of the effect of mutations introduced into the enzyme by genetic engineering of the yeast and bacterial proteins, have revealed the detailed mechanisms of this electron-driven proton pump.

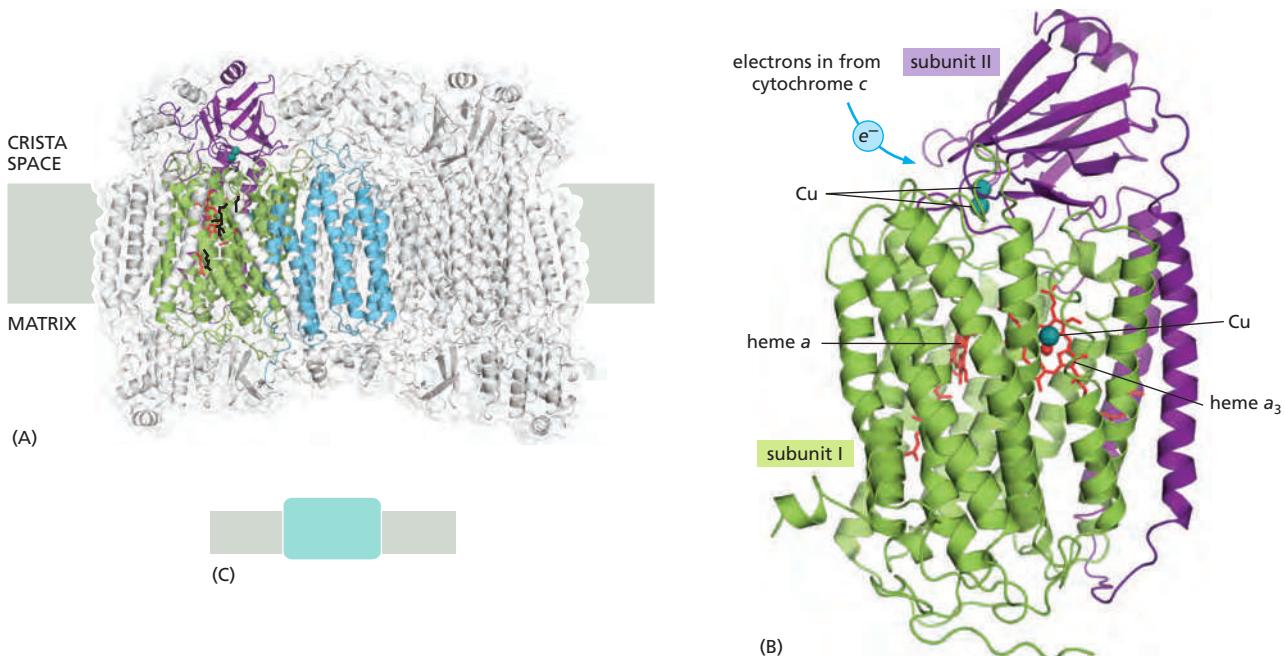
**Figure 14-23** The two-step mechanism of the cytochrome *c* reductase Q-cycle.

(A) In step 1, ubiquinol reduced by NADH dehydrogenase docks to the cytochrome *c* reductase complex. Oxidation of the quinol produces two protons and two electrons. The protons are released into the cristae space. One electron passes via an iron-sulfur cluster to heme  $c_1$ , and then to the soluble electron carrier protein cytochrome *c* on the membrane surface. The second electron passes via hemes  $b_L$  and  $b_H$  to a ubiquinone (*red* Q) bound at a separate site near the matrix side of the protein. Uptake of a proton from the matrix produces an ubisemiquinone radical (see Figure 14-17), which remains bound to this site (*red*  $QH^\bullet$ ) in B.

(B) In step 2, a second ubiquinol (blue  $QH_2$ ) docks and releases two protons and two electrons, as described for step 1. One electron is passed to a second cytochrome *c*, whereas the other electron is accepted by the ubisemiquinone. The ubisemiquinone takes up a proton from the matrix and is released into the lipid bilayer as fully reduced ubiquinol (*red*  $QH_2$ ).

On balance, the oxidation of one ubiquinol in the Q cycle pumps two protons through the membrane by a directional release and uptake of protons (see Figure 14-21), while releasing another two into the cristae space. In addition, in each of the two steps (A) and (B), one electron is transferred to a cytochrome *c* carrier (Movie 14.4).

Because oxygen has a high affinity for electrons, it can release a large amount of free energy when it is reduced to form water. Thus, the evolution of cellular respiration, in which  $O_2$  is converted to water, enabled organisms to harness much more energy than can be derived from anaerobic metabolism. As we discuss later, the availability of the large amount of energy released by the reduction of molecular oxygen to form water is thought to have been essential to the emergence of multicellular life: this would explain why all large organisms respire. The ability of biological systems to use  $O_2$  in this way, however, requires sophisticated chemistry. Once a molecule of  $O_2$  has picked up one electron, it forms a superoxide radical anion ( $O_2^\bullet-$ ) that is dangerously reactive and rapidly takes up an additional three electrons wherever it can get them, with destructive effects on its immediate environment. We can tolerate oxygen in the air we breathe only because the uptake of the first electron by the  $O_2$  molecule is slow, allowing cells to use enzymes to control electron uptake by oxygen. Thus, cytochrome *c* oxidase holds



**Figure 14-24** The structure of cytochrome *c* oxidase. The final complex in the mitochondrial electron-transfer chain consists of 13 different protein subunits, with a total mass of 204,000 daltons. (A) The entire dimeric complex is shown, positioned in the crista membrane. The highly conserved subunits I (green), II (purple), and III (blue) are encoded by the mitochondrial genome, and they form the functional core of the enzyme. (B) The functional core of the complex. Electrons pass through this structure from cytochrome *c* via bound copper ions (blue spheres) and hemes (red) to an  $O_2$  molecule bound between heme  $a_3$  and a copper ion. The four protons needed to reduce  $O_2$  to water are taken up from the matrix; see also Figure 14-25. (C) This shows the symbol for cytochrome *c* oxidase used throughout this chapter. (PDB code: 2OCC.)

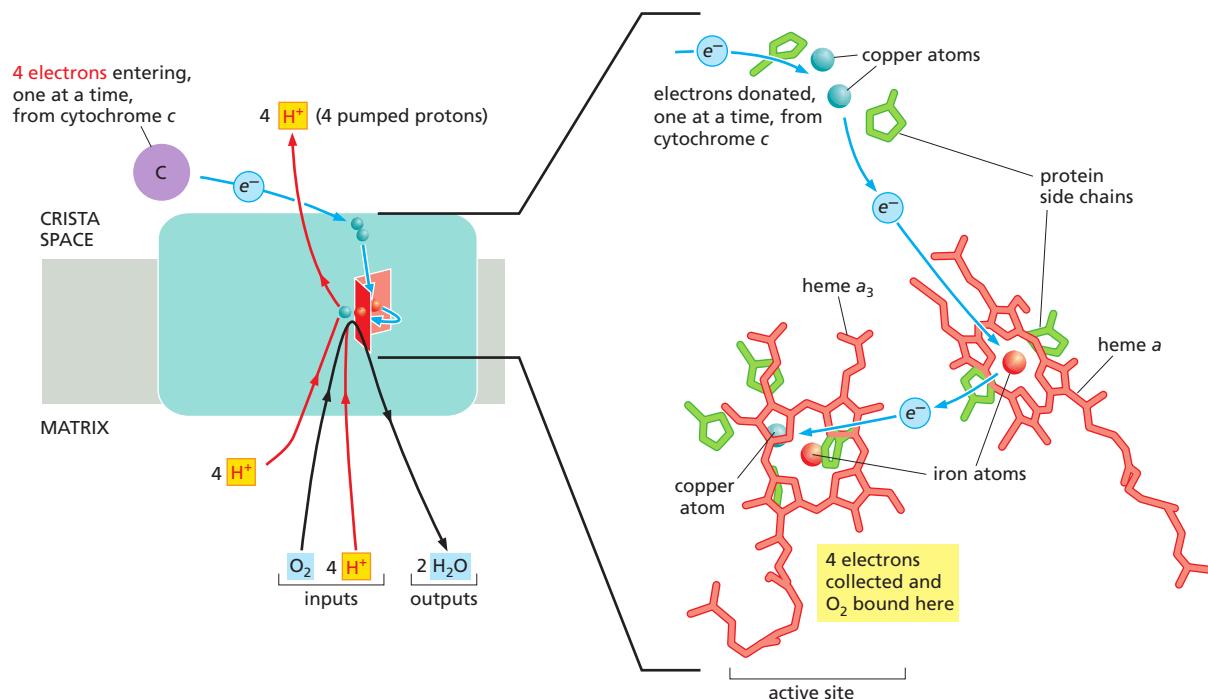
on to oxygen at a special bimetallic center, where it remains clamped between a heme-linked iron atom and a copper ion until it has picked up a total of four electrons. Only then are the two oxygen atoms of the oxygen molecule safely released as two molecules of water (**Figure 14–25**).

The cytochrome *c* oxidase reaction accounts for about 90% of the total oxygen uptake in most cells. This protein complex is therefore crucial for all aerobic life. Cyanide and azide are extremely toxic because they bind to the heme iron atoms in cytochrome *c* oxidase much more tightly than does oxygen, thereby greatly reducing ATP production.

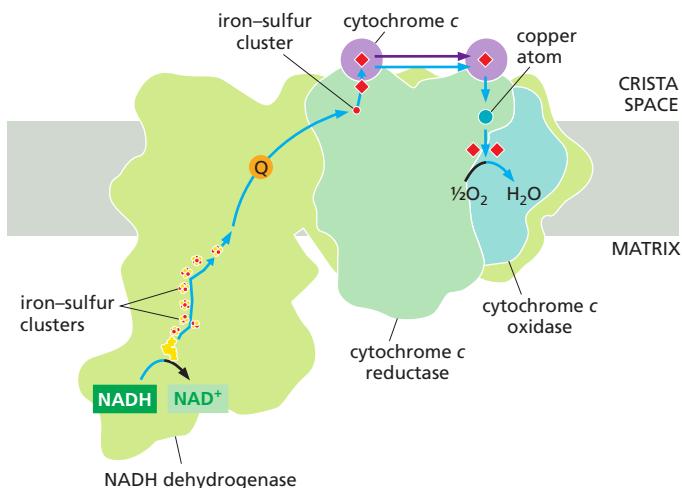
### The Respiratory Chain Forms a Supercomplex in the Crista Membrane

By using cryoelectron microscopy to examine proteins that have been very gently isolated, it can be shown that the three protein complexes that form the respiratory chain assemble into an even larger *supercomplex* in the crista membrane. As illustrated in **Figure 14–26**, this structure is thought to help the mobile electron carriers ubiquinone (in the crista membrane) and cytochrome *c* (in the crista space) transfer electrons with high efficiency. The formation of the supercomplex depends on the presence of the mitochondrial lipid cardiolipin (see Figure 14–11), which presumably works like a hydrophobic glue that holds the components together.

In addition to the three proton pumps in the supercomplex just discussed, one of the enzymes in the citric acid cycle, *succinate dehydrogenase*, is embedded in the mitochondrial crista membrane. In the course of oxidizing succinate to fumarate in the matrix, this enzyme complex captures electrons in the form of a tightly bound FADH<sub>2</sub> molecule (see Panel 2–9, pp. 106–107) and passes them to



**Figure 14–25** The reaction of O<sub>2</sub> with electrons in cytochrome *c* oxidase. Electrons from cytochrome *c* pass through the complex via bound copper ions (blue spheres) and hemes (red) to an O<sub>2</sub> molecule bound between heme *a*<sub>3</sub> and a copper ion. Iron atoms are shown as red spheres. The iron atom in heme *a* serves as an electron queuing point where electrons are held so that they can be released to an O<sub>2</sub> molecule (not shown) that is held at the bimetallic center active site, which is formed by the central iron of the other heme (heme *a*<sub>3</sub>) and a closely apposed copper atom. The four protons needed to reduce O<sub>2</sub> to water are removed from the matrix. For each O<sub>2</sub> molecule that undergoes the reaction  $4e^- + 4H^+ + O_2 \rightarrow 2H_2O$ , another four protons are pumped out of the matrix by mechanisms that are driven by allosteric changes in protein conformation (see Figure 14–28).



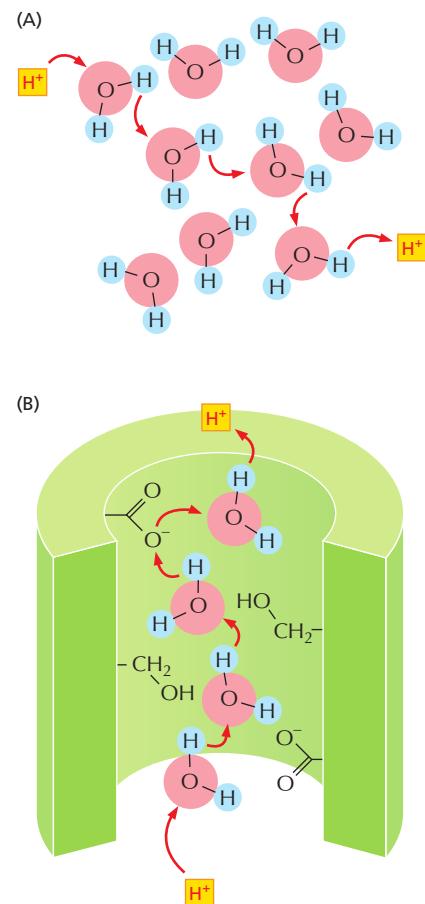
a molecule of ubiquinone. The reduced ubiquinol then passes its two electrons to the respiratory chain via cytochrome *c* reductase (see Figure 14–18). Succinate dehydrogenase is not a proton pump, and it does not contribute directly to the electrochemical potential utilized for ATP production in mitochondria. Thus, it is not considered to be an integral part of the respiratory chain.

### Protons Can Move Rapidly Through Proteins Along Predefined Pathways

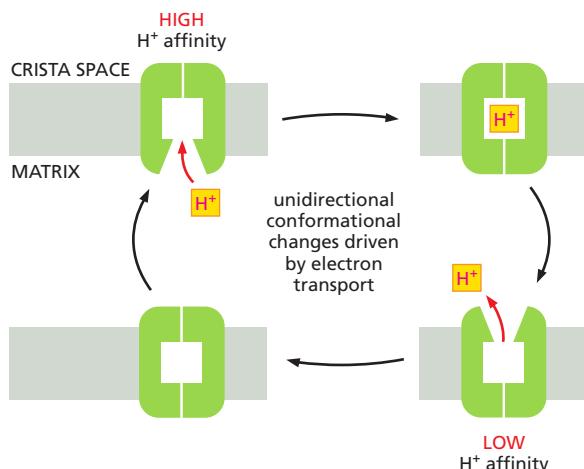
The protons in water are highly mobile: by rapidly dissociating from one water molecule and associating with its neighbor, they can rapidly flit through a hydrogen-bonded network of water molecules (see Figure 2–5). But how can a proton move through the hydrophobic interior of a protein embedded in the lipid bilayer? Proton-translocating proteins contain so-called *proton wires*, which are rows of polar or ionic side chains, or water molecules spaced at short distances, so that the protons can jump from one to the next (Figure 14–27). Along such predefined pathways, protons move up to 40 times faster than through bulk water. The three-dimensional structure of cytochrome *c* oxidase indicates two different proton-uptake pathways. This confirmed earlier mutagenesis studies, which had shown that replacing the side chains of particular aspartate or arginine residues, whose side chains can bind and release protons, made the cytochrome *c* oxidase less efficient as a proton pump.

But how can electron transport cause allosteric changes in protein conformations that pump protons? From the most basic point of view, if electron transport drives sequential allosteric changes in protein conformation that alter the redox state of the components, these conformational changes can be connected to protein wires that allow the protein to pump H<sup>+</sup> across the crista membrane. This type of H<sup>+</sup> pumping requires at least three distinct conformations for the pump protein, as schematically illustrated in Figure 14–28.

**Figure 14–26** The respiratory-chain supercomplex from bovine heart mitochondria. The three proton-pumping complexes of the mitochondrial respiratory chain of mammalian mitochondria assemble into large supercomplexes in the crista membrane. Supercomplexes can be isolated by mild detergent treatment of mitochondria, and their structure has been deciphered by single-particle cryo-electron microscopy. The bovine heart supercomplex has a total mass of 1.7 megadaltons. Shown is a schematic of such a complex that consists of NADH dehydrogenase, cytochrome *c* reductase, and cytochrome *c* oxidase, as indicated. The facing quinol-binding sites of NADH dehydrogenase and cytochrome *c* reductase, plus the short distance between the cytochrome *c*-binding sites in cytochrome *c* reductase and cytochrome *c* oxidase, facilitate fast, efficient electron transfer. Cofactors active in electron transport are marked as a yellow dot (flavin mononucleotide), red and yellow dots (iron-sulfur clusters), Q (quinone), red squares (hemes), and a blue dot (copper atom). Only cofactors participating in the linear flow of electrons from NADH to water are shown. Blue arrows indicate the path of the electrons through the supercomplex. (Adapted from T. Athoff et al., *EMBO J.* 30:4652–4664, 2011.)



**Figure 14–27** Proton movement through water and proteins. (A) Protons move rapidly through water, hopping from one H<sub>2</sub>O molecule to the next by the continuous formation and dissociation of hydronium ions, H<sub>3</sub>O<sup>+</sup> (see Chapter 2). In this diagram, proton jumps are indicated by red arrows. (B) Protons can move even more rapidly through a protein along “proton wires.” These are predefined proton paths consisting of suitably spaced amino acid side chains that accept and release protons easily (Asp, Glu) or carry a waterlike hydroxyl group (Ser, Thr), along with water molecules trapped in the protein interior.



**Figure 14–28** A general model for  $H^+$  pumping coupled to electron transport.

This mechanism for  $H^+$  pumping by a transmembrane protein is thought to be used by NADH dehydrogenase and cytochrome c oxidase, and by many other proton pumps. The protein is driven through a cycle of three conformations. In one of these conformations, the protein has a high affinity for  $H^+$ , causing it to pick up an  $H^+$  on the inside of the membrane. In another conformation, the protein has a low affinity for  $H^+$ , causing it to release an  $H^+$  on the outside of the membrane. As indicated, the transitions from one conformation to another occur only in one direction, because they are being driven by being allosterically coupled to the energetically favorable process of electron transport (discussed in Chapter 11).

## Summary

The respiratory chain embedded in the inner mitochondrial membrane contains three respiratory enzyme complexes, through which electrons pass on their way from NADH to  $O_2$ . In these complexes, electrons are transferred along a series of protein-bound electron carriers, including hemes and iron-sulfur clusters. The energy released as the electrons move to lower and lower energy levels is used to pump protons by different mechanisms in the three respiratory enzyme complexes, each coupling lateral electron transport to vectorial proton transport across the membrane. Electrons are shuttled between enzyme complexes by the mobile electron carriers ubiquinone and cytochrome c to complete the electron-transport chain. The path of electron flow is  $NADH \rightarrow NADH \text{ dehydrogenase complex} \rightarrow \text{ubiquinone} \rightarrow \text{cytochrome c reductase} \rightarrow \text{cytochrome c} \rightarrow \text{cytochrome c oxidase complex} \rightarrow \text{molecular oxygen} (O_2)$ .

## ATP PRODUCTION IN MITOCHONDRIA

As we have just discussed, the three proton pumps of the respiratory chain each contribute to the formation of an electrochemical proton gradient across the inner mitochondrial membrane. This gradient drives ATP synthesis by ATP synthase, a large membrane-bound protein complex that performs the extraordinary feat of converting the energy contained in this electrochemical gradient into biologically useful, chemical-bond energy in the form of ATP (see Figure 14–10). Protons flow down their electrochemical gradient through the membrane part of this proton turbine, thereby driving the synthesis of ATP from ADP and  $P_i$  in the extramembranous part of the complex. As discussed in Chapter 2, the formation of ATP from ADP and inorganic phosphate is highly unfavorable energetically. As we shall see, ATP synthase can produce ATP only because of allosteric shape changes in this protein complex that directly couple ATP synthesis to the energetically favorable flow of protons across its membrane.

### The Large Negative Value of $\Delta G$ for ATP Hydrolysis Makes ATP Useful to the Cell

An average person turns over roughly 50 kg of ATP per day. In athletes running a marathon, this figure can go up to several hundred kilograms. The ATP produced in mitochondria is derived from the energy available in the intermediates NADH,  $FADH_2$ , and GTP. These three energy-rich compounds are produced both by the oxidation of glucose (Table 14–1A), and by the oxidation of fats (Table 14–1B; see also Figure 2–56).

Glycolysis alone can produce only two molecules of ATP for every molecule of glucose that is metabolized, and this is the total energy yield for the fermentation processes that occur in the absence of  $O_2$  (discussed in Chapter 2). In oxidative

**TABLE 14–1 Product Yields from the Oxidation of Sugars and Fats**

A. Net products from oxidation of one molecule of glucose
In cytosol (glycolysis) 1 glucose → 2 pyruvate + 2 NADH + 2 ATP
In mitochondrion (pyruvate dehydrogenase and citric acid cycle) 2 pyruvate → 2 acetyl CoA + 2 NADH 2 acetyl CoA → 6 NADH + 2 FADH <sub>2</sub> + 2 GTP
<b>Net result in mitochondrion</b> 2 pyruvate → 8 NADH + 2 FADH <sub>2</sub> + 2 GTP
B. Net products from oxidation of one molecule of palmitoyl CoA (activated form of palmitate, a fatty acid)
In mitochondrion (fatty acid oxidation and citric acid cycle) 1 palmitoyl CoA → 8 acetyl CoA + 7 NADH + 7 FADH <sub>2</sub> 8 acetyl CoA → 24 NADH + 8 FADH <sub>2</sub> + 8 GTP
<b>Net result in mitochondrion</b> 1 palmitoyl CoA → 31 NADH + 15 FADH <sub>2</sub> + 8 GTP

phosphorylation, each pair of electrons donated by the NADH produced in mitochondria can provide energy for the formation of about 2.5 molecules of ATP. Oxidative phosphorylation also produces 1.5 ATP molecules per electron pair from the FADH<sub>2</sub> produced by succinate dehydrogenase in the mitochondrial matrix, and from the NADH molecules produced by glycolysis in the cytosol. From the product yields of glycolysis and the citric acid cycle, we can calculate that the complete oxidation of one molecule of glucose—starting with glycolysis and ending with oxidative phosphorylation—gives a net yield of about 30 molecules of ATP. Nearly all this ATP is produced by the mitochondrial ATP synthase.

In Chapter 2, we introduced the concept of free energy ( $G$ ). The free-energy change for a reaction,  $\Delta G$ , determines whether that reaction will occur in a cell. We showed on pp. 60–63 that the  $\Delta G$  for a given reaction can be written as the sum of two parts: the first, called the standard free-energy change,  $\Delta G^\circ$ , depends only on the intrinsic characters of the reacting molecules; the second depends only on their concentrations. For the simple reaction A → B,

$$\Delta G = \Delta G^\circ + RT \ln \frac{[B]}{[A]}$$

where [A] and [B] denote the concentrations of A and B, and ln is the natural logarithm.  $\Delta G^\circ$  is the standard reference value, which can be seen to be equal to the value of  $\Delta G$  when the molar concentrations of A and B are equal (since  $\ln 1 = 0$ ).

In Chapter 2, we discussed how the large, favorable free-energy change (large negative  $\Delta G$ ) for ATP hydrolysis is used, through coupled reactions, to drive many other chemical reactions in the cell that would otherwise not occur (see pp. 65–66). The ATP hydrolysis reaction produces two products, ADP and P<sub>i</sub>; it is therefore of the type A → B + C, where, as demonstrated in [Figure 14–29](#),

$$\Delta G = \Delta G^\circ + RT \ln \frac{[B][C]}{[A]}$$

When ATP is hydrolyzed to ADP and P<sub>i</sub> under the conditions that normally exist in a cell, the free-energy change is roughly -46 to -54 kJ/mole (-11 to -13 kcal/mole). This extremely favorable  $\Delta G$  depends on maintaining a high concentration of ATP compared with the concentrations of ADP and P<sub>i</sub>. When ATP, ADP, and P<sub>i</sub> are all present at the same concentration of 1 mole/liter (so-called standard conditions), the  $\Delta G$  for ATP hydrolysis drops to the standard free-energy change ( $\Delta G^\circ$ ), which is only -30.5 kJ/mole (-7.3 kcal/mole). At much lower concentrations of ATP relative to ADP and P<sub>i</sub>,  $\Delta G$  becomes zero. At this point, the rate at

**1**

$$\text{ATP} \xrightarrow{\text{hydrolysis}} \text{ADP} + \text{P}_i$$

$$\text{hydrolysis rate} = \text{hydrolysis rate constant} \times \text{concentration of ATP}$$

**2**

$$\text{ADP} + \text{P}_i \xrightarrow{\text{synthesis}} \text{ATP}$$

$$\text{synthesis rate} = \text{synthesis rate constant} \times \text{conc. of phosphate} \times \text{conc. of ADP}$$

**3**

**AT EQUILIBRIUM:**

$$\text{synthesis rate} = \text{hydrolysis rate}$$

$$\text{synthesis rate constant} \times \text{conc. of phosphate} \times \text{conc. of ADP} = \text{hydrolysis rate constant} \times \text{conc. of ATP}$$

$$\text{thus, } \frac{\text{conc. of ADP} \times \text{conc. of phosphate}}{\text{concentration of ATP}} = \frac{\text{hydrolysis rate constant}}{\text{synthesis rate constant}} = \text{equilibrium constant } K$$

$$\text{or abbreviated, } \frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} = K$$

**4**

For the reaction

$$\text{ATP} \longrightarrow \text{ADP} + \text{P}_i$$

the following equation applies:

$$\Delta G = \Delta G^\circ + RT \ln \frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]}$$

where  $\Delta G$  and  $\Delta G^\circ$  are in Joules per mole,  $R$  is the gas constant (8.3 J/mole K),  $T$  is the absolute temperature (K), and all the concentrations are in moles per liter.

When the concentrations of all reactants are at 1 M,  $\Delta G = \Delta G^\circ$  (since  $RT \ln 1 = 0$ ).  $\Delta G^\circ$  is thus a constant defined as the standard free-energy change for the reaction.

At equilibrium the reaction has no net effect on the disorder of the universe, so  $\Delta G = 0$ . Therefore, at equilibrium,

$$-RT \ln \frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} = \Delta G^\circ$$

But the concentrations of reactants at equilibrium must satisfy the equilibrium equation:

$$\frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} = K$$

Therefore, at equilibrium,

$$\Delta G^\circ = -RT \ln K$$

We thus see that whereas  $\Delta G^\circ$  indicates the equilibrium point for a reaction,  $\Delta G$  reveals how far the reaction is from equilibrium.  $\Delta G$  is a measure of the "driving force" for any chemical reaction, just as the proton-motive force is the driving force for the translocation of protons.

which ADP and  $P_i$  will join to form ATP will be equal to the rate at which ATP hydrolyzes to form ADP and  $P_i$ . In other words, when  $\Delta G = 0$ , the reaction is at equilibrium (see Figure 14–29).

It is  $\Delta G$ , not  $\Delta G^\circ$ , that indicates how far a reaction is from equilibrium and determines whether it can drive other reactions. Because the efficient conversion of ADP to ATP in mitochondria maintains such a high concentration of ATP relative to ADP and  $P_i$ , the ATP hydrolysis reaction in cells is kept very far from equilibrium and  $\Delta G$  is correspondingly very negative. Without this large disequilibrium, ATP hydrolysis could not be used to drive the reactions of the cell. At low ATP concentrations, many biosynthetic reactions would run backward and the cell would die.

### The ATP Synthase Is a Nanomachine that Produces ATP by Rotary Catalysis

The **ATP synthase** is a finely tuned nanomachine composed of 23 or more separate protein subunits, with a total mass of about 600,000 daltons. The ATP synthase can work both in the forward direction, producing ATP from ADP and phosphate in response to an electrochemical gradient, or in reverse, generating an electrochemical gradient by ATP hydrolysis. To distinguish it from other enzymes that hydrolyze ATP, it is also called an  $F_1F_0$  ATP synthase or F-type ATPase.

Resembling a turbine, ATP synthase is composed of both a rotor and a stator (Figure 14–30). To prevent the catalytic head from rotating, a stalk at the periphery of the complex (the stator stalk) connects the head to stator subunits embedded in the membrane. A second stalk in the center of the assembly (the rotor stalk) is connected to the rotor ring in the membrane that turns as protons flow through it, driven by the electrochemical gradient across the membrane. As a result, proton

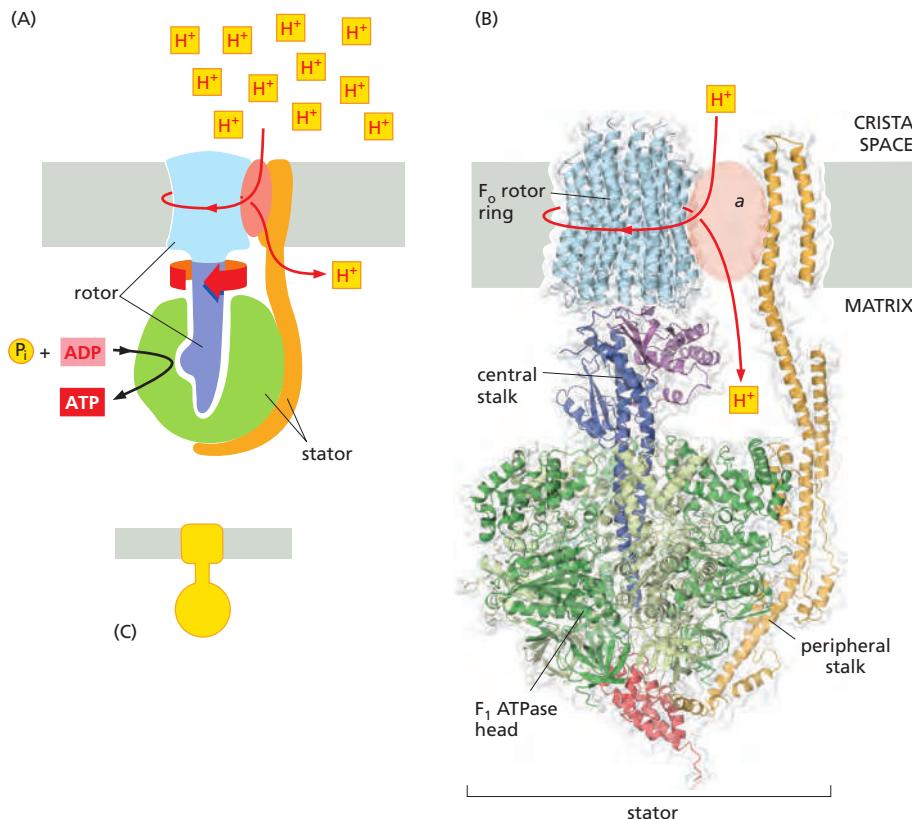
**Figure 14–29** The basic relationship between free-energy changes and equilibrium in the ATP hydrolysis reaction. The rate constants in boxes 1 and 2 are determined from experiments in which product accumulation is measured as a function of time (conc., concentration). The equilibrium constant shown here,  $K$ , is in units of moles per liter. (See Panel 2–7, pp. 102–103, for a discussion of free energy and see Figure 3–44 for a discussion of the equilibrium constant.)

flow makes the rotor stalk rotate inside the stationary head, where the catalytic sites that assemble ATP from ADP and  $P_i$  are located. Three  $\alpha$  and three  $\beta$  subunits of similar structure alternate to form the head. Each of the three  $\beta$  subunits has a catalytic nucleotide-binding site at the  $\alpha/\beta$  interface. These catalytic sites are all in different conformations, depending on their interaction with the rotor stalk. This stalk acts like a camshaft, the device that opens and closes the valves in a combustion engine. As it rotates within the head, the stalk changes the conformations of the  $\beta$  subunits sequentially. One of the possible conformations of the catalytic sites has high affinity for ADP and  $P_i$ , and as the rotor stalk pushes the binding site into a different conformation, these two substrates are driven to form ATP. In this way, the mechanical force exerted by the central rotor stalk is directly converted into the chemical energy of the ATP phosphate bond.

Serving as a proton-driven turbine, the ATP synthase is driven by  $H^+$  flow into the matrix to spin at about 8000 revolutions per minute, generating three molecules of ATP per turn. In this way, each ATP synthase can produce roughly 400 molecules of ATP per second.

### Proton-driven Turbines Are of Ancient Origin

The membrane-embedded rotors of ATP synthases consist of a ring of identical  $c$  subunits (Figure 14–31). Each  $c$  subunit is a hairpin of two membrane-spanning  $\alpha$  helices that contain a proton-binding site defined by a glutamate or aspartate in the middle of the lipid bilayer. The  $a$  subunit, which is part of the stator (see Figure 14–30), makes two narrow channels at the interface between the rotor and stator, each spanning half of the membrane and converging on the proton-binding site at the middle of the rotor subunit. Protons flow through the two half-channels down their electrochemical gradient from the crista space back into the matrix. A negatively charged side chain in the binding site accepts a proton arriving from the crista space through the first half-channel, as it rotates past the  $a$  subunit. The bound proton then rides round in the ring for a full cycle, whereupon it is thought to be displaced by a positively charged arginine in the  $a$  subunit, and escapes



**Figure 14–30 ATP synthase.** The three-dimensional structure of the  $F_1F_0$  ATP synthase, determined by x-ray crystallography. Also known as an F-type ATPase, it consists of an  $F_0$  part (from “oligomycin-sensitive factor”) in the membrane and the large, catalytic  $F_1$  head in the matrix. Under mild dissociation conditions, this complex separates into its  $F_1$  and  $F_0$  components, which can be isolated and studied individually. (A) Diagram of the enzyme complex showing how its globular head portion (green) is kept stationary as proton-flow across the membrane drives a rotor (blue) that turns inside it. (B) In bovine heart mitochondria, the  $F_0$  rotor ring in the membrane (light blue) has eight  $c$  subunits. It is attached to the  $\gamma$  subunit of the central stalk (dark blue) by the  $\epsilon$  subunit (purple). The catalytic  $F_1$  head consists of a ring of three  $\alpha$  and three  $\beta$  subunits (light and dark green), and it directly converts mechanical energy into chemical-bond energy in ATP, as described in the text. The elongated peripheral stalk of the stator (orange) is connected to the  $F_1$  head by the small  $\delta$  subunit (red) at one end, and to the  $a$  subunit in the membrane (pink oval) at the other. Together with the  $c$  subunits of the ring rotating past it, the  $a$  subunit creates a path for protons through the membrane. (C) The symbol for ATP synthase used throughout this book.

The closely related ATP synthases of mitochondria, chloroplasts, and bacteria synthesize ATP by harnessing the proton-motive force across a membrane. This powers the rotation of the rotor against the stator in a counterclockwise direction, as seen from the  $F_1$  head. The same enzyme complex can also pump protons against their electrochemical gradient by hydrolyzing ATP, which then drives the clockwise rotation of the rotor. The direction of operation depends on the net free-energy change ( $\Delta G$ ) for the coupled processes of  $H^+$  translocation across the membrane and the synthesis of ATP from ADP and  $P_i$  (Movie 14.5 and Movie 14.6).

Measurement of the torque that the ATP synthase can produce by ATP hydrolysis reveals that the ATP synthase is 60 times more powerful than a diesel engine of equal dimensions. (B, courtesy of K. Davies. PDB codes: 2WPD, 2OLY, 2WSS, 2BO5.)

through the second half-channel into the matrix. Thus proton flow causes the rotor ring to spin against the stator like a proton-driven turbine.

The mitochondrial ATP synthase is of ancient origin: essentially the same enzyme occurs in plant chloroplasts and in the plasma membrane of bacteria or archaea. The main difference between them is the number of *c* subunits in the rotor ring. In mammalian mitochondria, the ring has 8 subunits. In yeast mitochondria, the number is 10; in bacteria and archaea, it ranges from 11 to 13; in plant chloroplasts, there are 14; and the rings of some cyanobacteria contain 15 *c* subunits.

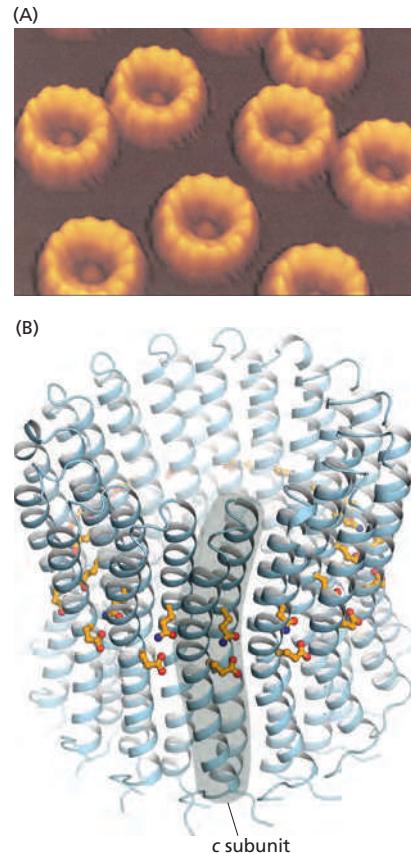
The *c* subunits in the rotor ring can be thought of as cogs in the gears of a bicycle. A high gear, with a small number of cogs, is advantageous when the supply of protons is limited, as in mitochondria, but a low gear, with a large number of cogs in the wheel, is preferable when the proton gradient is high. This is the case in chloroplasts and cyanobacteria, where protons produced through the action of sunlight are plentiful. Because each rotation produces three molecules of ATP in the head, the synthesis of one ATP requires around three protons in mitochondria but up to five in photosynthetic organisms. It is the number of *c* subunits in the ring that defines how many protons need to pass through this marvelous device to make each molecule of ATP, and thereby how high a ratio of ATP to ADP can be maintained by the ATP synthase.

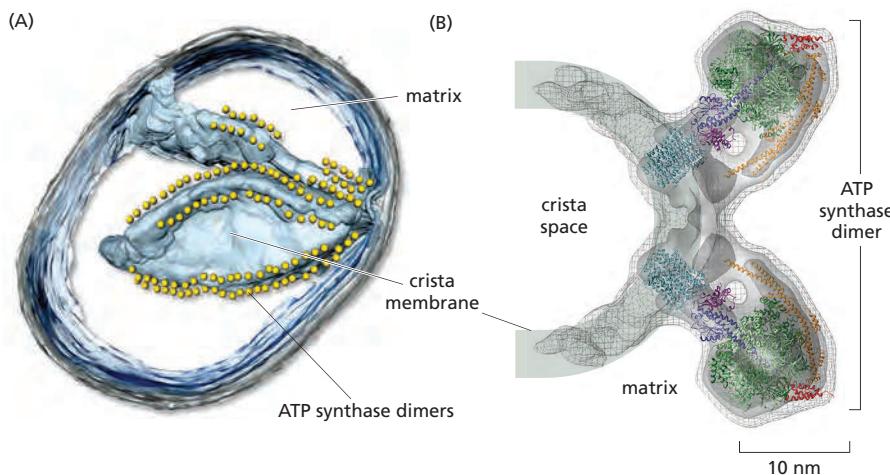
In principle, ATP synthase can also run in reverse as an ATP-powered proton pump that converts the energy of ATP back into a proton gradient across the membrane. In many bacteria, the rotor of the ATP synthase in the plasma membrane changes direction routinely, from ATP synthesis mode in aerobic respiration, to ATP hydrolysis mode in anaerobic metabolism. In this latter case, ATP hydrolysis serves to maintain the proton gradient across the plasma membrane, which is used to power many other essential cell functions including nutrient transport and the rotation of bacterial flagella. The V-type ATPases that acidify certain cellular organelles are architecturally similar to the F-type ATP synthases, but they normally function in reverse (see Figure 13–37).

### Mitochondrial Cristae Help to Make ATP Synthesis Efficient

In the electron microscope, the mitochondrial ATP synthase complexes can be seen to project like lollipops on the matrix side of cristae membranes. Recent studies by cryoelectron microscopy and tomography have shown that this large complex is not distributed randomly in the membrane, but forms long rows of dimers along the cristae ridges (Figure 14–32). The dimer rows induce or stabilize these regions of high membrane curvature, which are otherwise energetically unfavorable. Indeed, the formation of ATP synthase dimers and their assembly into rows are required for cristae formation and have far-reaching consequences for cellular fitness. By contrast with bacterial or chloroplast ATP synthases, which do not form dimers, the mitochondrial complex contains additional subunits, located mostly near the membrane end of the stator stalk. Several of these subunits are found to be dimer-specific. If these subunits are mutated in yeast, the ATP synthase in the membrane remains monomeric, the mitochondria have no cristae, cellular respiration drops by half, and the cells grow more slowly.

**Figure 14–31**  $F_0$  ATP synthase rotor rings. (A) Atomic force microscopy image of ATP synthase rotors from the cyanobacterium *Synechococcus elongatus* in a lipid bilayer. Whereas 8 *c* subunits form the rotor in Figure 14–30, there are 13 *c* subunits in this ring. (B) The x-ray structure of the  $F_0$  ring of the ATP synthase from *Spirulina platensis*, another cyanobacterium, shows that this rotor has 15 *c* subunits. In all ATP synthases, the *c* subunits are hairpins of two membrane-spanning  $\alpha$  helices (one subunit is highlighted in gray). The helices are highly hydrophobic, except for two glutamine and glutamate side chains (yellow) that create proton-binding sites in the membrane. (A, courtesy of Thomas Meier and Denys Pogoryelov; B, PDB code: 2WIE.)

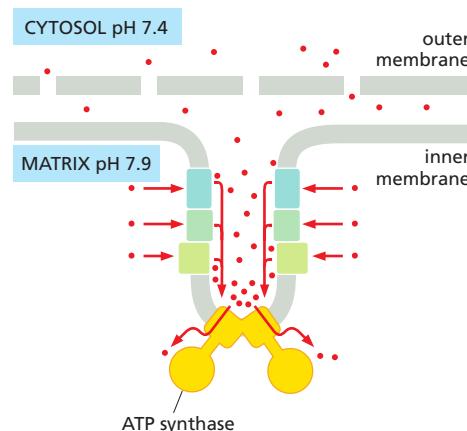




Electron tomography suggests that the proton pumps of the respiratory chain are located in the membrane regions at either side of the dimer rows. Protons pumped into the crista space by these respiratory-chain complexes are thought to diffuse very rapidly along the membrane surface, with the ATP synthase rows creating a proton “sink” at the cristae tips (Figure 14-33). *In vitro* studies suggest that the ATP synthase needs a proton gradient of about 2 pH units to produce ATP at the rate required by the cell, irrespective of the membrane potential. The H<sup>+</sup> gradient across the inner mitochondrial membrane is only 0.5 to 0.6 pH units. The cristae thus seem to work as proton traps that enable the ATP synthase to make efficient use of the protons pumped out of the mitochondrial matrix. As we shall see in the next section, this elaborate arrangement of membrane protein complexes is absent in chloroplasts, where the H<sup>+</sup> gradient is much higher.

### Special Transport Proteins Exchange ATP and ADP Through the Inner Membrane

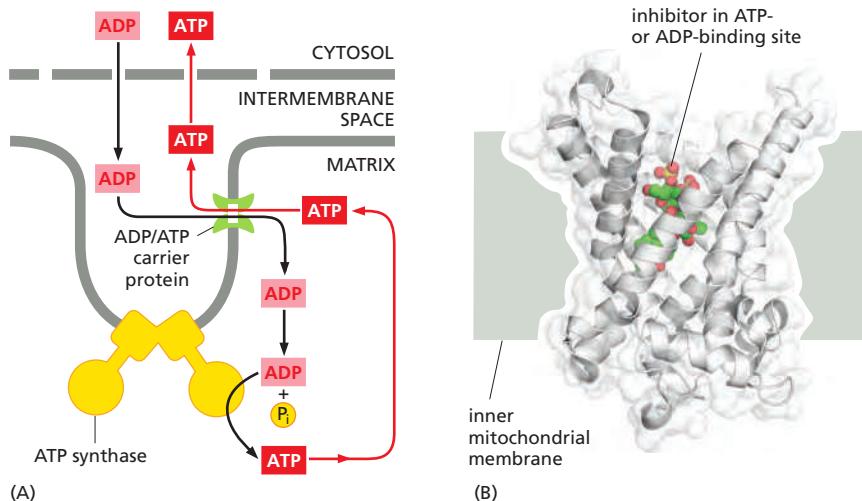
Like all biological membranes, the inner mitochondrial membrane contains numerous specific *transport proteins* that allow particular substances to pass through. One of the most abundant of these is the *ADP/ATP carrier protein* (Figure 14-34). This carrier shuttles the ATP produced in the matrix through the inner membrane to the intermembrane space, from where it diffuses through the outer mitochondrial membrane to the cytosol. In exchange, ADP passes from the cytosol into the matrix for recycling into ATP. ATP<sup>4-</sup> has one more negative charge than ADP<sup>3-</sup>, and the exchange of ATP and ADP is driven by the electrochemical gradient across the inner membrane, so that the more negatively charged ATP is pushed out of the matrix, and the less negatively charged ADP is pulled in. The ADP/ATP carrier is but one member of a *mitochondrial carrier family*: the inner



**Figure 14-32** Dimers of mitochondrial ATP synthase in cristae membranes.

(A) A three-dimensional map of a small mitochondrion obtained by electron microscope tomography shows that ATP synthases form long paired rows along cristae ridges. The outer membrane is gray, the inner membrane and cristae membranes have been colored light blue. Each head of an ATP synthase is indicated by a yellow sphere. (B) A three-dimensional map of a mitochondrial ATP synthase dimer in the crista membrane obtained by subtomogram averaging, with fitted x-ray structures (Movie 14.7). (A, from K. Davies et al., Proc. Natl Acad. Sci. USA 108:14121–14126, 2011. With permission from the National Academy of Sciences; B, from K. Davies et al., Proc. Natl Acad. Sci. USA 109:13602–13607, 2012. With permission from the National Academy of Sciences.)

**Figure 14-33** ATP synthase dimers at cristae ridges and ATP production. At the crista ridges, the ATP synthases (yellow) form a sink for protons (red). The proton pumps of the electron-transport chain (green) are located in the membrane regions on either side of the crista. As illustrated, protons tend to diffuse along the membrane from their source to the proton sink created by the ATP synthase. This allows efficient ATP production despite the small H<sup>+</sup> gradient between the cytosol and matrix. Red arrows show the direction of the proton flow.



**Figure 14–34 The ADP/ATP carrier protein.** (A) The ADP/ATP carrier protein is a small membrane protein that carries the ATP produced on the matrix side of the inner membrane to the intermembrane space, and the ADP that is needed for ATP synthesis into the matrix. (B) In the ADP/ATP carrier, six transmembrane  $\alpha$  helices define a cavity that binds either ADP or ATP. In this x-ray structure, the substrate is replaced by a tightly bound inhibitor instead (colored). When ADP binds from outside the inner membrane, it triggers a conformational change and is released into the matrix. In exchange, a molecule of ATP quickly binds to the matrix side of the carrier and is transported to the intermembrane space. From there the ATP diffuses through the outer mitochondrial membrane to the cytoplasm, where it powers the energy-requiring processes in the cell. (B, PDB code: 1OKC.)

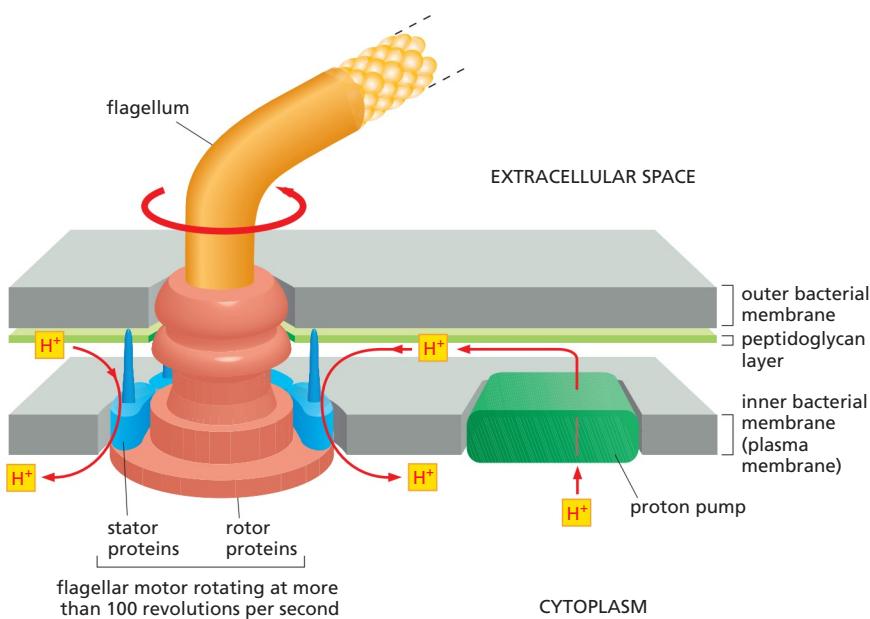
mitochondrial membrane contains about 20 related carrier proteins exchanging various other metabolites, including the phosphate that is required along with ADP for ATP synthesis.

In some specialized fat cells, mitochondrial respiration is uncoupled from ATP synthesis by the *uncoupling protein*, another member of the mitochondrial carrier family. In these cells, known as brown fat cells, most of the energy of oxidation is dissipated as heat rather than being converted into ATP. In the inner membranes of the large mitochondria in these cells, the uncoupling protein allows protons to move down their electrochemical gradient without passing through ATP synthase. This process is switched on when heat generation is required, causing the cells to oxidize their fat stores at a rapid rate and produce heat rather than ATP. Tissues containing brown fat serve as “heating pads,” helping to revive hibernating animals and to protect newborn human babies from the cold.

### Chemiosmotic Mechanisms First Arose in Bacteria

Bacteria use enormously diverse energy sources. Some, like animal cells, are aerobic; they synthesize ATP from sugars they oxidize to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  by glycolysis, the citric acid cycle, and a respiratory chain in their plasma membrane that is similar to the one in the inner mitochondrial membrane. Others are strict anaerobes, deriving their energy either from glycolysis alone (by fermentation, see Figure 2–47) or from an electron-transport chain that employs a molecule other than oxygen as the final electron acceptor. The alternative electron acceptor can be a nitrogen compound (nitrate or nitrite), a sulfur compound (sulfate or sulfite), or a carbon compound (fumarate or carbonate), for example. A series of electron carriers in the plasma membrane that are comparable to those in mitochondrial respiratory chains transfers the electrons to these acceptors.

Despite this diversity, the plasma membrane of the vast majority of bacteria contains an ATP synthase that is very similar to the one in mitochondria. In bacteria that use an electron-transport chain to harvest energy, the electron-transport chain pumps  $\text{H}^+$  out of the cell and thereby establishes a proton-motive force across the plasma membrane that drives the ATP synthase to make ATP. In other



**Figure 14–35** The rotation of the bacterial flagellum driven by H<sup>+</sup> flow.

The flagellum is attached to a series of protein rings (pink), which are embedded in the outer and inner membranes and rotate with the flagellum. The rotation is driven by a flow of protons through an outer ring of proteins (the stator) by mechanisms that may resemble those used by the ATP synthase. However, the flow of protons in the flagellar motor is always toward the cytosol, both during clockwise and counterclockwise rotation, whereas in ATP synthase this flow reverses with the direction of rotation (Movie 14.8).

bacteria, the ATP synthase works in reverse, using the ATP produced by glycolysis to pump H<sup>+</sup> and establish a proton gradient across the plasma membrane.

Bacteria, including the strict anaerobes, maintain a proton gradient across their plasma membrane that is harnessed to drive many other processes. It can be used to drive a flagellar motor, for example (Figure 14–35). This gradient is harnessed to pump Na<sup>+</sup> out of the bacterium via a Na<sup>+</sup>-H<sup>+</sup> antiporter that takes the place of the Na<sup>+</sup>-K<sup>+</sup> pump of eukaryotic cells. The gradient is also used for the active inward transport of nutrients, such as most amino acids and many sugars: each nutrient is dragged into the cell along with one or more protons through a specific symporter (Figure 14–36; see also Chapter 11). In animal cells, by contrast, most inward transport across the plasma membrane is driven by the Na<sup>+</sup> gradient (high Na<sup>+</sup> outside, low Na<sup>+</sup> inside) that is established by the Na<sup>+</sup>-K<sup>+</sup> pump (see Figure 11–15).

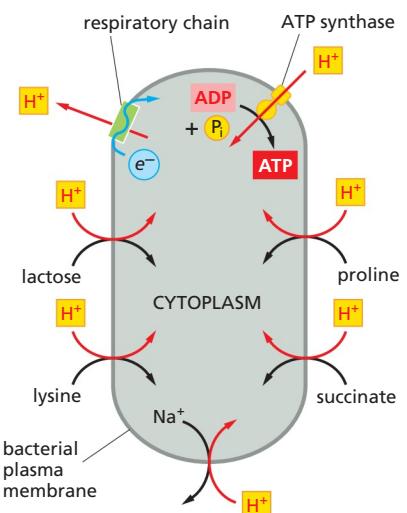
Some unusual bacteria have adapted to live in a very alkaline environment and yet must maintain their cytoplasm at a physiological pH. For these cells, any attempt to generate an electrochemical H<sup>+</sup> gradient would be opposed by a large H<sup>+</sup> concentration gradient in the wrong direction (H<sup>+</sup> higher inside than outside). Presumably for this reason, some of these bacteria substitute Na<sup>+</sup> for H<sup>+</sup> in all of their chemiosmotic mechanisms. The respiratory chain pumps Na<sup>+</sup> out of the cell, the transport systems and flagellar motor are driven by an inward flux of Na<sup>+</sup>, and a Na<sup>+</sup>-driven ATP synthase synthesizes ATP. The existence of such bacteria demonstrates a critical point: the principle of chemiosmosis is more fundamental than the proton-motive force on which it is normally based.

As we discuss next, an ATP synthase coupled to chemiosmotic processes is also a central feature of plants, where it plays critical roles in both mitochondria and chloroplasts.

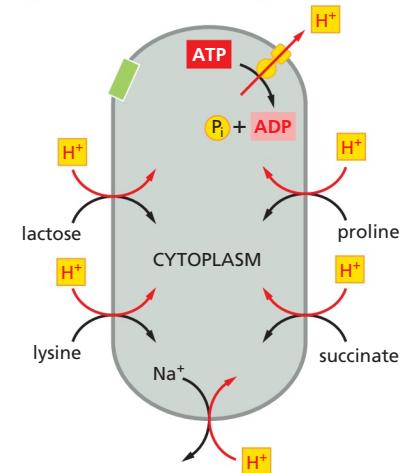
**Figure 14–36** The importance of H<sup>+</sup>-driven transport in bacteria.

A proton-motive force generated across the plasma membrane pumps nutrients into the cell and expels Na<sup>+</sup>. (A) In an aerobic bacterium, a respiratory chain fed by the oxidation of substrates produces an electrochemical proton gradient across the plasma membrane. This gradient is then harnessed to make ATP, as well as to transport nutrients (proline, succinate, lactose, and lysine) into the cell and to pump Na<sup>+</sup> out of the cell. (B) When the same bacterium grows under anaerobic conditions, it derives its ATP from glycolysis. As indicated, the ATP synthase in the plasma membrane then hydrolyzes some of this ATP to establish an electrochemical proton gradient that drives the same transport processes that depend on respiratory chain proton-pumping in (A).

**(A) AEROBIC CONDITIONS**



**(B) ANAEROBIC CONDITIONS**



## Summary

The large amount of free energy released when  $H^+$  flows back into the matrix from the cristae provides the basis for ATP production on the matrix side of mitochondrial cristae membranes by a remarkable protein machine—the ATP synthase. The ATP synthase functions like a miniature turbine, and it is a reversible device that can couple proton flow to either ATP synthesis or ATP hydrolysis. The transmembrane electrochemical gradient that drives ATP production in mitochondria also drives the active transport of selected metabolites across the inner mitochondrial membrane, including an efficient ADP/ATP exchange between the mitochondrion and the cytosol that keeps the cell's ATP pool highly charged. The resulting high cellular concentration of ATP makes the free-energy change for ATP hydrolysis extremely favorable, allowing this hydrolysis reaction to drive a very large number of energy-requiring processes throughout the cell. The universal presence of ATP synthase in bacteria, mitochondria, and chloroplasts testifies to the central importance of chemiosmotic mechanisms in cells.

## CHLOROPLASTS AND PHOTOSYNTHESIS

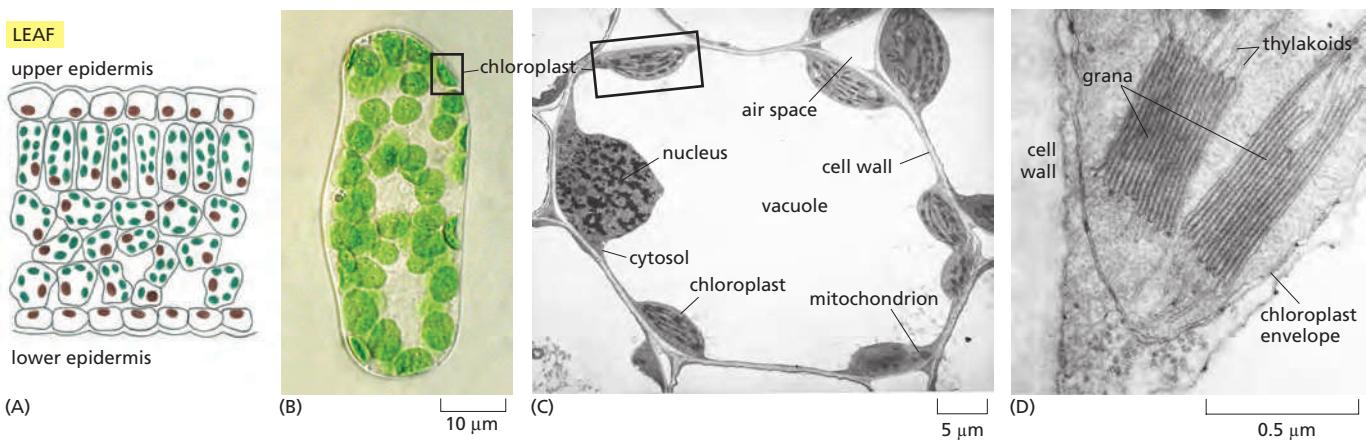
All animals and most microorganisms rely on the continual uptake of large amounts of organic compounds from their environment. These compounds provide both the carbon-rich building blocks for biosynthesis and the metabolic energy for life. It is likely that the first organisms on the primitive Earth had access to an abundance of organic compounds produced by geochemical processes, but it is clear that these were used up billions of years ago. Since that time, virtually all of the organic materials required by living cells have been produced by *photosynthetic organisms*, including plants and photosynthetic bacteria. The core machinery that drives all photosynthesis appears to have evolved more than 3 billion years ago in the ancestors of present-day bacteria; today it provides the only major solar energy storage mechanism on Earth.

The most advanced photosynthetic bacteria are the cyanobacteria, which have minimal nutrient requirements. They use electrons from water and the energy of sunlight to convert atmospheric  $CO_2$  into organic compounds—a process called *carbon fixation*. In the course of the overall reaction  $nH_2O + nCO_2 \rightarrow$  (light)  $(CH_2O)_n + nO_2$ , they also liberate into the atmosphere the molecular oxygen that then powers oxidative phosphorylation. In this way, it is thought that the evolution of cyanobacteria from more primitive photosynthetic bacteria eventually made possible the development of the many different aerobic life-forms that populate the Earth today.

### Chloroplasts Resemble Mitochondria But Have a Separate Thylakoid Compartment

Plants (including algae) developed much later than cyanobacteria, and their photosynthesis occurs in a specialized intracellular organelle—the **chloroplast** (Figure 14–37). Chloroplasts use chemiosmotic mechanisms to carry out their energy interconversions in much the same way that mitochondria do. Although much larger than mitochondria, they are organized on the same principles. They have a highly permeable outer membrane; a much less permeable inner membrane, in which membrane transport proteins are embedded; and a narrow intermembrane space in between. Together, these two membranes form the chloroplast envelope (Figure 14–37D). The inner chloroplast membrane surrounds a large space called the **stroma**, which is analogous to the mitochondrial matrix. The stroma contains many metabolic enzymes and, as for the mitochondrial matrix, it is the place where ATP is made by the head of an ATP synthase. Like the mitochondrion, the chloroplast has its own genome and genetic system. The stroma therefore also contains a special set of ribosomes, RNAs, and the chloroplast DNA.

An important difference between the organization of mitochondria and chloroplasts is highlighted in Figure 14–38. The inner membrane of the chloroplast is



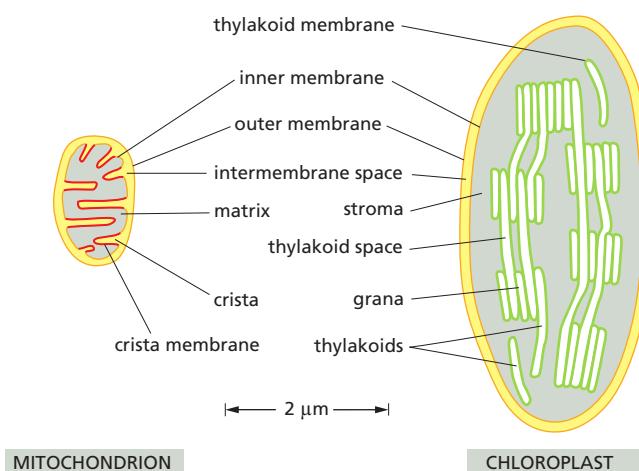
not folded into cristae and does not contain electron-transport chains. Instead, the electron-transport chains, photosynthetic light-capturing systems, and ATP synthase are all contained in the **thylakoid membrane**, a separate, distinct membrane that forms a set of flattened, disc-like sacs, the **thylakoids**. The thylakoid membrane is highly folded into numerous local stacks of flattened vesicles called **grana**, interconnected by nonstacked thylakoids. The lumen of each thylakoid is connected with the lumen of other thylakoids, thereby defining a third internal compartment called the **thylakoid space**. This space represents a separate compartment in each chloroplast that is not connected to either the intermembrane space or the stroma.

### Chloroplasts Capture Energy from Sunlight and Use It to Fix Carbon

We can group the reactions that occur during photosynthesis in chloroplasts into two broad categories:

1. The **photosynthetic electron-transfer reactions** (also called the “light reactions”) occur in two large protein complexes, called *reaction centers*, embedded in the thylakoid membrane. A photon (a quantum of light) knocks an electron out of the green pigment molecule **chlorophyll** in the first reaction center, creating a positively charged chlorophyll ion. This electron then moves along an electron-transport chain and through a second reaction center in much the same way that an electron moves along the respiratory chain in mitochondria. During this electron-transport process, H<sup>+</sup> is pumped across the thylakoid membrane, and the resulting

**Figure 14-37 Chloroplasts in the cell.**  
 (A) Schematic cross section through the leaf of a green plant. (B) Light microscopy of a plant leaf cell—here, a mesophyll cell from *Zinnia elegans*—shows chloroplasts as bright green bodies, measuring several micrometers across, in the transparent cell interior. (C) The electron micrograph of a thin, stained section through a wheat leaf cell shows a thin rim of cytoplasm—containing chloroplasts, the nucleus, and mitochondria—surrounding a large, water-filled vacuole. (D) At higher magnification, electron microscopy reveals the chloroplast envelope membrane and the thylakoid membrane within the chloroplast that is highly folded into grana stacks  
**(Movie 14.9).** (B, courtesy of John Innes Foundation; C and D, courtesy of K. Plaskitt.)



**Figure 14-38 A mitochondrion and chloroplast compared.** Chloroplasts are generally larger than mitochondria. In addition to an outer and inner envelope membrane, they contain the thylakoid membrane with its internal thylakoid space. The chloroplast thylakoid membrane, which is the site of solar energy conversion in plants and algae, corresponds to the mitochondrial cristae, which are the sites of energy conversion by cellular respiration. Unlike the cristae membrane, which is continuous with the inner mitochondrial membrane at cristae junctions, the thylakoid membrane is not connected to the inner chloroplast membrane at any point.

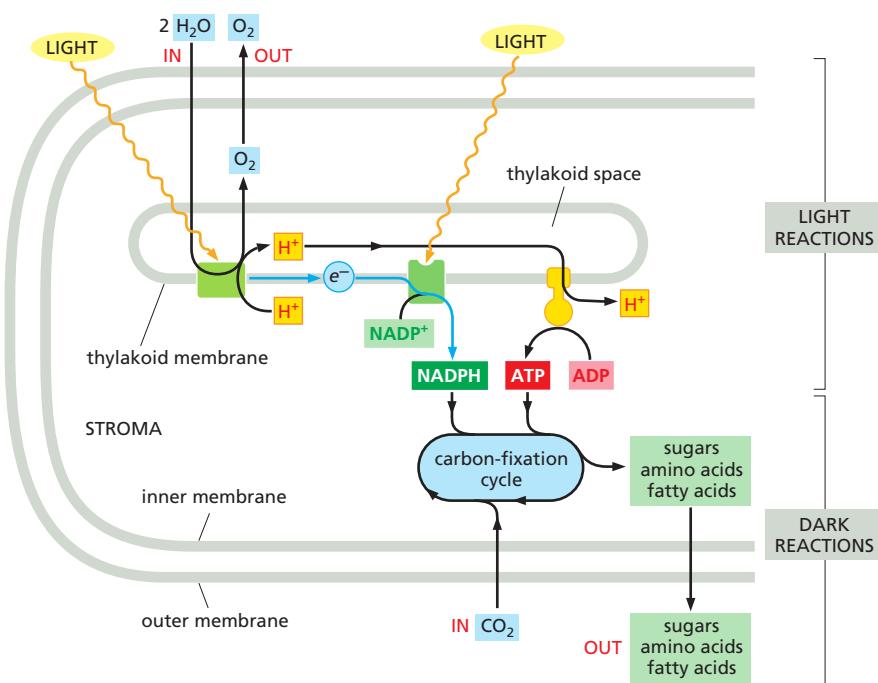
electrochemical proton gradient drives the synthesis of ATP in the stroma. As the final step in this series of reactions, electrons are loaded (together with H<sup>+</sup>) onto NADP<sup>+</sup>, converting it to the energy-rich NADPH molecule. Because the positively charged chlorophyll in the first reaction center quickly regains its electrons from water (H<sub>2</sub>O), O<sub>2</sub> gas is produced as a by-product. All of these reactions are confined to the chloroplast.

- The carbon-fixation reactions do not require sunlight. Here the ATP and NADPH generated by the light reactions serve as the source of energy and reducing power, respectively, to drive the conversion of CO<sub>2</sub> to carbohydrate. These carbon-fixation reactions begin in the chloroplast stroma, where they generate the three-carbon sugar *glyceraldehyde 3-phosphate*. This simple sugar is exported to the cytosol, where it is used to produce sucrose and many other organic metabolites in the leaves of the plant. The sucrose is then exported to meet the metabolic needs of the nonphotosynthetic plant tissues, serving as a source of both carbon skeletons and energy for growth.

Thus, the formation of ATP, NADPH, and O<sub>2</sub> (which requires light energy directly) and the conversion of CO<sub>2</sub> to carbohydrate (which requires light energy only indirectly) are separate processes (Figure 14–39). However, they are linked by elaborate feedback mechanisms that allow a plant to manufacture sugars only when it is appropriate to do so. Several of the chloroplast enzymes required for carbon fixation, for example, are inactive in the dark and reactivated by light-stimulated electron-transport processes.

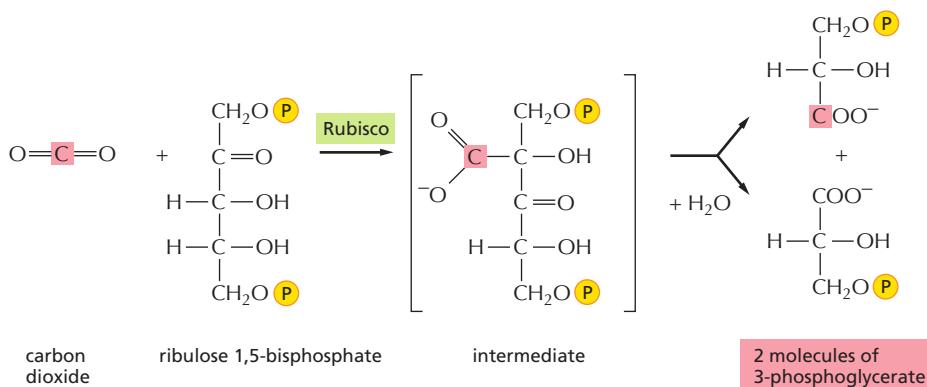
### Carbon Fixation Uses ATP and NADPH to Convert CO<sub>2</sub> into Sugars

We have seen earlier in this chapter how animal cells produce ATP by using the large amount of free energy released when carbohydrates are oxidized to CO<sub>2</sub> and H<sub>2</sub>O. The reverse reaction, in which plants make carbohydrate from CO<sub>2</sub> and H<sub>2</sub>O, takes place in the chloroplast stroma. The large amounts of ATP and NADPH produced by the photosynthetic electron-transfer reactions are required to drive this energetically unfavorable reaction.



**Figure 14–39** A summary of the energy-converting metabolism in chloroplasts. Chloroplasts require only water and carbon dioxide as inputs for their light-driven photosynthesis reactions, and they produce the nutrients for most other organisms on the planet. Each oxidation of two water molecules by a photochemical reaction center in the thylakoid membrane produces one molecule of oxygen, which is released into the atmosphere. At the same time, protons are concentrated in the thylakoid space. These protons create a large electrochemical gradient across the thylakoid membrane, which is utilized by the chloroplast ATP synthase to produce ATP from ADP and phosphate. The electrons withdrawn from water are transferred to a second type of photochemical reaction center to produce NADPH from NADP<sup>+</sup>. As indicated, the NADPH and ATP are fed into the *carbon-fixation cycle* to reduce carbon dioxide, thereby producing the precursors for sugars, amino acids, and fatty acids. The CO<sub>2</sub> that is taken up from the atmosphere here is the source of the carbon atoms for most organic molecules on Earth.

In a plant cell, a variety of metabolites produced in the chloroplast are exported to the cytoplasm for biosyntheses. Some of the sugar produced is stored in the form of starch granules in the chloroplast, but the rest is transported throughout the plant as sucrose or converted to starch in special storage tissues. These storage tissues serve as a major food source for animals.



**Figure 14–40** The initial reaction in carbon fixation. This carboxylation reaction allows one molecule each of carbon dioxide and water to be incorporated into organic carbon molecules. It is catalyzed in the chloroplast stroma by the abundant enzyme ribulose bisphosphate carboxylase, or Rubisco. As indicated, the product is two molecules of 3-phosphoglycerate.

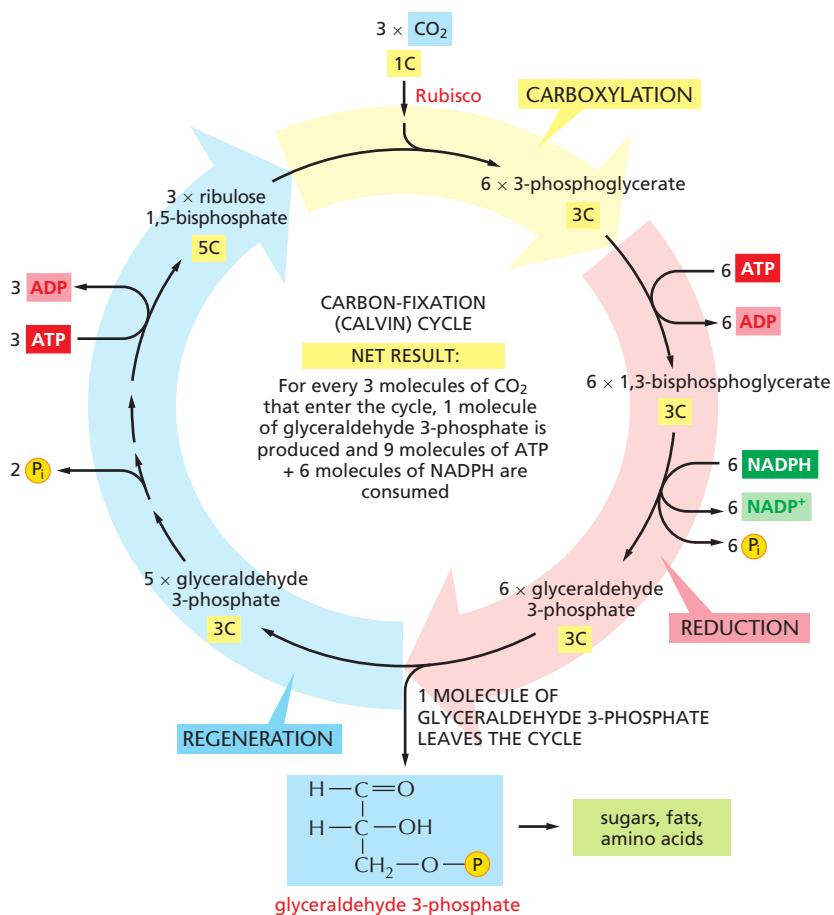
**Figure 14–40** illustrates the central reaction of **carbon fixation**, in which an atom of inorganic carbon is converted to organic carbon:  $\text{CO}_2$  from the atmosphere combines with the five-carbon compound ribulose 1,5-bisphosphate plus water to yield two molecules of the three-carbon compound 3-phosphoglycerate. This carboxylation reaction is catalyzed in the chloroplast stroma by a large enzyme called *ribulose bisphosphate carboxylase*, or *Rubisco* for short. Because the reaction is so slow (each Rubisco molecule turns over only about 3 molecules of substrate per second, compared to 1000 molecules per second for a typical enzyme), an unusually large number of enzyme molecules are needed. Rubisco often constitutes more than 50% of the chloroplast protein mass, and it is thought to be the most abundant protein on Earth. In a global context, Rubisco also keeps the amount of the greenhouse gas  $\text{CO}_2$  in the atmosphere at a low level.

Although the production of carbohydrates from  $\text{CO}_2$  and  $\text{H}_2\text{O}$  is energetically unfavorable, the fixation of  $\text{CO}_2$  catalyzed by Rubisco is an energetically favorable reaction. Carbon fixation is energetically favorable because a continuous supply of the energy-rich ribulose 1,5-bisphosphate is fed into the process. This compound is consumed by the addition of  $\text{CO}_2$ , and it must be replenished. The energy and reducing power needed to regenerate ribulose 1,5-bisphosphate come from the ATP and NADPH produced by the photosynthetic light reactions.

The elaborate series of reactions in which  $\text{CO}_2$  combines with ribulose 1,5-bisphosphate to produce a simple sugar—a portion of which is used to regenerate ribulose 1,5-bisphosphate—forms a cycle, called the *carbon-fixation cycle*, or the *Calvin cycle* (Figure 14–41). This cycle was one of the first metabolic pathways to be worked out by applying radioisotopes as tracers in biochemistry. As indicated, each turn of the cycle converts six molecules of 3-phosphoglycerate to three molecules of ribulose 1,5-bisphosphate plus one molecule of glyceraldehyde 3-phosphate. *Glyceraldehyde 3-phosphate*, the three-carbon sugar produced by the cycle, then provides the starting material for the synthesis of many other sugars and all of the other organic molecules that form the plant.

### Sugars Generated by Carbon Fixation Can Be Stored as Starch or Consumed to Produce ATP

The glyceraldehyde 3-phosphate generated by carbon fixation in the chloroplast stroma can be used in a number of ways, depending on the needs of the plant. During periods of excess photosynthetic activity, much of it is retained in the chloroplast stroma and converted to *starch*. Like glycogen in animal cells, starch is a large polymer of glucose that serves as a carbohydrate reserve, and it is stored as large granules in the chloroplast stroma. Starch forms an important part of the diet of all animals that eat plants. Other glyceraldehyde 3-phosphate molecules are converted to fat in the stroma. This material, which accumulates as fat droplets, likewise serves as an energy reserve. At night, this stored starch and fat can be broken down to sugars and fatty acids, which are exported to the cytosol to help support the metabolic needs of the plant. Some of the exported sugar enters the



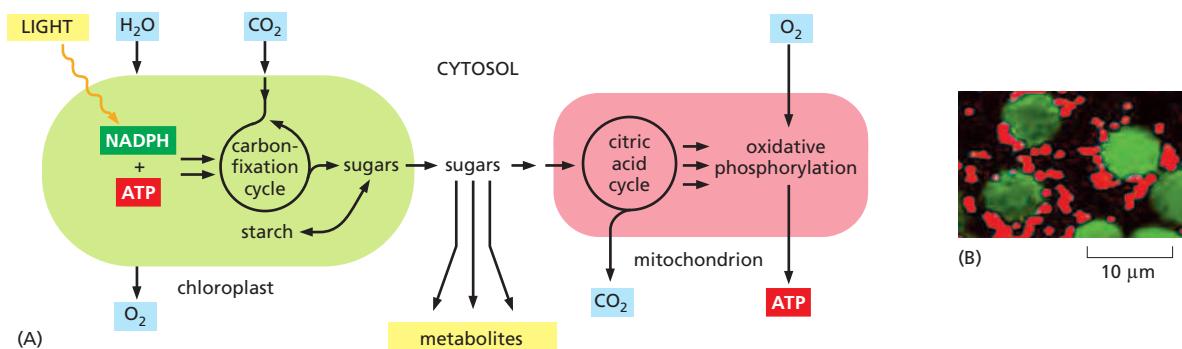
**Figure 14–41** The carbon-fixation cycle. This central metabolic pathway allows organic molecules to be produced from  $\text{CO}_2$  and  $\text{H}_2\text{O}$ . In the first stage of the cycle (carboxylation),  $\text{CO}_2$  is added to ribulose 1,5-bisphosphate, as shown in Figure 14–40. In the second stage (reduction), ATP and NADPH are consumed to produce glyceraldehyde 3-phosphate molecules. In the final stage (regeneration), some of the glyceraldehyde 3-phosphate produced is used to regenerate ribulose 1,5-bisphosphate. Other glyceraldehyde 3-phosphate molecules are either converted to starch and fat in the chloroplast stroma, or transported out of the chloroplast into the cytosol. The number of carbon atoms in each type of molecule is indicated in yellow. There are many intermediates between glyceraldehyde 3-phosphate and ribulose 5-phosphate, but they have been omitted here for clarity. The entry of water into the cycle is also not shown (but see Figure 14–40).

glycolytic pathway (see Figure 2–46), where it is converted to pyruvate. Both that pyruvate and the fatty acids can enter the plant cell mitochondria and be fed into the citric acid cycle, ultimately leading to the production of large amounts of ATP by oxidative phosphorylation (Figure 14–42). Plants use this ATP in the same way that animal cells and other nonphotosynthetic organisms do to power a variety of metabolic reactions.

The glyceraldehyde 3-phosphate exported from chloroplasts into the cytosol can also be converted into many other metabolites, including the disaccharide *sucrose*. Sucrose is the major form in which sugar is transported between the cells of a plant: just as glucose is transported in the blood of animals, so sucrose is exported from the leaves to provide carbohydrate to the rest of the plant.

### The Thylakoid Membranes of Chloroplasts Contain the Protein Complexes Required for Photosynthesis and ATP Generation

We next need to explain how the large amounts of ATP and NADPH required for carbon fixation are generated in the chloroplast. Chloroplasts are much larger and less dynamic than mitochondria, but they make use of chemiosmotic energy conversion in much the same way. As we saw in Figure 14–38, chloroplasts and mitochondria are organized on the same principles, although the chloroplast contains a separate thylakoid membrane system in which its chemiosmotic mechanisms occur. The thylakoid membranes contain two large membrane protein complexes, called *photosystems*, which endow plants and other photosynthetic organisms with the ability to capture and convert solar energy for their own use. Two other protein complexes in the thylakoid membrane that work together with the photosystems in photophosphorylation—the generation of ATP with sunlight—have mitochondrial equivalents. These are the heme-containing cytochrome



**Figure 14–42 How chloroplasts and mitochondria collaborate to supply cells with both metabolites and ATP.** (A) The inner chloroplast membrane is impermeable to the ATP and NADPH that are produced in the stroma during the light reactions of photosynthesis. These molecules are therefore funneled into the carbon-fixation cycle, where they are used to make sugars. The resulting sugars and their metabolites are either stored within the chloroplast—in the form of starch or fat—or exported to the rest of the plant cell. There, they can enter the energy-generating pathway that ends in ATP synthesis linked to oxidative phosphorylation inside the mitochondrion. Unlike the chloroplast, mitochondrial membranes contain a specific transporter that makes them permeable to ATP (see Figure 14–34). Note that the  $O_2$  released to the atmosphere by photosynthesis in chloroplasts is used for oxidative phosphorylation in mitochondria; similarly, the  $CO_2$  released by the citric acid cycle in mitochondria is used for carbon fixation in chloroplasts. (B) In a leaf, mitochondria (red) tend to cluster close to the chloroplasts (green), as seen in this light micrograph. (B, courtesy of Olivier Grandjean.)

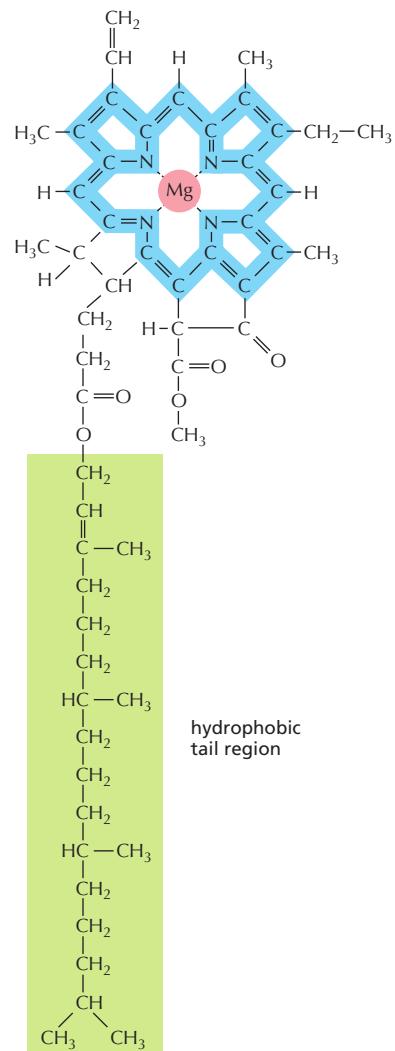
$b_{6-f}$  complex, which both functionally and structurally resembles cytochrome *c* reductase in the respiratory chain; and the chloroplast ATP synthase, which closely resembles the mitochondrial ATP synthase and works in the same way.

### Chlorophyll–Protein Complexes Can Transfer Either Excitation Energy or Electrons

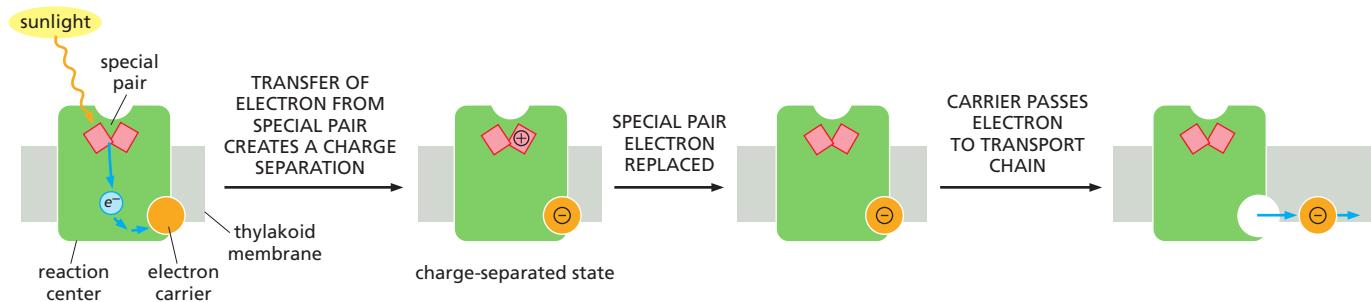
The photosystems in the thylakoid membrane are multiprotein assemblies of a complexity comparable to that of the protein complexes in the mitochondrial electron-transport chain. They contain large numbers of specifically bound chlorophyll molecules, in addition to cofactors that will be familiar from our discussion of mitochondria (heme, iron-sulfur clusters, and quinones). **Chlorophyll**, the green pigment of photosynthetic organisms, has a long hydrophobic tail that makes it behave like a lipid, plus a porphyrin ring that has a central Mg atom and an extensive system of delocalized electrons in conjugated double bonds (Figure 14–43). When a chlorophyll molecule absorbs a quantum of sunlight (a photon), the energy of the photon causes one of these electrons to move from a low-energy molecular orbital to another orbital of higher energy.

The excited electron in a chlorophyll molecule tends to return quickly to its ground state, which can occur in one of three ways:

1. By converting the extra energy into heat (molecular motion) or to some combination of heat and light of a longer wavelength (fluorescence); this is what usually happens when light is absorbed by an isolated chlorophyll molecule in solution.
2. By transferring the energy—but not the electron—directly to a neighboring chlorophyll molecule by a process called *resonance energy transfer*.
3. By transferring the excited electron with its negative charge to another nearby molecule, an *electron acceptor*, after which the positively charged chlorophyll returns to its original state by taking up an electron from some other molecule, an *electron donor*.



**Figure 14–43 The structure of chlorophyll.** A magnesium atom is held in a porphyrin ring, which is related to the porphyrin ring that binds iron in heme (see Figure 14–15). Electrons are delocalized over the bonds shaded in blue.



The latter two mechanisms occur when chlorophylls are attached to proteins in a *chlorophyll-protein complex*. The protein coordinates the central Mg atom in the chlorophyll porphyrin, most often through a histidine side chain located in the hydrophobic interior of a membrane, causing each of the chlorophylls in a protein complex to be held at exactly defined distances and orientations. The flow of excitation energy or electrons then depends on both the precise spatial arrangement and the local protein environment of the protein-bound chlorophylls.

When excited by a photon, most protein-bound chlorophylls simply transmit the absorbed energy to another nearby chlorophyll by the process of resonance energy transfer. However, in a few specially positioned chlorophylls, the energy difference between the ground state and the excited state is just right for the photon to trigger a light-induced chemical reaction. The special state of such chlorophyll molecules derives from their close interaction with a second chlorophyll molecule in the same chlorophyll-protein complex. Together, these two chlorophylls form a *special pair*.

The photosynthetic electron transfer process starts when a photon of suitable energy ionizes a chlorophyll molecule in such a special pair, dissociating it into an electron and a positively charged chlorophyll ion. The energized electron is passed rapidly to a quinone in the same protein complex, preventing its unproductive reassociation with the chlorophyll ion. This light-induced transfer of an electron from a chlorophyll to a mobile electron carrier is the central **charge-separation** step in photosynthesis, in which a chlorophyll becomes positively charged and an electron carrier becomes negatively charged (Figure 14-44). The chlorophyll ion is a very strong oxidant that is able to withdraw an electron from a low-energy substrate; in the first step of oxygenic photosynthesis, this low-energy substrate is water.

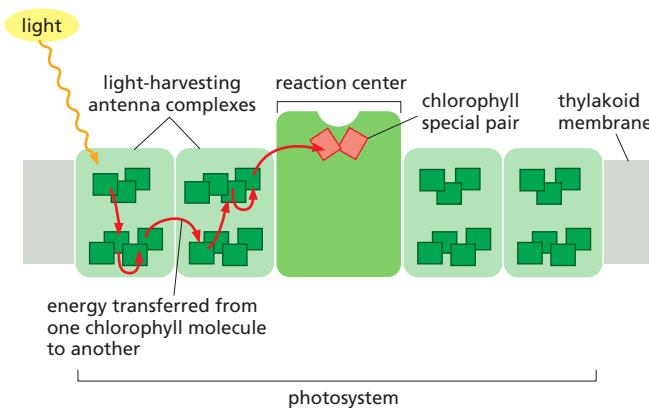
Upon transfer to a mobile carrier in the electron-transport chain, the electron is stabilized as part of a strong electron donor and made available for subsequent reactions. These subsequent reactions require more time to complete, and they result in light-generated energy-rich compounds.

### A Photosystem Consists of an Antenna Complex and a Reaction Center

There are two distinct types of chlorophyll-protein complexes in the photosynthetic membrane. One type, called a *photochemical reaction center*, contains the special pair of chlorophylls just described. The other type engages exclusively in light absorption and resonance energy transfer and is called an *antenna complex*. Together, the two types of complex make up a **photosystem** (Figure 14-45).

The role of the **antenna complex** in the photosystem is to collect the energy of a sufficient number of photons for photosynthesis. Without it, the process would be slow and inefficient, as each reaction-center chlorophyll would absorb only about one light quantum per second, even in broad daylight, whereas hundreds per second are needed for effective photosynthesis. When light excites a chlorophyll molecule in the antenna complex, the energy passes rapidly from one protein-bound chlorophyll to another by resonance energy transfer until it reaches the special pair in the reaction center. The antenna complex is also known as a

**Figure 14-44** A general scheme for the charge-separation step in a photosynthetic reaction center. In a reaction center, light energy is harnessed to generate electrons that are held at a high energy level by mobile electron carriers in a membrane. Light energy is thereby converted to chemical energy. The process starts when a photon absorbed by the special pair of chlorophylls in the reaction center knocks an electron out of one of the chlorophylls. The electron is taken up by a mobile electron carrier (orange) bound at the opposite membrane surface. A set of intermediary carriers embedded in the reaction center provide the path from the special pair to this carrier (not shown). The physical distance between the positively charged chlorophyll ion and the negatively charged electron carrier stabilizes the charge-separated state for a short time, during which the chlorophyll ion, a strong oxidant, withdraws an electron from a suitable compound (for example, from water, an event we will discuss in detail shortly). The electron carrier then diffuses away from the reaction center as a strong electron donor that will transfer its electron to an electron-transport chain.



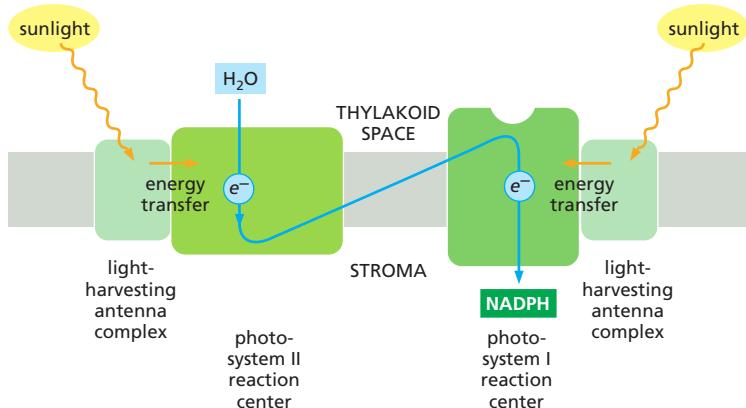
**Figure 14–45 A photosystem.** Each photosystem consists of a reaction center plus a number of light-harvesting antenna complexes. The solar energy for photosynthesis is collected by the antenna complexes, which account for most of the chlorophyll in a plant cell. The energy hops randomly by resonance energy transfer (red arrows) from one chlorophyll molecule to another, until it reaches the reaction center complex, where it ionizes a chlorophyll in the special pair. The chlorophyll special pair holds its electrons at a lower energy than the chlorophyll in the antenna complexes, causing the energy transferred to it from the antenna complex to become trapped there. Note that it is only energy that moves from one chlorophyll molecule to another in the antenna complex, not electrons (Movie 14.10).

**light-harvesting complex**, or LHC. In addition to many chlorophyll molecules, an LHC contains orange carotenoid pigments. The carotenoids collect light of a different wavelength from that absorbed by chlorophylls, helping to make the antenna complex more efficient. They also have an important protective role in preventing the formation of harmful oxygen radicals in the photosynthetic membrane.

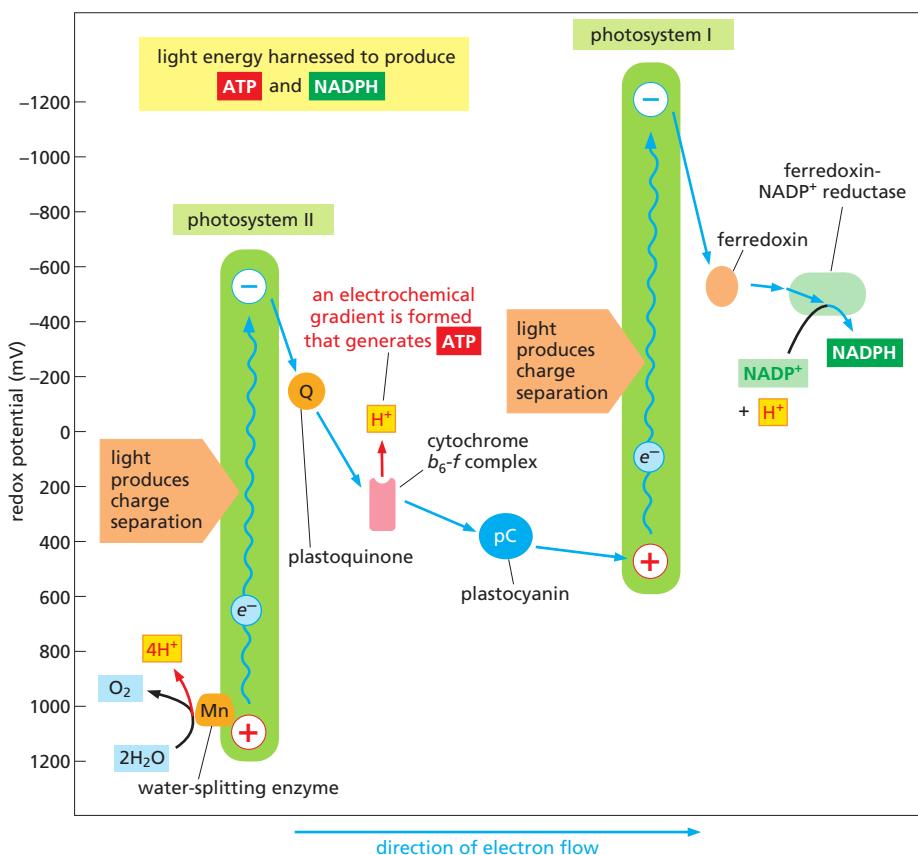
### The Thylakoid Membrane Contains Two Different Photosystems Working in Series

The excitation energy collected by the antenna complex is delivered to the special pair in the **photochemical reaction center**. The reaction center is a transmembrane chlorophyll-protein complex that lies at the heart of photosynthesis. It harbors the special pair of chlorophyll molecules, which acts as an irreversible trap for excitation energy (see Figure 14–45).

Chloroplasts contain two functionally different although structurally related photosystems, each of which feeds electrons generated by the action of sunlight into an electron-transfer chain. In the chloroplast thylakoid membrane, *photosystem I* is confined to the unstacked stroma thylakoids, while the stacked grana thylakoids contain *photosystem II*. The two photosystems were named in order of their discovery, not of their actions in the photosynthetic pathway, and electrons are first activated in *photosystem II* before being transferred to *photosystem I* (Figure 14–46). The path of the electron through the two photosystems can be described as a Z-like trajectory and is known as the *Z scheme*. In the Z scheme, the reaction center of *photosystem II* first withdraws an electron from water. The electron passes via an electron-transport chain (composed of the electron carrier plastoquinone, the cytochrome  $b_6-f$  complex, and the protein plastocyanin) to *photosystem I*, which propels the electron across the membrane in a second light-driven charge-separation reaction that leads to NADPH production.



**Figure 14–46 The Z scheme for photosynthesis.** The thylakoids of plants and cyanobacteria contain two different photosystems, known as photosystem I and photosystem II, which work in series. Each of the photosystem I and II reaction centers receives excitation energy from its own set of tightly associated antenna complexes, known as LHC-I and LHC-II, by resonance energy transfer. Note that, for historical reasons, the two photosystems were named opposite to the order in which they act, with photosystem II passing its electrons to photosystem I.



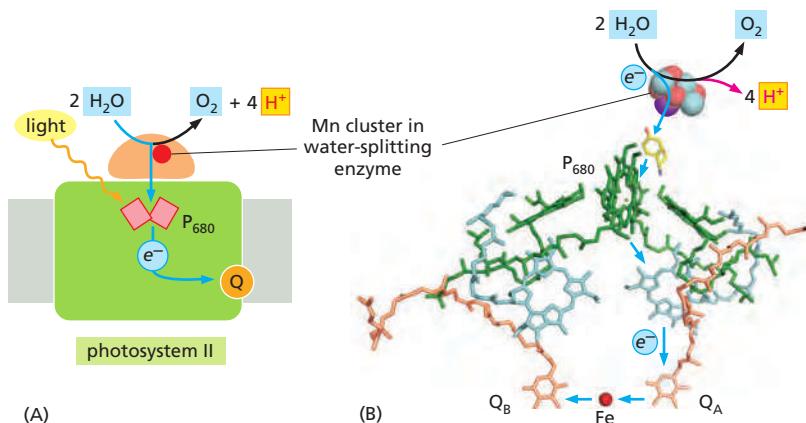
**Figure 14-47 Changes in redox potential during photosynthesis.** The redox potential for each molecule is indicated by its position along the vertical axis. Photosystem II passes electrons derived from water to photosystem I, which in turn passes them to NADP<sup>+</sup> through ferredoxin-NADP<sup>+</sup> reductase. The net electron flow through the two photosystems is from water to NADP<sup>+</sup>, and it produces NADPH as well as an electrochemical proton gradient. This proton gradient is used by the ATP synthase to produce ATP. Details in this figure will be explained in the subsequent text.

The Z scheme is necessary to bridge the very large energy gap between water and NADPH (Figure 14-47). A single quantum of visible light does not contain enough energy both to withdraw electrons from water, which holds on to its electrons very tightly (redox potential +820 mV) and therefore is a very poor electron donor, and to force them on to NADP<sup>+</sup>, which is a very poor electron acceptor (redox potential -320 mV). The Z scheme first evolved in cyanobacteria to enable them to use water as a universally available electron source. Other, simpler photosynthetic bacteria have only one photosystem. As we shall see, they cannot use water as an electron source and must rely on other, more energy-rich substrates instead, from which electrons are more readily withdrawn. The ability to extract electrons from water (and thereby to produce molecular oxygen) was acquired by plants when their ancestors took up the endosymbiotic cyanobacteria that later evolved into chloroplasts (see Figure 1-31).

### Photosystem II Uses a Manganese Cluster to Withdraw Electrons From Water

In biology, only photosystem II is able to withdraw electrons from water and to generate molecular oxygen as a waste product. This remarkable specialization of photosystem II is conferred by the unique properties of one of the two chlorophyll molecules of its special pair and by a *manganese cluster* linked to the protein. These chlorophyll molecules and the manganese cluster form the catalytic core of the photosystem II reaction center, whose mechanism is outlined in Figure 14-48.

Water is an inexhaustible source of electrons, but it is also extremely stable; therefore a large amount of energy is required to make it part with its electrons. The only compound in living organisms that is able to achieve this feat after its ionization by light, is the chlorophyll special pair called P<sub>680</sub> (P<sub>680</sub>/P<sub>680</sub><sup>+</sup> redox potential = +1270 mV). The reaction 2H<sub>2</sub>O + 4 photons → 4H<sup>+</sup> + 4e<sup>-</sup> + O<sub>2</sub> is catalyzed by its adjacent manganese cluster. The intermediates remain firmly attached to the manganese cluster until two water molecules have been fully oxidized to O<sub>2</sub>, thus



**Figure 14-48** The conversion of light energy to chemical energy in the photosystem II complex. (A) Schematic diagram of the photosystem II reaction center, whose special pair of chlorophyll molecules is designated as  $P_{680}$  based on the wavelength of its absorbance maximum (680 nm). (B) Cofactors and pigments at the core of the reaction center. Shown are the manganese (Mn) cluster, the tyrosine side chain that links it to the  $P_{680}$  special pair, four chlorophylls (green), two pheophytins (light blue), two plastoquinones (pink), and an iron atom (red). The path of electrons is shown by blue arrows. In the manganese cluster, four manganese atoms (light blue), one calcium atom (purple), and five oxygen atoms (red) work together to catalyze the oxidation of water. The water-splitting reaction occurs in four successive steps, each requiring the energy of one photon. Each photon turns a  $P_{680}$  reaction-center chlorophyll into a positively charged chlorophyll ion. Through an ionized tyrosine side chain (yellow), this chlorophyll ion pulls an electron away from a water molecule bound at the manganese cluster. In this way, a total of four electrons are withdrawn from two water molecules to generate molecular oxygen, which is released into the atmosphere.

ensuring that no dangerous oxygen radicals are released as the reaction proceeds. The protons released by the two water molecules are discharged to the thylakoid space, contributing to the proton gradient across the thylakoid membrane (pH lower in the thylakoid space than in the stroma). The unique protein environment that endows life with this all-important ability to oxidize water has remained essentially unchanged throughout billions of years of evolution (Figure 14-49).

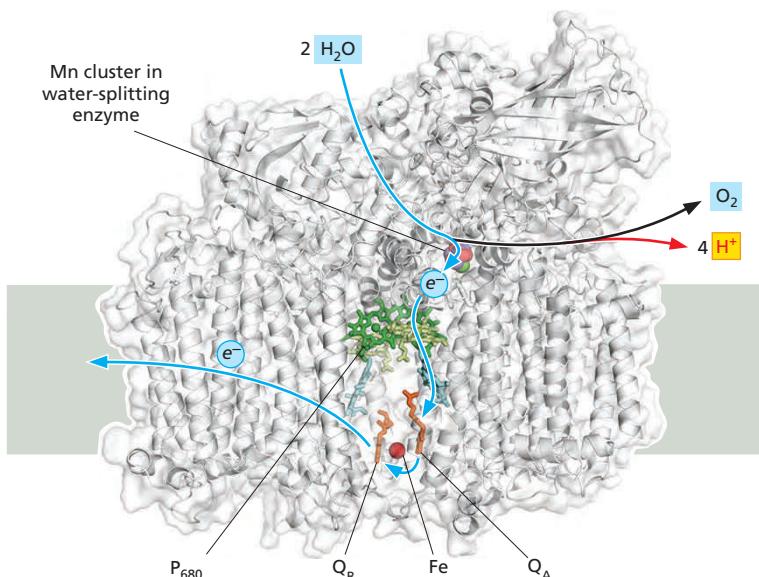
All of the oxygen in the Earth's atmosphere has been generated in this way. Although the exact details of the water-oxidation reaction in photosystem II are still not fully understood, scientists are trying to construct an artificial system that mimics the process. If successful, this might provide a virtually endless supply of clean energy, helping to solve the world's energy crisis.

### The Cytochrome $b_6-f$ Complex Connects Photosystem II to Photosystem I

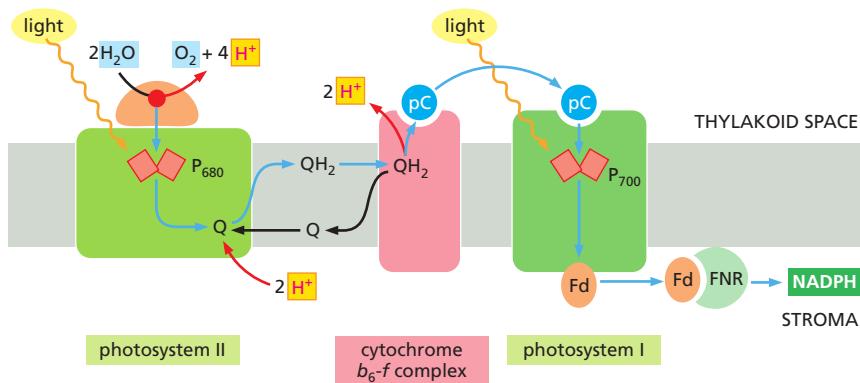
Following the path shown previously in Figure 14-48, the electrons extracted from water by photosystem II are transferred to plastoquinol, a strong electron donor similar to ubiquinol in mitochondria. This quinol, which can diffuse rapidly in the lipid bilayer of the thylakoid membrane, transfers its electrons to the *cytochrome  $b_6-f$  complex*, whose structure is homologous to the cytochrome  $c$  reductase in mitochondria. The cytochrome  $b_6-f$  complex pumps  $H^+$  into the thylakoid space using the same Q cycle that is utilized in mitochondria (see Figure 14-21), thereby adding to the proton gradient across the thylakoid membrane.

The cytochrome  $b_6-f$  complex forms the connecting link between photosystems II and I in the chloroplast electron-transport chain. It passes its electrons

Each electron that is energized by light passes from the special pair along an electron-transfer chain inside the complex, along the indicated path to the permanently bound plastoquinone  $Q_A$  and then to plastoquinone  $Q_B$  as electron acceptors. Once  $Q_B$  has picked up two electrons (plus two protons; see Figure 14-17), it dissociates from its binding site in the complex and enters the lipid bilayer as a mobile electron carrier, being immediately replaced by a new, nonreduced molecule of plastoquinone. Note that the chlorophylls and pheophytins form two symmetrical branches of a potential electron-transport chain. Only one branch is active, thus ensuring that the plastoquinones become fully reduced in minimum time.



**Figure 14-49** The structure of the complete photosystem II complex. This photosystem contains at least 16 protein subunits, along with 36 chlorophylls, two pheophytins, two hemes, and a number of protective carotenoids (colored). Most of these pigments and cofactors are deeply buried, tightly complexed to protein (gray). The path of electrons is indicated by the blue arrows, and is explained in Figure 14-48B. The photosystem II complex presented here is the cyanobacterial complex, which is simpler and more stable than the plant complex, which works in the same way. (PDB code: 3ARC.)

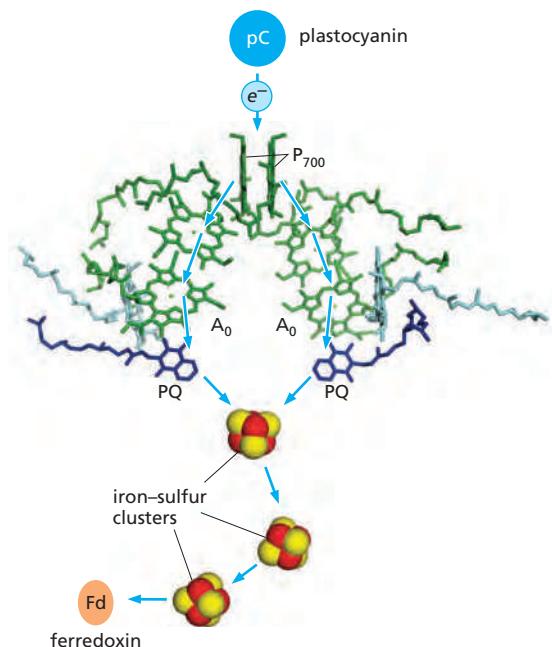


one at a time to the mobile electron carrier plastocyanin (a small copper-containing protein that takes the place of the cytochrome *c* in mitochondria), which will transfer them to photosystem I (Figure 14–50). As we discuss next, photosystem I then harnesses a second photon of light to further energize the electrons that it receives.

### Photosystem I Carries Out the Second Charge-Separation Step in the Z Scheme

Photosystem I receives electrons from plastocyanin in the thylakoid space and transfers them, via a second charge-separation reaction, to the small protein ferredoxin on the opposite membrane surface (Figure 14–51). Then, in a final step, ferredoxin feeds its electrons to a membrane-associated enzyme complex, the *ferredoxin-NADP<sup>+</sup> reductase*, which uses the electrons to produce NADPH from NADP<sup>+</sup> (see Figure 14–50).

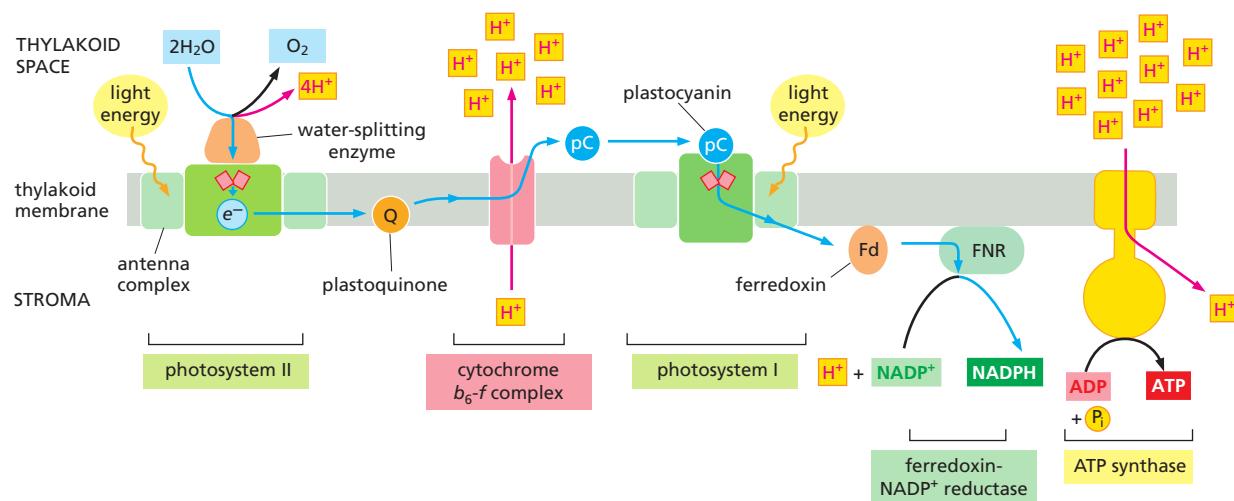
The redox potential of the NADP<sup>+</sup>/NADPH pair ( $-320$  mV) is already very low, and reduction of NADP<sup>+</sup> therefore requires a compound with an even lower redox potential. This turns out to be a chlorophyll molecule near the stromal membrane surface of photosystem I that has a redox potential of  $-1000$  mV (chlorophyll A<sub>0</sub>), making it the strongest known electron donor in biology. The reduced NADPH is released into the chloroplast stroma, where it is used for biosynthesis of glyceraldehyde 3-phosphate, amino acid precursors, and fatty acids, much of it to be exported to the cytoplasm.



**Figure 14–50** Electron flow through the cytochrome *b*<sub>6</sub>-*f* complex to NADPH.

The cytochrome *b*<sub>6</sub>-*f* complex is the functional equivalent of cytochrome *c* reductase (the cytochrome *b*-*c*<sub>1</sub> complex) in mitochondria (see Figure 14–22). Like its mitochondrial homolog, the *b*<sub>6</sub>-*f* complex receives its electrons from a quinone and engages in a complicated Q cycle that pumps two protons across the membrane (details not shown). It hands its electrons, one at a time, to plastocyanin (pC). Plastocyanin diffuses along the membrane surface to photosystem I and transfers the electrons via ferredoxin (Fd) to the ferredoxin-NADP<sup>+</sup> reductase (FNR), where they are utilized to produce NADPH. P<sub>700</sub> is a special pair of chlorophylls that absorbs light of wavelength 700 nm.

**Figure 14–51** Structure and function of photosystem I. At the heart of the photosystem I complex assembly is the electron-transfer chain shown. At one end is a special pair of chlorophylls called P<sub>700</sub> (because it absorbs light of 700 nm wavelength), receiving electrons from plastocyanin (pC). At the other end are the A<sub>0</sub> chlorophylls, which hand the electrons on to ferredoxin via two plastoquinones (PQ; purple) and three iron-sulfur clusters. Even though the roles of photosystems I and II in photosynthesis are very different, their central electron-transfer chains are structurally similar, indicating a common evolutionary origin (see Figure 14–53). Note that in photosystem I both branches of the electron-transfer chain are active, unlike in photosystem II (see Figure 14–48). (PDB code: 3LW5.)



### The Chloroplast ATP Synthase Uses the Proton Gradient Generated by the Photosynthetic Light Reactions to Produce ATP

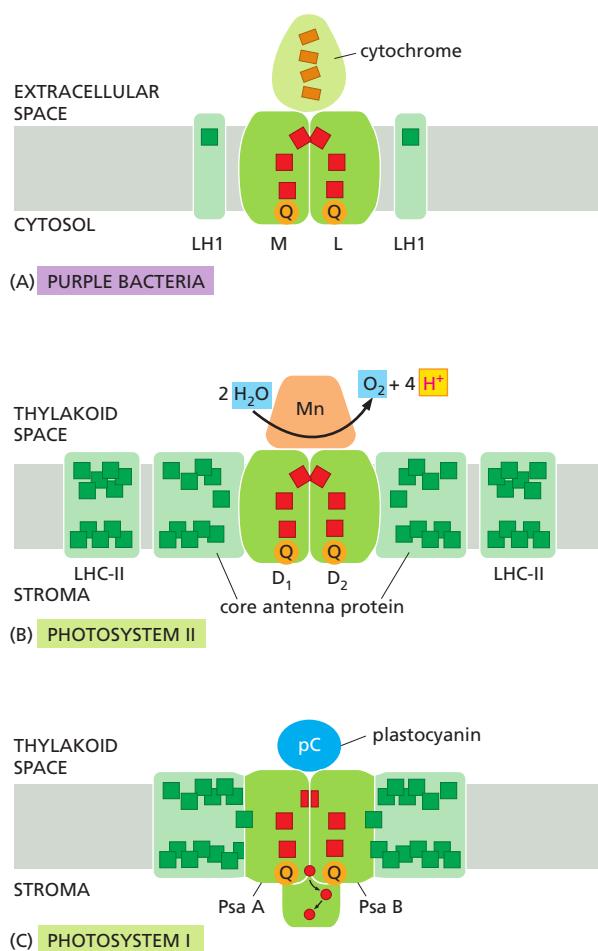
The sequence of events that results in light-driven production of ATP and NADPH in chloroplasts and cyanobacteria is summarized in **Figure 14–52**. Starting with the withdrawal of electrons from water, the light-driven charge-separation steps in photosystems II and I enable the energetically unfavorable (uphill) flow of electrons from water to NADPH (see Figure 14–47). Three small mobile electron carriers—plastoquinone, plastocyanin, and ferredoxin—participate in this process. Together with the electron-driven proton pump of the cytochrome  $b_6-f$  complex, the photosystems generate a large proton gradient across the thylakoid membrane. The ATP synthase molecules embedded in the thylakoid membranes then harness this proton gradient to produce large amounts of ATP in the chloroplast stroma, mimicking the synthesis of ATP in the mitochondrial matrix.

The linear Z scheme for photosynthesis thus far discussed can switch to a circular mode of electron flow through photosystem I and the  $b_6-f$  complex. Here, the reduced ferredoxin diffuses back to the  $b_6-f$  complex to reduce plastoquinone, instead of passing its electrons to the ferredoxin-NADP<sup>+</sup> reductase enzyme complex. This, in effect, turns photosystem I into a light-driven proton pump, thereby increasing the proton gradient and thus the amount of ATP made by the ATP synthase. An elaborate set of regulatory mechanisms control this switch, which enables the chloroplast to generate either more NADPH (linear mode) or more ATP (circular mode), depending on the metabolic needs of the cell.

### All Photosynthetic Reaction Centers Have Evolved From a Common Ancestor

Evidence for the prokaryotic origins of mitochondria and chloroplasts abounds in their genetic systems, as we will see in the next section. But strong and direct evidence for the evolutionary origins of chloroplasts can also be found in the molecular structures of photosynthetic reaction centers revealed in recent years by crystallography. The positions of the chlorophylls in the special pair and the two branches of the electron-transfer chain are basically the same in photosystem I, photosystem II, and the photochemical reaction centers of photosynthetic bacteria (**Movie 14.11**). As a result, one can conclude that they all have evolved from a common ancestor. Evidently, the molecular architecture of the photosynthetic reaction center originated only once and has remained essentially unchanged during evolution. By contrast, the less critical antenna systems have evolved in several different ways and are correspondingly diverse in present-day photosynthetic organisms (**Figure 14–53**).

**Figure 14–52 Summary of electron and proton movements during photosynthesis in the thylakoid membrane.** Electrons are withdrawn, through the action of light energy, from a water molecule that is held by the manganese cluster in photosystem II. The electrons pass on to plastocyanin, which delivers them to the cytochrome  $b_6-f$  complex that resembles the cytochrome  $c$  reductase of mitochondria and the  $b-c$  complex of bacteria. They are then carried to photosystem I by the soluble electron carrier plastocyanin, the functional equivalent of cytochrome  $c$  in mitochondria. From photosystem I they are transferred to ferredoxin-NADP<sup>+</sup> reductase (FNR) by the soluble carrier ferredoxin (Fd; a small protein containing an iron-sulfur center). Protons are pumped into the thylakoid space by the cytochrome  $b_6-f$  complex, in the same way that protons are pumped into mitochondrial cristae by cytochrome  $c$  reductase (see Figure 14–21). In addition, the  $H^+$  released into the thylakoid space by water oxidation, and the  $H^+$  consumed during NADPH formation in the stroma, contribute to the generation of the electrochemical  $H^+$  gradient across the thylakoid membrane. As illustrated, this gradient drives ATP synthesis by an ATP synthase that sits in the same membrane (see Figure 14–47).



**Figure 14-53 Evolution of photosynthetic reaction centers.**

Pigments involved in light-harvesting are colored green; those involved in the central photochemical events are colored red. (A) The primitive photochemical reaction center of purple bacteria contains two related protein subunits, L and M, that bind the pigments involved in the central process of photosynthesis, including a special pair of chlorophyll molecules. Electrons are fed into the excited chlorophylls by a cytochrome. LH1 is a bacterial antenna complex. (B) Photosystem II contains the D<sub>1</sub> and D<sub>2</sub> proteins, which are homologous to the L and M subunits in (A). The excited P<sub>680</sub> chlorophyll in the special pair withdraws electrons from water held by the manganese cluster. LHC-II is the light-harvesting complex that feeds energy into the core antenna proteins. (C) Photosystem I contains the Psa A and Psa B proteins, each of which is equivalent to a fusion of the D<sub>1</sub> or D<sub>2</sub> protein to a core antenna protein of photosystem II. The loosely bound plastocyanin (pC) feeds electrons into the excited chlorophyll pair. As indicated, in photosystem I, electrons are passed from a bound quinone (Q) through a series of three iron–sulfur centers (red circles). (Modified from K. Rhee, E. Morris, J. Barber and W. Kühlbrandt, *Nature* 396:283–286, 1998; and W. Kühlbrandt, *Nature* 411:896–899, 2001. With permission from Macmillan Publishers Ltd.)

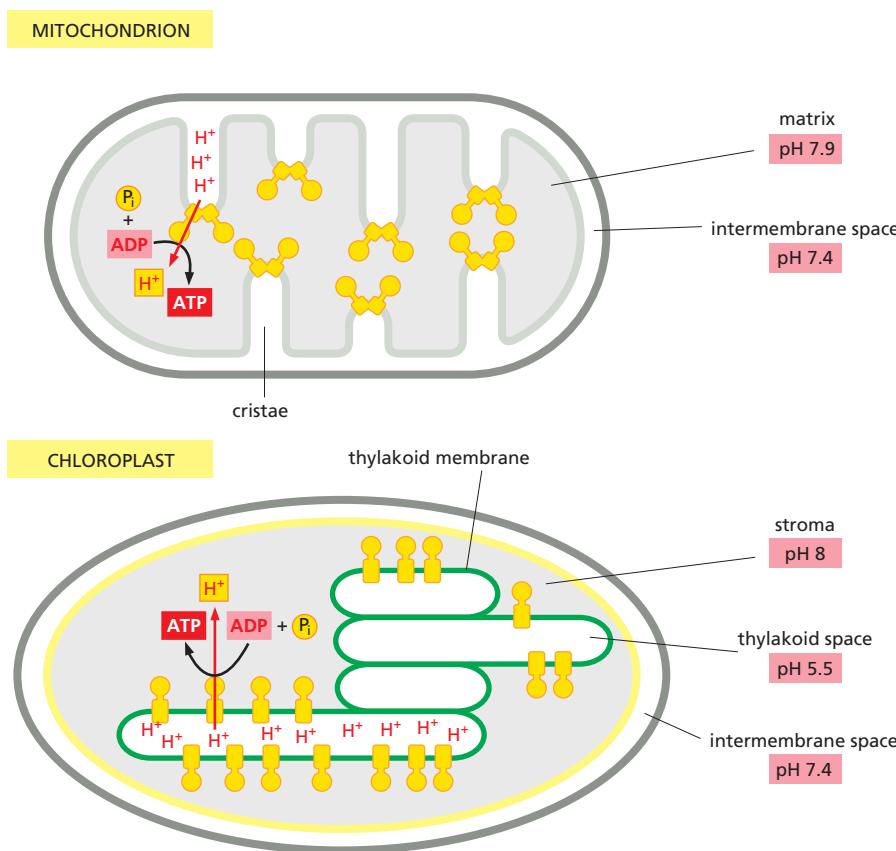
### The Proton-Motive Force for ATP Production in Mitochondria and Chloroplasts Is Essentially the Same

The proton gradient across the thylakoid membrane depends both on the proton-pumping activity of the cytochrome *b*<sub>6</sub>-*f* complex and on the photosynthetic activity of the two photosystems, which in turn depends on light intensity. In chloroplasts exposed to light, H<sup>+</sup> is pumped out of the stroma (pH around 8, similar to the mitochondrial matrix) into the thylakoid space (pH 5–6), creating a gradient of 2–3 pH units across the thylakoid membrane, representing a proton-motive force of about 180 mV. This is very similar to the proton-motive force in respiring mitochondria. However, a membrane potential across the inner mitochondrial membrane makes the largest contribution to the proton-motive force that drives the mitochondrial ATP synthase to make ATP, whereas a H<sup>+</sup> gradient predominates for chloroplasts.

In contrast to mitochondrial ATP synthase, which forms long rows of dimers along the cristae ridges, the chloroplast ATP synthase is monomeric and located in flat membrane regions (Figure 14-54). Evidently, the H<sup>+</sup> gradient across the thylakoid membrane is high enough for ATP synthesis without the need for the elaborate arrangement of ATP synthase seen in mitochondria.

### Chemiosmotic Mechanisms Evolved in Stages

The first living cells on Earth may have consumed geochemically produced organic molecules and generated their ATP by fermentation. Because oxygen was not yet present in the atmosphere, such anaerobic fermentation reactions would have dumped organic acids—such as lactic or formic acids, for example—into the



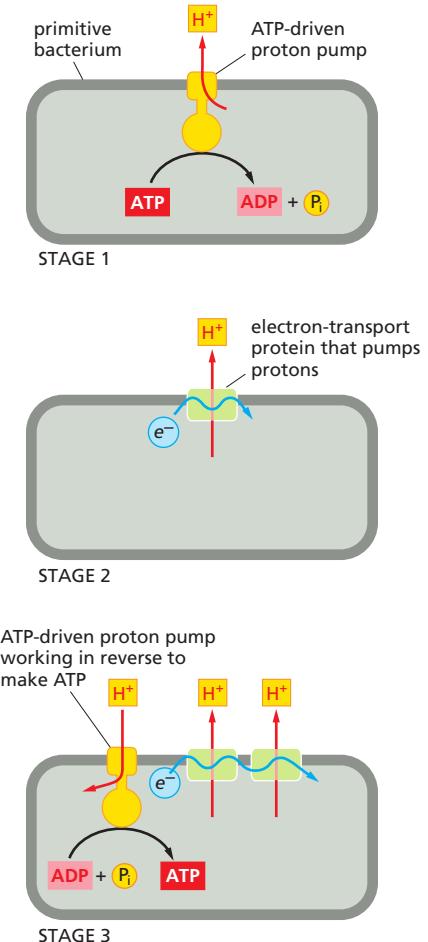
**Figure 14–54** A comparison of H<sup>+</sup> concentrations and the arrangement of ATP synthase in mitochondria and chloroplasts. In both organelles, the pH in the intermembrane space is 7.4, as in the cytoplasm. The pH of the mitochondrial matrix and the pH of the chloroplast stroma are both about 8 (light gray). The pH in the thylakoid space is around 5.5, depending on photosynthetic activity. This results in a high proton-motive force across the thylakoid membrane, consisting largely of the H<sup>+</sup> gradient (a high permeability of this membrane to Mg<sup>2+</sup> and Cl<sup>-</sup> ions allows the flow of these ions to dissipate most of the membrane potential).

In contrast to chloroplasts, the H<sup>+</sup> gradient across the inner mitochondrial membrane is insufficient for ATP production, and mitochondria need a membrane potential to bring the proton-motive force to the same level as in chloroplasts. The arrangement of the mitochondrial ATP synthase in rows of dimers along the cristae ridges (see Figure 14–32) next to the respiratory-chain proton pumps may help the flow of protons along the membrane surface toward the ATP synthase, as the availability of protons is limiting for ATP production. In the chloroplast, the ATP synthase is distributed randomly in thylakoid membranes.

environment (see Figure 2–47). Perhaps such acids lowered the pH of the environment, favoring the survival of cells that evolved transmembrane proteins that could pump H<sup>+</sup> out of the cytosol, thereby preventing the cell from becoming too acidic (stage 1 in Figure 14–55). One of these pumps may have used the energy available from ATP hydrolysis to eject H<sup>+</sup> from the cell; such a proton pump could have been the ancestor of present-day ATP synthases.

As the Earth's supply of geochemically produced nutrients began to dwindle, organisms that could find a way to pump H<sup>+</sup> without consuming ATP would have been at an advantage: they could save the small amounts of ATP they derived from the fermentation of increasingly scarce foodstuffs to fuel other important activities. This need to conserve resources might have led to the evolution of electron-transport proteins that allowed cells to use the movement of electrons between molecules of different redox potentials as a source of energy for pumping H<sup>+</sup> across the plasma membrane (stage 2 in Figure 14–55). Some of these cells might have used the nonfermentable organic acids that neighboring cells had excreted as waste to provide the electrons needed to feed this electron-transport system. Some present-day bacteria grow on formic acid, for example, using the small amount of redox energy derived from the transfer of electrons from formic acid to fumarate to pump H<sup>+</sup>.

**Figure 14–55** How ATP synthesis by chemiosmosis might have evolved in stages. The first stage could have involved the evolution of an ATPase that pumped protons out of the cell using the energy of ATP hydrolysis. Stage 2 could have involved the evolution of a different proton pump, driven by an electron-transport chain. Stage 3 would then have linked these two systems together to generate a primitive ATP synthase that used the protons pumped by the electron-transport chain to synthesize ATP. An early bacterium with this final system would have had a selective advantage over bacteria with neither of the systems or only one.



Eventually, some bacteria would have developed H<sup>+</sup>-pumping electron-transport systems that were so efficient that they could harvest more redox energy than they needed to maintain their internal pH. Such cells would probably have generated large electrochemical proton gradients, which they could then use to produce ATP. Protons could leak back into the cell through the ATP-driven H<sup>+</sup> pumps, essentially running them in reverse so that they synthesized ATP (stage 3 in Figure 14–55). Because such cells would require much less of the dwindling supply of fermentable nutrients, they would have proliferated at the expense of their neighbors.

### By Providing an Inexhaustible Source of Reducing Power, Photosynthetic Bacteria Overcame a Major Evolutionary Obstacle

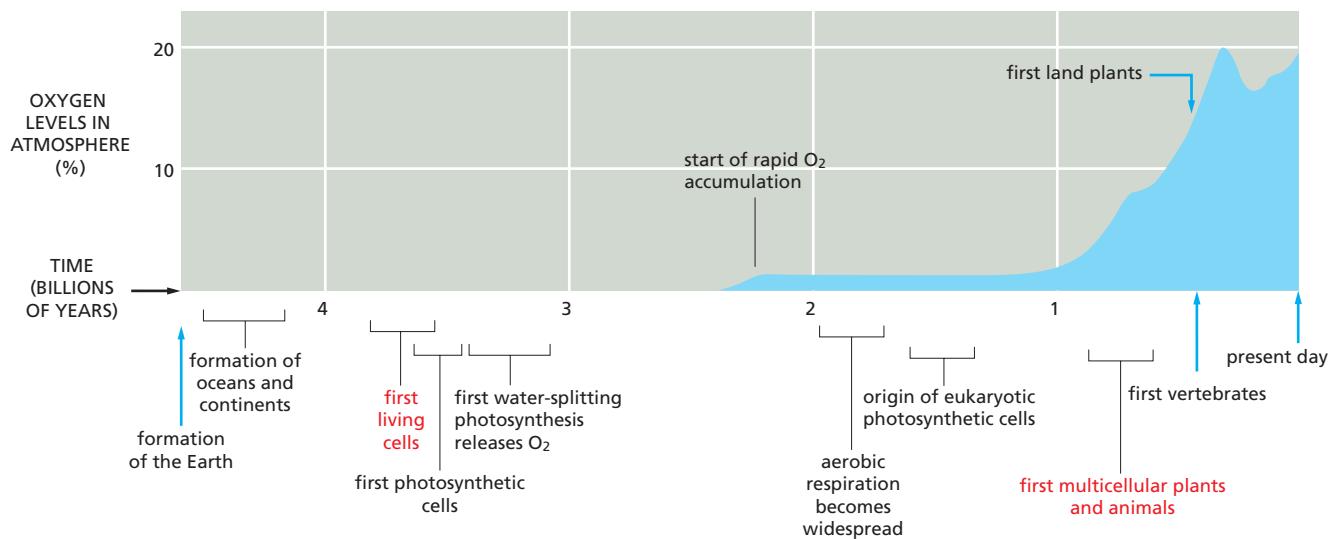
The gradual depletion of nutrients from the environment on the early Earth meant that organisms had to find some alternative source of carbon to make the sugars that serve as the precursors for so many other cell components. Although the CO<sub>2</sub> in the atmosphere provides an abundant potential carbon source, to convert it into an organic molecule such as a carbohydrate requires reducing the fixed CO<sub>2</sub> with a strong electron donor, such as NADPH, which can generate (CH<sub>2</sub>O) units from CO<sub>2</sub> (see Figure 14–41). Early in cellular evolution, strong reducing agents (electron donors) are thought to have been plentiful. But once an ancestor of ATP synthase began to generate most of the ATP, it would have become imperative for cells to evolve a new way of generating strong reducing agents.

A major evolutionary breakthrough in energy metabolism came with the development of photochemical reaction centers that could use the energy of sunlight to produce molecules such as NADPH. It is thought that this occurred early in the process of cellular evolution in the ancestors of the green sulfur bacteria. Present-day green sulfur bacteria use light energy to transfer hydrogen atoms (as an electron plus a proton) from H<sub>2</sub>S to NADPH, thereby producing the strong reducing power required for carbon fixation. Because the redox potential of H<sub>2</sub>S is much lower than that of H<sub>2</sub>O (−230 mV for H<sub>2</sub>S compared with +820 mV for H<sub>2</sub>O), one quantum of light absorbed by the single photosystem in these bacteria is sufficient to generate NADPH via a relatively simple photosynthetic electron-transport chain.

### The Photosynthetic Electron-Transport Chains of Cyanobacteria Produced Atmospheric Oxygen and Permitted New Life-Forms

The next evolutionary step, which is thought to have occurred with the development of the cyanobacteria perhaps 3 billion years ago, was the evolution of organisms capable of using water as the electron source for CO<sub>2</sub> reduction. This entailed the evolution of a water-splitting enzyme and also required the addition of a second photosystem, acting in series with the first, to bridge the large gap in redox potential between H<sub>2</sub>O and NADPH. The biological consequences of this evolutionary step were far-reaching. For the first time, there would have been organisms that could survive on water, CO<sub>2</sub>, and sunlight (plus a few trace elements). These cells would have been able to spread and evolve in ways denied to the earlier photosynthetic bacteria, which needed H<sub>2</sub>S or organic acids as a source of electrons. Consequently, large amounts of biologically synthesized, reduced organic materials accumulated and oxygen entered the atmosphere for the first time.

Oxygen is highly toxic because the oxidation of biological molecules alters their structure and properties indiscriminately and irreversibly. Most anaerobic bacteria, for example, are rapidly killed when exposed to air. Thus, organisms on the primitive Earth would have had to evolve protective mechanisms against the rising O<sub>2</sub> levels in the environment. Late evolutionary arrivals, such as ourselves, have numerous detoxifying mechanisms that protect our cells from the ill effects of oxygen. Even so, an accumulation of oxidative damage to our macromolecules has been postulated to contribute to human aging, as we discuss in the next section.

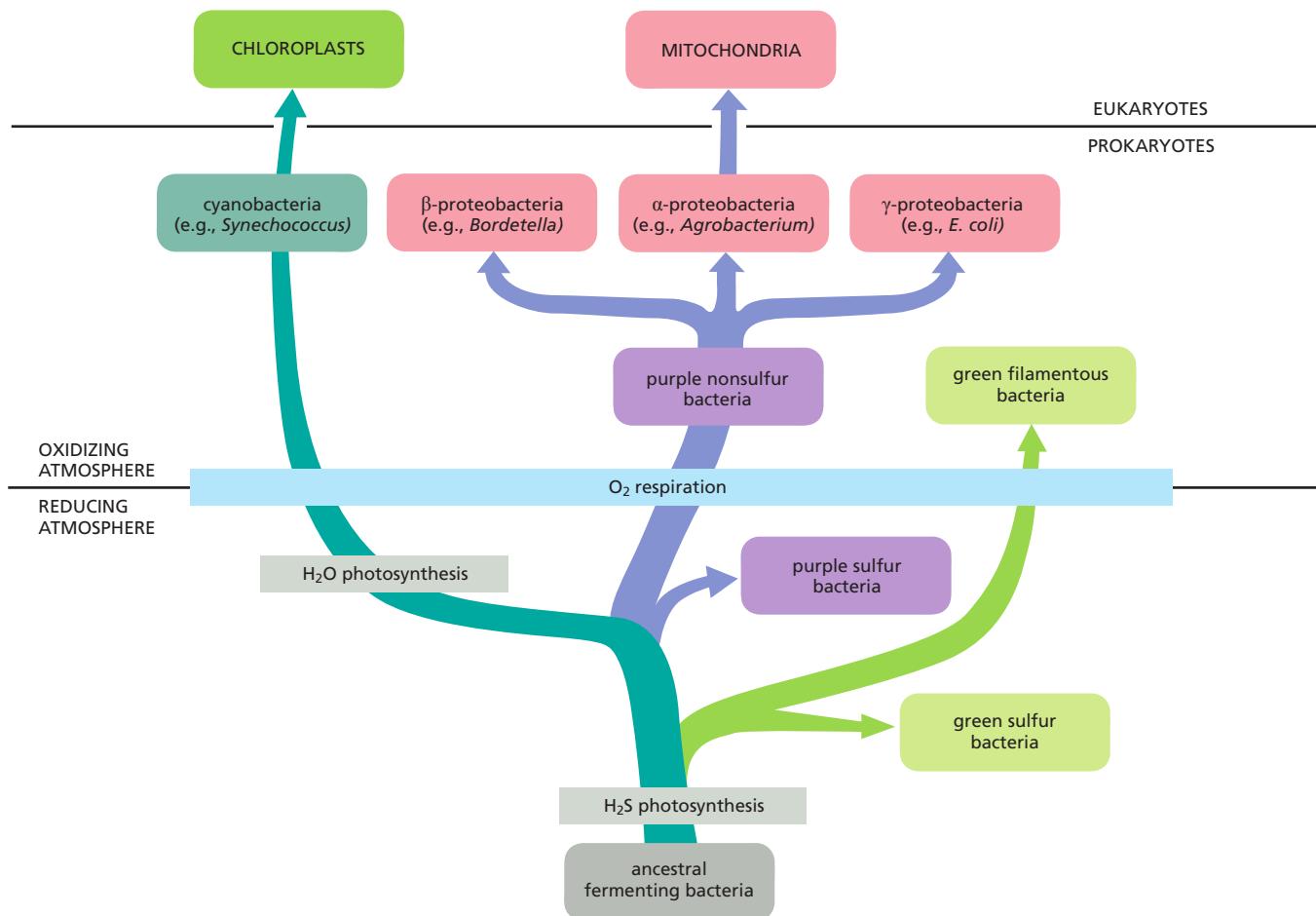


**Figure 14–56 Major events during the evolution of living organisms on Earth.** With the evolution of the membrane-based process of photosynthesis, organisms were able to make their own organic molecules from  $\text{CO}_2$  gas. The delay of more than  $10^9$  years between the appearance of bacteria that split water and released  $\text{O}_2$  during photosynthesis and the accumulation of high levels of  $\text{O}_2$  in the atmosphere is thought to be due to the initial reaction of the oxygen with the abundant ferrous iron ( $\text{Fe}^{2+}$ ) that was dissolved in the early oceans. Only when the ferrous iron was used up would oxygen have started to accumulate in the atmosphere. In response to the rising oxygen levels, nonphotosynthetic oxygen-consuming organisms evolved, and the concentration of oxygen in the atmosphere equilibrated at its present-day level.

The increase in atmospheric  $\text{O}_2$  was very slow at first and would have allowed a gradual evolution of protective devices. For example, the early seas contained large amounts of iron in its reduced, ferrous state ( $\text{Fe}^{2+}$ ), and nearly all the  $\text{O}_2$  produced by early photosynthetic bacteria would have been used up in oxidizing  $\text{Fe}^{2+}$  to ferric  $\text{Fe}^{3+}$ . This conversion caused the precipitation of huge amounts of stable oxides, and the extensive banded iron formations in sedimentary rocks, beginning about 2.7 billion years ago, help to date the spread of the cyanobacteria. By about 2 billion years ago, the supply of  $\text{Fe}^{2+}$  was exhausted, and the deposition of further iron precipitates ceased. Geological evidence reveals how  $\text{O}_2$  levels in the atmosphere have changed over billions of years, approximating current levels only about 0.5 billion years ago (Figure 14–56).

The availability of  $\text{O}_2$  enabled the rise of bacteria that developed an aerobic metabolism to make their ATP. These organisms could harness the large amount of energy released by breaking down carbohydrates and other reduced organic molecules all the way to  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , as explained when we discussed mitochondria. Components of preexisting electron-transport complexes were modified to produce a cytochrome oxidase, so that the electrons obtained from organic or inorganic substrates could be transported to  $\text{O}_2$  as the terminal electron acceptor. Some present-day purple photosynthetic bacteria can switch between photosynthesis and respiration depending on the availability of light and  $\text{O}_2$ , with only relatively minor reorganizations of their electron-transport chains.

In Figure 14–57, we relate these postulated evolutionary pathways to different types of bacteria. By necessity, evolution is always conservative, taking parts of the old and building on them to create something new. Thus, parts of the electron-transport chains that were derived to service anaerobic bacteria 3–4 billion years ago survive, in altered form, in the mitochondria and chloroplasts of today's higher eukaryotes. A good example is the overall similarity in structure and function between the cytochrome *c* reductase that pumps  $\text{H}^+$  in the central segment of the mitochondrial respiratory chain and the analogous cytochrome *b-f* complex in the electron-transport chains of both bacteria and chloroplasts, revealing their common evolutionary origin (Figure 14–58).



**Figure 14–57** Evolutionary scheme showing the postulated origins of mitochondria and chloroplasts and their bacterial ancestors. The consumption of oxygen by respiration is thought to have first developed about 2 billion years ago. Nucleotide-sequence analyses suggest that an endosymbiotic oxygen-evolving cyanobacterium (cyan) gave rise to chloroplasts (dark green), while mitochondria arose from an α-proteobacterium. The nearest relatives of mitochondria (pink) are members of three closely related groups of α-proteobacteria—the rhizobacteria, agrobacteria, and rickettsias—known to form intimate associations with present-day eukaryotic cells. Proteobacteria are pink, purple photosynthetic bacteria are purple, and other photosynthetic bacteria are light green.

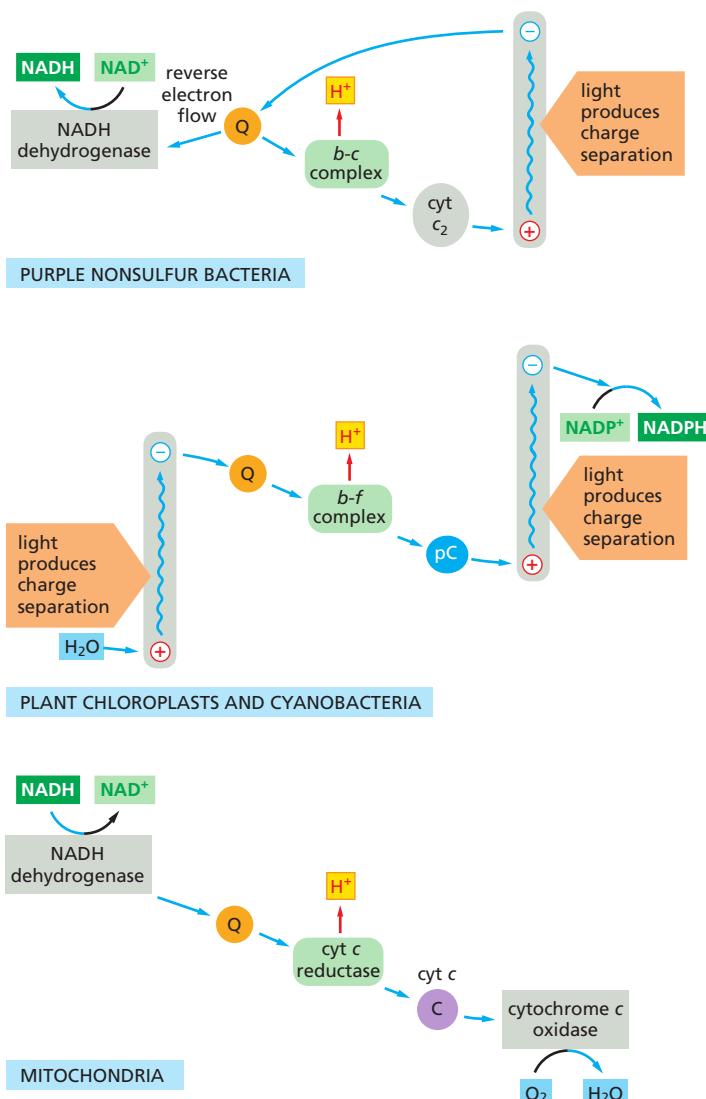
## Summary

Chloroplasts and photosynthetic bacteria have the unique ability to harness the energy of sunlight to produce energy-rich compounds. This is achieved by the photosystems, in which chlorophyll molecules attached to proteins are excited when hit by a photon. Photosystems are composed of an antenna complex that collects solar energy and a photochemical reaction center, in which the collected energy is funneled to a chlorophyll molecule held in a special position, enabling it to withdraw electrons from an electron donor. Chloroplasts and cyanobacteria contain two distinct photosystems. The two photosystems are normally linked in series in the Z scheme, and they transfer electrons from water to NADP<sup>+</sup> to form NADPH, generating a transmembrane electrochemical potential. One of the two photosystems—photosystem II—can split water by removing electrons from this ubiquitous, low-energy compound. All the molecular oxygen (O<sub>2</sub>) in our atmosphere is a by-product of the water-splitting reaction in this photosystem. The three-dimensional structures of photosystems I and II are strikingly similar to the photosystems of purple photosynthetic bacteria, demonstrating a remarkable degree of conservation over billions of years of evolution.

The two photosystems and the cytochrome b<sub>6</sub>-f complex reside in the thylakoid membrane, a separate membrane system in the central stroma compartment of the

chloroplast that is differentiated into stacked grana and unstacked stroma thylakoids. Electron-transport processes in the thylakoid membrane cause protons to be released into the thylakoid space. The backflow of protons through the chloroplast ATP synthase then generates ATP. This ATP is used in conjunction with the NADPH produced by photosynthesis to drive a large number of biosynthetic reactions in the chloroplast stroma, including the carbon-fixation cycle, which generates large amounts of carbohydrates from  $\text{CO}_2$ .

In the early evolution of life, cyanobacteria overcame a major obstacle in devising a way to use solar energy to split water and fix carbon dioxide. Cyanobacteria produced both abundant organic nutrients and molecular oxygen, enabling the rise of a multitude of aerobic life-forms. The chloroplasts in plants have evolved from a cyanobacterium that was endocytosed long ago by an aerobic eukaryotic host organism.



**Figure 14–58** A comparison of three electron-transport chains discussed in this chapter.

Bacteria, chloroplasts, and mitochondria all contain a membrane-bound enzyme complex that resembles the cytochrome *c* reductase of mitochondria. These complexes all accept electrons from a quinone carrier (*Q*) and pump  $\text{H}^+$  across their respective membranes. Moreover, in reconstituted *in vitro* systems, the different complexes can substitute for one another, and the structures of their protein components reveal that they are evolutionarily related. Note that the purple nonsulfur bacteria use a cyclic flow of electrons to produce a large electrochemical proton gradient that drives a reverse electron flow through NADH dehydrogenase to produce NADH from  $\text{NAD}^+ + \text{H}^+ + \text{e}^-$ .

## THE GENETIC SYSTEMS OF MITOCHONDRIA AND CHLOROPLASTS

As we discussed in Chapter 1, mitochondria and chloroplasts are thought to have evolved from endosymbiotic bacteria (see Figures 1–29 and 1–31). Both types of organelles still contain their own genomes (**Figure 14–59**). As we will discuss shortly, they also retain their own biosynthetic machinery for making RNA and organellar proteins.

Like bacteria, mitochondria and chloroplasts proliferate by growth and division of an existing organelle. In actively dividing cells, each type of organelle must double in mass in each cell generation and then be distributed into each daughter cell. In addition, nondividing cells must replenish organelles that are degraded as part of the continual process of organelle turnover, or produce additional organelles as the need arises. Organelle growth and proliferation are therefore carefully controlled. The process is complicated because mitochondrial and chloroplast proteins are encoded in two places: the nuclear genome and the separate genomes harbored in the organelles themselves. The biogenesis of mitochondria and chloroplasts thus requires contributions from two separate genetic systems, which must be closely coordinated.

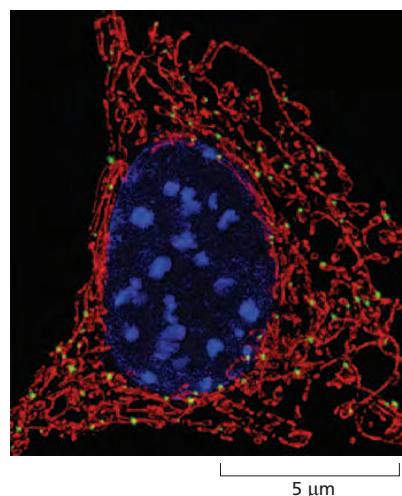
Most organellar proteins are encoded by the nuclear DNA. The organelle imports these proteins from the cytosol, after they have been synthesized on cytosolic ribosomes, through the mitochondrial protein translocases of the outer and inner mitochondrial membrane—TOM and TIM. In Chapter 12, we discussed how this happens. Here, we describe the organelle genomes and genetic systems, and consider the consequences of separate organelle genomes for the cell and the organism as a whole.

### The Genetic Systems of Mitochondria and Chloroplasts Resemble Those of Prokaryotes

As discussed in Chapter 12, it is thought that eukaryotic cells originated through a symbiotic relationship between an archaeon and an aerobic bacterium (a proteobacterium). The two organisms are postulated to have merged to form the ancestor of all nucleated cells, with the archaeon providing the nucleus and the proteobacterium serving as a respiration, ATP-producing endosymbiont—one that would eventually evolve into the mitochondrion (see Figure 12–3). This most likely occurred roughly 1.6 billion years ago, when oxygen had entered the atmosphere in substantial amounts (see Figure 14–56). The chloroplast was derived later, after the plant and animal lineages diverged, through endocytosis of an oxygen-producing cyanobacterium.

This *endosymbiont hypothesis* of organelle development receives strong support from the observation that the genetic systems of mitochondria and chloroplasts are similar to those of present-day bacteria. For example, chloroplast ribosomes are very similar to bacterial ribosomes, both in their structure and in their sensitivity to various antibiotics (such as chloramphenicol, streptomycin, erythromycin, and tetracycline). In addition, protein synthesis in chloroplasts starts with *N*-formylmethionine, as in bacteria, and not with methionine as in the cytosol of eukaryotic cells. Although mitochondrial genetic systems are much less similar to those of present-day bacteria than are the genetic systems of chloroplasts, their ribosomes are also sensitive to antibacterial antibiotics, and protein synthesis in mitochondria also starts with *N*-formylmethionine.

**Figure 14–59 Staining of nuclear and mitochondrial DNA.** In this confocal micrograph of a human fibroblast, the nuclear DNA is stained with the dye DAPI (blue) and mitochondrial DNA is visualized with fluorescent antibodies that bind DNA (green). The mitochondria are stained with fluorescent antibodies that recognize a specialized protein translocase specific to the outer mitochondrial membrane (red). Numerous copies of the mitochondrial genome are distributed in distinct nucleoids throughout the mitochondria that snake through the cytoplasm. (From C. Kukat et al., *Proc. Natl. Acad. Sci. USA* 108:13534–13539, 2011. With permission from the National Academy of Sciences.)

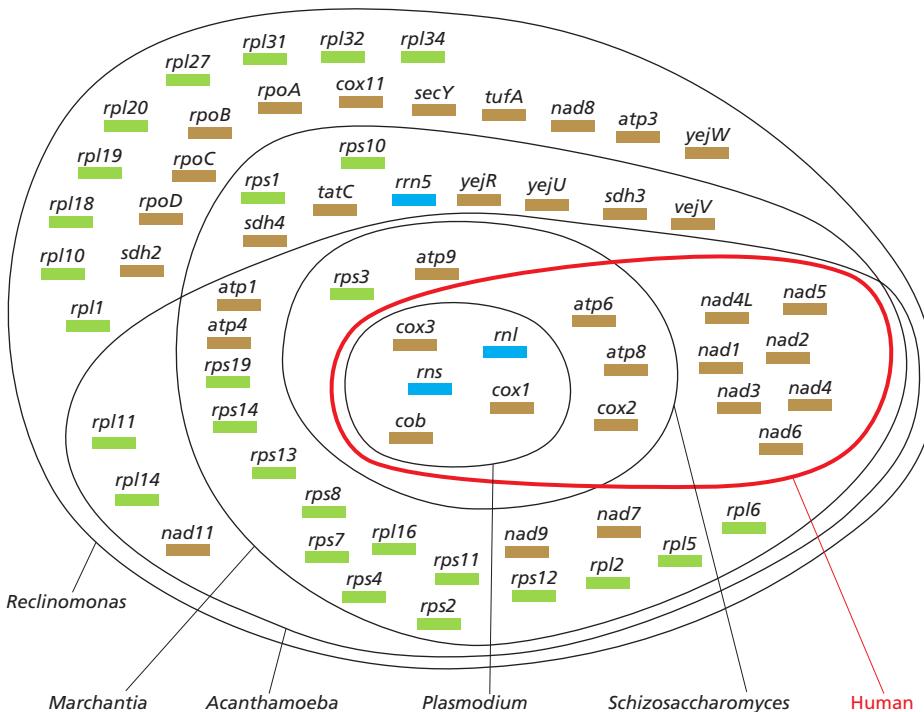


The processes of organelle DNA transcription, protein synthesis, and DNA replication take place where the genome is located: in the matrix of mitochondria or the stroma of chloroplasts. Although the enzymes that mediate these genetic processes are unique to the organelle, and resemble those of bacteria (or even of bacterial viruses) rather than their eukaryotic analogs, the nuclear genome encodes the vast majority of these enzymes. Indeed, most present-day mitochondrial and chloroplast proteins are encoded by genes that reside in the cell nucleus.

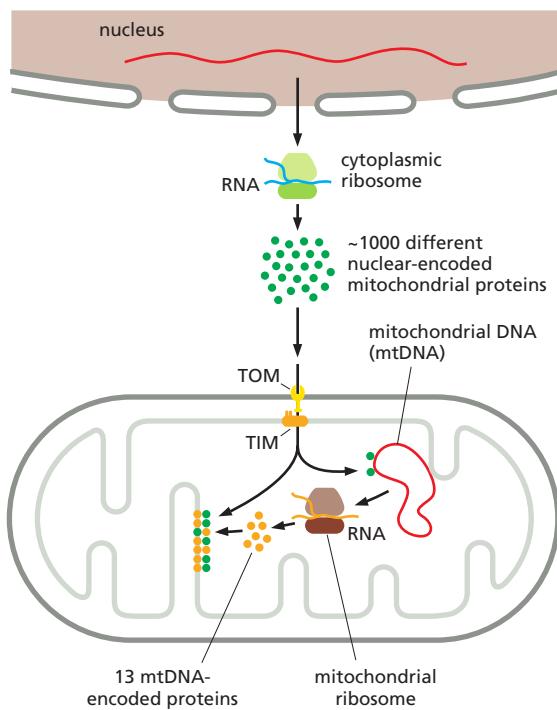
### Over Time, Mitochondria and Chloroplasts Have Exported Most of Their Genes to the Nucleus by Gene Transfer

The nature of the organelle genes located in the nucleus of the cell demonstrates that an extensive transfer of genes from organelle to nuclear DNA has occurred in the course of eukaryotic evolution. Such successful *gene transfer* is expected to be rare, because any gene moved from the organelle needs to adapt to both nuclear transcription and cytoplasmic translation requirements. In addition, the protein needs to acquire a signal sequence that directs it to the correct organelle after its synthesis in the cytosol. By comparing the genes in the mitochondria from different organisms, we can infer that some of the gene transfers to the nucleus occurred relatively recently. The smallest and presumably most highly evolved mitochondrial genomes, for example, encode only a few hydrophobic inner-membrane proteins of the electron-transport chain, plus ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). Other mitochondrial genomes that have remained more complex tend to contain this same subset of genes along with others (Figure 14–60). The most complex mitochondrial genomes include genes that encode components of the mitochondrial genetic system, such as RNA polymerase subunits and ribosomal proteins; these same genes are found in the cell nucleus in yeast and all animal cells.

The proteins that are encoded by genes in the organellar DNA are synthesized on ribosomes within the organelle, using organelle-produced messenger RNA (mRNA) to specify their amino acid sequence (Figure 14–61). The protein traffic between the cytosol and these organelles seems to be unidirectional: proteins are normally not exported from mitochondria or chloroplasts to the cytosol. An important exception occurs when a cell is about to undergo apoptosis. As will be



**Figure 14–60 Comparison of mitochondrial genomes.** Less complex mitochondrial genomes encode subsets of the proteins and ribosomal RNAs that are encoded by larger mitochondrial genomes. In this comparison, there are only five genes that are shared by the six mitochondrial genomes; these encode ribosomal RNAs (*rns* and *m1*), cytochrome *b* (*cob*), and two cytochrome oxidase subunits (*cox1* and *cox3*). Blue indicates ribosomal RNAs; green, ribosomal proteins; and brown, components of the respiratory chain and other proteins. (Adapted from M.W. Gray, G. Burger and B.F. Lang, *Science* 283:1476–1481, 1999.)



**Figure 14–61** Biogenesis of the respiratory-chain proteins in human mitochondria. Most of the protein components of the mitochondrial respiratory chain are encoded by nuclear DNA, with only a small number encoded by mitochondrial DNA (mtDNA). Transcription of mtDNA produces 13 mRNAs, all of which encode subunits of the oxidative phosphorylation system, and the 24 RNAs (22 transfer RNAs and 2 ribosomal RNAs) needed for translation of these mRNAs on the mitochondrial ribosomes (brown).

The mRNAs produced by transcription of nuclear genes are translated on cytoplasmic ribosomes (green), which are distinct from the mitochondrial ribosomes. The nuclear-encoded mitochondrial proteins (dark green) are imported into mitochondria through two protein translocases called TOM and TIM, and constitute the vast majority of the approximately 1000 different protein species present in mammalian mitochondria. The nuclear-encoded mitochondrial proteins in humans include the majority of the oxidative phosphorylation system subunits, all proteins needed for expression and maintenance of mtDNA, and all proteins of the mitochondrial ribosomes.

The mtDNA-encoded subunits (orange) assemble together with the nuclear subunits to form a functional oxidative phosphorylation system. (Adapted from N.G. Larsson, *Annu. Rev. Biochem.* 79:683–706, 2010.)

discussed in detail in Chapter 18, during apoptosis the mitochondrion releases proteins (most notably cytochrome *c*) from the crista space through its outer mitochondrial membrane, as part of an elaborate signaling pathway that is triggered to cause cells to undergo programmed cell death.

### The Fission and Fusion of Mitochondria Are Topologically Complex Processes

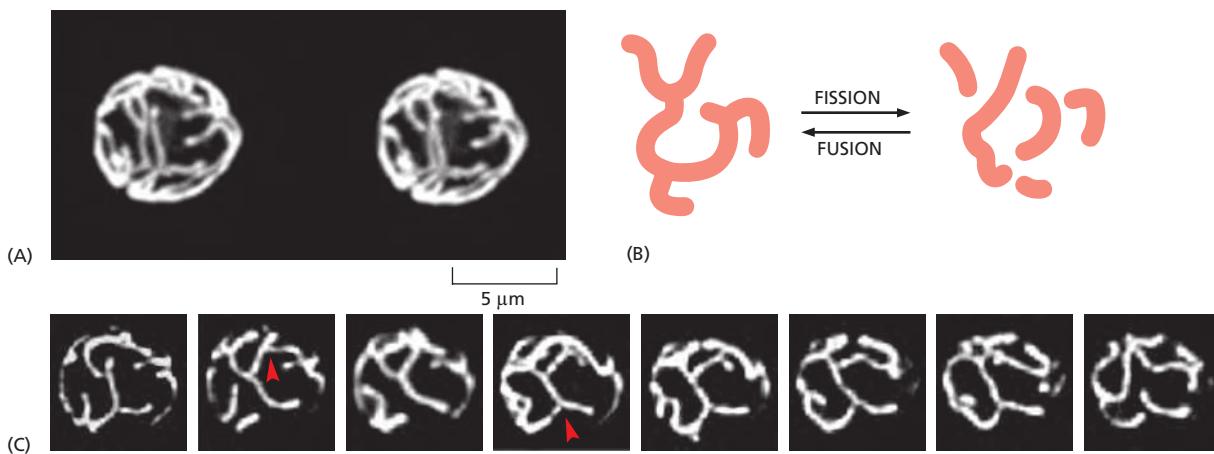
In mammalian cells, mitochondrial DNA makes up less than 1% of the total cellular DNA. In other cells, however, such as the leaves of higher plants or the very large egg cells of amphibians, a much larger fraction of the cellular DNA may be present in mitochondria or chloroplasts (Table 14–2), and a large fraction of the total RNA and protein synthesis takes place in the organelles.

Mitochondria and chloroplasts are large enough to be visible by light microscopy in living cells. For example, mitochondria can be visualized by expressing in cells a genetically engineered fusion of a mitochondrial protein linked to green

**TABLE 14–2** Relative Amounts of Organelle DNA in Some Cells and Tissues

Organism	Tissue or cell type	DNA molecules per organelle	Organelles per cell	Organelle DNA as percentage of total cellular DNA
<b>Mitochondrial DNA</b>				
Rat	Liver	5–10	1000	1
Yeast*	Vegetative	2–50	1–50	15
Frog	Egg	5–10	10 <sup>7</sup>	99
<b>Chloroplast DNA</b>				
<i>Chlamydomonas</i>	Vegetative	80	1	7
Maize	Leaves	0–300**	20–40	0–15**

\*The large variation in the number and size of mitochondria per cell in yeasts is due to mitochondrial fusion and fission. \*\*In maize, the amount of chloroplast DNA drops precipitously in mature leaves, after cell division ceases: the chloroplast DNA is degraded and stable mRNAs persist to provide for protein synthesis.



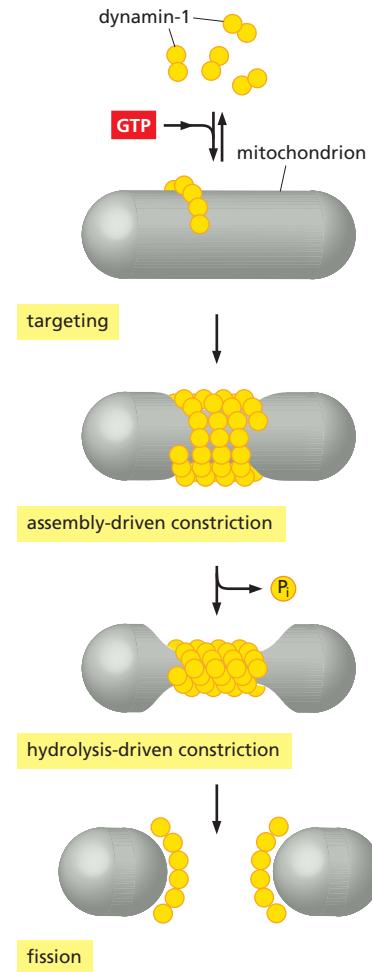
**Figure 14–62** The mitochondrial reticulum is dynamic. (A) In yeast cells, mitochondria form a continuous reticulum on the cytoplasmic side of the plasma membrane (stereo pair). (B) A balance between fission and fusion determines the arrangement of the mitochondria in different cells. (C) Time-lapse fluorescent microscopy shows the dynamic behavior of the mitochondrial network in a yeast cell. In addition to shape changes, fission and fusion constantly remodel the network (red arrows). These pictures were taken at 3-minute intervals. (A and C, from J. Nunnari et al., *Mol. Biol. Cell* 8:1233–1242, 1997. With permission from the American Society for Cell Biology.)

fluorescent protein (GFP), or cells can be incubated with a fluorescent dye that is specifically taken up by mitochondria because of their membrane potential. Such images demonstrate that the mitochondria in living cells are dynamic—frequently dividing by fission, fusing, and changing shape (Figure 14–62 and Movie 14.12). The fission of mitochondria may be necessary so that small parts of the network can pinch off and reach remote regions of the cell—for example in the thin, extended axon and dendrites of a neuron.

The fission and fusion of mitochondria are topologically complex processes that must ensure the integrity of the separate mitochondrial compartments defined by the inner and outer membranes. These processes control the number and shape of mitochondria, which can vary dramatically in different cell types, ranging from multiple spherical or wormlike organelles to a highly branched, net-shaped single organelle called a *reticulum*. Each depends on its own special set of proteins. The mitochondrial fission machine works by assembling dynamin-related GTPases (discussed in Chapter 13) into helical oligomers that cause local constrictions in tubular mitochondria. GTP hydrolysis then generates the mechanical force that severs the inner and outer mitochondrial membranes in one step (Figure 14–63). Mitochondrial fusion requires two separate machineries, one each for the outer and the inner membrane (Figure 14–64). In addition to GTP hydrolysis for force generation, both mechanisms also depend on the mitochondrial proton-motive force for reasons that are still unknown.

### Animal Mitochondria Contain the Simplest Genetic Systems Known

Comparisons of DNA sequences in different organisms reveal that, in vertebrates (including ourselves), the mutation rate during evolution has been roughly 100 times greater in the mitochondrial genome than in the nuclear genome. This difference is likely to be due to lower fidelity of mitochondrial DNA replication,



**Figure 14–63** A model for mitochondrial division. Dynamin-1 (yellow) exists as dimers in the cytosol, which form larger oligomeric structures in a process that requires GTP hydrolysis. Dynamin assemblies interact with the outer mitochondrial membrane through special adaptor proteins, forming a spiral of GTP-dynamin around the mitochondrion that causes a constriction. A concerted GTP-hydrolysis event in the dynamin subunits is then thought to produce the conformational changes that result in fission. (Adapted from S. Hoppins, L. Lackner and J. Nunnari, *Annu. Rev. Biochem.* 76:751–780, 2007.)

inefficient DNA repair, or both, given that the mechanisms that perform these processes in the organelle are relatively simple compared with those in the nucleus. As discussed in Chapter 4, the relatively high rate of evolution of animal mitochondrial genes makes a comparison of mitochondrial DNA sequences especially useful for estimating the dates of relatively recent evolutionary events, such as the steps in primate evolution.

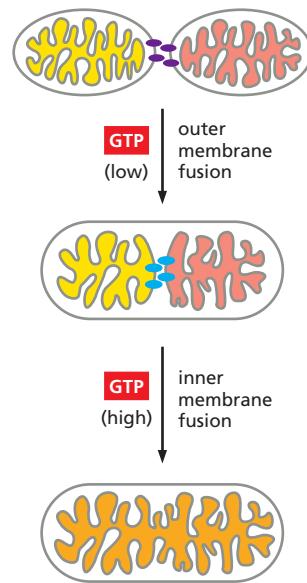
There are 13 protein-encoding genes in human mitochondrial DNA (Figure 14–65). These code for hydrophobic components of the respiratory-chain complexes and of ATP synthase. In contrast, roughly 1000 mitochondrial proteins are encoded in the nucleus, produced on cytosolic ribosomes, and imported by the protein import machinery in the outer and inner membrane (discussed in Chapter 12). It has been suggested that the cytosolic production of hydrophobic membrane proteins and their import into the organelle may present a problem to the cell, and that this is the reason why their genes have remained in the mitochondrion. However, some of the most hydrophobic mitochondrial proteins, such as the *c* subunit of the ATP synthase rotor ring, are imported from the cytosol in some species (though they are mitochondrially encoded in others). And the parasites *Plasmodium falciparum* and *Leishmania tarentolae*, which spend most of their life cycles inside cells of their host organisms, have retained only two or three mitochondrially encoded proteins.

The size range of mitochondrial DNAs is similar to that of viral DNAs. The mitochondrial DNA in *Plasmodium falciparum* (the human malaria parasite) has less than 6000 nucleotide pairs, whereas the mitochondrial DNAs of some land plants contain more than 300,000 nucleotide pairs (Figure 14–66). In animals, the mitochondrial genome is a simple DNA circle of about 16,600 nucleotide pairs (less than 0.001% of the nuclear genome), and it is nearly the same size in organisms as different from us as *Drosophila* and sea urchins.

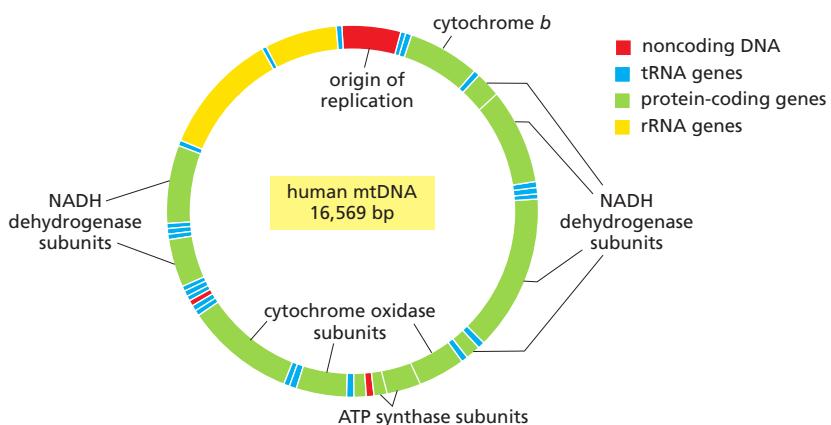
### Mitochondria Have a Relaxed Codon Usage and Can Have a Variant Genetic Code

The human mitochondrial genome has several surprising features that distinguish it from nuclear, chloroplast, and bacterial genomes:

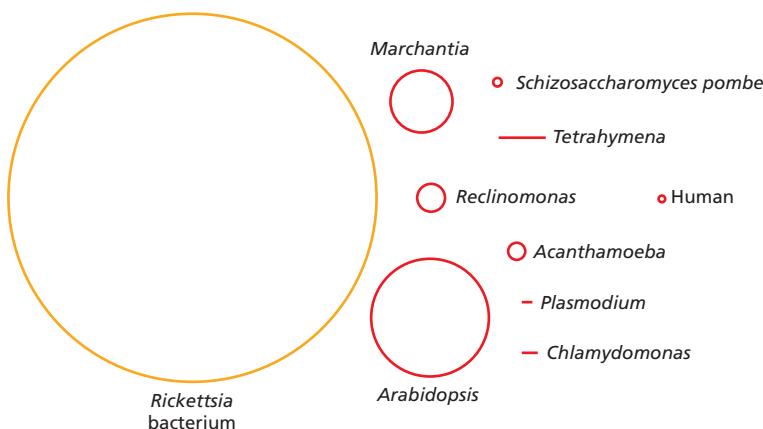
1. *Dense gene packing.* Unlike other genomes, the human mitochondrial genome seems to contain almost no noncoding DNA: nearly every nucleotide seems to be part of a coding sequence, either for a protein or for one of the rRNAs or tRNAs. Since these coding sequences run directly into each other, there is very little room left for regulatory DNA sequences.
2. *Relaxed codon usage.* Whereas 30 or more tRNAs specify amino acids in the cytosol and in chloroplasts, only 22 tRNAs are required for mitochondrial protein synthesis. The normal codon-anticodon pairing rules are relaxed in mitochondria, so that many tRNA molecules recognize any one of the four nucleotides in the third (wobble) position. Such “2 out of 3” pairing



**Figure 14–64 A model for mitochondrial fusion.** The fusions of the outer and inner mitochondrial membranes are coordinated sequential events, each of which requires a separate set of protein factors. Outer membrane fusion is brought about by an outer-membrane GTPase (purple), which forms an oligomeric complex that includes subunits anchored in the two membranes to be fused. Fusion of outer membranes requires GTP and an H<sup>+</sup> gradient across the inner membrane. For fusion of the inner membrane, a dynamin-related protein forms an oligomeric tethering complex (blue) that includes subunits anchored in the two inner membranes to be fused. Fusion of the inner membranes requires GTP and the electrical component of the potential across the inner membrane. (Adapted from S. Hoppins, L. Lackner and J. Nunnari, *Annu. Rev. Biochem.* 76:751–780, 2007.)



**Figure 14–65 The organization of the human mitochondrial genome.** The human mitochondrial genome of ≈16,600 nucleotide pairs contains 2 rRNA genes, 22 tRNA genes, and 13 protein-coding sequences. There are two transcriptional promoters, one for each strand of the mitochondrial DNA (mtDNA). The DNAs of many other animal mitochondrial genomes have been completely sequenced. Most of these animal mitochondrial DNAs encode precisely the same genes as humans, with the gene order being identical for animals ranging from fish to mammals.



**Figure 14–66 Comparison of various sizes of mitochondrial genomes with the genome of bacterial ancestors.** The complete DNA sequences for thousands of mitochondrial genomes have been determined. The lengths of a few of these mitochondrial DNAs are shown to scale—as circles for those genomes thought to be circular and lines for linear genomes. The largest circle represents the genome of *Rickettsia prowazekii*, a small pathogenic bacterium whose genome most closely resembles that of mitochondria. The size of mitochondrial genomes does not correlate well with the number of proteins encoded in them: while human mitochondrial DNA encodes 13 proteins, the 22-fold larger mitochondrial DNA of *Arabidopsis thaliana* encodes only 32 proteins—that is, about 2.5-fold as many as human mitochondrial DNA. The extra DNA that is found in *Arabidopsis*, *Marchantia*, and other plant mitochondria may be “junk DNA”—that is, noncoding DNA with no apparent function. The mitochondrial DNA of the protozoan *Reclinomonas americana* has 98 genes. (Adapted from M.W. Gray, G. Burger and B.F. Lang, *Science* 283:1476–1481, 1999.)

allows one tRNA to pair with any one of four codons and permits protein synthesis with fewer tRNA molecules.

3. *Variant genetic code.* Perhaps most surprising, comparisons of mitochondrial gene sequences and the amino acid sequences of the corresponding proteins indicate that the genetic code is different: 4 of the 64 codons have different “meanings” from those of the same codons in other genomes (Table 14–3).

The close similarity of the genetic code in all organisms provides strong evidence that they all have evolved from a common ancestor. How, then, do we explain the differences in the genetic code in many mitochondria? A hint comes from the finding that the mitochondrial genetic code in different organisms is not the same. In the mitochondrion with the largest number of genes in Figure 14–60, that of the protozoan *Reclinomonas*, the genetic code is unchanged from the standard genetic code of the cell nucleus. Yet UGA, which is a stop codon elsewhere, is read as tryptophan in the mitochondria of mammals, fungi, and invertebrates. Similarly, the codon AGG normally codes for arginine, but it codes for *stop* in the mitochondria of mammals and codes for serine in the mitochondria of *Drosophila* (see Table 14–3). Such variation suggests that a random drift can occur in the genetic code in mitochondria. Presumably, the unusually small number of proteins encoded by the mitochondrial genome makes an occasional change in the meaning of a rare codon tolerable, whereas such a change in a larger genome would alter the function of many proteins and thereby destroy the cell.

Interestingly, in many species, one or two tRNAs for mitochondrial protein synthesis are encoded in the nucleus. Some parasites, for example trypanosomes, have not retained any tRNA genes in their mitochondrial DNA. Instead, the required tRNAs are all produced in the cytosol and are thought to be imported into the mitochondrion by special tRNA translocases that are distinct from the mitochondrial protein import system.

**TABLE 14–3 Some Differences Between the “Universal” Code and Mitochondrial Genetic Codes\***

Codon	“Universal” code	Mitochondrial codes			
		Mammals	Invertebrates	Yeast	Plants
UGA	STOP	<i>Trp</i>	<i>Trp</i>	<i>Trp</i>	STOP
AUA	Ile	<i>Met</i>	<i>Met</i>	<i>Met</i>	Ile
CUA	Leu	Leu	Leu	<i>Thr</i>	Leu
AGA AGG	Arg	<i>STOP</i>	<i>Ser</i>	Arg	Arg

\*Red italics indicate that the code differs from the “Universal” code.

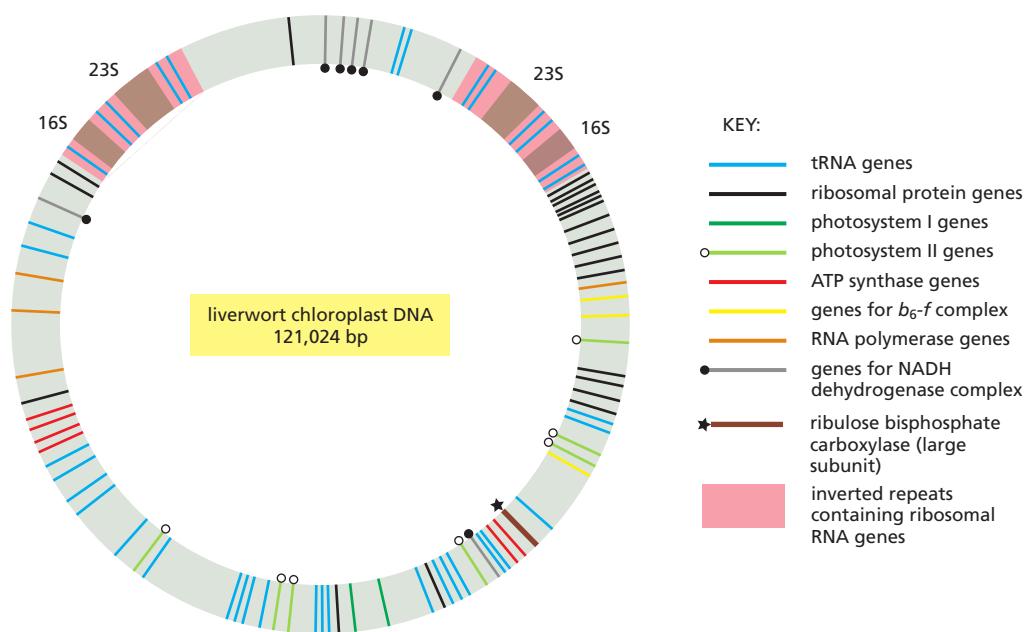
## Chloroplasts and Bacteria Share Many Striking Similarities

The chloroplast genomes of land plants range in size from 70,000 to 200,000 nucleotide pairs. More than 300 chloroplast genomes have now been sequenced. Many are surprisingly similar, even in distantly related plants (such as tobacco and liverwort), and even those of green algae are closely related (Figure 14–67). Chloroplast genes are involved in three main processes: transcription, translation, and photosynthesis. Plant chloroplast genomes typically encode 80–90 proteins and around 45 RNAs, including 37 or more tRNAs. As in mitochondria, most of the organelle-encoded proteins are part of larger protein complexes that also contain one or more subunits encoded in the nucleus and imported from the cytosol.

The genomes of chloroplasts and bacteria have striking similarities. Basic regulatory sequences, such as transcription promoters and terminators, are virtually identical. The amino acid sequences of the proteins encoded in chloroplasts are clearly recognizable as bacterial, and several clusters of genes with related functions (such as those encoding ribosomal proteins) are organized in the same way in the genomes of chloroplasts, the bacterium *E. coli*, and cyanobacteria.

The mechanisms by which chloroplasts and bacteria divide are also similar. Both utilize *FtsZ* proteins, which are self-assembling GTPases related to tubulins (see Chapter 16). Bacterial *FtsZ* is a soluble protein that assembles into a dynamic ring of membrane-attached protofilaments beneath the plasma membrane in the middle of the dividing cell. The *FtsZ* ring acts as a scaffold for recruitment of other cell-division proteins and generates a contractile force that results in membrane constriction and eventually in cell division. Presumably, chloroplasts divide in very much the same way. Although both employ membrane-interacting GTPases, the mechanisms by which mitochondria and chloroplasts divide are fundamentally different. The machinery for chloroplast division acts from the inside, as in bacteria, while the dynamin-like GTPases divide mitochondria from the outside (see Figure 14–63). The chloroplasts have remained closer to their bacterial origins than have mitochondria, since the eukaryotic mechanisms of membrane constriction and vesicle formation have been adapted for mitochondrial fission.

The RNA editing and RNA processing that is prevalent in chloroplasts owes everything to their eukaryotic hosts. This RNA processing includes the generation of transcript 5' and 3' termini and the cleavage of polycistronic transcripts. In addition, an RNA editing process converts specific C residues to U and can change the amino acid specified by the edited codon. These and other RNA-based



**Figure 14–67** The organization of the liverwort chloroplast genome. The chloroplast genome organization is similar in all higher plants, although the size varies from species to species—depending on how much of the DNA surrounding the genes encoding the chloroplast's 16S and 23S ribosomal RNAs is present in two copies.

processes are catalyzed by protein families that are not found in prokaryotes. One can ask why the expression of so few chloroplast genes needs to be so complex. One explanation is that the expression of chloroplast and nuclear genes must be closely coordinated. More generally, the bacterial concept of the operon as a co-regulated set of genes in a single transcription unit has been largely abandoned in chloroplasts. Polycistronic transcripts are cleaved into smaller fragments, which then require splicing or RNA editing to become functional.

### Organelle Genes Are Maternally Inherited in Animals and Plants

In *Saccharomyces cerevisiae* (baker's yeast), when two haploid cells mate, they are equal in size and contribute equal amounts of mitochondrial DNA to the diploid zygote. Mitochondrial inheritance in yeasts is therefore *biparental*: both parents contribute equally to the mitochondrial gene pool of the progeny. However, during the course of the subsequent asexual, vegetative growth, the mitochondria become distributed more or less randomly to daughter cells. After a few generations, the mitochondria of any given cell contain only the DNA from one or the other parent cell, because only a small sample of the mitochondrial DNA passes from the mother cell to the bud of the daughter cell. This process is known as *mitotic segregation*, and it gives rise to a distinct form of inheritance that is called non-Mendelian, or *cytoplasmic inheritance*, in contrast to the Mendelian inheritance of nuclear genes.

The inheritance of mitochondria in animals and plants is quite different. In these organisms, the egg cell contributes much more cytoplasm to the zygote than does the male gamete (sperm in animals, pollen in plants). For example, a typical human oocyte contains about 100,000 copies of maternal mitochondrial DNA, whereas a sperm cell contains only a few. In addition, an active process ensures that the sperm mitochondria do not compete with those in the egg. As sperm mature, the DNA is degraded in their mitochondria. Sperm mitochondria are also specifically recognized then eliminated from the fertilized egg cell by autophagy in very much the same way that damaged mitochondria are removed (by ubiquitylation followed by delivery to lysosomes, as discussed in Chapter 13). Because of these two processes, the mitochondrial inheritance in both animals and plants is *uniparental*. More precisely, the mitochondrial DNA passes from one generation to the next by **maternal inheritance**.

In about two-thirds of higher plants, the chloroplast precursors from the male parent (contained in pollen grains) fail to enter the zygote, so that chloroplast as well as mitochondrial DNA is maternally inherited. In other plants, the chloroplast precursors from the pollen grains enter the zygote, making chloroplast inheritance biparental. In such plants, defective chloroplasts are a cause of variegation: a mixture of normal and defective chloroplasts in a zygote may sort out by mitotic segregation during plant growth and development, thereby producing alternating green and white patches in leaves. Leaf cells in the green patches contain normal chloroplasts, while those in the white patches contain defective chloroplasts (**Figure 14–68**).

### Mutations in Mitochondrial DNA Can Cause Severe Inherited Diseases

In humans, as we have explained, all the mitochondrial DNA in a fertilized egg cell is inherited from the mother. Some mothers carry a mixed population of both mutant and normal mitochondrial genomes. Their daughters and sons will inherit this mixture of normal and mutant mitochondrial DNAs and be healthy unless the process of mitotic segregation results in a majority of defective mitochondria in a particular tissue. Muscle and the nervous system are most at risk. Because they need particularly large amounts of ATP, muscle and nerve cells are particularly dependent on fully functional mitochondria.

Numerous diseases in humans are caused by mutations in mitochondrial DNA. These diseases are recognized by their passage from affected mothers to



**Figure 14–68** A variegated leaf. In the white patches, the plant cells have inherited a defective chloroplast. (Courtesy of John Innes Foundation.)

both their daughters and their sons, with the daughters but not the sons producing children with the disease. As expected from the random nature of mitotic segregation, the symptoms of these diseases vary greatly between different family members—including not only the severity and age of onset, but also which tissue is affected. There are also mitochondrial diseases that are caused by mutations in nuclear-encoded mitochondrial proteins; these diseases are inherited in the regular, Mendelian fashion.

### The Accumulation of Mitochondrial DNA Mutations Is a Contributor to Aging

Mitochondria are marvels of efficiency in energy conversion, and they supply the cells of our body with a readily available source of energy in the form of ATP. But in highly developed, long-lived animals such as ourselves, the cells in our body age and eventually die. A factor in this inevitable process is the accumulation of deletions and point mutations in mitochondrial DNA. Oxidative damage to the cell by *reactive oxygen species* (ROS) such as H<sub>2</sub>O<sub>2</sub>, superoxide, or hydroxyl radicals also increases with age. The mitochondrial respiratory chain is the main source of ROS in animal cells, and animals in which mitochondrial superoxide dismutase—the main ROS scavenger—has been knocked out, die prematurely.

The less complex DNA replication and repair systems in mitochondria mean that accidents are corrected less efficiently. This results in a 100-fold higher occurrence of deletions and point mutations than in nuclear DNA. Mathematical modeling suggests that most of these mutations and lesions are acquired in childhood or early adult life, and then proliferate by *clonal expansion* in later life. Due to mitotic segregation, some cells will accumulate higher levels of faulty mitochondrial DNA than others. Above some threshold, serious deficiencies in respiratory-chain function will develop, producing cells that are *senescent*. In many organs of the human body, senescent cells with high levels of mitochondrial DNA damage are intermingled with normal cells, resulting in a mosaic of cells with and without respiratory-chain deficiency.

The main role of mitochondrial fusion in cellular physiology is most likely to ensure an even distribution of mitochondrial DNA throughout the mitochondrial reticulum, and to prevent the accumulation of damaged DNA in one part of the network. When the fusion machinery is defective, DNA is lost from a subset of the mitochondria in the cell. Loss of mitochondrial DNA leads to a loss of respiratory-chain function, and it can cause disease.

All of the considerations just discussed have suggested to some scientists that changes in our mitochondria are major contributors to human aging. However, there are many other processes that tend to go wrong as cells and tissues age, as one might expect given the incredible complexity of human cell biology. Despite intensive research, the issue remains unresolved.

### Why Do Mitochondria and Chloroplasts Maintain a Costly Separate System for DNA Transcription and Translation?

Why do mitochondria and chloroplasts require their own separate genetic systems, when other organelles that share the same cytoplasm, such as peroxisomes and lysosomes, do not? The question is not trivial, because maintaining a separate genetic system is costly: more than 90 proteins—including many ribosomal proteins, aminoacyl-tRNA synthetases, DNA polymerase, RNA polymerase, and RNA-processing and RNA-modifying enzymes—must be encoded by nuclear genes specifically for this purpose. Moreover, as we have seen, the mitochondrial genetic system entails the risk of aging and disease.

A possible reason for maintaining this costly and potentially hazardous arrangement is the highly hydrophobic nature of the nonribosomal proteins encoded by organelle genes. This may make their production in and import from the cytoplasm simply too difficult and energy-consuming. It is also possible that the evolution (and eventual elimination) of the organellar genetic systems is still

ongoing, but for now there is no alternative for the cell than to maintain separate genetic systems for its nuclear, mitochondrial, and chloroplast genes.

## Summary

*Mitochondria are organelles that allow eukaryotes to carry out oxidative phosphorylation, while chloroplasts are organelles that allow plants to carry out photosynthesis. Presumably as a result of their prokaryotic origins, each organelle maintains and reproduces itself in a highly coordinated process that requires the contribution of two separate genetic systems—one in the organelle and the other in the cell nucleus. The vast majority of the proteins in these organelles are encoded by nuclear DNA, synthesized in the cytosol, and then imported individually into the organelle. Other organelle proteins, as well as organelle ribosomal and transfer RNAs, are encoded by the organelle DNA; these are synthesized in the organelle itself.*

*The ribosomes of chloroplasts closely resemble bacterial ribosomes, while the origin of mitochondrial ribosomes is more difficult to trace. Extensive protein similarities, however, suggest that both organelles originated when a primitive eukaryotic cell entered into a stable endosymbiotic relationship with a bacterium. Although some of the genes of these former bacteria still function to make organelle proteins and RNA, most of them have been transferred into the nuclear genome, where they encode bacteria-like enzymes that are synthesized on cytosolic ribosomes and then imported into the organelle. The mitochondrial DNA replication and DNA repair processes are substantially less effective than the corresponding processes in the cell nucleus. Damage therefore accumulates in the genome of mitochondria over time; this damage may be a substantial contributor to the aging of cells and organisms, and it can cause serious diseases.*

## WHAT WE DON'T KNOW

- What structures are needed to form the barriers that separate and maintain the differentiated membrane domains in a single continuous membrane—as for the cristae and inner boundary membrane in mitochondria?
- How does a eukaryotic cell regulate the many functions of mitochondria, including ATP production?
- What are the origins and evolutionary history of photosynthetic complexes? Are there undiscovered types of photosynthesis present on Earth to help answer this question?
- Why is the mutation rate so much higher in mitochondria than in the nucleus (and chloroplasts)? Could this high rate have been useful to the cell?
- What mechanisms and pathways have been used during evolution to transfer genes from the mitochondrion to the nucleus?

## PROBLEMS

Which statements are true? Explain why or why not.

**14–1** The three respiratory enzyme complexes in the mitochondrial inner membrane tend to associate with each other in ways that facilitate the correct transfer of electrons between appropriate complexes.

**14–2** The number of *c* subunits in the rotor ring of ATP synthase defines how many protons need to pass through the turbine to make each molecule of ATP.

**14–3** Mutations that are inherited according to Mendelian rules affect nuclear genes; mutations whose inheritance violates Mendelian rules are likely to affect organelle genes.

Discuss the following problems.

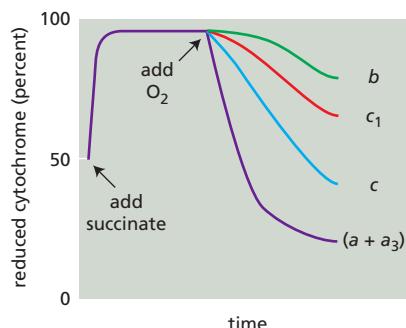
**14–4** In the 1860s, Louis Pasteur noticed that when he added O<sub>2</sub> to a culture of yeast growing anaerobically on glucose, the rate of glucose consumption declined dramatically. Explain the basis for this result, which is known as the Pasteur effect.

**14–5** Heart muscle gets most of the ATP needed to power its continual contractions through oxidative phosphorylation. When oxidizing glucose to CO<sub>2</sub>, heart muscle consumes O<sub>2</sub> at a rate of 10 μmol/min per g of tissue, in order to replace the ATP used in contraction and give a steady-state ATP concentration of 5 μmol/g of tissue. At this rate,

how many seconds would it take the heart to consume an amount of ATP equal to its steady-state levels? (Complete oxidation of one molecule of glucose to CO<sub>2</sub> yields 30 ATP, 26 of which are derived by oxidative phosphorylation using the 12 pairs of electrons captured in the electron carriers NADH and FADH<sub>2</sub>.)

**14–6** Both H<sup>+</sup> and Ca<sup>2+</sup> are ions that move through the cytosol. Why is the movement of H<sup>+</sup> ions so much faster than that of Ca<sup>2+</sup> ions? How do you suppose the speed of these two ions would be affected by freezing the solution? Would you expect them to move faster or slower? Explain your answer.

**14–7** If isolated mitochondria are incubated with a source of electrons such as succinate, but without oxygen, electrons enter the respiratory chain, reducing each of the electron carriers almost completely. When oxygen is then introduced, the carriers become oxidized at different rates (**Figure Q14–1**). How does this result allow you to order



**Figure Q14–1** Rapid spectrophotometric analysis of the rates of oxidation of electron carriers in the respiratory chain (Problem 14–7). Cytochromes *a* and *a*<sub>3</sub> cannot be distinguished and thus are listed as cytochrome (*a + a*<sub>3</sub>).

the electron carriers in the respiratory chain? What is their order?

**14–8** Normally, the flow of electrons to  $O_2$  is tightly linked to the production of ATP via the electrochemical gradient. If ATP synthase is inhibited, for example, electrons do not flow down the electron-transport chain and respiration ceases. Since the 1940s, several substances—such as 2,4-dinitrophenol—have been known to uncouple electron flow from ATP synthesis. Dinitrophenol was once prescribed as a diet drug to aid in weight loss. How would an uncoupler of oxidative phosphorylation promote weight loss? Why do you suppose dinitrophenol is no longer prescribed?

**14–9** In actively respiring liver mitochondria, the pH in the matrix is about half a pH unit higher than it is in the cytosol. Assuming that the cytosol is at pH 7 and the matrix is a sphere with a diameter of 1  $\mu m$  [ $V = (4/3)\pi r^3$ ], calculate the total number of protons in the matrix of a respiration liver mitochondrion. If the matrix began at pH 7 (equal to that in the cytosol), how many protons would have to be pumped out to establish a matrix pH of 7.5 (a difference of 0.5 pH units)?

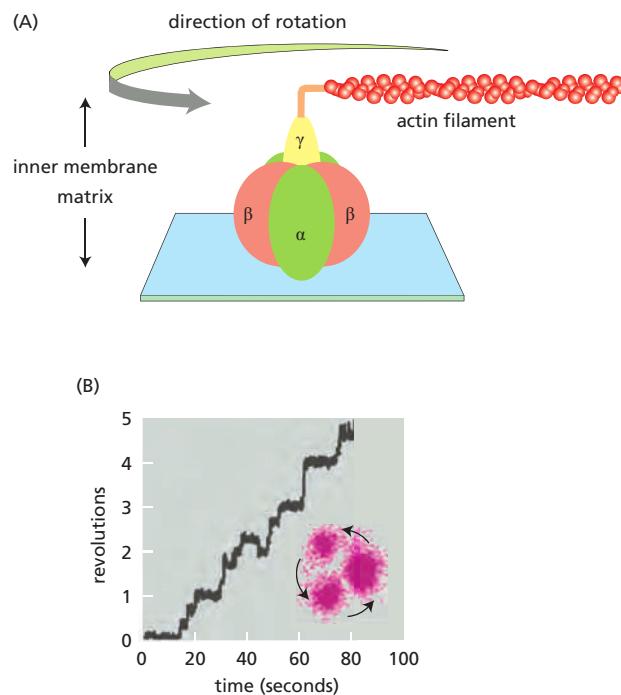
**14–10** ATP synthase is the world's smallest rotary motor. Passage of  $H^+$  ions through the membrane-embedded portion of ATP synthase (the  $F_o$  component) causes rotation of the single, central, axle-like  $\gamma$  subunit inside the head group. The tripartite head is composed of the three  $\alpha\beta$  dimers, the  $\beta$  subunit of which is responsible for synthesis of ATP. The rotation of the  $\gamma$  subunit induces conformational changes in the  $\alpha\beta$  dimers that allow ADP and Pi to be converted into ATP. A variety of indirect evidence had suggested rotary catalysis by ATP synthase, but seeing is believing.

To demonstrate rotary motion, a modified form of the  $\alpha_3\beta_3\gamma$  complex was used. The  $\beta$  subunits were modified so they could be firmly anchored to a solid support and the  $\gamma$  subunit was modified (on the end that normally inserts into the  $F_o$  component in the inner membrane) so that a fluorescently tagged, readily visible filament of actin could be attached (Figure Q14–2A). This arrangement allows rotations of the  $\gamma$  subunit to be visualized as revolutions of the long actin filament. In these experiments, ATP synthase was studied in the reverse of its normal mechanism by allowing it to hydrolyze ATP. At low ATP concentrations, the actin filament was observed to revolve in steps of 120° and then pause for variable lengths of time, as shown in Figure Q14–2B.

**A.** Why does the actin filament revolve in steps with pauses in between? What does this rotation correspond to in terms of the structure of the  $\alpha_3\beta_3\gamma$  complex?

**B.** In its normal mode of operation inside the cell, how many ATP molecules do you suppose would be synthesized for each complete 360° rotation of the  $\gamma$  subunit? Explain your answer.

**14–11** How much energy is available in visible light? How much energy does sunlight deliver to Earth? How efficient



**Figure Q14–2** Experimental set-up for observing rotation of the  $\gamma$  subunit of ATP synthase (Problem 14–10). (A) The immobilized  $\alpha_3\beta_3\gamma$  complex. The  $\beta$  subunits are anchored to a solid support and a fluorescent actin filament is attached to the  $\gamma$  subunit. (B) Stepwise revolution of the actin filament. The indicated trace is a typical example from one experiment. The inset shows the positions in the revolution at which the actin filament pauses. (B, from R. Yasuda et al., *Cell* 93:1117–1124, 1998. With permission from Elsevier.)

are plants at converting light energy into chemical energy? The answers to these questions provide an important backdrop to the subject of photosynthesis.

Each quantum or photon of light has energy  $h\nu$ , where  $h$  is Planck's constant ( $6.6 \times 10^{-37}$  kJ sec/photon) and  $\nu$  is the frequency in  $sec^{-1}$ . The frequency of light is equal to  $c/\lambda$ , where  $c$  is the speed of light ( $3.0 \times 10^{17}$  nm/sec) and  $\lambda$  is the wavelength in nm. Thus, the energy ( $E$ ) of a photon is

$$E = h\nu = hc/\lambda$$

**A.** Calculate the energy of a mole of photons ( $6 \times 10^{23}$  photons/mole) at 400 nm (violet light), at 680 nm (red light), and at 800 nm (near-infrared light).

**B.** Bright sunlight strikes Earth at the rate of about 1.3 kJ/sec per square meter. Assuming for the sake of calculation that sunlight consists of monochromatic light of wavelength 680 nm, how many seconds would it take for a mole of photons to strike a square meter?

**C.** Assuming that it takes eight photons to fix one molecule of  $CO_2$  as carbohydrate under optimal conditions (8–10 photons is the currently accepted value), calculate how long it would take a tomato plant with a leaf area of 1 square meter to make a mole of glucose from  $CO_2$ . Assume that photons strike the leaf at the rate calculated above and, furthermore, that all the photons are absorbed and used to fix  $CO_2$ .

**D.** If it takes 468 kJ/mole to fix a mole of  $\text{CO}_2$  into carbohydrate, what is the efficiency of conversion of light energy into chemical energy after photon capture? Assume again that eight photons of red light (680 nm) are required to fix one molecule of  $\text{CO}_2$ .

**14–12** In chloroplasts, protons are pumped out of the stroma across the thylakoid membrane, whereas in mitochondria, they are pumped out of the matrix across the crista membrane. Explain how this arrangement allows chloroplasts to generate a larger proton gradient across the thylakoid membrane than mitochondria can generate across the inner membrane.

**14–13** Examine the variegated leaf shown in **Figure Q14–3**. Yellow patches surrounded by green are common, but there are no green patches surrounded by yellow. Propose an explanation for this phenomenon.



**Figure Q14–3** A variegated leaf of *Aucuba japonica* with green and yellow patches (Problem 14–13).

## REFERENCES

### General

- Cramer WA & Knaff DB (1990) Energy Transduction in Biological Membranes: A Textbook of Bioenergetics. New York: Springer-Verlag.  
 Mathews CK, van Holde KE & Ahern K-G (2012) Biochemistry, 4th ed. San Francisco: Benjamin Cummings.  
 Nicholls DG & Ferguson SJ (2013) Bioenergetics, 4th ed. New York: Academic Press.  
 Schatz G (2012) The fires of life. *Annu. Rev. of Biochem.* 81, 34–59.

### The Mitochondrion

- Ernster L & Schatz G (1981) Mitochondria: a historical review. *J. Cell Biol.* 91, 227s–255s.  
 Friedman JR & Nunnari J (2014) Mitochondrial form and function. *Nature* 505, 335–43.  
 Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* 191, 144–148.  
 Pebay-Peyroula E & Brandolin G (2004) Nucleotide exchange in mitochondria: insight at a molecular level. *Curr. Opin. Struct. Biol.* 14, 420–425.  
 Scheffler IE (1999) Mitochondria. New York/Chichester: Wiley-Liss.

### The Proton Pumps of the Electron-Transport Chain

- Althoff T, Mills DJ, Popot J-L & Kühlbrandt W (2011) Arrangement of electron transport chain components in bovine mitochondrial supercomplex I<sub>1</sub>I<sub>1</sub>I<sub>2</sub>V<sub>1</sub>. *EMBO J.* 30, 4652–4664.  
 Baradaran R, Berrisford JM, Minhas GS & Sazanov LA (2013) Crystal structure of the entire respiratory complex I. *Nature* 494, 443–448.  
 Beinert H, Holm RH & Münck E (1997) Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* 277, 653–659.  
 Berry EA, Guergova-Kuras M, Huang LS & Crofts AR (2000) Structure and function of cytochrome bc complexes. *Annu. Rev. Biochem.* 69, 1005–1075.  
 Brandt U (2006) Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* 75, 69–92.  
 Chance B & Williams GR (1955) A method for the localization of sites for oxidative phosphorylation. *Nature* 176, 250–254.

- Cooley JW (2013) Protein conformational changes involved in the cytochrome bc<sub>1</sub> complex catalytic cycle. *Biochim. Biophys. Acta.* 1827, 1340–45.  
 Gottschalk G (1997) Bacterial Metabolism, 2nd ed. New York: Springer.  
 Gray HB & Winkler JR (1996) Electron transfer in proteins. *Annu. Rev. Biochem.* 65, 537–561.  
 Hirst J (2013) Mitochondrial complex I. *Ann. Rev. Biochem.* 82, 551–75.  
 Hosler JP, Ferguson-Miller S & Mills DA (2006) Energy transduction: proton transfer through the respiratory complexes. *Annu. Rev. Biochem.* 75, 165–187.  
 Hunte C, Zickermann V & Brandt U (2010) Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* 329, 448–451.  
 Keilin D (1966) The History of Cell Respiration and Cytochromes. Cambridge, UK: Cambridge University Press.  
 Rouault TA, Tracey A & Tong WH (2008) Iron-sulfur cluster biogenesis and human disease. *Trends Genet.* 24, 398–407.  
 Trumppower BL (2002) A concerted, alternating sites mechanism of ubiquinol oxidation by the dimeric cytochrome bc<sub>1</sub> complex. *Biochim. Biophys. Acta.* 1555, 166–173.  
 Tsukihara T, Aoyama H, Yamashita E et al. (1996) The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272, 1136–1144.

### ATP Production in Mitochondria

- Abrahams JP, Leslie AG, Lutter R & Walker JE (1994) Structure at 2.8 Å resolution of F<sub>1</sub>-ATPase from bovine heart mitochondria. *Nature* 370, 621–628.  
 Berg HC (2003) The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* 72, 19–54.  
 Boyer PD (1997) The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* 66, 717–749.  
 Meier T, Polzer P, Diederichs K et al. (2005) Structure of the rotor ring of F-type Na<sup>+</sup>-ATPase from *Illyobacter tartaricus*. *Science* 308, 659–662.  
 Stock D, Gibbons C, Arechaga I et al. (2000) The rotary mechanism of ATP synthase. *Curr. Opin. Struct. Biol.* 10, 672–679.

von Ballmoos C, Wiedenmann A & Dimroth P (2009) Essentials for ATP synthesis by F<sub>1</sub>F<sub>0</sub> ATP synthases. *Annu. Rev. Biochem.* 78, 649–672.

### Chloroplasts and Photosynthesis

- Barber J (2013) Photosystem II: the water-splitting enzyme of photosynthesis. *Cold Spring Harbor Symp. Quant. Biol.* 77, 295–307.
- Bassham JA (1962) The path of carbon in photosynthesis. *Sci. Am.* 206, 89–100.
- Blankenship RE (2002) Molecular Mechanisms of Photosynthesis. Oxford, UK: Blackwell Scientific.
- Blankenship RE & Bauer CE (eds) (1995) Anoxygenic Photosynthetic Bacteria. Dordrecht: Kluwer.
- Deisenhofer J & Michel H (1989) Nobel lecture. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *EMBO J.* 8, 2149–2170.
- De Las Rivas J, Balsera M & Barber J (2004) Evolution of oxygenic photosynthesis: genome-wide analysis of the OEC extrinsic proteins. *Trends Plant Sci.* 9:18–25.
- Hohmann-Marriott MF & Blankenship RE (2011) Evolution of photosynthesis. *Annu. Rev. Plant Biol.* 62, 515–548.
- Jordan P, Fromme P, Witt HT et al. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411, 909–917.
- Kühlbrandt W, Wang DN & Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367, 614–621.
- Lane N & Martin WF (2012) The origin of membrane bioenergetics. *Cell* 151, 1406–1416.
- Lyons TW, Reinhard CT & Planavsky NJ (2014) The rise of oxygen in earth's early ocean and atmosphere. *Nature* 506, 307–15.
- Nelson N & Ben-Shem A (2004) The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* 5, 971–982.
- Orgel LE (1998) The origin of life—a review of facts and speculations. *Trends Biochem. Sci.* 23, 491–495.
- Tang K-H, Tang YJ & Blankenship RE (2011) Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front. Microbiol.* 2, 165.
- Umena Y, Kawakami K, Shen J-R & Kamiya N (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* 473, 55–60.
- Vinyard DJ, Ananyev GM & Dismukes GC (2013) Photosystem II: the reaction center of oxygenic photosynthesis. *Annu. Rev. Biochem.* 82, 577–606.
- Yano J, Kern J, Sauer K et al. (2006) Where water is oxidized to dioxygen: structure of the photosynthetic Mn<sub>4</sub>Ca cluster. *Science* 314, 821–25.
- Yoon HS (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809–18.

### The Genetic Systems of Mitochondria and Chloroplasts

- Anderson S, Bankier AT, Barrell BG et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465.
- Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16, 1661–1666.
- Birky CW Jr (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl Acad. Sci. USA* 92, 11331–11338.
- Bullerwell CE & Gray MW (2004) Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Curr. Opin. Microbiol.* 7, 528–534.
- Chen XJ & Butow RA (2005) The organization and inheritance of the mitochondrial genome. *Nat. Rev. Genet.* 6, 815–825.
- Clayton DA (2000) Vertebrate mitochondrial DNA—a circle of surprises. *Exp. Cell Res.* 255, 4–9.
- Daley DO & Whelan J (2005) Why genes persist in organelle genomes. *Genome Biol.* 6, 110.
- de Duve C (2007) The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* 8, 395–403.
- Dyall SD, Brown MT & Johnson PJ (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304, 253–257.
- Falkenberg M, Larsson NG & Gustafsson CM (2007) DNA replication and transcription in mammalian mitochondria. *Annu. Rev. Biochem.* 76, 679–699.
- Harel A, Bromberg Y, Falkowski PG & Bhattacharya D (2014) Evolutionary history of redox metal-binding domains across the tree of life. *Proc. Natl Acad. Sci. USA* 111, 7042–47.
- Hoppins S, Lackner L & Nunnari J (2007) The machines that divide and fuse mitochondria. *Annu. Rev. Biochem.* 76, 751–780.
- Larsson NG (2010) Somatic mitochondrial DNA mutations in mammalian aging. *Annu. Rev. Biochem.* 79, 683–706.
- Ma H, Xu H & O'Farrell PH (2014) Transmission of mitochondrial mutations and action of purifying selection in *Drosophila melanogaster*. *Nat. Genet.* 46, 393–97.
- Neupert W & Herrmann JM (2007) Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* 76, 723–749.
- Rawi A, Louvet-Vallee SS, Djeddi D et al. (2011) Postfertilization autophagy of sperm organelles prevents paternal mitochondrial DNA transmission. *Science* 334, 1144–47.
- Taylor RW, Barron MJ, Borthwick GM et al. (2003) Mitochondrial DNA mutations in human colonic crypt stem cells. *J. Clin. Invest.* 112, 1351–60.
- Wallace DC (1999) Mitochondrial diseases in man and mouse. *Science* 283, 1482–1488.
- Williams TA, Foster PG, Cox CJ & Embley TM (2014) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–36.

# Cell Signaling

CHAPTER  
**15**

When things change, cells respond. Every cell, from the humble bacterium to the most sophisticated eukaryotic cell, monitors its intracellular and extracellular environment, processes the information it gathers, and responds accordingly. Unicellular organisms, for example, modify their behavior in response to changes in environmental nutrients or toxins. The cells of multicellular organisms detect and respond to countless internal and extracellular signals that control their growth, division, and differentiation during development, as well as their behavior in adult tissues. At the heart of all these communication systems are regulatory proteins that produce chemical signals, which are sent from one place to another in the body or within a cell, usually being processed along the way and integrated with other signals to provide clear and effective communication.

The study of cell signaling has traditionally focused on the mechanisms by which eukaryotic cells communicate with each other using *extracellular signal molecules* such as hormones and growth factors. In this chapter, we describe the features of some of these cell-cell communication systems, and we use them to illustrate the general principles by which any regulatory system, inside or outside the cell, is able to generate, process, and respond to signals. Our main focus is on animal cells, but we end by considering the special features of cell signaling in plants.

## PRINCIPLES OF CELL SIGNALING

Long before multicellular creatures roamed the Earth, unicellular organisms had developed mechanisms for responding to physical and chemical changes in their environment. These almost certainly included mechanisms for responding to the presence of other cells. Evidence comes from studies of present-day unicellular organisms such as bacteria and yeasts. Although these cells lead mostly independent lives, they can communicate and influence one another's behavior. Many bacteria, for example, respond to chemical signals that are secreted by their neighbors and accumulate at higher population density. This process, called *quorum sensing*, allows bacteria to coordinate their behavior, including their motility, antibiotic production, spore formation, and sexual conjugation. Similarly, yeast cells communicate with one another in preparation for mating. The budding yeast *Saccharomyces cerevisiae* provides a well-studied example: when a haploid individual is ready to mate, it secretes a peptide *mating factor* that signals cells of the opposite mating type to stop proliferating and prepare to mate. The subsequent fusion of two haploid cells of opposite mating type produces a diploid zygote.

Intercellular communication achieved an astonishing level of complexity during the evolution of multicellular organisms. These organisms are tight-knit societies of cells, in which the well-being of the individual cell is often set aside for the benefit of the organism as a whole. Complex systems of intercellular communication have evolved to allow the collaboration and coordination of different tissues and cell types. Bewildering arrays of signaling systems govern every conceivable feature of cell and tissue function during development and in the adult.

Communication between cells in multicellular organisms is mediated mainly by **extracellular signal molecules**. Some of these operate over long distances,

### IN THIS CHAPTER

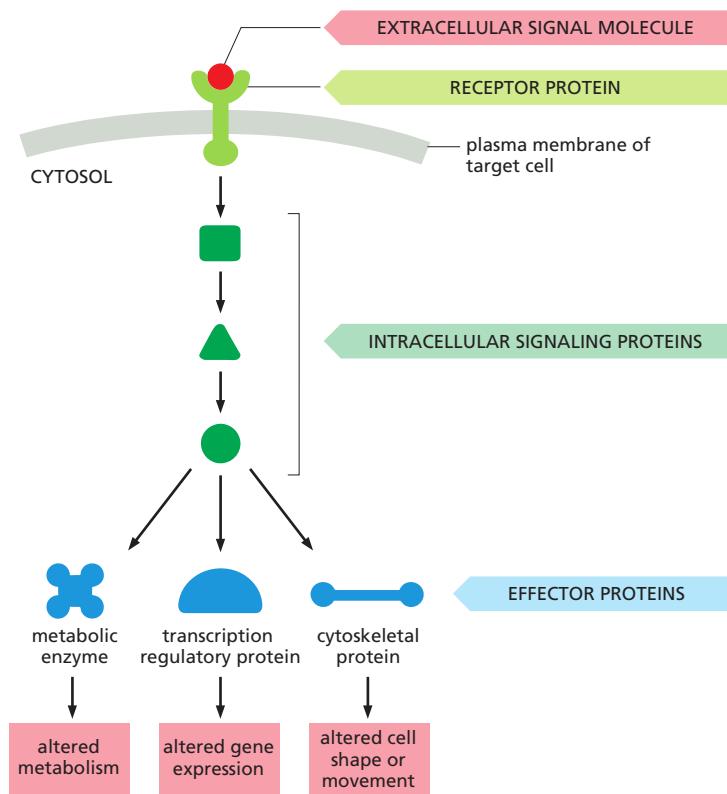
[PRINCIPLES OF CELL SIGNALING](#)

[SIGNALING THROUGH G-PROTEIN-COUPLED RECEPTORS](#)

[SIGNALING THROUGH ENZYME-COUPLED RECEPTORS](#)

[ALTERNATIVE SIGNALING ROUTES IN GENE REGULATION](#)

[SIGNALING IN PLANTS](#)



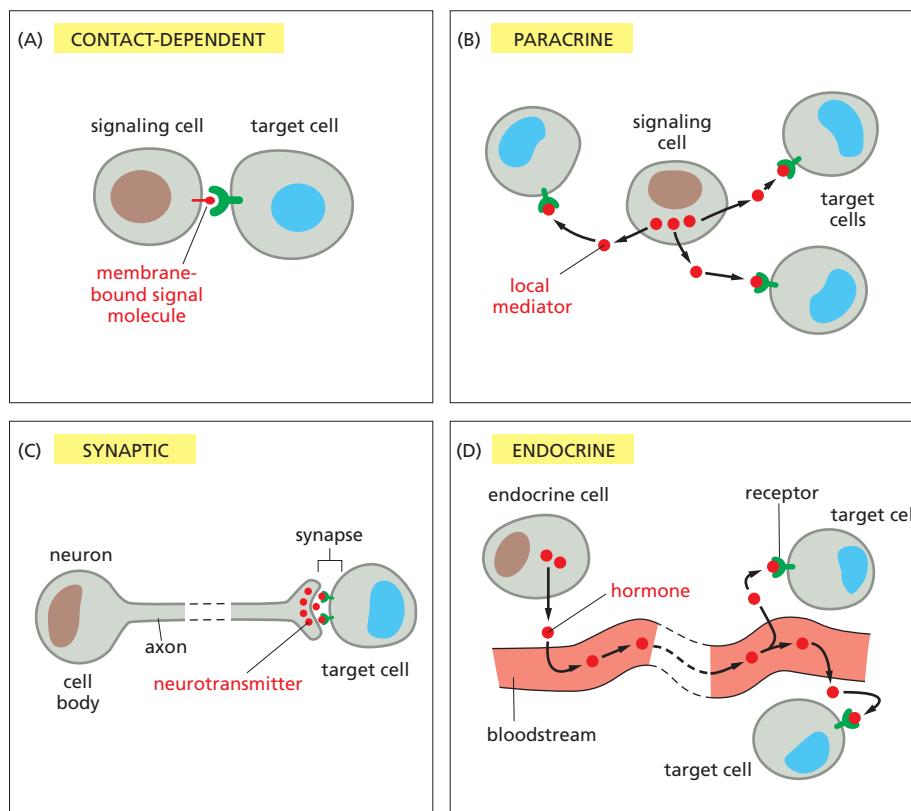
**Figure 15–1** A simple intracellular signaling pathway activated by an extracellular signal molecule. The signal molecule usually binds to a receptor protein that is embedded in the plasma membrane of the target cell. The receptor activates one or more intracellular signaling pathways, involving a series of signaling proteins. Finally, one or more of the intracellular signaling proteins alters the activity of effector proteins and thereby the behavior of the cell.

signaling to cells far away; others signal only to immediate neighbors. Most cells in multicellular organisms both emit and receive signals. Reception of the signals depends on *receptor proteins*, usually (but not always) at the cell surface, which bind the signal molecule. The binding activates the receptor, which in turn activates one or more *intracellular signaling pathways* or *systems*. These systems depend on *intracellular signaling proteins*, which process the signal inside the receiving cell and distribute it to the appropriate intracellular targets. The targets that lie at the end of signaling pathways are generally called *effector proteins*, which are altered in some way by the incoming signal and implement the appropriate change in cell behavior. Depending on the signal and the type and state of the receiving cell, these effectors can be transcription regulators, ion channels, components of a metabolic pathway, or parts of the cytoskeleton (Figure 15–1).

The fundamental features of cell signaling have been conserved throughout the evolution of the eukaryotes. In budding yeast, for example, the response to mating factor depends on cell-surface receptor proteins, intracellular GTP-binding proteins, and protein kinases that are clearly related to functionally similar proteins in animal cells. Through gene duplication and divergence, however, the signaling systems in animals have become much more elaborate than those in yeasts; the human genome, for example, contains more than 1500 genes that encode receptor proteins, and the number of different receptor proteins is further increased by alternative RNA splicing and post-translational modifications.

### Extracellular Signals Can Act Over Short or Long Distances

Many extracellular signal molecules remain bound to the surface of the signaling cell and influence only cells that contact it (Figure 15–2A). Such **contact-dependent signaling** is especially important during development and in immune responses. Contact-dependent signaling during development can sometimes operate over relatively large distances if the communicating cells extend long thin processes to make contact with one another.



**Figure 15-2 Four forms of intercellular signaling.** (A) Contact-dependent signaling requires cells to be in direct membrane–membrane contact. (B) Paracrine signaling depends on local mediators that are released into the extracellular space and act on neighboring cells. (C) Synaptic signaling is performed by neurons that transmit signals electrically along their axons and release neurotransmitters at synapses, which are often located far away from the neuronal cell body. (D) Endocrine signaling depends on endocrine cells, which secrete hormones into the bloodstream for distribution throughout the body. Many of the same types of signaling molecules are used in paracrine, synaptic, and endocrine signaling; the crucial differences lie in the speed and selectivity with which the signals are delivered to their targets.

In most cases, however, signaling cells secrete signal molecules into the extracellular fluid. Often, the secreted molecules are **local mediators**, which act only on cells in the local environment of the signaling cell. This is called **paracrine signaling** (Figure 15-2B). Usually, the signaling and target cells in paracrine signaling are of different cell types, but cells may also produce signals that they themselves respond to: this is referred to as *autocrine signaling*. Cancer cells, for example, often produce extracellular signals that stimulate their own survival and proliferation.

Large multicellular organisms like us need long-range signaling mechanisms to coordinate the behavior of cells in remote parts of the body. Thus, they have evolved cell types specialized for intercellular communication over large distances. The most sophisticated of these are nerve cells, or neurons, which typically extend long, branching processes (axons) that enable them to contact target cells far away, where the processes terminate at the specialized sites of signal transmission known as *chemical synapses*. When a neuron is activated by stimuli from other nerve cells, it sends electrical impulses (action potentials) rapidly along its axon; when the impulse reaches the synapse at the end of the axon, it triggers secretion of a chemical signal that acts as a **neurotransmitter**. The tightly organized structure of the synapse ensures that the neurotransmitter is delivered specifically to receptors on the postsynaptic target cell (Figure 15-2C). The details of this **synaptic signaling** process are discussed in Chapter 11.

A quite different strategy for signaling over long distances makes use of **endocrine cells**, which secrete their signal molecules, called **hormones**, into the bloodstream. The blood carries the molecules far and wide, allowing them to act on target cells that may lie anywhere in the body (Figure 15-2D).

### Extracellular Signal Molecules Bind to Specific Receptors

Cells in multicellular animals communicate by means of hundreds of kinds of extracellular signal molecules. These include proteins, small peptides, amino

acids, nucleotides, steroids, retinoids, fatty acid derivatives, and even dissolved gases such as nitric oxide and carbon monoxide. Most of these signal molecules are released into the extracellular space by exocytosis from the signaling cell, as discussed in Chapter 13. Some, however, are emitted by diffusion through the signaling cell's plasma membrane, whereas others are displayed on the external surface of the cell and remain attached to it, providing a signal to other cells only when they make contact. Transmembrane signal proteins may operate in this way, or their extracellular domains may be released from the signaling cell's surface by proteolytic cleavage and then act at a distance.

Regardless of the nature of the signal, the *target cell* responds by means of a **receptor**, which binds the signal molecule and then initiates a response in the target cell. The binding site of the receptor has a complex structure that is shaped to recognize the signal molecule with high specificity, helping to ensure that the receptor responds only to the appropriate signal and not to the many other signaling molecules surrounding the cell. Many signal molecules act at very low concentrations (typically  $\leq 10^{-8}$  M), and their receptors usually bind them with high affinity (dissociation constant  $K_d \leq 10^{-8}$  M; see Figure 3–44).

In most cases, receptors are transmembrane proteins on the target-cell surface. When these proteins bind an extracellular signal molecule (a *ligand*), they become activated and generate various intracellular signals that alter the behavior of the cell. In other cases, the receptor proteins are inside the target cell, and the signal molecule has to enter the cell to bind to them: this requires that the signal molecule be sufficiently small and hydrophobic to diffuse across the target cell's plasma membrane (Figure 15–3). This chapter focuses primarily on signaling through cell-surface receptors, but we will briefly describe signaling through intracellular receptors later in the chapter.

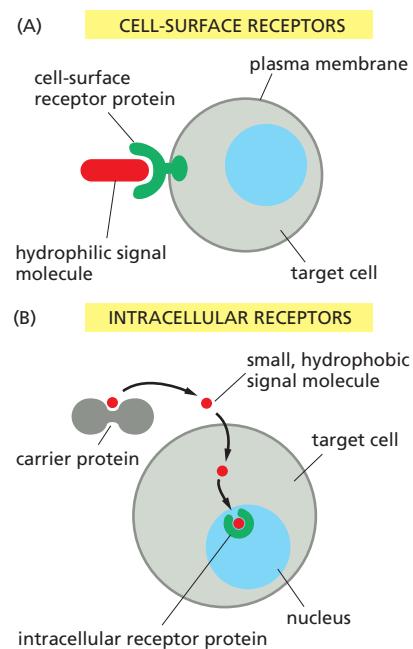
### Each Cell Is Programmed to Respond to Specific Combinations of Extracellular Signals

A typical cell in a multicellular organism is exposed to hundreds of different signal molecules in its environment. The molecules can be soluble, bound to the extracellular matrix, or bound to the surface of a neighboring cell; they can be stimulatory or inhibitory; they can act in innumerable different combinations; and they can influence almost any aspect of cell behavior. The cell responds to this blizzard of signals selectively, in large part by expressing only those receptors and intracellular signaling systems that respond to the signals that are required for the regulation of that cell.

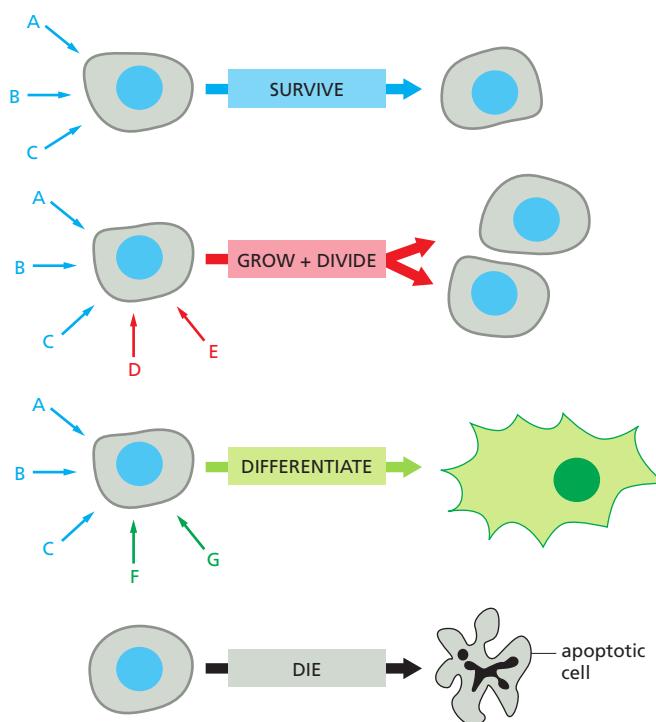
Most cells respond to many different signals in the environment, and some of these signals may influence the response to other signals. One of the key challenges in cell biology is to determine how a cell integrates all of this signaling information in order to make decisions—to divide, to move, to differentiate, and so on. Many cells, for example, require a specific combination of extracellular survival factors to allow the cell to continue living; when deprived of these signals, the cell activates a suicide program and kills itself—usually by *apoptosis*, a form of *programmed cell death*, as discussed in Chapter 18. Cell proliferation often depends on a combination of signals that promote both cell division and survival, as well as signals that stimulate cell growth (Figure 15–4). On the other hand, differentiation into a nondividing state (called *terminal differentiation*) frequently requires a different combination of survival and differentiation signals that must override any signal to divide.

In principle, the hundreds of signal molecules that an animal makes can be used in an almost unlimited number of combinations to control the diverse behaviors of its cells in highly specific ways. Relatively small numbers of types of signal molecules and receptors are sufficient. The complexity lies in the ways in which cells respond to the combinations of signals that they receive.

A signal molecule often has different effects on different types of target cells. The neurotransmitter acetylcholine (Figure 15–5A), for example, decreases the

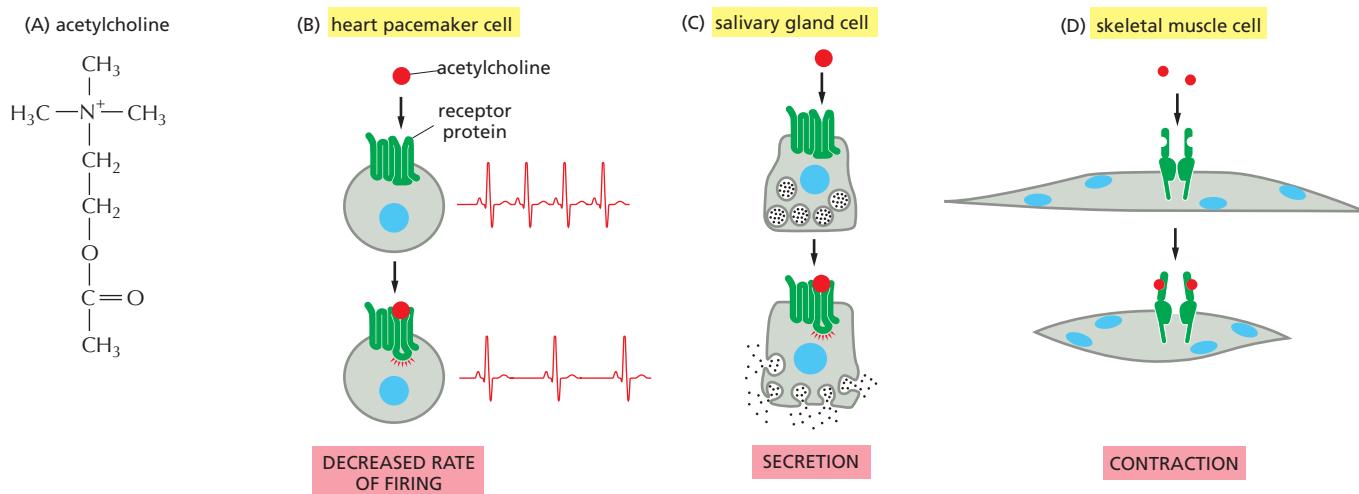


**Figure 15–3** The binding of extracellular signal molecules to either cell-surface or intracellular receptors. (A) Most signal molecules are hydrophilic and are therefore unable to cross the target cell's plasma membrane directly; instead, they bind to cell-surface receptors, which in turn generate signals inside the target cell (see Figure 15–1). (B) Some small signal molecules, by contrast, diffuse across the plasma membrane and bind to receptor proteins inside the target cell—either in the cytosol or in the nucleus (as shown here). Many of these small signal molecules are hydrophobic and poorly soluble in aqueous solutions; they are therefore transported in the bloodstream and other extracellular fluids bound to carrier proteins, from which they dissociate before entering the target cell.



**Figure 15–4** An animal cell’s dependence on multiple extracellular signal molecules. Each cell type displays a set of receptors that enables it to respond to a corresponding set of signal molecules produced by other cells. These signal molecules work in various combinations to regulate the behavior of the cell. As shown here, an individual cell often requires multiple signals to survive (blue arrows) and additional signals to grow and divide (red arrows) or differentiate (green arrows). If deprived of appropriate survival signals, a cell will undergo a form of cell suicide known as apoptosis. The actual situation is even more complex. Although not shown, some extracellular signal molecules act to inhibit these and other cell behaviors, or even to induce apoptosis.

rate of action potential firing in heart pacemaker cells (Figure 15–5B) and stimulates the production of saliva by salivary gland cells (Figure 15–5C), even though the receptors are the same on both cell types. In skeletal muscle, acetylcholine causes the cells to contract by binding to a different receptor protein (Figure 15–5D). The different effects of acetylcholine in these cell types result from differences in the intracellular signaling proteins, effector proteins, and genes that are activated. Thus, an extracellular signal itself has little information content; it simply induces the cell to respond according to its predetermined state, which depends on the cell’s developmental history and the specific genes it expresses.



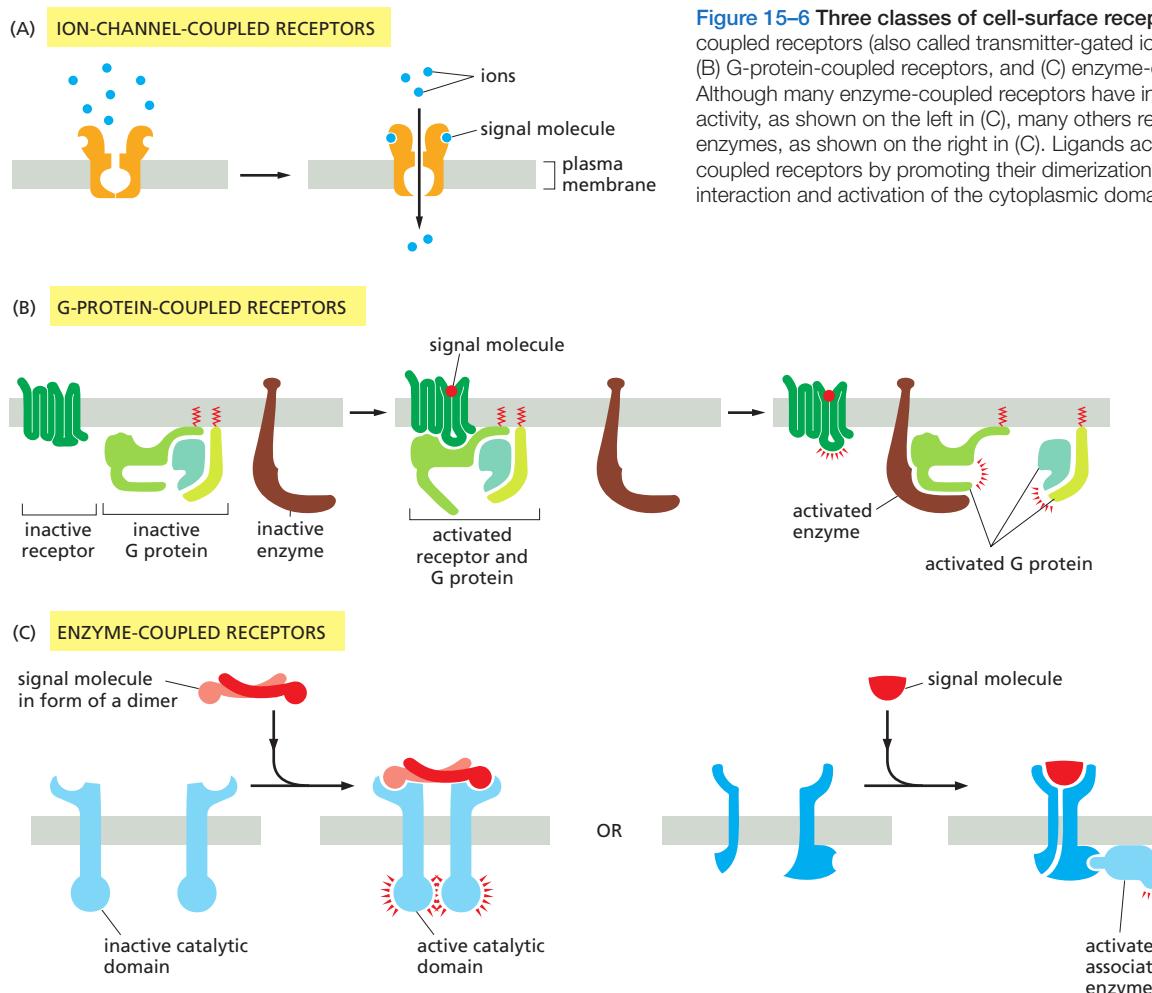
**Figure 15–5** Various responses induced by the neurotransmitter acetylcholine. (A) The chemical structure of acetylcholine. (B–D) Different cell types are specialized to respond to acetylcholine in different ways. In some cases (B and C), acetylcholine binds to similar receptor proteins (G-protein-coupled receptors; see Figure 15–6), but the intracellular signals produced are interpreted differently in cells specialized for different functions. In other cases (D), the receptor protein is also different (here, an ion-channel-coupled receptor; see Figure 15–6).

## There Are Three Major Classes of Cell-Surface Receptor Proteins

Most extracellular signal molecules bind to specific receptor proteins on the surface of the target cells they influence and do not enter the cytosol or nucleus. These cell-surface receptors act as *signal transducers* by converting an extracellular ligand-binding event into intracellular signals that alter the behavior of the target cell.

Most cell-surface receptor proteins belong to one of three classes, defined by their transduction mechanism. **Ion-channel-coupled receptors**, also known as *transmitter-gated ion channels* or *ionotropic receptors*, are involved in rapid synaptic signaling between nerve cells and other electrically excitable target cells such as nerve and muscle cells (Figure 15–6A). This type of signaling is mediated by a small number of neurotransmitters that transiently open or close an ion channel formed by the protein to which they bind, briefly changing the ion permeability of the plasma membrane and thereby changing the excitability of the postsynaptic target cell. Most ion-channel-coupled receptors belong to a large family of homologous, multipass transmembrane proteins. Because they are discussed in detail in Chapter 11, we will not consider them further here.

G-protein-coupled receptors act by indirectly regulating the activity of a separate plasma-membrane-bound target protein, which is generally either an enzyme or an ion channel. A *trimeric GTP-binding protein* (*G protein*) mediates the interaction between the activated receptor and this target protein (Figure 15–6B). The activation of the target protein can change the concentration of one or



**Figure 15–6** Three classes of cell-surface receptors. (A) Ion-channel-coupled receptors (also called transmitter-gated ion channels), (B) G-protein-coupled receptors, and (C) enzyme-coupled receptors. Although many enzyme-coupled receptors have intrinsic enzymatic activity, as shown on the left in (C), many others rely on associated enzymes, as shown on the right in (C). Ligands activate most enzyme-coupled receptors by promoting their dimerization, which results in the interaction and activation of the cytoplasmic domains.

more small intracellular signaling molecules (if the target protein is an enzyme), or it can change the ion permeability of the plasma membrane (if the target protein is an ion channel). The small intracellular signaling molecules act in turn to alter the behavior of yet other signaling proteins in the cell.

Enzyme-coupled receptors either function as enzymes or associate directly with enzymes that they activate (Figure 15–6C). They are usually single-pass transmembrane proteins that have their ligand-binding site outside the cell and their catalytic or enzyme-binding site inside. Enzyme-coupled receptors are heterogeneous in structure compared with the other two classes; the great majority, however, are either protein kinases or associate with protein kinases, which phosphorylate specific sets of proteins in the target cell when activated.

There are also some types of cell-surface receptors that do not fit easily into any of these classes but have important functions in controlling the specialization of different cell types during development and in tissue renewal and repair in adults. We discuss these in a later section, after we explain how G-protein-coupled receptors and enzyme-coupled receptors operate. First, we continue our general discussion of the principles of signaling via cell-surface receptors.

### Cell-Surface Receptors Relay Signals Via Intracellular Signaling Molecules

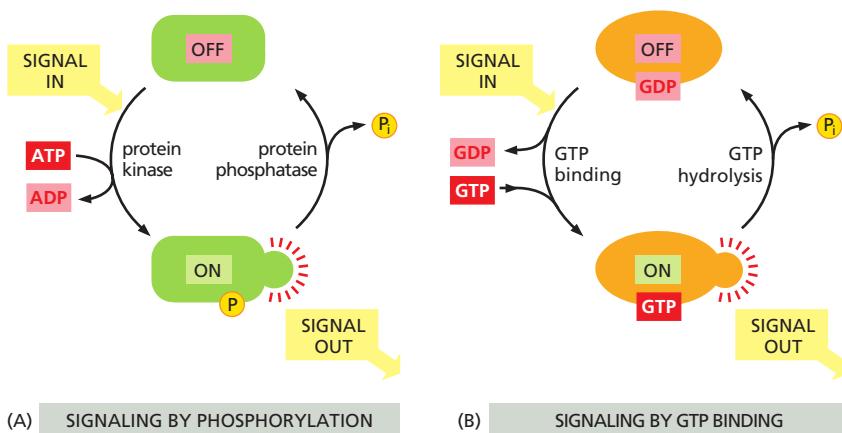
Numerous intracellular signaling molecules relay signals received by cell-surface receptors into the cell interior. The resulting chain of intracellular signaling events ultimately alters effector proteins that are responsible for modifying the behavior of the cell (see Figure 15–1).

Some intracellular signaling molecules are small chemicals, which are often called **second messengers** (the “first messengers” being the extracellular signals). They are generated in large amounts in response to receptor activation and diffuse away from their source, spreading the signal to other parts of the cell. Some, such as *cyclic AMP* and  $\text{Ca}^{2+}$ , are water-soluble and diffuse in the cytosol, while others, such as *diacylglycerol*, are lipid-soluble and diffuse in the plane of the plasma membrane. In either case, they pass the signal on by binding to and altering the behavior of selected signaling or effector proteins.

Most intracellular signaling molecules are proteins, which help relay the signal into the cell by either generating second messengers or activating the next signaling or effector protein in the pathway. Many of these proteins behave like *molecular switches*. When they receive a signal, they switch from an inactive to an active state, until another process switches them off, returning them to their inactive state. The switching off is just as important as the switching on. If a signaling pathway is to recover after transmitting a signal so that it can be ready to transmit another, every activated molecule in the pathway must return to its original, unactivated state.

The largest class of molecular switches consists of proteins that are activated or inactivated by **phosphorylation** (discussed in Chapter 3). For these proteins, the switch is thrown in one direction by a **protein kinase**, which covalently adds one or more phosphate groups to specific amino acids on the signaling protein, and in the other direction by a **protein phosphatase**, which removes the phosphate groups (Figure 15–7A). The activity of any protein regulated by phosphorylation depends on the balance between the activities of the kinases that phosphorylate it and of the phosphatases that dephosphorylate it. About 30–50% of human proteins contain covalently attached phosphate, and the human genome encodes about 520 protein kinases and about 150 protein phosphatases. A typical mammalian cell makes use of hundreds of distinct types of protein kinases at any moment.

Protein kinases attach phosphate to the hydroxyl group of specific amino acids on the target protein. There are two main types of protein kinase. The great majority are **serine/threonine kinases**, which phosphorylate the hydroxyl groups of serines and threonines in their targets. Others are **tyrosine kinases**, which



**Figure 15-7** Two types of intracellular signaling proteins that act as molecular switches. (A) A protein kinase covalently adds a phosphate from ATP to the signaling protein, and a protein phosphatase removes the phosphate. Although not shown, many signaling proteins are activated by dephosphorylation rather than by phosphorylation. (B) A GTP-binding protein is induced to exchange its bound GDP for GTP, which activates the protein; the protein then inactivates itself by hydrolyzing its bound GTP to GDP.

phosphorylate proteins on tyrosines. The two types of protein kinase are closely related members of a large family, differing primarily in the structure of their protein substrate binding sites.

Many intracellular signaling proteins controlled by phosphorylation are themselves protein kinases, and these are often organized into **kinase cascades**. In such a cascade, one protein kinase, activated by phosphorylation, phosphorylates the next protein kinase in the sequence, and so on, relaying the signal onward and, in some cases, amplifying it or spreading it to other signaling pathways.

The other important class of molecular switches consists of **GTP-binding proteins** (discussed in Chapter 3). These proteins switch between an “on” (actively signaling) state when GTP is bound and an “off” state when GDP is bound. In the “on” state, they usually have intrinsic GTPase activity and shut themselves off by hydrolyzing their bound GTP to GDP (Figure 15-7B). There are two major types of GTP-binding proteins. Large, *trimeric GTP-binding proteins* (also called *G proteins*) help relay signals from G-protein-coupled receptors that activate them (see Figure 15-6B). Small **monomeric GTPases** (also called *monomeric GTP-binding proteins*) help relay signals from many classes of cell-surface receptors.

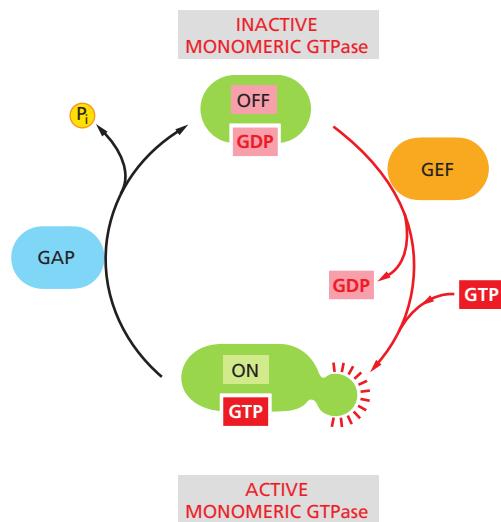
Specific regulatory proteins control both types of GTP-binding proteins. **GTPase-activating proteins (GAPs)** drive the proteins into an “off” state by increasing the rate of hydrolysis of bound GTP. Conversely, **guanine nucleotide exchange factors (GEFs)** activate GTP-binding proteins by promoting the release of bound GDP, which allows a new GTP to bind. In the case of trimeric G proteins, the activated receptor serves as the GEF. Figure 15-8 illustrates the regulation of monomeric GTPases.

Not all molecular switches in signaling systems depend on phosphorylation or GTP binding. We see later that some signaling proteins are switched on or off by the binding of another signaling protein or a second messenger such as cyclic AMP or  $\text{Ca}^{2+}$ , or by covalent modifications other than phosphorylation or dephosphorylation, such as ubiquitylation (discussed in Chapter 3).

For simplicity, we often portray a signaling pathway as a series of activation steps (see Figure 15-1). It is important to note, however, that most signaling pathways contain inhibitory steps, and a sequence of two inhibitory steps can have the same effect as one activating step (Figure 15-9). This *double-negative* activation is very common in signaling systems, as we will see when we describe specific pathways later in this chapter.

### Intracellular Signals Must Be Specific and Precise in a Noisy Cytoplasm

Ideally, an activated intracellular signaling molecule should interact only with the appropriate downstream targets, and, likewise, the targets should only be activated by the appropriate upstream signal. In reality, however, intracellular

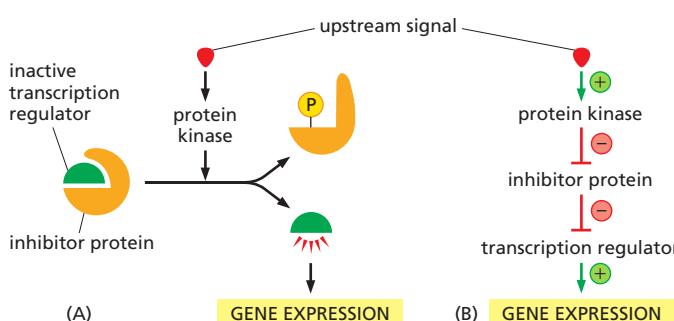


**Figure 15–8 The regulation of a monomeric GTPase.** GTPase-activating proteins (GAPs) inactivate the protein by stimulating it to hydrolyze its bound GTP to GDP, which remains tightly bound to the inactivated GTPase. Guanine nucleotide exchange factors (GEFs) activate the inactive protein by stimulating it to release its GDP; because the concentration of GTP in the cytosol is 10 times greater than the concentration of GDP, the protein rapidly binds GTP and is thereby activated.

signaling molecules share the cytoplasm with a crowd of closely related signaling molecules that control a diverse array of cellular processes. It is inevitable that an occasional signaling molecule will bind or modify the wrong partner, potentially creating unwanted cross-talk and interference between signaling systems. How does a signal remain strong, precise, and specific under these noisy conditions?

The first line of defense comes from the high affinity and specificity of the interactions between signaling molecules and their correct partners compared to the relatively low affinity of the interactions between inappropriate partners. The binding of a signaling molecule to the correct target is determined by precise and complex interactions between complementary surfaces on the two molecules. Protein kinases, for example, contain active sites that recognize a specific amino acid sequence around the phosphorylation site on the correct target protein, and they often contain additional *docking sites* that promote a specific, high-affinity interaction with the target. These and related mechanisms help provide a strong and persistent interaction between the correct partners, reducing the likelihood of inappropriate interactions with other proteins.

Another important way that cells avoid responses to unwanted background signals depends on the ability of many downstream target proteins to simply ignore such signals. These proteins respond only when the upstream signal reaches a high concentration or activity level. Consider a signaling pathway in which a protein kinase activates some downstream target protein by phosphorylation. If a response is triggered only when more than half of the target proteins are phosphorylated, then there will be little harm done if a small number of them are occasionally phosphorylated by some inappropriate protein kinase. Furthermore, constitutively active protein phosphatases will further reduce the impact of background phosphorylation by rapidly removing much of it. In these and other ways, intracellular signaling systems filter out noise, generating little or no response to low levels of stimuli.



**Figure 15–9 A sequence of two inhibitory signals produces a positive signal.** (A) In this simple signaling system, a transcription regulator is kept in an inactive state by a bound inhibitor protein. In response to some upstream signal, a protein kinase is activated and phosphorylates the inhibitor, causing its dissociation from the transcription regulator and thereby activating gene expression. (B) This signaling pathway can be diagrammed as a sequence of four steps, including two sequential inhibitory steps that are equivalent to a single activating step.

Cells in a population often exhibit random variation in the concentration or activity of their intracellular signaling molecules. Similarly, individual molecules in a large population of molecules vary in their activity or interactions with other molecules. This *signal variability* introduces another form of noise that can interfere with the precision and efficiency of signaling. Most signaling systems, however, are built to generate remarkably robust and precise responses even when upstream signals are variable or even when some components of the system are disabled. In many cases, this *robustness* depends on the presence of backup mechanisms: for example, a signal might employ two parallel pathways to activate a single common downstream target protein, allowing the response to occur even if one pathway is crippled.

### Intracellular Signaling Complexes Form at Activated Receptors

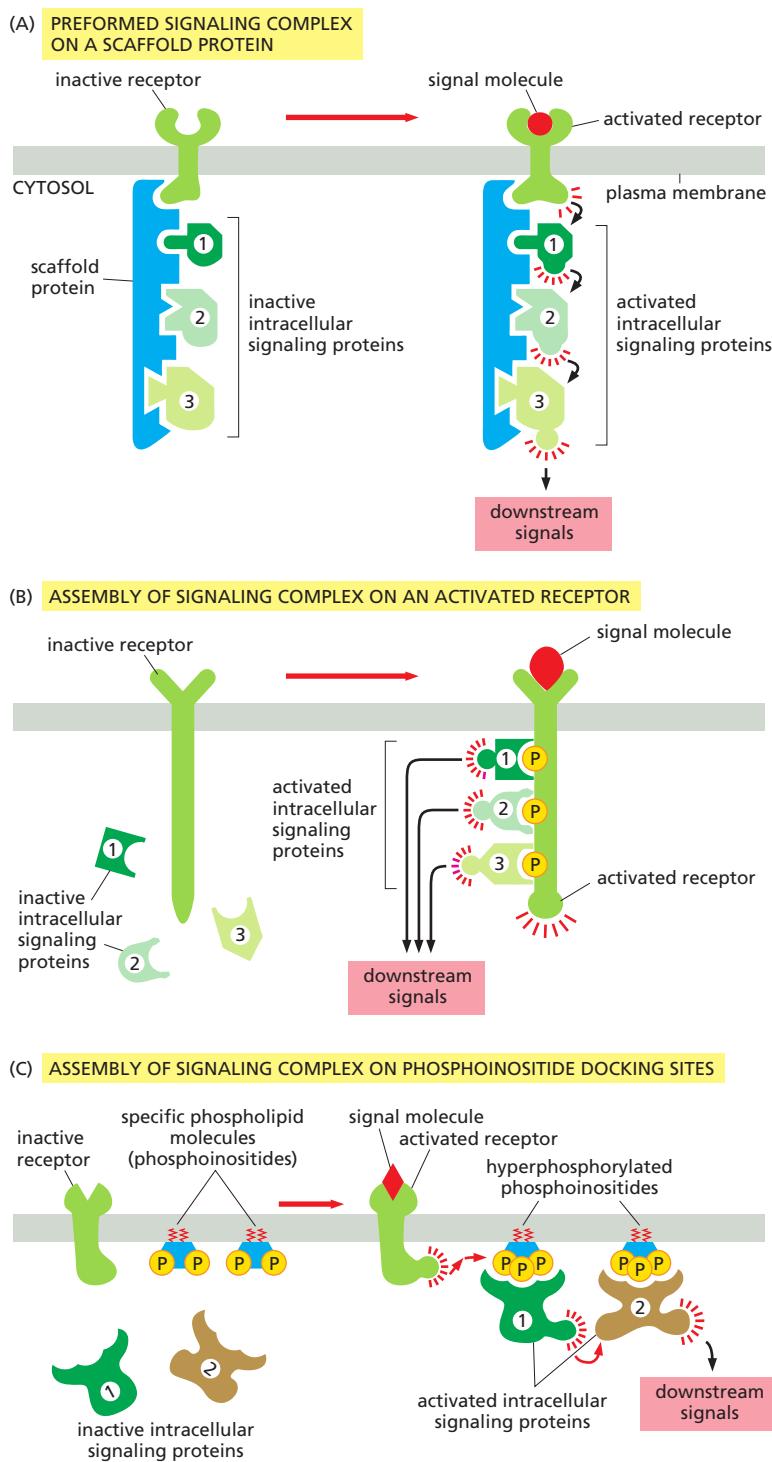
One simple and effective strategy for enhancing the specificity of interactions between signaling molecules is to localize them in the same part of the cell or even within large protein complexes, thereby ensuring that they interact only with each other and not with inappropriate partners. Such mechanisms often involve **scaffold proteins**, which bring together groups of interacting signaling proteins into *signaling complexes*, often before a signal has been received (Figure 15–10A). Because the scaffold holds the proteins in close proximity, they can interact at high local concentrations and be sequentially activated rapidly, efficiently, and selectively in response to an appropriate extracellular signal, avoiding unwanted cross-talk with other signaling pathways.

In other cases, such signaling complexes form only transiently in response to an extracellular signal and rapidly disassemble when the signal is gone. They often assemble around a receptor after an extracellular signal molecule has activated it. In many of these cases, the cytoplasmic tail of the activated receptor is phosphorylated during the activation process, and the phosphorylated amino acids then serve as docking sites for the assembly of other signaling proteins (Figure 15–10B). In yet other cases, receptor activation leads to the production of modified phospholipid molecules (called phosphoinositides) in the adjacent plasma membrane, which then recruit specific intracellular signaling proteins to this region of membrane, where they are activated (Figure 15–10C).

### Modular Interaction Domains Mediate Interactions Between Intracellular Signaling Proteins

Simply bringing intracellular signaling proteins together into close proximity is sometimes sufficient to activate them. Thus, *induced proximity*, where a signal triggers assembly of a signaling complex, is commonly used to relay signals from protein to protein along a signaling pathway. The assembly of such signaling complexes depends on various highly conserved, small **interaction domains**, which are found in many intracellular signaling proteins. Each of these compact protein modules binds to a particular structural motif in another protein or lipid. The recognized motif in the interacting protein can be a short peptide sequence, a covalent modification (such as a phosphorylated amino acid), or another protein domain. The use of modular interaction domains presumably facilitated the evolution of new signaling pathways; because it can be inserted at many locations in a protein without disturbing the protein's folding or function, a new interaction domain added to an existing signaling protein could connect the protein to additional signaling pathways.

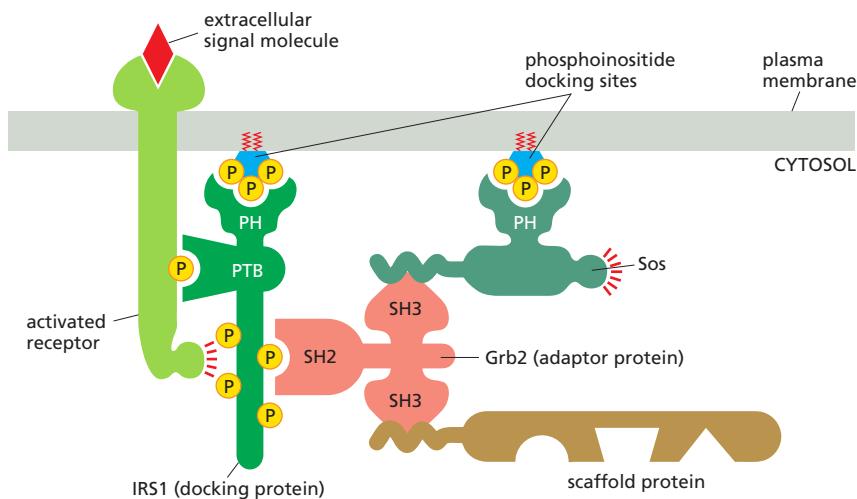
There are many types of interaction domains in signaling proteins. *Src homology 2 (SH2) domains* and *phosphotyrosine-binding (PTB) domains*, for example, bind to phosphorylated tyrosines in a particular peptide sequence on activated receptors or intracellular signaling proteins. *Src homology 3 (SH3) domains* bind to short, proline-rich amino acid sequences. Some *pleckstrin homology (PH)* domains bind to the charged head groups of specific phosphoinositides that are produced in the plasma membrane in response to an extracellular signal; they



**Figure 15–10** Three types of intracellular signaling complexes. (A) A receptor and some of the intracellular signaling proteins it activates in sequence are preassembled into a signaling complex on the inactive receptor by a large scaffold protein. (B) A signaling complex assembles transiently on a receptor only after the binding of an extracellular signal molecule has activated the receptor; here, the activated receptor phosphorylates itself at multiple sites, which then act as docking sites for intracellular signaling proteins. (C) Activation of a receptor leads to the increased phosphorylation of specific phospholipids (phosphoinositides) in the adjacent plasma membrane; these then serve as docking sites for specific intracellular signaling proteins, which can now interact with each other.

enable the protein they are part of to dock on the membrane and interact with other similarly recruited signaling proteins (see Figure 15–10C). Some signaling proteins consist solely of two or more interaction domains and function only as **adaptors** to link two other proteins together in a signaling pathway.

Interaction domains enable signaling proteins to bind to one another in multiple specific combinations. Like Lego® bricks, the proteins can form linear or branching chains or three-dimensional networks, which determine the route followed by the signaling pathway. As an example, Figure 15–11 illustrates how some interaction domains mediate the formation of a large signaling complex around the receptor for the hormone *insulin*.



**Figure 15–11 A specific signaling complex formed using modular interaction domains.** This example is based on the insulin receptor, which is an enzyme-coupled receptor (a receptor tyrosine kinase, discussed later). First, the activated receptor phosphorylates itself on tyrosines, and one of the phosphotyrosines then recruits a docking protein called insulin receptor substrate-1 (IRS1) via a PTB domain of IRS1; the PH domain of IRS1 also binds to specific phosphoinositides on the inner surface of the plasma membrane. Then, the activated receptor phosphorylates IRS1 on tyrosines, and one of these phosphotyrosines binds the SH2 domain of the adaptor protein Grb2. Next, Grb2 uses one of its two SH3 domains to bind to a proline-rich region of a protein called Sos, which relays the signal downstream by acting as a GEF (see Figure 15–8) to activate a monomeric GTPase called Ras (not shown). Sos also binds to phosphoinositides in the plasma membrane via its PH domain. Grb2 uses its other SH2 domain to bind to a proline-rich sequence in a scaffold protein. The scaffold protein binds several other signaling proteins, and the other phosphorylated tyrosines on IRS1 recruit additional signaling proteins that have SH2 domains (not shown).

Another way of bringing receptors and intracellular signaling proteins together is to concentrate them in a specific region of the cell. An important example is the **primary cilium** that projects like an antenna from the surface of most vertebrate cells (discussed in Chapter 16). It is usually short and nonmotile and has microtubules in its core, and a number of surface receptors and signaling proteins are concentrated there. We shall see later that light and smell receptors are also highly concentrated in specialized cilia.

### The Relationship Between Signal and Response Varies in Different Signaling Pathways

The function of an intracellular signaling system is to detect and measure a specific stimulus in one location of a cell and then generate an appropriately timed and measured response at another location. The system accomplishes this task by sending information in the form of molecular “signals” from the sensor to the target, often through a series of intermediaries that do not simply pass the signal along but process it in various ways. All signaling systems do not work in precisely the same way: each has evolved specialized behaviors that produce a response that is appropriate for the cell function that system controls. In the following paragraphs, we list some of these behaviors and describe how they vary in different systems, as a foundation for more detailed discussions later.

1. *Response timing* varies dramatically in different signaling systems, according to the speed required for the response. In some cases, such as synaptic signaling (see Figure 15–2C), the response can occur within milliseconds. In other cases, as in the control of cell fate by morphogens during development, a full response can require hours or days.
2. *Sensitivity* to extracellular signals can vary greatly. Hormones tend to act at very low concentrations on their distant target cells, which are therefore highly sensitive to low concentrations of signal. Neurotransmitters, on the other hand, operate at much higher concentrations at a synapse, reducing the need for high sensitivity in postsynaptic receptors. Sensitivity is often controlled by changes in the number or affinity of the receptors on the target cell. A particularly important mechanism for increasing the sensitivity of a signaling system is *signal amplification*, whereby a small number of activated cell-surface receptors evoke a large intracellular response either by producing large amounts of a second messenger or by activating many copies of a downstream signaling protein.
3. *Dynamic range* of a signaling system is related to its sensitivity. Some systems, like those involved in simple developmental decisions, are responsive

over a narrow range of extracellular signal concentrations. Other systems, like those controlling vision or the metabolic response to some hormones, are highly responsive over a much broader range of signal strengths. We will see that broad dynamic range is often achieved by *adaptation* mechanisms that adjust the responsiveness of the system according to the prevailing amount of signal.

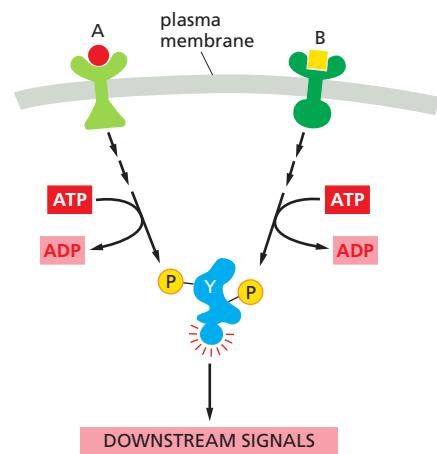
4. *Persistence* of a response can vary greatly. A transient response of less than a second is appropriate in some synaptic responses, for example, while a prolonged or even permanent response is required in cell fate decisions during development. Numerous mechanisms, including positive feedback, can be used to alter the duration and reversibility of a response.
5. *Signal processing* can convert a simple signal into a complex response. In many systems, for example, a gradual increase in an extracellular signal is converted into an abrupt, switchlike response. In other cases, a simple input signal is converted into an oscillatory response, produced by a repeating series of transient intracellular signals. Feedback usually lies at the heart of biochemical switches and oscillators, as we describe later.
6. *Integration* allows a response to be governed by multiple inputs. As discussed earlier, for example, specific combinations of extracellular signals are generally required to stimulate complex cell behaviors such as cell survival and proliferation (see Figure 15–4). The cell therefore has to integrate information coming from multiple signals, which often depends on intracellular *coincidence detectors*; these proteins are equivalent to *AND gates* in the microprocessor of a computer, in that they are only activated if they receive multiple converging signals (Figure 15–12).
7. *Coordination* of multiple responses in one cell can be achieved by a single extracellular signal. Some extracellular signal molecules, for example, stimulate a cell to both grow and divide. This coordination generally depends on mechanisms for distributing a signal to multiple effectors, by creating branches in the signaling pathway. In some cases, the branching of signaling pathways can allow one signal to *modulate* the strength of a response to other signals.

Given the complexity that arises from behaviors like signal integration, distribution, and feedback, it is clear that signaling systems rarely depend on a simple linear sequence of steps but are often more like a network, in which information flows not just forward but in multiple directions—and sometimes even backward. A major research challenge is to understand the nature of these networks and the response behaviors they can achieve.

### The Speed of a Response Depends on the Turnover of Signaling Molecules

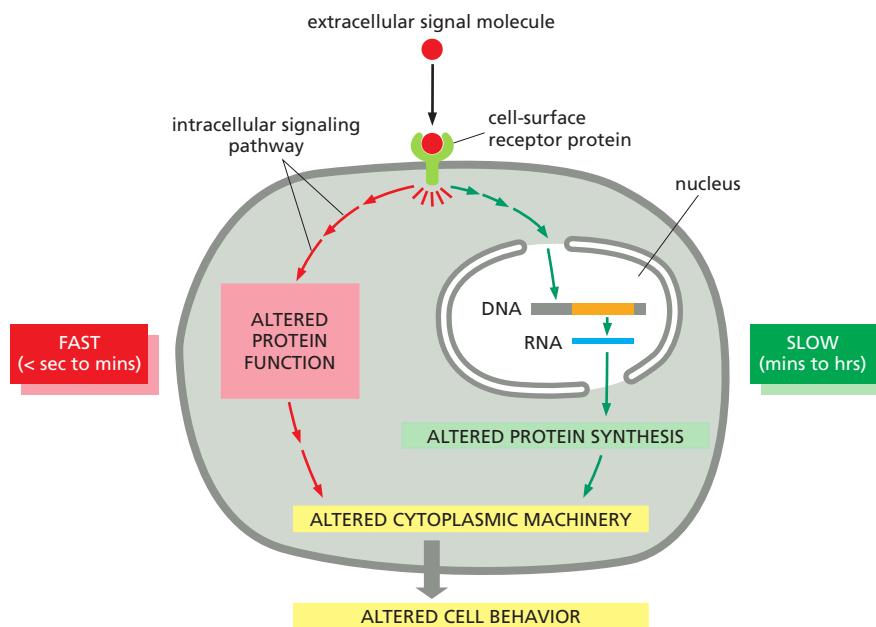
The speed of any signaling response depends on the nature of the intracellular signaling molecules that carry out the target cell's response. When the response requires only changes in proteins already present in the cell, it can occur very rapidly: an allosteric change in a neurotransmitter-gated ion channel (discussed in Chapter 11), for example, can alter the plasma membrane electrical potential in milliseconds, and responses that depend solely on protein phosphorylation can occur within seconds. When the response involves changes in gene expression and the synthesis of new proteins, however, it usually requires many minutes or hours, regardless of the mode of signal delivery (Figure 15–13).

It is natural to think of intracellular signaling systems in terms of the changes produced when an extracellular signal is delivered. But it is just as important to consider what happens when the signal is withdrawn. During development, transient extracellular signals often produce lasting effects: they can trigger a change in the cell's development that persists indefinitely through cell memory mechanisms, as we discuss later (and in Chapters 7 and 22). In most cases in adult tissues, however, the response fades when a signal ceases. Often the effect is transient



**Figure 15–12** Signal integration.

Extracellular signals A and B activate different intracellular signaling pathways, each of which leads to the phosphorylation of protein Y but at different sites on the protein. Protein Y is activated only when both of these sites are phosphorylated, and therefore it becomes active only when signals A and B are simultaneously present. Such proteins are often called coincidence detectors.



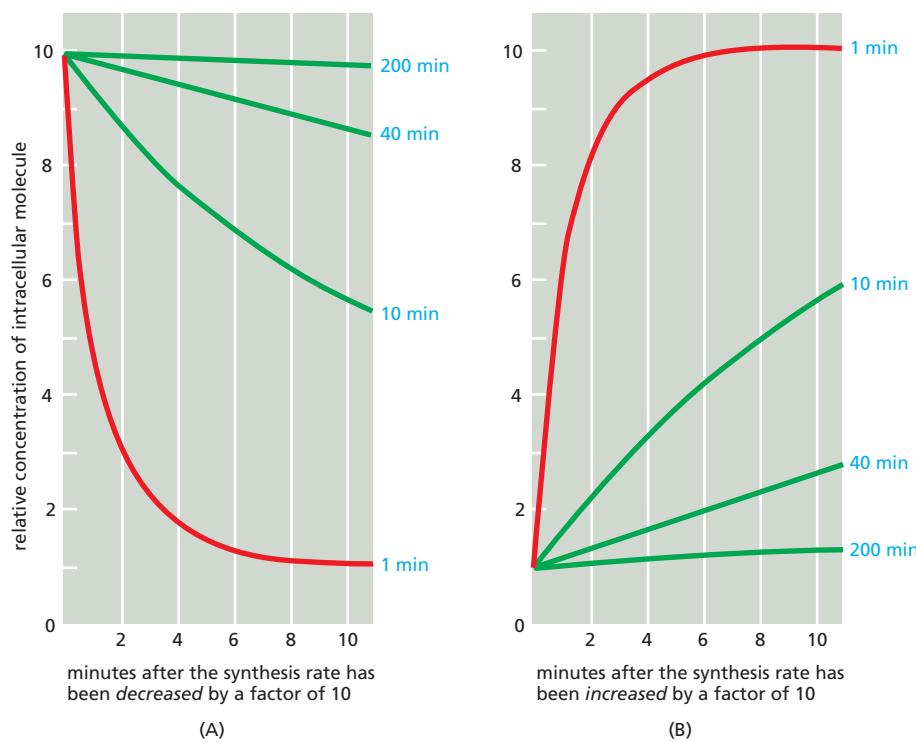
**Figure 15–13 Slow and rapid responses to an extracellular signal.** Certain types of signal-induced cellular responses, such as increased cell growth and division, involve changes in gene expression and the synthesis of new proteins; they therefore occur slowly, often starting an hour or more after the signal is received. Other responses—such as changes in cell movement, secretion, or metabolism—need not involve changes in gene transcription and therefore occur much more quickly, often starting in seconds or minutes; they may involve the rapid phosphorylation of effector proteins in the cytoplasm, for example. Synaptic responses mediated by changes in membrane potential are even quicker and can occur in milliseconds (not shown). Some signaling systems generate both rapid and slow responses as shown here, allowing the cell to respond quickly to a signal while simultaneously initiating a more long-term, persistent response.

because the signal exerts its effects by altering the concentrations of intracellular molecules that are short-lived (unstable), undergoing continual turnover. Thus, once the extracellular signal is gone, the degradation of the old molecules quickly wipes out all traces of the signal's action. It follows that the speed with which a cell responds to signal removal depends on the rate of destruction, or turnover, of the intracellular molecules that the signal affects.

It is also true, although much less obvious, that this turnover rate can determine the promptness of the response when an extracellular signal arrives. Consider, for example, two intracellular signaling molecules, X and Y, both of which are normally maintained at a steady-state concentration of 1000 molecules per cell. The cell synthesizes and degrades molecule Y at a rate of 100 molecules per second, with each molecule having an average lifetime of 10 seconds. Molecule X has a turnover rate that is 10 times slower than that of Y: it is both synthesized and degraded at a rate of 10 molecules per second, so that each molecule has an average lifetime in the cell of 100 seconds. If a signal acting on the cell causes a tenfold increase in the synthesis rates of both X and Y with no change in the molecular lifetimes, at the end of 1 second the concentration of Y will have increased by nearly 900 molecules per cell ( $10 \times 100 - 100$ ), while the concentration of X will have increased by only 90 molecules per cell. In fact, after a molecule's synthesis rate has been either increased or decreased abruptly, the time required for the molecule to shift halfway from its old to its new equilibrium concentration is equal to its half-life—that is, equal to the time that would be required for its concentration to fall by half if all synthesis were stopped (Figure 15–14).

The same principles apply to proteins and small molecules, whether the molecules are in the extracellular space or inside cells. Many intracellular proteins have short half-lives, some surviving for less than 10 minutes. In most cases, these are key regulatory proteins whose concentrations are rapidly controlled in the cell by changes in their rates of synthesis.

As we have seen, many cell responses to extracellular signals depend on the conversion of intracellular signaling proteins from an inactive to an active form, rather than on their synthesis or degradation. Phosphorylation or the binding of GTP, for example, commonly activates signaling proteins. Even in these cases, however, the activation must be rapidly and continuously reversed (by dephosphorylation or GTP hydrolysis to GDP, respectively, in these examples) to make rapid signaling possible. These inactivation processes play a crucial part in determining the magnitude, rapidity, and duration of the response.



**Figure 15–14 The importance of rapid turnover.** The graphs show the predicted relative rates of change in the intracellular concentrations of molecules with differing turnover times when their synthesis rates are either (A) decreased or (B) increased suddenly by a factor of 10. In both cases, the concentrations of those molecules that are normally degraded rapidly in the cell (red lines) change quickly, whereas the concentrations of those that are normally degraded slowly (green lines) change proportionally more slowly. The numbers (in blue) on the right are the half-lives assumed for each of the different molecules.

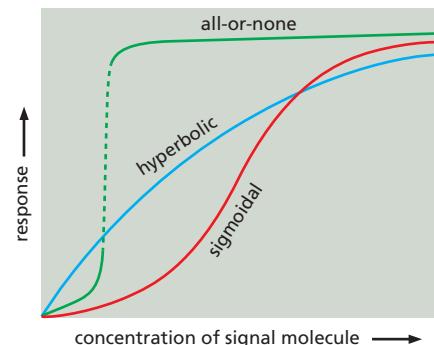
### Cells Can Respond Abruptly to a Gradually Increasing Signal

Some signaling systems are capable of generating a smoothly graded response over a wide range of extracellular signal concentrations (Figure 15–15, blue line); such systems are useful, for example, in the fine tuning of metabolic processes by some hormones. Other signaling systems generate significant responses only when the signal concentration rises beyond some threshold value. These abrupt responses are of two types. One is a *sigmoidal* response, in which low concentrations of stimulus do not have much effect, but then the response rises steeply and continuously at intermediate stimulus levels (Figure 15–15, red line). Such systems provide a filter to reduce inappropriate responses to low-level background signals but respond with high sensitivity when the stimulus falls within a small range of physiological signal concentrations. A second type of abrupt response is the *discontinuous* or *all-or-none* response, in which the response switches on completely (and often irreversibly) when the signal reaches some threshold concentration (Figure 15–15, green line). Such responses are particularly useful for controlling the choice between two alternative cell states, and they generally involve positive feedback, as we describe in more detail shortly.

Cells use a variety of molecular mechanisms to produce a sigmoidal response to increasing signal concentrations. In one mechanism, more than one intracellular signaling molecule must bind to its downstream target protein to induce a response. As we discuss later, for example, four molecules of the second messenger cyclic AMP must bind simultaneously to each molecule of *cyclic-AMP-dependent protein kinase (PKA)* to activate the kinase. A similar sharpening of response is seen when the activation of an intracellular signaling protein requires phosphorylation at more than one site. Such responses become sharper as the number of required molecules or phosphate groups increases, and if the number is large enough, responses become almost all-or-none (Figure 15–16).

Responses are also sharpened when an intracellular signaling molecule activates one enzyme and also inhibits another enzyme that catalyzes the opposite reaction. A well-studied example of this common type of regulation is the stimulation of glycogen breakdown in skeletal muscle cells induced by the hormone *adrenaline* (epinephrine). Adrenaline's binding to a G-protein-coupled

**Figure 15–15 Signal processing can produce smoothly graded or switchlike responses.** Some cell responses increase gradually as the concentration of extracellular signal molecule increases, eventually reaching a plateau as the signaling pathway is saturated, resulting in a *hyperbolic* response curve (blue line). In other cases, the signaling system reduces the response at low signal concentrations and then produces a steeper response at some intermediate signal concentration—resulting in a *sigmoidal* response curve (red line). In still other cases, the response is more abrupt and switchlike; the cell switches completely between a low and high response, without any stable intermediate response (green line).

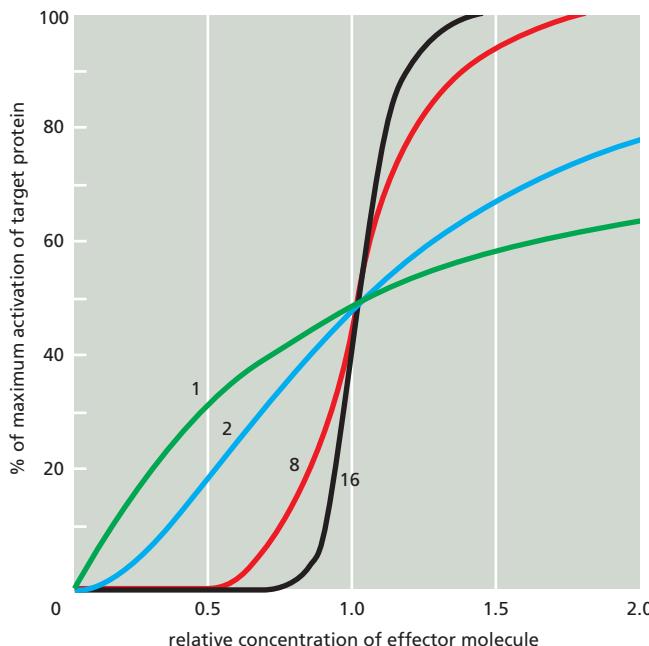


cell-surface receptor increases the intracellular concentration of cyclic AMP, which both activates an enzyme that promotes glycogen breakdown and inhibits an enzyme that promotes glycogen synthesis.

### Positive Feedback Can Generate an All-or-None Response

Like intracellular metabolic pathways (discussed in Chapter 2) and the systems controlling gene activity (Chapter 7), most intracellular signaling systems incorporate feedback loops, in which the output of a process acts back to regulate that same process. We discussed the mathematical analysis of feedback loops in Chapter 8. In *positive feedback*, the output stimulates its own production; in *negative feedback*, the output inhibits its own production (Figure 15–17). Feedback loops are of great general importance in biology, and they regulate many chemical and physical processes in cells. Those that regulate cell signaling can either operate exclusively within the target cell or involve the secretion of extracellular signals. Here, we focus on those feedback loops that operate entirely within the target cell; even the simplest of these loops can produce complex and interesting effects.

Positive feedback in a signaling pathway can transform the behavior of the responding cell. If the positive feedback is of only moderate strength, its effect will be simply to steepen the response to the signal, generating a sigmoidal response like those described earlier; but if the feedback is strong enough, it can produce an all-or-none response (see Figure 15–15). This response goes hand in hand with a further property: once the responding system has switched to the high level of activation, this condition is often self-sustaining and can persist even after the



**Figure 15–16 Activation curves for an allosteric protein as a function of effector molecule concentration.** The curves show how the sharpness of the activation response increases with an increase in the number of allosteric effector molecules that must be bound simultaneously to activate the target protein. The curves shown are those expected, under certain conditions, if the activation requires the simultaneous binding of 1, 2, 8, or 16 effector molecules.

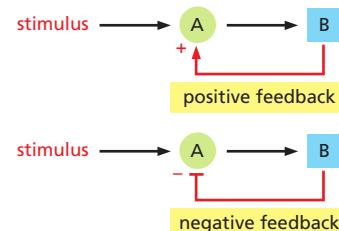
signal strength drops back below its critical value. In such a case, the system is said to be *bistable*: it can exist in either a “switched-off” or a “switched-on” state, and a transient stimulus can flip it from the one state to the other (Figure 15–18A and B).

Through positive feedback, a transient extracellular signal can induce long-term changes in cells and their progeny that can persist for the lifetime of the organism. The signals that trigger muscle-cell specification, for example, turn on the transcription of a series of genes that encode muscle-specific transcription regulatory proteins, which stimulate the transcription of their own genes, as well as genes encoding various other muscle-cell proteins; in this way, the decision to become a muscle cell is made permanent. This type of cell memory, which depends on positive feedback, is one of the basic ways in which a cell can undergo a lasting change of character without any alteration in its DNA sequence.

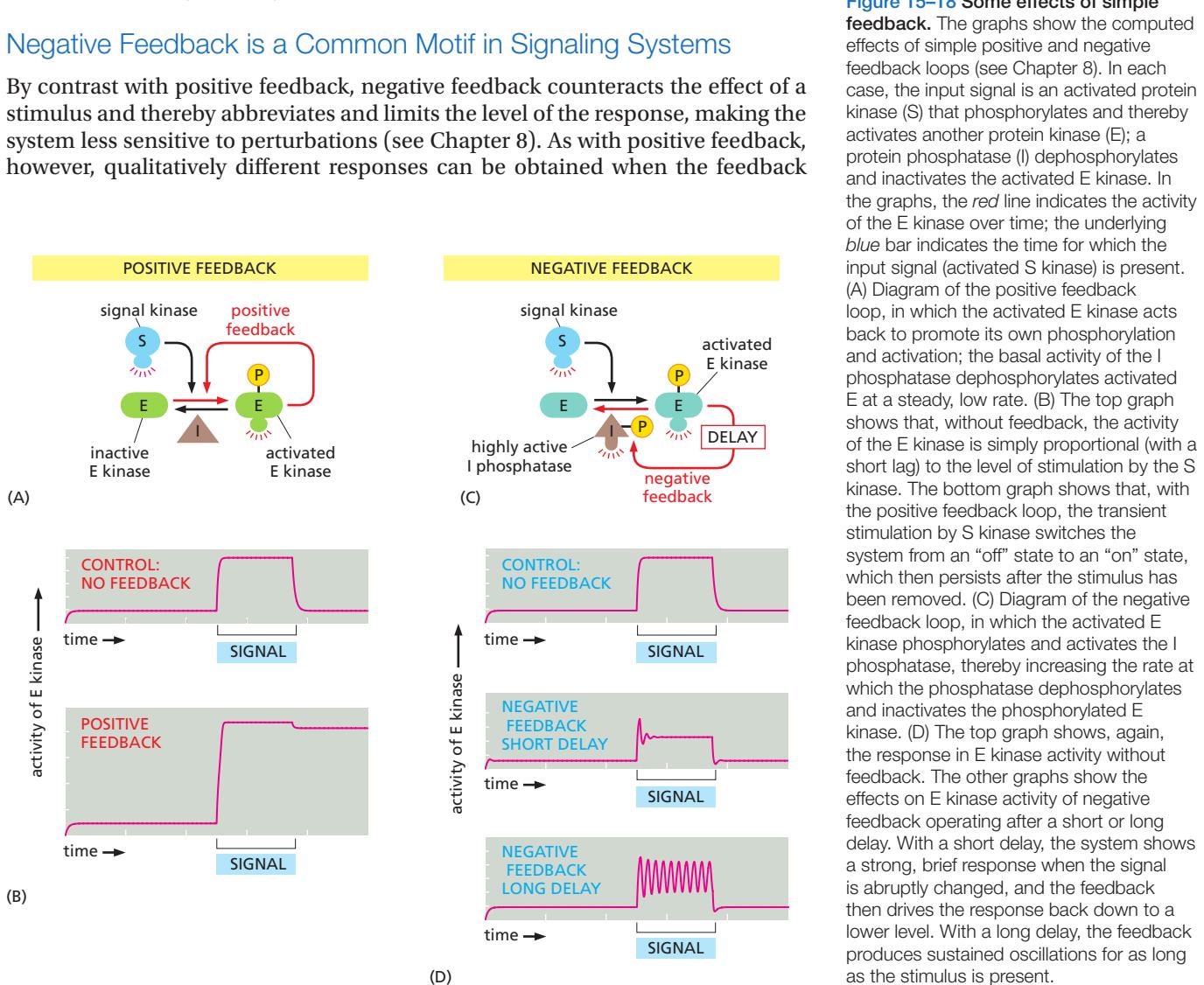
Studies of signaling responses in large populations of cells can give the false impression that a response is smoothly graded, even when strong positive feedback is causing an abrupt, discontinuous switch in the response in individual cells. Only by studying the response in single cells is it possible to see its all-or-none character (Figure 15–19). The misleading smooth response in a cell population is due to the random, intrinsic variability in signaling systems that we described earlier: all cells in a population do not respond identically to the same concentration of extracellular signal, especially at intermediate signal concentrations where the receptor is only partially occupied.

### Negative Feedback is a Common Motif in Signaling Systems

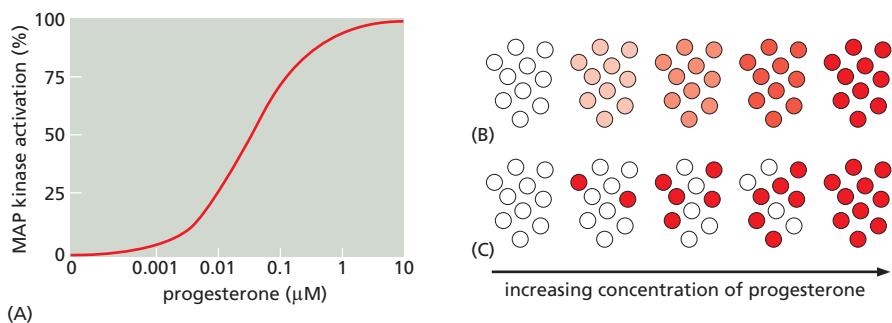
By contrast with positive feedback, negative feedback counteracts the effect of a stimulus and thereby abbreviates and limits the level of the response, making the system less sensitive to perturbations (see Chapter 8). As with positive feedback, however, qualitatively different responses can be obtained when the feedback



**Figure 15–17 Positive and negative feedback.** In these simple examples, a stimulus activates protein A, which, in turn, activates protein B. Protein B then acts back to either increase or decrease the activity of A.



**Figure 15–18 Some effects of simple feedback.** The graphs show the computed effects of simple positive and negative feedback loops (see Chapter 8). In each case, the input signal is an activated protein kinase (S) that phosphorylates and thereby activates another protein kinase (E); a protein phosphatase (I) dephosphorylates and inactivates the activated E kinase. In the graphs, the red line indicates the activity of the E kinase over time; the underlying blue bar indicates the time for which the input signal (activated S kinase) is present. (A) Diagram of the positive feedback loop, in which the activated E kinase acts back to promote its own phosphorylation and activation; the basal activity of the I phosphatase dephosphorylates activated E at a steady, low rate. (B) The top graph shows that, without feedback, the activity of the E kinase is simply proportional (with a short lag) to the level of stimulation by the S kinase. The bottom graph shows that, with the positive feedback loop, the transient stimulation by S kinase switches the system from an “off” state to an “on” state, which then persists after the stimulus has been removed. (C) Diagram of the negative feedback loop, in which the activated E kinase phosphorylates and activates the I phosphatase, thereby increasing the rate at which the phosphatase dephosphorylates and inactivates the phosphorylated E kinase. (D) The top graph shows, again, the response in E kinase activity without feedback. The other graphs show the effects on E kinase activity of negative feedback operating after a short or long delay. With a short delay, the system shows a strong, brief response when the signal is abruptly changed, and the feedback then drives the response back down to a lower level. With a long delay, the feedback produces sustained oscillations for as long as the stimulus is present.



**Figure 15–19** The importance of examining individual cells to detect all-or-none responses to increasing concentrations of an extracellular signal. In these experiments, immature frog eggs (oocytes) were stimulated with increasing concentrations of the hormone progesterone. The response was assessed by analyzing the activation of MAP kinase (discussed later), which is one of the protein kinases activated by phosphorylation in the response. The amount of phosphorylated (activated) MAP kinase in extracts of the oocytes was assessed biochemically. In (A), extracts of populations of stimulated oocytes were analyzed, and the activation of MAP kinase appeared to increase progressively with increasing progesterone concentration. There are two possible ways of explaining this result: (B) MAP kinase could have increased gradually in each individual cell with increasing progesterone concentration; or (C) individual cells could have responded in an all-or-none way, with the gradual increase in total MAP kinase activation reflecting the increasing number of cells responding with increasing progesterone concentration. When extracts of individual oocytes were analyzed, it was found that cells had either very low amounts or very high amounts, but not intermediate amounts, of the activated kinase, indicating that the response was essentially all-or-none at the level of individual cells, as diagrammed in (C). Subsequent studies revealed that this all-or-none response is due in part to strong positive feedback in the progesterone signaling system. (Adapted from J.E. Ferrell and E.M. Machleder, *Science* 280:895–898, 1998. With permission from AAAS.)

operates more powerfully. A delayed negative feedback with a long enough delay can produce responses that oscillate. The oscillations may persist for as long as the stimulus is present (Figure 15–18C and D) or they may even be generated spontaneously, without need of an external signal to drive them. Many such oscillators also contain positive feedback loops that generate sharper oscillations. Later in this chapter, we will encounter specific examples of oscillatory behavior in the intracellular responses to extracellular signals; all of them depend on negative feedback, generally accompanied by positive feedback.

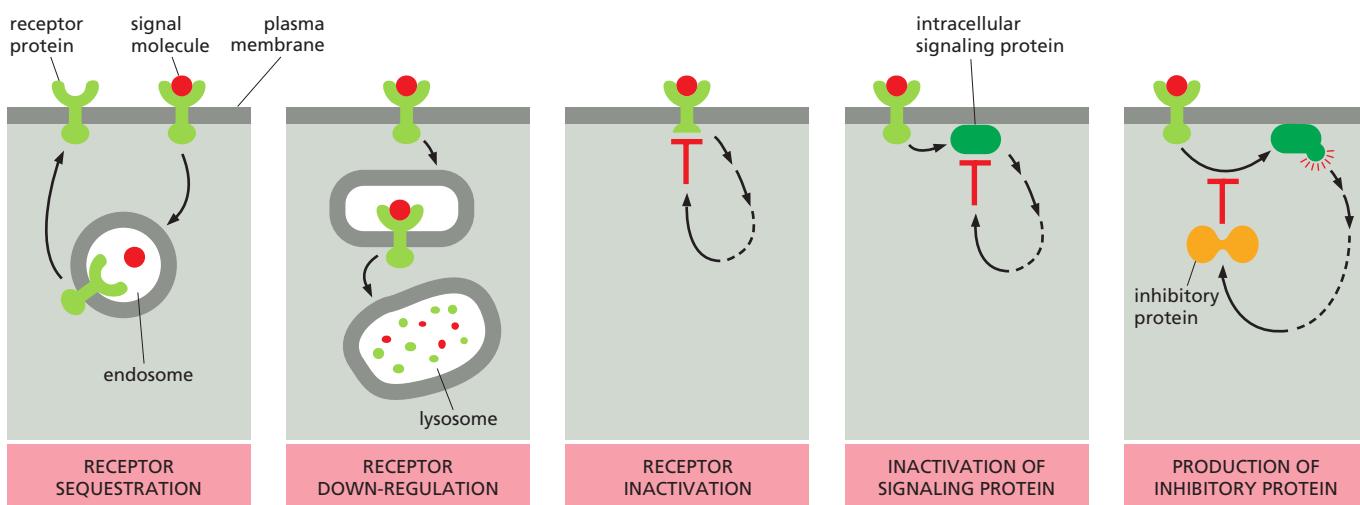
If negative feedback operates with a short delay, the system behaves like a change detector. It gives a strong response to a stimulus, but the response rapidly decays even while the stimulus persists; if the stimulus is suddenly increased, however, the system responds strongly again, but, again, the response rapidly decays. This is the phenomenon of *adaptation*, which we now discuss.

### Cells Can Adjust Their Sensitivity to a Signal

In responding to many types of stimuli, cells and organisms are able to detect the same percentage of change in a signal over a wide range of stimulus strengths. The target cells accomplish this through a reversible process of **adaptation**, or **desensitization**, whereby a prolonged exposure to a stimulus decreases the cells' response to that level of stimulus. In chemical signaling, adaptation enables cells to respond to *changes* in the concentration of an extracellular signal molecule (rather than to the absolute concentration of the signal) over a very wide range of signal concentrations. The underlying mechanism is negative feedback that operates with a short delay: a strong response modifies the signaling machinery involved, such that the machinery resets itself to become less responsive to the same level of signal (see Figure 15–18D, middle graph). Owing to the delay, however, a sudden increase in the signal is able to stimulate the cell again for a short period before the negative feedback has time to kick in.

Adaptation to a signal molecule can occur in various ways. It can result from inactivation of the receptors themselves. The binding of signal molecules to cell-surface receptors, for example, may induce the endocytosis and temporary sequestration of the receptors in endosomes. In some cases, such signal-induced receptor endocytosis leads to the destruction of the receptors in lysosomes, a process referred to as *receptor down-regulation* (in other cases, however, activated receptors continue to signal after they have been endocytosed). Receptors can also become inactivated on the cell surface—for example, by becoming phosphorylated—with a short delay following their activation. Adaptation can also occur at sites downstream of the receptors, either by a change in intracellular signaling proteins involved in transducing the extracellular signal or by the production of an inhibitor protein that blocks the signal transduction process. These various adaptation mechanisms are compared in Figure 15–20.

Though bewildering in their complexity, the multiple cross-regulatory signaling pathways and feedback loops that we describe in this chapter are not just a haphazard tangle, but a highly evolved system for processing and interpreting



**Figure 15–20** Some ways in which target cells can become adapted (desensitized) to an extracellular signal molecule. The mechanisms shown here that operate at the level of the receptor often involve phosphorylation or ubiquitylation of the receptor proteins.

the vast number of signals that impinge upon animal cells. The whole molecular control network, leading from the receptors at the cell surface to the genes in the nucleus, can be viewed as a computing device; and, like that other biological computing device, the brain, it presents one of the hardest problems in biology. We can identify the components and discover how they work individually. We can understand how small subsets of components work together as regulatory modules, noise filters, or adaptation mechanisms, as we have seen. However, it is a much more difficult task to understand how the system works as a whole. This is not only because the system is complex; it is also because the way in which it behaves is strongly dependent on the quantitative details of the molecular interactions, and, for most animal cells, we have only rough qualitative information. A major challenge for the future of signaling research is to develop more sophisticated quantitative and computational methods for the analysis of signaling systems, as described in Chapter 8.

## Summary

*Each cell in a multicellular animal is programmed to respond to a specific set of extracellular signal molecules produced by other cells. The signal molecules act by binding to a complementary set of receptor proteins expressed by the target cells. Most extracellular signal molecules activate cell-surface receptor proteins, which act as signal transducers, converting the extracellular signal into intracellular ones that alter the behavior of the target cell. Activated receptors relay the signal into the cell interior by activating intracellular signaling proteins. Some of these signaling proteins transduce, amplify, or spread the signal as they relay it, while others integrate signals from different signaling pathways. Some function as switches that are transiently activated by phosphorylation or GTP binding. Large signaling complexes form by means of modular interaction domains in the signaling proteins, which allow the proteins to form functional signaling networks.*

*Target cells use various mechanisms, including feedback loops, to adjust the ways in which they respond to extracellular signals. Positive feedback loops can help cells to respond in an all-or-none fashion to a gradually increasing concentration of an extracellular signal and to convert a short-lasting signal into a long-lasting, or even irreversible, response. Negative feedback allows cells to adapt to a signal molecule, which enables them to respond to small changes in the concentration of the signal molecule over a large concentration range.*

## SIGNALING THROUGH G-PROTEIN-COUPLED RECEPTORS

**G-protein-coupled receptors (GPCRs)** form the largest family of cell-surface receptors, and they mediate most responses to signals from the external world, as well as signals from other cells, including hormones, neurotransmitters, and local mediators. Our senses of sight, smell, and taste depend on them. There are more than 800 GPCRs in humans, and in mice there are about 1000 concerned with the sense of smell alone. The signal molecules that act on GPCRs are as varied in structure as they are in function and include proteins and small peptides, as well as derivatives of amino acids and fatty acids, not to mention photons of light and all the molecules that we can smell or taste. The same signal molecule can activate many different GPCR family members; for example, adrenaline activates at least 9 distinct GPCRs, acetylcholine another 5, and the neurotransmitter serotonin at least 14. The different receptors for the same signal are usually expressed in different cell types and elicit different responses.

Despite the chemical and functional diversity of the signal molecules that activate them, all GPCRs have a similar structure. They consist of a single polypeptide chain that threads back and forth across the lipid bilayer seven times, forming a cylindrical structure, often with a deep ligand-binding site at its center (**Figure 15–21**). In addition to their characteristic orientation in the plasma membrane, they all use G proteins to relay the signal into the cell interior.

The GPCR superfamily includes *rhodopsin*, the light-activated protein in the vertebrate eye, as well as the large number of olfactory receptors in the vertebrate nose. Other family members are found in unicellular organisms: the receptors in yeasts that recognize secreted mating factors are an example. It is likely that the GPCRs that mediate cell-cell signaling in multicellular organisms evolved from the sensory receptors in their unicellular eukaryotic ancestors.

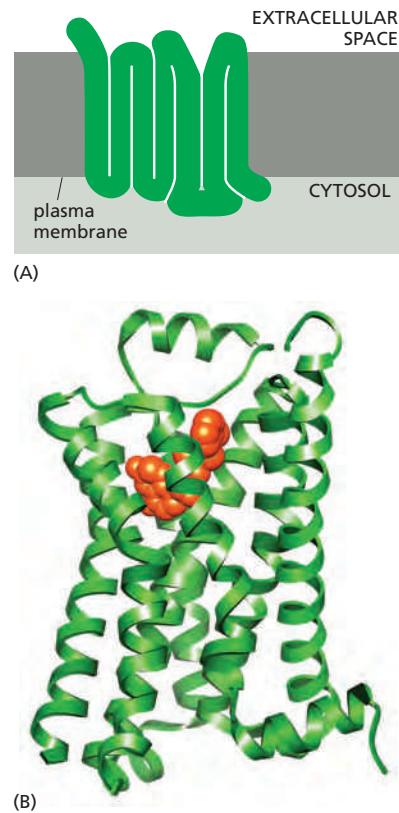
It is remarkable that almost half of all known drugs work through GPCRs or the signaling pathways GPCRs activate. Of the many hundreds of genes in the human genome that encode GPCRs, about 150 encode orphan receptors, for which the ligand is unknown. Many of them are likely targets for new drugs that remain to be discovered.

### Trimeric G Proteins Relay Signals From GPCRs

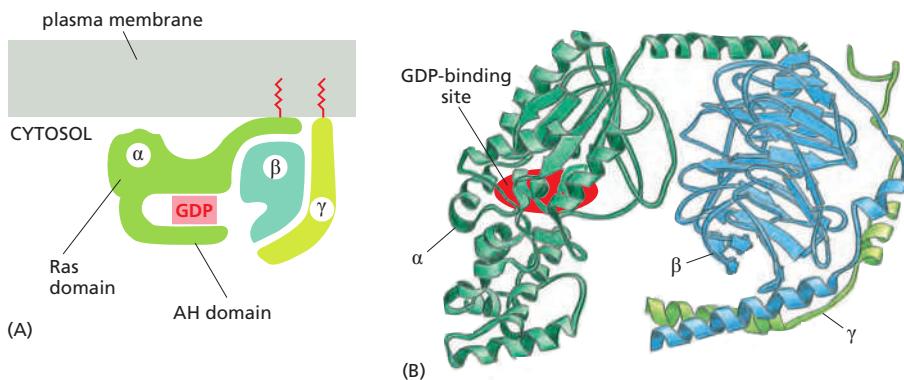
When an extracellular signal molecule binds to a GPCR, the receptor undergoes a conformational change that enables it to activate a **trimeric GTP-binding protein (G protein)**, which couples the receptor to enzymes or ion channels in the membrane. In some cases, the G protein is physically associated with the receptor before the receptor is activated, whereas in others it binds only after receptor activation. There are various types of G proteins, each specific for a particular set of GPCRs and for a particular set of target proteins in the plasma membrane. They all have a similar structure, however, and operate similarly.

G proteins are composed of three protein subunits— $\alpha$ ,  $\beta$ , and  $\gamma$ . In the unstimulated state, the  $\alpha$  subunit has GDP bound and the G protein is inactive (**Figure 15–22**). When a GPCR is activated, it acts like a guanine nucleotide exchange factor (GEF) and induces the  $\alpha$  subunit to release its bound GDP, allowing GTP to bind in its place. GTP binding then causes an activating conformational change in the  $\text{G}\alpha$  subunit, releasing the G protein from the receptor and triggering dissociation of the GTP-bound  $\text{G}\alpha$  subunit from the  $\text{G}\beta\gamma$  pair—both of which then interact with various targets, such as enzymes and ion channels in the plasma membrane, which relay the signal onward (**Figure 15–23**).

The  $\alpha$  subunit is a GTPase and becomes inactive when it hydrolyzes its bound GTP to GDP. The time required for GTP hydrolysis is usually short because the GTPase activity is greatly enhanced by the binding of the  $\alpha$  subunit to a second protein, which can be either the target protein or a specific **regulator of G protein signaling (RGS)**. RGS proteins act as  $\alpha$ -subunit-specific GTPase-activating proteins (GAPs) (see Figure 15–8), and they help shut off G-protein-mediated responses in all eukaryotes. There are about 25 RGS proteins encoded in the human genome, each of which interacts with a particular set of G proteins.



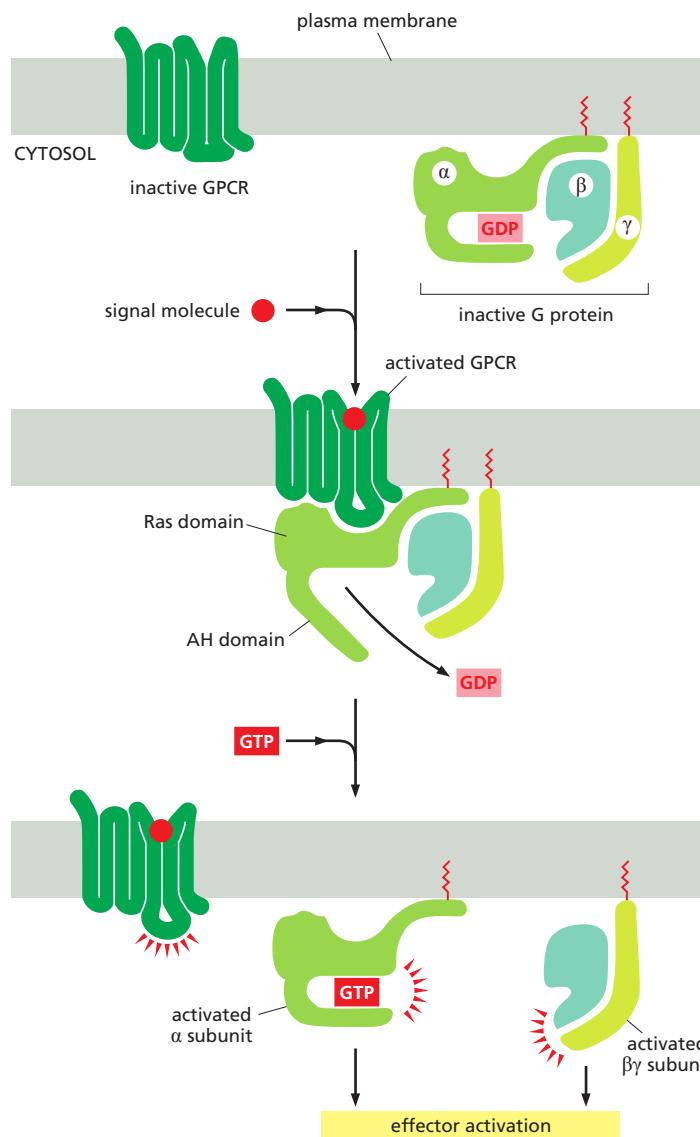
**Figure 15–21 A G-protein-coupled receptor (GPCR).** (A) GPCRs that bind small ligands such as adrenaline have small extracellular domains, and the ligand usually binds deep within the plane of the membrane to a site that is formed by amino acids from several transmembrane segments. GPCRs that bind protein ligands have a large extracellular domain (not shown here) that contributes to ligand binding. (B) The structure of the  $\beta_2$ -adrenergic receptor, a receptor for the neurotransmitter adrenaline, illustrates the typical cylindrical arrangement of the seven transmembrane helices in a GPCR. The ligand (orange) binds in a pocket between the helices, resulting in conformational changes on the cytoplasmic surface of the receptor that promote G-protein activation (not shown). (PDB code: 3P0G.)



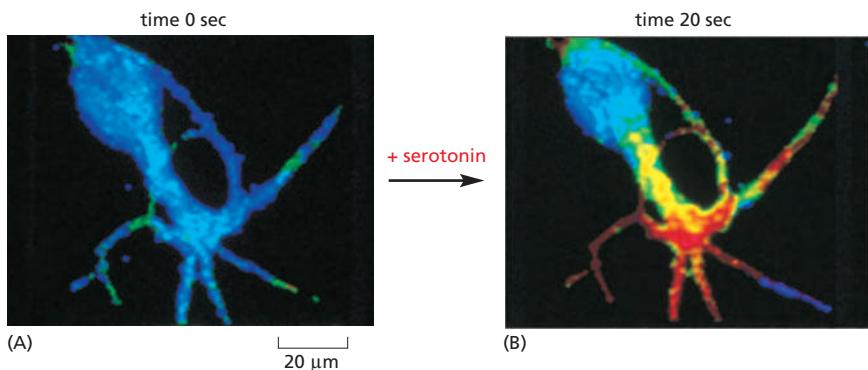
### Some G Proteins Regulate the Production of Cyclic AMP

**Cyclic AMP (cAMP)** acts as a second messenger in some signaling pathways. An extracellular signal can increase cAMP concentration more than twentyfold in seconds (Figure 15–24). As explained earlier (see Figure 15–14), such a rapid response requires balancing a rapid synthesis of the molecule with its rapid breakdown or removal. Cyclic AMP is synthesized from ATP by an enzyme called

**Figure 15–22 The structure of an inactive G protein.** (A) Note that both the  $\alpha$  and the  $\gamma$  subunits have covalently attached lipid molecules (red tails) that help bind them to the plasma membrane, and the  $\alpha$  subunit has GDP bound. (B) The three-dimensional structure of the inactive, GDP-bound form of a G protein called  $G_s$ , which interacts with numerous GPCRs, including the  $\beta_2$ -adrenergic receptor shown in Figures 15–21 and 15–23. The  $\alpha$  subunit contains the GTPase domain and binds to one side of the  $\beta$  subunit. The  $\gamma$  subunit binds to the opposite side of the  $\beta$  subunit, and the  $\beta$  and  $\gamma$  subunits together form a single functional unit. The GTPase domain of the  $\alpha$  subunit contains two major subdomains: the “Ras” domain, which is related to other GTPases and provides one face of the nucleotide-binding pocket; and the alpha-helical or “AH” domain, which clamps the nucleotide in place. (B, based on D.G. Lombright et al., *Nature* 379:311–319, 1996. With permission from Macmillan Publishers Ltd.)



**Figure 15–23 Activation of a G protein by an activated GPCR.** Binding of an extracellular signal molecule to a GPCR changes the conformation of the receptor, which allows the receptor to bind and alter the conformation of a trimeric G protein. The AH domain of the G protein  $\alpha$  subunit moves outward to open the nucleotide-binding site, thereby promoting dissociation of GDP. GTP binding then promotes closure of the nucleotide-binding site, triggering conformational changes that cause dissociation of the  $\alpha$  subunit from the receptor and from the  $\beta\gamma$  complex. The GTP-bound  $\alpha$  subunit and the  $\beta\gamma$  complex each regulate the activities of downstream signaling molecules (not shown). The receptor stays active while the extracellular signal molecule is bound to it, and it can therefore catalyze the activation of many G-protein molecules (Movie 15.1).



**adenylyl cyclase**, and it is rapidly and continuously destroyed by **cyclic AMP phosphodiesterases** (Figure 15–25). Adenylyl cyclase is a large, multipass transmembrane protein with its catalytic domain on the cytosolic side of the plasma membrane. There are at least eight isoforms in mammals, most of which are regulated by both G proteins and  $\text{Ca}^{2+}$ .

Many extracellular signals work by increasing cAMP concentrations inside the cell. These signals activate GPCRs that are coupled to a **stimulatory G protein ( $\text{G}_s$ )**. The activated  $\alpha$  subunit of  $\text{G}_s$  binds and thereby activates adenylyl cyclase. Other extracellular signals, acting through different GPCRs, reduce cAMP levels by activating an **inhibitory G protein ( $\text{G}_i$ )**, which then inhibits adenylyl cyclase.

Both  $\text{G}_s$  and  $\text{G}_i$  are targets for medically important bacterial toxins. *Cholera toxin*, which is produced by the bacterium that causes cholera, is an enzyme that catalyzes the transfer of ADP ribose from intracellular  $\text{NAD}^+$  to the  $\alpha$  subunit of  $\text{G}_s$ . This ADP ribosylation alters the  $\alpha$  subunit so that it can no longer hydrolyze its bound GTP, causing it to remain in an active state that stimulates adenylyl cyclase indefinitely. The resulting prolonged elevation in cAMP concentration within intestinal epithelial cells causes a large efflux of  $\text{Cl}^-$  and water into the gut, thereby causing the severe diarrhea that characterizes cholera. *Pertussis toxin*, which is made by the bacterium that causes pertussis (whooping cough), catalyzes the ADP ribosylation of the  $\alpha$  subunit of  $\text{G}_i$ , preventing the protein from interacting with receptors; as a result, the G protein remains in the inactive GDP-bound state and is unable to regulate its target proteins. These two toxins are widely used in experiments to determine whether a cell's GPCR-dependent response to a signal is mediated by  $\text{G}_s$  or by  $\text{G}_i$ .

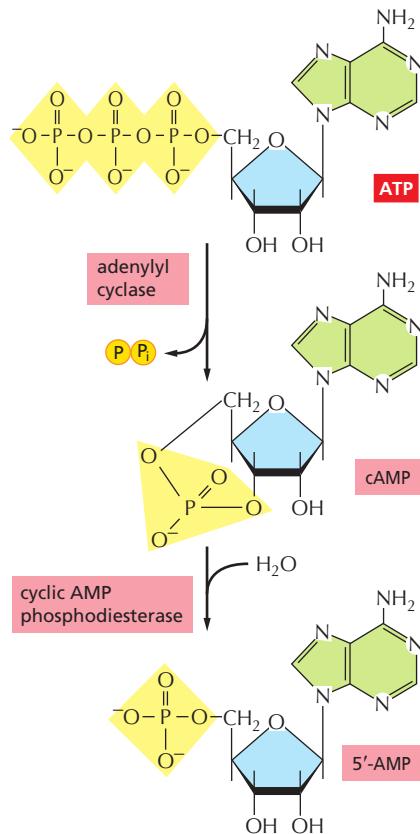
Some of the responses mediated by a  $\text{G}_s$ -stimulated increase in cAMP concentration are listed in Table 15–1. As the table shows, different cell types respond differently to an increase in cAMP concentration. Some cell types, such as fat cells, activate adenylyl cyclase in response to multiple hormones, all of which thereby stimulate the breakdown of triglyceride (the storage form of fat) to fatty acids. Individuals with genetic defects in the  $\text{G}_s \alpha$  subunit show decreased responses to certain hormones, resulting in metabolic abnormalities, abnormal bone development, and mental retardation.

### Cyclic-AMP-Dependent Protein Kinase (PKA) Mediates Most of the Effects of Cyclic AMP

In most animal cells, cAMP exerts its effects mainly by activating **cyclic-AMP-dependent protein kinase (PKA)**. This kinase phosphorylates specific serines or

**Figure 15–24** An increase in cyclic AMP in response to an extracellular signal.

This nerve cell in culture is responding to the neurotransmitter serotonin, which acts through a GPCR to cause a rapid rise in the intracellular concentration of cyclic AMP. To monitor the cyclic AMP level, the cell has been loaded with a fluorescent protein that changes its fluorescence when it binds cyclic AMP. Blue indicates a low level of cyclic AMP, yellow an intermediate level, and red a high level. (A) In the resting cell, the cyclic AMP level is about  $5 \times 10^{-8}$  M. (B) Twenty seconds after the addition of serotonin to the culture medium, the intracellular level of cyclic AMP has increased to more than  $10^{-6}$  M in the relevant parts of the cell, an increase of more than twentyfold. (From B.J. Bacskai et al., *Science* 260:222–226, 1993. With permission from AAAS.)



**Figure 15–25** The synthesis and degradation of cyclic AMP. In a reaction catalyzed by the enzyme adenylyl cyclase, cyclic AMP (cAMP) is synthesized from ATP through a cyclization reaction that removes two phosphate groups as pyrophosphate (PP<sub>i</sub>); a pyrophosphatase drives this synthesis by hydrolyzing the released pyrophosphate to phosphate (not shown). Cyclic AMP is short-lived (unstable) in the cell because it is hydrolyzed by specific phosphodiesterases to form 5'-AMP, as indicated.

**TABLE 15–1** Some Hormone-induced Cell Responses Mediated by Cyclic AMP

Target tissue	Hormone	Major response
Thyroid gland	Thyroid-stimulating hormone (TSH)	Thyroid hormone synthesis and secretion
Adrenal cortex	Adrenocorticotropic hormone (ACTH)	Cortisol secretion
Ovary	Luteinizing hormone (LH)	Progesterone secretion
Muscle	Adrenaline	Glycogen breakdown
Bone	Parathormone	Bone resorption
Heart	Adrenaline	Increase in heart rate and force of contraction
Liver	Glucagon	Glycogen breakdown
Kidney	Vasopressin	Water resorption
Fat	Adrenaline, ACTH, glucagon, TSH	Triglyceride breakdown

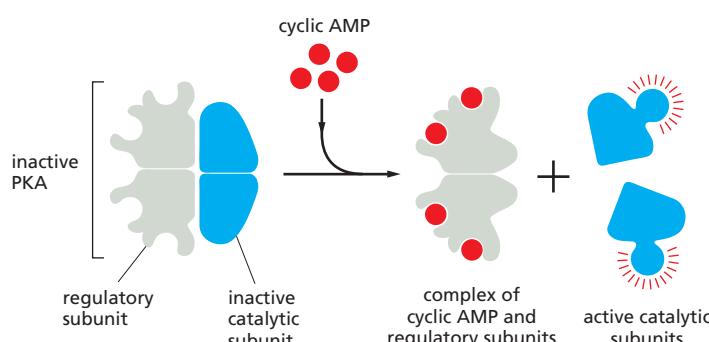
threonines on selected target proteins, including intracellular signaling proteins and effector proteins, thereby regulating their activity. The target proteins differ from one cell type to another, which explains why the effects of cAMP vary so markedly depending on the cell type (see Table 15–1).

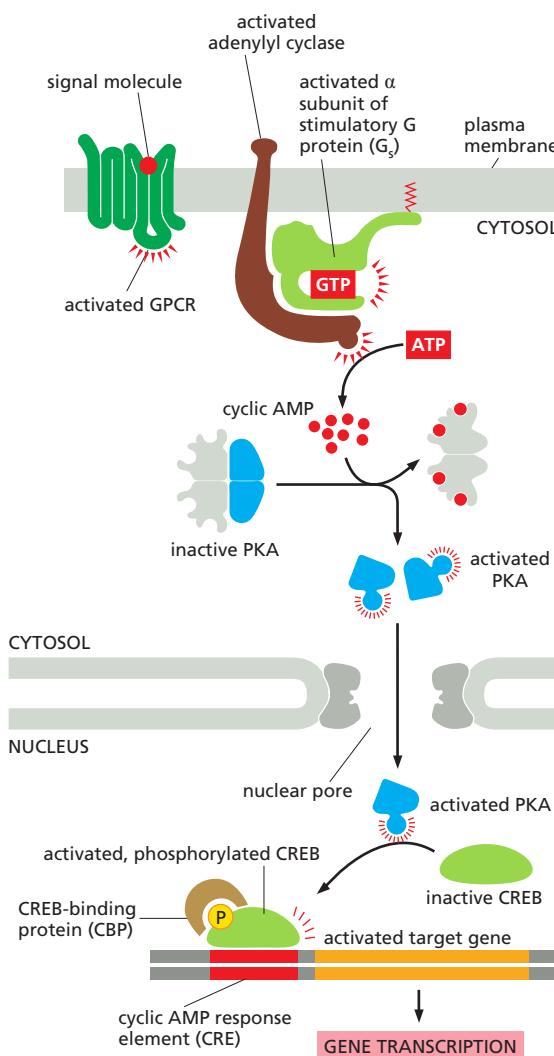
In the inactive state, PKA consists of a complex of two catalytic subunits and two regulatory subunits. The binding of cAMP to the regulatory subunits alters their conformation, causing them to dissociate from the complex. The released catalytic subunits are thereby activated to phosphorylate specific target proteins (Figure 15–26). The regulatory subunits of PKA (also called A-kinase) are important for localizing the kinase inside the cell: special *A-kinase anchoring proteins* (AKAPs) bind both to the regulatory subunits and to a component of the cytoskeleton or a membrane of an organelle, thereby tethering the enzyme complex to a particular subcellular compartment. Some AKAPs also bind other signaling proteins, forming a signaling complex. An AKAP located around the nucleus of heart muscle cells, for example, binds both PKA and a phosphodiesterase that hydrolyzes cAMP. In unstimulated cells, the phosphodiesterase keeps the local cAMP concentration low, so that the bound PKA is inactive; in stimulated cells, cAMP concentration rapidly rises, overwhelming the phosphodiesterase and activating the PKA. Among the target proteins that PKA phosphorylates and activates in these cells is the adjacent phosphodiesterase, which rapidly lowers the cAMP concentration again. This negative feedback arrangement converts what might otherwise be a prolonged PKA response into a brief, local pulse of PKA activity.

Whereas some responses mediated by cAMP occur within seconds (see Figure 15–24), others depend on changes in the transcription of specific genes and take hours to develop fully. In cells that secrete the peptide hormone *somatostatin*,

**Figure 15–26** The activation of cyclic-AMP-dependent protein kinase (PKA).

The binding of cAMP to the regulatory subunits of the PKA tetramer induces a conformational change, causing these subunits to dissociate from the catalytic subunits, thereby activating the kinase activity of the catalytic subunits. The release of the catalytic subunits requires the binding of more than two cAMP molecules to the regulatory subunits in the tetramer. This requirement greatly sharpens the response of the kinase to changes in cAMP concentration, as discussed earlier (see Figure 15–16). Mammalian cells have at least two types of PKAs: type I is mainly in the cytosol, whereas type II is bound via its regulatory subunits and special anchoring proteins to the plasma membrane, nuclear membrane, mitochondrial outer membrane, and microtubules. In both types, once the catalytic subunits are freed and active, they can migrate into the nucleus (where they can phosphorylate transcription regulatory proteins), while the regulatory subunits remain in the cytoplasm.





**Figure 15–27** How a rise in intracellular cyclic AMP concentration can alter gene transcription. The binding of an extracellular signal molecule to its GPCR activates adenyl cyclase via  $G_s$  and thereby increases cAMP concentration in the cytosol. This rise activates PKA, and the released catalytic subunits of PKA can then enter the nucleus, where they phosphorylate the transcription regulatory protein CREB. Once phosphorylated, CREB recruits the coactivator CBP, which stimulates gene transcription. In some cases, at least, the inactive CREB protein is bound to the cyclic AMP response element (CRE) in DNA before it is phosphorylated (not shown). See Movie 15.2.

for example, cAMP activates the gene that encodes this hormone. The regulatory region of the somatostatin gene contains a short *cis*-regulatory sequence, called the *cyclic AMP response element (CRE)*, which is also found in the regulatory region of many other genes activated by cAMP. A specific transcription regulator called **CRE-binding (CREB) protein** recognizes this sequence. When PKA is activated by cAMP, it phosphorylates CREB on a single serine; phosphorylated CREB then recruits a transcriptional coactivator called *CREB-binding protein (CBP)*, which stimulates the transcription of the target genes (Figure 15–27). Thus, CREB can transform a short cAMP signal into a long-term change in a cell, a process that, in the brain, is thought to play an important part in some forms of learning and memory.

### Some G Proteins Signal Via Phospholipids

Many GPCRs exert their effects through G proteins that activate the plasma-membrane-bound enzyme **phospholipase C-β (PLC-β)**. Table 15–2 lists some examples of responses activated in this way. The phospholipase acts on a phosphorylated inositol phospholipid (a *phosphoinositide*) called **phosphatidylinositol 4,5-bisphosphate [PI(4,5)P<sub>2</sub>]**, which is present in small amounts in the inner half of the plasma membrane lipid bilayer (Figure 15–28). Receptors that activate this **inositol phospholipid signaling pathway** mainly do so via a G protein called  $G_q$ , which activates phospholipase C-β in much the same way that  $G_s$  activates adenyl cyclase. The activated phospholipase then cleaves the PI(4,5)P<sub>2</sub> to generate two

**TABLE 15–2 Some Cell Responses in Which GPCRs Activate PLC $\beta$** 

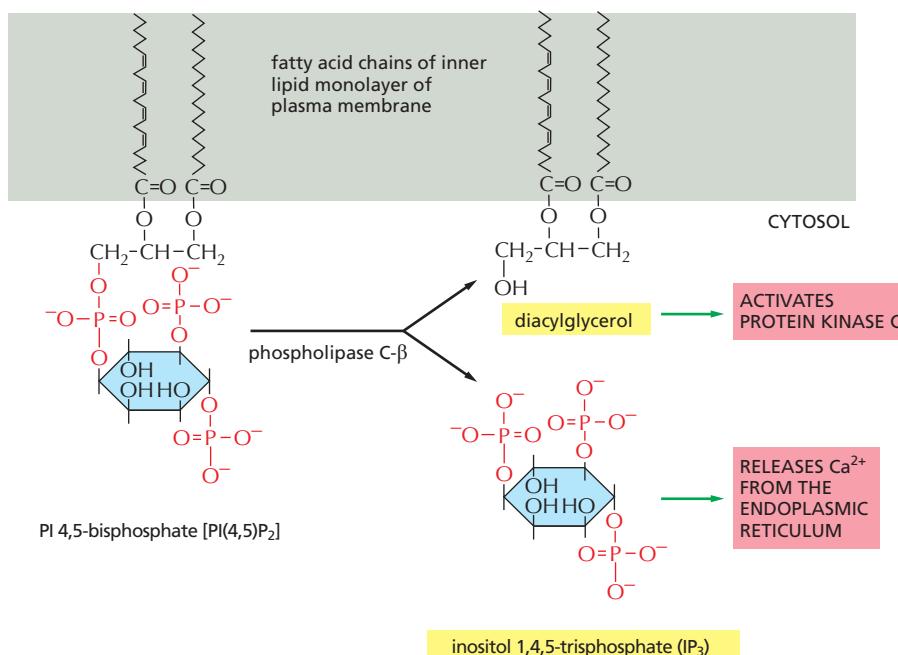
Target tissue	Signal molecule	Major response
Liver	Vasopressin	Glycogen breakdown
Pancreas	Acetylcholine	Amylase secretion
Smooth muscle	Acetylcholine	Muscle contraction
Blood platelets	Thrombin	Platelet aggregation

products: **inositol 1,4,5-trisphosphate (IP<sub>3</sub>)** and **diacylglycerol**. At this step, the signaling pathway splits into two branches.

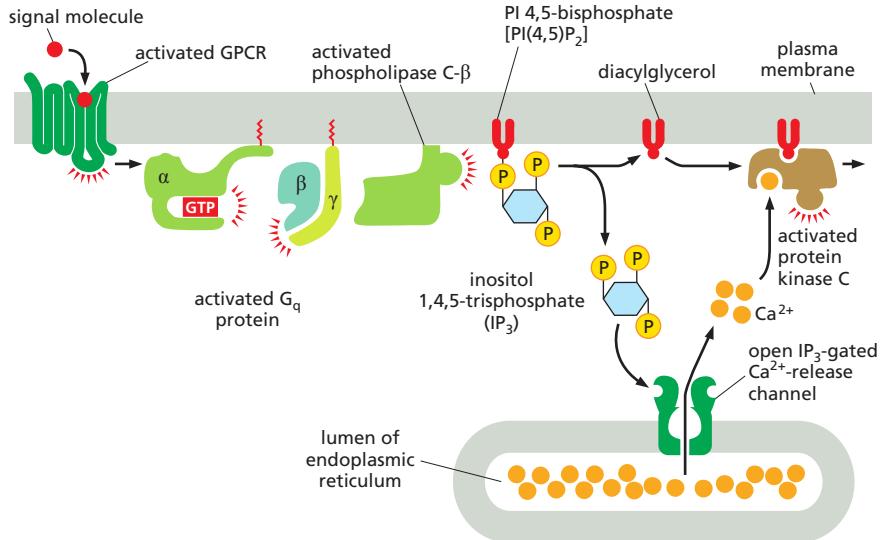
IP<sub>3</sub> is a water-soluble molecule that leaves the plasma membrane and diffuses rapidly through the cytosol. When it reaches the endoplasmic reticulum (ER), it binds to and opens **IP<sub>3</sub>-gated Ca<sup>2+</sup>-release channels** (also called **IP<sub>3</sub> receptors**) in the ER membrane. Ca<sup>2+</sup> stored in the ER is released through the open channels, quickly raising the concentration of Ca<sup>2+</sup> in the cytosol (Figure 15–29). The increase in cytosolic Ca<sup>2+</sup> propagates the signal by influencing the activity of Ca<sup>2+</sup>-sensitive intracellular proteins, as we describe shortly.

At the same time that the IP<sub>3</sub> produced by the hydrolysis of PI(4,5)P<sub>2</sub> is increasing the concentration of Ca<sup>2+</sup> in the cytosol, the other cleavage product of the PI(4,5)P<sub>2</sub>, diacylglycerol, is exerting different effects. It also acts as a second messenger, but it remains embedded in the plasma membrane, where it has several potential signaling roles. One of its major functions is to activate a protein kinase called **protein kinase C (PKC)**, so named because it is Ca<sup>2+</sup>-dependent. The initial rise in cytosolic Ca<sup>2+</sup> induced by IP<sub>3</sub> alters the PKC so that it translocates from the cytosol to the cytoplasmic face of the plasma membrane. There it is activated by the combination of Ca<sup>2+</sup>, diacylglycerol, and the negatively charged membrane phospholipid phosphatidylserine (see Figure 15–29). Once activated, PKC phosphorylates target proteins that vary depending on the cell type. The principles are the same as discussed earlier for PKA, although most of the target proteins are different.

Diacylglycerol can be further cleaved to release arachidonic acid, which can either act as a signal in its own right or be used in the synthesis of other small lipid signal molecules called *eicosanoids*. Most vertebrate cell types make eicosanoids, including *prostaglandins*, which have many biological activities. They participate



**Figure 15–28** The hydrolysis of PI(4,5)P<sub>2</sub> by phospholipase C- $\beta$ . Two second messengers are produced directly from the hydrolysis of PI(4,5)P<sub>2</sub>: inositol 1,4,5-trisphosphate (IP<sub>3</sub>), which diffuses through the cytosol and releases Ca<sup>2+</sup> from the endoplasmic reticulum, and diacylglycerol, which remains in the membrane and helps to activate protein kinase C (PKC; see Figure 15–29). There are several classes of phospholipase C: these include the  $\beta$  class, which is activated by GPCRs; as we see later, the  $\gamma$  class is activated by a class of enzyme-coupled receptors called receptor tyrosine kinases (RTKs).



in pain and inflammatory responses, for example, and many anti-inflammatory drugs (such as aspirin, ibuprofen, and cortisone) act in part by inhibiting their synthesis.

### $\text{Ca}^{2+}$ Functions as a Ubiquitous Intracellular Mediator

Many extracellular signals, and not just those that work via G proteins, trigger an increase in cytosolic  $\text{Ca}^{2+}$  concentration. In muscle cells,  $\text{Ca}^{2+}$  triggers contraction, and in many secretory cells, including nerve cells, it triggers secretion.  $\text{Ca}^{2+}$  has numerous other functions in a variety of cell types.  $\text{Ca}^{2+}$  is such an effective signaling mediator because its concentration in the cytosol is normally very low ( $\sim 10^{-7}$  M), whereas its concentration in the extracellular fluid ( $\sim 10^{-3}$  M) and in the lumen of the ER [and sarcoplasmic reticulum (SR) in muscle] is high. Thus, there is a large gradient tending to drive  $\text{Ca}^{2+}$  into the cytosol across both the plasma membrane and the ER or SR membrane. When a signal transiently opens  $\text{Ca}^{2+}$  channels in these membranes,  $\text{Ca}^{2+}$  rushes into the cytosol, and the resulting 10–20-fold increase in the local  $\text{Ca}^{2+}$  concentration activates  $\text{Ca}^{2+}$ -responsive proteins in the cell.

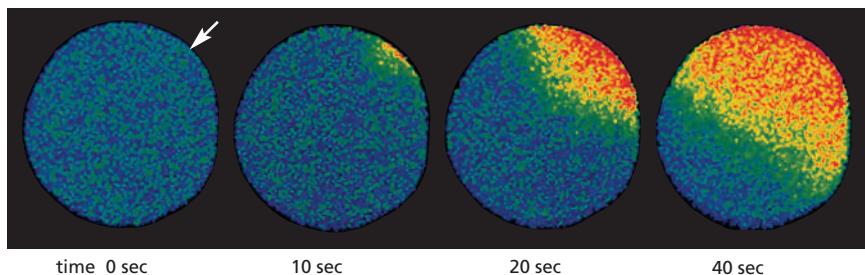
Some stimuli, including membrane depolarization, membrane stretch, and certain extracellular signals, activate  $\text{Ca}^{2+}$  channels in the plasma membrane, resulting in  $\text{Ca}^{2+}$  influx from outside the cell. Other signals, including the GPCR-mediated signals described earlier, act primarily through IP<sub>3</sub> receptors to stimulate  $\text{Ca}^{2+}$  release from intracellular stores in the ER (see Figure 15–29). The ER membrane also contains a second type of regulated  $\text{Ca}^{2+}$  channel called the **ryanodine receptor** (so called because it is sensitive to the plant alkaloid ryanodine), which opens in response to rising  $\text{Ca}^{2+}$  levels and thereby amplifies the  $\text{Ca}^{2+}$  signal, as we describe shortly.

Several mechanisms rapidly terminate the  $\text{Ca}^{2+}$  signal and are also responsible for keeping the concentration of  $\text{Ca}^{2+}$  in the cytosol low in resting cells. Most importantly, there are  $\text{Ca}^{2+}$ -pumps in the plasma membrane and the ER membrane that use the energy of ATP hydrolysis to pump  $\text{Ca}^{2+}$  out of the cytosol. Cells such as muscle and nerve cells, which make extensive use of  $\text{Ca}^{2+}$  signaling, have an additional  $\text{Ca}^{2+}$  transporter (a  $\text{Na}^+$ -driven  $\text{Ca}^{2+}$  exchanger) in their plasma membrane that couples the efflux of  $\text{Ca}^{2+}$  to the influx of  $\text{Na}^+$ .

### Feedback Generates $\text{Ca}^{2+}$ Waves and Oscillations

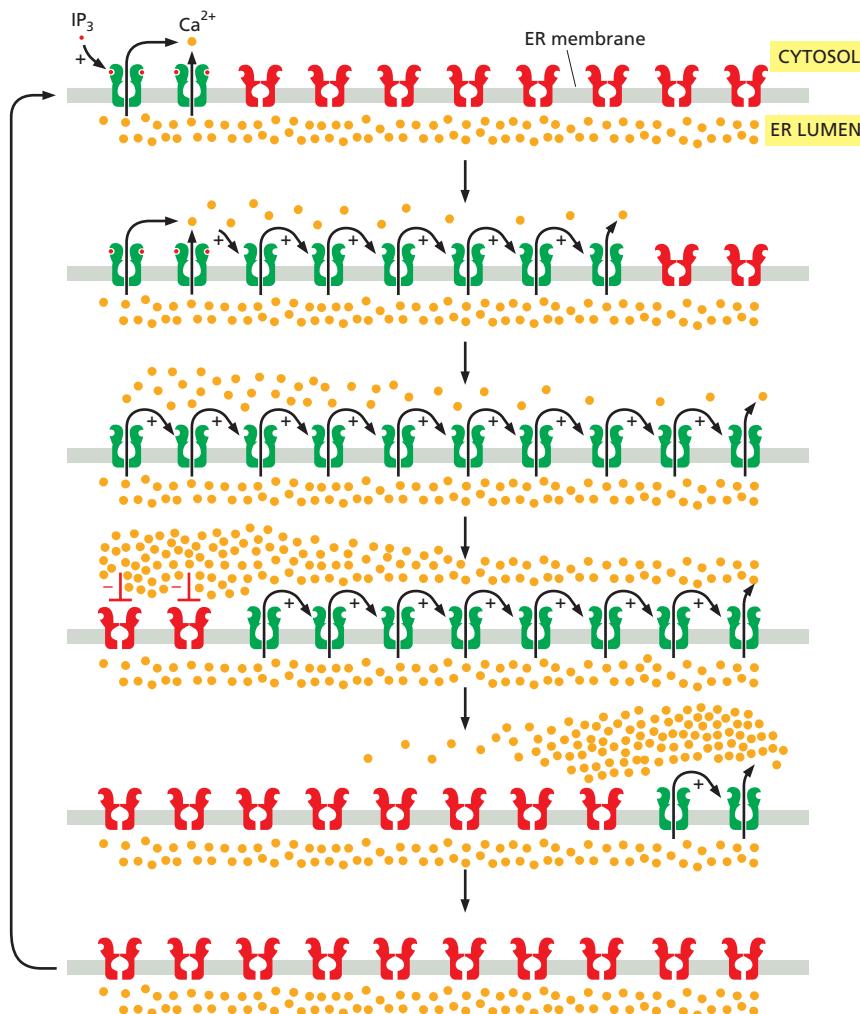
The IP<sub>3</sub> receptors and ryanodine receptors of the ER membrane have an important feature: they are both stimulated by low to moderate cytoplasmic  $\text{Ca}^{2+}$  concentrations. This  $\text{Ca}^{2+}$ -induced calcium release (CICR) results in positive feedback,

**Figure 15–29** How GPCRs increase cytosolic  $\text{Ca}^{2+}$  and activate protein kinase C. The activated GPCR stimulates the plasma-membrane-bound phospholipase C- $\beta$  (PLC $\beta$ ) via a G protein called G<sub>q</sub>. The  $\alpha$  subunit and  $\beta\gamma$  complex of G<sub>q</sub> are both involved in this activation. Two second messengers are produced when PI(4,5)P<sub>2</sub> is hydrolyzed by activated PLC $\beta$ . Inositol 1,4,5-trisphosphate (IP<sub>3</sub>) diffuses through the cytosol and releases  $\text{Ca}^{2+}$  from the ER by binding to and opening IP<sub>3</sub>-gated  $\text{Ca}^{2+}$ -release channels (IP<sub>3</sub> receptors) in the ER membrane. The large electrochemical gradient for  $\text{Ca}^{2+}$  across this membrane causes  $\text{Ca}^{2+}$  to escape into the cytosol when the release channels are opened. Diacylglycerol remains in the plasma membrane and, together with phosphatidylserine (not shown) and  $\text{Ca}^{2+}$ , helps to activate protein kinase C (PKC), which is recruited from the cytosol to the cytosolic face of the plasma membrane. Of the 10 or more distinct isoforms of PKC in humans, at least 4 are activated by diacylglycerol (Movie 15.3).



which has a major impact on the properties of the  $\text{Ca}^{2+}$  signal. The importance of this feedback is seen clearly in studies with  $\text{Ca}^{2+}$ -sensitive fluorescent indicators, such as *aequorin* or *fura-2* (discussed in Chapter 9), which allow researchers to monitor cytosolic  $\text{Ca}^{2+}$  in individual cells under a microscope (Figure 15–30 and Movie 15.4).

When cells carrying a  $\text{Ca}^{2+}$  indicator are treated with a small amount of an extracellular signal molecule that stimulates  $\text{IP}_3$  production, tiny bursts of  $\text{Ca}^{2+}$  are seen in one or more discrete regions of the cell. These  $\text{Ca}^{2+}$  puffs or sparks reflect the local opening of small groups of  $\text{IP}_3$ -gated  $\text{Ca}^{2+}$ -release channels in the ER. Because various  $\text{Ca}^{2+}$ -binding proteins act as  $\text{Ca}^{2+}$  buffers and restrict the diffusion of  $\text{Ca}^{2+}$ , the signal often remains localized to the site where the  $\text{Ca}^{2+}$  enters the cytosol. If the extracellular signal is sufficiently strong and persistent, however, the local  $\text{Ca}^{2+}$  concentration can reach a sufficient level to activate nearby  $\text{IP}_3$  receptors and ryanodine receptors, resulting in a regenerative wave of  $\text{Ca}^{2+}$  release that moves through the cytosol (Figure 15–31), much like an action potential in an axon.



**Figure 15–30** The fertilization of an egg by a sperm triggers a wave of cytosolic  $\text{Ca}^{2+}$ . This starfish egg was injected with a  $\text{Ca}^{2+}$ -sensitive fluorescent dye before it was fertilized. A wave of cytosolic  $\text{Ca}^{2+}$  (red), released from the ER, sweeps across the egg from the site of sperm entry (arrow). This  $\text{Ca}^{2+}$  wave changes the egg cell surface, preventing the entry of other sperm, and it also initiates embryonic development (Movie 15.5). The initial increase in  $\text{Ca}^{2+}$  is thought to be caused by a sperm-specific form of PLC ( $\text{PLC}\zeta$ ) that the sperm brings into the egg cytoplasm when it fuses with the egg; the  $\text{PLC}\zeta$  cleaves  $\text{PI}(4,5)\text{P}_2$  to produce  $\text{IP}_3$ , which releases  $\text{Ca}^{2+}$  from the egg ER. The released  $\text{Ca}^{2+}$  stimulates further  $\text{Ca}^{2+}$  release from the ER, producing the spreading wave, as we explain in Figure 15–31. (Courtesy of Stephen A. Stricker.)

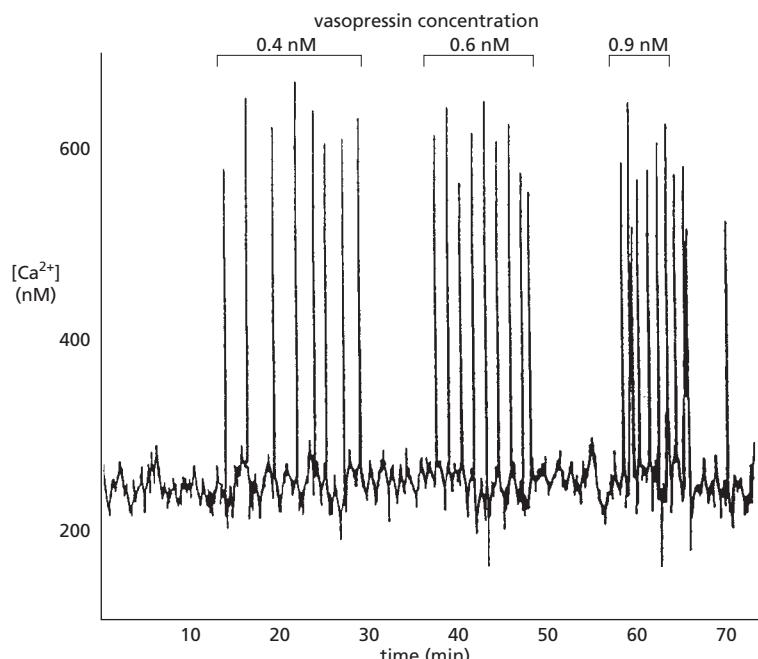
**Figure 15–31** Positive and negative feedback produce  $\text{Ca}^{2+}$  waves and oscillations. This diagram shows  $\text{IP}_3$  receptors and ryanodine receptors on a portion of the ER membrane: active receptors are in green; inactive receptors are in red. When a small amount of cytosolic  $\text{IP}_3$  activates a cluster of  $\text{IP}_3$  receptors at one site on the ER membrane (top), the local release of  $\text{Ca}^{2+}$  promotes the opening of nearby  $\text{IP}_3$  and ryanodine receptors, resulting in more  $\text{Ca}^{2+}$  release. This positive feedback (indicated by positive signs) produces a regenerative wave of  $\text{Ca}^{2+}$  release that spreads across the cell (see Figure 15–30). These waves of  $\text{Ca}^{2+}$  release move more quickly across the cell than would be possible by simple diffusion. Also, unlike a diffusing burst of  $\text{Ca}^{2+}$  ions, which will become more dilute as it spreads, the regenerative wave produces a high  $\text{Ca}^{2+}$  concentration across the entire cell. Eventually, the local  $\text{Ca}^{2+}$  concentration inactivates  $\text{IP}_3$  receptors and ryanodine receptors (middle; indicated by red negative signs), shutting down the  $\text{Ca}^{2+}$  release.  $\text{Ca}^{2+}$ -pumps reduce the local cytosolic  $\text{Ca}^{2+}$  concentration to its normal low levels. The result is a  $\text{Ca}^{2+}$  spike: positive feedback drives a rapid rise in cytosolic  $\text{Ca}^{2+}$ , and negative feedback sends it back down again. The  $\text{Ca}^{2+}$  channels remain refractory to further stimulation for some period of time, delaying the generation of another  $\text{Ca}^{2+}$  spike (bottom). Eventually, however, the negative feedback wears off, allowing  $\text{IP}_3$  to trigger another  $\text{Ca}^{2+}$  wave. The end result is repeated  $\text{Ca}^{2+}$  oscillations (see Figure 15–32). Under some conditions, these oscillations can be seen as repeating narrow waves of  $\text{Ca}^{2+}$  moving across the cell.

Another important property of IP<sub>3</sub> receptors and ryanodine receptors is that they are inhibited, after some delay, by high Ca<sup>2+</sup> concentrations (a form of negative feedback). Thus, the rise in Ca<sup>2+</sup> in a stimulated cell leads to inhibition of Ca<sup>2+</sup> release; because Ca<sup>2+</sup> pumps remove the cytosolic Ca<sup>2+</sup>, the Ca<sup>2+</sup> concentration falls (see Figure 15–31). The decline in Ca<sup>2+</sup> eventually relieves the negative feedback, allowing cytosolic Ca<sup>2+</sup> to rise again. As in other cases of delayed negative feedback (see Figure 15–18), the result is an oscillation in the Ca<sup>2+</sup> concentration. These oscillations persist for as long as receptors are activated at the cell surface, and their frequency reflects the strength of the extracellular stimulus (Figure 15–32). The frequency, amplitude, and breadth of oscillations can also be modulated by other signaling mechanisms, such as phosphorylation, which influence the Ca<sup>2+</sup> sensitivity of Ca<sup>2+</sup> channels or affect other components in the signaling system.

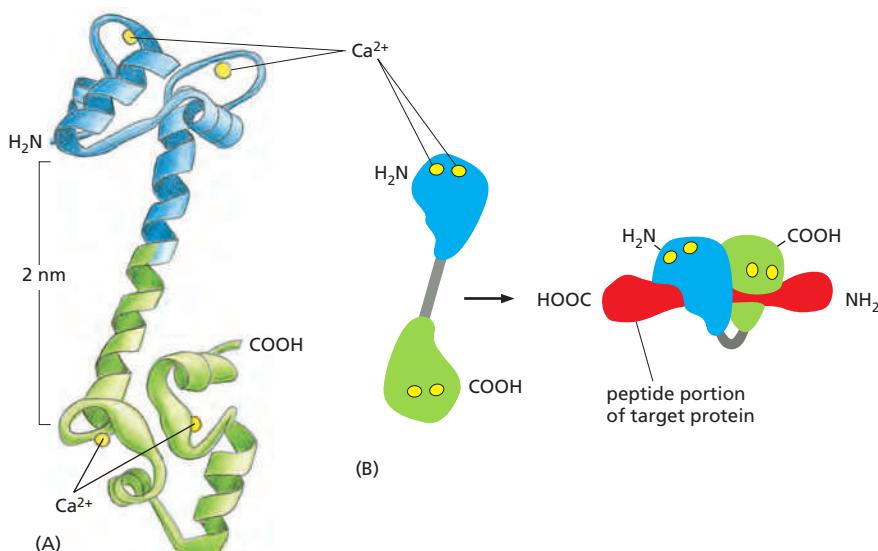
The frequency of Ca<sup>2+</sup> oscillations can be translated into a frequency-dependent cell response. In some cases, the frequency-dependent response itself is also oscillatory: in hormone-secreting pituitary cells, for example, stimulation by an extracellular signal induces repeated Ca<sup>2+</sup> spikes, each of which is associated with a burst of hormone secretion. In other cases, the frequency-dependent response is non-oscillatory: in some types of cells, for instance, one frequency of Ca<sup>2+</sup> spikes activates the transcription of one set of genes, while a higher frequency activates the transcription of a different set. How do cells sense the frequency of Ca<sup>2+</sup> spikes and change their response accordingly? The mechanism presumably depends on Ca<sup>2+</sup>-sensitive proteins that change their activity as a function of Ca<sup>2+</sup>-spike frequency. A protein kinase that acts as a molecular memory device seems to have this remarkable property, as we discuss next.

### Ca<sup>2+</sup>/Calmodulin-Dependent Protein Kinases Mediate Many Responses to Ca<sup>2+</sup> Signals

Various Ca<sup>2+</sup>-binding proteins help to relay the cytosolic Ca<sup>2+</sup> signal. The most important is **calmodulin**, which is found in all eukaryotic cells and can constitute as much as 1% of a cell's total protein mass. Calmodulin functions as a multipurpose intracellular Ca<sup>2+</sup> receptor, governing many Ca<sup>2+</sup>-regulated processes. It consists of a highly conserved, single polypeptide chain with four high-affinity Ca<sup>2+</sup>-binding sites (Figure 15–33A). When activated by Ca<sup>2+</sup> binding, it undergoes a conformational change. Because two or more Ca<sup>2+</sup> ions must bind before



**Figure 15–32** Vasopressin-induced Ca<sup>2+</sup> oscillations in a liver cell. The cell was loaded with the Ca<sup>2+</sup>-sensitive protein aequorin and then exposed to increasing concentrations of the peptide signal molecule vasopressin, which activates a GPCR and thereby PLC $\beta$  (see Table 15–2). Note that the frequency of the Ca<sup>2+</sup> spikes increases with an increasing concentration of vasopressin but that the amplitude of the spikes is not affected. Each spike lasts about 7 seconds. (Adapted from N.M. Woods, K.S.R. Cuthbertson and P.H. Cobbold, *Nature* 319:600–602, 1986. With permission from Macmillan Publishers Ltd.)



calmodulin adopts its active conformation, the protein displays a sigmoidal response to increasing concentrations of  $\text{Ca}^{2+}$  (see Figure 15–16).

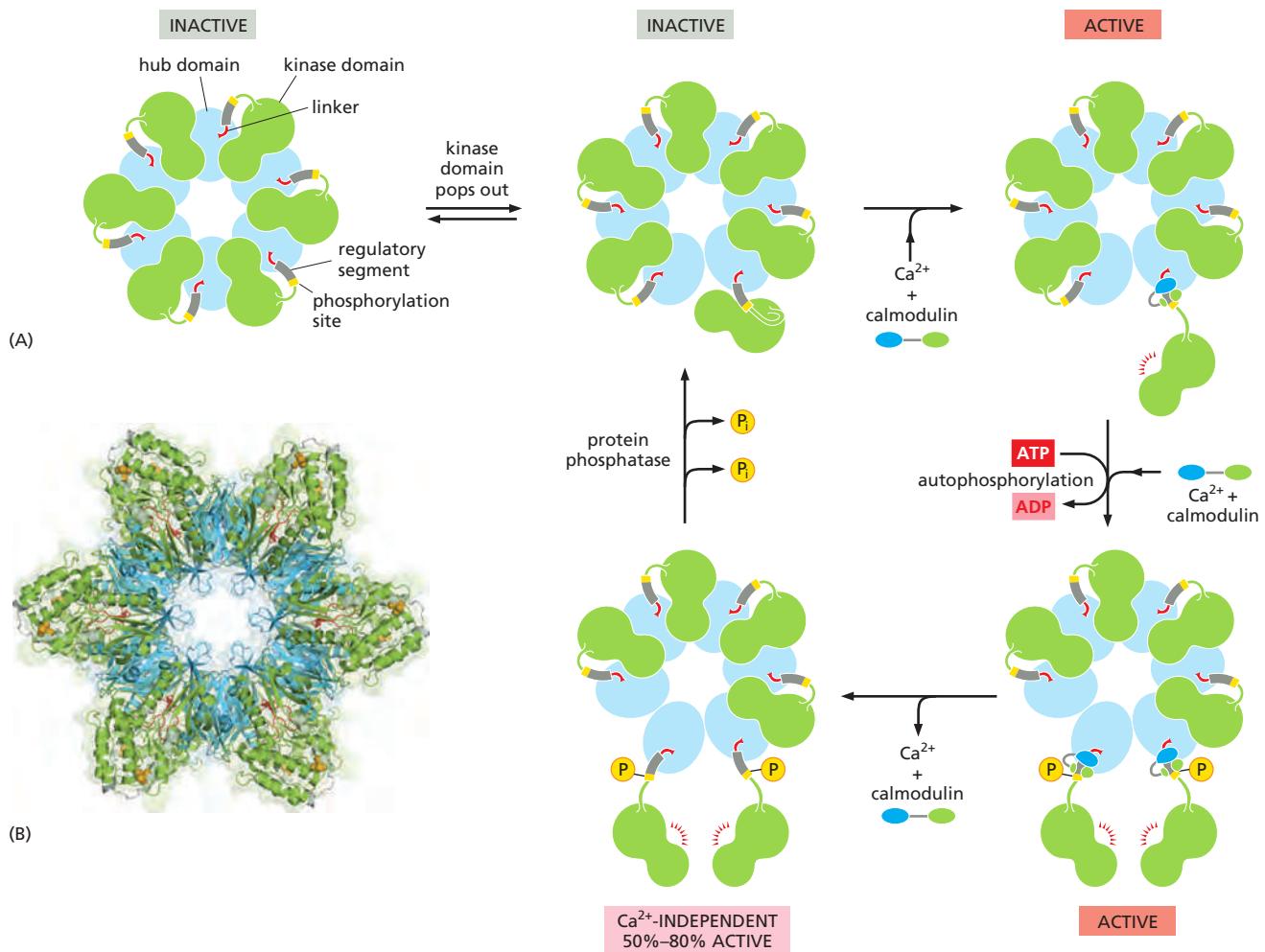
The allosteric activation of calmodulin by  $\text{Ca}^{2+}$  is analogous to the activation of PKA by cyclic AMP, except that  $\text{Ca}^{2+}$ /calmodulin has no enzymatic activity itself but instead acts by binding to and activating other proteins. In some cases, calmodulin serves as a permanent regulatory subunit of an enzyme complex, but usually the binding of  $\text{Ca}^{2+}$  instead enables calmodulin to bind to various target proteins in the cell to alter their activity.

When an activated molecule of  $\text{Ca}^{2+}$ /calmodulin binds to its target protein, the calmodulin further changes its conformation, the nature of which depends on the specific target protein (Figure 15–33B). Among the many targets calmodulin regulates are enzymes and membrane transport proteins. As one example,  $\text{Ca}^{2+}$ /calmodulin binds to and activates the plasma membrane  $\text{Ca}^{2+}$ -pump that uses ATP hydrolysis to pump  $\text{Ca}^{2+}$  out of cells. Thus, whenever the concentration of  $\text{Ca}^{2+}$  in the cytosol rises, the pump is activated, which helps to return the cytosolic  $\text{Ca}^{2+}$  level to resting levels.

Many effects of  $\text{Ca}^{2+}$ , however, are more indirect and are mediated by protein phosphorylations catalyzed by a family of protein kinases called  **$\text{Ca}^{2+}$ /calmodulin-dependent kinases (CaM-kinases)**. Some CaM-kinases phosphorylate transcription regulators, such as the CREB protein (see Figure 15–27), and in this way activate or inhibit the transcription of specific genes.

One of the best-studied CaM-kinases is **CaM-kinase II**, which is found in most animal cells but is especially enriched in the nervous system. It constitutes up to 2% of the total protein mass in some regions of the brain, and it is highly concentrated in synapses. CaM-kinase II has several remarkable properties. To begin with, it has a spectacular quaternary structure: twelve copies of the enzyme are assembled into a stacked pair of rings, with kinase domains on the outside linked to a central hub (Figure 15–34). This structure helps the enzyme function as a molecular memory device, switching to an active state when exposed to  $\text{Ca}^{2+}$ /calmodulin and then remaining active even after the  $\text{Ca}^{2+}$  signal has decayed. This is because adjacent kinase subunits can phosphorylate each other (a process called *autophosphorylation*) when  $\text{Ca}^{2+}$ /calmodulin activates them (Figure 15–34). Once a kinase subunit is autophosphorylated, it remains active even in the absence of  $\text{Ca}^{2+}$ , thereby prolonging the duration of the kinase activity beyond that of the initial activating  $\text{Ca}^{2+}$  signal. The enzyme maintains this activity until a protein phosphatase removes the autophosphorylation and shuts the kinase off. CaM-kinase II activation can thereby serve as a memory trace of a prior  $\text{Ca}^{2+}$  pulse, and it seems to have a role in some types of memory and learning in the vertebrate nervous system. Mutant mice that lack a brain-specific form of the enzyme have specific defects in their ability to remember where things are.

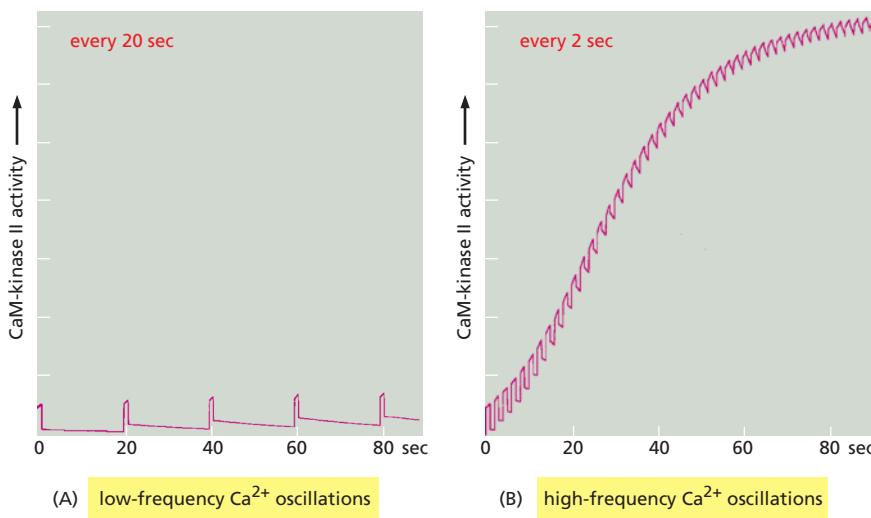
**Figure 15–33 The structure of  $\text{Ca}^{2+}$ /calmodulin.** (A) The molecule has a dumbbell shape, with two globular ends, which can bind to many target proteins. The globular ends are connected by a long, exposed  $\alpha$  helix, which allows the protein to adopt a number of different conformations, depending on the target protein it interacts with. Each globular head has two  $\text{Ca}^{2+}$ -binding sites (Movie 15.6). (B) Shown is the major structural change that occurs in  $\text{Ca}^{2+}$ /calmodulin when it binds to a target protein (in this example, a peptide that consists of the  $\text{Ca}^{2+}$ /calmodulin-binding domain of a  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase). Note that the  $\text{Ca}^{2+}$ /calmodulin has “jack-knifed” to surround the peptide. When it binds to other targets, it can adopt different conformations. (A, based on x-ray crystallographic data from Y.S. Babu et al., *Nature* 315:37–40, 1985. With permission from Macmillan Publishers Ltd; B, based on x-ray crystallographic data from W.E. Meador, A.R. Means, and F.A. Quirocho, *Science* 257:1251–1255, 1992, and on nuclear magnetic resonance (NMR) spectroscopy data from M. Ikura et al., *Science* 256:632–638, 1992.)



**Figure 15–34** The stepwise activation of CaM-kinase II. (A) Each CaM-kinase II protein has two major domains: an amino-terminal kinase domain (green) and a carboxyl-terminal hub domain (blue), linked by a regulatory segment. Six CaM-kinase II proteins are assembled into a giant ring in which the hub domains interact tightly to produce a central structure that is surrounded by kinase domains. The complete enzyme contains two stacked rings, for a total of 12 kinase proteins, but only one ring is shown here for clarity. When the enzyme is inactive, the ring exists in a dynamic equilibrium between two states. The first (*upper left*) is a compact state, in which the kinase domains interact with the hub, so that the regulatory segment is buried in the kinase active site and thereby blocks catalytic activity. In the second inactive state (*upper middle*), a kinase domain has popped out and is linked to the central hub by its regulatory segment, which continues to inhibit the kinase but is now accessible to  $\text{Ca}^{2+}$ /calmodulin. If present,  $\text{Ca}^{2+}$ /calmodulin will bind the regulatory segment and prevent it from inhibiting the kinase, thereby locking the kinase in an active state (*upper right*). If the adjacent kinase subunit also pops out from the hub, it will also be activated by  $\text{Ca}^{2+}$ /calmodulin, and the two kinases will then phosphorylate each other on their regulatory segments (*lower right*). This autophosphorylation further activates the enzyme. It also prolongs the activity of the enzyme in two ways. First, it traps the bound  $\text{Ca}^{2+}$ /calmodulin so that it does not dissociate from the enzyme until cytosolic  $\text{Ca}^{2+}$  levels return to basal values for at least 10 seconds (not shown). Second, it converts the enzyme to a  $\text{Ca}^{2+}$ -independent form, so that the kinase remains active even after the  $\text{Ca}^{2+}$ /calmodulin dissociates from it (*lower left*). This activity continues until the action of a protein phosphatase overrides the autophosphorylation activity of CaM-kinase II. (B) This structural model of the enzyme is based on x-ray crystallographic analysis.

The remarkable dodecameric structure of the enzyme allows it to achieve a broad range of intermediate activity states in response to different  $\text{Ca}^{2+}$  oscillation frequencies: higher frequencies tend to cause more subunits in the enzyme to reach the phosphorylated active state (see Figure 15–35). The behavior of CaM-kinase II is also controlled by the length of the linker segment between the kinase and hub domains. The linker is longer in some isoforms of the enzyme; in these isoforms, the kinase domains tend to pop out of the ring more frequently, making it more sensitive to  $\text{Ca}^{2+}$ . These and other mechanisms allow the cell to tailor the responsiveness of the enzyme to the needs of different types of neurons. (Adapted from L.H. Chao et al., *Cell* 146:732–745, 2011. PDB code: 3SOA.)

Another remarkable property of CaM-kinase II is that the enzyme can use its intrinsic memory mechanism to decode the frequency of  $\text{Ca}^{2+}$  oscillations. This property is thought to be especially important at a nerve cell synapse, where changes in intracellular  $\text{Ca}^{2+}$  levels in a postsynaptic cell as a result of neural activity can lead to long-term changes in the subsequent effectiveness of that synapse



**Figure 15–35** CaM-kinase II as a frequency decoder of  $\text{Ca}^{2+}$  oscillations. (A) At low frequencies of  $\text{Ca}^{2+}$  spikes, the enzyme becomes inactive after each spike, as the autophosphorylation induced by  $\text{Ca}^{2+}$ /calmodulin binding does not maintain the enzyme's activity long enough for the enzyme to remain active until the next  $\text{Ca}^{2+}$  spike arrives. (B) At higher spike frequencies, however, the enzyme fails to inactivate completely between  $\text{Ca}^{2+}$  spikes, so its activity ratchets up with each spike. If the spike frequency is high enough, this progressive increase in enzyme activity will continue until the enzyme is autophosphorylated on all subunits and is therefore maximally activated. Although not shown, once enough of its subunits are autophosphorylated, the enzyme can be maintained in a highly active state even with a relatively low frequency of  $\text{Ca}^{2+}$  spikes (a form of cell memory). The binding of  $\text{Ca}^{2+}$ /calmodulin to the enzyme is enhanced by the CaM-kinase II autophosphorylation (an additional form of positive feedback), helping to generate a more switchlike response to repeated  $\text{Ca}^{2+}$  spikes. (From P.I. Hanson, T. Meyer, L. Stryer, and H. Schulman, *Neuron* 12:943–956, 1994. With permission from Elsevier.)

(discussed in Chapter 11). When CaM-kinase II is exposed to both a protein phosphatase and repetitive pulses of  $\text{Ca}^{2+}$ /calmodulin at different frequencies that mimic those observed in stimulated cells, the enzyme's activity increases steeply as a function of pulse frequency (Figure 15–35).

### Some G Proteins Directly Regulate Ion Channels

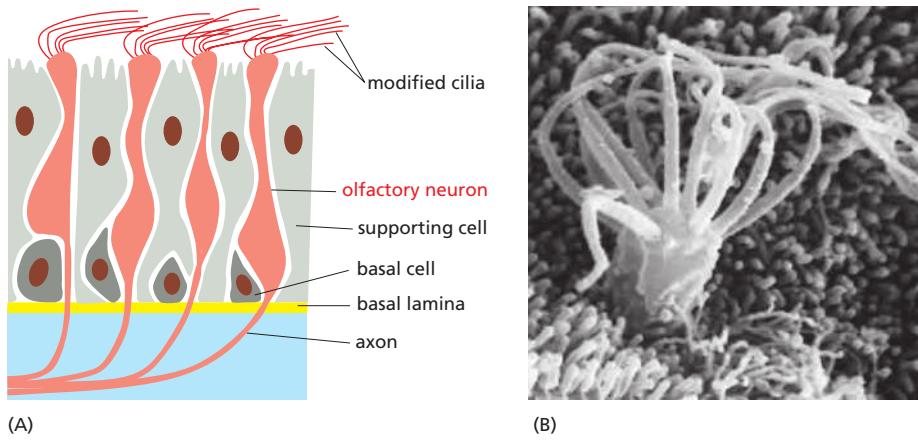
G proteins do not act exclusively by regulating the activity of membrane-bound enzymes that alter the concentration of cyclic AMP or  $\text{Ca}^{2+}$  in the cytosol. The  $\alpha$  subunit of one type of G protein (called  $G_{12}$ ), for example, activates a guanine nucleotide exchange factor (GEF) that activates a monomeric GTPase of the *Rho family* (discussed later and in Chapter 16), which regulates the actin cytoskeleton.

In some other cases, G proteins directly activate or inactivate ion channels in the plasma membrane of the target cell, thereby altering the ion permeability—and hence the electrical excitability—of the membrane. As an example, acetylcholine released by the vagus nerve reduces the heart rate (see Figure 15–5B). This effect is mediated by a special class of acetylcholine receptors that activate the  $G_i$  protein discussed earlier. Once activated, the  $\alpha$  subunit of  $G_i$  inhibits adenylyl cyclase (as described previously), while the  $\beta\gamma$  subunits bind to  $K^+$  channels in the heart muscle cell plasma membrane and open them. The opening of these  $K^+$  channels makes it harder to depolarize the cell and thereby contributes to the inhibitory effect of acetylcholine on the heart. (These acetylcholine receptors, which can be activated by the fungal alkaloid muscarine, are called *muscarinic acetylcholine receptors* to distinguish them from the very different *nicotinic acetylcholine receptors*, which are ion-channel-coupled receptors on skeletal muscle and nerve cells that can be activated by the binding of nicotine, as well as by acetylcholine.)

Other G proteins regulate the activity of ion channels less directly, either by stimulating channel phosphorylation (by PKA, PKC, or CaM-kinase, for example) or by causing the production or destruction of cyclic nucleotides that directly activate or inactivate ion channels. These *cyclic-nucleotide-gated ion channels* have a crucial role in both smell (olfaction) and vision, as we now discuss.

### Smell and Vision Depend on GPCRs That Regulate Ion Channels

Humans can distinguish more than 10,000 distinct smells, which they detect using specialized olfactory receptor neurons in the lining of the nose. These cells use specific GPCRs called **olfactory receptors** to recognize odors; the receptors are displayed on the surface of the modified cilia that extend from each cell (Figure 15–36). The receptors act through cAMP. When stimulated by odorant binding,



**Figure 15-36 Olfactory receptor neurons.** (A) A section of olfactory epithelium in the nose. Olfactory receptor neurons possess modified cilia, which project from the surface of the epithelium and contain the olfactory receptors, as well as the signal transduction machinery. The axon, which extends from the opposite end of the receptor neuron, conveys electrical signals to the brain when an odorant activates the cell to produce an action potential. In rodents, at least, the basal cells act as stem cells, producing new receptor neurons throughout life, to replace the neurons that die. (B) A scanning electron micrograph of the cilia on the surface of an olfactory neuron. (B, from E.E. Morrison and R.M. Costanzo, *J. Comp. Neurol.* 297:1–13, 1990. With permission from Wiley-Liss.)

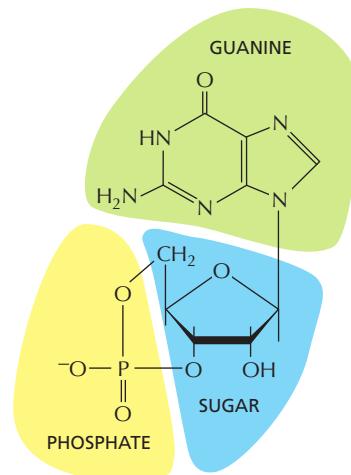
they activate an olfactory-specific G protein (known as  $G_{olf}$ ), which in turn activates adenylyl cyclase. The resulting increase in cAMP opens cyclic-AMP-gated cation channels, thereby allowing an influx of  $\text{Na}^+$ , which depolarizes the olfactory receptor neuron and initiates a nerve impulse that travels along its axon to the brain.

There are about 1000 different olfactory receptors in a mouse and about 350 in a human, each encoded by a different gene and each recognizing a different set of odorants. Each olfactory receptor neuron produces only one of these receptors; the neuron responds to a specific set of odorants by means of the specific receptor it displays, and each odorant activates its own characteristic set of olfactory receptor neurons. The same receptor also helps direct the elongating axon of each developing olfactory neuron to the specific target neurons that it will connect to in the brain. A different set of GPCRs acts in a similar way in some vertebrates to mediate responses to *pheromones*, chemical signals detected in a different part of the nose that are used in communication between members of the same species. Humans, however, are thought to lack functional pheromone receptors.

Vertebrate vision employs a similarly elaborate, highly sensitive, signal-detection process. Cyclic-nucleotide-gated ion channels are also involved, but the crucial cyclic nucleotide is **cyclic GMP** (Figure 15-37) rather than cAMP. As with cAMP, a continuous rapid synthesis (by *guanylyl cyclase*) and rapid degradation (by *cyclic GMP phosphodiesterase*) controls the concentration of cyclic GMP in the cytosol.

In visual transduction responses, which are the fastest G-protein-mediated responses known in vertebrates, the receptor activation stimulated by light causes a fall rather than a rise in the level of the cyclic nucleotide. The pathway has been especially well studied in **rod photoreceptors** (rods) in the vertebrate retina. Rods are responsible for noncolor vision in dim light, whereas **cone photoreceptors** (cones) are responsible for color vision in bright light. A rod photoreceptor is a highly specialized cell with outer and inner segments, a cell body, and a synaptic region where the rod passes a chemical signal to a retinal nerve cell (Figure 15-38). This nerve cell relays the signal to another nerve cell in the retina, which in turn relays it to the brain.

The phototransduction apparatus is in the outer segment of the rod, which contains a stack of *discs*, each formed by a closed sac of membrane that is densely packed with photosensitive **rhodopsin** molecules. The plasma membrane surrounding the outer segment contains *cyclic-GMP-gated cation channels*. Cyclic GMP bound to these channels keeps them open in the dark. Paradoxically, light causes a hyperpolarization (which inhibits synaptic signaling) rather than a depolarization of the plasma membrane (which would stimulate synaptic signaling). Hyperpolarization (that is, the membrane potential moves to a more negative value—discussed in Chapter 11) results because the light-induced activation of rhodopsin molecules in the disc membrane decreases the cyclic GMP concentration and closes the cation channels in the surrounding plasma membrane (Figure 15-39).



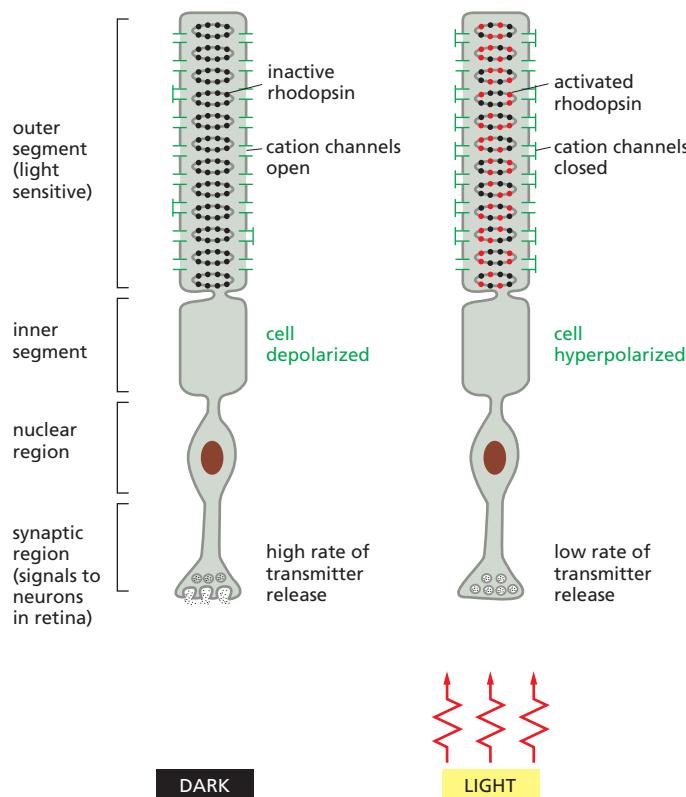
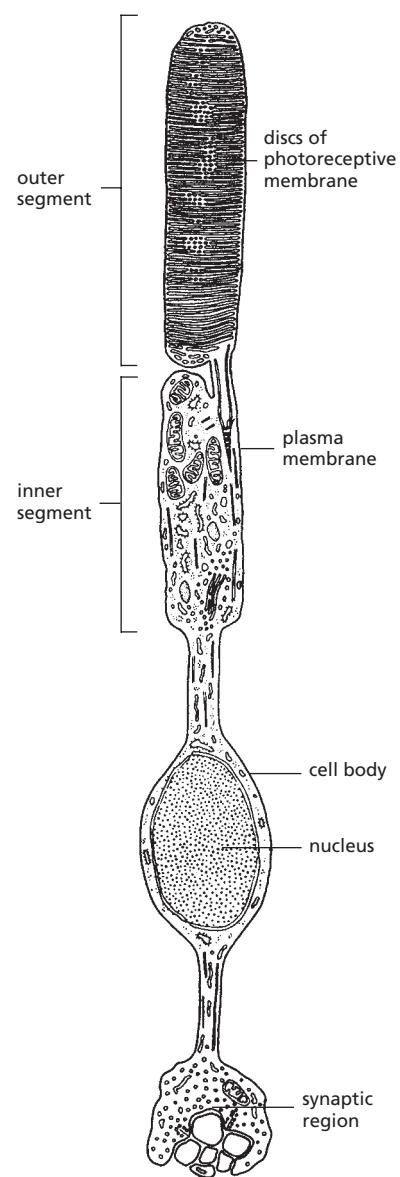
**Figure 15-37 Cyclic GMP.**

**Figure 15–38 A rod photoreceptor cell.** There are about 1000 discs in the outer segment. The disc membranes are not connected to the plasma membrane. The inner and outer segments are specialized parts of a *primary cilium* (discussed in Chapter 16). A primary cilium extends from the surface of most vertebrate cells, where it serves as a signaling organelle.

Rhodopsin is a member of the GPCR family, but the activating extracellular signal is not a molecule but a photon of light. Each rhodopsin molecule contains a covalently attached chromophore, 11-*cis* retinal, which isomerizes almost instantaneously to all-*trans* retinal when it absorbs a single photon. The isomerization alters the shape of the retinal, forcing a conformational change in the protein (opsin). The activated rhodopsin molecule then alters the conformation of the G protein *transducin* ( $G_t$ ), causing the transducin  $\alpha$  subunit to activate **cyclic GMP phosphodiesterase**. The phosphodiesterase then hydrolyzes cyclic GMP, so that cyclic GMP levels in the cytosol fall. This drop in cyclic GMP concentration decreases the amount of cyclic GMP bound to the plasma membrane cation channels, allowing more of these cyclic-GMP-sensitive channels to close. In this way, the signal quickly passes from the disc membrane to the plasma membrane, and a light signal is converted into an electrical one, through a hyperpolarization of the rod cell plasma membrane.

Rods use several negative feedback loops to allow the cells to revert quickly to a resting, dark state in the aftermath of a flash of light—a requirement for perceiving the shortness of the flash. A rhodopsin-specific protein kinase called *rhodopsin kinase* ( $RK$ ) phosphorylates the cytosolic tail of activated rhodopsin on multiple serines, partially inhibiting the ability of the rhodopsin to activate transducin. An inhibitory protein called *arrestin* (discussed later) then binds to the phosphorylated rhodopsin, further inhibiting rhodopsin's activity. Mice or humans with a mutation that inactivates the gene encoding RK have a prolonged light response.

At the same time as arrestin shuts off rhodopsin, an RGS protein (discussed earlier) binds to activated transducin, stimulating the transducin to hydrolyze its bound GTP to GDP, which returns transducin to its inactive state. In addition, the cation channels that close in response to light are permeable to  $\text{Ca}^{2+}$ , as well as



**Figure 15–39 The response of a rod photoreceptor cell to light.** Rhodopsin molecules in the outer-segment discs absorb photons. Photon absorption closes cation channels in the plasma membrane, which hyperpolarizes the membrane and reduces the rate of neurotransmitter release from the synaptic region. Because the neurotransmitter inhibits many of the postsynaptic retinal neurons, illumination serves to free the neurons from inhibition and thus, in effect, excites them. The neural connections of the retina lie between the light source and the outer segment, and so the light must pass through the synapses and rod cell nucleus to reach the light sensors.

to  $\text{Na}^+$ , so that when they close, the normal influx of  $\text{Ca}^{2+}$  is inhibited, causing the  $\text{Ca}^{2+}$  concentration in the cytosol to fall. The decrease in  $\text{Ca}^{2+}$  concentration stimulates guanylyl cyclase to replenish the cyclic GMP, rapidly returning its level to where it was before the light was switched on. A specific  $\text{Ca}^{2+}$ -sensitive protein mediates the activation of guanylyl cyclase in response to the fall in  $\text{Ca}^{2+}$  levels. In contrast to calmodulin, this protein is inactive when  $\text{Ca}^{2+}$  is bound to it and active when it is  $\text{Ca}^{2+}$ -free. It therefore stimulates the cyclase when  $\text{Ca}^{2+}$  levels fall following a light response.

Negative feedback mechanisms do more than just return the rod to its resting state after a transient light flash; they also help the rod to *adapt*, stepping down the response when the rod is exposed to light continuously. Adaptation, as we discussed earlier, allows the receptor cell to function as a sensitive detector of *changes* in stimulus intensity over an enormously wide range of baseline levels of stimulation. It is why we can see faint stars in a dark sky, or a camera flash in bright sunlight.

The various trimeric G proteins we have discussed in this chapter fall into four major families, as summarized in **Table 15–3**.

### Nitric Oxide Is a Gaseous Signaling Mediator That Passes Between Cells

Signaling molecules like cyclic nucleotides and calcium are hydrophilic small molecules that generally act within the cell where they are produced. Some signaling molecules, however, are hydrophobic enough, small enough, or both, to pass readily across the plasma membrane and carry signals to nearby cells. An important and remarkable example is the gas **nitric oxide (NO)**, which acts as a signal molecule in many tissues of both animals and plants.

In mammals, one of NO's many functions is to relax smooth muscle in the walls of blood vessels. The neurotransmitter acetylcholine stimulates NO synthesis by

TABLE 15–3 Four Major Families of Trimeric G Proteins*			
Family	Some family members	Subunits that mediate action	Some functions
I	$G_s$	$\alpha$	Activates adenylyl cyclase; activates $\text{Ca}^{2+}$ channels
	$G_{olf}$	$\alpha$	Activates adenylyl cyclase in olfactory sensory neurons
II	$G_i$	$\alpha$	Inhibits adenylyl cyclase
		$\beta\gamma$	Activates $\text{K}^+$ channels
	$G_o$	$\beta\gamma$	Activates $\text{K}^+$ channels; inactivates $\text{Ca}^{2+}$ channels
		$\alpha$ and $\beta\gamma$	Activates phospholipase C- $\beta$
	$G_t$ (transducin)	$\alpha$	Activates cyclic GMP phosphodiesterase in vertebrate rod photoreceptors
III	$G_q$	$\alpha$	Activates phospholipase C- $\beta$
IV	$G_{12/13}$	$\alpha$	Activates Rho family monomeric GTPases (via Rho-GEF) to regulate the actin cytoskeleton

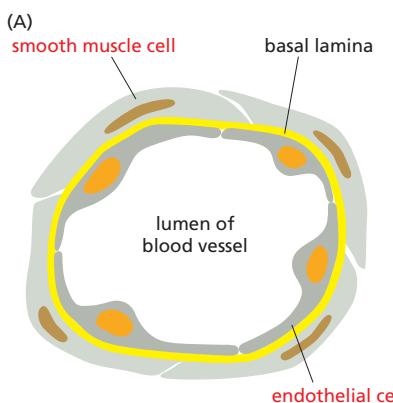
\*Families are determined by amino acid sequence relatedness of the  $\alpha$  subunits. Only selected examples are included. About 20  $\alpha$  subunits and at least 6  $\beta$  subunits and 11  $\gamma$  subunits have been described in humans.

activating a GPCR on the membranes of the endothelial cells that line the interior of the vessel. The activated receptor triggers IP<sub>3</sub> synthesis and Ca<sup>2+</sup> release (see Figure 15–29), leading to stimulation of an enzyme that synthesizes NO. Because dissolved NO passes readily across membranes, it diffuses out of the cell where it is produced and into neighboring smooth muscle cells, where it causes muscle relaxation and thereby vessel dilation (Figure 15–40). It acts only locally because it has a short half-life—about 5–10 seconds—in the extracellular space before oxygen and water convert it to nitrates and nitrites.

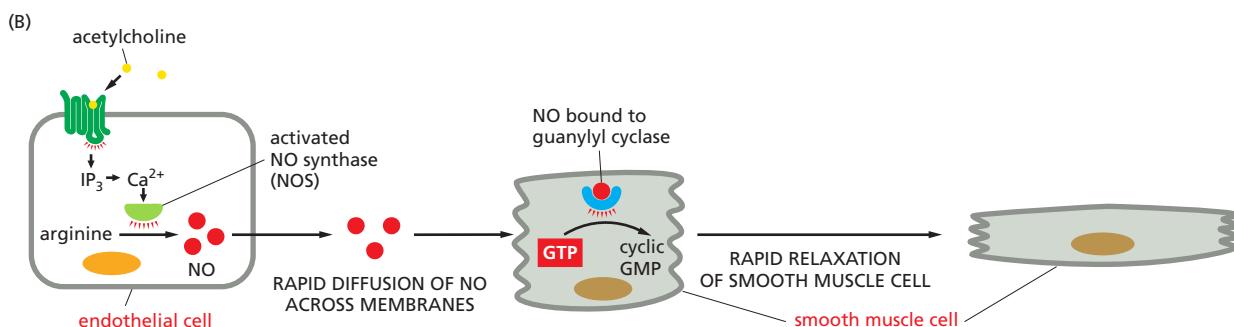
The effect of NO on blood vessels provides an explanation for the mechanism of action of nitroglycerine, which has been used for about 100 years to treat patients with angina (pain resulting from inadequate blood flow to the heart muscle). The nitroglycerine is converted to NO, which relaxes blood vessels. This reduces the workload on the heart and, as a consequence, reduces the oxygen requirement of the heart muscle.

NO is made by the deamination of the amino acid arginine, catalyzed by enzymes called **NO synthases (NOS)** (see Figure 15–40). The NOS in endothelial cells is called *eNOS*, while that in nerve and muscle cells is called *nNOS*. Both eNOS and nNOS are stimulated by Ca<sup>2+</sup>. Macrophages, by contrast, make yet another NOS, called inducible NOS (*iNOS*), that is constitutively active but synthesized only when the cells are activated, usually in response to an infection.

In some target cells, including smooth muscle cells, NO binds reversibly to iron in the active site of guanylyl cyclase, stimulating synthesis of cyclic GMP. NO can increase cyclic GMP in the cytosol within seconds, because the normal rate of turnover of cyclic GMP is high: rapid degradation to GMP by a phosphodiesterase constantly balances the production of cyclic GMP by guanylyl cyclase. The drug Viagra® and its relatives inhibit the cyclic GMP phosphodiesterase in the penis, thereby increasing the amount of time that cyclic GMP levels remain elevated in the smooth muscle cells of penile blood vessels after NO production is induced by local nerve terminals. The cyclic GMP, in turn, keeps the blood vessels relaxed and thereby the penis erect. NO can also signal cells independently of cyclic GMP. It can, for example, alter the activity of an intracellular protein by covalently nitrosylating thiol (-SH) groups on specific cysteines in the protein.



**Figure 15–40 The role of nitric oxide (NO) in smooth muscle relaxation in a blood vessel wall.** (A) Simplified cross section of a blood vessel, showing the endothelial cells lining the lumen and the smooth muscle cells around them. (B) The neurotransmitter acetylcholine stimulates blood vessel dilation by activating a GPCR—the *muscarinic acetylcholine receptor*—on the surface of endothelial cells. This receptor activates a G protein, G<sub>q</sub>, thereby stimulating IP<sub>3</sub> synthesis and Ca<sup>2+</sup> release by the mechanisms illustrated in Figure 15–29. Ca<sup>2+</sup> activates nitric oxide synthase, causing the endothelial cells to produce NO from arginine. The NO diffuses out of the endothelial cells and into the neighboring smooth muscle cells, where it activates guanylyl cyclase to produce cyclic GMP. The cyclic GMP triggers a response that causes the smooth muscle cells to relax, increasing blood flow through the vessel.



## Second Messengers and Enzymatic Cascades Amplify Signals

Despite the differences in molecular details, the different intracellular signaling pathways that GPCRs trigger share certain features and obey similar general principles. They depend on relay chains of intracellular signaling proteins and second messengers. These relay chains provide numerous opportunities for amplifying the responses to extracellular signals. In the visual transduction cascade, for example, a single activated rhodopsin molecule catalyzes the activation of hundreds of molecules of transducin at a rate of about 1000 transducin molecules per second. Each activated transducin molecule activates a molecule of cyclic GMP phosphodiesterase, each of which hydrolyzes about 4000 molecules of cyclic GMP per second. This catalytic cascade lasts for about 1 second and results in the hydrolysis of more than  $10^5$  cyclic GMP molecules for a single quantum of light absorbed, and the resulting drop in the concentration of cyclic GMP in turn transiently closes hundreds of cation channels in the plasma membrane (Figure 15–41). As a result, a rod cell can respond to even a single photon of light in a way that is highly reproducible in its timing and magnitude.

Likewise, when an extracellular signal molecule binds to a receptor that indirectly activates adenylyl cyclase via  $G_s$ , each receptor protein may activate many molecules of  $G_s$  protein, each of which can activate a cyclase molecule. Each cyclase molecule, in turn, can catalyze the conversion of a large number of ATP molecules to cAMP molecules. A similar amplification operates in the IP<sub>3</sub> signaling pathway. In these ways, a nanomolar ( $10^{-9}$  M) change in the concentration of an extracellular signal can induce micromolar ( $10^{-6}$  M) changes in the concentration of a second messenger such as cAMP or Ca<sup>2+</sup>. Because these messengers function as allosteric effectors to activate specific enzymes or ion channels, a single extracellular signal molecule can alter many thousands of protein molecules within the target cell.

Any such amplifying cascade of stimulatory signals requires counterbalancing mechanisms at every step of the cascade to restore the system to its resting state when stimulation ceases. As emphasized earlier, the response to stimulation can be rapid only if the inactivating mechanisms are also rapid. Cells therefore have efficient mechanisms for rapidly degrading (and resynthesizing) cyclic nucleotides and for buffering and removing cytosolic Ca<sup>2+</sup>, as well as for inactivating the responding enzymes and ion channels once they have been activated. This is not only essential for turning a response off, but is also important for defining the resting state from which a response begins.

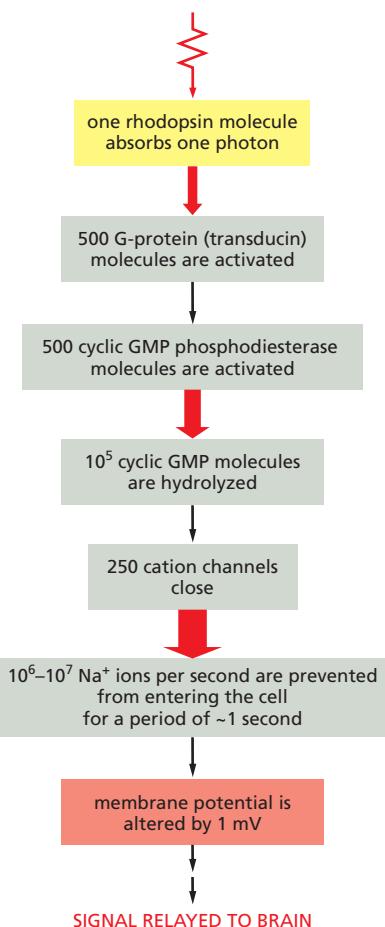
Each protein in the signaling relay chain can be a separate target for regulation, including the receptor itself, as we discuss next.

## GPCR Desensitization Depends on Receptor Phosphorylation

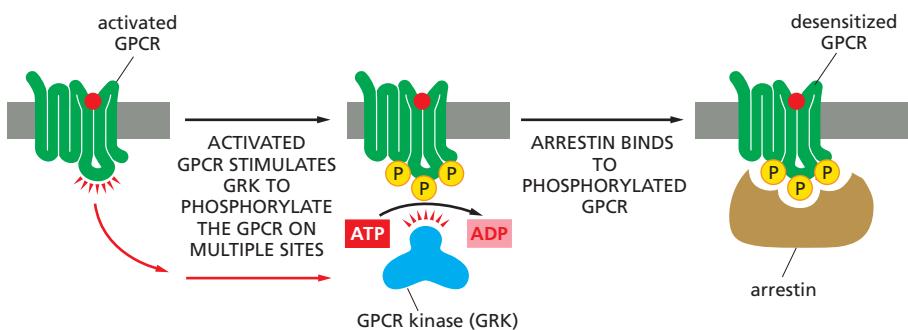
As discussed earlier, when target cells are exposed to a high concentration of a stimulating ligand for a prolonged period, they can become *desensitized*, or *adapted*, in several different ways. An important class of adaptation mechanisms depends on alteration of the quantity or condition of the receptor molecules themselves.

For GPCRs, there are three general modes of adaptation (see Figure 15–20): (1) In *receptor sequestration*, they are temporarily moved to the interior of the cell (internalized) so that they no longer have access to their ligand. (2) In *receptor down-regulation*, they are destroyed in lysosomes after internalization. (3) In *receptor inactivation*, they become altered so that they can no longer interact with G proteins.

In each case, the desensitization of the GPCRs depends on their phosphorylation by PKA, PKC, or a member of the family of **GPCR kinases (GRKs)**, which includes the rhodopsin-specific kinase RK involved in rod photoreceptor desensitization discussed earlier. The GRKs phosphorylate multiple serines and threonines on a GPCR, but they do so only after ligand binding has activated the receptor, because it is the activated receptor that allosterically activates the GRK.



**Figure 15–41** Amplification in the light-induced catalytic cascade in vertebrate rods. The red arrows indicate the steps where amplification occurs, with the thickness of the arrow roughly indicating the magnitude of the amplification.



**Figure 15–42** The roles of GPCR kinases (GRKs) and arrestins in GPCR desensitization.

A GRK phosphorylates only activated receptors because it is the activated GPCR that activates the GRK. The binding of an arrestin to the phosphorylated receptor prevents the receptor from binding to its G protein and also directs its endocytosis (not shown). Mice that are deficient in one form of arrestin fail to desensitize in response to morphine, for example, attesting to the importance of arrestins for desensitization.

As with rhodopsin, once a receptor has been phosphorylated by a GRK, it binds with high affinity to a member of the **arrestin** family of proteins (Figure 15–42).

The bound arrestin can contribute to the desensitization process in at least two ways. First, it prevents the activated receptor from interacting with G proteins. Second, it serves as an adaptor protein to help couple the receptor to the clathrin-dependent endocytosis machinery (discussed in Chapter 13), inducing receptor-mediated endocytosis. The fate of the internalized GPCR–arrestin complex depends on other proteins in the complex. In some cases, the receptor is dephosphorylated and recycled back to the plasma membrane for reuse. In others, it is ubiquitylated, endocytosed, and degraded in lysosomes (discussed later).

Receptor endocytosis does not necessarily stop the receptor from signaling. In some cases, the bound arrestin recruits other signaling proteins to relay the signal onward from the internalized GPCRs along new pathways.

## Summary

GPCRs can indirectly activate or inactivate either plasma-membrane-bound enzymes or ion channels via G proteins. When an activated receptor stimulates a G protein, the G protein undergoes a conformational change that activates its  $\alpha$  subunit, thereby triggering release of a  $\beta\gamma$  complex. Either component can then directly regulate the activity of target proteins in the plasma membrane. Some GPCRs either activate or inactivate adenylyl cyclase, thereby altering the intracellular concentration of the second messenger cyclic AMP. Others activate a phosphoinositide-specific phospholipase C (PLC $\beta$ ), which generates two second messengers. One is inositol 1,4,5-trisphosphate (IP<sub>3</sub>), which releases Ca<sup>2+</sup> from the ER and thereby increases the concentration of Ca<sup>2+</sup> in the cytosol. The other is diacylglycerol, which remains in the plasma membrane and helps activate protein kinase C (PKC). An increase in cytosolic cyclic AMP or Ca<sup>2+</sup> levels affects cells mainly by stimulating cAMP-dependent protein kinase (PKA) and Ca<sup>2+</sup>/calmodulin-dependent protein kinases (CaM-kinases), respectively.

PKC, PKA, and CaM-kinases phosphorylate specific target proteins and thereby alter the activity of the proteins. Each type of cell has its own characteristic set of target proteins that is regulated in these ways, enabling the cell to make its own distinctive response to the second messengers. The intracellular signaling cascades activated by GPCRs greatly amplify the responses, so that many thousands of target protein molecules are changed for each molecule of extracellular signaling ligand bound to its receptor. The responses mediated by GPCRs are rapidly turned off when the extracellular signal is removed, and activated GPCRs are inactivated by phosphorylation and association with arrestins.

## SIGNALING THROUGH ENZYME-COUPLED RECEPTORS

Like GPCRs, **enzyme-coupled receptors** are transmembrane proteins with their ligand-binding domain on the outer surface of the plasma membrane. Instead of having a cytosolic domain that associates with a trimeric G protein, however, their cytosolic domain either has intrinsic enzyme activity or associates directly with an enzyme. Whereas a GPCR has seven transmembrane segments, each subunit of an enzyme-coupled receptor typically has only one. GPCRs and enzyme-coupled receptors often activate some of the same signaling pathways. In this section, we describe some of the important features of signaling by enzyme-coupled receptors, with an emphasis on the most common class of these proteins, the *receptor tyrosine kinases*.

### Activated Receptor Tyrosine Kinases (RTKs) Phosphorylate Themselves

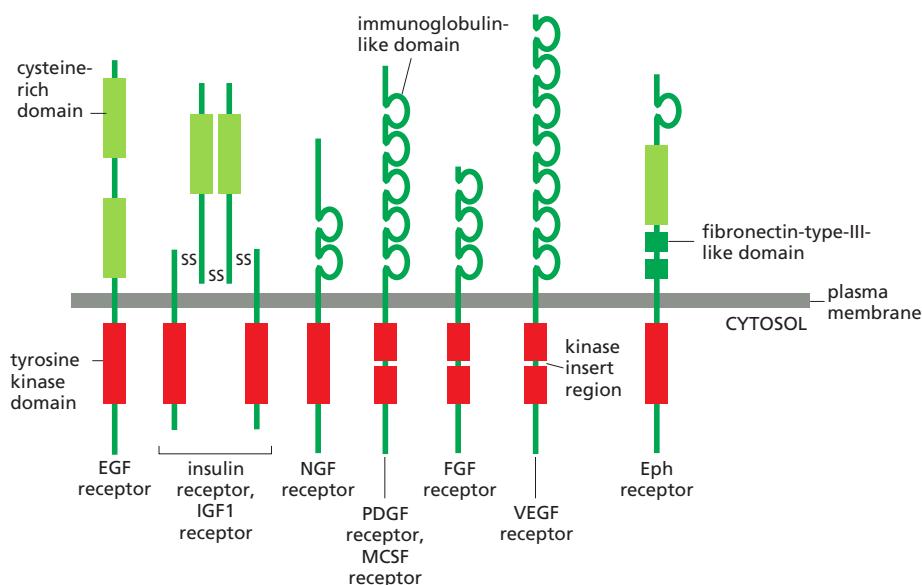
Many extracellular signal proteins act through **receptor tyrosine kinases (RTKs)**. These include many secreted and cell-surface-bound proteins that control cell behavior in developing and adult animals. Some of these signal proteins and their RTKs are listed in **Table 15–4**.

There are about 60 human RTKs, which can be classified into about 20 structural subfamilies, each dedicated to its complementary family of protein ligands. **Figure 15–43** shows the basic structural features of a number of the families that operate in mammals. In all cases, the binding of the signal protein to the ligand-binding domain on the extracellular side of the receptor activates the tyrosine kinase domain on the cytosolic side. This leads to phosphorylation of tyrosine side chains on the cytosolic part of the receptor, creating phosphotyrosine docking sites for various intracellular signaling proteins that relay the signal.

How does the binding of an extracellular ligand activate the kinase domain on the other side of the plasma membrane? For a GPCR, ligand binding is thought to change the relative orientation of several of the transmembrane  $\alpha$  helices, thereby shifting the position of the cytoplasmic loops relative to one another. It is unlikely, however, that a conformational change could propagate across the lipid bilayer

**TABLE 15–4** Some Signal Proteins That Act Via RTKs

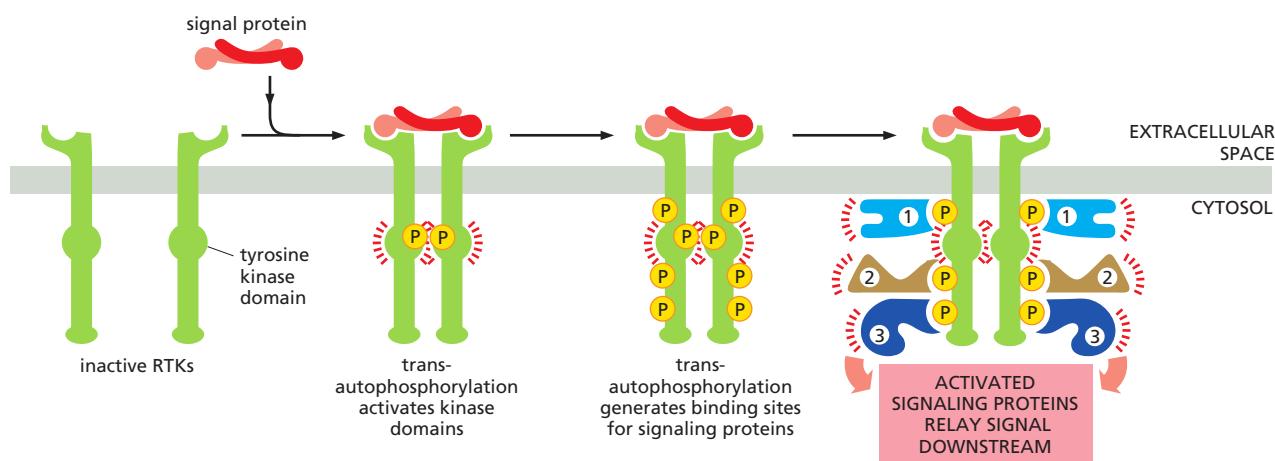
Signal protein family	Receptor family	Some representative responses
Epidermal growth factor (EGF)	EGF receptors	Stimulates cell survival, growth, proliferation, or differentiation of various cell types; acts as inductive signal in development
Insulin	Insulin receptor	Stimulates carbohydrate utilization and protein synthesis
Insulin-like growth factor (IGF1)	IGF receptor-1	Stimulates cell growth and survival in many cell types
Nerve growth factor (NGF)	Trk receptors	Stimulates survival and growth of some neurons
Platelet-derived growth factor (PDGF)	PDGF receptors	Stimulates survival, growth, proliferation, and migration of various cell types
Macrophage-colony-stimulating factor (M-CSF)	M-CSF receptor	Stimulates monocyte/macrophage proliferation and differentiation
Fibroblast growth factor (FGF)	FGF receptors	Stimulates proliferation of various cell types; inhibits differentiation of some precursor cells; acts as inductive signal in development
Vascular endothelial growth factor (VEGF)	VEGF receptors	Stimulates angiogenesis
Ephrin	Eph receptors	Stimulates angiogenesis; guides cell and axon migration



**Figure 15–43 Some subfamilies of RTKs.** Only one or two members of each subfamily are indicated. Note that in some cases, the tyrosine kinase domain is interrupted by a “kinase insert region” that is an extra segment emerging from the folded kinase domain. The functions of most of the cysteine-rich, immunoglobulin-like, and fibronectin-type-III-like domains are not known. Some of the ligands for the receptors shown are listed in Table 15–4, along with some representative responses that they mediate.

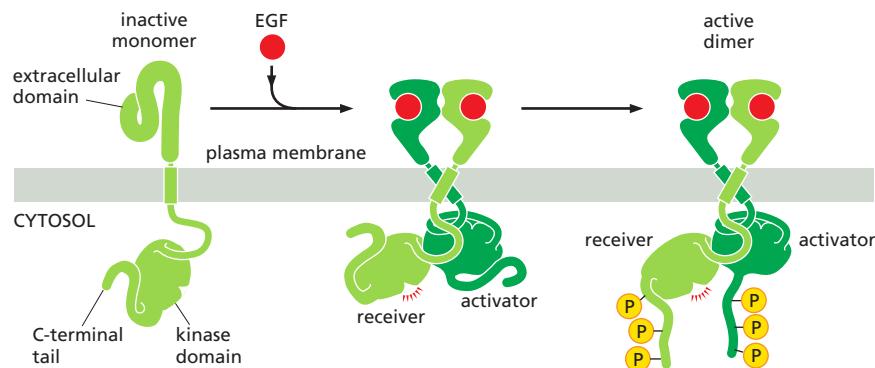
through a single transmembrane  $\alpha$  helix. Instead, for most RTKs, ligand binding causes the receptors to dimerize, bringing the two cytoplasmic kinase domains together and thereby promoting their activation (Figure 15–44).

Dimerization stimulates kinase activity by a variety of mechanisms. In many cases, such as the insulin receptor, dimerization simply brings the kinase domains close to each other in an orientation that allows them to phosphorylate each other on specific tyrosines in the kinase active sites, thereby promoting conformational changes that fully activate both kinase domains. In other cases, such as the receptor for *epidermal growth factor* (EGF), the kinase is not activated by phosphorylation but by conformational changes brought about by interactions between the two kinase domains outside their active sites (Figure 15–45).



**Figure 15–44 Activation of RTKs by dimerization.** In the absence of extracellular signals, most RTKs exist as monomers in which the internal kinase domain is inactive. Binding of ligand brings two monomers together to form a dimer. In most cases, the close proximity in the dimer leads the two kinase domains to phosphorylate each other, which has two effects. First, phosphorylation at some tyrosines in the kinase domains promotes the complete activation of the domains. Second, phosphorylation at tyrosines in other parts of the receptors generates docking sites for intracellular signaling proteins, resulting in the formation of large signaling complexes that can then broadcast signals along multiple signaling pathways.

Mechanisms of dimerization vary widely among different RTK family members. In some cases, as shown here, the ligand itself is a dimer and brings two receptors together by binding them simultaneously. In other cases, a monomeric ligand can interact with two receptors simultaneously to bring them together, or two ligands can bind independently on two receptors to promote dimerization. In some RTKs—notably those in the insulin receptor family—the receptor is always a dimer (see Figure 15–43), and ligand binding causes a conformational change that brings the two internal kinase domains closer together. Although many RTKs are activated by transautophosphorylation as shown here, there are some important exceptions, including the EGF receptor illustrated in Figure 15–45.



**Figure 15–45 Activation of the EGF receptor kinase.** In the absence of ligand, the EGF receptor exists primarily as an inactive monomer. EGF binding results in a conformational change that promotes dimerization of the external domains. The receptor kinase domain, unlike that of many RTKs, is not activated by transautophosphorylation. Instead, dimerization orients the internal kinase domains into an asymmetric dimer, in which one kinase domain (the “activator”) pushes against the other kinase domain (the “receiver”), thereby causing an activating conformational change in the receiver. The active receiver domain then phosphorylates multiple tyrosines in the C-terminal tails of both receptors, generating docking sites for intracellular signaling proteins (see Figure 15–44).

### Phosphorylated Tyrosines on RTKs Serve as Docking Sites for Intracellular Signaling Proteins

Once the kinase domains of an RTK dimer are activated, they phosphorylate multiple additional sites in the cytosolic parts of the receptors, typically in disordered regions outside the kinase domain (see Figure 15–44). This phosphorylation creates high-affinity docking sites for intracellular signaling proteins. Each signaling protein binds to a particular phosphorylated site on the activated receptors because it contains a specific phosphotyrosine-binding domain that recognizes surrounding features of the polypeptide chain in addition to the phosphotyrosine.

Once bound to the activated RTK, a signaling protein may become phosphorylated on tyrosines and thereby activated. In many cases, however, the binding alone may be sufficient to activate the docked signaling protein, by either inducing a conformational change in the protein or simply bringing it near the protein that is next in the signaling pathway. Thus, receptor phosphorylation serves as a switch to trigger the assembly of an intracellular signaling complex, which can then relay the signal onward, often along multiple routes, to various destinations in the cell. Because different RTKs bind different combinations of these signaling proteins, they activate different responses.

Some RTKs use additional docking proteins to enlarge the signaling complex at activated receptors. Insulin and IGF1 receptor signaling, for example, depend on a specialized docking protein called *insulin receptor substrate 1* (*IRS1*). *IRS1* associates with phosphorylated tyrosines on the activated receptor and is then phosphorylated at multiple sites, thereby creating many more docking sites than could be accommodated on the receptor alone (see Figure 15–11).

### Proteins with SH2 Domains Bind to Phosphorylated Tyrosines

A whole menagerie of intracellular signaling proteins can bind to the phosphotyrosines on activated RTKs (or on docking proteins such as *IRS1*). They help to relay the signal onward, mainly through chains of protein–protein interactions mediated by modular *interaction domains*, as discussed earlier. Some of the docked proteins are enzymes, such as **phospholipase C-γ** (**PLC $\gamma$** ), which functions in the same way as phospholipase C-β—activating the inositol phospholipid signaling pathway discussed earlier in connection with GPCRs (see Figures 15–28 and 15–29). Through this pathway, RTKs can increase cytosolic Ca<sup>2+</sup> levels and activate PKC. Another enzyme that docks on these receptors is the cytoplasmic tyrosine kinase *Src*, which phosphorylates other signaling proteins on tyrosines. Yet another is **phosphoinositide 3-kinase** (**PI 3-kinase**), which phosphorylates lipids rather than proteins; as we discuss later, the phosphorylated lipids then serve as docking sites to attract various signaling proteins to the plasma membrane.

The intracellular signaling proteins that bind to phosphotyrosines have varied structures and functions. However, they usually share highly conserved phosphotyrosine-binding domains. These can be either **SH2 domains** (for *Src homology*

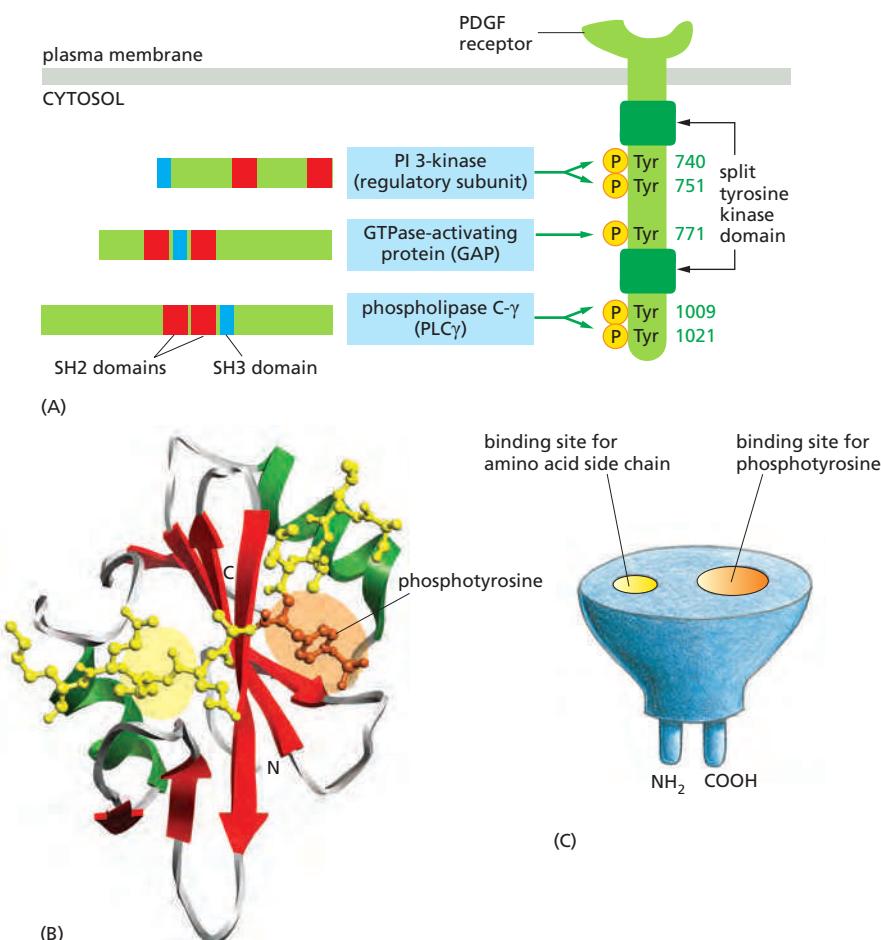
region) or, less commonly, *PTB domains* (for *phosphotyrosine-binding*). By recognizing specific phosphorylated tyrosines, these small interaction domains enable the proteins that contain them to bind to activated RTKs, as well as to many other intracellular signaling proteins that have been transiently phosphorylated on tyrosines (Figure 15–46). As discussed previously, many signaling proteins also contain other interaction domains that allow them to interact specifically with other proteins as part of the signaling process. These domains include the *SH3 domain*, which binds to proline-rich motifs in intracellular proteins (see Figure 15–11).

Not all proteins that bind to activated RTKs via SH2 domains help to relay the signal onward. Some act to decrease the signaling process, providing negative feedback. One example is the *c-Cbl protein*, which can dock on some activated receptors and catalyze their ubiquitylation, covalently adding one or more ubiquitin molecules to specific sites on the receptor. This promotes the endocytosis and degradation of the receptors in lysosomes—an example of receptor down-regulation (see Figure 15–20). Endocytic proteins that contain *ubiquitin-interaction motifs (UIMs)* recognize the ubiquitylated RTKs and direct them into clathrin-coated vesicles and, ultimately, into lysosomes (discussed in Chapter 13). Mutations that inactivate c-Cbl-dependent RTK down-regulation cause prolonged RTK signaling and thereby promote the development of cancer.

As is the case for GPCRs, ligand-induced endocytosis of RTKs does not always decrease signaling. In some cases, RTKs are endocytosed with their bound signaling proteins and continue to signal from endosomes or other intracellular compartments. This mechanism, for example, allows *nerve growth factor (NGF)* to bind to its specific RTK (called TrkA) at the end of a long nerve cell axon and signal to the cell body of the same cell a long distance away. Here, signaling endocytic

**Figure 15–46** The binding of SH2-containing intracellular signaling proteins to an activated RTK.

(A) This drawing of a receptor for *platelet-derived growth factor (PDGF)* shows five phosphotyrosine docking sites, three in the kinase insert region and two on the C-terminal tail, to which the three signaling proteins shown bind as indicated. The numbers on the right indicate the positions of the tyrosines in the polypeptide chain. These binding sites have been identified by using recombinant DNA technology to mutate specific tyrosines in the receptor. Mutation of tyrosines 1009 and 1021, for example, prevents the binding and activation of PLC $\gamma$ , so that receptor activation no longer stimulates the inositol phospholipid signaling pathway. The locations of the SH2 (red) and SH3 (blue) domains in the three signaling proteins are indicated. (Additional phosphotyrosine docking sites on this receptor are not shown, including those that bind the cytoplasmic tyrosine kinase Src and two adaptor proteins.) It is unclear how many signaling proteins can bind simultaneously to a single RTK. (B) The three-dimensional structure of an SH2 domain, as determined by x-ray crystallography. The binding pocket for phosphotyrosine is shaded in orange on the right, and a pocket for binding a specific amino acid side chain (isoleucine, in this case) is shaded in yellow on the left. The RTK polypeptide segment that binds the SH2 domain is shown in yellow (see also Figure 3–40). (C) The SH2 domain is a compact, “plug-in” module, which can be inserted almost anywhere in a protein without disturbing the protein’s folding or function (discussed in Chapter 3). Because each domain has distinct sites for recognizing phosphotyrosine and for recognizing a particular amino acid side chain, different SH2 domains recognize phosphotyrosine in the context of different flanking amino acid sequences. (B, based on data from G. Waksman et al., *Cell* 72:779–790, 1993. With permission from Elsevier. PDB code: 2SRC.)



vesicles containing TrkA, with NGF bound on the luminal side and signaling proteins docked on the cytosolic side, are transported along the axon to the cell body, where they signal the cell to survive.

Some signaling proteins are composed almost entirely of SH2 and SH3 domains and function as *adaptors* to couple tyrosine-phosphorylated proteins to other proteins that do not have their own SH2 domains (see Figure 15–11). Adapter proteins of this type help to couple activated RTKs to the important signaling protein *Ras*, a monomeric GTPase that, in turn, can activate various downstream signaling pathways, as we now discuss.

### The GTPase Ras Mediates Signaling by Most RTKs

The **Ras superfamily** consists of various families of monomeric GTPases, but only the Ras and Rho families relay signals from cell-surface receptors (Table 15–5). By interacting with different intracellular signaling proteins, a single Ras or Rho family member can coordinately spread the signal along several distinct downstream signaling pathways, thereby acting as a *signaling hub*.

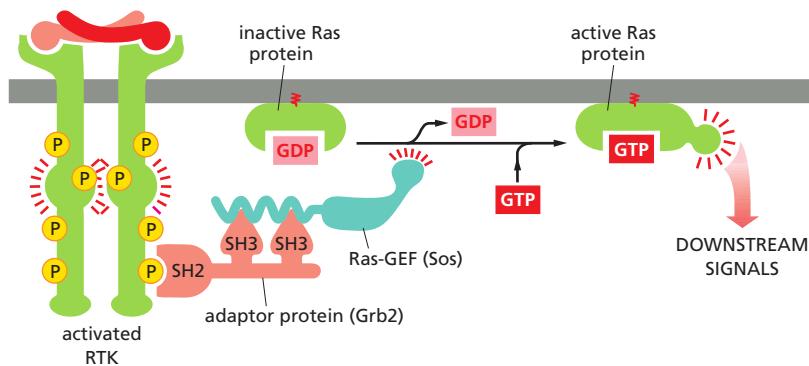
There are three major, closely related Ras proteins in humans: H-, K-, and N-Ras (see Table 15–5). Although they have subtly different functions, they are thought to work in the same way, and we will refer to them simply as **Ras**. Like many monomeric GTPases, Ras contains one or more covalently attached lipid groups that help anchor the protein to the cytoplasmic face of the membrane, from where it relays signals to other parts of the cell. Ras is often required, for example, when RTKs signal to the nucleus to stimulate cell proliferation or differentiation, both of which require changes in gene expression. If Ras function is inhibited by various experimental approaches, the cell proliferation or differentiation responses normally induced by the activated RTKs do not occur. Conversely, 30% of human tumors express hyperactive mutant forms of Ras, which contribute to the uncontrolled proliferation of the cancer cells.

Like other GTP-binding proteins, Ras functions as a molecular switch, cycling between two distinct conformational states—active when GTP is bound and inactive when GDP is bound (Movie 15.7). As discussed earlier for monomeric GTPases in general, two classes of signaling proteins regulate Ras activity by influencing its transition between active and inactive states (see Figure 15–8).

**TABLE 15–5 The Ras Superfamily of Monomeric GTPases**

Family	Some family members	Some functions
Ras	H-Ras, K-Ras, N-Ras	Relay signals from RTKs
	Rheb	Activates mTOR to stimulate cell growth
	Rap1	Activated by a cyclic-AMP-dependent GEF; influences cell adhesion by activating integrins
Rho*	Rho, Rac, Cdc42	Relay signals from surface receptors to the cytoskeleton and elsewhere
ARF*	ARF1–ARF6	Regulate assembly of protein coats on intracellular vesicles
Rab*	Rab1–60	Regulate intracellular vesicle traffic
Ran*	Ran	Regulates mitotic spindle assembly and nuclear transport of RNAs and proteins

\*The Rho family is discussed in Chapter 16, the ARF and Rab proteins in Chapter 13, and Ran in Chapters 12 and 17. The three-dimensional structure of Ras is shown in Figure 3–67.



**Ras guanine nucleotide exchange factors (Ras-GEFs)** stimulate the dissociation of GDP and the subsequent uptake of GTP from the cytosol, thereby activating Ras. **Ras GTPase-activating proteins (Ras-GAPs)** increase the rate of hydrolysis of bound GTP by Ras, thereby inactivating Ras. Hyperactive mutant forms of Ras are resistant to GAP-mediated GTPase stimulation and are locked permanently in the GTP-bound active state, which is why they promote the development of cancer.

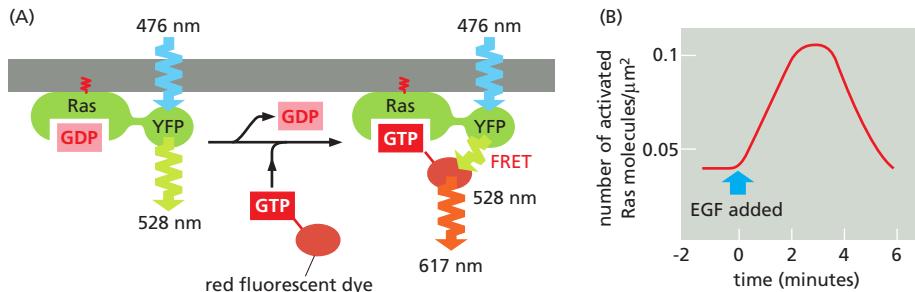
But how do RTKs normally activate Ras? In principle, they could either activate a Ras-GEF or inhibit a Ras-GAP. Even though some GAPs bind directly (via their SH2 domains) to activated RTKs (see Figure 15–46A), it is the indirect coupling of the receptor to a Ras-GEF that drives Ras into its active state. The loss of function of a Ras-GEF has a similar effect to the loss of function of Ras itself. Activation of the other Ras superfamily proteins, including those of the Rho family, also occurs through the activation of GEFs. The particular GEF determines in which membrane the GTPase is activated and, by acting as a scaffold, it can also determine which downstream proteins the GTPase activates.

The GEF that mediates Ras activation by RTKs was discovered by genetic studies of eye development in *Drosophila*, where an RTK called *Sevenless* (*Sev*) is required for the formation of a photoreceptor cell called R7. Genetic screens for components of this signaling pathway led to the discovery of a Ras-GEF called *Son-of-sevenless* (*Sos*). Further genetic screens uncovered another protein, now called *Grb2*, which is an adaptor protein that links the *Sev* receptor to the *Sos* protein; the SH2 domain of the *Grb2* adaptor binds to the activated receptor, while its two SH3 domains bind to *Sos*. *Sos* then promotes Ras activation. Biochemical and cell biological studies have shown that *Grb2* and *Sos* also link activated RTKs to Ras in mammalian cells, revealing that this is a highly conserved mechanism in RTK signaling (Figure 15–47). Once activated, Ras activates various other signaling proteins to relay the signal downstream, as we discuss next.

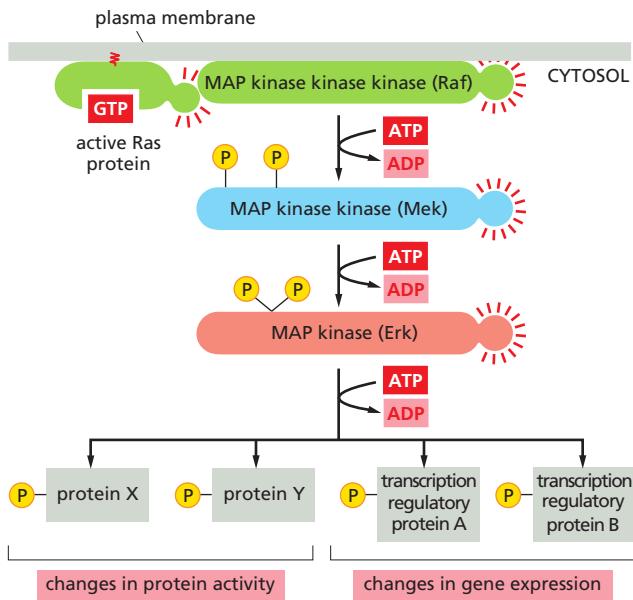
### Ras Activates a MAP Kinase Signaling Module

Both the tyrosine phosphorylations and the activation of Ras triggered by activated RTKs are usually short-lived (Figure 15–48). Tyrosine-specific protein phosphatases quickly reverse the phosphorylations, and Ras-GAPs induce activated

**Figure 15–47 How an RTK activates Ras.** Grb2 recognizes a specific phosphorylated tyrosine on the activated receptor by means of an SH2 domain and recruits Sos by means of two SH3 domains. Sos stimulates the inactive Ras protein to replace its bound GDP by GTP, which activates Ras to relay the signal downstream.



**Figure 15–48 Transient activation of Ras revealed by single-molecule fluorescence resonance energy transfer (FRET).** (A) Schematic drawing of the experimental strategy. Cells of a human cancer cell line are genetically engineered to express a Ras protein that is covalently linked to yellow fluorescent protein (YFP). GTP that is labeled with a red fluorescent dye is microinjected into some of the cells. The cells are then stimulated with the extracellular signal protein EGF, and single fluorescent molecules of Ras-YFP at the inner surface of the plasma membrane are followed by video fluorescence microscopy in individual cells. When a fluorescent Ras-YFP molecule becomes activated, it exchanges unlabeled GDP for fluorescently labeled GTP; the energy emitted by the YFP now activates the fluorescent GTP to emit red light (called fluorescence resonance energy transfer, or FRET; see Figure 9–26). Thus, the activation of single Ras molecules can be followed by the emission of red fluorescence from a previously yellow-green fluorescent spot at the plasma membrane. As shown in (B), activated Ras molecules can be detected after about 30 seconds of EGF stimulation. The red signal peaks at about 3 minutes and then decreases to baseline by 6 minutes. As Ras-GAP is found to be recruited to the same spots at the plasma membrane as Ras, it presumably plays a major part in rapidly shutting off the Ras signal. (Modified from H. Murakoshi et al., Proc. Natl. Acad. Sci. USA 101:7317–7322, 2004. With permission from National Academy of Sciences.)



**Figure 15–49** The MAP kinase module activated by Ras. The three-component module begins with a MAP kinase kinase kinase called Raf. Ras recruits Raf to the plasma membrane and helps activate it. Raf then activates the MAP kinase kinase Mek, which then activates the MAP kinase Erk. Erk in turn phosphorylates a variety of downstream proteins, including other protein kinases, as well as transcription regulators in the nucleus. The resulting changes in protein activities and gene expression cause complex changes in cell behavior.

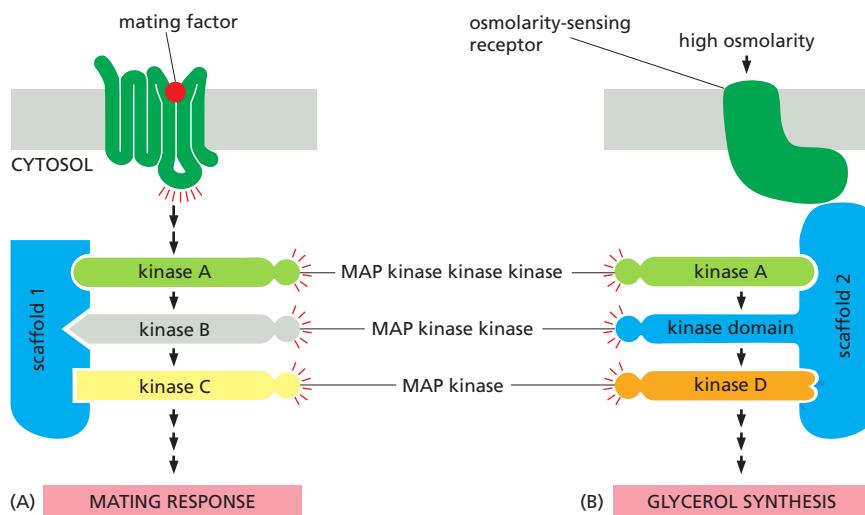
Ras to inactivate itself by hydrolyzing its bound GTP to GDP. To stimulate cells to proliferate or differentiate, these short-lived signaling events must be converted into longer-lasting ones that can sustain the signal and relay it downstream to the nucleus to alter the pattern of gene expression. One of the key mechanisms used for this purpose is a system of proteins called the *mitogen-activated protein kinase module* (MAP kinase module) (Figure 15–49). The three components of this system form a functional signaling module that has been remarkably well conserved during evolution and is used, with variations, in many different signaling contexts.

The three components are all protein kinases. The final kinase in the series is called simply MAP kinase (MAPK). The next one upstream from this is MAP kinase kinase (MAPKK): it phosphorylates and thereby activates MAP kinase. Next above that, receiving an activating signal directly from Ras, is MAP kinase kinase kinase (MAPKKK): it phosphorylates and thereby activates MAPKK. In the mammalian **Ras–MAP–kinase signaling pathway**, these three kinases are known by shorter names: Raf (= MAPKKK), Mek (= MAPKK), and Erk (=MAPK).

Once activated, the MAP kinase relays the signal downstream by phosphorylating various proteins in the cell, including transcription regulators and other protein kinases (see Figure 15–49). Erk, for example, enters the nucleus and phosphorylates one or more components of a transcription regulatory complex. This activates the transcription of a set of *immediate early genes*, so named because they turn on within minutes after an RTK receives an extracellular signal, even if protein synthesis is experimentally blocked with drugs. Some of these genes encode other transcription regulators that turn on other genes, a process that requires both protein synthesis and more time. In this way, the Ras–MAP–kinase signaling pathway conveys signals from the cell surface to the nucleus and alters the pattern of gene expression. Among the genes activated by this pathway are some that stimulate cell proliferation, such as the genes encoding *G<sub>1</sub> cyclins* (discussed in Chapter 17).

Extracellular signals usually activate MAP kinases only transiently, and the period during which the kinase remains active influences the response. When EGF activates its receptors in a neural precursor cell line, for example, Erk MAP kinase activity peaks at 5 minutes and rapidly declines, and the cells later go on to divide. By contrast, when NGF activates its receptors on the same cells, Erk activity remains high for many hours, and the cells stop proliferating and differentiate into neurons.

Many factors influence the duration and other features of the signaling response, including positive and negative feedback loops, which can combine to



**Figure 15–50** The organization of two MAP kinase modules by scaffold proteins in budding yeast. Budding yeast have at least six three-component MAP kinase modules involved in a variety of biological processes, including the two responses illustrated here—a mating response and the response to high osmolarity. (A) The mating response is triggered when a mating factor secreted by a yeast of opposite mating type binds to a GPCR. This activates a G protein, the  $\beta\gamma$  complex of which indirectly activates the MAPKKK (kinase A), which then relays the response onward. Once activated, the MAP kinase (kinase C) phosphorylates and thereby activates several proteins that mediate the mating response, in which the yeast cell stops dividing and prepares for fusion. The three kinases in this module are bound to scaffold protein 1. (B) In a second response, a yeast cell exposed to a high-osmolarity environment is induced to synthesize glycerol to increase its internal osmolarity. This response is mediated by an osmolarity-sensing receptor protein and a different MAP kinase module bound to a second scaffold protein. (Note that the kinase domain of scaffold 2 provides the MAPKK activity of this module.) Although both pathways use the same MAPKKK (kinase A, green), there is no cross-talk between them because the kinases in each module are bound to different scaffold proteins, and the osmosensor is bound to the same scaffold protein as the particular kinase it activates.

give responses that are either graded or switchlike and either brief or long lasting. In an example illustrated earlier, in Figure 15–19, MAP kinase activates a complex positive feedback loop to produce an all-or-none, irreversible response when frog oocytes are stimulated to mature by a brief exposure to the extracellular signal molecule progesterone. In many cells, MAP kinases activate a negative feedback loop by increasing the concentration of a protein phosphatase that removes the phosphate from MAP kinase. The increase in the phosphatase results from both an increase in the transcription of the phosphatase gene and the stabilization of the enzyme against degradation. In the Ras-MAP-kinase pathway shown in Figure 15–49, Erk also phosphorylates and inactivates Raf, providing another negative feedback loop that helps shut off the MAP kinase module.

### Scaffold Proteins Help Prevent Cross-talk Between Parallel MAP Kinase Modules

Three-component MAP kinase signaling modules operate in all eukaryotic cells, with different modules mediating different responses in the same cell. In budding yeast, for example, one such module mediates the response to mating pheromone, another the response to starvation, and yet another the response to osmotic shock. Some of these MAP kinase modules use one or more of the same kinases and yet manage to activate different effector proteins and hence different responses. As discussed earlier, one way in which cells avoid cross-talk between the different parallel signaling pathways and ensure that each response is specific is to use scaffold proteins (see Figure 15–10A). In budding yeast cells, such scaffolds bind all or some of the kinases in each MAP kinase module to form a complex and thereby help to ensure response specificity (**Figure 15–50**).

Mammalian cells also use this scaffold strategy to prevent cross-talk between different MAP kinase modules. At least five parallel MAP kinase modules can operate in a mammalian cell. These modules make use of at least 12 MAP kinases, 7 MAPKKs, and 7 MAPKKKs. Two of these modules (terminating in MAP kinases called JNK and p38) are activated by different kinds of cell stresses, such as ultraviolet (UV) irradiation, heat shock, and osmotic stress, as well as by inflammatory cytokines; others mainly mediate responses to signals from other cells.

Although the scaffold strategy provides precision and avoids cross-talk, it reduces the opportunities for amplification and spreading of the signal to different parts of the cell, which require at least some of the components to be diffusible. It is unclear to what extent the individual components of MAP kinase modules can dissociate from the scaffold during the activation process to permit amplification.

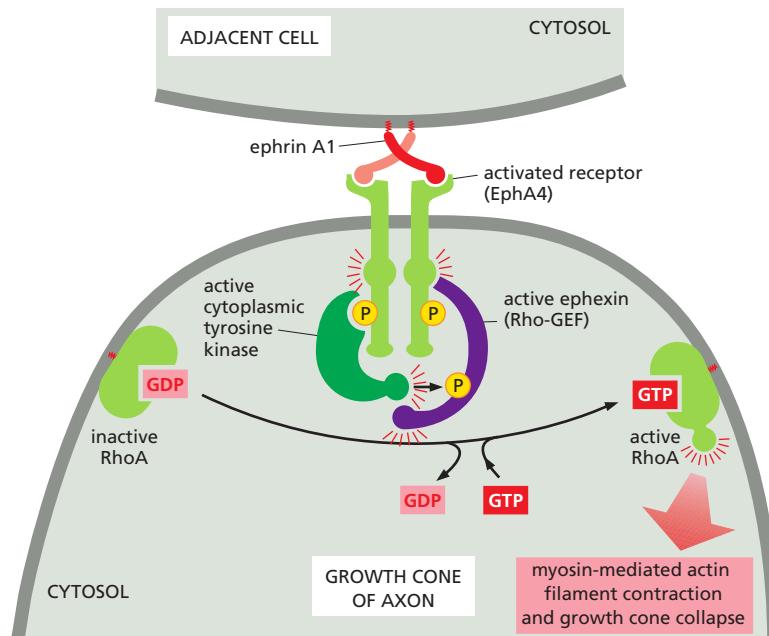
## Rho Family GTPases Functionally Couple Cell-Surface Receptors to the Cytoskeleton

Besides the Ras proteins, the other class of Ras superfamily GTPases that relays signals from cell-surface receptors is the large **Rho family** (see Table 15–5). Rho family monomeric GTPases regulate both the actin and microtubule cytoskeletons, controlling cell shape, polarity, motility, and adhesion (discussed in Chapter 16); they also regulate cell-cycle progression, gene transcription, and membrane transport. They play a key part in the guidance of cell migration and nerve axon outgrowth, mediating cytoskeletal responses to the activation of a special class of guidance receptors. We focus on this aspect of Rho family function here.

The three best-characterized family members are **Rho** itself, **Rac**, and **Cdc42**, each of which affects multiple downstream target proteins. In the same way as for Ras, GEFs activate and GAPs inactivate the Rho family GTPases; there are more than 80 Rho-GEFs and more than 70 Rho-GAPs in humans. Some of the GEFs and GAPs are specific for one particular family member, whereas others are less specific. Unlike Ras, which is membrane-associated even when inactive (with GDP bound), inactive Rho family GTPases are often bound to *guanine nucleotide dissociation inhibitors* (*GDI*s) in the cytosol, which prevent the GTPases from interacting with their Rho-GEFs at the plasma membrane.

Signaling by extracellular signaling proteins of the **ephrin** family provides an example of how RTKs can activate a Rho GTPase. Ephrins bind and thereby activate members of the *Eph* family of RTKs (see Figure 15–43). One member of the Eph family is found on the surface of motor neurons and helps guide the migrating tip of the axon (called a *growth cone*) to its muscle target. The binding of a cell-surface *ephrin* protein activates the Eph receptor, causing the growth cones to collapse, thereby repelling them from inappropriate regions and keeping them on track. The response depends on a Rho-GEF called *ephexin*, which is stably associated with the cytosolic tail of the Eph receptor. When ephrin binding activates the Eph receptor, the receptor activates a cytoplasmic tyrosine kinase that phosphorylates ephexin on a tyrosine, enhancing the ability of ephexin to activate the Rho protein RhoA. The activated RhoA (RhoA-GTP) then regulates various downstream target proteins, including some effector proteins that control the actin cytoskeleton, causing the growth cone to collapse (Figure 15–51).

Having considered how RTKs use GEFs and monomeric GTPases to relay signals into the cell, we now consider a second major strategy that RTKs use that depends on a quite different intracellular relay mechanism.



**Figure 15–51** Growth cone collapse mediated by Rho family GTPases. The binding of ephrin A1 proteins on an adjacent cell activates EphA4 RTKs on the growth cone. Phosphotyrosines on the activated Eph receptors recruit and activate a cytosolic tyrosine kinase to phosphorylate the receptor-associated Rho-GEF ephexin on a tyrosine. This enhances the ability of the ephexin to activate RhoA. RhoA then induces the growth cone to collapse by stimulating the myosin-dependent contraction of the actin cytoskeleton.

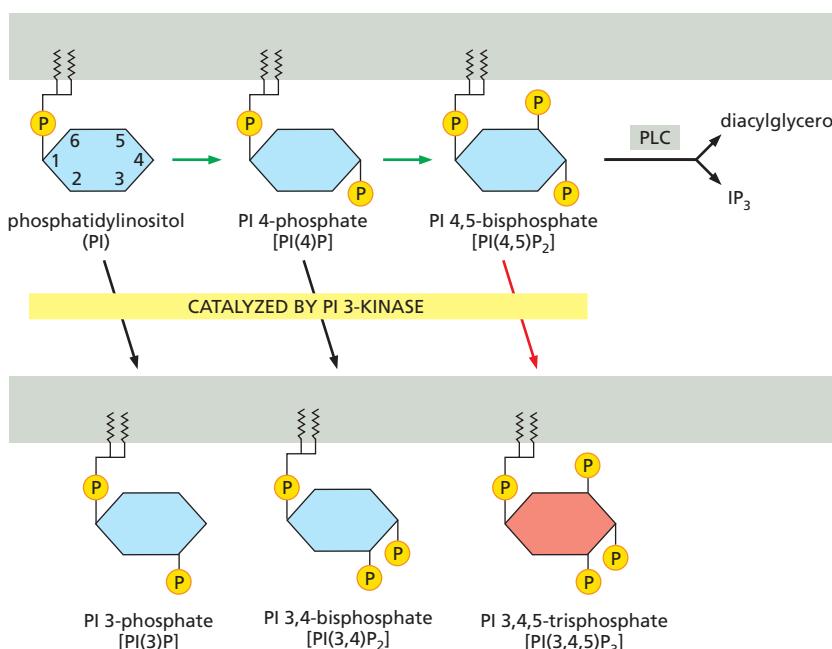
## PI 3-Kinase Produces Lipid Docking Sites in the Plasma Membrane

As mentioned earlier, one of the proteins that binds to the intracellular tail of RTK molecules is the plasma-membrane-bound enzyme **phosphoinositide 3-kinase (PI 3-kinase)**. This kinase principally phosphorylates inositol phospholipids rather than proteins, and both RTKs and GPCRs can activate it. It plays a central part in promoting cell survival and growth.

*Phosphatidylinositol (PI)* is unique among membrane lipids because it can undergo reversible phosphorylation at multiple sites on its inositol head group to generate a variety of phosphorylated PI lipids called **phosphoinositides**. When activated, PI 3-kinase catalyzes phosphorylation at the 3 position of the inositol ring to generate several phosphoinositides (Figure 15–52). The production of PI(3,4,5)P<sub>3</sub> matters most because it can serve as a docking site for various intracellular signaling proteins, which assemble into signaling complexes that relay the signal into the cell from the cytosolic face of the plasma membrane (see Figure 15–10C).

Notice the difference between this use of phosphoinositides and their use described earlier, in which PI(4,5)P<sub>2</sub> is cleaved by PLC $\beta$  (in the case of GPCRs) or PLC $\gamma$  (in the case of RTKs) to generate soluble IP<sub>3</sub> and membrane-bound diacylglycerol (see Figures 15–28 and 15–29). By contrast, PI(3,4,5)P<sub>3</sub> is not cleaved by either PLC. It is made from PI(4,5)P<sub>2</sub> and then remains in the plasma membrane until specific *phosphoinositide phosphatases* dephosphorylate it. Prominent among these is the *PTEN* phosphatase, which dephosphorylates the 3 position of the inositol ring. Mutations in PTEN are found in many cancers: by prolonging signaling by PI 3-kinase, they promote uncontrolled cell growth.

There are various types of PI 3-kinases. Those activated by RTKs and GPCRs belong to class I. These are heterodimers composed of a common catalytic subunit and different regulatory subunits. RTKs activate *class Ia PI 3-kinases*, in which the regulatory subunit is an adaptor protein that binds to two phosphotyrosines on activated RTKs through its two SH2 domains (see Figure 15–46A). GPCRs activate *class Ib PI 3-kinases*, which have a regulatory subunit that binds to the  $\beta\gamma$  complex of an activated trimeric G protein (G<sub>q</sub>) when GPCRs are activated by their extracellular ligand. The direct binding of activated Ras can also activate the common class I catalytic subunit.



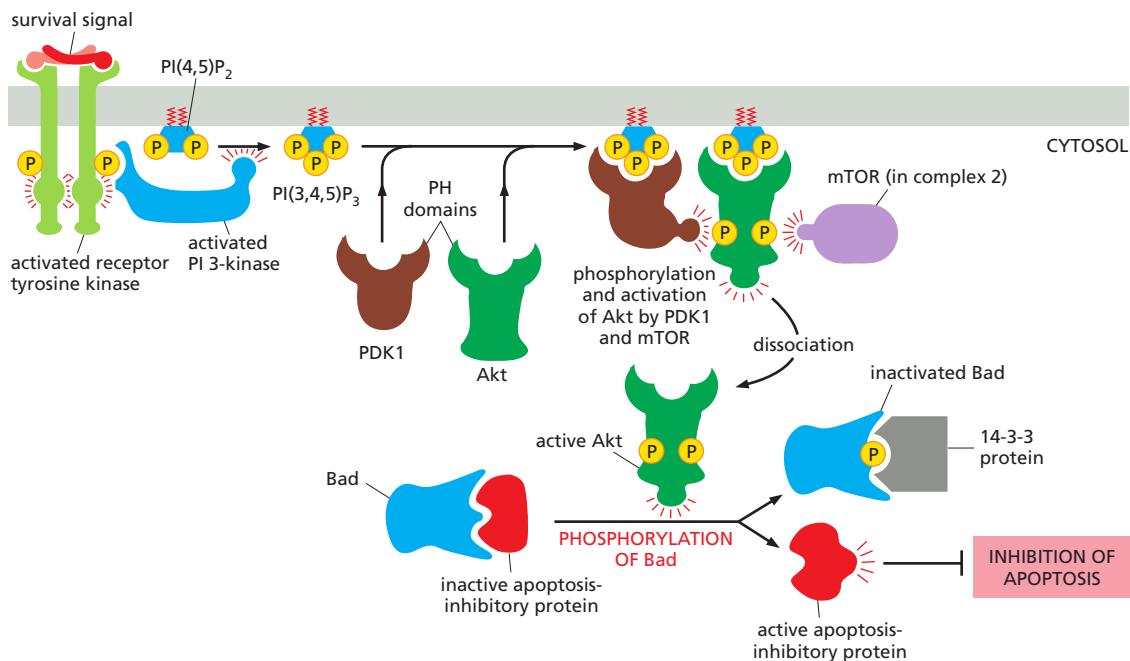
**Figure 15–52** The generation of phosphoinositide docking sites by PI 3-kinase. PI 3-kinase phosphorylates the inositol ring on carbon atom 3 to generate the phosphoinositides shown at the bottom of the figure (diverting them away from the pathway leading to IP<sub>3</sub> and diacylglycerol; see Figure 15–28). The most important phosphorylation (indicated in red) is of PI(4,5)P<sub>2</sub> to PI(3,4,5)P<sub>3</sub>, which can serve as a docking site for signaling proteins with PI(3,4,5)P<sub>3</sub>-binding PH domains. Other inositol phospholipid kinases (not shown) catalyze the phosphorylations indicated by the green arrows.

Intracellular signaling proteins bind to PI(3,4,5)P<sub>3</sub> produced by activated PI 3-kinase via a specific interaction domain, such as a **pleckstrin homology (PH) domain**, first identified in the platelet protein pleckstrin. PH domains function mainly as protein–protein interaction domains, and it is only a small subset of them that bind to PI(3,4,5)P<sub>3</sub>; at least some of these also recognize a specific membrane-bound protein as well as the PI(3,4,5)P<sub>3</sub>, which greatly increases the specificity of the binding and helps to explain why the signaling proteins with PI(3,4,5)P<sub>3</sub>-binding PH domains do not all dock at all PI(3,4,5)P<sub>3</sub> sites. PH domains occur in about 200 human proteins, including the Ras-GEF Sos discussed earlier (see Figure 15–11).

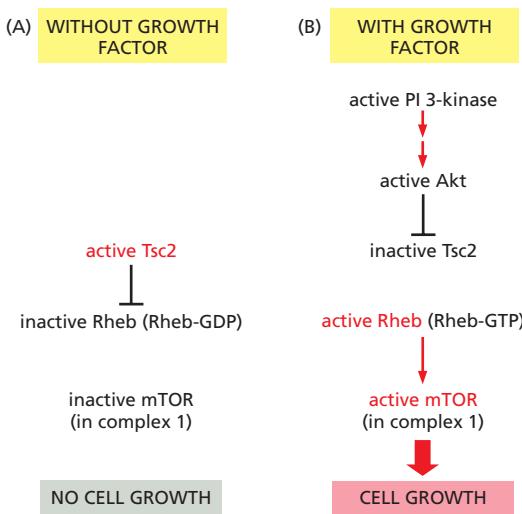
One especially important PH-domain-containing protein is the serine/threonine protein kinase *Akt*. The *PI-3-kinase–Akt signaling pathway* is the major pathway activated by the hormone *insulin*. It also plays a key part in promoting the survival and growth of many cell types in both invertebrates and vertebrates, as we now discuss.

### The PI-3-Kinase–Akt Signaling Pathway Stimulates Animal Cells to Survive and Grow

As discussed earlier, extracellular signals are usually required for animal cells to grow and divide, as well as to survive (see Figure 15–4). Members of the *insulin-like growth factor (IGF)* family of signal proteins, for example, stimulate many types of animal cells to survive and grow. They bind to specific RTKs (see Figure 15–43), which activate PI 3-kinase to produce PI(3,4,5)P<sub>3</sub>. The PI(3,4,5)P<sub>3</sub> recruits two protein kinases to the plasma membrane via their PH domains—Akt (also called *protein kinase B*, or *PKB*) and *phosphoinositide-dependent protein kinase 1* (*PDK1*), and this leads to the activation of Akt (Figure 15–53). Once activated,



**Figure 15–53** One way in which signaling through PI 3-kinase promotes cell survival. An extracellular survival signal activates an RTK, which recruits and activates PI 3-kinase. The PI 3-kinase produces PI(3,4,5)P<sub>3</sub>, which serves as a docking site for two serine/threonine kinases with PH domains—Akt and the phosphoinositide-dependent kinase PDK1—and brings them into proximity at the plasma membrane. The Akt is phosphorylated on a serine by a third kinase (usually mTOR in complex 2), which alters the conformation of the Akt so that it can be phosphorylated on a threonine by PDK1, which activates the Akt. The activated Akt now dissociates from the plasma membrane and phosphorylates various target proteins, including the Bad protein. When unphosphorylated, Bad holds one or more apoptosis-inhibitory proteins (of the Bcl2 family—discussed in Chapter 18) in an inactive state. Once phosphorylated, Bad releases the inhibitory proteins, which now can block apoptosis and thereby promote cell survival. As shown, the phosphorylated Bad binds to a ubiquitous cytosolic protein called 14-3-3, which keeps Bad out of action.



Akt phosphorylates various target proteins at the plasma membrane, as well as in the cytosol and nucleus. The effect on most of the known targets is to inactivate them; but the targets are such that these actions of Akt all conspire to enhance cell survival and growth, as illustrated for one cell survival pathway in Figure 15–53.

The control of cell growth by the **PI-3-kinase-Akt pathway** depends in part on a large protein kinase called **TOR** (named as the target of *rapamycin*, a bacterial toxin that inactivates the kinase and is used clinically as both an immunosuppressant and anticancer drug). TOR was originally identified in yeasts in genetic screens for rapamycin resistance; in mammalian cells, it is called **mTOR**, which exists in cells in two functionally distinct multiprotein complexes. *mTOR complex 1* contains the protein *raptor*; this complex is sensitive to rapamycin, and it stimulates cell growth—both by promoting ribosome production and protein synthesis and by inhibiting protein degradation. Complex 1 also promotes both cell growth and cell survival by stimulating nutrient uptake and metabolism. *mTOR complex 2* contains the protein *rictor* and is insensitive to rapamycin; it helps to activate Akt (see Figure 15–53), and it regulates the actin cytoskeleton via Rho family GTPases.

The mTOR in complex 1 integrates inputs from various sources, including extracellular signal proteins referred to as *growth factors* and nutrients such as amino acids, both of which help activate mTOR and promote cell growth. The growth factors activate mTOR mainly via the PI-3-kinase-Akt pathway. Akt activates mTOR in complex 1 indirectly by phosphorylating, and thereby inhibiting, a GAP called Tsc2. Tsc2 acts on a monomeric Ras-related GTPase called **Rheb** (see Table 15–5). Rheb in its active form (Rheb-GTP) activates mTOR in complex 1. The net result is that Akt activates mTOR and thereby promotes cell growth (Figure 15–54). We discuss how mTOR stimulates ribosome production and protein synthesis in Chapter 17 (see Figure 17–64).

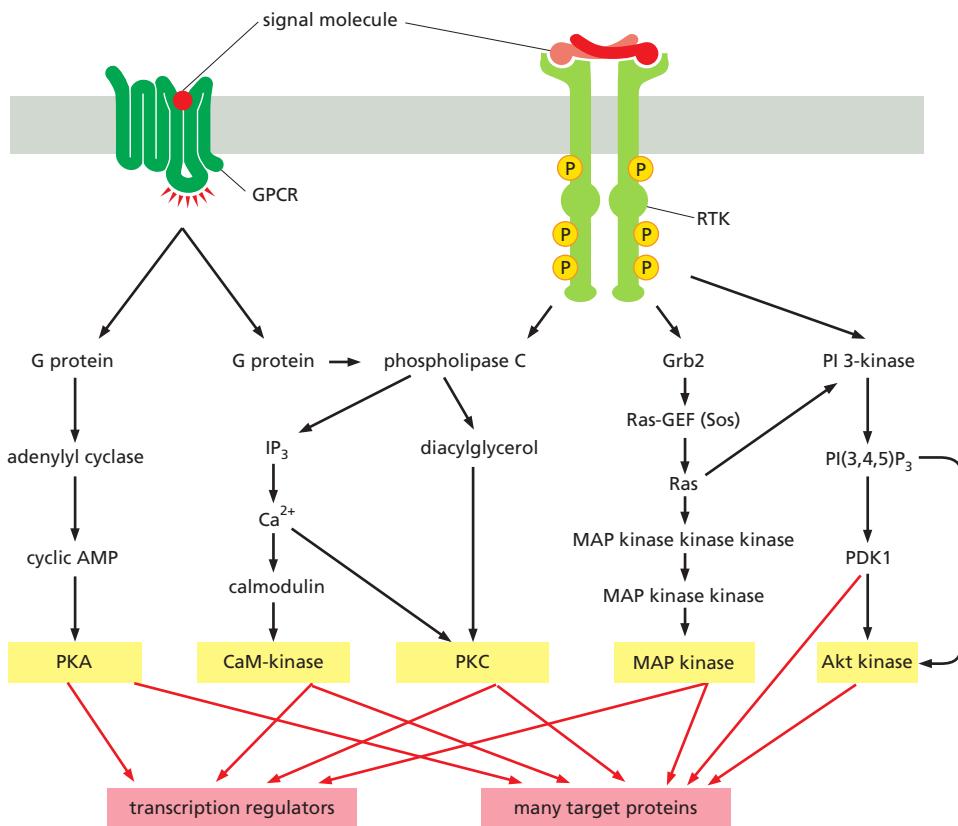
### RTKs and GPCRs Activate Overlapping Signaling Pathways

As mentioned earlier, RTKs and GPCRs activate some of the same intracellular signaling pathways. Both, for example, can activate the inositol phospholipid pathway triggered by phospholipase C. Moreover, even when they activate different pathways, the different pathways can converge on the same target proteins. Figure 15–55 illustrates both of these types of signaling overlaps: it summarizes five parallel intracellular signaling pathways that we have discussed so far—one triggered by GPCRs, two triggered by RTKs, and two triggered by both kinds of receptors. Interactions among these pathways allow different extracellular signal molecules to modulate and coordinate each other's effects.

**Figure 15–54 Activation of mTOR by the PI-3-kinase–Akt signaling pathway.**

(A) In the absence of extracellular growth factors, Tsc2 (a Rheb-GAP) keeps Rheb inactive; mTOR in complex 1 is inactive, and there is no cell growth. (B) In the presence of growth factors, activated Akt phosphorylates and inhibits Tsc2, thereby promoting the activation of Rheb. Activated Rheb (Rheb-GTP) helps activate mTOR in complex 1, which in turn stimulates cell growth. Figure 15–53 shows how growth factors (or survival signals) activate Akt. The Erk MAP kinase (see Figure 15–49) can also phosphorylate and inhibit Tsc2 and thereby activate mTOR. Thus, both the PI-3-kinase–Akt and Ras–MAP-kinase signaling pathways converge on mTOR in complex 1 to stimulate cell growth.

Tsc2 is short for *tuberous sclerosis protein 2*, and it is one component of a heterodimer composed of Tsc1 and Tsc2 (not shown); these proteins are so called because mutations in either gene encoding them cause the genetic disease *tuberous sclerosis*, which is associated with benign tumors that contain abnormally large cells.



**Figure 15–55** Five parallel intracellular signaling pathways activated by GPCRs, RTKs, or both. In this simplified example, the five kinases (shaded yellow) at the end of each signaling pathway phosphorylate target proteins (shaded red), many of which are phosphorylated by more than one of the kinases. The phospholipase C activated by the two types of receptors is different: GPCRs activate PLC $\beta$ , whereas RTKs activate PLC $\gamma$  (not shown). Although not shown, some GPCRs can also activate Ras, but they do so independently of Grb2, via a Ras-GEF that is activated by Ca $^{2+}$  and diacylglycerol.

### Some Enzyme-Coupled Receptors Associate with Cytoplasmic Tyrosine Kinases

Many cell-surface receptors depend on tyrosine phosphorylation for their activity and yet lack a tyrosine kinase domain. These receptors act through **cytoplasmic tyrosine kinases**, which are associated with the receptors and phosphorylate various target proteins, often including the receptors themselves, when the receptors bind their ligand. These **tyrosine-kinase-associated receptors** thus function in much the same way as RTKs, except that their kinase domain is encoded by a separate gene and is noncovalently associated with the receptor polypeptide chain. A variety of receptor classes belong in this category, including the receptors for antigen and interleukins on lymphocytes (discussed in Chapter 24), integrins (discussed in Chapter 19), and receptors for various cytokines and some hormones. As with RTKs, many of these receptors are either preformed dimers or are cross-linked into dimers by ligand binding.

Some of these receptors depend on members of the largest family of mammalian cytoplasmic tyrosine kinases, the **Src family** (see Figures 3–10 and 3–64), which includes *Src*, *Yes*, *Fgr*, *Fyn*, *Lck*, *Lyn*, *Hck*, and *Blk*. These protein kinases all contain SH2 and SH3 domains and are located on the cytoplasmic side of the plasma membrane, held there partly by their interaction with transmembrane receptor proteins and partly by covalently attached lipid chains. Different family members are associated with different receptors and phosphorylate overlapping but distinct sets of target proteins. Lyn, Fyn, and Lck, for example, are each associated with different sets of receptors on lymphocytes. In each case, the kinase is activated when an extracellular ligand binds to the appropriate receptor protein. Src itself, as well as several other family members, can also bind to activated RTKs; in these cases, the receptor and cytoplasmic kinases mutually stimulate each other's catalytic activity, thereby strengthening and prolonging the signal (see Figure 15–51). There are even some G proteins (G<sub>s</sub> and G<sub>i</sub>) that can activate Src, which is one way that the activation of GPCRs can lead to tyrosine phosphorylation of intracellular signaling proteins and effector proteins.

Another type of cytoplasmic tyrosine kinase associates with *integrins*, the main receptors that cells use to bind to the extracellular matrix (discussed in Chapter 19). The binding of matrix components to integrins activates intracellular signaling pathways that influence the behavior of the cell. When integrins cluster at sites of matrix contact, they help trigger the assembly of cell-matrix junctions called *focal adhesions*. Among the many proteins recruited into these junctions is the cytoplasmic tyrosine kinase called **focal adhesion kinase (FAK)**, which binds to the cytosolic tail of one of the integrin subunits with the assistance of other proteins. The clustered FAK molecules phosphorylate each other, creating phosphotyrosine docking sites where the Src kinase can bind. Src and FAK then phosphorylate each other and other proteins that assemble in the junction, including many of the signaling proteins used by RTKs. In this way, the two tyrosine kinases signal to the cell that it has adhered to a suitable substratum, where the cell can now survive, grow, divide, migrate, and so on.

The largest and most diverse class of receptors that rely on cytoplasmic tyrosine kinases to relay signals into the cell is the class of *cytokine receptors*, which we consider next.

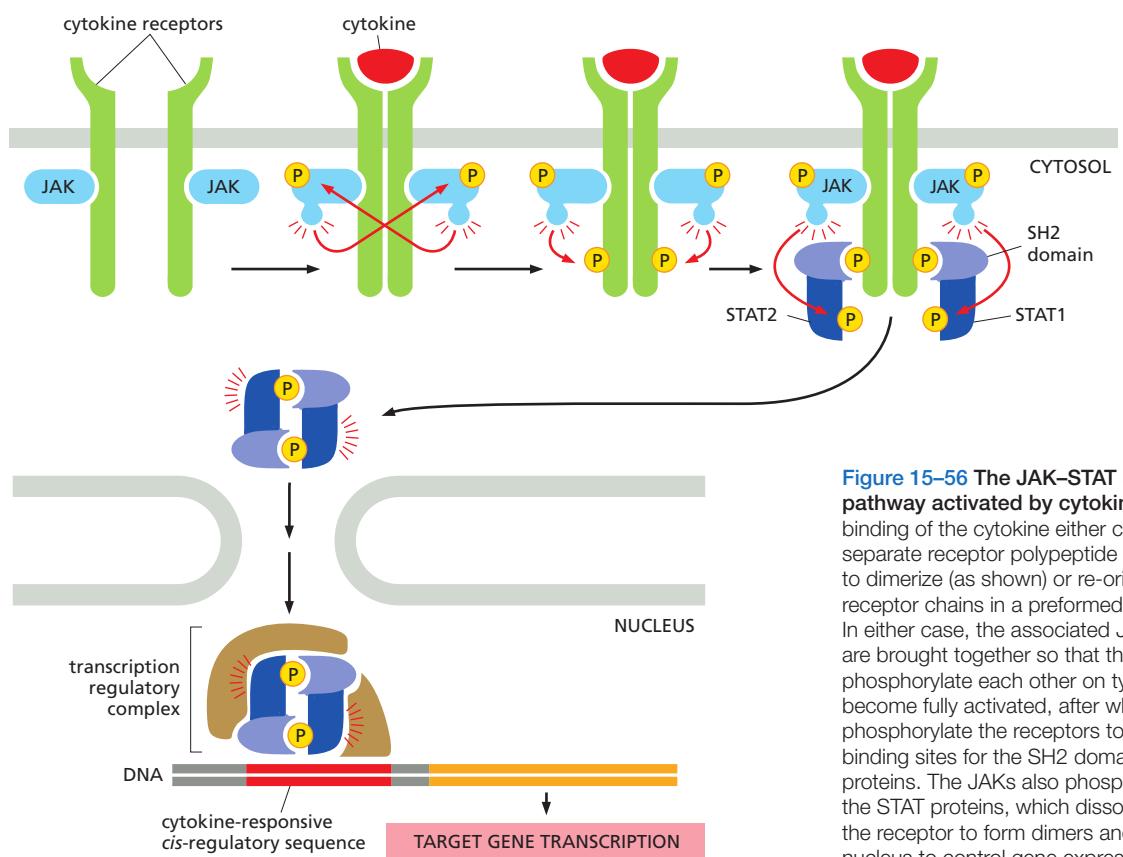
### Cytokine Receptors Activate the JAK–STAT Signaling Pathway

The large family of **cytokine receptors** includes receptors for many kinds of local mediators (collectively called *cytokines*), as well as receptors for some hormones, such as *growth hormone* and *prolactin* ([Movie 15.8](#)). These receptors are stably associated with cytoplasmic tyrosine kinases called **Janus kinases (JAKs)** (after the two-faced Roman god), which phosphorylate and activate transcription regulators called **STATs** (signal transducers and activators of transcription). STAT proteins are located in the cytosol and are referred to as *latent transcription regulators* because they migrate into the nucleus and regulate gene transcription only after they are activated.

Although many intracellular signaling pathways lead from cell-surface receptors to the nucleus, where they alter gene transcription (see Figure 15–55), the **JAK–STAT signaling pathway** provides one of the more direct routes. Cytokine receptors are dimers or trimers and are stably associated with one or two of the four known JAKs (JAK1, JAK2, JAK3, and Tyk2). Cytokine binding alters the arrangement so as to bring two JAKs into close proximity so that they phosphorylate each other, thereby increasing the activity of their tyrosine kinase domains. The JAKs then phosphorylate tyrosines on the cytoplasmic tails of cytokine receptors, creating phosphotyrosine docking sites for STATs ([Figure 15–56](#)). Some adaptor proteins can also bind to some of these sites and couple cytokine receptors to the Ras–MAP-kinase signaling pathway discussed earlier, but these will not be discussed here.

There are at least six STATs in mammals. Each has an SH2 domain that performs two functions. First, it mediates the binding of the STAT protein to a phosphotyrosine docking site on an activated cytokine receptor. Once bound, the JAKs phosphorylate the STAT on tyrosines, causing the STAT to dissociate from the receptor. Second, the SH2 domain on the released STAT now mediates its binding to a phosphotyrosine on another STAT molecule, forming either a STAT homodimer or a heterodimer. The STAT dimer then translocates to the nucleus, where, in combination with other transcription regulatory proteins, it binds to a specific *cis*-regulatory sequence in various genes and stimulates their transcription (see Figure 15–56). In response to the hormone prolactin, for example, which stimulates breast cells to produce milk, activated STAT5 stimulates the transcription of genes that encode milk proteins. [Table 15–6](#) lists some of the more than 30 cytokines and hormones that activate the JAK–STAT pathway by binding to cytokine receptors.

Negative feedback regulates the responses mediated by the JAK–STAT pathway. In addition to activating genes that encode proteins mediating the cytokine-induced response, the STAT dimers can also activate genes that encode inhibitory proteins that help shut off the response. Some of these proteins bind to



**Figure 15–56** The JAK–STAT signaling pathway activated by cytokines. The binding of the cytokine either causes two separate receptor polypeptide chains to dimerize (as shown) or re-orient the receptor chains in a preformed dimer. In either case, the associated JAKs are brought together so that they can phosphorylate each other on tyrosines to become fully activated, after which they phosphorylate the receptors to generate binding sites for the SH2 domains of STAT proteins. The JAKs also phosphorylate the STAT proteins, which dissociate from the receptor to form dimers and enter the nucleus to control gene expression.

and inactivate phosphorylated JAKs and their associated phosphorylated receptors; others bind to phosphorylated STAT dimers and prevent them from binding to their DNA targets. Such negative feedback mechanisms, however, are not enough on their own to turn off the response. Inactivation of the activated JAKs and STATs requires dephosphorylation of their phosphotyrosines.

### Protein Tyrosine Phosphatases Reverse Tyrosine Phosphorylations

In all signaling pathways that use tyrosine phosphorylation, the tyrosine phosphorylations are reversed by **protein tyrosine phosphatases**. These phosphatases are as important in the signaling process as the protein tyrosine kinases that add the phosphates. Whereas only a few types of *serine/threonine protein phosphatase*

**TABLE 15–6** Some Extracellular Signal Proteins That Act Through Cytokine Receptors and the JAK–STAT Signaling Pathway

Signal protein	Receptor-associated JAKs	STATs activated	Some responses
Interferon- $\gamma$ (IFN $\gamma$ )	JAK1 and JAK2	STAT1	Activates macrophages
Interferon- $\alpha$ (IFN $\alpha$ )	Tyk2 and JAK2	STAT1 and STAT2	Increases cell resistance to viral infection
Erythropoietin	JAK2	STAT5	Stimulates production of erythrocytes
Prolactin	JAK1 and JAK2	STAT5	Stimulates milk production
Growth hormone	JAK2	STAT1 and STAT5	Stimulates growth by inducing IGF1 production
Granulocyte–Macrophage-Colony-Stimulating Factor (GMCSF)	JAK2	STAT5	Stimulates production of granulocytes and macrophages

catalytic subunits are responsible for removing phosphate groups from phosphorylated serines and threonines on proteins, there are about 100 protein tyrosine phosphatases encoded in the human genome, including some *dual-specificity phosphatases* that also dephosphorylate serines and threonines.

Like tyrosine kinases, the tyrosine phosphatases occur in both cytoplasmic and transmembrane forms. Unlike serine/threonine protein phosphatases, which generally have broad specificity, most tyrosine phosphatases display exquisite specificity for their substrates, removing phosphate groups from only selected phosphotyrosines on a subset of proteins. Together, these phosphatases ensure that tyrosine phosphorylations are short-lived and that the level of tyrosine phosphorylation in resting cells is very low. They do not, however, simply continuously reverse the effects of protein tyrosine kinases; they are often regulated to act only at the appropriate time and place.

Having discussed the crucial role of tyrosine phosphorylation and dephosphorylation in the intracellular signaling pathways activated by many enzyme-coupled receptors, we now turn to a class of enzyme-coupled receptors that rely on serine and threonine phosphorylation. These *receptor serine/threonine kinases* activate an even more direct signaling pathway to the nucleus than does the JAK-STAT pathway. They directly phosphorylate latent transcription regulators called *Smads*, which then translocate into the nucleus to control gene transcription.

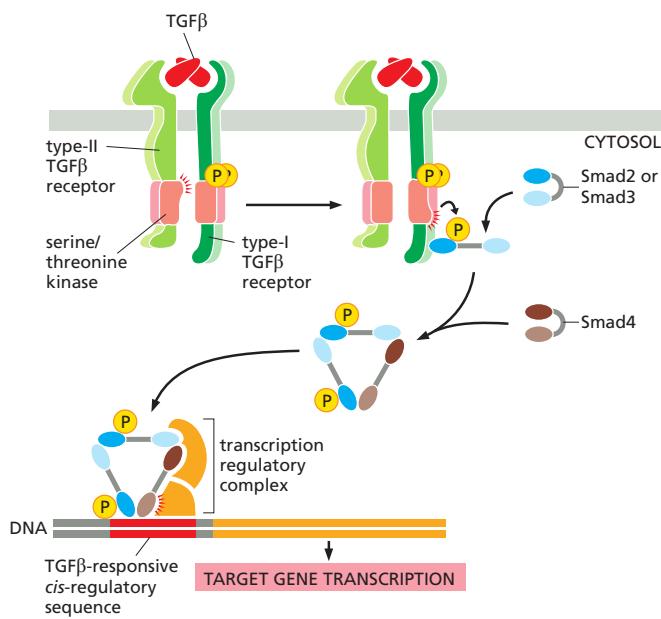
### Signal Proteins of the TGF $\beta$ Superfamily Act Through Receptor Serine/Threonine Kinases and Smads

The **transforming growth factor- $\beta$  (TGF $\beta$ ) superfamily** consists of a large number (33 in humans) of structurally related, secreted, dimeric proteins. They act either as hormones or, more commonly, as local mediators to regulate a wide range of biological functions in all animals. During development, they regulate pattern formation and influence various cell behaviors, including proliferation, specification and differentiation, extracellular matrix production, and cell death. In adults, they are involved in tissue repair and in immune regulation, as well as in many other processes. The superfamily consists of the TGF $\beta$ /*activin* family and the larger *bone morphogenetic protein (BMP)* family.

All of these proteins act through enzyme-coupled receptors that are single-pass transmembrane proteins with a serine/threonine kinase domain on the cytosolic side of the plasma membrane. There are two classes of these **receptor serine/threonine kinases**—*type I* and *type II*—which are structurally similar homodimers. Each member of the TGF $\beta$  superfamily binds to a characteristic combination of type-I and type-II receptor dimers, bringing the kinase domains together so that the type-II receptor can phosphorylate and activate the type-I receptor, forming an active tetrameric receptor complex.

Once activated, the receptor complex uses a strategy for rapidly relaying the signal to the nucleus that is very similar to the JAK-STAT strategy used by cytokine receptors. The activated type-I receptor directly binds and phosphorylates a latent transcription regulator of the **Smad family** (named after the first two proteins identified, Sma in *C. elegans* and Mad in *Drosophila*). Activated TGF $\beta$ /activin receptors phosphorylate Smad2 or Smad3, while activated BMP receptors phosphorylate Smad1, Smad5, or Smad8. Once one of these *receptor-activated Smads (R-Smads)* has been phosphorylated, it dissociates from the receptor and binds to Smad4 (called a *co-Smad*), which can form a complex with any of the five R-Smads. The Smad complex then translocates into the nucleus, where it associates with other transcription regulators and controls the transcription of specific target genes (**Figure 15–57**). Because the partner proteins in the nucleus vary depending on the cell type and state of the cell, the genes affected vary.

Activated TGF $\beta$  receptors and their bound ligand are endocytosed by two distinct routes, one leading to further activation and the other leading to inactivation. The activation route depends on clathrin-coated vesicles and leads to early endosomes (discussed in Chapter 13), where most of the Smad activation occurs. An anchoring protein called *SARA* (for *Smad anchor for receptor activation*) has



**Figure 15–57** The Smad-dependent signaling pathway activated by TGF $\beta$ . The TGF $\beta$  dimer promotes the assembly of a tetrameric receptor complex containing two copies each of the type-I and type-II receptors. The type-II receptors phosphorylate specific sites on the type-I receptors, thereby activating their kinase domains and leading to phosphorylation of R-Smads such as Smad2 and Smad3. Smads open up to expose a dimerization surface when they are phosphorylated, leading to the formation of a trimeric Smad complex containing two R-Smads and the co-Smad, Smad4. The phosphorylated Smad complex enters the nucleus and collaborates with other transcription regulators to control the transcription of specific target genes.

an important role in this pathway; it is concentrated in early endosomes and binds to both activated TGF $\beta$  receptors and Smads, increasing the efficiency of receptor-mediated Smad phosphorylation. The inactivation route depends on *caveolae* (discussed in Chapter 13) and leads to receptor ubiquitylation and degradation in proteasomes.

During the signaling response, the Smads shuttle continuously between the cytoplasm and the nucleus: they are dephosphorylated in the nucleus and exported to the cytoplasm, where they can be rephosphorylated by activated receptors. In this way, the effect exerted on the target genes reflects both the concentration of the extracellular signal and the time the signal continues to act on the cell-surface receptors (often several hours). Cells exposed to a morphogen at high concentration, or for a long time, or both, will switch on one set of genes, whereas cells receiving a lower or more transient exposure will switch on another set.

As in other signaling systems, negative feedback regulates the Smad pathway. Among the target genes activated by Smad complexes are those that encode *inhibitory Smads*, either Smad6 or Smad7. Smad7 (and possibly Smad6) binds to the cytosolic tail of the activated receptor and inhibits its signaling ability in at least three ways: (1) it competes with R-Smads for binding sites on the receptor, decreasing R-Smad phosphorylation; (2) it recruits a ubiquitin ligase called *Smurf*, which ubiquitylates the receptor, leading to receptor internalization and degradation (it is because Smurfs also ubiquitylate and promote the degradation of Smads that they are called *Smad ubiquitylation regulatory factors*, or Smurfs); and (3) it recruits a protein phosphatase that dephosphorylates and inactivates the receptor. In addition, the inhibitory Smads bind to the co-Smad, Smad4, and inhibit it, either by preventing its binding to R-Smads or by promoting its ubiquitylation and degradation.

Although receptor serine/threonine kinases operate mainly through the Smad pathway just described, they can also stimulate other intracellular signaling proteins such as MAP kinases and PI 3-kinase. Conversely, signaling proteins in other pathways can phosphorylate Smads and thereby influence signaling along the Smad pathway.

## Summary

There are various classes of enzyme-coupled receptors, the most common of which are receptor tyrosine kinases (RTKs), tyrosine-kinase-associated receptors, and receptor serine/threonine kinases.

Ligand binding to RTKs causes their dimerization, which leads to activation of their kinase domains. These activated kinase domains phosphorylate multiple tyrosines on the receptors, producing a set of phosphotyrosines that serve as docking sites for a set of intracellular signaling proteins, which bind via their SH2 (or PTB) domains. One such signaling protein serves as an adaptor to couple some activated receptors to a Ras-GEF (Sos), which activates the monomeric GTPase Ras; Ras, in turn, activates a three-component MAP kinase signaling module, which relays the signal to the nucleus by phosphorylating transcription regulatory proteins. Another important signaling protein that can dock on activated RTKs is PI 3-kinase, which phosphorylates specific phosphoinositides to produce lipid docking sites in the plasma membrane for signaling proteins with phosphoinositide-binding PH domains, including the serine/threonine protein kinase Akt (PKB), which plays a key part in the control of cell survival and cell growth. Many receptor classes, including some RTKs, activate Rho family monomeric GTPases, which functionally couple the receptors to the cytoskeleton.

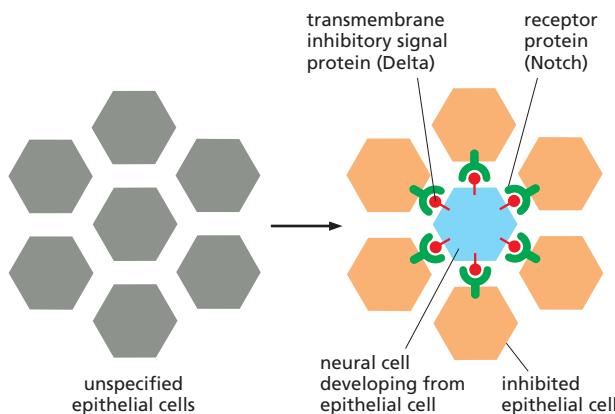
Tyrosine-kinase-associated receptors depend on various cytoplasmic tyrosine kinases for their action. These kinases include members of the Src family, which associate with many kinds of receptors, and the focal adhesion kinase (FAK), which associates with integrins at focal adhesions. The cytoplasmic tyrosine kinases then phosphorylate a variety of signaling proteins to relay the signal onward. The largest family of receptors in this class is the cytokine receptor family. When stimulated by ligand binding, these receptors activate JAK cytoplasmic tyrosine kinases, which phosphorylate STATs. The STATs then dimerize, translocate to the nucleus, and activate the transcription of specific genes. Receptor serine/threonine kinases, which are activated by signal proteins of the TGF $\beta$  superfamily, act similarly: they directly phosphorylate and activate Smads, which then oligomerize with another Smad, translocate to the nucleus, and regulate gene transcription.

## ALTERNATIVE SIGNALING ROUTES IN GENE REGULATION

Major changes in the behavior of a cell tend to depend on changes in the expression of numerous genes. Thus, many extracellular signaling molecules carry out their effects, in whole or in part, by initiating signaling pathways that change the activities of transcription regulators. There are numerous examples of gene regulation in both GPCR and enzyme-coupled receptor pathways (see Figures 15–27 and 15–49). In this section, we describe some of the less common signaling mechanisms by which gene expression can be controlled. We begin with several pathways that depend on *regulated proteolysis* to control the activity and location of latent transcription regulators. We then turn to a class of extracellular signal molecules that do not employ cell-surface receptors but enter the cell and interact directly with transcription regulators to perform their functions. Finally, we briefly discuss some of the mechanisms by which gene expression is controlled by the *circadian rhythm*: the daily cycle of light and dark.

### The Receptor Notch Is a Latent Transcription Regulatory Protein

Signaling through the **Notch** receptor protein is used widely in animal development. As discussed in Chapter 22, it has a general role in controlling cell fate choices and regulating pattern formation during the development of most tissues, as well as in the continual renewal of tissues such as the lining of the gut. It is best known, however, for its role in the production of *Drosophila* neural cells, which usually arise as isolated single cells within an epithelial sheet of precursor cells. During this process, when a precursor cell commits to becoming a neural cell, it signals to its immediate neighbors not to do the same; the inhibited cells develop into epidermal cells instead. This process, called *lateral inhibition*, depends on a contact-dependent signaling mechanism that is activated by a single-pass transmembrane signal protein called **Delta**, displayed on the surface of the future neural cell. By binding to the Notch receptor protein on a neighboring cell, Delta



**Figure 15–58** Lateral inhibition mediated by Notch and Delta during neural cell development in *Drosophila*. When individual cells in the epithelium begin to develop as neural cells, they signal to their neighbors not to do the same. This inhibitory, contact-dependent signaling is mediated by the ligand Delta, which appears on the surface of the future neural cell and binds to Notch receptor proteins on the neighboring cells. In many tissues, all the cells in a cluster initially express both Delta and Notch, and a competition occurs, with one cell emerging as winner, expressing Delta strongly and inhibiting its neighbors from doing likewise. In other cases, additional factors interact with Delta or Notch to make some cells susceptible to the lateral inhibition signal and others unresponsive to it.

signals to the neighbor not to become neural (Figure 15–58). When this signaling process is defective, a huge excess of neural cells is produced at the expense of epidermal cells, which is lethal.

Notch is a single-pass transmembrane protein that requires proteolytic processing to function. It acts as a latent transcription regulator and provides the simplest and most direct signaling pathway known from a cell-surface receptor to the nucleus. When activated by the binding of Delta on another cell, a plasma-membrane-bound protease cleaves off the cytoplasmic tail of Notch, and the released tail translocates into the nucleus to activate the transcription of a set of Notch-response genes. The Notch tail fragment acts by binding to a DNA-binding protein, converting it from a transcriptional repressor into a transcriptional activator.

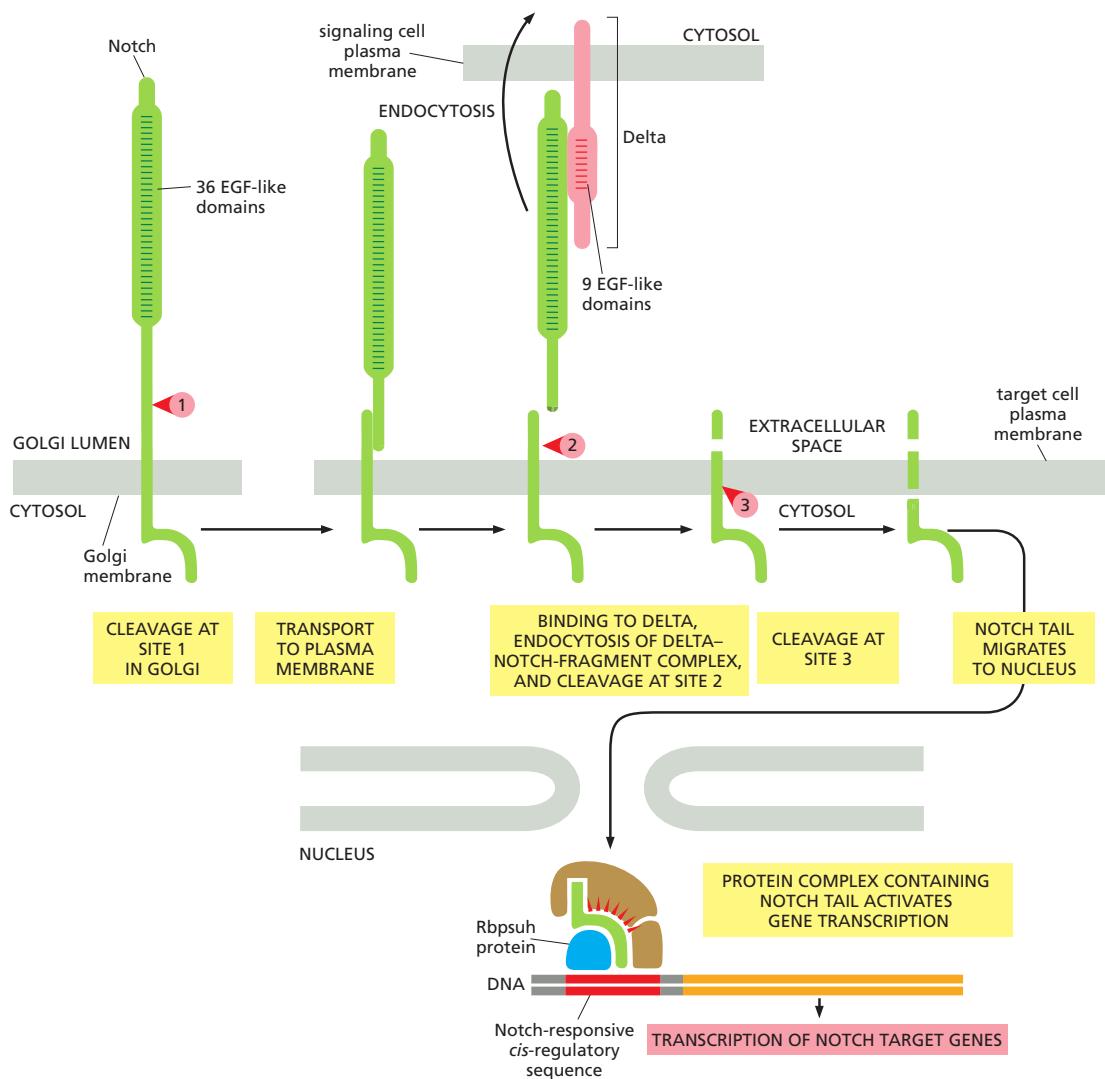
The Notch receptor undergoes three successive proteolytic cleavage steps, but only the last two depend on Delta binding. As part of its normal biosynthesis, it is cleaved in the Golgi apparatus to form a heterodimer, which is then transported to the cell surface as the mature receptor. The binding of Delta to Notch induces a second cleavage in the extracellular domain, mediated by an extracellular protease. A final cleavage quickly follows, cutting free the cytoplasmic tail of the activated receptor (Figure 15–59). Note that, unlike most receptors, the activation of Notch is irreversible; once activated by ligand binding, the protein cannot be used again.

This final cleavage of the Notch tail occurs just within the transmembrane segment, and it is mediated by a protease complex called  $\gamma$ -secretase, which is also responsible for the intramembrane cleavage of various other proteins. One of its essential subunits is *Presenilin*, so called because mutations in the gene encoding it are a frequent cause of early-onset, familial Alzheimer's disease, a form of presenile dementia. The protease complex is thought to contribute to this and other forms of Alzheimer's disease by generating extracellular peptide fragments from a transmembrane neuronal protein; the fragments accumulate in excessive amounts and form aggregates of misfolded protein called amyloid plaques, which may injure nerve cells and contribute to their degeneration and loss.

Both Notch and Delta are glycoproteins, and their interaction is regulated by the glycosylation of Notch. The *Fringe family* of glycosyl transferases, in particular, adds extra sugars to the O-linked oligosaccharide (discussed in Chapter 13) on Notch, which alters the specificity of Notch for its ligands. This has provided the first example of the modulation of ligand-receptor signaling by differential receptor glycosylation.

### Wnt Proteins Bind to Frizzled Receptors and Inhibit the Degradation of $\beta$ -Catenin

**Wnt proteins** are secreted signal molecules that act as local mediators and morphogens to control many aspects of development in all animals that have been studied. They were discovered independently in flies and in mice: in *Drosophila*, the *Wingless* (*Wg*) gene originally came to light because of its role as a morphogen



**Figure 15–59** The processing and activation of Notch by proteolytic cleavage. The numbered red arrowheads indicate the sites of proteolytic cleavage. The first proteolytic processing step occurs within the *trans* Golgi network to generate the mature heterodimeric Notch receptor that is then displayed on the cell surface. The binding to Delta on a neighboring cell triggers the next two proteolytic steps: the complex of Delta and the Notch fragment to which it is bound is endocytosed by the Delta-expressing cell, exposing the extracellular cleavage site in the transmembrane Notch subunit. Note that Notch and Delta interact through their repeated EGF-like domains. The released Notch tail migrates into the nucleus, where it binds to the Rbpsuh protein, which converts from a transcriptional repressor to a transcriptional activator.

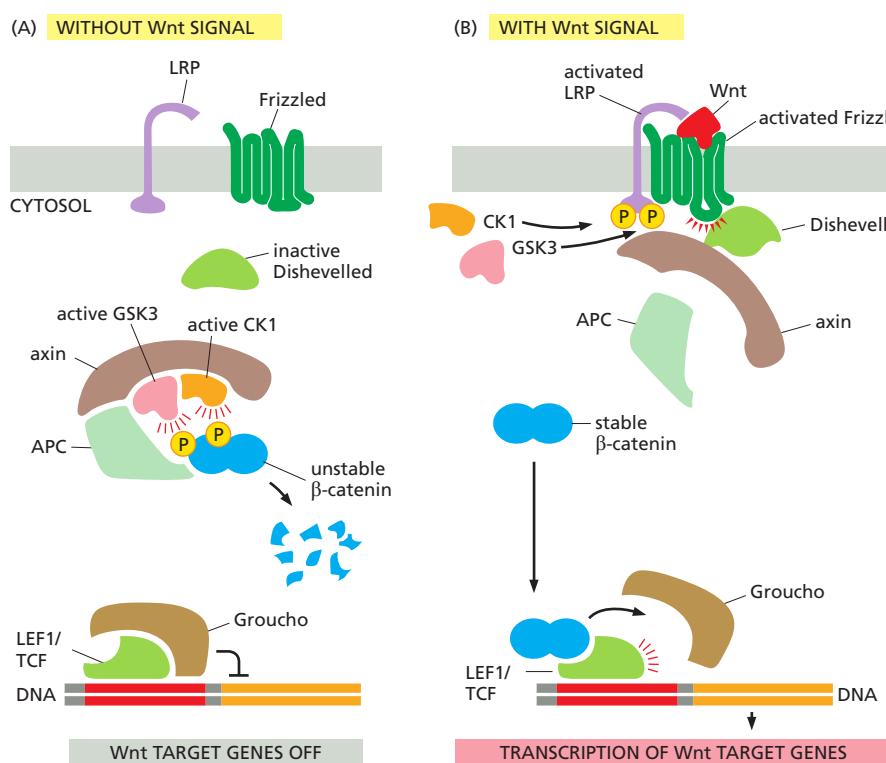
in wing development, while in mice, the *Int1* gene was found because it promoted the development of breast tumors when activated by the integration of a virus next to it. Both of these genes encode Wnt proteins. Wnts are unusual as secreted proteins in that they have a fatty acid chain covalently attached to their N-terminus, which increases their binding to cell surfaces. There are 19 Wnts in humans, each having distinct, but often overlapping, functions.

Wnts can activate at least two types of intracellular signaling pathways. Our primary focus here is the *Wnt/β-catenin pathway* (also known as the *canonical Wnt pathway*), which is centered on the latent transcription regulator *β-catenin*. A second pathway, called the *planar polarity pathway*, coordinates the polarization of cells in the plane of a developing epithelium and depends on Rho family GTPases. Both of these pathways begin with the binding of Wnts to **Frizzled**

family cell-surface receptors, which are seven-pass transmembrane proteins that resemble GPCRs in structure but do not generally work through the activation of G proteins. Instead, when activated by Wnt binding, Frizzled proteins recruit the scaffold protein **Dishevelled**, which helps relay the signal to other signaling molecules.

The **Wnt/β-catenin pathway** acts by regulating the proteolysis of the multi-functional protein **β-catenin** (or *Armadillo* in flies). A portion of the cell's β-catenin is located at cell-cell junctions and thereby contributes to the control of cell-cell adhesion (discussed in Chapter 19), while the remaining β-catenin is rapidly degraded in the cytoplasm. Degradation depends on a large protein *degradation complex*, which binds β-catenin and keeps it out of the nucleus while promoting its degradation. The complex contains at least four other proteins: a protein kinase called *casein kinase 1 (CK1)* phosphorylates the β-catenin on a serine, priming it for further phosphorylation by another protein kinase called *glycogen synthase kinase 3 (GSK3)*; this final phosphorylation marks the protein for ubiquitylation and rapid degradation in proteasomes. Two scaffold proteins called *axin* and *Adenomatous polyposis coli (APC)* hold the protein complex together (Figure 15–60A). APC gets its name from the finding that the gene encoding it is often mutated in a type of benign tumor (adenoma) of the colon; the tumor projects into the lumen as a polyp and can eventually become malignant. (This APC should not be confused with the anaphase-promoting complex, or APC/C, that plays a central part in selective protein degradation during the cell cycle—see Figure 17–15A.)

Wnt proteins regulate β-catenin proteolysis by binding to both a Frizzled protein and a co-receptor that is related to the low-density lipoprotein (LDL) receptor (discussed in Chapter 13) and is therefore called an **LDL-receptor-related protein (LRP)**. In a poorly understood process, the activated receptor complex recruits the Dishevelled scaffold and promotes the phosphorylation of the LRP receptor by the two protein kinases, GSK3 and CK1. Axin is brought to the receptor complex and inactivated, thereby disrupting the β-catenin degradation complex in the cytoplasm. In this way, the phosphorylation and degradation of β-catenin are prevented, enabling unphosphorylated β-catenin to accumulate and translocate to the nucleus, where it alters the pattern of gene transcription (Figure 15–60B).



**Figure 15–60 The Wnt/β-catenin signalling pathway.** (A) In the absence of a Wnt signal, β-catenin that is not bound to cell-cell adherens junctions (not shown) interacts with a degradation complex containing APC, axin, GSK3, and CK1. In this complex, β-catenin is phosphorylated by CK1 and then by GSK3, triggering its ubiquitylation and degradation in proteasomes. Wnt-responsive genes are kept inactive by the Groucho co-repressor protein bound to the transcription regulator LEF1/TCF. (B) Wnt binding to Frizzled and LRP clusters the two co-receptors together, and the cytosolic tail of LRP is phosphorylated by GSK3 and then by CK1. Axin binds to the phosphorylated LRP and is inactivated and/or degraded, resulting in disassembly of the degradation complex. The phosphorylation of β-catenin is thereby prevented, and unphosphorylated β-catenin accumulates and translocates to the nucleus, where it binds to LEF1/TCF, displaces the co-repressor Groucho, and acts as a coactivator to stimulate the transcription of Wnt target genes. The scaffold protein Dishevelled is required for the signaling pathway to operate; it binds to Frizzled and becomes phosphorylated (not shown), but its precise role is unknown.

In the absence of Wnt signaling, Wnt-responsive genes are kept silent by an inhibitory complex of transcription regulatory proteins. The complex includes proteins of the *LEF1/TCF* family bound to a co-repressor protein of the *Groucho* family (see Figure 15–60A). In response to a Wnt signal,  $\beta$ -catenin enters the nucleus and binds to the LEF1/TCF proteins, displacing Groucho. The  $\beta$ -catenin now functions as a coactivator, inducing the transcription of the Wnt target genes (see Figure 15–60B). Thus, as in the case of Notch signaling, Wnt/ $\beta$ -catenin signaling triggers a switch from transcriptional repression to transcriptional activation.

Among the genes activated by  $\beta$ -catenin is *Myc*, which encodes a protein (Myc) that is an important regulator of cell growth and proliferation (discussed in Chapter 17). Mutations of the *Apc* gene occur in 80% of human colon cancers (discussed in Chapter 20). These mutations inhibit the protein's ability to bind  $\beta$ -catenin, so that  $\beta$ -catenin accumulates in the nucleus and stimulates the transcription of *c-Myc* and other Wnt target genes, even in the absence of Wnt signaling. The resulting uncontrolled cell growth and proliferation promote the development of cancer.

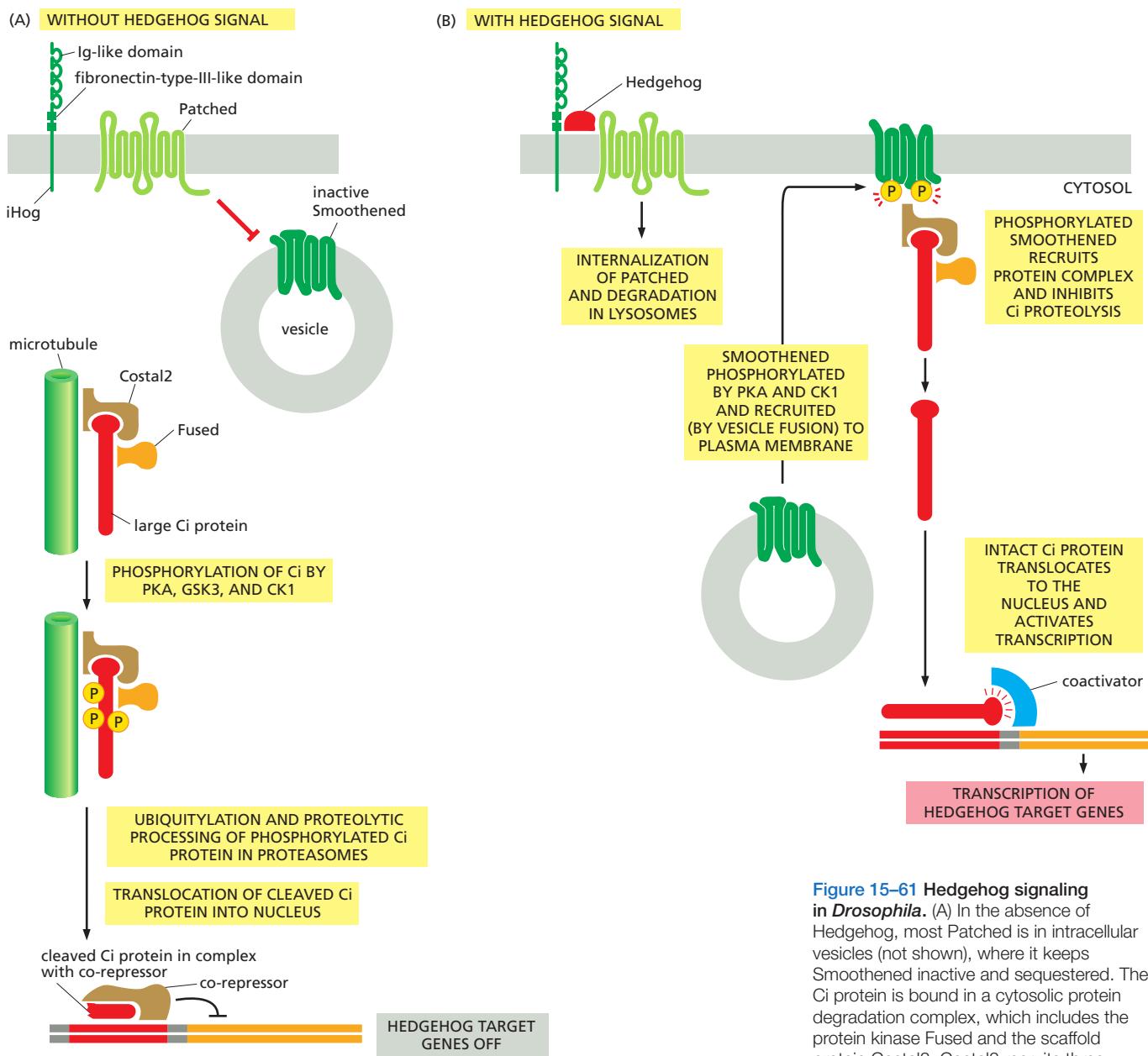
Various secreted inhibitory proteins regulate Wnt signaling in development. Some bind to the LRP receptors and promote their down-regulation, whereas others compete with Frizzled receptors for secreted Wnts. In *Drosophila* at least, Wnts activate negative feedback loops, in which Wnt target genes encode proteins that help shut the response off; some of these proteins inhibit Dishevelled, and others are secreted inhibitors.

### Hedgehog Proteins Bind to Patched, Relieving Its Inhibition of Smoothened

Hedgehog proteins and Wnt proteins act in similar ways. Both are secreted signal molecules, which act as local mediators and morphogens in many developing invertebrate and vertebrate tissues. Both proteins are modified by covalently attached lipids, depend on secreted or cell-surface-bound heparan sulfate proteoglycans (discussed in Chapter 19) for their action, and activate latent transcription regulators by inhibiting their degradation. They both trigger a switch from transcriptional repression to transcriptional activation, and excessive signaling along either pathway in adult cells can lead to cancer. They even use some of the same intracellular signaling proteins and sometimes collaborate to mediate a response.

The **Hedgehog proteins** were discovered in *Drosophila*, where this protein family has only one member. Mutation of the *Hedgehog* gene produces a larva covered with spiky processes (denticles), like a hedgehog. At least three genes encode Hedgehog proteins in vertebrates—*Sonic*, *Desert*, and *Indian hedgehog*. The active forms of all Hedgehog proteins are covalently coupled to cholesterol, as well as to a fatty acid chain. The cholesterol is added during an unusual processing step, in which a precursor protein cleaves itself to produce a smaller, cholesterol-containing signal protein. Most of what we know about the Hedgehog signaling pathway came initially from genetic studies in flies, and it is the fly pathway that we summarize here.

The effects of Hedgehog are mediated by a latent transcription regulator called **Cubitus interruptus (Ci)**, the regulation of which is reminiscent of the regulation of  $\beta$ -catenin by Wnts. In the absence of a Hedgehog signal, Ci is ubiquitylated and proteolytically cleaved in proteasomes. Instead of being completely degraded, however, Ci is processed to form a smaller fragment, which accumulates in the nucleus, where it acts as a transcriptional repressor, helping to keep Hedgehog-responsive genes silent. The proteolytic processing of the Ci protein depends on its phosphorylation by three protein kinases—PKA and two kinases also used in the Wnt pathway, namely GSK3 and CK1. As in the Wnt pathway, the proteolytic processing occurs in a multiprotein complex. The complex includes the protein kinase *Fused* and a scaffold protein *Costal2*, which stably associates with Ci, recruits the three other kinases, and binds the complex to microtubules, thereby keeping unprocessed Ci out of the nucleus (Figure 15–61A).



**Figure 15–61 Hedgehog signaling in *Drosophila*.** (A) In the absence of Hedgehog, most Patched is in intracellular vesicles (not shown), where it keeps Smoothened inactive and sequestered. The Ci protein is bound in a cytosolic protein degradation complex, which includes the protein kinase Fused and the scaffold protein Costal2. Costal2 recruits three other protein kinases (PKA, GSK3, and CK1; not shown), which phosphorylate Ci. Phosphorylated Ci is ubiquitylated and then cleaved in proteasomes (not shown) to form a transcriptional repressor, which accumulates in the nucleus to help keep Hedgehog target genes inactive. (B) Hedgehog binding to iHog and Patched removes the inhibition of Smoothened by Patched. Smoothened is phosphorylated by PKA and CK1 and translocates to the plasma membrane, where it recruits the complex containing Fused, Costal2, and Ci. Costal2 releases unprocessed Ci, which accumulates in the nucleus and activates the transcription of Hedgehog target genes. Many details in the pathway are poorly understood, including the role of Fused.

Hedgehog functions by blocking the proteolytic processing of Ci, thereby changing it into a transcriptional activator. It does this by a convoluted signaling process that depends on three transmembrane proteins: Patched, iHog, and Smoothened. **Patched** is predicted to cross the plasma membrane 12 times, and, although much of it is in intracellular vesicles, some is on the cell surface where it can bind the Hedgehog protein. **iHog** is also on the cell surface and is thought to serve as a co-receptor for Hedgehog. **Smoothened** is a seven-pass transmembrane protein with a structure very similar to a GPCR, but it does not seem to act as a Hedgehog receptor or even as an activator of G proteins; it is controlled by Patched and iHog.

In the absence of a Hedgehog signal, Patched employs an unknown mechanism to keep Smoothened sequestered and inactive in intracellular vesicles (see Figure 15–61A). The binding of Hedgehog to iHog and Patched inhibits the activity of Patched and induces its endocytosis and degradation. The result is that Smoothened is liberated from inhibition and translocates to the plasma membrane, where it recruits the protein complex containing Ci, Fused, and Costal2.

Costal2 is no longer able to bind the other three kinases, and so Ci is no longer cleaved and can now enter the nucleus and activate the transcription of Hedgehog target genes (Figure 15–61B). Among the genes activated by Ci is *Patched* itself; the resulting increase in Patched protein on the cell surface inhibits further Hedgehog signaling—providing another example of negative feedback.

Many gaps remain in our understanding of the Hedgehog signaling pathway. It is not known, for example, how Patched keeps Smoothened inactive and intracellular. As the structure of Patched resembles a transmembrane transporter protein, it has been proposed that it may transport a small molecule into the cell that keeps Smoothened sequestered in vesicles.

Even less is known about the more complex Hedgehog pathway in vertebrate cells. In addition to there being at least three types of vertebrate Hedgehog proteins, there are three Ci-like transcription regulator proteins (*Gli1*, *Gli2*, and *Gli3*) downstream of Smoothened. *Gli2* and *Gli3* are most similar to Ci in structure and function, and *Gli3* has been shown to undergo proteolytic processing like Ci and to act as either a transcriptional repressor or a transcriptional activator. Moreover, in vertebrates, Smoothened, upon activation, becomes localized to the surface of the primary cilium (discussed in Chapter 16), where the Gli proteins are also concentrated, thereby increasing the speed and efficiency of signaling.

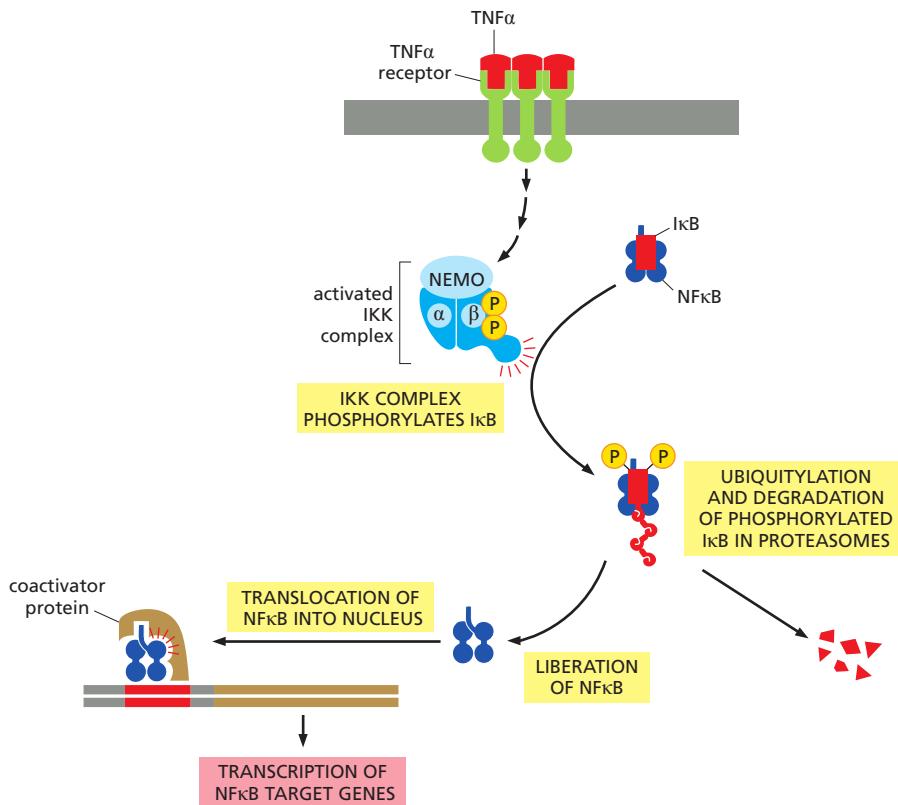
Hedgehog signaling can promote cell proliferation, and excessive Hedgehog signaling can lead to cancer. Inactivating mutations in one of the two human *Patched* genes, for example, which lead to excessive Hedgehog signaling, occur frequently in *basal cell carcinoma* of the skin, the most common form of cancer in Caucasians. A small molecule called *cyclopamine*, made by a meadow lily, is being used to treat cancers associated with excessive Hedgehog signaling. It blocks Hedgehog signaling by binding tightly to Smoothened and inhibiting its activity. It was originally identified because it causes severe developmental defects in the progeny of sheep grazing on such lilies; these include the presence of a single central eye (a condition called *cyclopia*), which is also seen in mice that are deficient in Hedgehog signaling.

### Many Stressful and Inflammatory Stimuli Act Through an NF $\kappa$ B-Dependent Signaling Pathway

The NF $\kappa$ B proteins are latent transcription regulators that are present in most animal cells and are central to many stressful, inflammatory, and innate immune responses. These responses occur as a reaction to infection or injury and help protect stressed multicellular organisms and their cells (discussed in Chapter 24). An excessive or inappropriate inflammatory response in animals can also damage tissue and cause severe pain, and chronic inflammation can lead to cancer; as in the case of Wnt and Hedgehog signaling, excessive NF $\kappa$ B signaling is found in a number of human cancers. NF $\kappa$ B proteins also have important roles during normal animal development: the *Drosophila* NF $\kappa$ B family member *Dorsal*, for example, has a crucial role in specifying the dorsal–ventral axis of the developing fly embryo (discussed in Chapter 22).

Various cell-surface receptors activate the NF $\kappa$ B signaling pathway in animal cells. *Toll receptors* in *Drosophila* and *Toll-like receptors* in vertebrates, for example, recognize pathogens and activate this pathway in triggering innate immune responses (discussed in Chapter 24). The receptors for *tumor necrosis factor  $\alpha$*  (*TNF $\alpha$* ) and *interleukin-1* (*IL1*), which are vertebrate cytokines especially important in inducing inflammatory responses, also activate this signaling pathway. The Toll, Toll-like, and IL1 receptors belong to the same family of proteins, whereas TNF receptors belong to a different family; all of them, however, act in similar ways to activate NF $\kappa$ B. When activated, they trigger a multiprotein ubiquitylation and phosphorylation cascade that releases NF $\kappa$ B from an inhibitory protein complex, so that it can translocate to the nucleus and turn on the transcription of hundreds of genes that participate in inflammatory and innate immune responses.

There are five NF $\kappa$ B proteins in mammals (*RelA*, *RelB*, *c-Rel*, *NFKB1*, and *NFKB2*), and they form a variety of homodimers and heterodimers, each of which



**Figure 15–62** The activation of the **NFκB pathway by TNFα**. Both TNFα and its receptors are trimers. The binding of TNFα causes a rearrangement of the clustered cytosolic tails of the receptors, which now recruit various signaling proteins, resulting in the activation of a protein kinase that phosphorylates and activates IκB kinase kinase (IKK). IKK is a heterotrimer composed of two kinase subunits (IKK $\alpha$  and IKK $\beta$ ) and a regulatory subunit called NEMO. IKK $\beta$  then phosphorylates IκB on two serines, which marks the protein for ubiquitylation and degradation in proteasomes. The released NFκB translocates into the nucleus, where, in collaboration with coactivator proteins, it stimulates the transcription of its target genes.

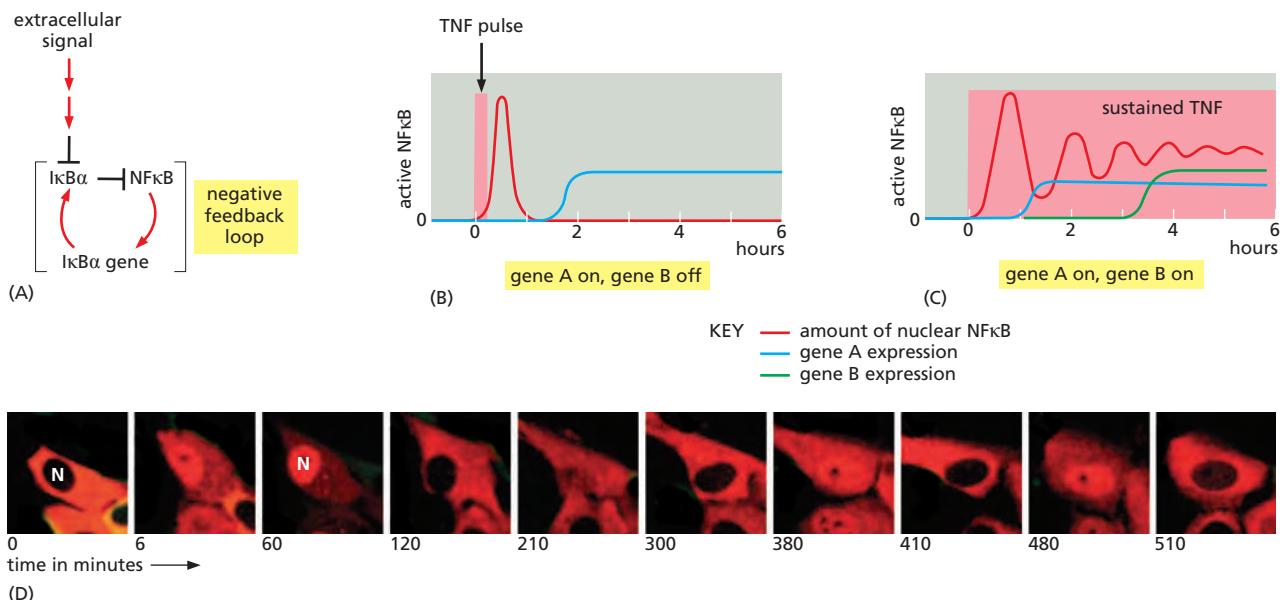
activates its own characteristic set of genes. Inhibitory proteins called IκB bind tightly to the dimers and hold them in an inactive state within the cytoplasm of unstimulated cells. There are three major IκB proteins in mammals (IκB $\alpha$ ,  $\beta$ , and  $\epsilon$ ), and the signals that release NFκB dimers do so by triggering a signaling pathway that leads to the phosphorylation, ubiquitylation, and consequent degradation of the IκB proteins (Figure 15–62).

Among the genes activated by the released NFκB is the gene that encodes IκB $\alpha$ . This activation leads to increased synthesis of IκB $\alpha$  protein, which binds to NFκB and inactivates it, creating a negative feedback loop (Figure 15–63A). Experiments on TNFα-induced responses, as well as computer modeling studies of the responses, indicate that the negative feedback produces two types of NFκB responses, depending on the duration of the TNFα stimulus; importantly, the two types of responses induce different patterns of gene expression (Figure 15–63B, C, and D). The negative feedback through IκB $\alpha$  is required for both types of responses: in cells deficient in IκB $\alpha$ , even a short exposure to TNFα induces a sustained activation of NFκB, without oscillations, and all of the NFκB-responsive genes are activated.

Thus far, we have focused on the mechanisms by which extracellular signal molecules use cell-surface receptors to initiate changes in gene expression. We now turn to a class of extracellular signals that bypasses the plasma membrane entirely and controls, in the most direct way possible, transcription regulatory proteins inside the cell.

### Nuclear Receptors Are Ligand-Modulated Transcription Regulators

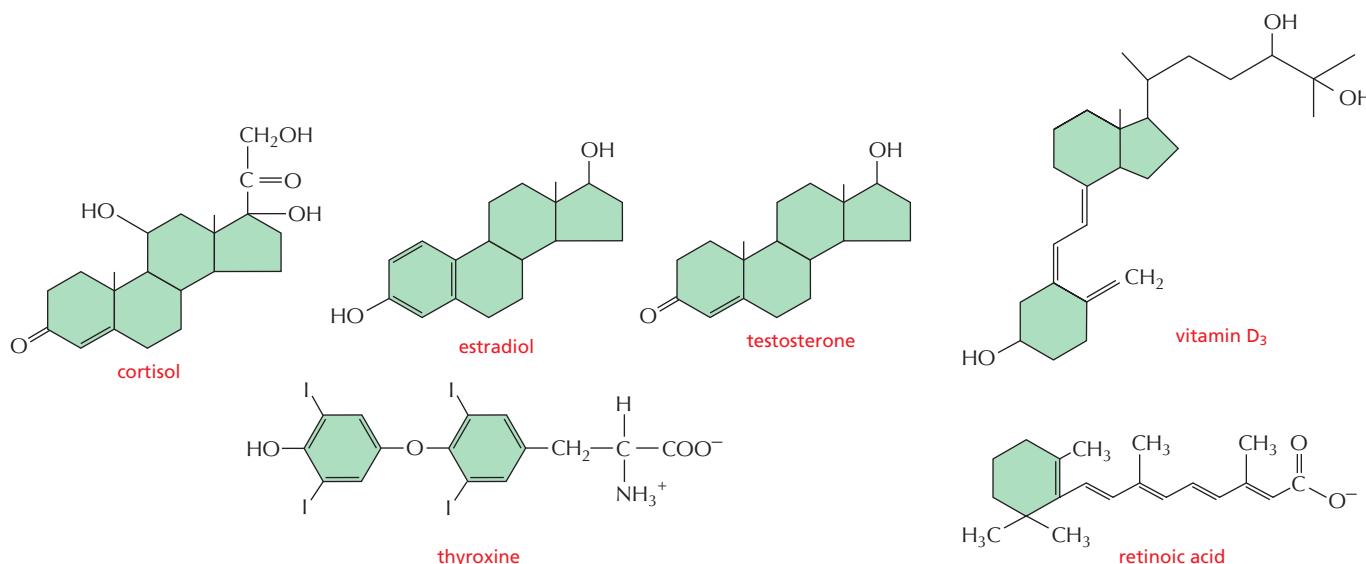
Various small, hydrophobic signal molecules diffuse directly across the plasma membrane of target cells and bind to intracellular receptors that are transcription regulators. These signal molecules include steroid hormones, thyroid hormones, retinoids, and vitamin D. Although they differ greatly from one another in both



chemical structure (Figure 15–64) and function, they all act by a similar mechanism. They bind to their respective intracellular receptor proteins and alter the ability of these proteins to control the transcription of specific genes. Thus, these proteins serve both as intracellular receptors and as intracellular effectors for the signal.

The receptors are all structurally related, being part of the very large **nuclear receptor superfamily**. Many family members have been identified by DNA sequencing only, and their ligand is not yet known; they are therefore referred to as *orphan nuclear receptors*, and they make up large fractions of the nuclear receptors encoded in the genomes of humans, *Drosophila*, and the nematode *C. elegans*. Some mammalian nuclear receptors are regulated by intracellular metabolites rather than by secreted signal molecules; the *peroxisome proliferation-activated receptors* (PPARs), for example, bind intracellular lipid metabolites and regulate the transcription of genes involved in lipid metabolism and fat-cell differentiation. It seems likely that the nuclear receptors for hormones evolved from such receptors for intracellular metabolites, which would help explain their intracellular location.

**Steroid hormones**—which include cortisol, the steroid sex hormones, vitamin D (in vertebrates), and the molting hormone *ecdysone* (in insects)—are all made from cholesterol. *Cortisol* is produced in the cortex of the adrenal glands and influences the metabolism of many types of cells. The **steroid sex hormones** are made in the testes and ovaries and are responsible for the secondary sex characteristics that distinguish males from females. **Vitamin D** is synthesized in the skin in response to sunlight; after it has been converted to its active form in



the liver or kidneys, it regulates  $\text{Ca}^{2+}$  metabolism, promoting  $\text{Ca}^{2+}$  uptake in the gut and reducing its excretion in the kidneys. The *thyroid hormones*, which are made from the amino acid tyrosine, act to increase the metabolic rate of many cell types, while the *retinoids*, such as retinoic acid, are made from vitamin A and have important roles as local mediators in vertebrate development. Although all of these signal molecules are relatively insoluble in water, they are made soluble for transport in the bloodstream and other extracellular fluids by binding to specific carrier proteins, from which they dissociate before entering a target cell (see Figure 15–3B).

The nuclear receptors bind to specific DNA sequences adjacent to the genes that the ligand regulates. Some of the receptors, such as those for cortisol, are located primarily in the cytosol and enter the nucleus only after ligand binding; others, such as the thyroid and retinoid receptors, are bound to DNA in the nucleus even in the absence of ligand. In either case, the inactive receptors are usually bound to inhibitory protein complexes. Ligand binding alters the conformation of the receptor protein, causing the inhibitory complex to dissociate, while also causing the receptor to bind coactivator proteins that stimulate gene transcription (Figure 15–65). In other cases, however, ligand binding to a nuclear receptor inhibits transcription: some thyroid hormone receptors, for example, act as transcriptional activators in the absence of their hormone and become transcriptional repressors when hormone binds.

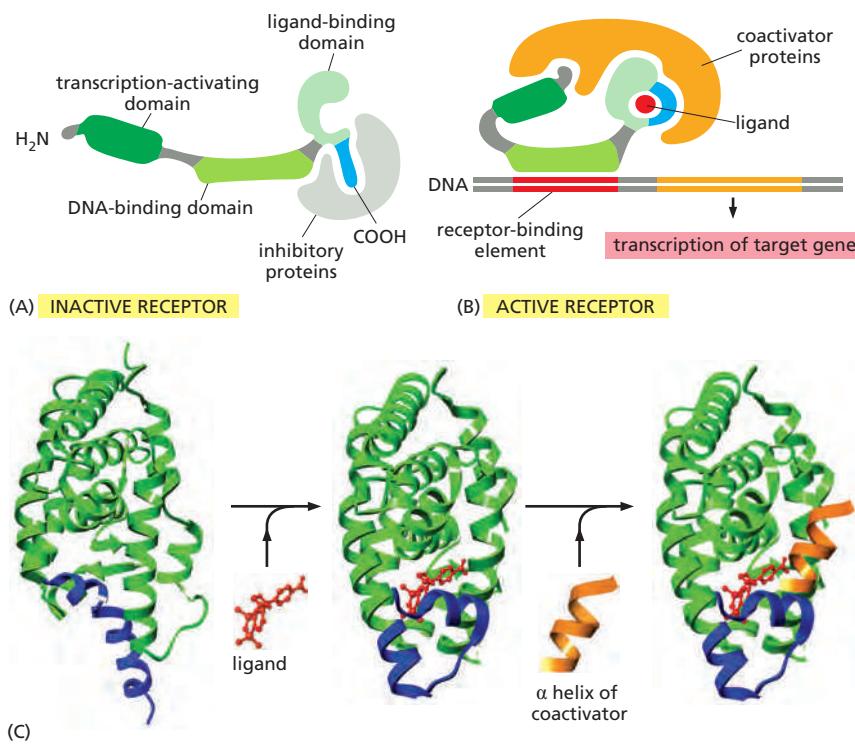
Thus far, we have focused on the control of gene expression by extracellular signal molecules produced by other cells. We now turn to gene regulation by a more global environmental signal: the cycle of light and darkness that results from the Earth's rotation.

### Circadian Clocks Contain Negative Feedback Loops That Control Gene Expression

Life on Earth evolved in the presence of a daily cycle of day and night, and many present-day organisms (ranging from archaea to plants and humans) possess an internal rhythm that dictates different behaviors at different times of day. These behaviors range from the cyclical change in metabolic enzyme activities of a bacterium to the elaborate sleep–wake cycles of humans. The internal oscillators that control such diurnal rhythms are called **circadian clocks**.

Having a circadian clock enables an organism to anticipate the regular daily changes in its environment and take appropriate action in advance. Of course, the internal clock cannot be perfectly accurate, and so it must be capable of being reset by external cues such as the light of day. Thus, circadian clocks keep running

**Figure 15–64** Some signal molecules that bind to intracellular receptors. Note that all of them are small and hydrophobic. The active, hydroxylated form of vitamin D<sub>3</sub> is shown. Estradiol and testosterone are steroid sex hormones.



**Figure 15–65 The activation of nuclear receptors.** All nuclear receptors bind to DNA as either homodimers or heterodimers, but for simplicity we show them as monomers. (A) The receptors all have a related structure, which includes three major domains, as shown. An inactive receptor is bound to inhibitory proteins. (B) Typically, the binding of ligand to the receptor causes the ligand-binding domain of the receptor to clamp shut around the ligand, the inhibitory proteins to dissociate, and coactivator proteins to bind to the receptor's transcription-activating domain, thereby increasing gene transcription. In other cases, ligand binding has the opposite effect, causing co-repressor proteins to bind to the receptor, thereby decreasing transcription (not shown). (C) The structure of the ligand-binding domain of the retinoic acid receptor is shown in the absence (left) and presence (middle) of ligand (shown in red). When ligand binds, the blue  $\alpha$  helix acts as a lid that snaps shut, trapping the ligand in place. The shift in the conformation of the receptor upon ligand binding also creates a binding site for a small  $\alpha$  helix (orange) on the surface of coactivator proteins. (PDB codes: 1LBD, 2ZYQ, and 2ZXZ.)

even when the environmental cues (changes in light and dark) are removed, but the period of this free-running rhythm is generally a little less or more than 24 hours. External signals indicating the time of day cause small adjustments in the running of the clock, so as to keep the organism in synchrony with its environment. Following more drastic shifts, circadian cycles become gradually reset (entrained) by the new cycle of light and dark, as anyone who has experienced jet lag can attest.

We might expect that the circadian clock would be a complex multicellular device, with different groups of cells responsible for different parts of the oscillation mechanism. Remarkably, however, in almost all multicellular organisms, including humans, the timekeepers are individual cells. Thus, a clock that operates in each member of a specialized group of brain cells (the SCN cells in the suprachiasmatic nucleus of the hypothalamus) controls our diurnal cycles of sleeping and waking, body temperature, and hormone release. Even if these cells are removed from the brain and dispersed in a culture dish, they will continue to oscillate individually, showing a cyclic pattern of gene expression with a period of approximately 24 hours. In the intact body, the SCN cells receive neural cues from the retina, entraining the SCN cells to the daily cycle of light and dark; they also send information about the time of day to another brain area, the pineal gland, which relays the time signal to the rest of the body by releasing the hormone melatonin in time with the clock.

Although the SCN cells have a central role as timekeepers in mammals, almost all the other cells in the mammalian body have an internal circadian rhythm, which has the ability to reset in response to light. Similarly, in *Drosophila*, many different types of cells have a similar circadian clock, which continues to cycle when they have been dissected away from the rest of the fly and can be reset by externally imposed light and dark cycles.

The working of circadian clocks, therefore, is a fundamental problem in cell biology. Although we do not yet understand all the details, studies in a wide variety of organisms have revealed the basic principles and molecular components. The key principle is that circadian clocks generally depend on *negative feedback loops*. As discussed earlier, oscillations in the activity of an intracellular signaling protein can occur if that protein inhibits its own activity with a long delay (see

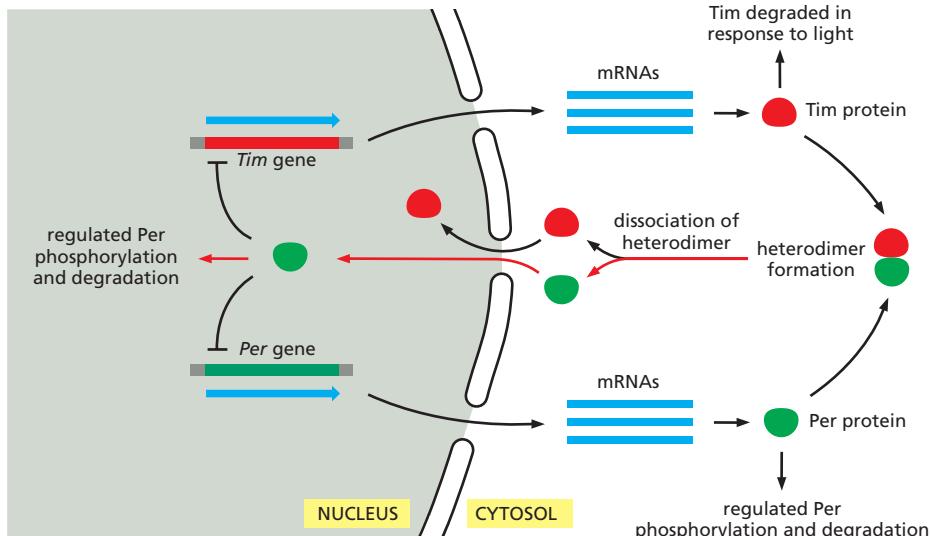


Figure 15–18C and D). In *Drosophila* and many other animals, including humans, the heart of the circadian clock is a delayed negative feedback loop based on transcription regulators: accumulation of certain gene products switches off the transcription of their own genes, but with a delay, so that the cell oscillates between a state in which the products are present and transcription is switched off, and one in which the products are absent and transcription is switched on (Figure 15–66). The negative feedback underlying circadian rhythms does not have to be based on transcription regulators. In some cell types, the circadian clock is constructed of proteins that govern their own activities through post-translational mechanisms, as we discuss next.

### Three Proteins in a Test Tube Can Reconstitute a Cyanobacterial Circadian Clock

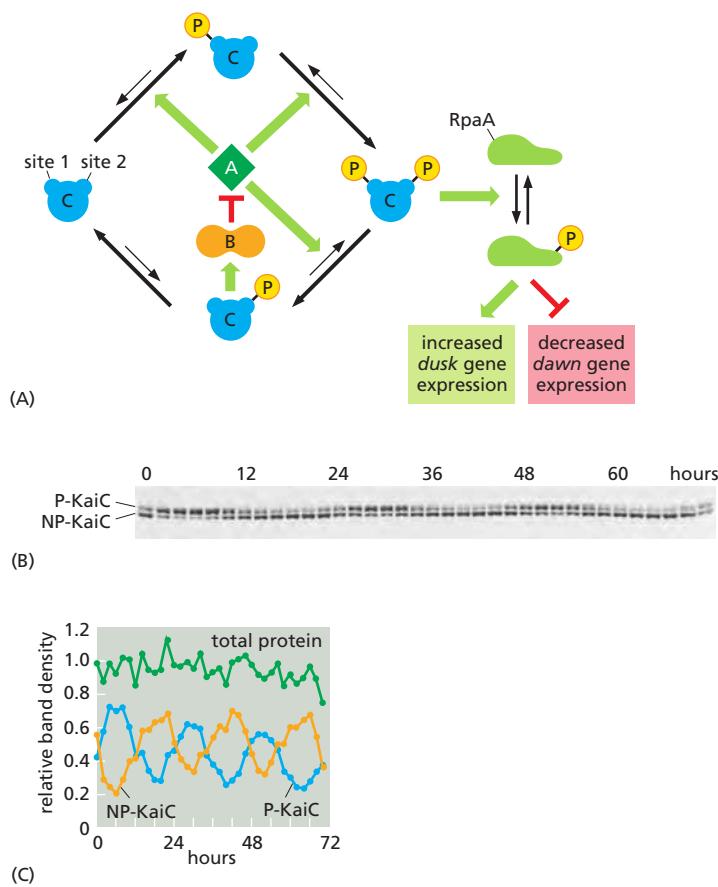
The best understood circadian clock is found in the photosynthetic cyanobacterium, *Synechococcus elongatus*. The core oscillator in this organism is remarkably simple, being composed of just three proteins—*KaiA*, *KaiB*, and *KaiC*. The central player is *KaiC*, a multifunctional enzyme that catalyzes its own phosphorylation and dephosphorylation in a 24-hour cycle: it gradually phosphorylates itself sequentially at two sites during the day and dephosphorylates itself during the night. This timing depends on interactions with the two other *Kai* proteins: *KaiA* binds to unphosphorylated *KaiC* and stimulates *KaiC* autophosphorylation, first at one site and then, with a delay, at the other. The second phosphorylation promotes the binding of the third protein, *KaiB*, which blocks the stimulatory effect of *KaiA* and thereby allows *KaiC* to dephosphorylate itself, bringing *KaiC* back to its dephosphorylated state. This clock depends on a negative feedback loop: *KaiC* drives its own phosphorylation until, after a delay, it recruits an inhibitor, *KaiB*, that stimulates *KaiC* to dephosphorylate itself. Amazingly, when the three *Kai* proteins are purified and incubated in a test tube with ATP, *KaiC* phosphorylation and dephosphorylation occur with roughly 24-hour timing over a period of several days (Figure 15–67).

Circadian oscillations in *KaiC* phosphorylation lead to parallel rhythms in the expression of large numbers of genes involved in controlling metabolic activities and cell division (see Figure 15–67). As a result, many aspects of cell behavior are synchronized with the circadian cycle.

Even in continuous darkness, cyanobacterial cells generate free-running oscillations of *KaiC* phosphorylation with roughly 24-hour periods. As in other circadian clocks, the cyanobacterial clock is entrained by the environmental light/dark

**Figure 15–66** Simplified outline of the mechanism of the circadian clock in *Drosophila* cells. A central feature of the clock is the periodic accumulation and decay of two transcription regulatory proteins, Tim (short for timeless, based on the phenotype of a gene mutation) and Per (short for period). The mRNAs encoding these proteins rise gradually during the day and are translated in the cytosol, where the two proteins associate to form a heterodimer. After a time delay, the heterodimer dissociates and Tim and Per are transported into the nucleus, where Per represses the *Tim* and *Per* genes, resulting in negative feedback that causes the levels of Tim and Per to fall. In addition to this transcriptional feedback, the clock depends on numerous other proteins. For example, the controlled degradation of Per indicated in the diagram imposes delays in the accumulation of Tim and Per, which are crucial to the functioning of the clock. Steps at which specific delays are imposed are shown in red.

Entrainment (or resetting) of the clock occurs in response to new light-dark cycles. Although most *Drosophila* cells do not have true photoreceptors, light is sensed by intracellular flavoproteins, also called cryptochromes. In the presence of light, these proteins associate with the Tim protein and cause its degradation, thereby resetting the clock. (Adapted from J.C. Dunlap, *Science* 311:184–186, 2006.)



cycle. Light is thought to affect the circadian clock indirectly: the activities of Kai proteins are influenced by changes in intracellular redox potential, which occur as a result of increased photosynthetic activity during the day.

### Summary

Some signaling pathways that are especially important in animal development depend on proteolysis to control the activity and location of latent transcription regulatory proteins. Notch receptors are themselves such proteins, which are activated by cleavage when Delta on another cell binds to them; the cleaved cytosolic tail of Notch migrates into the nucleus, where it stimulates the transcription of Notch-responsive genes. In the Wnt/β-catenin signaling pathway, by contrast, the proteolysis of the latent transcription regulatory protein β-catenin is inhibited when a secreted Wnt protein binds to both a Frizzled and LRP receptor protein; as a result, β-catenin accumulates in the nucleus and activates the transcription of Wnt target genes.

Hedgehog signaling in flies works much like Wnt signaling. In the absence of a signal, a bifunctional, cytoplasmic transcription regulator, Ci, is proteolytically cleaved to form a transcriptional repressor that keeps Hedgehog target genes silenced. The binding of Hedgehog to its receptors (Patched and iHog) inhibits the proteolytic processing of Ci; as a result, the intact Ci protein accumulates in the nucleus and activates the transcription of Hedgehog-responsive genes. In Notch, Wnt, and Hedgehog signaling, the extracellular signal triggers a switch from transcriptional repression to transcriptional activation.

Signaling through the latent transcription regulator NFκB also depends on proteolysis. NFκB proteins are normally held in an inactive state by inhibitory IκB proteins in the cytoplasm. A variety of extracellular stimuli, including proinflammatory cytokines, trigger the phosphorylation and ubiquitylation of IκB, marking it for degradation; this enables the NFκB to translocate to the nucleus and activate

**Figure 15–67 The core circadian oscillator of cyanobacteria.** (A) KaiC is a combined kinase and phosphatase that phosphorylates and dephosphorylates itself on two adjacent sites. In the absence of other proteins, the phosphatase activity is dominant, and the protein is mostly unphosphorylated. The binding of KaiA to KaiC suppresses the phosphatase activity and promotes the kinase activity, leading to KaiC phosphorylation, first at site 1 and then at site 2, resulting in diphosphorylated KaiC. KaiC then dephosphorylates itself slowly at site 1, even in the presence of KaiA, so that KaiC is phosphorylated only at site 2. This form of KaiC interacts with KaiB, which blocks the stimulatory effects of KaiA, thereby reducing the rate of KaiC phosphorylation and allowing dephosphorylation to occur. Diphosphorylated KaiC increases in abundance during the day and peaks around dusk. It activates other proteins that phosphorylate a transcription regulator (RpaA), which then stimulates expression of some genes (the *dusk* genes that peak in early evening) and inhibits expression of other genes (the *dawn* genes that peak in the morning). When KaiC dephosphorylation gradually occurs during the night, these effects are reversed: *dusk* genes are turned off and *dawn* genes are turned on.

(B) In this experiment, the three Kai proteins were purified and mixed in a test tube with ATP (which is required for KaiC kinase activity). Every two hours over the next 3 days, the KaiC protein was analyzed by polyacrylamide gel electrophoresis, in which the phosphorylated form of KaiC migrates more slowly (*upper band*, P-KaiC) than the nonphosphorylated form (*lower band*, NP-KaiC). The three different phosphorylated forms of KaiC are not distinguished by this method. The phosphorylation of KaiC oscillates with a roughly 24-hour period. (C) The amount of phosphorylated and unphosphorylated KaiC in the experiment in B is plotted on this graph, along with the amount of total protein. (B and C, from M. Nakajima et al., Science 308:414–415, 2005. With permission from AAAS.)

the transcription of its target genes.  $\text{NF}\kappa\text{B}$  also activates the transcription of the gene that encodes  $\text{IkB}\alpha$ , creating a negative feedback loop, which can produce prolonged oscillations in  $\text{NF}\kappa\text{B}$  activity with sustained extracellular signaling.

Some small, hydrophobic signal molecules, including steroid and thyroid hormones, diffuse across the plasma membrane of the target cell and activate intracellular receptor proteins that directly regulate the transcription of specific genes.

In many cell types, gene expression is governed by circadian clocks, in which delayed negative feedback produces 24-hour oscillations in the activities of transcription regulators, anticipating the cell's changing needs during the day and night.

## SIGNALING IN PLANTS

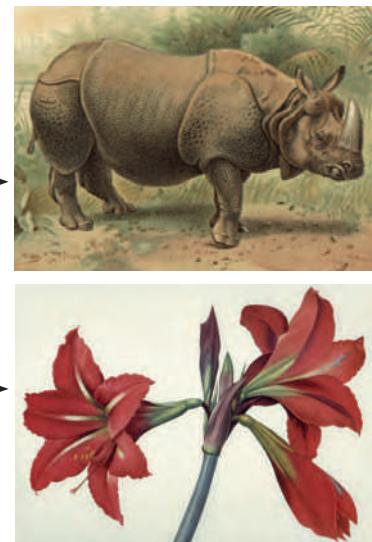
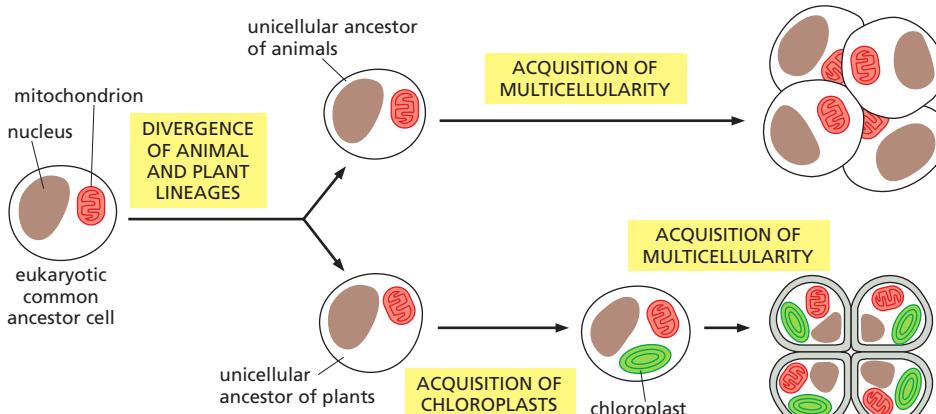
In plants, as in animals, cells are in constant communication with one another. Plant cells communicate to coordinate their activities in response to the changing conditions of light, dark, and temperature, which guide the plant's cycle of growth, flowering, and fruiting. Plant cells also communicate to coordinate activities in their roots, stems, and leaves. In this final section, we consider how plant cells signal to one another and how they respond to light. Less is known about the receptors and intracellular signaling mechanisms involved in cell communication in plants than is known in animals, and we will concentrate mainly on how the receptors and intracellular signaling mechanisms differ from those used by animals.

### Multicellularity and Cell Communication Evolved Independently in Plants and Animals

Although plants and animals are both eukaryotes, they have evolved separately for more than a billion years. Their last common ancestor is thought to have been a unicellular eukaryote that had mitochondria but no chloroplasts; the plant lineage acquired chloroplasts after plants and animals diverged. The earliest fossils of multicellular animals and plants date from almost 600 million years ago. Thus, it seems that plants and animals evolved multicellularity independently, each starting from a different unicellular eukaryote, some time between 1.6 and 0.6 billion years ago (Figure 15–68).

If multicellularity evolved independently in plants and animals, the molecules and mechanisms used for cell communication will have evolved separately and would be expected to be different. There should be some degree of resemblance, however, because the genes in both plants and animals diverged from those contained by their last common unicellular ancestor. Thus, whereas both plants and animals use nitric oxide, cyclic GMP,  $\text{Ca}^{2+}$ , and Rho family GTPases for signaling, there are no homologs of the nuclear receptor family, Ras, JAK, STAT,  $\text{TGF}\beta$ ,

**Figure 15–68** The proposed divergence of plant and animal lineages from a common unicellular eukaryotic ancestor. The plant lineage acquired chloroplasts after the two lineages diverged. Both lineages independently gave rise to multicellular organisms—plants and animals. (Paintings courtesy of John Innes Foundation.)



Notch, Wnt, or Hedgehog encoded by the completely sequenced genome of *Arabidopsis thaliana*, the small flowering plant. Similarly, plants do not seem to use cyclic AMP for intracellular signaling. Nevertheless, the general strategies underlying signaling are frequently very similar in plants and animals. Both, for example, use enzyme-coupled cell-surface receptors, as we now discuss.

### Receptor Serine/Threonine Kinases Are the Largest Class of Cell-Surface Receptors in Plants

Most cell-surface receptors in plants are enzyme-coupled. However, whereas the largest class of enzyme-coupled receptors in animals is the receptor tyrosine kinase (RTK) class, this type of receptor is extremely rare in plants. Instead, plants rely largely on a great diversity of transmembrane *receptor serine/threonine kinases*, which have a typical serine/threonine kinase cytoplasmic domain and an extracellular ligand-binding domain. The most abundant types of these receptors have a tandem array of extracellular leucine-rich repeat structures and are therefore called **leucine-rich repeat (LRR) receptor kinases**.

There are about 175 LRR receptor kinases encoded by the *Arabidopsis* genome. These include a protein called *Bri1*, which forms part of a cell-surface steroid hormone receptor. Plants synthesize a class of steroids that are called **brassinosteroids** because they were originally identified in the mustard family *Brassicaceae*, which includes *Arabidopsis*. These signal molecules regulate the growth and differentiation of plants throughout their life cycle. Binding of a brassinosteroi d to a *Bri1* cell-surface receptor kinase initiates an intracellular signaling cascade that uses a GSK3 protein kinase and a protein phosphatase to regulate the phosphorylation and degradation of specific transcription regulatory proteins in the nucleus, and thereby specific gene transcription. Mutant plants that are deficient in the *Bri1* receptor kinase are insensitive to brassinosteroids and are therefore dwarfs.

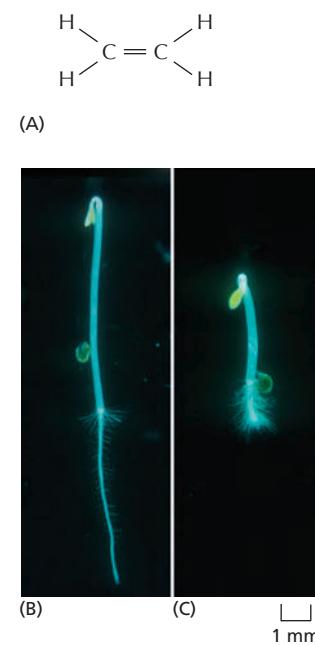
The LRR receptor kinases are only one of many classes of transmembrane receptor serine/threonine kinases in plants. There are at least six additional families, each with its own characteristic set of extracellular domains. The *lectin receptor kinases*, for example, have extracellular domains that bind carbohydrate signal molecules. The *Arabidopsis* genome encodes over 300 receptor serine/threonine kinases, which makes them the largest family of receptors known in plants. Many are involved in defense responses against pathogens.

### Ethylene Blocks the Degradation of Specific Transcription Regulatory Proteins in the Nucleus

Various **plant growth regulators** (also called **plant hormones**) help to coordinate plant development. They include *ethylene*, *auxin*, *cytokinins*, *gibberellins*, and *abscisic acid*, as well as brassinosteroids. Growth regulators are all small molecules made by most plant cells. They diffuse readily through cell walls and can either act locally or be transported to influence cells further away. Each growth regulator can have multiple effects. The specific effect depends on environmental conditions, the nutritional state of the plant, the responsiveness of the target cells, and which other growth regulators are acting.

**Ethylene** is an important example. This small gas molecule (Figure 15–69A) can influence plant development in various ways; it can, for example, promote fruit ripening, leaf abscission, and plant senescence. It also functions as a stress signal in response to wounding, infection, flooding, and so on. When the shoot of a germinating seedling, for instance, encounters an obstacle, ethylene promotes a complex response that allows the seedling to safely bypass the obstacle (Figure 15–69B and C).

Plants have various ethylene receptors, which are located in the endoplasmic reticulum and are all structurally related. They are dimeric, multipass transmembrane proteins, with a copper-containing ethylene-binding domain and a domain that interacts with a cytoplasmic protein called *CTR1*, which is closely related



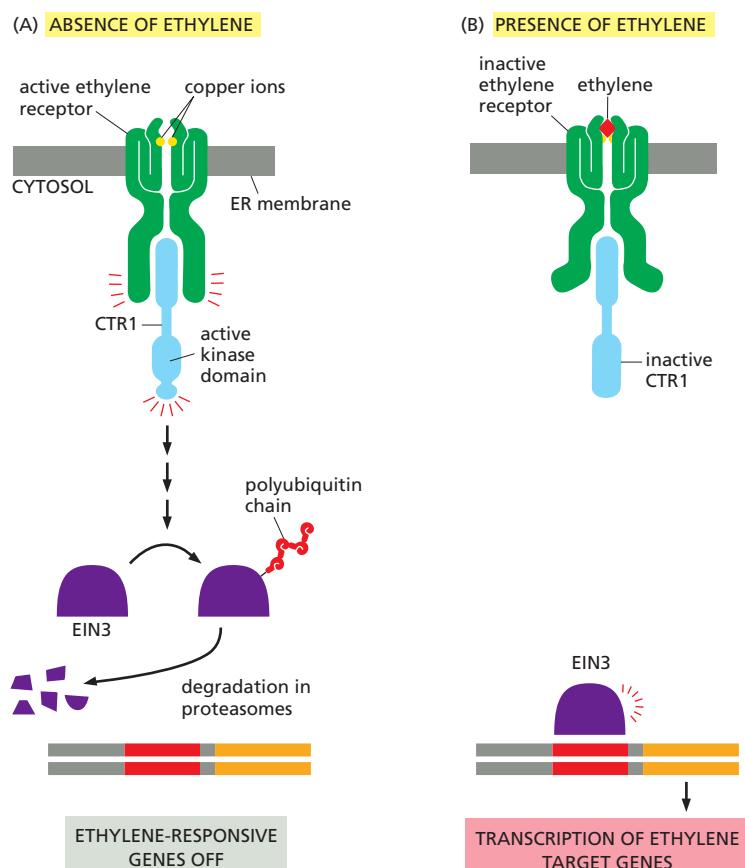
**Figure 15–69** The ethylene-mediated triple response that occurs when the growing shoot of a germinating seedling encounters an obstacle underground. (A) The structure of ethylene. (B) In the absence of obstacles, the shoot grows upward and is long and thin. (C) If the shoot encounters an obstacle, such as a piece of gravel in the soil, the seedling responds to the encounter in three ways. First, it thickens its stem, which can then exert more force on the obstacle. Second, it shields the tip of the shoot (at top) by increasing the curvature of a specialized hook structure. Third, it reduces the shoot's tendency to grow away from the direction of gravity, so as to avoid the obstacle. (Courtesy of Melanie Webb.)

in sequence to the Raf MAP kinase kinase kinase discussed earlier (see Figure 15–49). Surprisingly, it is the empty receptors that are active and keep CTR1 active. By an unknown signaling mechanism, active CTR1 stimulates the ubiquitylation and degradation in proteasomes of a nuclear transcription regulator called *EIN3*, which is required for the transcription of ethylene-responsive genes. In this way, the empty but active receptors keep ethylene-response genes off. Ethylene binding inactivates the receptors, altering their conformation so that they no longer activate CTR1. The *EIN3* protein is no longer ubiquitylated and degraded and can now activate the transcription of the large number of ethylene-responsive genes (Figure 15–70).

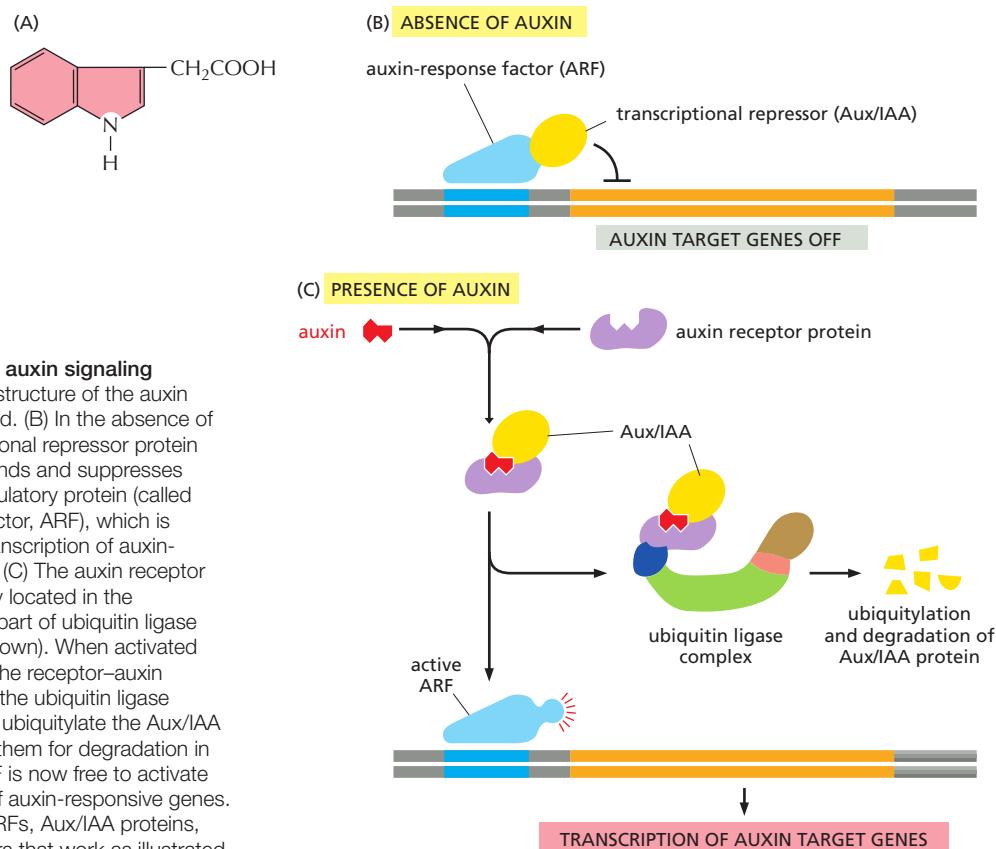
### Regulated Positioning of Auxin Transporters Patterns Plant Growth

The plant hormone **auxin**, which is generally indole-3-acetic acid (Figure 15–71A), binds to receptor proteins in the nucleus. It helps plants grow toward light, grow upward rather than branch out, and grow their roots downward. It also regulates organ initiation and positioning and helps plants flower and bear fruit. Like ethylene (and like some of the animal signal molecules we have described in this chapter), auxin influences gene expression by controlling the degradation of transcription regulators. It works by stimulating the ubiquitylation and degradation of repressor proteins that block the transcription of auxin target genes in unstimulated cells (Figure 15–71B and C).

Auxin is unique in the way that it is transported. Unlike animal hormones, which are usually secreted by a specific endocrine organ and transported to target



**Figure 15–70** The ethylene signaling pathway. (A) In the absence of ethylene, the receptors and CTR1 are active, causing the ubiquitylation and destruction of EIN3, the transcription regulatory protein in the nucleus that is responsible for the transcription of ethylene-responsive genes. (B) The binding of ethylene inactivates the receptors and disrupts the activation of CTR1. The EIN3 protein is not degraded and can therefore activate the transcription of ethylene-responsive genes.



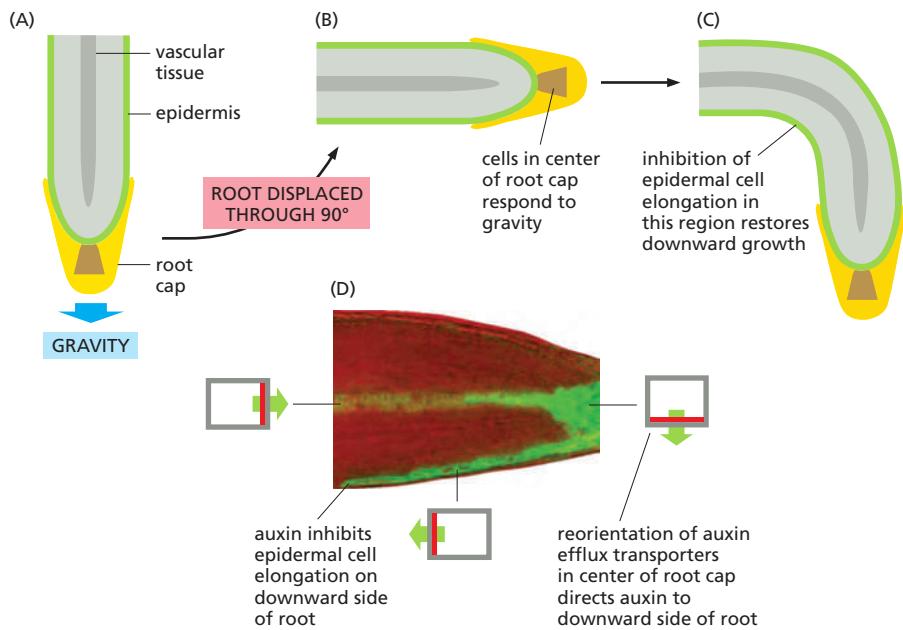
**Figure 15–71** The auxin signaling pathway. (A) The structure of the auxin indole-3-acetic acid. (B) In the absence of auxin, a transcriptional repressor protein (called Aux/IAA) binds and suppresses a transcription regulatory protein (called auxin-response factor, ARF), which is required for the transcription of auxin-responsive genes. (C) The auxin receptor proteins are mainly located in the nucleus and form part of ubiquitin ligase complexes (not shown). When activated by auxin binding, the receptor–auxin complexes recruit the ubiquitin ligase complexes, which ubiquitylate the Aux/IAA proteins, marking them for degradation in proteasomes. ARF is now free to activate the transcription of auxin-responsive genes. There are many ARFs, Aux/IAA proteins, and auxin receptors that work as illustrated.

cells via the circulatory system, auxin has its own transport system. Specific plasma-membrane-bound *influx transporter proteins* and *efflux transporter proteins* move auxin into and out of plant cells, respectively. The efflux transporters can be distributed asymmetrically in the plasma membrane to make the efflux of auxin directional. A row of cells with their auxin efflux transporters confined to the basal plasma membrane, for example, will transport auxin from the top of the plant to the bottom.

In some regions of the plant, the localization of the auxin transporters, and therefore the direction of auxin flow, is highly dynamic and regulated. A cell can rapidly redistribute transporters by controlling the traffic of vesicles containing them. The auxin efflux transporters, for example, normally recycle between intracellular vesicles and the plasma membrane. A cell can redistribute these transporters on its surface by inhibiting their endocytosis in one domain of the plasma membrane, causing the transporters to accumulate there. One example occurs in the root, where gravity influences the direction of growth. The auxin efflux transporters are normally distributed symmetrically in the cap cells of the root. Within minutes of a change in the direction of the gravity vector, however, the efflux transporters redistribute to one side of the cells, so that auxin is pumped out toward the side of the root pointing downward. Because auxin inhibits root-cell elongation, this redirection of auxin transport causes the root tip to reorient, so that it grows downward again (Figure 15–72).

### Phytochromes Detect Red Light, and Cryptochromes Detect Blue Light

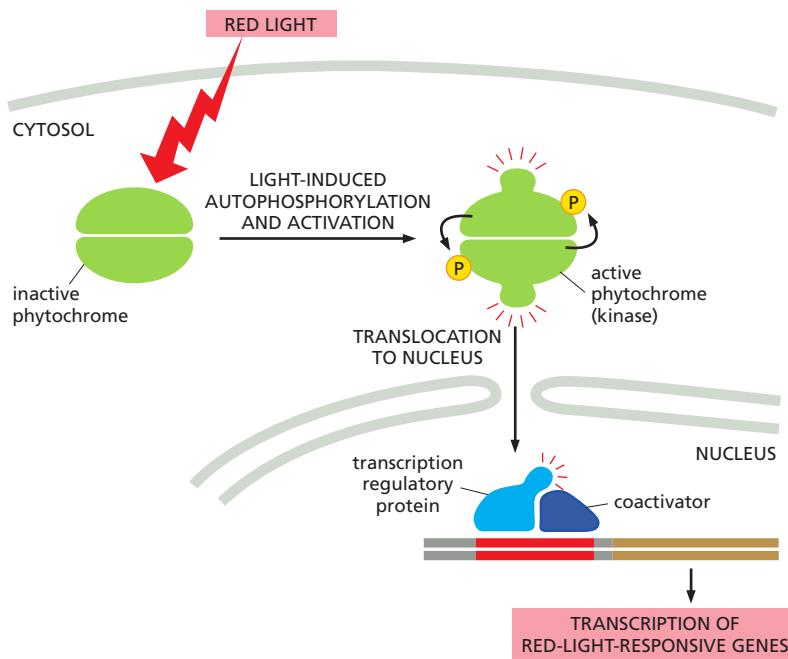
Plant development is greatly influenced by environmental conditions. Unlike animals, plants cannot move when conditions become unfavorable; they have to adapt or they die. The most important environmental influence on plants is



**Figure 15–72 Auxin transport and root gravitropism.** (A–C) Roots respond to a 90° change in the gravity vector and adjust their direction of growth so that they grow downward again. The cells that respond to gravity are in the center of the root cap, while it is the epidermal cells further back (on the lower side) that decrease their rate of elongation to restore downward growth. (D) The gravity-responsive cells in the root cap redistribute their auxin efflux transporters in response to the displacement of the root. This redirects the auxin flux mainly to the lower part of the displaced root, where it inhibits the elongation of the epidermal cells. The resulting asymmetrical distribution of auxin in the *Arabidopsis* root tip shown here is assessed indirectly, using an auxin-responsive reporter gene that encodes a protein fused to green fluorescent protein (GFP); the epidermal cells on the downward side of the root are green, whereas those on the upper side are not, reflecting the asymmetrical distribution of auxin. The distribution of auxin efflux transporters in the plasma membrane of cells in different regions of the root (shown as gray rectangles) is indicated in red, and the direction of auxin efflux is indicated by a green arrow. (The fluorescence photograph in D is from T. Paciorek et al., *Nature* 435:1251–1256, 2005. With permission from Macmillan Publishers Ltd.)

light, which is their energy source and has a major role throughout their entire life cycle—from germination, through seedling development, to flowering and senescence. Plants have thus evolved a large set of light-sensitive proteins to monitor the quantity, quality, direction, and duration of light. These are usually referred to as *photoreceptors*. However, because the term photoreceptor is also used for light-sensitive cells in the animal retina (see Figure 15–38), we shall use the term *photoprotein* instead.

All photoproteins sense light by means of a covalently attached light-absorbing chromophore, which changes its shape in response to light and then induces a change in the protein's conformation. The best-known plant photoproteins are the **phytochromes**, which are present in all plants and in some algae but are absent in animals. These are dimeric, cytoplasmic serine/threonine kinases, which respond differentially and reversibly to red and far-red light: whereas red light usually activates the kinase activity of the phytochrome, far-red light inactivates it. When activated by red light, the phytochrome is thought to phosphorylate itself and then to phosphorylate one or more other proteins in the cell. In some light responses, the activated phytochrome translocates into the nucleus, where it activates transcription regulators to alter gene transcription (Figure 15–73). In other cases, the activated phytochrome activates a latent transcription regulator in the cytoplasm, which then translocates into the nucleus to regulate gene transcription. In still other cases, the photoprotein triggers signaling pathways in the cytosol that alter the cell's behavior without involving the nucleus.



**Figure 15–73** One way in which phytochromes mediate a light response in plant cells. When activated by red light, the phytochrome, which is a dimeric protein kinase, phosphorylates itself and then moves into the nucleus, where it activates transcription regulatory proteins to stimulate the transcription of red-light-responsive genes.

Plants sense blue light using photoproteins of two other sorts, phototropin and cryptochromes. **Phototropin** is associated with the plasma membrane and is partly responsible for *phototropism*, the tendency of plants to grow toward light. Phototropism occurs by directional cell elongation, which is stimulated by auxin, but the links between phototropin and auxin are unknown.

**Cryptochromes** are flavoproteins that are sensitive to blue light. They are structurally related to blue-light-sensitive enzymes called *photolyases*, which are involved in the repair of ultraviolet-induced DNA damage in all organisms, except most mammals. Unlike phytochromes, cryptochromes are also found in animals, where they have an important role in circadian clocks (see Figure 15–66). Although cryptochromes are thought to have evolved from the photolyases, they do not have a role in DNA repair.

## Summary

Plants and animals are thought to have evolved multicellularity and cell communication mechanisms independently, each starting from a different unicellular eukaryote, which in turn evolved from a common unicellular eukaryotic ancestor. Not surprisingly, therefore, the mechanisms used to signal between cells in animals and in plants have both similarities and differences. Whereas animals rely heavily on GPCRs and RTKs, plants rely mainly on enzyme-coupled receptors of the receptor serine/threonine kinase type, especially ones with extracellular leucine-rich repeats. Various plant hormones, or growth regulators, including ethylene and auxin, help coordinate plant development. Ethylene acts through intracellular receptors to stop the degradation of specific nuclear transcription regulators, which can then activate the transcription of ethylene-responsive genes. The receptors for some other plant hormones, including auxin, also regulate the degradation of specific transcription regulators, although the details vary. Auxin signaling is unusual in that it has its own highly regulated transport system, in which the dynamic positioning of plasma-membrane-bound auxin transporters controls the direction of auxin flow and thereby the direction of plant growth. Light has an important role in regulating plant development. These light responses are mediated by a variety of light-sensitive photoproteins, including phytochromes, which are responsive to red light, and cryptochromes and phototropin, which are sensitive to blue light.

## WHAT WE DON'T KNOW

- How does a cell integrate the information received from its many different cell-surface receptors to make all-or-none decisions?
- Much of what we know about cell signaling comes from biochemical studies of isolated proteins in test tubes. What is the precise quantitative behavior of intracellular signaling networks in an intact cell, or in an intact animal, where countless other signals and cell components might influence signaling specificity and strength?
- How do intracellular signaling circuits generate specific and dynamic signaling patterns such as oscillations and waves, and how are these patterns sensed and interpreted by the cell?
- Scaffold proteins and activated receptor tyrosine kinases nucleate the assembly of large intracellular signaling complexes. What is the dynamic behavior of these complexes, and how does this behavior influence downstream signaling?

## PROBLEMS

Which statements are true? Explain why or why not.

**15–1** All second messengers are water-soluble and diffuse freely through the cytosol.

**15–2** In the regulation of molecular switches, protein kinases and guanine nucleotide exchange factors (GEFs) always turn proteins on, whereas protein phosphatases and GTPase-activating proteins (GAPs) always turn proteins off.

**15–3** Most intracellular signaling pathways provide numerous opportunities for amplifying the responses to extracellular signals.

**15–4** Binding of extracellular ligands to receptor tyrosine kinases (RTKs) activates the intracellular catalytic domain by propagating a conformational change across the lipid bilayer through a single transmembrane  $\alpha$  helix.

**15–5** Protein tyrosine phosphatases display exquisite specificity for their substrates, unlike most serine/threonine protein phosphatases, which have rather broad specificity.

**15–6** Even though plants and animals independently evolved multicellularity, they use virtually all the same signaling proteins and second messengers for cell-cell communication.

### Discuss the following problems.

**15–7** Suppose that the circulating concentration of hormone is  $10^{-10}$  M and the  $K_d$  for binding to its receptor is  $10^{-8}$  M. What fraction of the receptors will have hormone bound? If a meaningful physiological response occurs when 50% of the receptors have bound a hormone molecule, how much will the concentration of hormone have to rise to elicit a response? The fraction of receptors (R) bound to hormone (H) to form a receptor-hormone complex (R-H) is  $[R-H]/([R] + [R-H]) = [R-H]/[R]_{TOT} = [H]/([H] + K_d)$ .

**15–8** Cells communicate in ways that resemble human communication. Decide which of the following forms of human communication are analogous to autocrine, paracrine, endocrine, and synaptic signaling by cells.

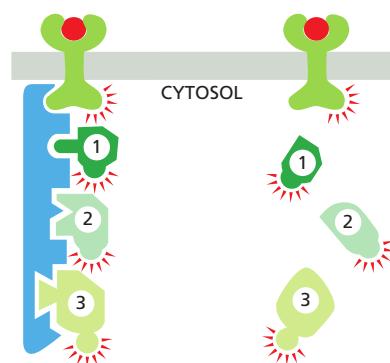
- A. Telephone conversation
- B. Talking to people at a cocktail party
- C. A radio announcement
- D. Talking to yourself

**15–9** Why do signaling responses that involve changes in proteins already present in the cell occur in milliseconds to seconds, whereas responses that require changes in gene expression require minutes to hours?

**15–10** How is it that different cells can respond in different ways to exactly the same signaling molecule even when they have identical receptors?

**15–11** Why do you suppose that phosphorylation/dephosphorylation, as opposed to allosteric binding of small molecules, for example, has evolved to play such a prominent role in switching proteins on and off in signaling pathways?

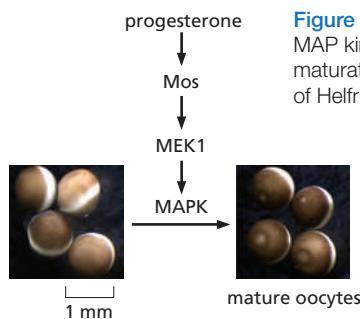
**15–12** Consider a signaling pathway that proceeds through three protein kinases that are sequentially activated by phosphorylation. In one case, the kinases are held in a signaling complex by a scaffolding protein; in the other, the kinases are freely diffusible (Figure Q15–1). Discuss the properties of these two types of organization in terms of signal amplification, speed, and potential for cross-talk between signaling pathways.



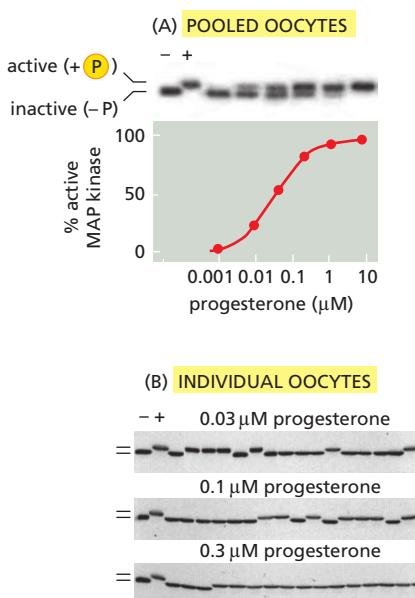
**Figure Q15–1** A kinase cascade organized by a scaffolding protein or composed of freely diffusing components (Problem 15–12).

**15–13** Describe three ways in which a gradual increase in an extracellular signal can be sharpened by the target cell to produce an abrupt or nearly all-or-none response.

**15–14** Activation (“maturation”) of frog oocytes is signaled through a MAP kinase signaling module. An increase in the hormone progesterone triggers the module by stimulating the translation of Mos mRNA, which is the frog’s MAP kinase kinase kinase (Figure Q15–2). Maturation is easy to score visually by the presence of a white spot in the middle of the brown surface of the oocyte (see Figure Q15–2). To determine the dose-response curve for progesterone-induced activation of MAP kinase, you place 16 oocytes in each of six plastic dishes and add various concentrations of progesterone. After an overnight incubation, you crush the oocytes, prepare an extract, and determine the state of MAP kinase phosphorylation (hence, activation) by SDS polyacrylamide-gel electrophoresis (Figure Q15–3A). This analysis shows a graded response of MAP kinase to increasing concentrations of progesterone.



**Figure Q15–2** Progesterone-induced MAP kinase activation, leading to oocyte maturation (Problem 15–14). (Courtesy of Helfrid Hochegger.)

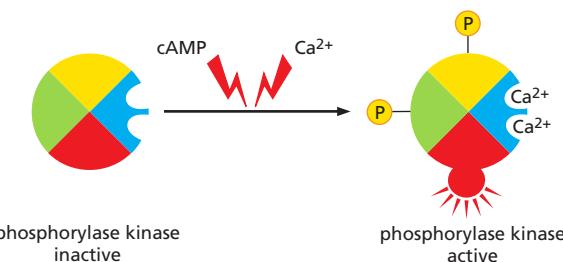


**Figure Q15-3** Activation of frog oocytes (Problem 15–14). (A) Phosphorylation of MAP kinase in pooled oocytes. (B) Phosphorylation of MAP kinase in individual oocytes. MAP kinase was detected by immunoblotting using a MAP-kinase-specific antibody. The first two lanes in each gel contain nonphosphorylated, inactive MAP kinase (−) and phosphorylated, active MAP kinase (+). (From J.E. Ferrell, Jr., and E.M. Machleder, *Science* 280:895–898, 1998. With permission from AAAS.)

Before you crushed the oocytes, you noticed that not all oocytes in individual dishes had white spots. Had some oocytes undergone partial activation and not yet reached the white-spot stage? To answer this question, you repeat the experiment, but this time you analyze MAP kinase activation in individual oocytes. You are surprised to find that each oocyte has either a fully activated or a completely inactive MAP kinase (Figure Q15-3B). How can an all-or-none response in individual oocytes give rise to a graded response in the population?

**15–15** Propose specific types of mutations in the gene for the regulatory subunit of cyclic-AMP-dependent protein kinase (PKA) that could lead to either a permanently active PKA or a permanently inactive PKA.

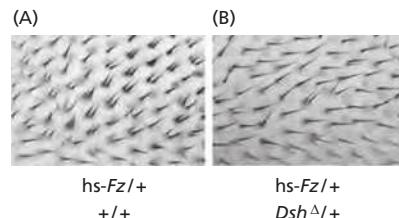
**15–16** Phosphorylase kinase integrates signals from the cyclic-AMP-dependent and  $\text{Ca}^{2+}$ -dependent signaling pathways that control glycogen breakdown in liver and muscle cells (Figure Q15-4). Phosphorylase kinase is composed of four subunits. One is the protein kinase that catalyzes the addition of phosphate to glycogen phosphorylase to activate it for glycogen breakdown. The other three subunits are regulatory proteins that control the activity of the



**Figure Q15-4** Integration of cyclic-AMP-dependent and  $\text{Ca}^{2+}$ -dependent signaling pathways by phosphorylase kinase in liver and muscle cells (Problem 15–16).

catalytic subunit. Two contain sites for phosphorylation by PKA, which is activated by cyclic AMP. The remaining subunit is calmodulin, which binds  $\text{Ca}^{2+}$  when the cytosolic  $\text{Ca}^{2+}$  concentration rises. The regulatory subunits control the equilibrium between the active and inactive conformations of the catalytic subunit, with each phosphate and  $\text{Ca}^{2+}$  nudging the equilibrium toward the active conformation. How does this arrangement allow phosphorylase kinase to serve its role as an integrator protein for the multiple pathways that stimulate glycogen breakdown?

**15–17** The Wnt planar polarity signaling pathway normally ensures that each wing cell in *Drosophila* has a single hair. Overexpression of the *Frizzled* gene from a heat-shock promoter (hs-*Fz*) causes multiple hairs to grow from many cells (Figure Q15-5A). This phenotype is suppressed if hs-*Fz* is combined with a heterozygous deletion (*Dsh* $^{\Delta 1}$ ) of the *Dishevelled* gene (Figure Q15-5B). Do these results allow you to order the action of Frizzled and Dishevelled in the signaling pathway? If so, what is the order? Explain your reasoning.



**Figure Q15-5** Pattern of hair growth on wing cells in genetically different *Drosophila* (Problem 15–17). (From C.G. Winter et al., *Cell* 105:81–91, 2001. With permission from Elsevier.)

## REFERENCES

### General

- Marks F, Klingmüller U & Müller-Decker K (2009) Cellular Signal Processing: An Introduction to the Molecular Mechanisms of Signal Transduction. New York: Garland Science.  
Lim W, Mayer B & Pawson T (2015) Cell Signaling: Principles and Mechanisms. New York: Garland Science.

### Principles of Cell Signaling

- Alon U (2007) Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.

- Ben-Shlomo I, Yu Hsu S, Rauch R et al (2003) Signaling receptors: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci. STKE* 187, RE9.  
Endicott JA, Noble ME & Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. *Annu. Rev. Biochem.* 81, 587–613.  
Ferrell JE, Jr (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* 14, 140–148.

- Good MC, Zalatan JG & Lim WA (2011) Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332, 680–686.
- Ladbury JE & Arold ST (2012) Noise in cellular signaling pathways: causes and effects. *Trends Biochem. Sci.* 37, 173–178.
- Mehta S & Zhang J (2011) Reporting from the field: genetically encoded fluorescent reporters uncover signaling dynamics in living biological systems. *Annu. Rev. Biochem.* 80, 375–401.
- Pires-daSilva A & Sommer RJ (2003) The evolution of signalling pathways in animal development. *Nature Rev. Genet.* 4, 39–49.
- Rosse C, Linch M, Kermorgant S, Cameron AJ et al. (2010) PKC and the control of localized signal dynamics. *Nat. Rev. Mol. Cell Biol.* 11, 103–112.
- Scott JD & Pawson T (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science* 326, 1220–1224.
- Seet BT, Dikic I, Zhou MM & Pawson T (2006) Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* 7, 473–483.
- Tyson JJ, Chen KC & Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231.
- Ubersax JA & Ferrell JE, Jr (2007) Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* 8, 530–541.
- Wittinghofer A & Vetter IR (2011) Structure-function relationships of the G domain, a canonical switch motif. *Annu. Rev. Biochem.* 80, 943–971.
- ### Signaling Through G-protein-coupled Receptors
- Audet M & Bouvier M (2012) Restructuring G-protein-coupled receptor activation. *Cell* 151, 14–23.
- Berridge MJ, Bootman MD & Roderick HL (2003) Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* 4, 517–529.
- Breer H (2003) Sense of smell: recognition and transduction of olfactory signals. *Biochem. Soc. Trans.* 31, 113–116.
- Hoeflich KP & Ikura M (2002) Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* 108, 739–742.
- Kamenetsky M, Middelhaufe S, Bank EM et al. (2006) Molecular details of cAMP generation in mammalian cells: a tale of two systems. *J. Mol. Biol.* 362, 623–639.
- McConnachie G, Langeberg LK & Scott JD (2006) AKAP signaling complexes: getting to the heart of the matter. *Trends Mol. Med.* 12, 317–323.
- Murad F (2006) Shattuck Lecture. Nitric oxide and cyclic GMP in cell signaling and drug development. *N. Engl. J. Med.* 355, 2003–2011.
- Parker PJ (2004) The ubiquitous phosphoinositides. *Biochem. Soc. Trans.* 32, 893–898.
- Rasmussen SG, DeVree BT, Zou Y et al. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477, 549–555.
- Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annu. Rev. Biochem.* 70, 281–312.
- Robishaw JD & Berlot CH (2004) Translating G protein subunit diversity into functional specificity. *Curr. Opin. Cell Biol.* 16, 206–209.
- Rosenbaum DM, Rasmussen SG & Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363.
- Shenoy SK & Lefkowitz RJ (2011) beta-Arrestin-mediated receptor trafficking and signal transduction. *Trends Pharmacol. Sci.* 32, 521–533.
- Sorkin A & von Zastrow M (2009) Endocytosis and signalling: intertwining molecular networks. *Nat. Rev. Mol. Cell Biol.* 10, 609–622.
- Stratton MM, Chao LH, Schulman H & Kuriyan J (2013) Structural studies on the regulation of Ca<sup>2+</sup>/calmodulin dependent protein kinase II. *Curr. Opin. Struct. Biol.* 23, 292–301.
- Sung CH & Chuang JZ (2010) The cell biology of vision. *J. Cell Biol.* 190, 953–963.
- Willoughby D & Cooper DM (2008) Live-cell imaging of cAMP dynamics. *Nat. Methods* 5, 29–36.
- ### Signaling Through Enzyme-coupled Receptors
- Jura N, Zhang X, Endres NF et al. (2011) Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. *Mol. Cell* 42, 9–22.
- Lemmon MA & Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. *Cell* 141, 1117–1134.
- Massagué J (2012) TGF $\beta$  signalling in context. *Nat. Rev. Mol. Cell Biol.* 13, 616–630.
- Manning BD & Cantley LC (2007) AKT/PKB signaling: navigating downstream. *Cell* 129, 1261–1274.
- Mitin N, Rossman KL & Der CJ (2005) Signaling interplay in Ras superfamily function. *Curr. Biol.* 15, R563–R574.
- Pitulescu ME & Adams RH (2010) Eph/ephrin molecules—a hub for signaling and endocytosis. *Genes Dev.* 24, 2480–2492.
- Qi M & Elion EA (2005) MAP kinase pathways. *J. Cell Sci.* 118, 3569–3572.
- Rawlings JS, Rosler KM & Harrison DA (2004) The JAK/STAT signaling pathway. *J. Cell Sci.* 117, 1281–1283.
- Roskoski R, Jr (2004) Src protein-tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* 324, 1155–1164.
- Schwartz MA & Madhani HD (2004) Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* 35, 725–748.
- Shaw RJ & Cantley LC (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 44, 424–430.
- Zoncu R, Efeyan A & Sabatini DM (2011) mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat. Rev. Mol. Cell Biol.* 12, 21–35.
- ### Alternative Signaling Routes in Gene Regulation
- Bray SJ (2006) Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689.
- Briscoe J & Therond PP (2013) The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* 14, 416–429.
- Hoffmann A & Baltimore D (2006) Circuitry of nuclear factor  $\kappa$ B signaling. *Immunolet.* 210, 171–186.
- Huang P, Chandra V & Rastinejad F (2010) Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu. Rev. Physiol.* 72, 247–272.
- Markson JS & O'Shea EK (2009) The molecular clockwork of a protein-based circadian oscillator. *FEBS Lett.* 583, 3938–3947.
- Mohawk JA, Green CB & Takahashi JS (2012) Central and peripheral circadian clocks in mammals. *Annu. Rev. Neurosci.* 35, 445–462.
- Niehrs C (2012) The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.* 13, 767–779.
- ### Signaling in Plants
- Chen M, Chory J & Fankhauser C (2004) Light signal transduction in higher plants. *Annu. Rev. Genet.* 38, 87–117.
- Dievart A & Clark SE (2004) LRR-containing receptors regulating plant development and defense. *Development* 131, 251–261.
- Kim TW & Wang ZY (2010) Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* 61, 681–704.
- Mockaitis K & Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. *Annu. Rev. Cell Dev. Biol.* 24, 55–80.
- Stepanova AN & Alonso JM (2009) Ethylene signaling and response: where different regulatory modules meet. *Curr. Opin. Plant Biol.* 12, 548–555.

# The Cytoskeleton

# CHAPTER 16

For cells to function properly, they must organize themselves in space and interact mechanically with each other and with their environment. They have to be correctly shaped, physically robust, and properly structured internally. Many have to change their shape and move from place to place. All cells have to be able to rearrange their internal components as they grow, divide, and adapt to changing circumstances. These spatial and mechanical functions depend on a remarkable system of filaments called the **cytoskeleton** (Figure 16–1).

The cytoskeleton's varied functions depend on the behavior of three families of protein filaments—*actin filaments*, *microtubules*, and *intermediate filaments*. Each type of filament has distinct mechanical properties, dynamics, and biological roles, but all share certain fundamental features. Just as we require our ligaments, bones, and muscles to work together, so all three cytoskeletal filament systems must normally function collectively to give a cell its strength, its shape, and its ability to move.

In this chapter, we describe the function and conservation of the three main filament systems. We explain the basic principles underlying filament assembly and disassembly, and how other proteins interact with the filaments to alter their dynamics, enabling the cell to establish and maintain internal order, to shape and remodel its surface, and to move organelles in a directed manner from one place to another. Finally, we discuss how the integration and regulation of the cytoskeleton allows a cell to move to new locations.

## FUNCTION AND ORIGIN OF THE CYTOSKELETON

The three major cytoskeletal filaments are responsible for different aspects of the cell's spatial organization and mechanical properties. Actin filaments determine the shape of the cell's surface and are necessary for whole-cell locomotion; they also drive the pinching of one cell into two. Microtubules determine the positions of membrane-enclosed organelles, direct intracellular transport, and form the mitotic spindle that segregates chromosomes during cell division. Intermediate filaments provide mechanical strength. All of these cytoskeletal filaments interact with hundreds of accessory proteins that regulate and link the filaments to other cell components, as well as to each other. The accessory proteins are essential for the controlled assembly of the cytoskeletal filaments in particular locations, and they include the *motor proteins*, remarkable molecular machines that convert the energy of ATP hydrolysis into mechanical force that can either move organelles along the filaments or move the filaments themselves.

In this section, we discuss the general features of the proteins that make up the filaments of the cytoskeleton. We focus on their ability to form intrinsically

### IN THIS CHAPTER

#### FUNCTION AND ORIGIN OF THE CYTOSKELETON

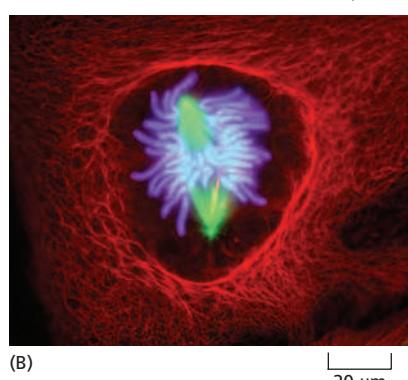
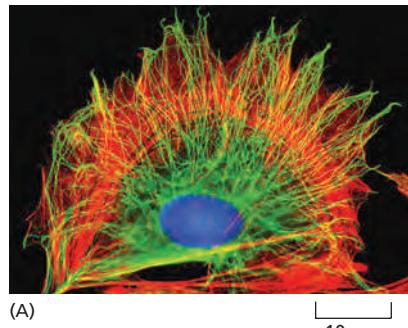
#### ACTIN AND ACTIN-BINDING PROTEINS

#### MYOSIN AND ACTIN

#### MICROTUBULES

#### INTERMEDIATE FILAMENTS AND SEPTINS

#### CELL POLARIZATION AND MIGRATION



**Figure 16–1 The cytoskeleton.** (A) A cell in culture has been fixed and labeled to show its cytoplasmic arrays of microtubules (green) and actin filaments (red). (B) This dividing cell has been labeled to show its spindle microtubules (green) and surrounding cage of intermediate filaments (red). The DNA in both cells is labeled in blue. (A, courtesy of Albert Tousson; B, courtesy of Conly Rieder.)

polarized and self-organized structures that are highly dynamic, allowing the cell to rapidly modify cytoskeletal structure and function under different conditions.

### Cytoskeletal Filaments Adapt to Form Dynamic or Stable Structures

Cytoskeletal systems are dynamic and adaptable, organized more like ant trails than interstate highways. A single trail of ants may persist for many hours, extending from the ant nest to a delectable picnic site, but the individual ants within the trail are anything but static. If the ant scouts find a new and better source of food, or if the picnickers clean up and leave, the dynamic structure adapts with astonishing rapidity. In a similar way, large-scale cytoskeletal structures can change or persist, according to need, lasting for lengths of time ranging from less than a minute up to the cell's lifetime. But the individual macromolecular components that make up these structures are in a constant state of flux. Thus, like the alteration of an ant trail, a structural rearrangement in a cell requires little extra energy when conditions change.

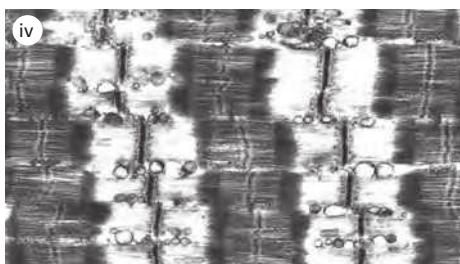
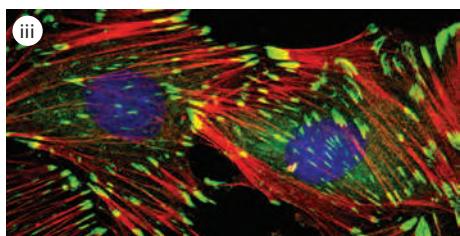
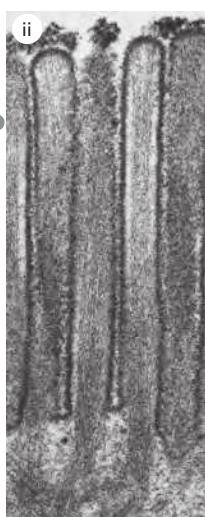
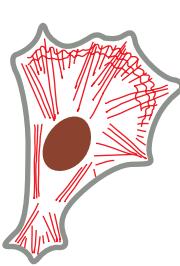
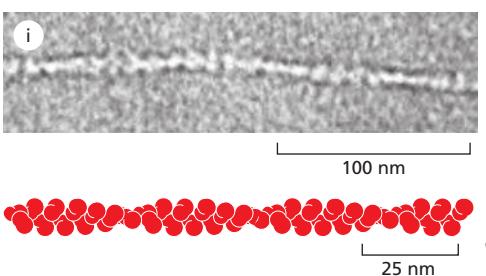
Regulation of the dynamic behavior and assembly of cytoskeletal filaments allows eukaryotic cells to build an enormous range of structures from the three basic filament systems. The micrographs in **Panel 16-1** illustrate some of these structures. Actin filaments underlie the plasma membrane of animal cells, providing strength and shape to its thin lipid bilayer. They also form many types of cell-surface projections. Some of these are dynamic structures, such as the *lamellipodia* and *filopodia* that cells use to explore territory and move around. More stable arrays allow cells to brace themselves against an underlying substratum and enable muscle to contract. The regular bundles of *stereocilia* on the surface of hair cells in the inner ear contain stable bundles of actin filaments that tilt as rigid rods in response to sound, and similarly organized *microvilli* on the surface of intestinal epithelial cells vastly increase the apical cell-surface area to enhance nutrient absorption. In plants, actin filaments drive the rapid streaming of cytoplasm inside cells.

Microtubules, which are frequently found in a cytoplasmic array that extends to the cell periphery, can quickly rearrange themselves to form a bipolar *mitotic spindle* during cell division. They can also form *cilia*, which function as motile whips or sensory devices on the surface of the cell, or tightly aligned bundles that serve as tracks for the transport of materials down long neuronal axons. In plant cells, organized arrays of microtubules help to direct the pattern of cell wall synthesis, and in many protozoans they form the framework upon which the entire cell is built.

Intermediate filaments line the inner face of the nuclear envelope, forming a protective cage for the cell's DNA; in the cytosol, they are twisted into strong cables that can hold epithelial cell sheets together or help nerve cells to extend long and robust axons, and they allow us to form tough appendages such as hair and fingernails.

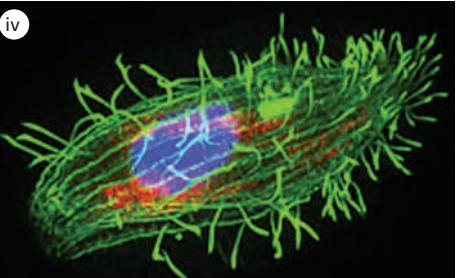
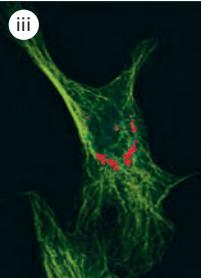
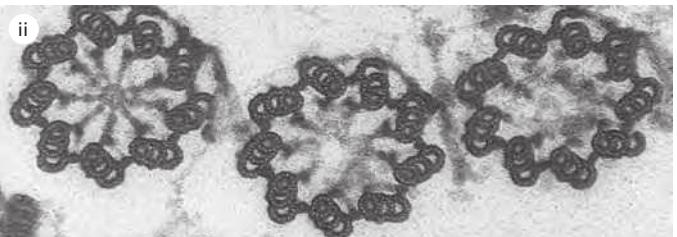
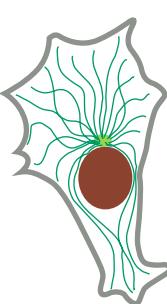
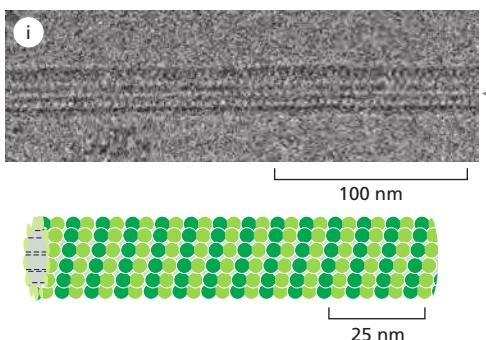
An important and dramatic example of rapid reorganization of the cytoskeleton occurs during cell division, as shown in **Figure 16-2** for a fibroblast growing in a tissue-culture dish. After the chromosomes have replicated, the interphase microtubule array that spreads throughout the cytoplasm is reconfigured into the bipolar *mitotic spindle*, which segregates the two copies of each chromosome into daughter nuclei. At the same time, the specialized actin structures that enable the fibroblast to crawl across the surface of the dish rearrange so that the cell stops moving and assumes a more spherical shape. Actin and its associated motor protein myosin then form a belt around the middle of the cell, the *contractile ring*, which constricts like a tiny muscle to pinch the cell in two. When division is complete, the cytoskeletons of the two daughter fibroblasts reassemble into their interphase structures to convert the two rounded-up daughter cells into smaller versions of the flattened, crawling mother cell.

Many cells require rapid cytoskeletal rearrangements for their normal functioning during interphase as well. For example, the *neutrophil*, a type of white

**ACTIN FILAMENTS**

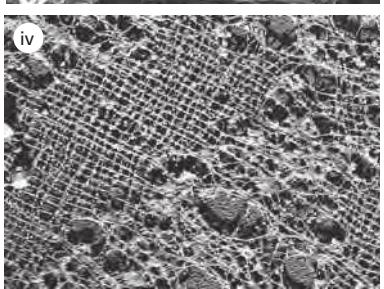
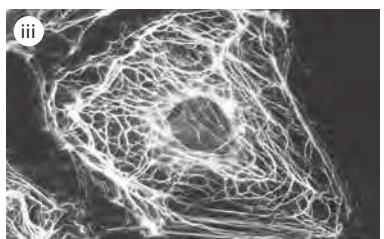
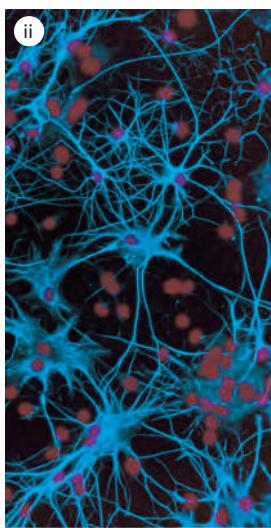
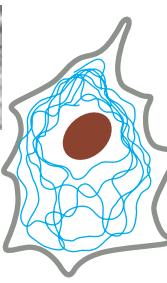
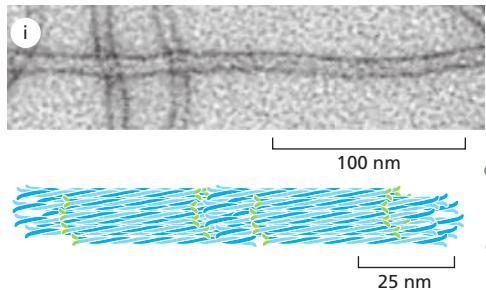
**Actin filaments** (also known as *microfilaments*) are helical polymers of the protein actin. They are flexible structures with a diameter of 8 nm that organize into a variety of linear bundles, two-dimensional networks, and three-dimensional gels. Although actin filaments are dispersed throughout the cell, they are most highly concentrated in the cortex, just beneath the plasma membrane. (i) Single actin filament; (ii) microvilli; (iii) stress fibers (red) terminating in focal adhesions (green); (iv) striated muscle.

Micrographs courtesy of R. Craig (i and iv); P.T. Matsudaira and D.R. Burgess (ii); K. Burridge (iii).

**MICROTUBULES**

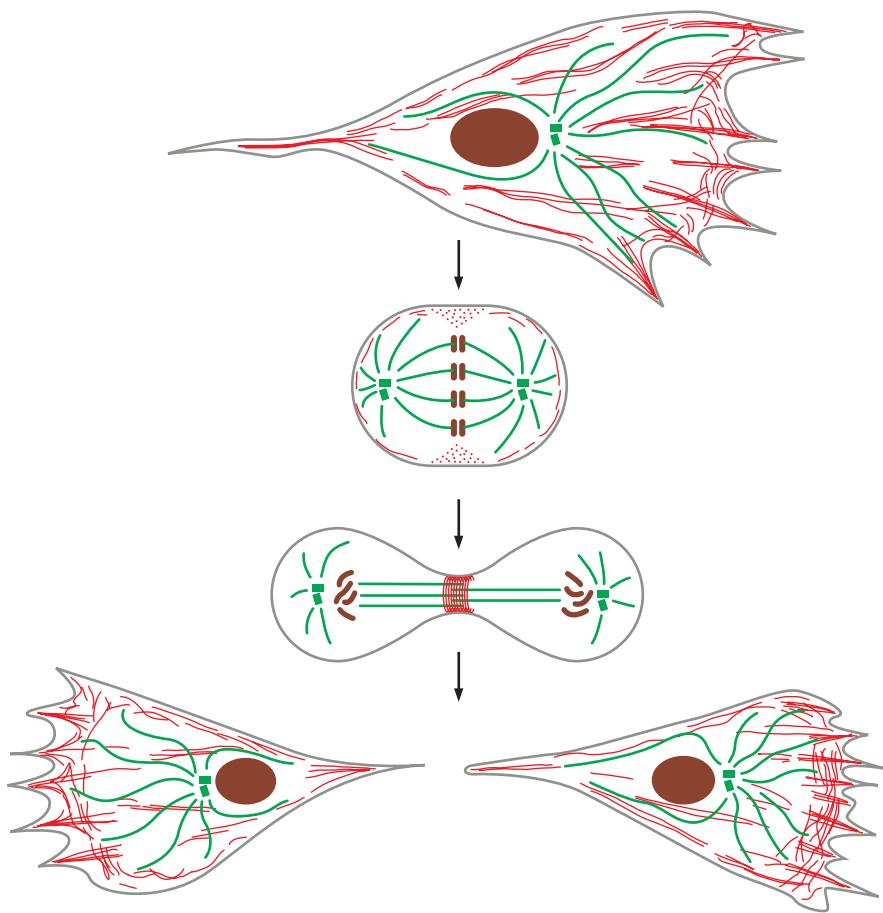
**Microtubules** are long, hollow cylinders made of the protein tubulin. With an outer diameter of 25 nm, they are much more rigid than actin filaments. Microtubules are long and straight and frequently have one end attached to a microtubule-organizing center (MTOC) called a *centrosome*. (i) Single microtubule; (ii) cross section at the base of three cilia showing triplet microtubules; (iii) interphase microtubule array (green) and organelles (red); (iv) ciliated protozoan.

Micrographs courtesy of R. Wade (i); D.T. Woodrow and R.W. Linck (ii); D. Shima (iii); D. Burnette (iv).

**INTERMEDIATE FILAMENTS**

**Intermediate filaments** are ropelike fibers with a diameter of about 10 nm; they are made of intermediate filament proteins, which constitute a large and heterogeneous family. One type of intermediate filament forms a meshwork called the nuclear lamina just beneath the inner nuclear membrane. Other types extend across the cytoplasm, giving cells mechanical strength. In an epithelial tissue, they span the cytoplasm from one cell-cell junction to another, thereby strengthening the entire epithelium. (i) Individual intermediate filaments; (ii) Intermediate filaments (blue) in neurons and (iii) epithelial cell; (iv) nuclear lamina.

Micrographs courtesy of R. Quinlan (i); N. L. Kedersha (ii); M. Osborn (iii); U. Aebi (iv).

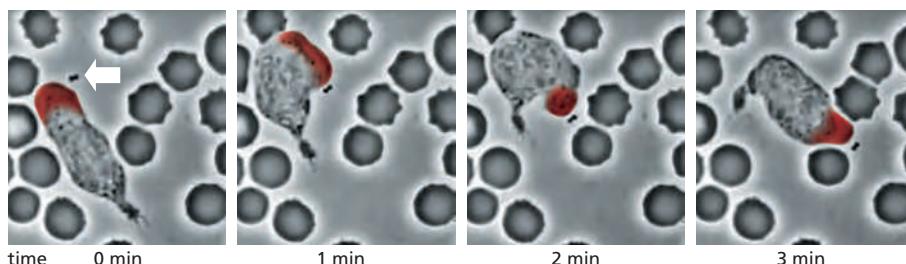


**Figure 16–2** Diagram of changes in cytoskeletal organization associated with cell division. The crawling fibroblast drawn here has a polarized, dynamic actin cytoskeleton (shown in red) that assembles lamellipodia and filopodia to push its leading edge toward the right. The polarization of the actin cytoskeleton is assisted by the microtubule cytoskeleton (green), consisting of long microtubules that emanate from a single microtubule-organizing center located in front of the nucleus. When the cell divides, the polarized microtubule array rearranges to form a bipolar mitotic spindle, which is responsible for aligning and then segregating the duplicated chromosomes (brown). The actin filaments form a contractile ring at the center of the cell that pinches the cell in two after the chromosome segregation. After cell division is complete, the two daughter cells reorganize both the microtubule and actin cytoskeletons into smaller versions of those that were present in the mother cell, enabling them to crawl their separate ways.

blood cell, chases and engulfs bacterial and fungal cells that accidentally gain access to the normally sterile parts of the body, as through a cut in the skin. Like most crawling cells, neutrophils advance by extending a protrusive structure filled with newly polymerized actin filaments. When the elusive bacterial prey moves in a different direction, the neutrophil is poised to reorganize its polarized protrusive structures within seconds ([Figure 16–3](#)).

### The Cytoskeleton Determines Cellular Organization and Polarity

In cells that have achieved a stable, differentiated morphology—such as mature neurons or epithelial cells—the dynamic elements of the cytoskeleton must also provide stable, large-scale structures for cellular organization. On specialized epithelial cells that line organs such as the intestine and the lung, cytoskeletal-based cell-surface protrusions including microvilli and cilia are able to maintain a constant location, length, and diameter over the entire lifetime of the cell. For the actin bundles at the cores of microvilli on intestinal epithelial cells, this is only a few days. But the actin bundles at the cores of stereocilia on the hair cells of the inner ear must maintain their stable organization for the entire lifetime of the animal, since these cells do not turn over. Nonetheless, the individual actin filaments



**Figure 16–3** A neutrophil in pursuit of bacteria. In this preparation of human blood, a clump of bacteria (white arrow) is about to be captured by a neutrophil. As the bacteria move, the neutrophil quickly reassembles the dense actin network at its leading edge (highlighted in red) to push toward the location of the bacteria ([Movie 16.1](#)). Rapid disassembly and reassembly of the actin cytoskeleton in this cell enables it to change its orientation and direction of movement within a few minutes. (From a video recorded by David Rogers.)

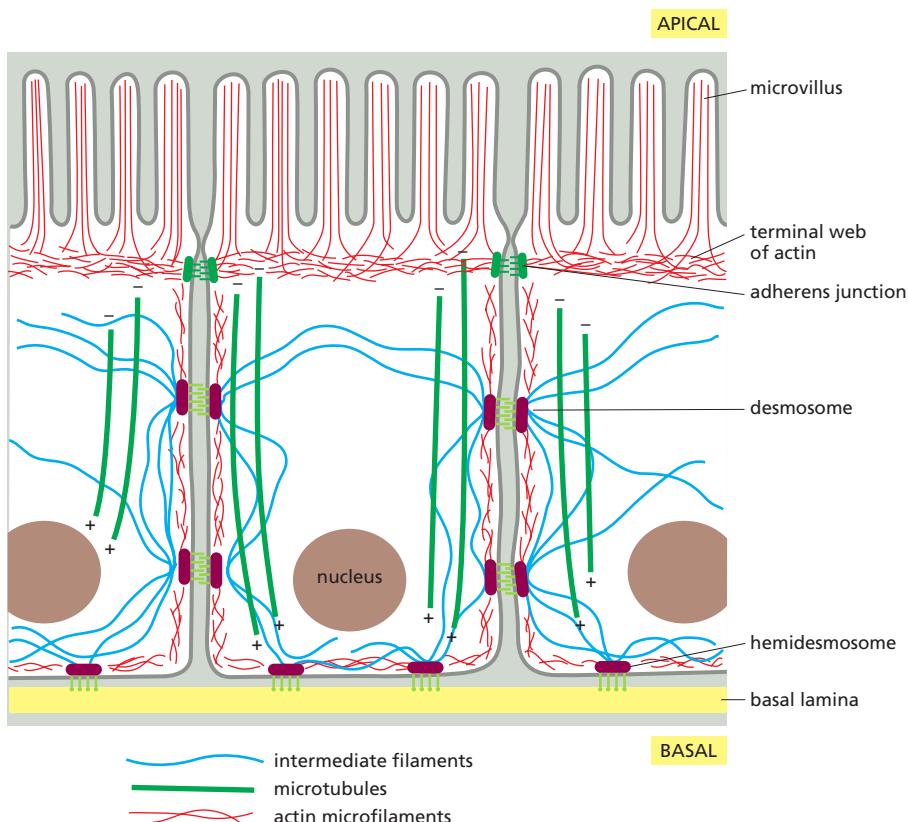
remain strikingly dynamic and are continuously remodeled and replaced every 48 hours, even within stable cell-surface structures that persist for decades.

Besides forming stable, specialized cell-surface protrusions, the cytoskeleton is also responsible for large-scale cellular polarity, enabling cells to tell the difference between top and bottom, or front and back. The large-scale polarity information conveyed by cytoskeletal organization is often maintained over the lifetime of the cell. Polarized epithelial cells use organized arrays of microtubules, actin filaments, and intermediate filaments to maintain the critical differences between the *apical surface* and the *basolateral surface*. They also must maintain strong adhesive contacts with one another to enable this single layer of cells to serve as an effective physical barrier (Figure 16–4).

### Filaments Assemble from Protein Subunits That Impart Specific Physical and Dynamic Properties

Cytoskeletal filaments can reach from one end of the cell to the other, spanning tens or even hundreds of micrometers. Yet the individual protein molecules that form the filaments are only a few nanometers in size. The cell builds the filaments by assembling large numbers of the small subunits, like building a skyscraper out of bricks. Because these subunits are small, they can diffuse rapidly in the cytosol, whereas the assembled filaments cannot. In this way, cells can undergo rapid structural reorganizations, disassembling filaments at one site and reassembling them at another site far away.

Actin filaments and microtubules are built from subunits that are compact and globular—*actin subunits* for actin filaments and *tubulin subunits* for microtubules—whereas intermediate filaments are made up of smaller subunits that are themselves elongated and fibrous. All three major types of cytoskeletal filaments form as helical assemblies of subunits (see Figure 3–22) that self-associate, using a combination of end-to-end and side-to-side protein contacts. Differences in the structures of the subunits and the strengths of the attractive forces between



**Figure 16–4 Organization of the cytoskeleton in polarized epithelial cells.**

All the components of the cytoskeleton cooperate to produce the characteristic shapes of specialized cells, including the epithelial cells that line the small intestine, diagrammed here. At the apical (upper) surface, facing the intestinal lumen, bundled actin filaments (red) form microvilli that increase the cell surface area available for absorbing nutrients from food. Below the microvilli, a circumferential band of actin filaments is connected to cell–cell adherens junctions that anchor the cells to each other. Intermediate filaments (blue) are anchored to other kinds of adhesive structures, including desmosomes and hemidesmosomes, that connect the epithelial cells into a sturdy sheet and attach them to the underlying extracellular matrix; these structures are discussed in Chapter 19. Microtubules (green) run vertically from the top of the cell to the bottom and provide a global coordinate system that enables the cell to direct newly synthesized components to their proper locations.

them produce important differences in the stability and mechanical properties of each type of filament. Whereas covalent linkages between their subunits hold together the backbones of many biological polymers—including DNA, RNA, and proteins—it is weak noncovalent interactions that hold together the three types of cytoskeletal polymers. Consequently, their assembly and disassembly can occur rapidly, without covalent bonds being formed or broken.

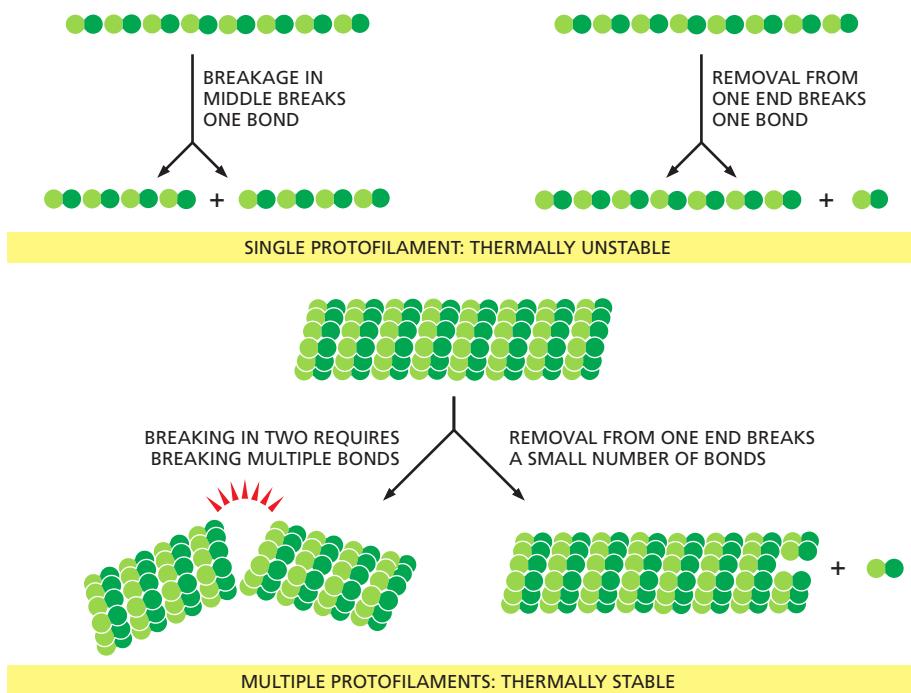
The subunits of actin filaments and microtubules are asymmetrical and bind to one another head-to-tail so that they all point in one direction. This subunit polarity gives the filaments structural polarity along their length, and makes the two ends of each polymer behave differently. In addition, actin and tubulin subunits are both enzymes that catalyze the hydrolysis of a nucleoside triphosphate—ATP and GTP, respectively. As we discuss later, the energy derived from nucleotide hydrolysis enables the filaments to remodel rapidly. By controlling when and where actin and microtubules assemble, the cell harnesses the polar and dynamic properties of these filaments to generate force in a specific direction, to move the leading edge of a migrating cell forward, for example, or to pull chromosomes apart during cell division. In contrast, the subunits of intermediate filaments are symmetrical, and thus do not form polarized filaments with two different ends. Intermediate filament subunits also do not catalyze the hydrolysis of nucleotides. Nevertheless, intermediate filaments can be disassembled rapidly when required. In mitosis, for example, kinases phosphorylate the subunits, leading to their dissociation.

Cytoskeletal filaments in living cells are not built by simply stringing subunits together in single file. A thousand tubulin subunits lined up end-to-end, for example, would span the diameter of a small eukaryotic cell, but a filament formed in this way would lack the strength to avoid breakage by ambient thermal energy, unless each subunit in the filament was bound extremely tightly to its two neighbors. Such tight binding would limit the rate at which the filaments could disassemble, making the cytoskeleton a static and less useful structure. To provide both strength and adaptability, microtubules are built of 13 **protofilaments**—linear strings of subunits joined end-to-end—that associate with one another laterally to form a hollow cylinder. The addition or loss of a subunit at the end of one protofilament makes or breaks a small number of bonds. In contrast, loss of a subunit from the middle of the filament requires breaking many more bonds, while breaking it in two requires breaking bonds in multiple protofilaments all at the same time (**Figure 16–5**). The greater energy required to break multiple noncovalent bonds simultaneously allows microtubules to resist thermal breakage, while allowing rapid subunit addition and loss at the filament ends. Helical actin filaments are much thinner and therefore require much less energy to break. However, multiple actin filaments are often bundled together inside cells, providing mechanical strength, while allowing dynamic behavior of filament ends.

As with other specific protein–protein interactions, many hydrophobic interactions and noncovalent bonds hold the subunits in a cytoskeletal filament together (see Figure 3–4). The locations and types of subunit–subunit contacts differ for the different filaments. Intermediate filaments, for example, assemble by forming strong lateral contacts between  $\alpha$ -helical coiled-coils, which extend over most of the length of each elongated fibrous subunit. Because the individual subunits are staggered in the filament, intermediate filaments form strong, ropelike structures that tolerate stretching and bending to a greater extent than do either actin filaments or microtubules (**Figure 16–6**).

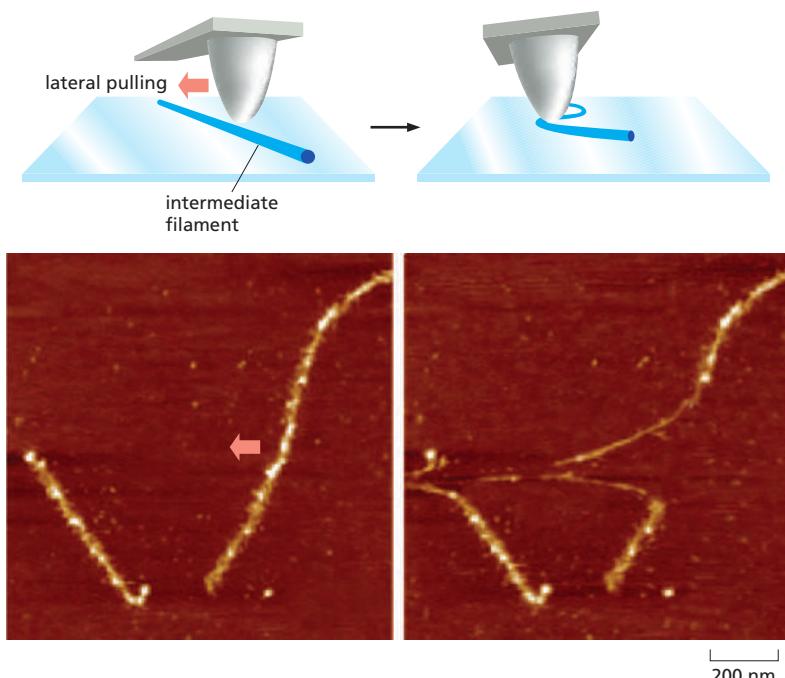
### Accessory Proteins and Motors Regulate Cytoskeletal Filaments

The cell regulates the length and stability of its cytoskeletal filaments, as well as their number and geometry. It does so largely by regulating their attachments to one another and to other components of the cell, so that the filaments can form a wide variety of higher-order structures. Direct covalent modification of the filament subunits regulates some filament properties, but most of the regulation is performed by hundreds of accessory proteins that determine the spatial



**Figure 16–5 The thermal stability of cytoskeletal filaments with dynamic ends.** A protofilament consisting of a single strand of subunits is thermally unstable, since breakage of a single bond between subunits is sufficient to break the filament. In contrast, formation of a cytoskeletal filament from more than one protofilament allows the ends to be dynamic, while enabling the filaments themselves to be resistant to thermal breakage. In a microtubule, for example, removing a single subunit dimer from the end of the filament requires breaking noncovalent bonds with a maximum of three other subunits, whereas fracturing the filament in the middle requires breaking noncovalent bonds in all thirteen protofilaments.

distribution and the dynamic behavior of the filaments, converting information received through signaling pathways into cytoskeletal action. These accessory proteins bind to the filaments or their subunits to determine the sites of assembly of new filaments, to regulate the partitioning of polymer proteins between filament and subunit forms, to change the kinetics of filament assembly and disassembly, to harness energy to generate force, and to link filaments to one another or to other cell structures such as organelles and the plasma membrane. In these processes, the accessory proteins bring cytoskeletal structure under the control of extracellular and intracellular signals, including those that trigger the dramatic transformations of the cytoskeleton that occur during each cell cycle. Acting together, the accessory proteins enable a eukaryotic cell to maintain a highly organized but flexible internal structure and, in many cases, to move.



**Figure 16–6 Flexibility and stretch in an intermediate filament.** Intermediate filaments are formed from elongated fibrous subunits with strong lateral contacts, resulting in resistance to stretching forces. When a tiny mechanical probe is dragged across an intermediate filament, the filament is stretched over three times its length before it breaks, as illustrated by the fluorescently labeled filaments in the photomicrographs. This technique is termed atomic force microscopy (see Figure 9–33). (Adapted from L. Kreplak et al., *J. Mol. Biol.* 354:569–577, 2005. With permission from Elsevier.)

Among the most fascinating proteins that associate with the cytoskeleton are the **motor proteins**. These proteins bind to a polarized cytoskeletal filament and use the energy derived from repeated cycles of ATP hydrolysis to move along it. Dozens of different motor proteins coexist in every eukaryotic cell. They differ in the type of filament they bind to (either actin or microtubules), the direction in which they move along the filament, and the “cargo” they carry. Many motor proteins carry membrane-enclosed organelles—such as mitochondria, Golgi stacks, or secretory vesicles—to their appropriate locations in the cell. Other motor proteins cause cytoskeletal filaments to exert tension or to slide against each other, generating the force that drives such phenomena as muscle contraction, ciliary beating, and cell division.

Cytoskeletal motor proteins that move unidirectionally along an oriented polymer track are reminiscent of some other proteins and protein complexes discussed elsewhere in this book, such as DNA and RNA polymerases, helicases, and ribosomes. All of these proteins have the ability to use chemical energy to propel themselves along a linear track, with the direction of sliding dependent on the structural polarity of the track. All of them generate motion by coupling nucleoside triphosphate hydrolysis to a large-scale conformational change (see Figure 3–75).

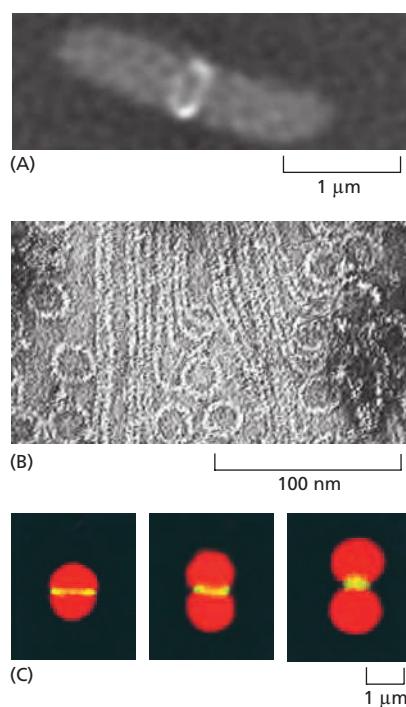
### Bacterial Cell Organization and Division Depend on Homologs of Eukaryotic Cytoskeletal Proteins

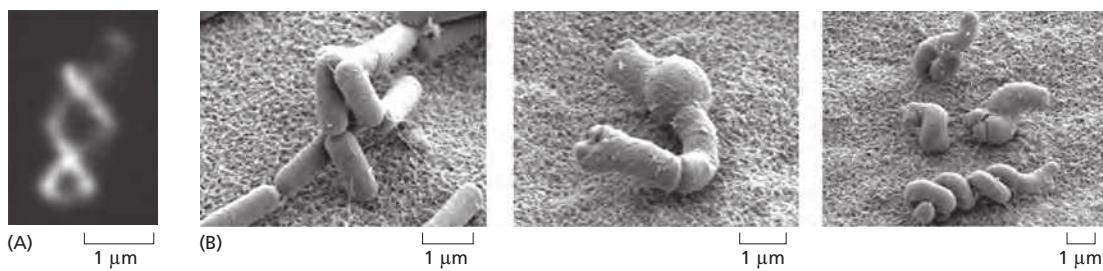
While eukaryotic cells are typically large and morphologically complex, bacterial cells are usually only a few micrometers long and assume simple shapes such as spheres or rods. Bacteria also lack elaborate networks of intracellular membrane-enclosed organelles. Historically, biologists assumed that a cytoskeleton was not necessary in such simple cells. We now know, however, that bacteria contain homologs of all three of the eukaryotic cytoskeletal filaments. Furthermore, bacterial actins and tubulins are more diverse than their eukaryotic versions, both in the types of assemblies they form and in the functions they carry out.

Nearly all bacteria and many archaea contain a homolog of tubulin called FtsZ, which can polymerize into filaments and assemble into a ring (called the Z-ring) at the site where the **septum** forms during cell division (Figure 16–7). Although the Z-ring persists for many minutes, the individual filaments within it are highly dynamic, with an average filament half-life of about thirty seconds. As the bacterium divides, the Z-ring becomes smaller until it has completely disassembled. FtsZ filaments in the Z-ring are thought to generate a bending force that drives the membrane invagination necessary to complete cell division. The Z-ring may also serve as a site for localization of enzymes required for building the septum between the two daughter cells.

Many bacteria also contain homologs of actin. Two of these, MreB and Mbl, are found primarily in rod-shaped or spiral-shaped cells where they assemble to form dynamic patches that move circumferentially along the length of the cell (Figure 16–8A). These proteins contribute to cell shape by serving as a scaffold to direct the synthesis of the peptidoglycan cell wall, in much the same way that microtubules help organize the synthesis of the cellulose cell wall in higher

**Figure 16–7 The bacterial FtsZ protein, a tubulin homolog in prokaryotes.** (A) A band of FtsZ protein forms a ring in a dividing bacterial cell. This ring has been labeled by fusing the FtsZ protein to green fluorescent protein (GFP), which allows it to be observed in living *E. coli* cells with a fluorescence microscope. (B) FtsZ filaments and circles, formed *in vitro*, as visualized using electron microscopy. (C) Dividing chloroplasts (red) from a red alga also cleave using a protein ring made from FtsZ (yellow). (A, from X. Ma, D.W. Ehrhardt and W. Margolin, *Proc. Natl Acad. Sci. USA* 93:12998–13003, 1996; B, from H.P. Erickson et al., *Proc. Natl Acad. Sci. USA* 93:519–523, 1996. Both with permission from National Academy of Sciences; C, from S. Miyagishima et al., *Plant Cell* 13:2257–2268, 2001, with permission from American Society of Plant Biologists.)





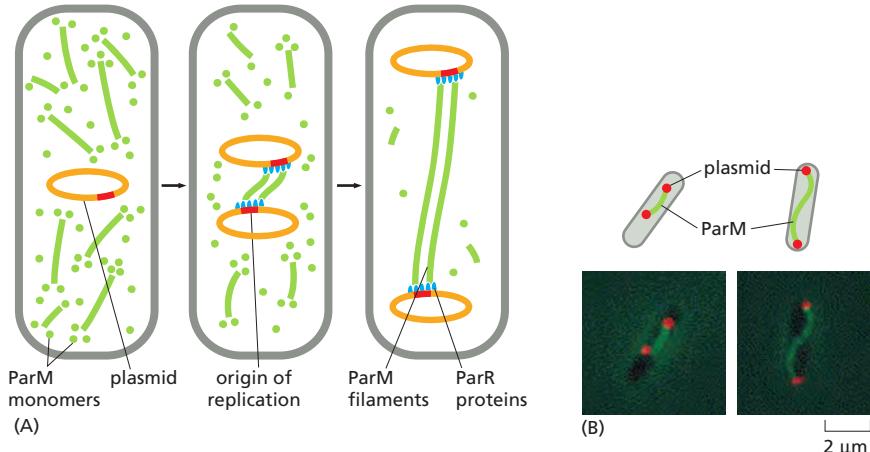
**Figure 16–8** Actin homologs in bacteria determine cell shape. (A) The MreB protein forms abundant patches made up of many short, interwoven linear or helical filaments that are seen to move circumferentially along the length of the bacterium and are associated with sites of cell wall synthesis. (B) The common soil bacterium *Bacillus subtilis* normally forms cells with a regular rodlike shape when viewed by scanning electron microscopy (*left*). In contrast, *B. subtilis* cells lacking the actin homolog MreB or Mbl grow in distorted or twisted shapes and eventually die (*center* and *right*). (A, from P. Vats and L. Rothfield, *Proc. Natl. Acad. Sci. USA* 104:17795–17800, 2007. With permission from National Academy of Sciences; B, from A. Chastanet and R. Carballido-Lopez, *Front. Biosci.* 4S:1582–1606, 2012. With permission Frontiers in Bioscience.)

plant cells (see Figure 19–65). As with FtsZ, MreB and Mbl filaments are highly dynamic, with half-lives of a few minutes, and nucleotide hydrolysis accompanies the polymerization process. Mutations disrupting MreB or Mbl expression cause extreme abnormalities in cell shape and defects in chromosome segregation (Figure 16–8B).

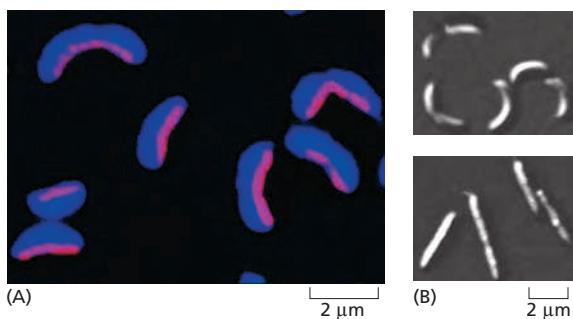
Relatives of MreB and Mbl have more specialized roles. A particularly intriguing bacterial actin homolog is ParM, which is encoded by a gene on certain bacterial plasmids that also carry genes responsible for antibiotic resistance and cause the spread of multidrug resistance in epidemics. Bacterial plasmids typically encode all the gene products that are necessary for their own segregation, presumably as a strategy to ensure their inheritance and propagation in bacterial hosts following plasmid replication. ParM assembles into filaments that associate at each end with a copy of the plasmid, and growth of the ParM filament pushes the replicated plasmid copies apart (Figure 16–9). This spindle-like structure apparently arises from the selective stabilization of filaments that bind to specialized proteins recruited to the origins of replication on the plasmids. A distant relative of both tubulin and FtsZ, called TubZ, has a similar function in other bacterial species.

Thus, self-association of nucleotide-binding proteins into dynamic filaments is used in all cells, and the actin and tubulin families are very ancient, predating the split between the eukaryotic and bacterial kingdoms.

At least one bacterial species, *Caulobacter crescentus*, appears to harbor a protein with significant structural similarity to the third major class of cytoskeletal filaments found in animal cells, the intermediate filaments. A protein called crescentin forms a filamentous structure that influences the unusual crescent shape of this species; when the gene encoding crescentin is deleted, the *Caulobacter* cells grow as straight rods (Figure 16–10).



**Figure 16–9** Role of the actin homolog ParM in plasmid segregation in bacteria. (A) Some bacterial drug-resistance plasmids (orange) encode an actin homolog, ParM, that will spontaneously nucleate to form small, dynamic filaments (green) throughout the bacterial cytoplasm. A second plasmid-encoded protein called ParR (blue) binds to specific DNA sequences in the plasmid and also stabilizes the dynamic ends of the ParM filaments. When the plasmid duplicates, both ends of the ParM filaments become stabilized, and the growing ParM filaments push the duplicated plasmids to opposite ends of the cell. (B) In these bacterial cells harboring a drug-resistance plasmid, the plasmids are labeled in red and the ParM protein in green. Left, a short ParM filament bundle connects the two daughter plasmids shortly after their duplication. Right, the fully assembled ParM filament has pushed the duplicated plasmids to the cell poles. (A, adapted from E.C. Garner, C.S. Campbell and R.D. Mullins, *Science* 306:1021–1025, 2004; B, from J. Møller-Jensen et al., *Mol. Cell* 12:1477–1487, 2003. With permission from Elsevier.)



**Figure 16-10 Caulobacter and crescentin.** The sickle-shaped bacterium *Caulobacter crescentus* expresses a protein, crescentin, with a series of coiled-coil domains similar in size and organization to the domains of eukaryotic intermediate filaments. (A) The crescentin protein forms a fiber (labeled in red) that runs down the inner side of the curving bacterial cell wall. (B) When the gene is disrupted, the bacteria grow as straight rods (bottom). (From N. Ausmees, J.R. Kuhn and C. Jacobs-Wagner, *Cell* 115:705–713, 2003. With permission from Elsevier.)

## Summary

The cytoplasm of eukaryotic cells is spatially organized by a network of protein filaments known as the cytoskeleton. This network contains three principal types of filaments: actin filaments, microtubules, and intermediate filaments. All three types of filaments form as helical assemblies of subunits that self-associate using a combination of end-to-end and side-to-side protein contacts. Differences in the structure of the subunits and the manner of their self-assembly give the filaments different mechanical properties. Subunit assembly and disassembly constantly remodel all three types of cytoskeletal filaments. Actin and tubulin (the subunits of actin filaments and microtubules, respectively) bind and hydrolyze nucleoside triphosphates (ATP and GTP, respectively), and assemble head-to-tail to generate polarized filaments capable of generating force. In living cells, accessory proteins modulate the dynamics and organization of cytoskeletal filaments, resulting in complex events such as cell division or migration, and generating elaborate cellular architecture to form polarized tissues such as epithelia. Bacterial cells also contain homologs of actin, tubulin, and intermediate filaments that form dynamic structures that help control cell shape and division.

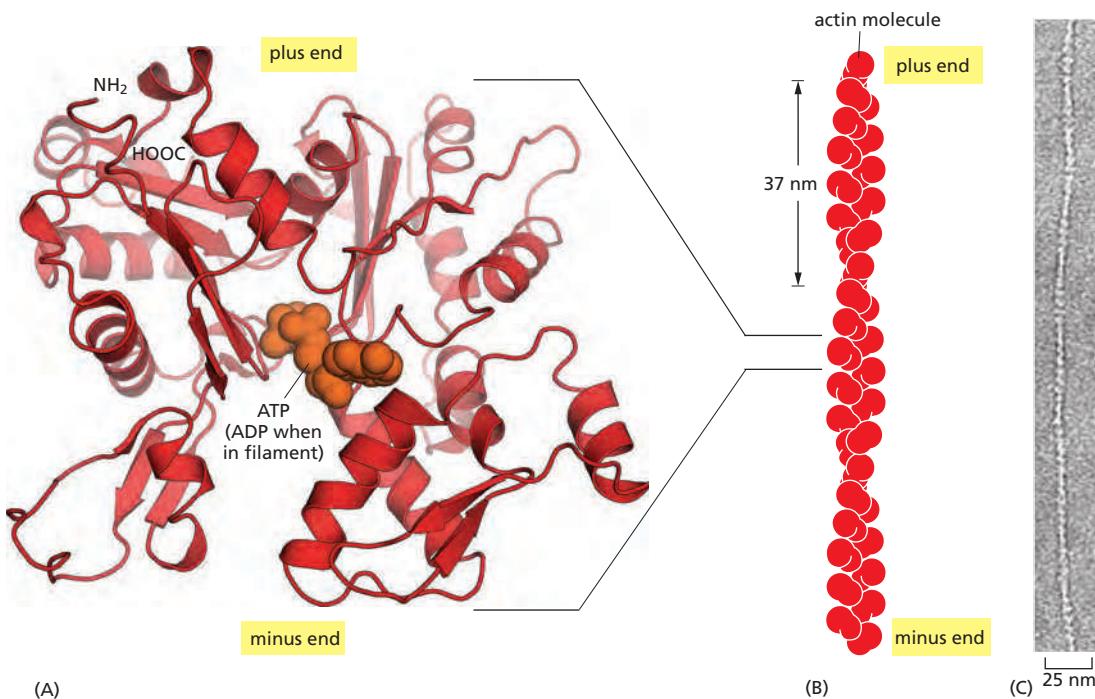
## ACTIN AND ACTIN-BINDING PROTEINS

The actin cytoskeleton performs a wide range of functions in diverse cell types. Each actin subunit, sometimes called globular or G-actin, is a 375-amino-acid polypeptide carrying a tightly associated molecule of ATP or ADP (Figure 16-11A). Actin is extraordinarily well conserved among eukaryotes. The amino acid sequences of actins from different eukaryotic species are usually about 90% identical. Small variations in actin amino acid sequence can cause significant functional differences: In vertebrates, for example, there are three isoforms of actin, termed  $\alpha$ ,  $\beta$ , and  $\gamma$ , that differ slightly in their amino acid sequences and have distinct functions.  $\alpha$ -Actin is expressed only in muscle cells, while  $\beta$ - and  $\gamma$ -actins are found together in almost all non-muscle cells.

### Actin Subunits Assemble Head-to-Tail to Create Flexible, Polar Filaments

Actin subunits assemble head-to-tail to form a tight, right-handed helix, forming a structure about 8 nm wide called filamentous or F-actin (Figure 16-11B and C). Because the asymmetrical actin subunits of a filament all point in the same direction, filaments are polar and have structurally different ends: a slower-growing *minus end* and a faster-growing *plus end*. The minus end is also referred to as the “pointed end” and the plus end as the “barbed end,” because of the “arrowhead” appearance of the complex formed between actin filaments and the motor protein myosin (Figure 16-12). Within the filament, the subunits are positioned with their nucleotide-binding cleft directed toward the minus end.

Individual actin filaments are quite flexible. The stiffness of a filament can be characterized by its *persistence length*, the minimum filament length at which random thermal fluctuations are likely to cause it to bend. The persistence length of an actin filament is only a few tens of micrometers. In a living cell, however,



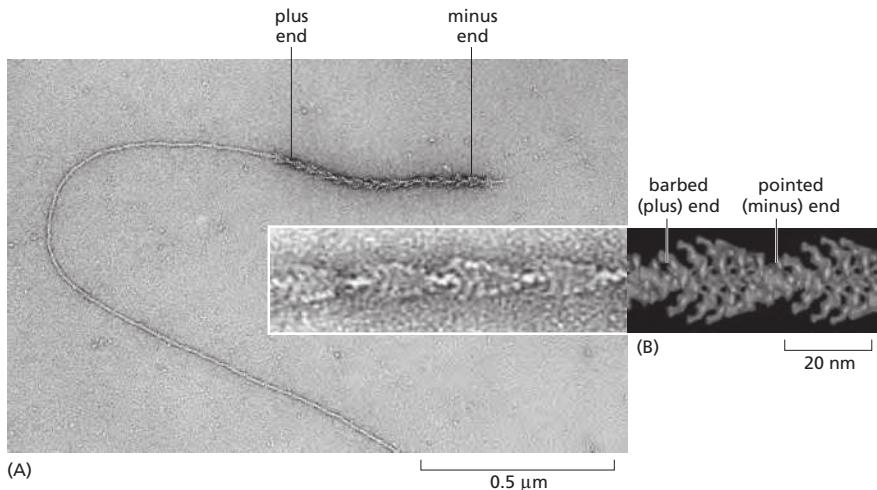
accessory proteins cross-link and bundle the filaments together, making large-scale actin structures that are much more rigid than an individual actin filament.

### Nucleation Is the Rate-Limiting Step in the Formation of Actin Filaments

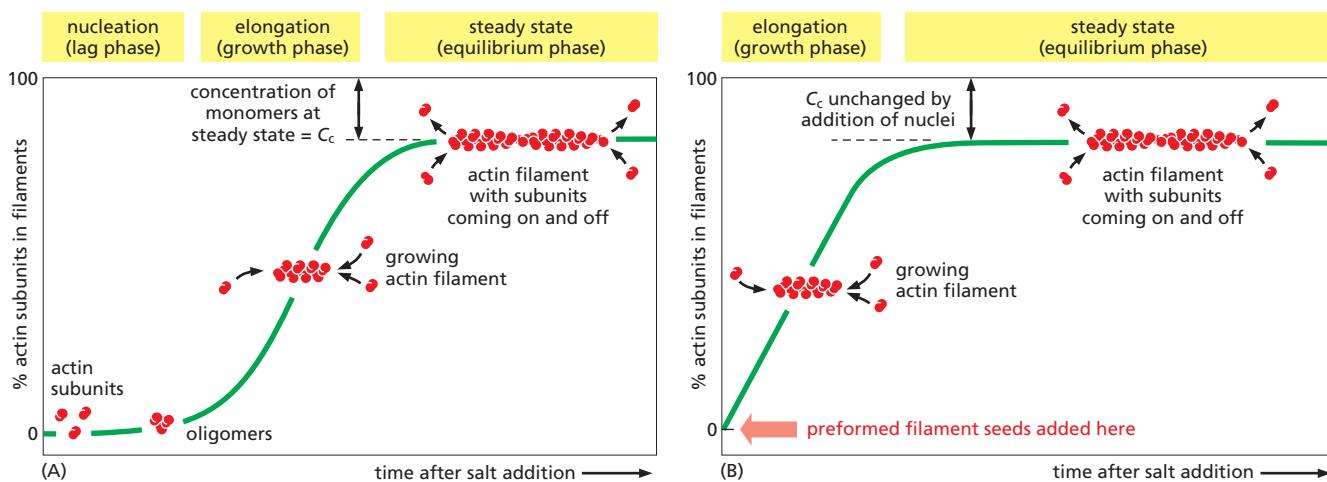
The regulation of actin filament formation is an important mechanism by which cells control their shape and movement. Small oligomers of actin subunits can assemble spontaneously, but they are unstable and disassemble readily because each monomer is bound to only one or two other monomers. For a new actin filament to form, subunits must assemble into an initial aggregate, or nucleus, that is stabilized by multiple subunit–subunit contacts and can then elongate rapidly by addition of more subunits. This process is called *filament nucleation*.

Many features of actin nucleation and polymerization have been studied with purified actin in a test tube (Figure 16-13). The instability of smaller actin aggregates creates a kinetic barrier to nucleation. When polymerization is initiated, this barrier results in a lag phase during which no filaments are observed. During this lag phase, however, a few of the small, unstable aggregates succeed in making the transition to a more stable form that resembles an actin filament. This leads to a

**Figure 16-11** The structures of an actin monomer and actin filament. (A) The actin monomer has a nucleotide (either ATP or ADP) bound in a deep cleft in the center of the molecule. (B) Arrangement of monomers in a filament consisting of two protofilaments, held together by lateral contacts, which wind around each other as two parallel strands of a helix, with a twist repeating every 37 nm. All the subunits within the filament have the same orientation. (C) Electron micrograph of negatively stained actin filament.  
(C, courtesy of Roger Craig.)



**Figure 16-12** Structural polarity of the actin filament. (A) This electron micrograph shows an actin filament polymerized from a short actin filament seed that was decorated with myosin motor domains, resulting in an arrowhead pattern. The filament has grown much faster at the barbed (plus) end than at the pointed (minus) end. (B) Enlarged image and model showing the arrowhead pattern.  
(A, courtesy of Tom Pollard; B, adapted from M. Whittaker, B.O. Carragher and K.A. Milligan, *Ultramicro*. 54:245–260, 1995.)



**Figure 16–13** The time course of actin polymerization in a test tube. (A) Polymerization of pure actin subunits into filaments occurs after a lag phase. (B) Polymerization occurs more rapidly in the presence of preformed fragments of actin filaments, which act as nuclei for filament growth. As indicated, the % free subunits after polymerization reflects the critical concentration ( $C_c$ ), at which there is no net change in polymer. Actin polymerization is often studied by observing the change in the light emission from a fluorescent probe, called pyrene, that has been covalently attached to the actin. Pyrene–actin fluoresces more brightly when it is incorporated into actin filaments.

phase of rapid filament elongation during which subunits are added quickly to the ends of the nucleated filaments (Figure 16–13A). Finally, as the concentration of actin monomers declines, the system approaches a steady state at which the rate of addition of new subunits to the filament ends exactly balances the rate of subunit dissociation. The concentration of free subunits left in solution at this point is called the *critical concentration*,  $C_c$ . As explained in Panel 16–2, the value of the critical concentration is equal to the rate constant for subunit loss divided by the rate constant for subunit addition—that is,  $C_c = k_{\text{off}}/k_{\text{on}}$ , which is equal to the dissociation constant,  $K_d$ , and the inverse of the equilibrium constant,  $K$  (see Figure 3–44). In a test tube, the  $C_c$  for actin polymerization—that is, the free actin monomer concentration at which the fraction of actin in the polymer stops increasing—is about 0.2  $\mu\text{M}$ . Inside the cell, the concentration of unpolymerized actin is much higher than this, and the cell has evolved mechanisms to prevent most of its monomeric actin from assembling into filaments, as we discuss later.

The lag phase in filament growth is eliminated if preexisting seeds (such as fragments of actin filaments that have been chemically cross-linked) are added to the solution at the beginning of the polymerization reaction (Figure 16–13B). The cell takes great advantage of this nucleation requirement: it uses special proteins to catalyze filament nucleation at specific sites, thereby determining the location at which new actin filaments are assembled.

### Actin Filaments Have Two Distinct Ends That Grow at Different Rates

Due to the uniform orientation of asymmetric actin subunits in the filament, the structures at its two ends are different. This orientation makes the two ends of each polymer different in ways that have a profound effect on filament growth rates. The kinetic rate constants for actin subunit association and dissociation— $k_{\text{on}}$  and  $k_{\text{off}}$ , respectively—are much greater at the plus end than the minus end. This can be seen when an excess of purified actin monomers is allowed to assemble onto polarity-marked filaments—the plus end of the filament elongates up to ten times faster (see Figure 16–12). If filaments are rapidly diluted so that the free subunit concentration drops below the critical concentration, the plus end also depolymerizes faster.

It is important to note, however, that the two ends of an actin filament have the same net affinity for actin subunits, if all of the subunits are in the same nucleotide

state. Addition of a subunit to either end of a filament of  $n$  subunits results in a filament of  $n + 1$  subunits. Thus, the free-energy difference, and therefore the equilibrium constant (and the critical concentration), must be the same for addition of subunits at either end of the polymer. In this case, the ratio of the rate constants,  $k_{\text{off}}/k_{\text{on}}$ , must be identical at the two ends, even though the absolute values of these rate constants are very different at each end (see Panel 16–2).

The cell takes advantage of actin filament dynamics and polarity to do mechanical work. Filament elongation proceeds spontaneously when the free-energy change ( $\Delta G$ ) for addition of the soluble subunit is less than zero. This is the case when the concentration of subunits in solution exceeds the critical concentration. A cell can couple an energetically unfavorable process to this spontaneous process; thus, the cell can use free energy released during spontaneous filament polymerization to move an attached load. For example, by orienting the fast-growing plus ends of actin filaments toward its leading edge, a motile cell can push its plasma membrane forward, as we discuss later.

### ATP Hydrolysis Within Actin Filaments Leads to Treadmilling at Steady State

Thus far in our discussion of actin filament dynamics, we have ignored the critical fact that actin can catalyze the hydrolysis of the nucleoside triphosphate ATP. For free actin subunits, this hydrolysis proceeds very slowly; however, it is accelerated when the subunits are incorporated into filaments. Shortly after ATP hydrolysis occurs, the free phosphate group is released from each subunit, but the ADP remains trapped in the filament structure. Thus, two different types of filament structures can exist, one with the “T form” of the nucleotide bound (ATP), and one with the “D form” bound (ADP).

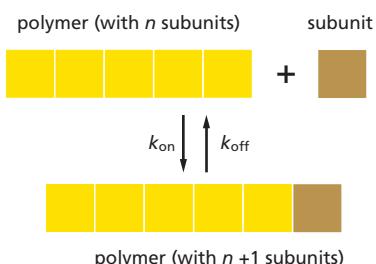
When the nucleotide is hydrolyzed, much of the free energy released by cleavage of the phosphate-phosphate bond is stored in the polymer. This makes the free-energy change for dissociation of a subunit from the D-form polymer more negative than the free-energy change for dissociation of a subunit from the T-form polymer. Consequently, the ratio of  $k_{\text{off}}/k_{\text{on}}$  for the D-form polymer, which is numerically equal to its critical concentration [ $C_c(\text{D})$ ], is larger than the corresponding ratio for the T-form polymer. Thus,  $C_c(\text{D})$  is greater than  $C_c(\text{T})$ . At certain concentrations of free subunits, D-form polymers will therefore shrink while T-form polymers grow.

In living cells, most soluble actin subunits are in the T form, as the free concentration of ATP is about tenfold higher than that of ADP. However, the longer the time that subunits have been in the actin filament, the more likely they are to have hydrolyzed their ATP. Whether the subunit at each end of a filament is in the T or the D form depends on the rate of this hydrolysis compared with the rate of subunit addition. If the concentration of actin monomers is greater than the critical concentration for both the T-form and D-form polymer, then subunits will add to the polymer at both ends before the nucleotides in the previously added subunits are hydrolyzed; as a result, the tips of the actin filament will remain in the T form. On the other hand, if the subunit concentration is less than the critical concentrations for both the T-form and D-form polymer, then hydrolysis may occur before the next subunit is added and both ends of the filament will be in the D form and will shrink. At intermediate concentrations of actin subunits, it is possible for the rate of subunit addition to be faster than nucleotide hydrolysis at the plus end, but slower than nucleotide hydrolysis at the minus end. In this case, the plus end of the filament remains in the T conformation, while the minus end adopts the D conformation. The filament then undergoes a net addition of subunits at the plus end, while simultaneously losing subunits from the minus end. This leads to the remarkable property of filament **treadmilling** (Figure 16–14; see Panel 16–2).

At a particular intermediate subunit concentration, the filament growth at the plus end exactly balances the filament shrinkage at the minus end. Under these conditions, the subunits cycle rapidly between the free and filamentous states,

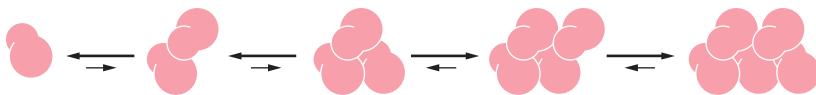
### ON RATES AND OFF RATES

A linear polymer of protein molecules, such as an actin filament or a microtubule, assembles (polymerizes) and disassembles (depolymerizes) by the addition and removal of subunits at the ends of the polymer. The rate of addition of these subunits (called monomers) is given by the rate constant  $k_{on}$ , which has units of  $M^{-1} \text{ sec}^{-1}$ . The rate of loss is given by  $k_{off}$  (units of  $\text{sec}^{-1}$ ).



### NUCLEATION

A helical polymer is stabilized by multiple contacts between adjacent subunits. In the case of actin, two actin molecules bind relatively weakly to each other, but addition of a third actin monomer to form a trimer makes the entire group more stable.



Further monomer addition can take place onto this trimer, which therefore acts as a **nucleus** for polymerization. For tubulin, the nucleus is larger and has a more complicated structure (possibly a ring of 13 or more tubulin molecules)—but the principle is the same.

The assembly of a nucleus is relatively slow, which explains the lag phase seen during polymerization. The lag phase can be reduced or abolished entirely by adding premade nuclei, such as fragments of already polymerized microtubules or actin filaments.

### THE CRITICAL CONCENTRATION

The number of monomers that add to the polymer (actin filament or microtubule) per second will be proportional to the concentration of the free subunit ( $k_{on}C$ ), but the subunits will leave the polymer end at a constant rate ( $k_{off}$ ) that does not depend on  $C$ . As the polymer grows, subunits are used up, and  $C$  is observed to drop until it reaches a constant value, called the **critical concentration** ( $C_c$ ). At this concentration, the rate of subunit addition equals the rate of subunit loss.

At this equilibrium,

$$k_{on} C = k_{off}$$

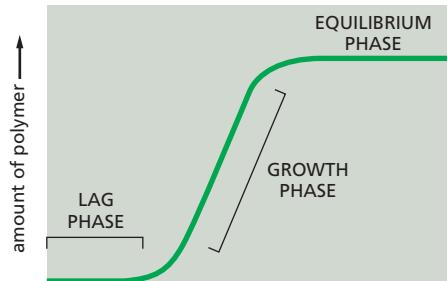
so that

$$C_c = \frac{k_{off}}{k_{on}} = K_d$$

(where  $K_d$  is the dissociation constant; see Figure 3–44).

### TIME COURSE OF POLYMERIZATION

The assembly of a protein into a long helical polymer such as a cytoskeletal filament or a bacterial flagellum typically shows the following time course:



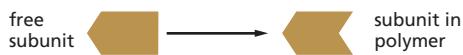
The lag phase corresponds to time taken for nucleation.

The growth phase occurs as monomers add to the exposed ends of the growing filament, causing filament elongation.

The equilibrium phase, or steady state, is reached when the growth of the polymer due to monomer addition precisely balances the shrinkage of the polymer due to disassembly back to monomers.

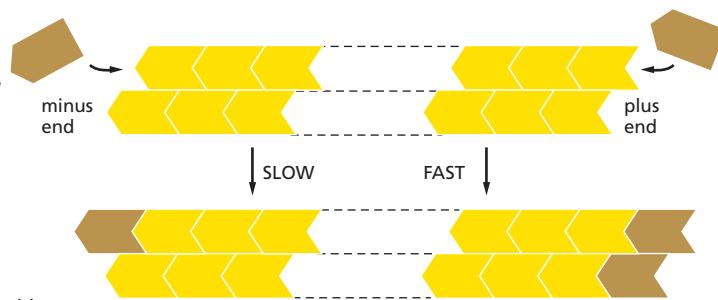
### PLUS AND MINUS ENDS

The two ends of an actin filament or microtubule polymerize at different rates. The fast-growing end is called the **plus end**, whereas the slow-growing end is called the **minus end**. The difference in the rates of growth at the two ends is made possible by changes in the conformation of each subunit as it enters the polymer.



This conformational change affects the rates at which subunits add to the two ends.

Even though  $k_{on}$  and  $k_{off}$  will have different values for the plus and minus ends of the polymer, their ratio  $k_{off}/k_{on}$ —and hence  $C_c$ —must be the same at both ends for a simple polymerization reaction (no ATP or GTP hydrolysis). This is because exactly the same subunit interactions are broken when a subunit is lost at either end, and the final state of the subunit after dissociation is identical. Therefore, the  $\Delta G$  for subunit



loss, which determines the equilibrium constant for its association with the end, is identical at both ends: if the plus end grows four times faster than the minus end, it must also shrink four times faster. Thus, for  $C > C_c$ , both ends grow; for  $C < C_c$ , both ends shrink.

The nucleoside triphosphate hydrolysis that accompanies actin and tubulin polymerization removes this constraint.

### NUCLEOTIDE HYDROLYSIS

Each actin molecule carries a tightly bound ATP molecule that is hydrolyzed to a tightly bound ADP molecule soon after its assembly into the polymer. Similarly, each tubulin molecule carries a tightly bound GTP that is converted to a tightly bound GDP molecule soon after the molecule assembles into the polymer.



Hydrolysis of the bound nucleotide reduces the binding affinity of the subunit for neighboring subunits and makes it more likely to dissociate from each end of the filament (see Figure 16–44 for a possible mechanism). It is usually the form that adds to the filament and the form that leaves.

Considering events at the plus end only:



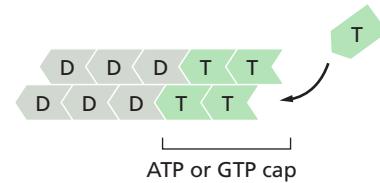
As before, the polymer will grow until  $C = C_c$ . For illustrative purposes, we can ignore  $k^D_{on}$  and  $k^T_{off}$  since they are usually very small, so that polymer growth ceases when

$$k^T_{on} C = k^D_{off} \quad \text{or} \quad C_c = \frac{k^D_{off}}{k^T_{on}}$$

This is a steady state and not a true equilibrium, because the ATP or GTP that is hydrolyzed must be replenished by a nucleotide exchange reaction of the free subunit (  $\rightarrow$  ).

### ATP CAPS AND GTP CAPS

The rate of addition of subunits to a growing actin filament or microtubule can be faster than the rate at which their bound nucleotide is hydrolyzed. Under such conditions, the end has a “cap” of subunits containing the nucleoside triphosphate—an ATP cap on an actin filament or a GTP cap on a microtubule.



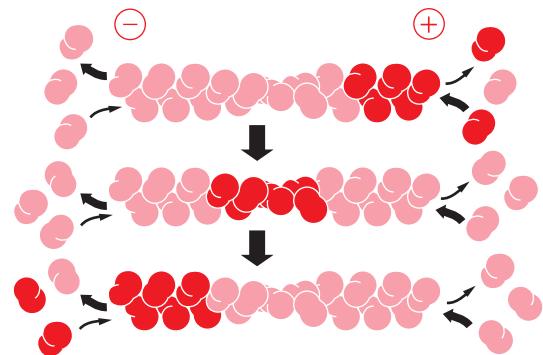
**DYNAMIC INSTABILITY** and **TREADMILLING** are two behaviors observed in cytoskeletal polymers. Both are associated with nucleoside triphosphate hydrolysis. Dynamic instability is believed to predominate in microtubules, whereas treadmilling may predominate in actin filaments.

### TREADMILLING

One consequence of the nucleotide hydrolysis that accompanies polymer formation is to change the critical concentration at the two ends of the polymer. Since  $k^D_{off}$  and  $k^T_{on}$  refer to different reactions, their ratio  $k^D_{off}/k^T_{on}$  need not be the same at both ends of the polymer, so that:

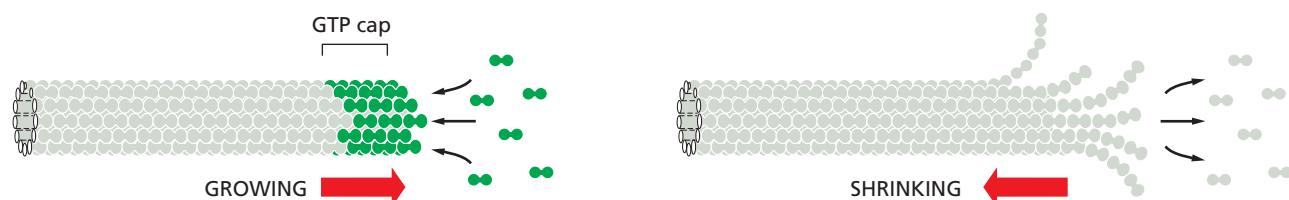
$$C_c (\text{minus end}) > C_c (\text{plus end})$$

Thus, if both ends of a polymer are exposed, polymerization proceeds until the concentration of free monomer reaches a value that is above  $C_c$  for the plus end but below  $C_c$  for the minus end. At this steady state, subunits undergo a net assembly at the plus end and a net disassembly at the minus end at an identical rate. The polymer maintains a constant length, even though there is a net flux of subunits through the polymer, known as **treadmilling**.

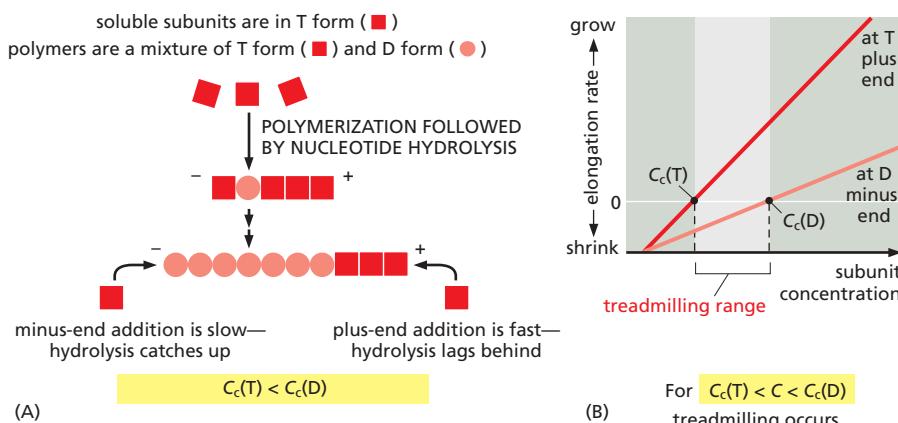


### DYNAMIC INSTABILITY

Microtubules depolymerize about 100 times faster from an end containing GDP-tubulin than from one containing GTP-tubulin. A GTP cap favors growth, but if it is lost, then depolymerization ensues.



Individual microtubules can therefore alternate between a period of slow growth and a period of rapid disassembly, a phenomenon called **dynamic instability**.



**Figure 16-14** Treadmilling of an actin filament, made possible by the ATP hydrolysis that follows subunit addition.

(A) Explanation for the different critical concentrations ( $C_c$ ) at the plus and minus ends. Subunits with bound ATP (T-form subunits) polymerize at both ends of a growing filament, and then undergo nucleotide hydrolysis within the filament. As the filament grows, elongation is faster than hydrolysis at the plus end in this example, and the terminal subunits at this end are therefore always in the T form. However, hydrolysis is faster than elongation at the minus end, and so terminal subunits at this end are in the D form. (B) Treadmilling occurs at intermediate concentrations of free subunits. The critical concentration for polymerization on a filament end in the T form is lower than for a filament end in the D form. If the actual subunit concentration is somewhere between these two values, the plus end grows while the minus end shrinks, resulting in treadmilling.

while the total length of the filament remains unchanged. This “steady-state treadmilling” requires a constant consumption of energy in the form of ATP hydrolysis.

### The Functions of Actin Filaments Are Inhibited by Both Polymer-stabilizing and Polymer-destabilizing Chemicals

Chemical compounds that stabilize or destabilize actin filaments are important tools in studies of the filaments’ dynamic behavior and function in cells. The *cytochalasins* are fungal products that prevent actin polymerization by binding to the plus end of actin filaments. *Latrunculin* prevents actin polymerization by binding to actin subunits. The *phalloidins* are toxins isolated from the *Amanita* mushroom that bind tightly all along the side of actin filaments and stabilize them against depolymerization. All of these compounds cause dramatic changes in the actin cytoskeleton and are toxic to cells, indicating that the function of actin filaments depends on a dynamic equilibrium between filaments and actin monomers (Table 16-1).

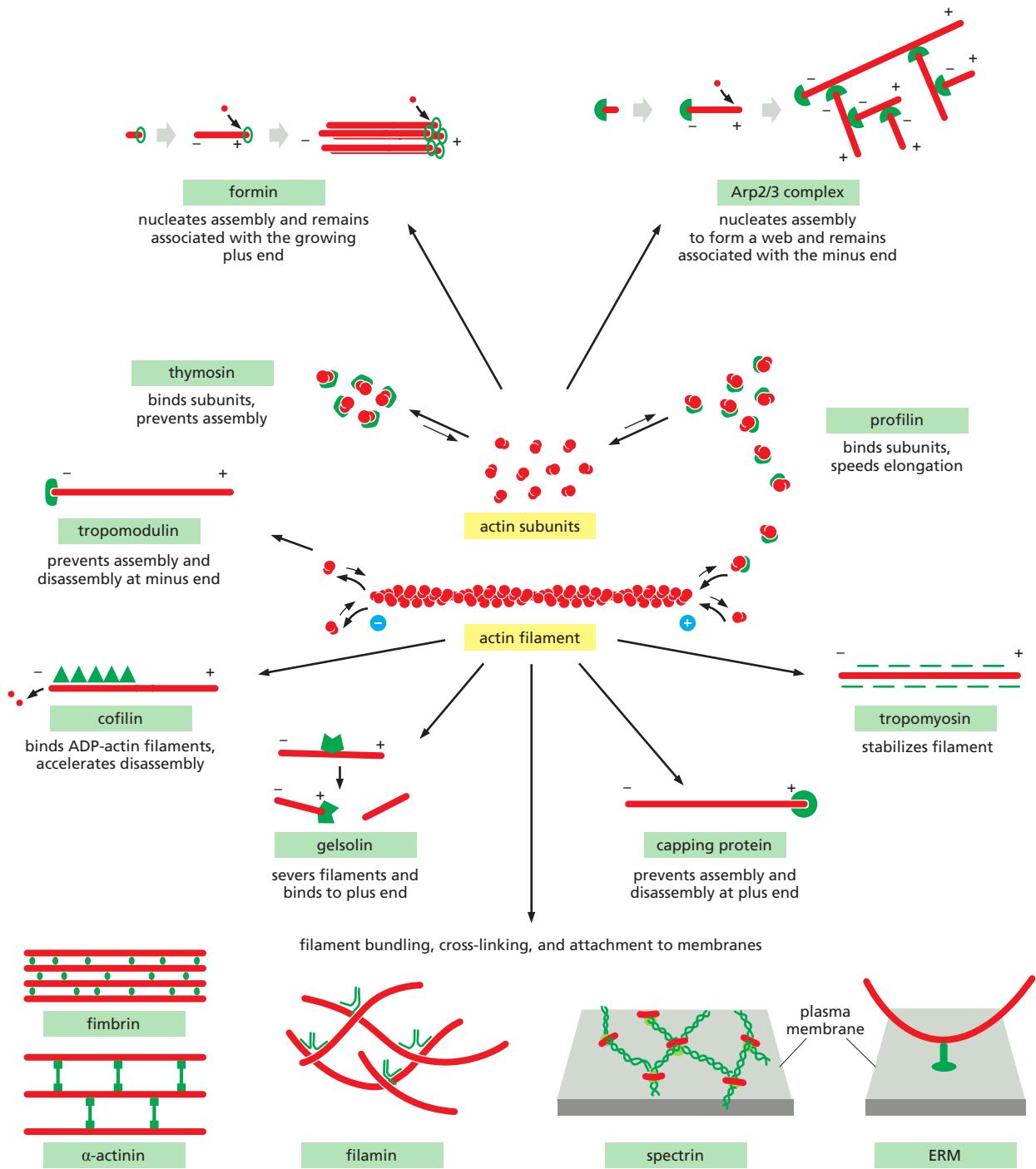
### Actin-Binding Proteins Influence Filament Dynamics and Organization

In a test tube, polymerization of actin is controlled simply by its concentration, as described above, and by pH and the concentrations of salts and ATP. Within a cell, however, actin behavior is also regulated by numerous accessory proteins that bind actin monomers or filaments (summarized in Panel 16-3). At steady state

**TABLE 16-1 Chemical Inhibitors of Actin and Microtubules**

Chemical	Effect on filaments	Mechanism	Original source
<b>Actin</b>			
Latrunculin	Depolymerizes	Binds actin subunits	Sponges
Cytochalasin B	Depolymerizes	Caps filament plus ends	Fungi
Phalloidin	Stabilizes	Binds along filaments	<i>Amanita</i> mushroom
<b>Microtubules</b>			
Taxol® (paclitaxel)	Stabilizes	Binds along filaments	Yew tree
Nocodazole	Depolymerizes	Binds tubulin subunits	Synthetic
Colchicine	Depolymerizes	Caps filament ends	Autumn crocus

## ACTIN FILAMENTS



Some of the major accessory proteins of the actin cytoskeleton. Except for the myosin motor proteins, an example of each major type is shown. Each of these is discussed in the text. However, most cells contain more than a hundred different actin-binding proteins, and it is likely that there are important types of actin-associated proteins that are not yet recognized.

*in vitro*, when the monomer concentration is 0.2  $\mu\text{M}$ , filament half-life, a measure of how long an individual actin monomer spends in a filament as it treadmills, is approximately 30 minutes. In a non-muscle vertebrate cell, actin half-life in filaments is only 30 seconds, demonstrating that cellular factors modify the dynamic behavior of actin filaments. Actin-binding proteins dramatically alter actin filament dynamics and organization through spatial and temporal control of monomer availability, filament nucleation, elongation, and depolymerization. In the following sections, we describe the ways in which these accessory proteins modify actin function in the cell.

### Monomer Availability Controls Actin Filament Assembly

In most non-muscle vertebrate cells, approximately 50% of the actin is in filaments and 50% is soluble—and yet the soluble monomer concentration is 50–200  $\mu\text{M}$ , well above the critical concentration. Why does so little of the actin polymerize into filaments? The reason is that the cell contains proteins that bind to the actin monomers and make polymerization much less favorable (an action similar to that of the drug latrunculin). A small protein called *thymosin* is the most abundant of these proteins. Actin monomers bound to thymosin are in a locked state, where they cannot associate with either the plus or minus ends of actin filaments and can neither hydrolyze nor exchange their bound nucleotide.

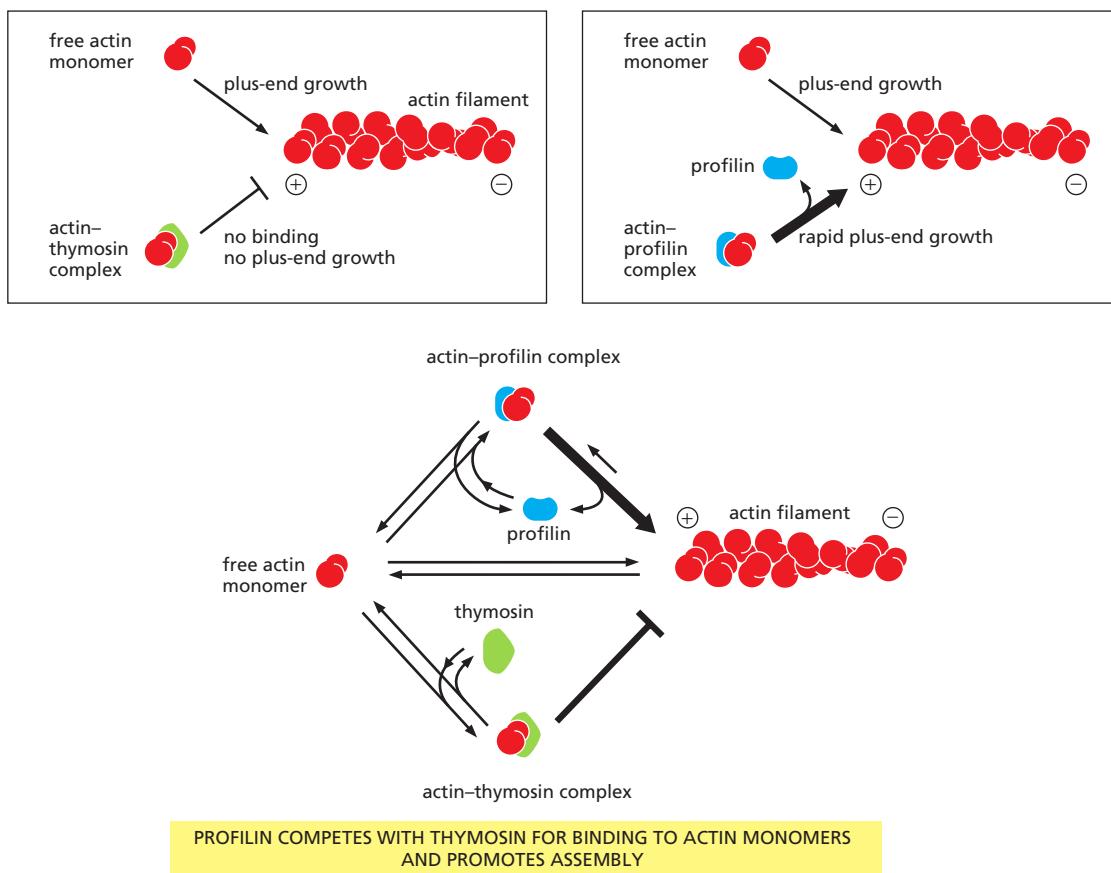
How do cells recruit actin monomers from this buffered storage pool and use them for polymerization? The answer depends on another monomer-binding protein called *profilin*. Profilin binds to the face of the actin monomer opposite the ATP-binding cleft, blocking the side of the monomer that would normally associate with the filament minus end, while leaving exposed the site on the monomer that binds to the plus end (Figure 16–15). When the profilin–actin complex binds a free plus end, a conformational change in actin reduces its affinity for profilin and the profilin falls off, leaving the actin filament one subunit longer. Profilin competes with thymosin for binding to individual actin monomers. Thus, by regulating the local activity of profilin, cells can control the movement of actin subunits from the sequestered thymosin-bound pool onto filament plus ends.

Several mechanisms regulate profilin activity, including profilin phosphorylation and profilin binding to inositol phospholipids. These mechanisms can define the sites where profilin acts. For example, profilin is required for filament assembly at the plasma membrane, where it is recruited by an interaction with acidic membrane phospholipids. At this location, extracellular signals can activate profilin to produce local actin polymerization and the extension of actin-rich motile structures such as filopodia and lamellipodia.

### Actin-Nucleating Factors Accelerate Polymerization and Generate Branched or Straight Filaments

In addition to the availability of active actin subunits, a second prerequisite for cellular actin polymerization is filament nucleation. Proteins that contain actin monomer binding motifs linked in tandem mediate the simplest mechanism of filament nucleation. These actin-nucleating proteins bring several actin subunits together to form a seed. In most cases, actin nucleation is catalyzed by one of two different types of factors: the Arp 2/3 complex or the formins. The first of these is a complex of proteins that includes two *actin-related proteins*, or ARPs, each of which is about 45% identical to actin. The **Arp 2/3 complex** nucleates actin filament growth from the minus end, allowing rapid elongation at the plus end (Figure 16–16A and B). The complex can attach to the side of another actin filament while remaining bound to the minus end of the filament that it has nucleated, thereby building individual filaments into a treelike web (Figure 16–16C and D).

**Formins** are dimeric proteins that nucleate the growth of straight, unbranched filaments that can be cross-linked by other proteins to form parallel bundles. Each formin subunit has a binding site for monomeric actin, and the formin dimer appears to nucleate actin filament polymerization by capturing two monomers.



**Figure 16–15 Effects of thymosin and profilin on actin polymerization.** An actin monomer bound to thymosin is sterically prevented from binding to and elongating the plus end of an actin filament (*left*). An actin monomer bound to profilin, on the other hand, is capable of elongating a filament (*right*). Thymosin and profilin cannot both bind to a single actin monomer at the same time. In a cell in which most of the actin monomer is bound to thymosin, the activation of a small amount of profilin can produce rapid filament assembly. As indicated (*bottom*), profilin binds to actin monomers that are transiently released from the thymosin-bound monomer pool, shuttles them onto the plus ends of actin filaments, and is then released and recycled for further rounds of filament elongation.

As the newly nucleated filament grows, the formin dimer remains associated with the rapidly growing plus end while still allowing the addition of new subunits at that end (**Figure 16–17**). This mechanism of filament assembly is clearly different from that used by the Arp 2/3 complex, which remains stably bound to the filament minus end, preventing subunit addition or loss at that end. Formin-dependent actin filament growth is strongly enhanced by the association of actin monomers with profilin (**Figure 16–18**).

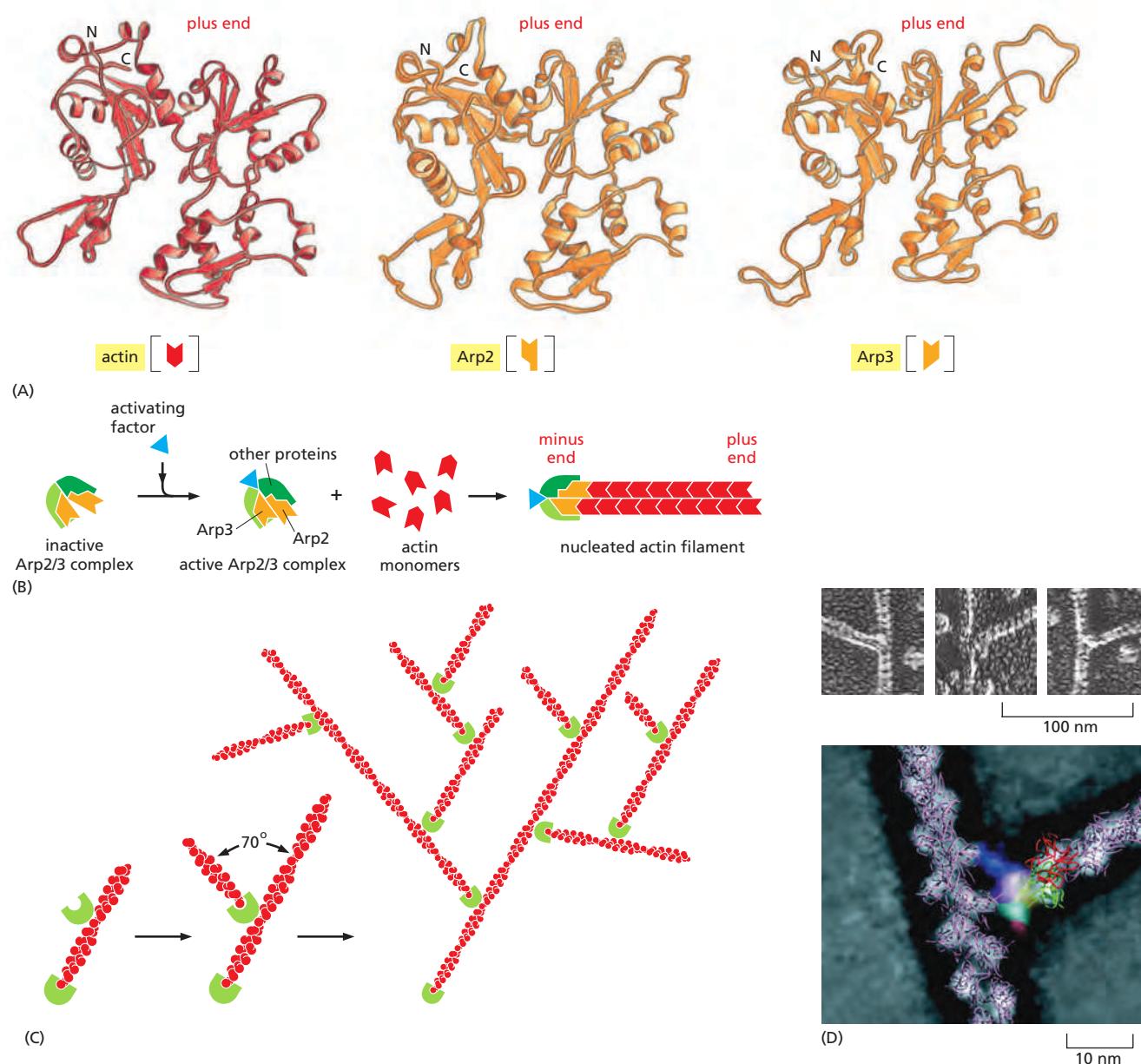
Like profilin activation, actin filament nucleation by Arp 2/3 complexes and formins occurs primarily at the plasma membrane, and the highest density of actin filaments in most cells is at the cell periphery. The layer just beneath the plasma membrane is called the **cell cortex**, and the actin filaments in this region determine the shape and movement of the cell surface, allowing the cell to change its shape and stiffness rapidly in response to changes in its external environment.

### Actin-Filament-Binding Proteins Alter Filament Dynamics

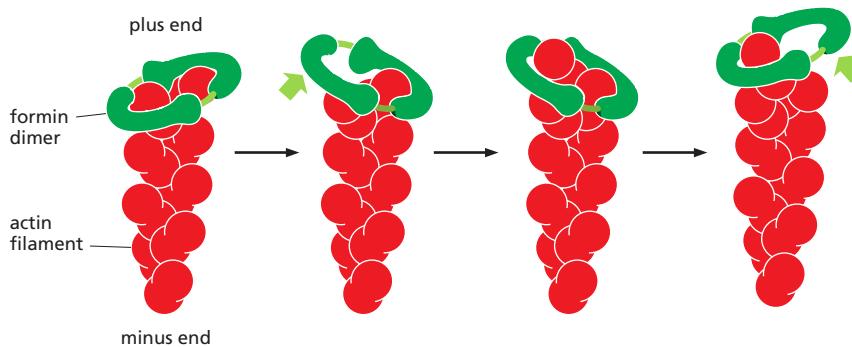
Actin filament behavior is regulated by two major classes of binding proteins: those that bind along the side of a filament and those that bind to the ends (see Panel 16–3). Side-binding proteins include *tropomyosin*, an elongated protein that binds simultaneously to six or seven adjacent actin subunits along each of the two grooves of the helical actin filament. In addition to stabilizing and stiffening

the filament, the binding of tropomyosin can prevent the actin filament from interacting with other proteins; this aspect of tropomyosin function is important in the control of muscle contraction, as we discuss later.

An actin filament that stops growing and is not specifically stabilized in the cell will depolymerize rapidly, particularly at its plus end, once the actin molecules



**Figure 16–16 Nucleation and actin web formation by the Arp 2/3 complex.** (A) The structures of Arp2 and Arp3 compared to the structure of actin. Although the face of the molecule equivalent to the plus end (top) in both Arp2 and Arp3 is very similar to the plus end of actin itself, differences on the sides and minus end prevent these actin-related proteins from forming filaments on their own or coassembling into filaments with actin. (B) A model for actin filament nucleation by the Arp 2/3 complex. In the absence of an activating factor, Arp2 and Arp3 are held by their accessory proteins in an orientation that prevents them from nucleating a new actin filament. When an activating factor (indicated by the blue triangle) binds the complex, Arp2 and Arp3 are brought together into a new configuration that resembles the plus end of an actin filament. Actin subunits can then assemble onto this structure, bypassing the rate-limiting step of filament nucleation. (C) The Arp 2/3 complex nucleates filaments most efficiently when it is bound to the side of a preexisting actin filament. The result is a filament branch that grows at a 70° angle relative to the original filament. Repeated rounds of branching nucleation result in a treelike web of actin filaments. (D, top) electron micrographs of branched actin filaments formed by mixing purified actin subunits with purified Arp 2/3 complexes. (D, bottom) Reconstructed image of a branch where the crystal structures of actin (pink) and the Arp 2/3 complex have been fitted to the electron density. The mother filament runs from top to bottom, and the daughter filament branches off to the right where the Arp 2/3 complex binds to three actin subunits in the mother filament. (D, top, from R.D. Mullins et al., *Proc. Natl. Acad. Sci. USA* 95:6181–6186, 1998, with permission from National Academy of Sciences; bottom, from N. Volkmann et al., *Science* 293:2456–2459, 2001, with permission from AAAS.)



have hydrolyzed their ATP. The binding of plus-end *capping protein* (also called *CapZ* for its location in the muscle Z band) stabilizes an actin filament at its plus end by rendering it inactive, greatly reducing the rates of filament growth and depolymerization (Figure 16–19). At the minus end, an actin filament may be capped by the Arp 2/3 complex that was responsible for its nucleation, although many minus ends in a typical cell are released from the Arp 2/3 complex and are uncapped.

*Tropomodulin*, best known for its function in the capping of exceptionally long-lived actin filaments in muscle, binds tightly to the minus ends of actin filaments that have been coated and thereby stabilized by tropomyosin. It can also transiently cap pure actin filaments and significantly reduce their elongation and depolymerization rates. A large family of tropomodulin proteins regulates actin filament length and stability in many cell types.

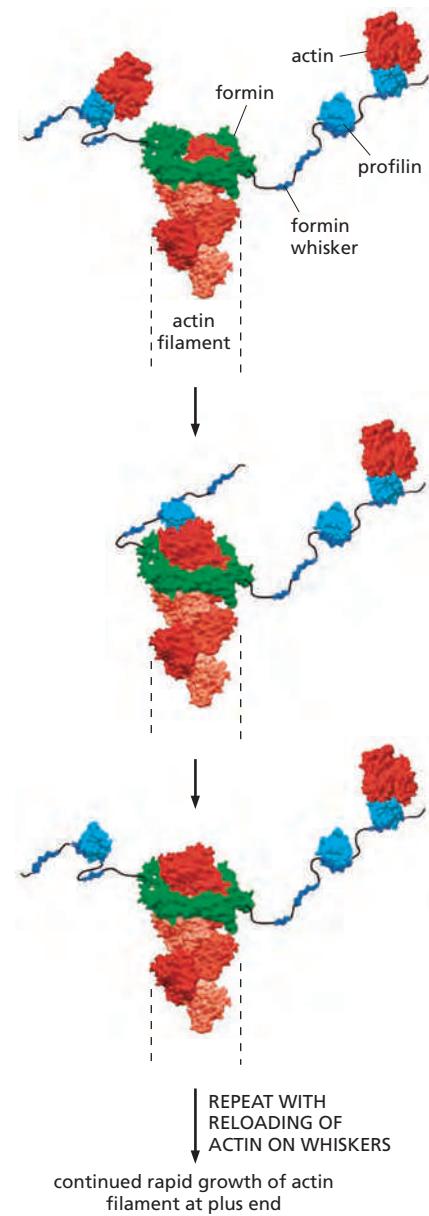
For maximum effect, proteins that bind the side of actin filaments coat the filament completely, and must therefore be present in high amounts. In contrast, end-binding proteins can affect filament dynamics even when they are present at very low levels. Since subunit addition and loss occur primarily at filament ends, one molecule of an end-binding protein per actin filament (roughly one molecule per 200–500 actin subunits) can be enough to transform the architecture of an actin filament network.

### Severing Proteins Regulate Actin Filament Depolymerization

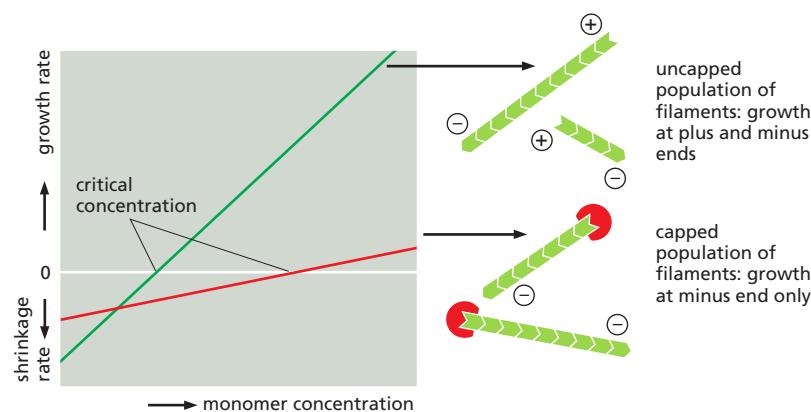
Another important mechanism of actin filament regulation depends on proteins that break an actin filament into many smaller filaments, thereby generating a large number of new filament ends. The fate of these new ends depends on the presence of other accessory proteins. Under some conditions, newly formed ends nucleate filament elongation, thereby accelerating the assembly of new filament structures. Under other conditions, severing promotes the depolymerization of old filaments, speeding up the depolymerization rate by tenfold or more. In addition, severing changes the physical and mechanical properties of the cytoplasm: stiff, large bundles and gels become more fluid.

One class of actin-severing proteins is the *gelsolin superfamily*. These proteins are activated by high levels of cytosolic  $\text{Ca}^{2+}$ . Gelsolin interacts with the side of the actin filament and contains subdomains that bind to two different sites: one that is exposed on the surface of the filament and one that is hidden between adjacent

**Figure 16–17** Actin elongation mediated by formins. Formin proteins (green) form a dimeric complex that can nucleate the formation of a new actin filament (red) and remain associated with the rapidly growing plus end as it elongates. The formin protein maintains its binding to one of the two actin subunits exposed at the plus end as it allows each new subunit to assemble. Only part of the large dimeric formin molecule is shown here. Other regions regulate its activity and link it to particular structures in the cell. Many formins are indirectly connected to the cell plasma membrane and aid the insertional polymerization of the actin filament directly beneath the membrane surface.



**Figure 16–18** Profilin and formins. Some members of the formin protein family have unstructured domains or "whiskers" that contain several binding sites for profilin or the profilin–actin complex. These flexible domains serve as a staging area for addition of actin to the growing plus end of the actin filament when formin is bound. Under some conditions, this can enhance the rate of actin filament elongation so that filament growth is faster than that expected for a diffusion-controlled reaction, and faster in the presence of formin and profilin than the rate for pure actin alone (see also Figure 3–78).

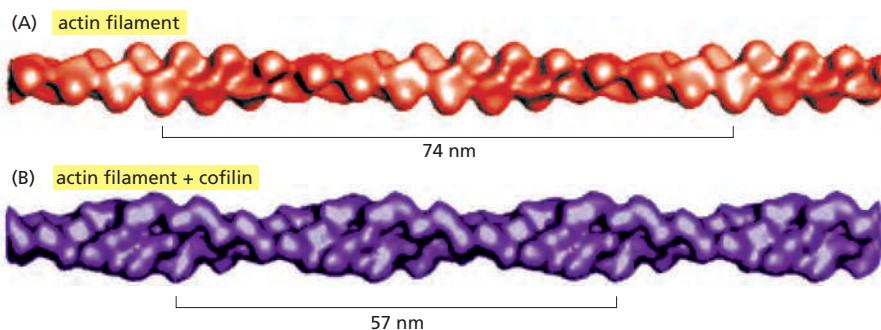


**Figure 16–19** Filament capping and its effects on filament dynamics. A population of uncapped filaments adds and loses subunits at both the plus and minus ends, resulting in rapid growth or shrinkage, depending on the concentration of available free monomers (green line). In the presence of a protein that caps the plus end (red line), only the minus end is able to add or lose subunits; consequently, filament growth will be slower at all monomer concentrations above the critical concentration, and filament shrinkage will be slower at all monomer concentrations below the critical concentration. In addition, the critical concentration for the population shifts to that of the filament minus end.

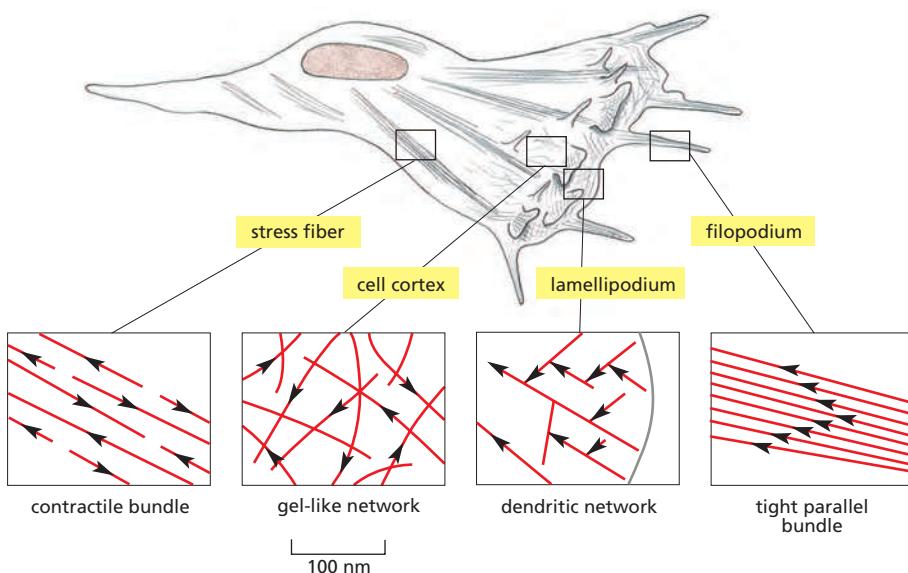
subunits. According to one model, gelsolin binds the side of an actin filament until a thermal fluctuation creates a small gap between neighboring subunits, at which point gelsolin inserts itself into the gap to break the filament. After the severing event, gelsolin remains attached to the actin filament and caps the new plus end.

Another important actin-filament destabilizing protein, found in all eukaryotic cells, is *cofilin*. Also called *actin depolymerizing factor*, cofilin binds along the length of the actin filament, forcing the filament to twist a little more tightly (Figure 16–20). This mechanical stress weakens the contacts between actin subunits in the filament, making the filament brittle and more easily severed by thermal motions, generating filament ends that undergo rapid disassembly. As a result, most of the actin filaments inside cells are shorter lived than are filaments formed from pure actin in a test tube.

Cofilin binds preferentially to ADP-containing actin filaments rather than to ATP-containing filaments. Since ATP hydrolysis is usually slower than filament assembly, the newest actin filaments in the cell still contain mostly ATP and are resistant to depolymerization by cofilin. Cofilin therefore tends to dismantle the older filaments in the cell. As we will discuss later, the cofilin-mediated disassembly of old but not new actin filaments is critical for the polarized, directed growth of the actin network that is responsible for unidirectional cell crawling and the intracellular motility of pathogens. Actin filaments can be protected from cofilin by tropomyosin binding. Thus, the dynamics of actin in different subcellular locations depends on the balance of stabilizing and destabilizing accessory proteins.



**Figure 16–20** Twisting of an actin filament induced by cofilin. (A) Three-dimensional reconstruction from cryoelectron micrographs of filaments made of pure actin. The bracket shows the span of two twists of the actin helix. (B) Reconstruction of an actin filament coated with cofilin, which binds in a 1:1 stoichiometry to actin subunits all along the filament. Cofilin is a small protein (14 kD) compared to actin (43 kD), and so the filament appears only slightly thicker. The energy of cofilin binding serves to deform the actin filament, twisting it more tightly and reducing the distance spanned by each twist of the helix. (From A. McGough et al., J. Cell Biol. 138:771–781, 1997. With permission from the authors.)



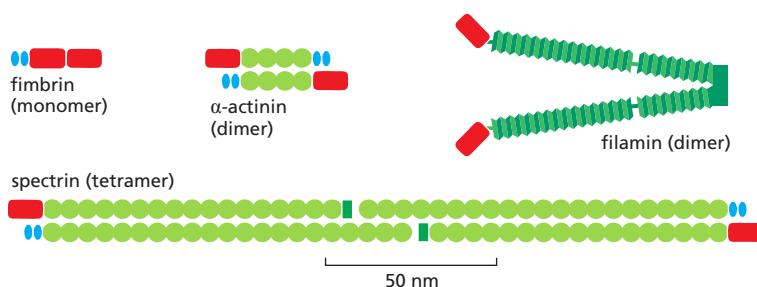
**Figure 16–21 Actin arrays in a cell.** A fibroblast crawling in a tissue-culture dish is shown with four areas enlarged to show the arrangement of actin filaments. The actin filaments are shown in red, with arrowheads pointing toward the minus end. Stress fibers are contractile and exert tension. The actin cortex underlies the plasma membrane and consists of gel-like networks or dendritic actin networks that enable membrane protrusion at lamellipodia. Filopodia are spike-like projections of the plasma membrane that allow a cell to explore its environment.

## Higher-Order Actin Filament Arrays Influence Cellular Mechanical Properties and Signaling

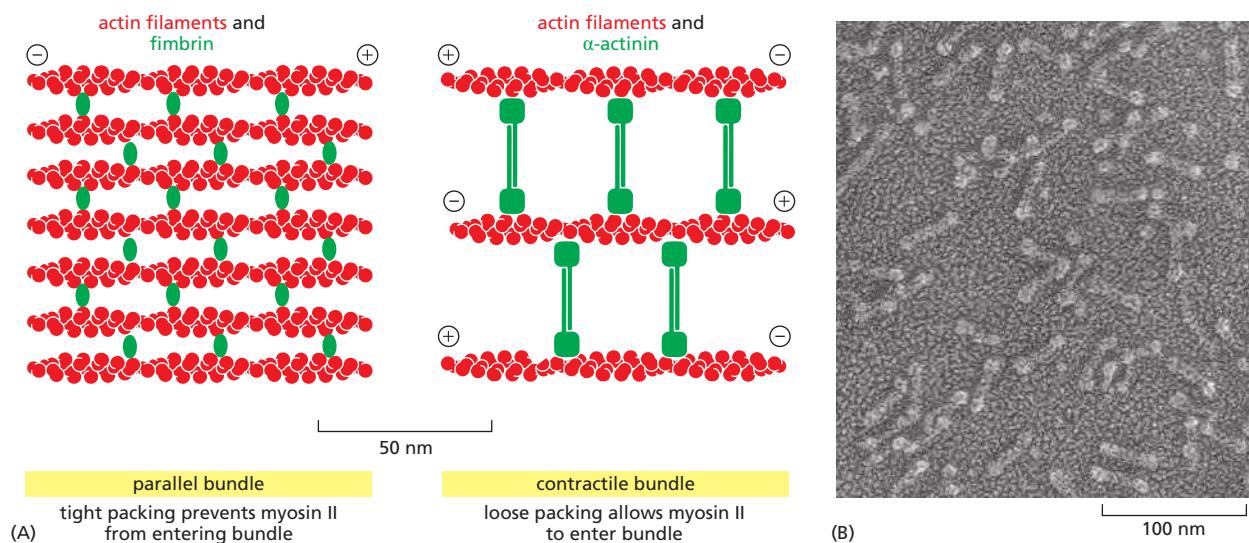
Actin filaments in animal cells are organized into several types of arrays: dendritic networks, bundles, and weblike (gel-like) networks (Figure 16–21). Different structures are initiated by the action of distinct nucleating proteins: the actin filaments of dendritic networks are nucleated by the Arp 2/3 complex, while bundles are made of the long, straight filaments produced by formins. The proteins nucleating the filaments in the gel-like networks are not yet well defined.

The structural organization of different actin networks depends on specialized accessory proteins. As explained earlier, Arp 2/3 organizes filaments into dendritic networks by attaching filament minus ends to the side of other filaments. Other actin filament structures are assembled and maintained by two classes of proteins: *bundling proteins*, which cross-link actin filaments into a parallel array, and *gel-forming proteins*, which hold two actin filaments together at a large angle to each other, thereby creating a looser meshwork. Both bundling and gel-forming proteins generally have two similar actin-filament-binding sites, which can either be part of a single polypeptide chain or contributed by each of two polypeptide chains held together in a dimer (Figure 16–22). The spacing and arrangement of these two filament-binding domains determine the type of actin structure that a given cross-linking protein forms.

Each type of bundling protein also determines which other molecules can interact with the cross-linked actin filaments. Myosin II is the motor protein that enables stress fibers and other contractile arrays to contract. The very close packing of actin filaments caused by the small monomeric bundling protein *fimbrin* apparently excludes myosin, and thus the parallel actin filaments held together by fimbrin are not contractile. On the other hand,  $\alpha$ -actinin cross-links oppositely



**Figure 16–22 The modular structures of four actin-cross-linking proteins.** Each of the proteins shown has two actin-binding sites (red) that are related in sequence. Fimbrin has two directly adjacent actin-binding sites, so that it holds its two actin filaments very close together (14 nm apart), aligned with the same polarity (see Figure 16–23A). The two actin-binding sites in  $\alpha$ -actinin are separated by a spacer around 30 nm long, so that it forms more loosely packed actin bundles (see Figure 16–23A). Filamin has two actin-binding sites with a V-shaped linkage between them, so that it cross-links actin filaments into a network with the filaments oriented almost at right angles to one another (see Figure 16–24). Spectrin is a tetramer of two  $\alpha$  and two  $\beta$  subunits, and the tetramer has two actin-binding sites spaced about 200 nm apart (see Figure 10–38).

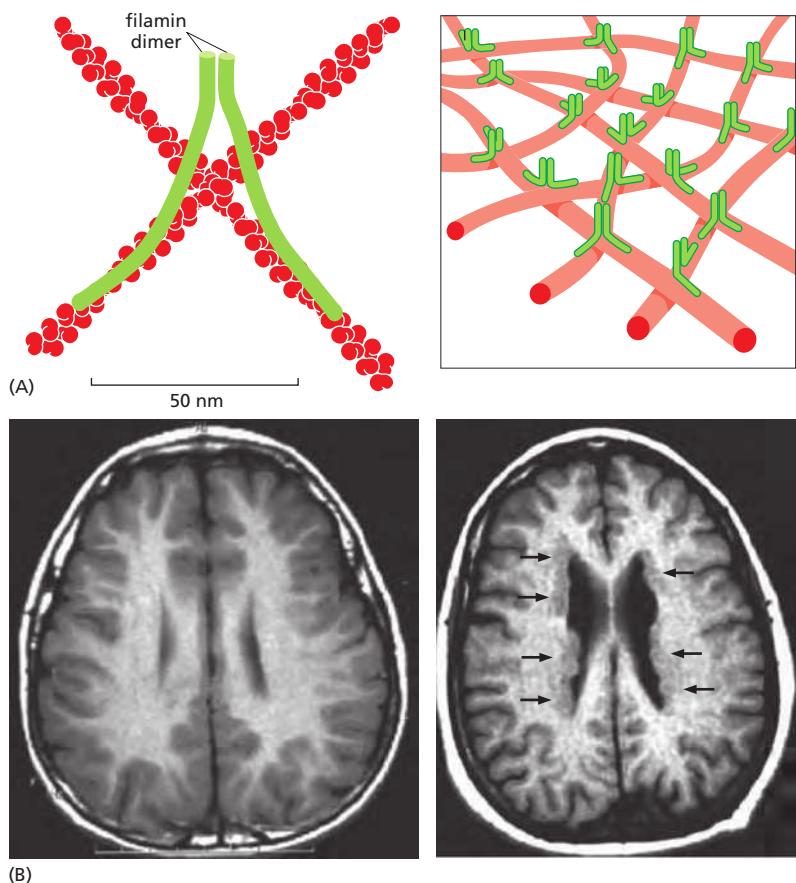


polarized actin filaments into loose bundles, allowing the binding of myosin and formation of contractile actin bundles (Figure 16–23). Because of the very different spacing and orientation of the actin filaments, bundling by fimbrin automatically discourages bundling by  $\alpha$ -actinin, and vice versa, so that the two types of bundling protein are mutually exclusive.

The bundling proteins that we have discussed so far have straight, stiff connections between their two actin-filament-binding domains. Other actin cross-linking proteins have either a flexible or a stiff, bent connection between their two binding domains, allowing them to form actin filament webs or gels, rather than actin bundles. Filamin (see Figure 16–22) promotes the formation of a loose and highly viscous gel by clamping together two actin filaments roughly at right angles (Figure 16–24A). Cells require the actin gels formed by filamin to extend the thin, sheetlike membrane projections called *lamellipodia* that help them to crawl across solid surfaces. In humans, mutations in the filamin A gene cause defects in nerve-cell migration during early embryonic development. Cells in the periventricular region of the brain fail to migrate to the cortex and instead form nodules, causing a syndrome called periventricular heterotopia (Figure 16–24B). Interestingly, in addition to binding actin, filamins have been reported to interact with a large number of cellular proteins of great functional diversity, including membrane receptors for signaling molecules, and filamin mutations can also lead to defects in development of bone, the cardiovascular system, and other organs. Thus, filamins may also function as signaling scaffolds by connecting and coordinating a wide variety of cellular processes with the actin cytoskeleton.

A very different, well-studied web-forming protein is *spectrin*, which was first identified in red blood cells. Spectrin is a long, flexible protein made out of four elongated polypeptide chains (two  $\alpha$  subunits and two  $\beta$  subunits), arranged so that the two actin-filament-binding sites are about 200 nm apart (compared with 14 nm for fimbrin and about 30 nm for  $\alpha$ -actinin; see Figure 16–23). In the red blood cell, spectrin is concentrated just beneath the plasma membrane, where it forms a two-dimensional weblike network held together by short actin filaments whose precise lengths are tightly regulated by capping proteins at each end; spectrin links this web to the plasma membrane because it has separate binding sites for peripheral membrane proteins, which are themselves positioned near the lipid bilayer by integral membrane proteins (see Figure 10–38). The resulting network creates a strong, yet flexible cell cortex that provides mechanical support for the overlying plasma membrane, allowing the red blood cell to spring back to its original shape after squeezing through a capillary. Close relatives of spectrin are found in the cortex of most other vertebrate cell types, where they also help to shape and stiffen the surface membrane. A particularly striking example of spectrin's role

**Figure 16–23** The formation of two types of actin filament bundles.  
 (A) Fimbrin cross-links actin filaments into tight bundles, which exclude the motor protein myosin II from participating in the assembly. In contrast,  $\alpha$ -actinin, which is a homodimer, cross-links actin filaments into loose bundles, which allow myosin (not shown) to incorporate into the bundle. Fimbrin and  $\alpha$ -actinin tend to exclude one another because of the very different spacing of the actin filament bundles that they form. (B) Electron micrograph of purified  $\alpha$ -actinin molecules. (B, courtesy of John Heuser.)



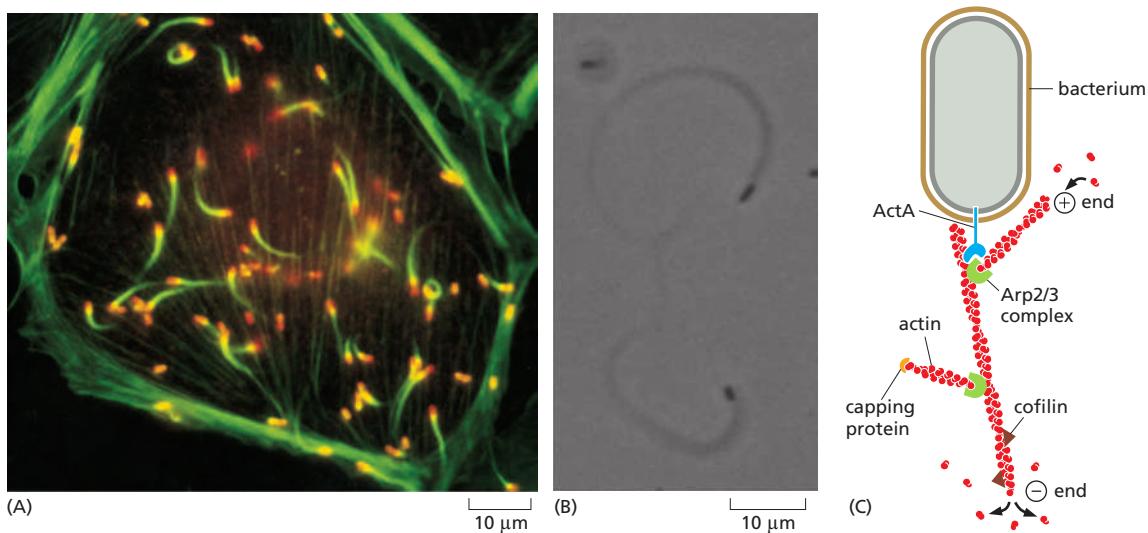
**Figure 16-24** Filamin cross-links actin filaments into a three-dimensional network and is required for normal neuronal migration. (A) Each filamin homodimer is about 160 nm long when fully extended and forms a flexible, high-angle link between two adjacent actin filaments. A set of actin filaments cross-linked by filamin forms a mechanically strong web or gel. (B) Magnetic resonance imaging of a normal human brain (left) and of a patient with periventricular heterotopia (right) caused by mutation in the filamin A gene. In contrast to the smooth ventricular surface in the normal brain, a rough zone of cortical neurons (arrowheads) is seen along the lateral walls of the ventricles, representing neurons that have failed to migrate to the cortex during brain development. Remarkably, although many neurons are not in the right place, the intelligence of affected individuals is frequently normal or only mildly compromised, and the major clinical syndrome is epilepsy that often starts in the second decade of life. (B, adapted from Y. Feng and C.A. Walsh, *Nat. Cell Biol.* 6:1034–1038, 2004. With permission from Macmillan Publishers Ltd.)

in promoting mechanical stability is the long, thin axon of neurons in the nematode worm *Caenorhabditis elegans*, where spectrin is required to keep them from breaking during the twisting motions the worms make during crawling.

The connections of the cortical actin cytoskeleton to the plasma membrane are only partially understood. Members of the *ERM* family (named for its first three members, ezrin, radixin, and moesin), help organize membrane domains through their ability to interact with transmembrane proteins and the underlying cytoskeleton. In so doing, they not only provide structural links to strengthen the cell cortex, but also regulate the activities of signal transduction pathways. Moesin also increases cortical stiffness to promote cell rounding during mitosis. Measurements by atomic force microscopy indicate that the cell cortex remains soft during mitosis when moesin is depleted. ERM proteins are thought to bind to and organize the cortical actin cytoskeleton in a variety of contexts, thereby affecting the shape and stiffness of the membrane as well as the localization and activity of signaling molecules.

### Bacteria Can Hijack the Host Actin Cytoskeleton

The importance of accessory proteins in actin-based motility and force production is illustrated beautifully by studies of certain bacteria and viruses that use components of the host-cell actin cytoskeleton to move through the cytoplasm. The cytoplasm of mammalian cells is extremely viscous, containing organelles and cytoskeletal elements that inhibit diffusion of large particles like bacteria or viruses. To move around in a cell and invade neighboring cells, several pathogens, including *Listeria monocytogenes* (which causes a rare but serious form of food poisoning), overcome this problem by recruiting and activating the Arp 2/3 complex at their surface. The Arp 2/3 complex nucleates the assembly of actin filaments that generate a substantial force and push the bacterium through the

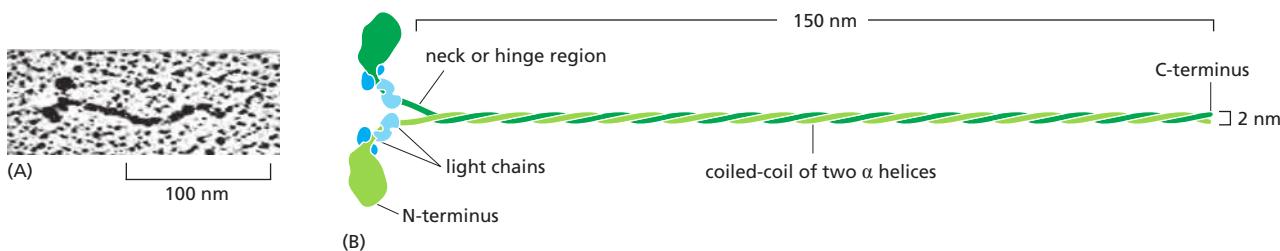


**Figure 16–25** The actin-based movement of *Listeria monocytogenes*. (A) Fluorescence micrograph of an infected cell that has been stained to reveal bacteria in red and actin filaments in green. Note the cometlike tail of actin filaments behind each moving bacterium. Regions of overlap between red and green fluorescence appear yellow. (B) *Listeria* motility can be reconstituted in a test tube with ATP and just four purified proteins: actin, Arp 2/3 complex, capping protein, and cofilin. This micrograph shows the dense actin tails behind bacteria (black). (C) The ActA protein on the bacterial surface activates the Arp 2/3 complex to nucleate new filament assembly along the sides of existing filaments. Filaments grow at their plus end until capped by capping protein. Actin is recycled through the action of cofilin, which enhances depolymerization at the minus ends of the filaments. By this mechanism, polymerization is focused at the rear surface of the bacterium, propelling it forward (see Movie 23.7). (A, courtesy of Julie Theriot and Tim Mitchison; B, from T.P. Loisel et al., *Nature* 401:613–616, 1999. With permission from Macmillan Publishers Ltd.)

cytoplasm at rates of up to 1  $\mu\text{m/sec}$ , leaving behind a long actin “comet tail” (Figure 16–25; see also Figures 23–28 and 23–29). This motility can be reconstituted in a test tube by adding the bacteria to a mixture of pure actin, Arp 2/3 complex, cofilin, and capping protein, illustrating how actin polymerization dynamics generate movement through spatial regulation of filament assembly and disassembly. As we shall see, actin-based movement of this sort also underlies membrane protrusion at the leading edge of motile cells.

### Summary

Actin is a highly conserved cytoskeletal protein that is present in high concentrations in nearly all eukaryotic cells. Nucleation presents a kinetic barrier to actin polymerization, but once formed, actin filaments undergo dynamic behavior due to hydrolysis of the bound nucleotide ATP. Actin filaments are polarized and can undergo treadmilling when a filament assembles at the plus end while simultaneously depolymerizing at the minus end. In cells, actin filament dynamics are regulated at every step, and the varied forms and functions of actin depend on a versatile repertoire of accessory proteins. Approximately half of the actin is kept in a monomeric form through association with sequestering proteins such as thymosin. Nucleation factors such as the Arp 2/3 complex and formins promote formation of branched and parallel filaments, respectively. Interplay between proteins that bind or cap actin filaments and those that promote filament severing or depolymerization can slow or accelerate the kinetics of filament assembly and disassembly. Another class of accessory proteins assembles the filaments into larger ordered structures by cross-linking them to one another in geometrically defined ways. Connections between these actin arrays and the plasma membrane of cells give an animal cell mechanical strength and permit the elaboration of cortical cellular structures such as lamellipodia, filopodia, and microvilli. By inducing actin filament polymerization at their surface, intracellular pathogens can hijack the host-cell cytoskeleton and move around inside the cell.



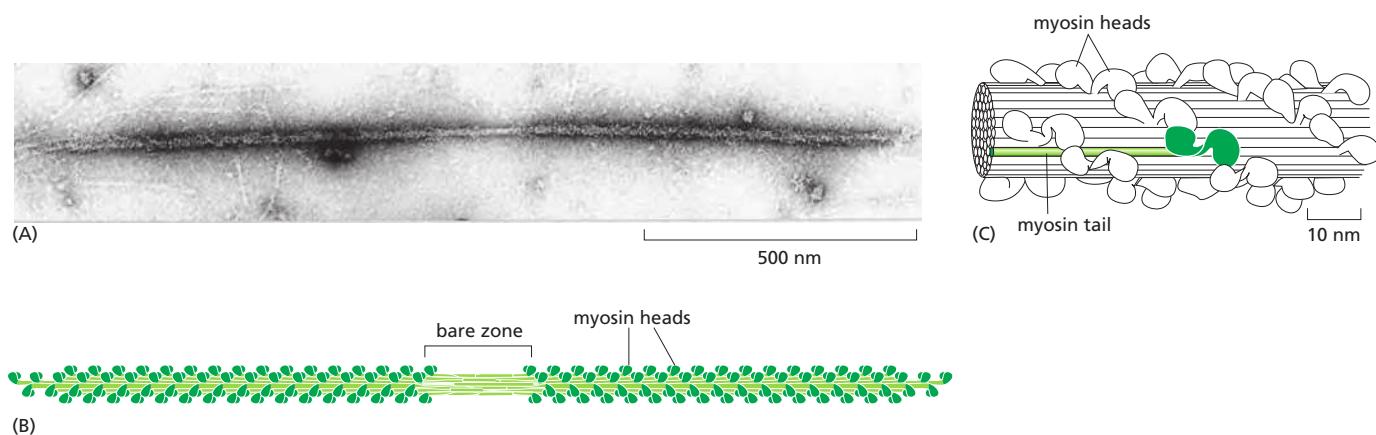
**Figure 16–26 Myosin II.** (A) The two globular heads and long tail of a myosin II molecule shadowed with platinum can be seen in this electron micrograph. (B) A myosin II molecule is composed of two heavy chains (each about 2000 amino acids long; green) and four light chains (blue). The light chains are of two distinct types, and one copy of each type is present on each myosin head. Dimerization occurs when the two  $\alpha$  helices of the heavy chains wrap around each other to form a coiled-coil, driven by the association of regularly spaced hydrophobic amino acids (see Figure 3–9). The coiled-coil arrangement makes an extended rod in solution, and this part of the molecule forms the tail. (A, courtesy of David Shotton.)

## MYOSIN AND ACTIN

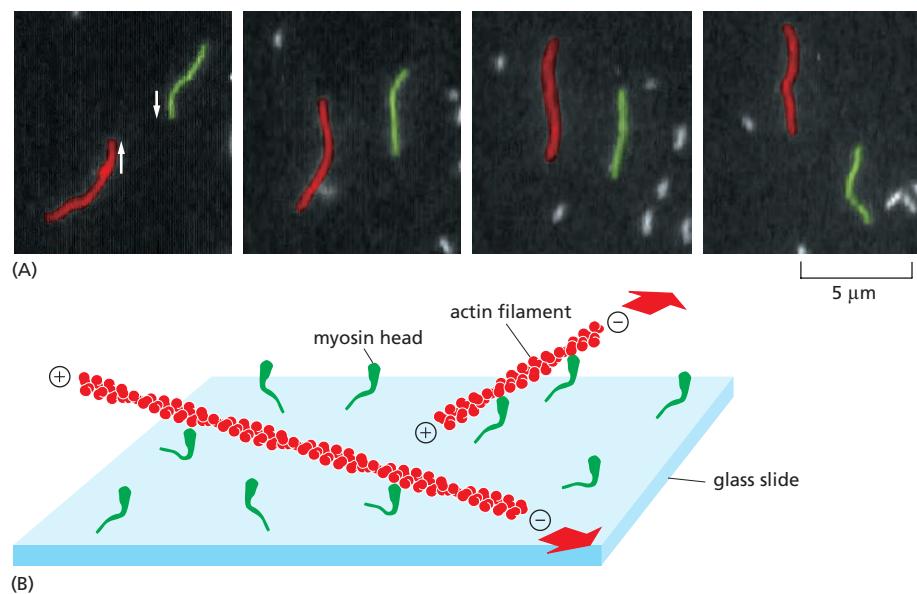
A crucial feature of the actin cytoskeleton is that it can form contractile structures that cross-link and slide actin filaments relative to one another through the action of **myosin** motor proteins. In addition to driving muscle contraction, actin-myosin assemblies perform important functions in non-muscle cells.

### Actin-Based Motor Proteins Are Members of the Myosin Superfamily

The first motor protein to be identified was skeletal muscle myosin, which generates the force for muscle contraction. This protein, now called *myosin II*, is an elongated protein formed from two heavy chains and two copies of each of two light chains. Each heavy chain has a globular head domain at its N-terminus that contains the force-generating machinery, followed by a very long amino acid sequence that forms an extended coiled-coil that mediates heavy-chain dimerization (Figure 16–26). The two light chains bind close to the N-terminal head domain, while the long coiled-coil tail bundles itself with the tails of other myosin molecules. These tail-tail interactions form large, bipolar “thick filaments” that have several hundred myosin heads, oriented in opposite directions at the two ends of the thick filament (Figure 16–27).



**Figure 16–27 The myosin II bipolar thick filament in muscle.** (A) Electron micrograph of a myosin II thick filament isolated from frog muscle. Note the central bare zone, which is free of head domains. (B) Schematic diagram, not drawn to scale. The myosin II molecules aggregate by means of their tail regions, with their heads projecting to the outside of the filament. The bare zone in the center of the filament consists entirely of myosin II tails. (C) A small section of a myosin II filament as reconstructed from electron micrographs. An individual myosin molecule is highlighted in green. The cytoplasmic myosin II filaments in non-muscle cells are much smaller, although similarly organized (see Figure 16–39). (A, courtesy of Murray Stewart; C, based on R.A. Crowther, R. Padrón and R. Craig, *J. Mol. Biol.* 184:429–439, 1985. With permission from Academic Press.)



**Figure 16–28 Direct evidence for the motor activity of the myosin head.** In this experiment, purified myosin heads were attached to a glass slide, and then actin filaments labeled with fluorescent phalloidin were added and allowed to bind to the myosin heads. (A) When ATP was added, the actin filaments began to glide along the surface, owing to the many individual steps taken by each of the dozens of myosin heads bound to each filament. The video frames shown in this sequence were recorded about 0.6 second apart; the two actin filaments shown (one red and one green) were moving in opposite directions at a rate of about 4  $\mu\text{m/sec}$ . (B) Diagram of the experiment. The large red arrows indicate the direction of actin filament movement (Movie 16.2). (A, courtesy of James Spudich.)

Each myosin head binds and hydrolyzes ATP, using the energy of ATP hydrolysis to walk toward the plus end of an actin filament (Figure 16–28). The opposing orientation of the heads in the thick filament makes the filament efficient at sliding pairs of oppositely oriented actin filaments toward each other, shortening the muscle. In skeletal muscle, in which carefully arranged actin filaments are aligned in “thin filament” arrays surrounding the myosin thick filaments, the ATP-driven sliding of actin filaments results in a powerful contraction. Cardiac and smooth muscle contain myosin II molecules that are similarly arranged, although different genes encode them.

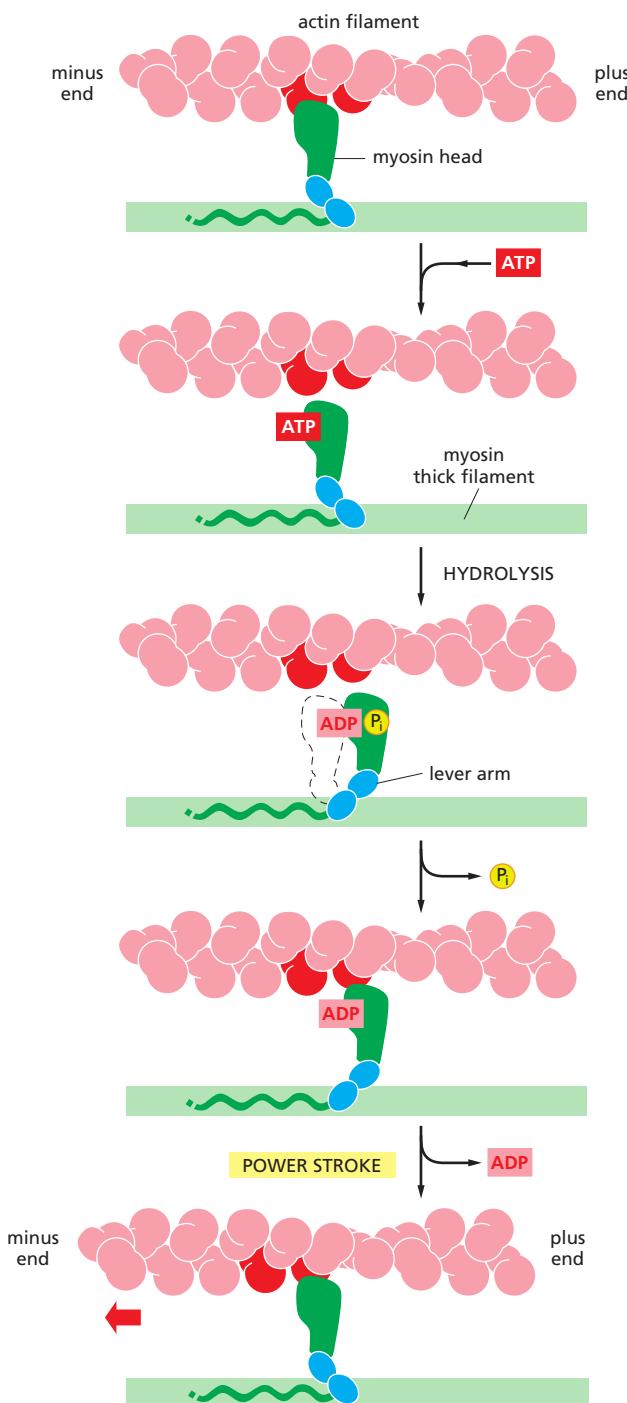
### Myosin Generates Force by Coupling ATP Hydrolysis to Conformational Changes

Motor proteins use structural changes in their ATP-binding sites to produce cyclic interactions with a cytoskeletal filament. Each cycle of ATP binding, hydrolysis, and release propels them forward in a single direction to a new binding site along the filament. For myosin II, each step of the movement along actin is generated by the swinging of an 8.5-nm-long  $\alpha$  helix, or *lever arm*, which is structurally stabilized by the binding of light chains. At the base of this lever arm next to the head, there is a pistonlike helix that connects movements at the ATP-binding cleft in the head to small rotations of the so-called converter domain. A small change at this point can swing the helix like a long lever, causing the far end of the helix to move by about 5.0 nm.

These changes in the conformation of the myosin are coupled to changes in its binding affinity for actin, allowing the myosin head to release its grip on the actin filament at one point and snatch hold of it again at another. The full mechanochemical cycle of nucleotide binding, nucleotide hydrolysis, and phosphate release (which causes the “power stroke”) produces a single step of movement (Figure 16–29). At low ATP concentrations, the interval between the force-producing step and the binding of the next ATP is long enough that single steps can be observed (Figure 16–30).

### Sliding of Myosin II Along Actin Filaments Causes Muscles to Contract

Muscle contraction is the most familiar and best-understood form of movement in animals. In vertebrates, running, walking, swimming, and flying all depend on the rapid contraction of skeletal muscle on its scaffolding of bone, while involuntary



**Figure 16–29** The cycle of structural changes used by myosin II to walk along an actin filament. In the myosin II cycle, the head remains bound to the actin filament for only about 5% of the entire cycle time, allowing many myosins to work together to move a single actin filament (Movie 16.3). (Based on I. Rayment et al., *Science* 261:50–58, 1993.)

movements such as heart pumping and gut peristalsis depend on the contraction of cardiac muscle and smooth muscle, respectively. All these forms of muscle contraction depend on the ATP-driven sliding of highly organized arrays of actin filaments against arrays of myosin II filaments.

Skeletal muscle was a relatively late evolutionary development, and muscle cells are highly specialized for rapid and efficient contraction. The long, thin muscle fibers of skeletal muscle are actually huge single cells that form during

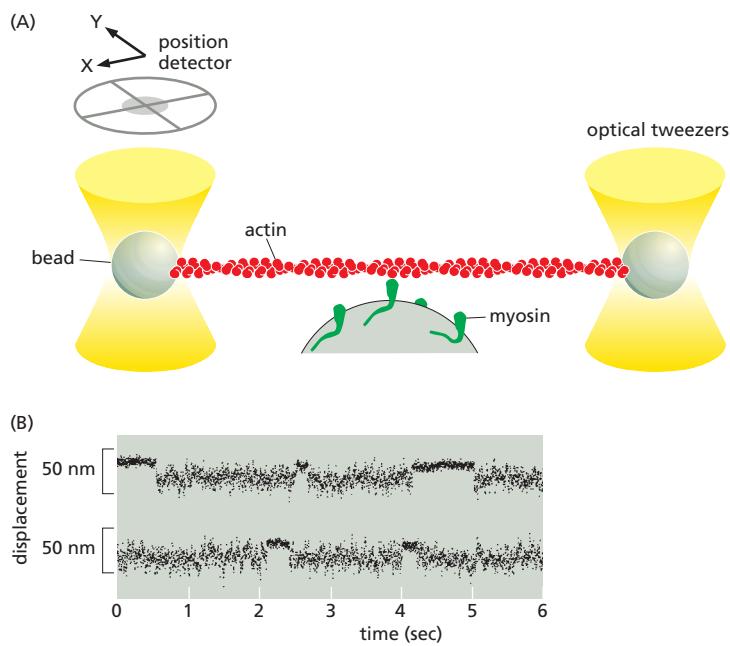
**ATTACHED** At the start of the cycle shown in this figure, a myosin head lacking a bound nucleotide is locked tightly onto an actin filament in a *rigor* configuration (so named because it is responsible for *rigor mortis*, the rigidity of death). In an actively contracting muscle, this state is very short-lived, being rapidly terminated by the binding of a molecule of ATP.

**RELEASED** A molecule of ATP binds to the large cleft on the “back” of the head (that is, on the side furthest from the actin filament) and immediately causes a slight change in the conformation of the actin-binding site, reducing the affinity of the head for actin and allowing it to move along the filament. (The space drawn here between the head and actin emphasizes this change, although in reality the head probably remains very close to the actin.)

**COCKED** The cleft closes like a clam shell around the ATP molecule, triggering a movement in the lever arm that causes the head to be displaced along the filament by a distance of about 5 nm. Hydrolysis of ATP occurs, but the ADP and inorganic phosphate ( $P_i$ ) remain tightly bound to the protein.

**FORCE-GENERATING** Weak binding of the myosin head to a new site on the actin filament causes release of the inorganic phosphate produced by ATP hydrolysis, concomitantly with the tight binding of the head to actin. This release triggers the power stroke—the force-generating change in shape during which the head regains its original conformation. In the course of the power stroke, the head loses its bound ADP, thereby returning to the start of a new cycle.

**ATTACHED** At the end of the cycle, the myosin head is again locked tightly to the actin filament in a rigor configuration. Note that the head has moved to a new position on the actin filament.



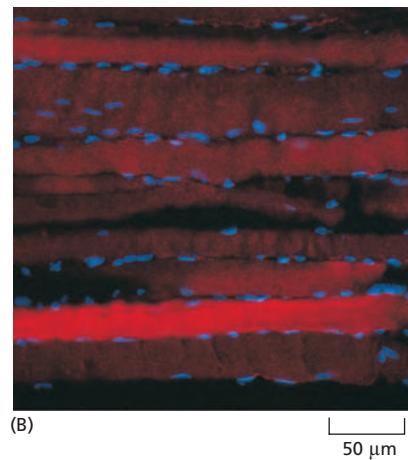
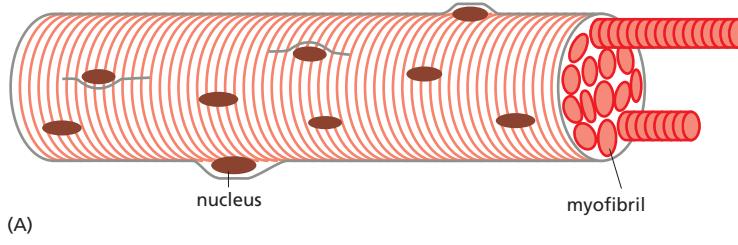
**Figure 16-30** The force of a single myosin molecule moving along an actin filament measured using an optical trap. (A) Schematic of the experiment, showing an actin filament with beads attached at both ends and held in place by focused beams of light called optical tweezers ([Movie 16.4](#)). The tweezers trap and move the bead, and can also be used to measure the force exerted on the bead through the filament. In this experiment, the filament was positioned over another bead to which myosin II motors were attached, and the optical tweezers were used to determine the effects of myosin binding on movement of the actin filament.

(B) These traces show filament movement in two separate experiments. Initially, when the actin filament is unattached to myosin, thermal motion of the filament produces noisy fluctuations in filament position. When a single myosin binds to the actin filament, thermal motion decreases abruptly and a roughly 10-nm displacement results from movement of the filament by the motor. The motor then releases the filament. Because the ATP concentration is very low in this experiment, the myosin remains attached to the actin filament for much longer than it would in a muscle cell. (Adapted from C. Rüegg et al., *Physiology* 17:213–218, 2002. With permission from the American Physiological Society.)

development by the fusion of many separate cells. The large muscle cell retains the many nuclei of the contributing cells. These nuclei lie just beneath the plasma membrane ([Figure 16-31](#)). The bulk of the cytoplasm inside is made up of myofibrils, which is the name given to the basic contractile elements of the muscle cell. A **myofibril** is a cylindrical structure 1–2  $\mu\text{m}$  in diameter that is often as long as the muscle cell itself. It consists of a long, repeated chain of tiny contractile units—called *sarcomeres*, each about 2.2  $\mu\text{m}$  long—which give the vertebrate myofibril its striated appearance ([Figure 16-32](#)).

Each sarcomere is formed from a miniature, precisely ordered array of parallel and partly overlapping thin and thick filaments. The *thin filaments* are composed of actin and associated proteins, and they are attached at their plus ends to a *Z disc* at each end of the sarcomere. The capped minus ends of the actin filaments extend in toward the middle of the sarcomere, where they overlap with *thick filaments*, the bipolar assemblies formed from specific muscle isoforms of myosin II (see Figure 16-27). When this region of overlap is examined in cross section by electron microscopy, the myosin filaments are arranged in a regular hexagonal lattice, with the actin filaments evenly spaced between them ([Figure 16-33](#)). Cardiac muscle and smooth muscle also contain sarcomeres, although the organization is not as regular as that in skeletal muscle.

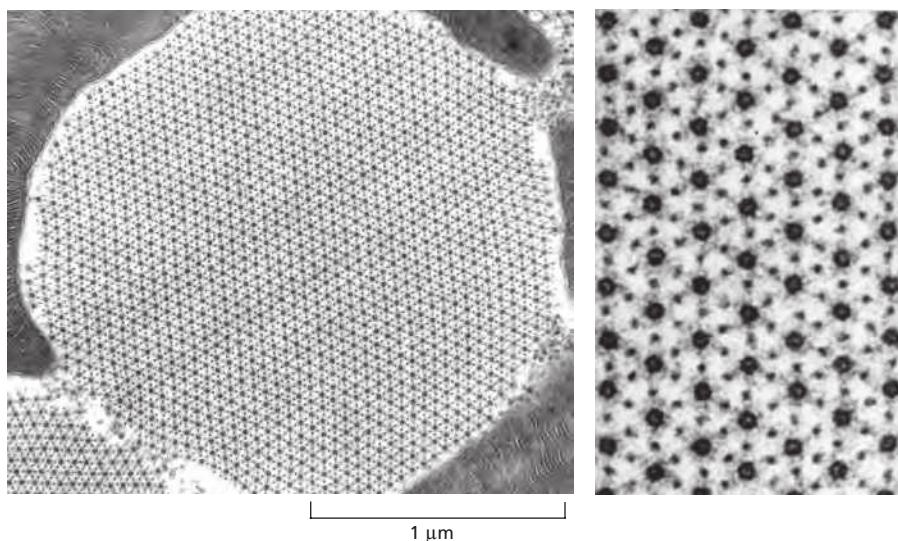
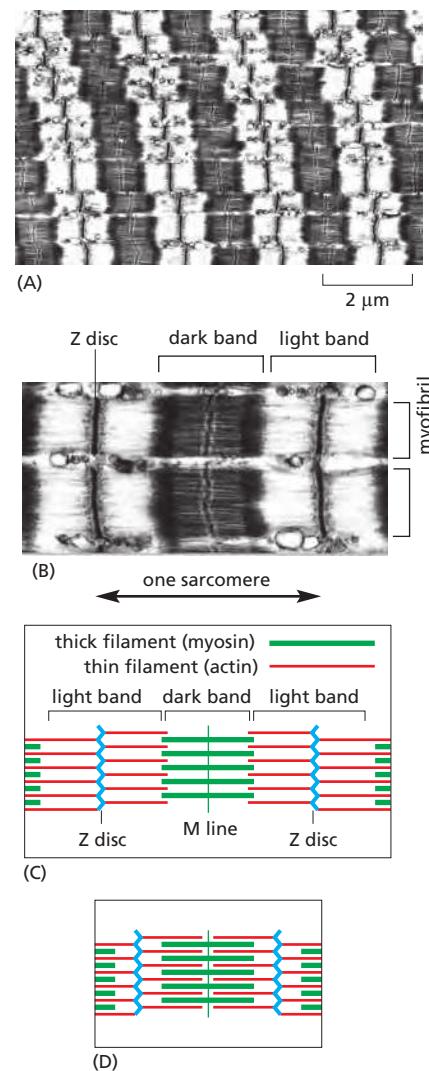
**Figure 16-31** Skeletal muscle cells (also called muscle fibers). (A) These huge multinucleated cells form by the fusion of many muscle cell precursors, called myoblasts. Here, a single muscle cell is depicted. In an adult human, a muscle cell is typically 50  $\mu\text{m}$  in diameter and can be up to several centimeters long. (B) Fluorescence micrograph of rat muscle, showing the peripherally located nuclei (blue) in these giant cells. Myofibrils are stained red. (B, courtesy of Nancy L. Kedersha.)



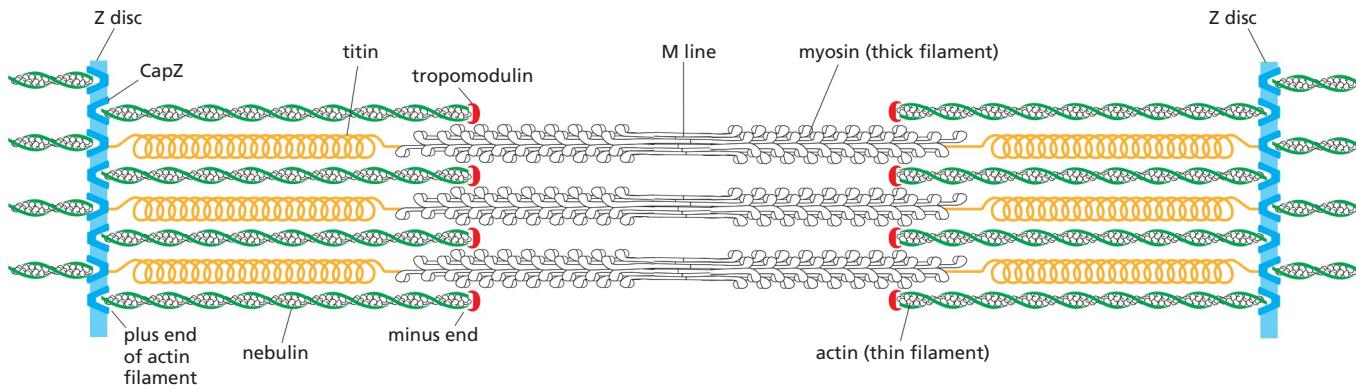
**Figure 16–32 Skeletal muscle myofibrils.** (A) Low-magnification electron micrograph of a longitudinal section through a skeletal muscle cell of a rabbit, showing the regular pattern of cross-striations. The cell contains many myofibrils aligned in parallel (see Figure 16–31). (B) Detail of the skeletal muscle shown in (A), showing portions of two adjacent myofibrils and the definition of a sarcomere (black arrow). (C) Schematic diagram of a single sarcomere, showing the origin of the dark and light bands seen in the electron micrographs. The Z discs, at each end of the sarcomere, are attachment sites for the plus ends of actin filaments (thin filaments); the M line, or midline, is the location of proteins that link adjacent myosin II filaments (thick filaments) to one another. (D) When the sarcomere contracts, the actin and myosin filaments slide past one another without shortening. (A and B, courtesy of Roger Craig.)

Sarcomere shortening is caused by the myosin filaments sliding past the actin thin filaments, with no change in the length of either type of filament (see Figure 16–32C and D). Bipolar thick filaments walk toward the plus ends of two sets of thin filaments of opposite orientations, driven by dozens of independent myosin heads that are positioned to interact with each thin filament. Because there is no coordination among the movements of the myosin heads, it is critical that they remain tightly bound to the actin filament for only a small fraction of each ATPase cycle so that they do not hold one another back. Each myosin thick filament has about 300 heads (294 in frog muscle), and each head cycles about five times per second in the course of a rapid contraction—sliding the myosin and actin filaments past one another at rates of up to 15  $\mu\text{m/sec}$  and enabling the sarcomere to shorten by 10% of its length in less than one-fiftieth of a second. The rapid synchronized shortening of the thousands of sarcomeres lying end-to-end in each myofibril enables skeletal muscle to contract rapidly enough for running and flying, or for playing the piano.

Accessory proteins produce the remarkable uniformity in filament organization, length, and spacing in the sarcomere (Figure 16–34). The actin filament plus ends are anchored in the Z disc, which is built from CapZ and  $\alpha$ -actinin; the Z disc caps the filaments (preventing depolymerization), while holding them together in a regularly spaced bundle. The precise length of each thin filament is influenced by a protein of enormous size, called *nebulin*, which consists almost entirely of a repeating 35-amino-acid actin-binding motif. Nebulin stretches from the Z disc toward the minus end of each thin filament, which is capped and stabilized by tropomodulin. Although there is some slow exchange of actin subunits at both ends of the muscle thin filament, such that the components of the thin filament turn over with a half-life of several days, the actin filaments in sarcomeres are remarkably stable compared with those found in most other cell types, whose dynamic actin filaments turn over with half-lives of a few minutes or less.



**Figure 16–33 Electron micrographs of an insect flight muscle viewed in cross section.** The myosin and actin filaments are packed together with almost crystalline regularity. Unlike their vertebrate counterparts, these myosin filaments have a hollow center, as seen in the enlargement on the right. The geometry of the hexagonal lattice is slightly different in vertebrate muscle. (From J. Auber, *J. de Microsc.* 8:197–232, 1969. With permission from Société Française de Microscopie Électronique.)



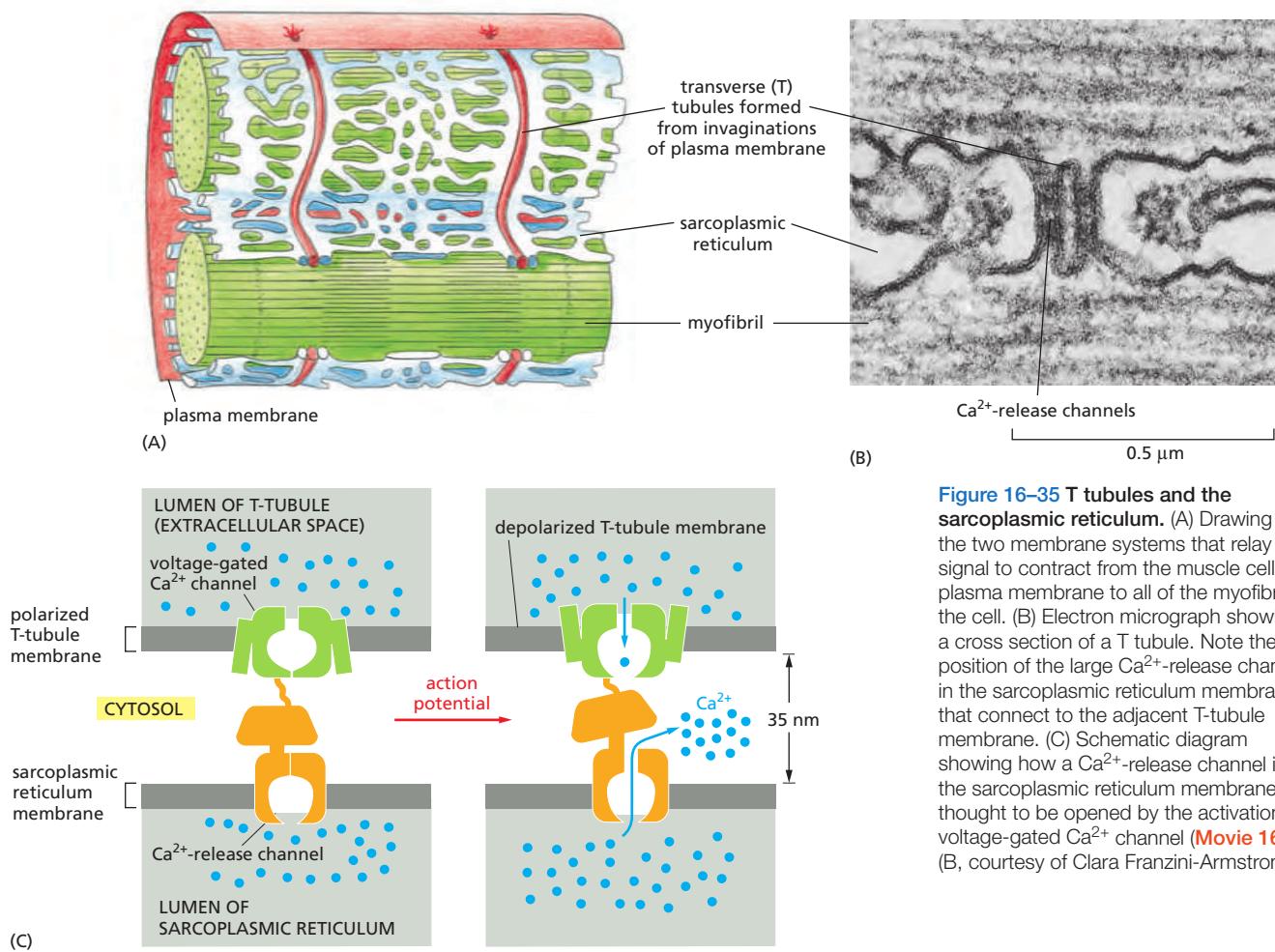
Opposing pairs of an even longer template protein, called *titin*, position the thick filaments midway between the Z discs. Titin acts as a molecular spring, with a long series of immunoglobulin-like domains that can unfold one by one as stress is applied to the protein. A springlike unfolding and refolding of these domains keeps the thick filaments poised in the middle of the sarcomere and allows the muscle fiber to recover after being overstretched. In *C. elegans*, whose sarcomeres are longer than those in vertebrates, titin is longer as well, suggesting that it serves also as a molecular ruler, determining in this case the overall length of each sarcomere.

### A Sudden Rise in Cytosolic $\text{Ca}^{2+}$ Concentration Initiates Muscle Contraction

The force-generating molecular interaction between myosin thick filaments and actin thin filaments takes place only when a signal passes to the skeletal muscle from the nerve that stimulates it. Immediately upon arrival of the signal, the muscle cell needs to be able to contract very rapidly, with all the sarcomeres shortening simultaneously. Two major features of the muscle cell make extremely rapid contraction possible. First, as previously discussed, the individual myosin motor heads in each thick filament spend only a small fraction of the ATP cycle time bound to the filament and actively generating force, so many myosin heads can act in rapid succession on the same thin filament without interfering with one another. Second, a specialized membrane system relays the incoming signal rapidly throughout the entire cell. The signal from the nerve triggers an action potential in the muscle cell plasma membrane (discussed in Chapter 11), and this electrical excitation spreads swiftly into a series of membranous folds—the transverse tubules, or *T* tubules—that extend inward from the plasma membrane around each myofibril. The signal is then relayed across a small gap to the *sarcoplasmic reticulum*, an adjacent weblike sheath of modified endoplasmic reticulum that surrounds each myofibril like a net stocking (Figure 16–35A and B).

When the incoming action potential activates a  $\text{Ca}^{2+}$  channel in the T-tubule membrane,  $\text{Ca}^{2+}$  influx triggers the opening of  $\text{Ca}^{2+}$ -release channels in the sarcoplasmic reticulum (Figure 16–35C).  $\text{Ca}^{2+}$  flooding into the cytosol then initiates the contraction of each myofibril. Because the signal from the muscle cell plasma membrane is passed within milliseconds (via the T tubules and sarcoplasmic reticulum) to every sarcomere in the cell, all of the myofibrils in the cell contract at once. The increase in  $\text{Ca}^{2+}$  concentration is transient because the  $\text{Ca}^{2+}$  is rapidly pumped back into the sarcoplasmic reticulum by an abundant, ATP-dependent  $\text{Ca}^{2+}$ -pump (also called a  $\text{Ca}^{2+}$ -ATPase) in its membrane (see Figure 11–13). Typically, the cytoplasmic  $\text{Ca}^{2+}$  concentration is restored to resting levels within 30 msec, allowing the myofibrils to relax. Thus, muscle contraction depends on two processes that consume enormous amounts of ATP: filament sliding, driven by the ATPase of the myosin motor domain, and  $\text{Ca}^{2+}$  pumping, driven by the  $\text{Ca}^{2+}$ -pump.

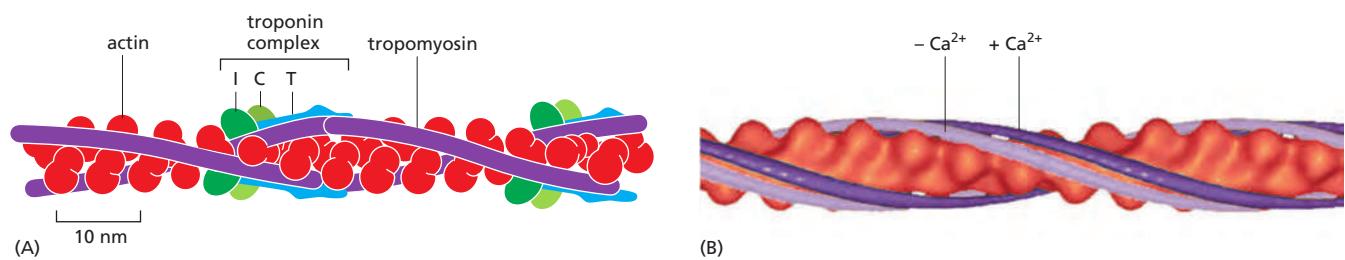
**Figure 16–34 Organization of accessory proteins in a sarcomere.** Each giant titin molecule extends from the Z disc to the M line—a distance of over 1  $\mu\text{m}$ . Part of each titin molecule is closely associated with a myosin thick filament (which switches polarity at the M line); the rest of the titin molecule is elastic and changes length as the sarcomere contracts and relaxes. Each nebulin molecule is exactly the length of a thin filament. The actin filaments are also coated with tropomyosin and troponin (not shown; see Figure 16–36) and are capped at both ends. Tropomodulin caps the minus end of the actin filaments, and CapZ anchors the plus end at the Z disc, which also contains  $\alpha$ -actinin (not shown).



**Figure 16-35 T tubules and the sarcoplasmic reticulum.** (A) Drawing of the two membrane systems that relay the signal to contract from the muscle cell plasma membrane to all of the myofibrils in the cell. (B) Electron micrograph showing a cross section of a T tubule. Note the position of the large  $\text{Ca}^{2+}$ -release channels in the sarcoplasmic reticulum membrane that connect to the adjacent T-tubule membrane. (C) Schematic diagram showing how a  $\text{Ca}^{2+}$ -release channel in the sarcoplasmic reticulum membrane is thought to be opened by the activation of a voltage-gated  $\text{Ca}^{2+}$  channel ([Movie 16.5](#)). (B, courtesy of Clara Franzini-Armstrong.)

The  $\text{Ca}^{2+}$ -dependence of vertebrate skeletal muscle contraction, and hence its dependence on commands transmitted via nerves, is due entirely to a set of specialized accessory proteins that are closely associated with the actin thin filaments. One of these accessory proteins is a muscle form of *tropomyosin*, the elongated protein that binds along the groove of the actin filament helix. The other is *troponin*, a complex of three polypeptides, troponins T, I, and C (named for their tropomyosin-binding, inhibitory, and  $\text{Ca}^{2+}$ -binding activities, respectively). Troponin I binds to actin as well as to troponin T. In a resting muscle, the troponin I-T complex pulls the tropomyosin out of its normal binding groove into a position along the actin filament that interferes with the binding of myosin heads, thereby preventing any force-generating interaction. When the level of  $\text{Ca}^{2+}$  is raised, troponin C—which binds up to four molecules of  $\text{Ca}^{2+}$ —causes troponin I to release its hold on actin. This allows the tropomyosin molecules to slip back into their normal position so that the myosin heads can walk along the actin filaments ([Figure 16-36](#)). Troponin C is closely related to the ubiquitous  $\text{Ca}^{2+}$ -binding protein calmodulin (see [Figure 15-33](#)); it can be thought of as a specialized form of calmodulin that has acquired binding sites for troponin I and troponin T, thereby ensuring that the myofibril responds extremely rapidly to an increase in  $\text{Ca}^{2+}$  concentration.

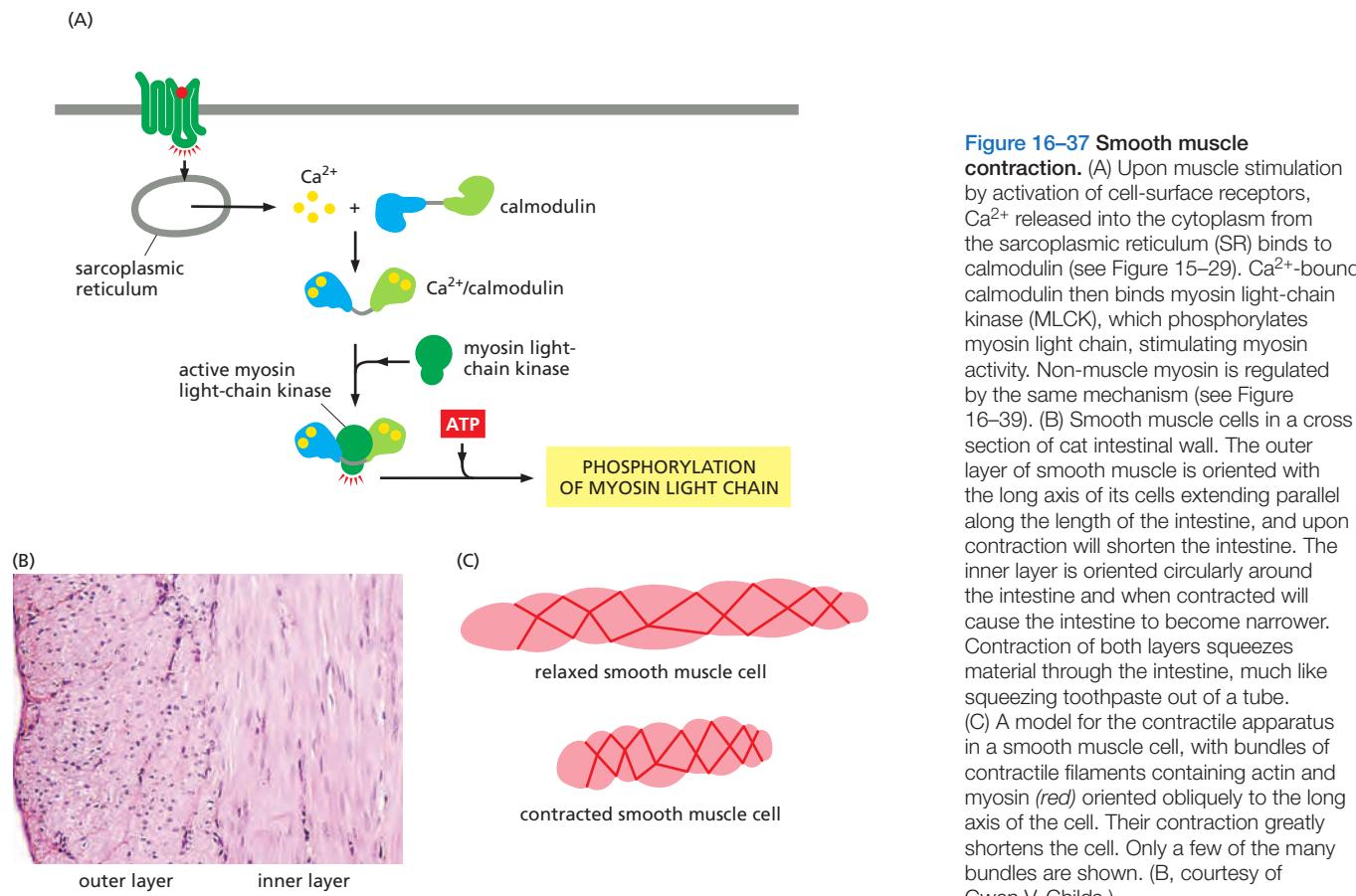
In smooth muscle cells, so called because they lack the regular striations of skeletal muscle, contraction is also triggered by an influx of calcium ions, but the regulatory mechanism is different. Smooth muscle forms the contractile portion of the stomach, intestine, and uterus, as well as the walls of arteries and many other structures requiring slow and sustained contractions. Smooth muscle is



**Figure 16–36 The control of skeletal muscle contraction by troponin.** (A) A skeletal-muscle-cell thin filament, showing the positions of tropomyosin and troponin along the actin filament. Each tropomyosin molecule has seven evenly spaced regions with similar amino acid sequences, each of which is thought to bind to an actin subunit in the filament. (B) Reconstructed cryo-electron microscopy image of an actin filament showing the relative position of a superimposed tropomyosin strand in the presence ( $-Ca^{2+}$ , light purple) or absence ( $+Ca^{2+}$ , dark purple) of calcium. (A, adapted from G.N. Phillips, J.P. Fillers and C. Cohen, *J. Mol. Biol.* 192:111–131, 1986. With permission from Academic Press; B, adapted from C. Xu et al., *Biophys. J.* 77: 985–992, 1999. With permission from Elsevier.)

composed of sheets of highly elongated spindle-shaped cells, each with a single nucleus. Smooth muscle cells do not express the troponins. Instead, elevated intracellular  $Ca^{2+}$  levels regulate contraction by a mechanism that depends on calmodulin (Figure 16–37).  $Ca^{2+}$ -bound calmodulin activates myosin light-chain kinase (MLCK), thereby inducing the phosphorylation of smooth muscle myosin on one of its two light chains. When the light chain is phosphorylated, the myosin head can interact with actin filaments and cause contraction; when it is dephosphorylated, the myosin head tends to dissociate from actin and becomes inactive.

The phosphorylation events that regulate contraction in smooth muscle cells occur relatively slowly, so that maximum contraction often requires nearly a second (compared with the few milliseconds required for contraction of a skeletal muscle cell). But rapid activation of contraction is not important in smooth



**Figure 16–37 Smooth muscle contraction.** (A) Upon muscle stimulation by activation of cell-surface receptors,  $Ca^{2+}$  released into the cytoplasm from the sarcoplasmic reticulum (SR) binds to calmodulin (see Figure 15–29).  $Ca^{2+}$ -bound calmodulin then binds myosin light-chain kinase (MLCK), which phosphorylates myosin light chain, stimulating myosin activity. Non-muscle myosin is regulated by the same mechanism (see Figure 16–39). (B) Smooth muscle cells in a cross section of cat intestinal wall. The outer layer of smooth muscle is oriented with the long axis of its cells extending parallel along the length of the intestine, and upon contraction will shorten the intestine. The inner layer is oriented circularly around the intestine and when contracted will cause the intestine to become narrower. Contraction of both layers squeezes material through the intestine, much like squeezing toothpaste out of a tube. (C) A model for the contractile apparatus in a smooth muscle cell, with bundles of contractile filaments containing actin and myosin (red) oriented obliquely to the long axis of the cell. Their contraction greatly shortens the cell. Only a few of the many bundles are shown. (B, courtesy of Gwen V. Childs.)

muscle: its myosin II hydrolyzes ATP about 10 times more slowly than skeletal muscle myosin, producing a slow cycle of myosin conformational changes that results in slow contraction.

### Heart Muscle Is a Precisely Engineered Machine

The heart is the most heavily worked muscle in the body, contracting about 3 billion ( $3 \times 10^9$ ) times during the course of a human lifetime (Movie 16.6). Heart cells express several specific isoforms of cardiac muscle myosin and cardiac muscle actin. Even subtle changes in these cardiac-specific contractile proteins—changes that would not cause any noticeable consequences in other tissues—can cause serious heart disease (Figure 16–38).

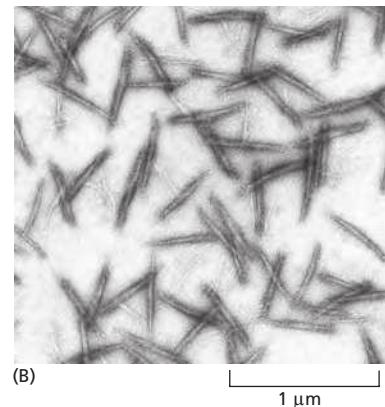
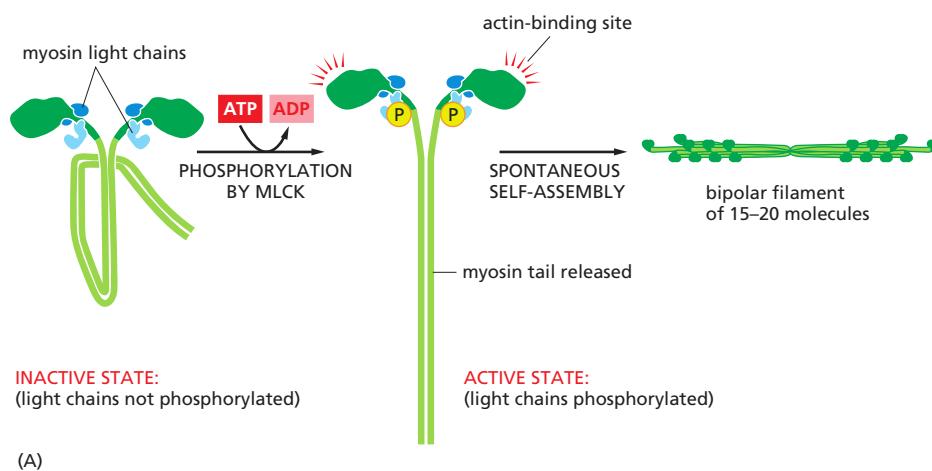
The normal cardiac contractile apparatus is such a highly tuned machine that a tiny abnormality anywhere in the works can be enough to gradually wear it down over years of repetitive motion. *Familial hypertrophic cardiomyopathy* is a common cause of sudden death in young athletes. It is a genetically dominant inherited condition that affects about two out of every thousand people, and it is associated with heart enlargement, abnormally small coronary vessels, and disturbances in heart rhythm (cardiac arrhythmias). The cause of this condition is either any one of over 40 subtle point mutations in the genes encoding cardiac  $\beta$  myosin heavy chain (almost all causing changes in or near the motor domain) or one of about a dozen mutations in other genes encoding contractile proteins—including myosin light chains, cardiac troponin, and tropomyosin. Minor missense mutations in the cardiac actin gene cause another type of heart condition, called *dilated cardiomyopathy*, which can also result in early heart failure.

### Actin and Myosin Perform a Variety of Functions in Non-Muscle Cells

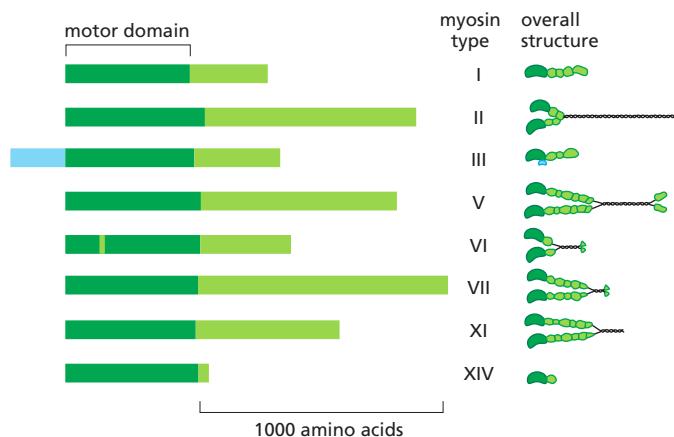
Most non-muscle cells contain small amounts of contractile actin–myosin II bundles that form transiently under specific conditions and are much less well organized than muscle fibers. Non-muscle contractile bundles are regulated by myosin phosphorylation rather than the troponin mechanism (Figure 16–39). These contractile bundles function to provide mechanical support to cells, for example, by assembling into cortical **stress fibers** that connect the cell to the extracellular



**Figure 16–38** Effect on the heart of a subtle mutation in cardiac myosin. *Left*, normal heart from a 6-day-old mouse pup. *Right*, heart from a pup with a point mutation in both copies of its cardiac myosin gene, changing Arg403 to Gln. The arrows indicate the atria. In the heart from the pup with the cardiac myosin mutation, both atria are greatly enlarged (hypertrophic), and the mice die within a few weeks of birth. (From D. Fink et al., *J. Clin. Invest.* 103:147–153, 1999. With permission from The American Society for Clinical Investigation.)



**Figure 16–39** Light-chain phosphorylation and the regulation of the assembly of myosin II into thick filaments. (A) The controlled phosphorylation by the enzyme myosin light-chain kinase (MLCK) of one of the two light chains (the so-called regulatory light chain, shown in light blue) on non-muscle myosin II in a test tube has at least two effects: it causes a change in the conformation of the myosin head, exposing its actin-binding site, and it releases the myosin tail from a “sticky patch” on the myosin head, thereby allowing the myosin molecules to assemble into short, bipolar, thick filaments. Smooth muscle is regulated by the same mechanism (see Figure 16–37). (B) Electron micrograph of negatively stained short filaments of myosin II that have been induced to assemble in a test tube by phosphorylation of their light chains. These myosin II filaments are much smaller than those found in skeletal muscle cells (see Figure 16–27). (B, courtesy of John Kendrick-Jones.)



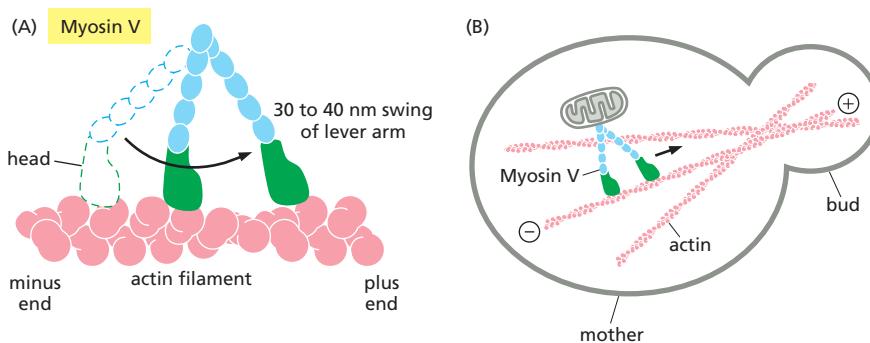
**Figure 16–40 Myosin superfamily members.** Comparison of the domain structure of the heavy chains of some myosins. All myosins share similar motor domains (shown in dark green), but their C-terminal tails (light green) and N-terminal extensions (light blue) are very diverse. On the right are depictions of the molecular structure for these family members. Many myosins form dimers, with two motor domains per molecule, but a few (such as I, III, and XIV) seem to function as monomers, with just one motor domain. Myosin VI, despite its overall structural similarity to other family members, is unique in moving toward the minus end (instead of the plus end) of an actin filament. The small insertion within its motor head domain, not found in other myosins, is probably responsible for this change in direction.

matrix through *focal adhesions* or by forming a *circumferential belt* in an epithelial cell, connecting it to adjacent cells through *adherens junctions* (discussed in Chapter 19). As described in Chapter 17, actin and myosin II in the *contractile ring* generate the force for cytokinesis, the final stage in cell division. Finally, as discussed later, contractile bundles also contribute to the adhesion and forward motion of migrating cells.

Non-muscle cells also express a large family of other myosin proteins, which have diverse structures and functions in the cell. Following the discovery of conventional muscle myosin, a second member of the family was found in the freshwater amoeba *Acanthamoeba castellanii*. This protein had a different tail structure and seemed to function as a monomer, and so it was named *myosin I* (for one-headed). Conventional muscle myosin was renamed *myosin II* (for two-headed). Subsequently, many other myosin types were discovered. The heavy chains generally start with a recognizable myosin motor domain at the N-terminus and then diverge widely with a variety of C-terminal tail domains (Figure 16–40). The myosin family includes a number of one-headed and two-headed varieties that are about equally related to myosin I and myosin II, and the nomenclature now reflects their approximate order of discovery (myosin III through at least myosin XVIII). Sequence comparisons among diverse eukaryotes indicate that there are at least 37 distinct myosin families in the superfamily. All of the myosins except one move toward the plus end of an actin filament, although they do so at different speeds. The exception is myosin VI, which moves toward the minus end. The myosin tails (and the tails of motor proteins generally) have apparently diversified during evolution to permit the proteins to bind other subunits and to interact with different cargoes.

Some myosins are found only in plants, and some are found only in vertebrates. Most, however, are found in all eukaryotes, suggesting that myosins arose early in eukaryotic evolution. The human genome includes about 40 myosin genes. Nine of the human myosins are expressed primarily or exclusively in the hair cells of the inner ear, and mutations in five of them are known to cause hereditary deafness. These extremely specialized myosins are important for the construction and function of the complex and beautiful bundles of actin found in stereocilia that project from the apical surface of these cells (see Figure 9–51); these cellular protrusions tilt in response to sound and convert sound waves into electrical signals.

The functions of most of the myosins remain to be determined, but several are well characterized. The myosin I proteins often contain either a second actin-binding site or a membrane-binding site in their tails, and they are generally involved in intracellular organization—including the protrusion of actin-rich structures at the cell surface, such as microvilli (see Panel 16–1 and Figure 16–4), and endocytosis. Myosin V is a two-headed myosin with a large step size (Figure 16–41A) and is involved in organelle transport along actin filaments. In contrast to myosin II motors, which work in ensembles and are attached only transiently to actin filaments so as not to interfere with one another, myosin V moves continuously,



**Figure 16–41 Myosin V carries cargo along actin filaments.** (A) The lever arm of myosin V is long, allowing it to take a bigger step along an actin filament than myosin II (see Figure 16–29). (B) Myosin V transports cargo and organelles along actin cables, in this example moving a mitochondrion into the growing bud of a yeast cell.

or *processively*, along actin filaments without letting go. Myosin V functions are well studied in the yeast *Saccharomyces cerevisiae*, which undergoes a stereotypical pattern of growth and division called budding. Actin cables in the mother cell point toward the bud, where actin is found in patches that concentrate where cell wall growth is taking place. Myosin V motors carry a wide range of cargoes—including mRNA, endoplasmic reticulum, and secretory vesicles—along the actin cables and into the bud. In addition, myosin V mediates the correct partitioning of organelles such as peroxisomes and mitochondria between mother and daughter cells (see Figure 16–41B).

### Summary

Using their neck domain as a lever arm, myosins convert ATP hydrolysis into mechanical work to move along actin filaments in a stepwise fashion. Skeletal muscle is made up of myofibrils containing thousands of sarcomeres assembled from highly ordered arrays of actin and myosin II filaments, together with many accessory proteins. Muscle contraction is stimulated by calcium, which causes the actin-filament-associated protein tropomyosin to move, uncovering myosin binding sites and allowing the filaments to slide past one another. Smooth muscle and non-muscle cells have less well-ordered contractile bundles of actin and myosin, which are regulated by myosin light-chain phosphorylation. Myosin V transports cargo by walking along actin filaments.

## MICROTUBULES

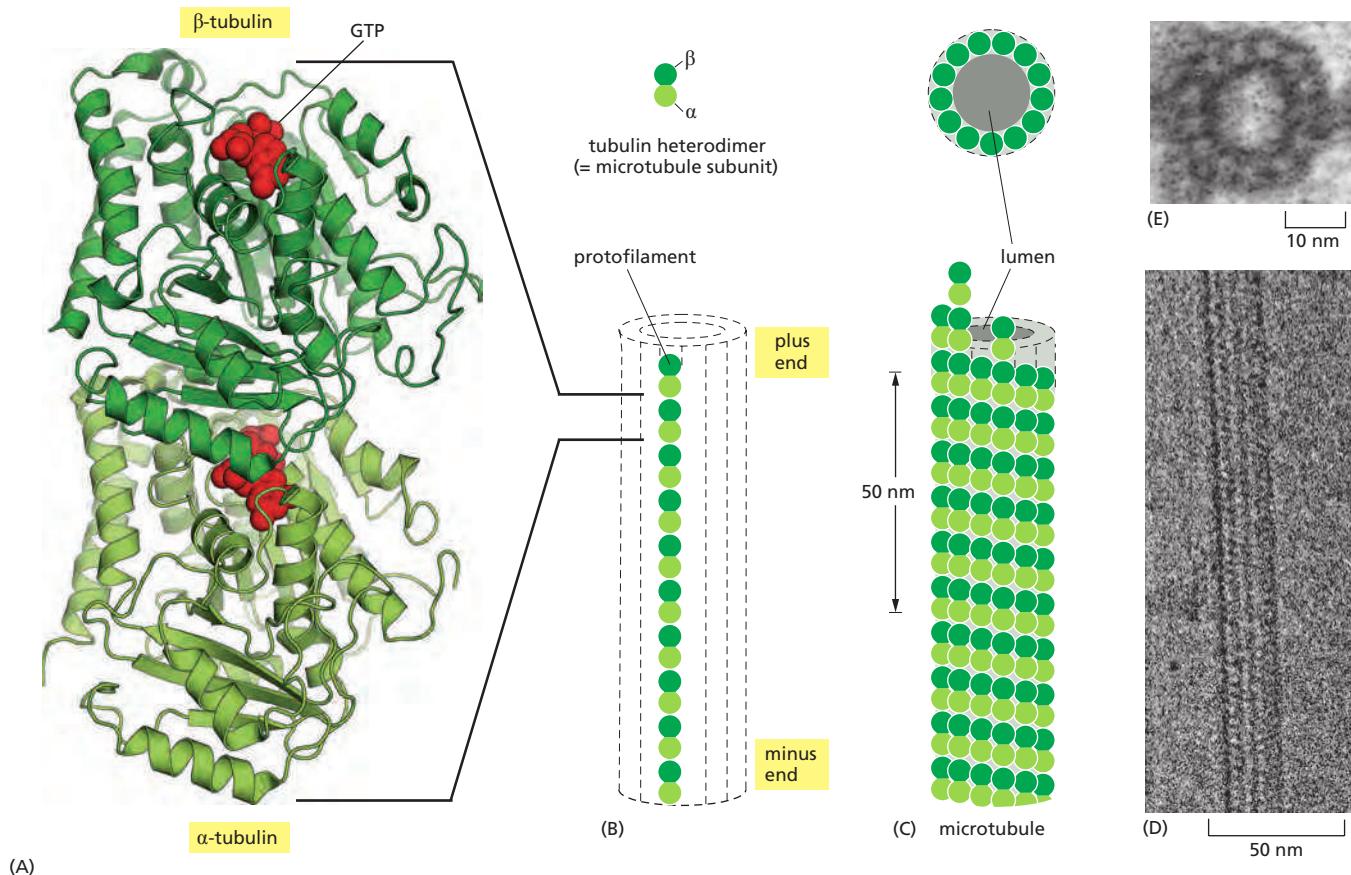
Microtubules are structurally more complex than actin filaments, but they are also highly dynamic and play comparably diverse and important roles in the cell. Microtubules are polymers of the protein **tubulin**. The tubulin subunit is itself a heterodimer formed from two closely related globular proteins called  $\alpha$ -tubulin and  $\beta$ -tubulin, each comprising 445–450 amino acids, which are tightly bound together by noncovalent bonds (Figure 16–42A). These two tubulin proteins are found only in this heterodimer, and each  $\alpha$  or  $\beta$  monomer has a binding site for one molecule of GTP. The GTP that is bound to  $\alpha$ -tubulin is physically trapped at the dimer interface and is never hydrolyzed or exchanged; it can therefore be considered to be an integral part of the tubulin heterodimer structure. The nucleotide on the  $\beta$ -tubulin, in contrast, may be in either the GTP or the GDP form and is exchangeable within the soluble (unpolymerized) tubulin dimer.

Tubulin is found in all eukaryotic cells, and it exists in multiple isoforms. Yeast and human tubulins are 75% identical in amino acid sequence. In mammals, there are at least six forms of  $\alpha$ -tubulin and a similar number of  $\beta$ -tubulins, each encoded by a different gene. The different forms of tubulin are very similar, and they generally copolymerize into mixed microtubules in the test tube. However, they can have distinct locations in cells and tissues and perform subtly different functions. As a striking example, mutations in a particular human  $\beta$ -tubulin gene give rise to a paralytic eye-movement disorder due to loss of ocular nerve function. Numerous human neurological diseases have been linked to specific mutations in different tubulin genes.

## Microtubules Are Hollow Tubes Made of Protofilaments

A microtubule is a hollow cylindrical structure built from 13 parallel protofilaments, each composed of  $\alpha\beta$ -tubulin heterodimers stacked head to tail and then folded into a tube (Figure 16–42B–D). Microtubule assembly generates two new types of protein–protein contacts. Along the longitudinal axis of the microtubule, the “top” of one  $\beta$ -tubulin molecule forms an interface with the “bottom” of the  $\alpha$ -tubulin molecule in the adjacent heterodimer. This interface is very similar to the interface holding the  $\alpha$  and  $\beta$  monomers together in the dimer subunit, and the binding energy is high. Perpendicular to these interactions, neighboring protofilaments form lateral contacts. In this dimension, the main lateral contacts are between monomers of the same type ( $\alpha$ - $\alpha$  and  $\beta$ - $\beta$ ). As longitudinal and lateral contacts are repeated during assembly, a slight stagger in lateral contacts gives rise to the helical microtubule lattice. Because multiple contacts within the lattice hold most of the subunits in a microtubule in place, the addition and loss of subunits occurs almost exclusively at the microtubule ends (see Figure 16–5). These multiple contacts among subunits make microtubules stiff and difficult to bend. The persistence length of a microtubule is several millimeters, making microtubules the stiffest and straightest structural elements found in most animal cells.

The subunits in each protofilament in a microtubule all point in the same direction, and the protofilaments themselves are aligned in parallel (see Figure



**Figure 16–42 The structure of a microtubule and its subunit.** (A) The subunit of each protofilament is a tubulin heterodimer, formed from a tightly linked pair of  $\alpha$ - and  $\beta$ -tubulin monomers. The GTP molecule in the  $\alpha$ -tubulin monomer is so tightly bound that it can be considered an integral part of the protein. The GTP molecule in the  $\beta$ -tubulin monomer, however, is less tightly bound and has an important role in filament dynamics. Both nucleotides are shown in red. (B) One tubulin subunit ( $\alpha\beta$ -heterodimer) and one protofilament are shown schematically. Each protofilament consists of many adjacent subunits with the same orientation. (C) The microtubule is a stiff hollow tube formed from 13 protofilaments aligned in parallel. (D) A short segment of a microtubule viewed in an electron microscope. (E) Electron micrograph of a cross section of a microtubule showing a ring of 13 distinct protofilaments. (D, courtesy of Richard Wade; E, courtesy of Richard Linck.)

**Figure 16–43** The preferential growth of microtubules at the plus end.

Microtubules grow faster at one end than at the other. A stable bundle of microtubules obtained from the core of a cilium (called an axoneme) was incubated for a short time with tubulin subunits under polymerizing conditions. Microtubules grew fastest from the plus end of the microtubule bundle, the end at the top in this micrograph. (Courtesy of Gary Borisy.)

16–42). Therefore, the microtubule lattice itself has a distinct structural polarity, with  $\alpha$ -tubulins exposed at the minus end and  $\beta$ -tubulins exposed at the plus end. As for actin filaments, the regular, parallel orientation of their subunits gives microtubules structural and dynamic polarity (**Figure 16–43**), with plus ends growing and shrinking more rapidly.

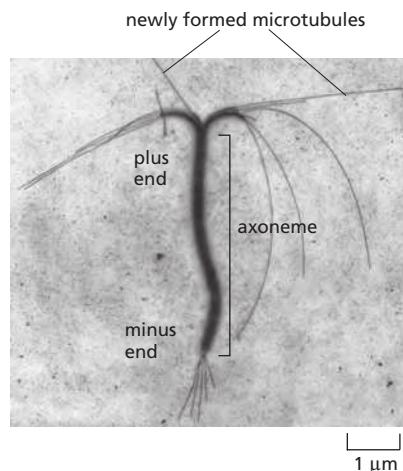
### Microtubules Undergo Dynamic Instability

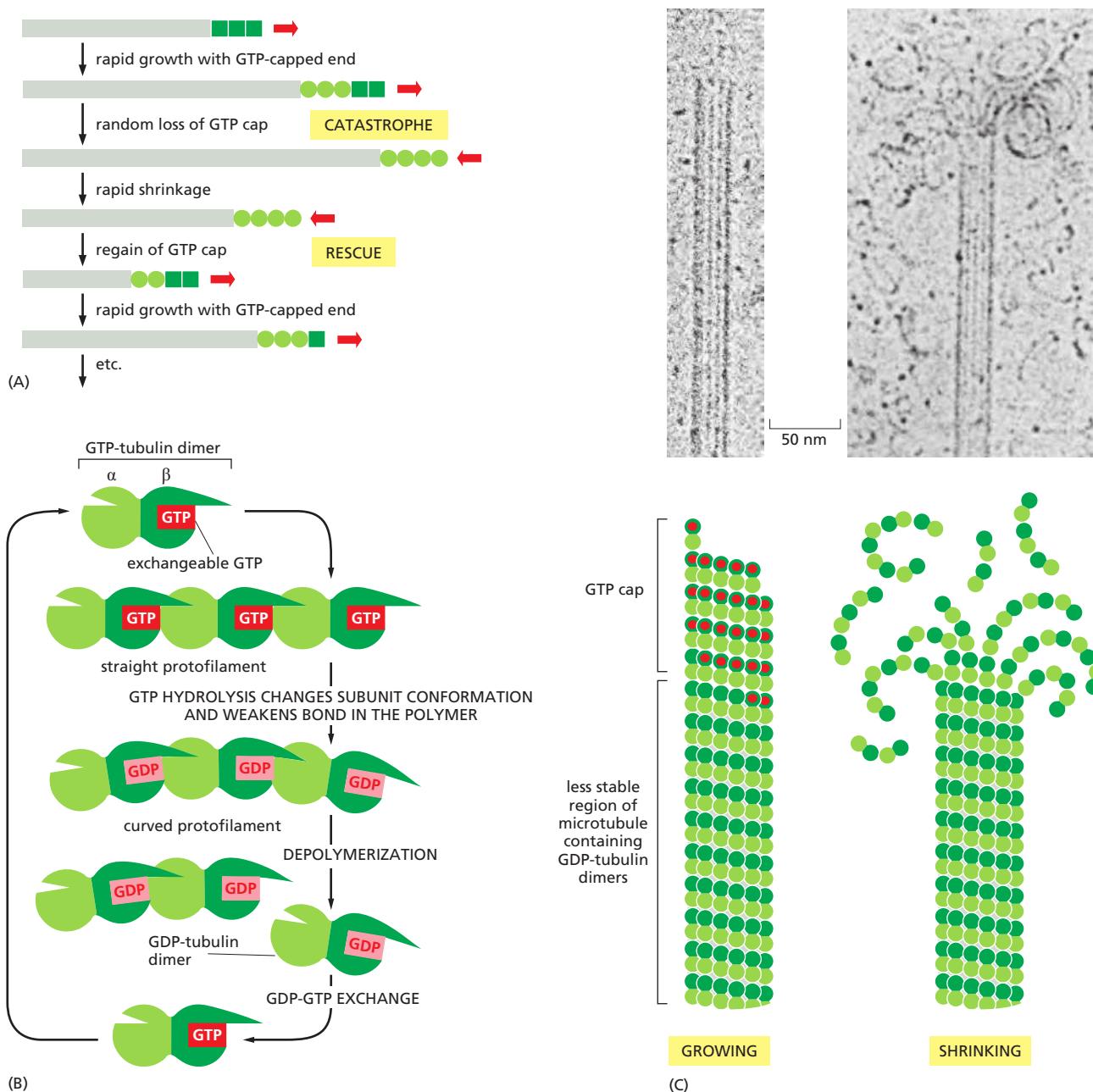
Microtubule dynamics, like those of actin filaments, are profoundly influenced by the binding and hydrolysis of nucleotide—GTP in this case. GTP hydrolysis occurs only within the  $\beta$ -tubulin subunit of the tubulin dimer. It proceeds very slowly in free tubulin subunits but is accelerated when they are incorporated into microtubules. Following GTP hydrolysis, the free phosphate group is released and the GDP remains bound to  $\beta$ -tubulin within the microtubule lattice. Thus, as in the case of actin filaments, two different types of microtubule structures can exist, one with the “T form” of the nucleotide bound (GTP) and one with the “D form” bound (GDP). The energy of nucleotide hydrolysis is stored as elastic strain in the polymer lattice, making the free-energy change for dissociation of a subunit from the D-form polymer more negative than the free-energy change for dissociation of a subunit from the T-form polymer. In consequence, the ratio of  $k_{\text{off}}/k_{\text{on}}$  for GDP-tubulin (its critical concentration  $[C_c(\text{D})]$ ) is much higher than that of GTP-tubulin. Thus, under physiological conditions, GTP-tubulin tends to polymerize and GDP-tubulin to depolymerize.

Whether the tubulin subunits at the very end of a microtubule are in the T or the D form depends on the relative rates of GTP hydrolysis and tubulin addition. If the rate of subunit addition is high—and thus the filament is growing rapidly—then it is likely that a new subunit will be added to the polymer before the nucleotide in the previously added subunit has been hydrolyzed. In this case, the tip of the polymer remains in the T form, forming a *GTP cap*. However, if the rate of subunit addition is low, hydrolysis may occur before the next subunit is added, and the tip of the filament will then be in the D form. If GTP-tubulin subunits assemble at the end of the microtubule at a rate similar to the rate of GTP hydrolysis, then hydrolysis will sometimes “catch up” with the rate of subunit addition and transform the end to a D form. This transformation is sudden and random, with a certain probability per unit time that depends on the concentration of free GTP-tubulin subunits.

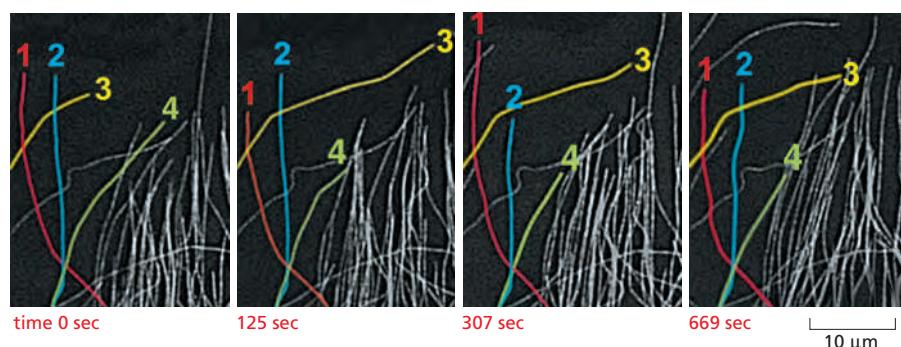
Suppose that the concentration of free tubulin is intermediate between the critical concentration for a T-form end and the critical concentration for a D-form end (that is, above the concentration necessary for T-form assembly, but below that for the D form). Now, any end that happens to be in the T form will grow, whereas any end that happens to be in the D form will shrink. On a single microtubule, an end might grow for a certain length of time in a T form, but then suddenly change to the D form and begin to shrink rapidly, even while the free subunit concentration is held constant. At some later time, it might then regain a T-form end and begin to grow again. This rapid interconversion between a growing and shrinking state, at a uniform free subunit concentration, is called **dynamic instability** (**Figure 16–44A** and **Figure 16–45**; see Panel 16–2). The change from growth to shrinkage is called a *catastrophe*, while the change to growth is called a *rescue*.

In a population of microtubules, at any instant some of the ends are in the T form and some are in the D form, with the ratio depending on the hydrolysis rate and the free subunit concentration. *In vitro*, the structural difference between a T-form end and a D-form end is dramatic. Tubulin subunits with GTP bound to





**Figure 16–44** Dynamic instability due to the structural differences between a growing and a shrinking microtubule end. (A) If the free tubulin concentration in solution is between the critical concentrations of the GTP- and GDP-bound forms, a single microtubule end may undergo transitions between a growing state and a shrinking state. A growing microtubule has GTP-containing subunits at its end, forming a GTP cap. If nucleotide hydrolysis proceeds more rapidly than subunit addition, the cap is lost and the microtubule begins to shrink, an event called a “catastrophe.” But GTP-containing subunits may still add to the shrinking end, and if enough add to form a new cap, then microtubule growth resumes, an event called “rescue.” (B) Model for the structural consequences of GTP hydrolysis in the microtubule lattice. The addition of GTP-containing tubulin subunits to the end of a protofilament causes the end to grow in a linear conformation that can readily pack into the cylindrical wall of the microtubule. Hydrolysis of GTP after assembly changes the conformation of the subunits and tends to force the protofilament into a curved shape that is less able to pack into the microtubule wall. (C) In an intact microtubule, protofilaments made from GDP-containing subunits are forced into a linear conformation by the many lateral bonds within the microtubule wall, given a stable cap of GTP-containing subunits. Loss of the GTP cap, however, allows the GDP-containing protofilaments to relax into their more curved conformation. This leads to a progressive disruption of the microtubule. Above the drawings of a growing and a shrinking microtubule, electron micrographs show actual microtubules in each of these two states. Note particularly the curling, disintegrating GDP-containing protofilaments at the end of the shrinking microtubule. (C, from E.M. Mandelkow, E. Mandelkow and R.A. Milligan, *J. Cell Biol.* 114:977–991, 1991. With permission from The Rockefeller University Press.)



**Figure 16–45** Direct observation of the dynamic instability of microtubules in a living cell. Microtubules in a newt lung epithelial cell were observed after the cell was injected with a small amount of rhodamine-labeled tubulin. Notice the dynamic instability of microtubules at the edge of the cell. Four individual microtubules are highlighted for clarity; each of these shows alternating shrinkage and growth (**Movie 16.7**). (Courtesy of Wendy C. Salmon and Clare Waterman-Storer.)

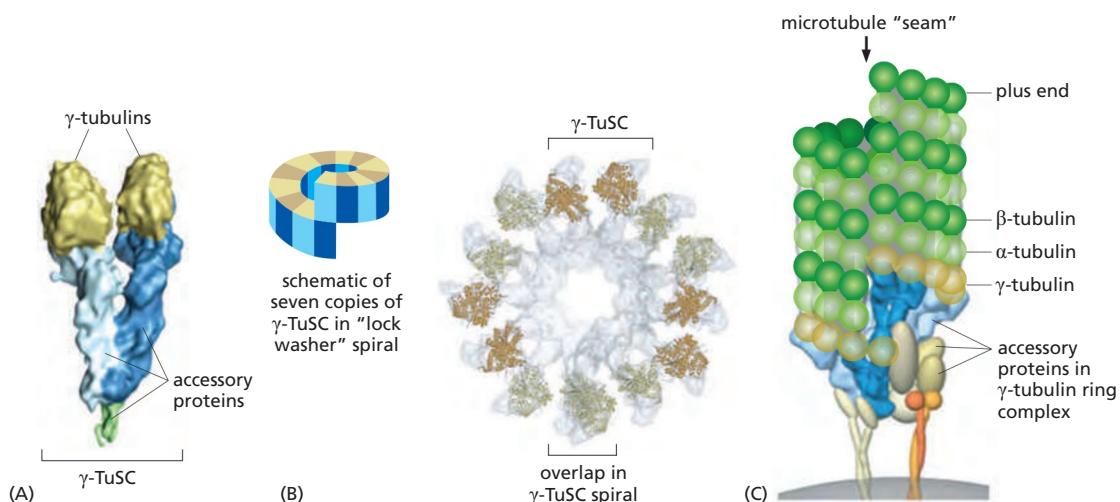
the  $\beta$ -monomer produce straight protofilaments that make strong and regular lateral contacts with one another. But the hydrolysis of GTP to GDP is associated with a subtle conformational change in the protein, which makes the protofilaments curved (Figure 16–44B). On a rapidly growing microtubule, the GTP cap is thought to constrain the curvature of the protofilaments, and the ends appear straight. But when the terminal subunits have hydrolyzed their nucleotides, this constraint is removed, and the curved protofilaments spring apart. This cooperative release of the energy of hydrolysis stored in the microtubule lattice causes the curled protofilaments to peel off rapidly, and curved oligomers of GDP-containing tubulin are seen near the ends of depolymerizing microtubules (Figure 16–44C).

### Microtubule Functions Are Inhibited by Both Polymer-stabilizing and Polymer-destabilizing Drugs

Chemical compounds that impair polymerization or depolymerization of microtubules are powerful tools for investigating the roles of these polymers in cells. Whereas *colchicine* and *nocodazole* interact with tubulin subunits and lead to microtubule depolymerization, *Taxol* binds to and stabilizes microtubules, causing a net increase in tubulin polymerization (see Table 16–1). Drugs like these have a rapid and profound effect on the organization of the microtubules in living cells. Both microtubule-depolymerizing drugs (such as nocodazole) and microtubule-polymerizing drugs (such as Taxol) preferentially kill dividing cells, since microtubule dynamics are crucial for correct function of the mitotic spindle (discussed in Chapter 17). Some of these drugs efficiently kill certain types of tumor cells in a human patient, although not without toxicity to rapidly dividing normal cells, including those in the bone marrow, intestine, and hair follicles. Taxol in particular has been widely used to treat cancers of the breast and lung, and it is frequently successful in treatment of tumors that are resistant to other chemotherapeutic agents.

### A Protein Complex Containing $\gamma$ -Tubulin Nucleates Microtubules

Because formation of a microtubule requires the interaction of many tubulin heterodimers, the concentration of tubulin subunits required for spontaneous nucleation of microtubules is very high. Microtubule nucleation therefore requires help from other factors. While  $\alpha$ - and  $\beta$ -tubulins are the regular building blocks of microtubules, another type of tubulin, called  $\gamma$ -tubulin, is present in much smaller amounts than  $\alpha$ - and  $\beta$ -tubulin and is involved in the nucleation of microtubule growth in organisms ranging from yeasts to humans. Microtubules are generally nucleated from a specific intracellular location known as a **microtubule-organizing center (MTOC)** where  $\gamma$ -tubulin is most enriched. Nucleation in many cases depends on the  **$\gamma$ -tubulin ring complex (TuRC)**. Within this complex, two accessory proteins bind directly to the  $\gamma$ -tubulin, along with several other proteins that help create a spiral ring of  $\gamma$ -tubulin molecules, which serves as a template that creates a microtubule with 13 protofilaments (**Figure 16–46**).



**Figure 16–46** Microtubule nucleation by the  $\gamma$ -tubulin ring complex. (A) Two copies of  $\gamma$ -tubulin associate with a pair of accessory proteins to form the  $\gamma$ -tubulin small complex ( $\gamma$ -TuSC). This image was generated by high-resolution electron microscopy of individual purified complexes. (B) Seven copies of the  $\gamma$ -TuSC associate to form a spiral structure in which the last  $\gamma$ -tubulin lies beneath the first, resulting in 13 exposed  $\gamma$ -tubulin subunits in a circular orientation that matches the orientation of the 13 protofilaments in a microtubule. (C) In many cell types, the  $\gamma$ -TuSC spiral associates with additional accessory proteins to form the  $\gamma$ -tubulin ring complex ( $\gamma$ -TuRC), which is likely to nucleate the minus end of a microtubule as shown here. Note the longitudinal discontinuity between two protofilaments, which results from the spiral orientation of the  $\gamma$ -tubulin subunits. Microtubules often have one such “seam” breaking the otherwise uniform helical packing of the protofilaments. (A and B, from J.M. Kollman et al., *Nature* 466:879–883, 2010. With permission from Macmillan Publishers Ltd.)

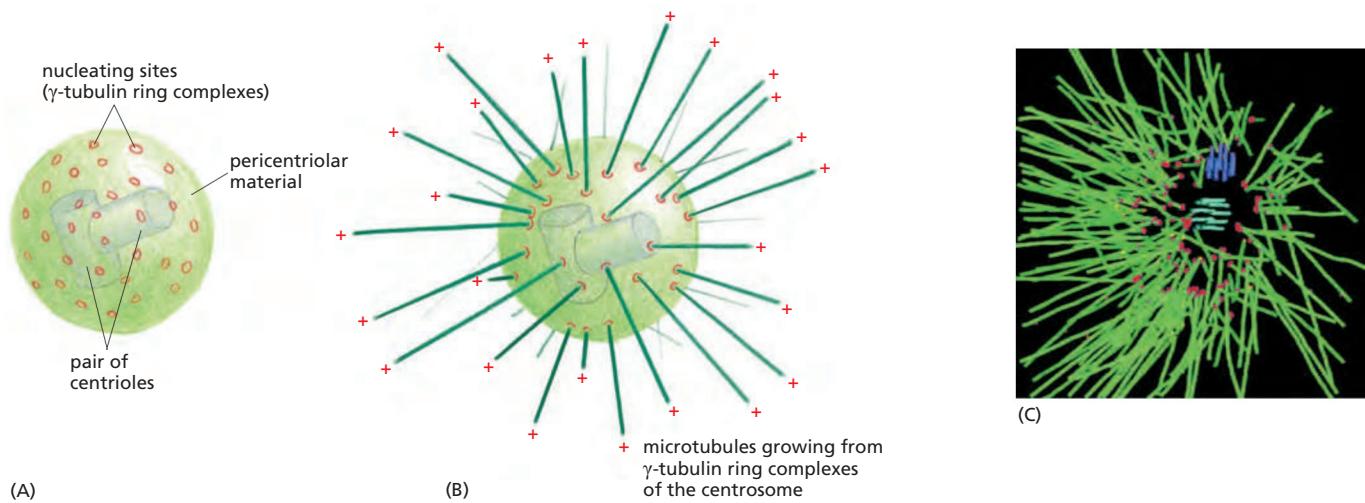
### Microtubules Emanate from the Centrosome in Animal Cells

Many animal cells have a single, well-defined MTOC called the **centrosome**, which is located near the nucleus and from which microtubules are nucleated at their minus ends, so the plus ends point outward and continuously grow and shrink, probing the entire three-dimensional volume of the cell. A centrosome typically recruits more than fifty copies of  $\gamma$ -TuRC. In addition,  $\gamma$ -TuRC molecules are found in the cytoplasm, and centrosomes are not absolutely required for microtubule nucleation, since destroying them with a laser pulse does not prevent microtubule nucleation elsewhere in the cell. A variety of proteins have been identified that anchor  $\gamma$ -TuRC to the centrosome, but mechanisms that activate microtubule nucleation at MTOCs and at other sites in the cell are poorly understood.

Embedded in the centrosome are the **centrioles**, a pair of cylindrical structures arranged at right angles to each other in an L-shaped configuration (Figure 16–47). A centriole consists of a cylindrical array of short, modified microtubules arranged into a barrel shape with striking ninefold symmetry (Figure 16–48). Together with a large number of accessory proteins, the centrioles organize the *pericentriolar material*, where microtubule nucleation takes place. As described in Chapter 17, the centrosome duplicates and splits into two parts before mitosis, each containing a duplicated centriole pair. The two centrosomes move to opposite sides of the nucleus when mitosis begins, and they form the two poles of the mitotic spindle (see Panel 17–1).

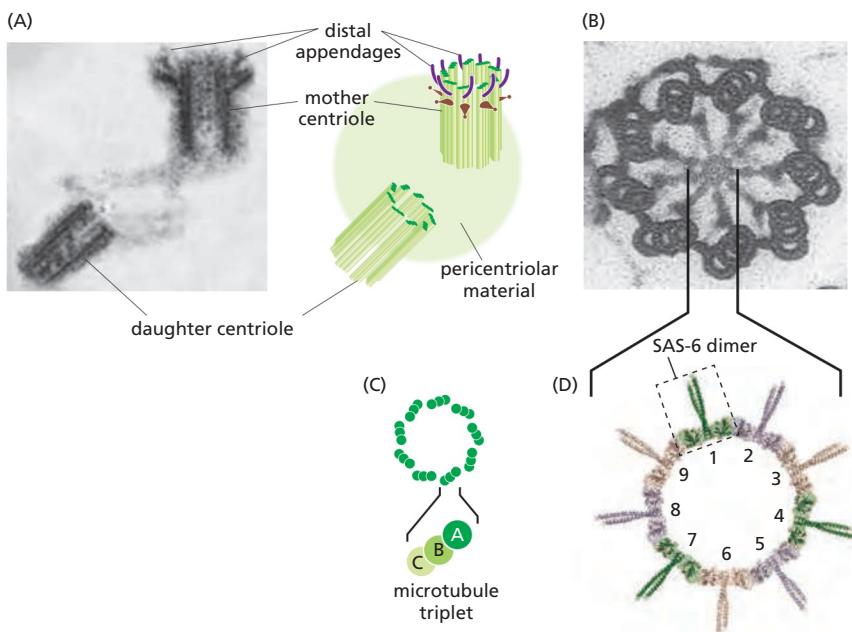
Microtubule organization varies widely among different species and cell types. In budding yeast, microtubules are nucleated at an MTOC that is embedded in the nuclear envelope as a small, multilayered structure called the *spindle pole body*, also found in other fungi and diatoms. Higher-plant cells appear to nucleate microtubules at sites distributed all around the nuclear envelope and at the cell cortex. Neither fungi nor most plant cells contain centrioles. Despite these differences, all these cells seem to use  $\gamma$ -tubulin to nucleate their microtubules.

In cultured animal cells, the aster-like configuration of microtubules is robust, with dynamic plus ends pointing outward toward the cell periphery and stable minus ends collected near the nucleus. The system of microtubules radiating from

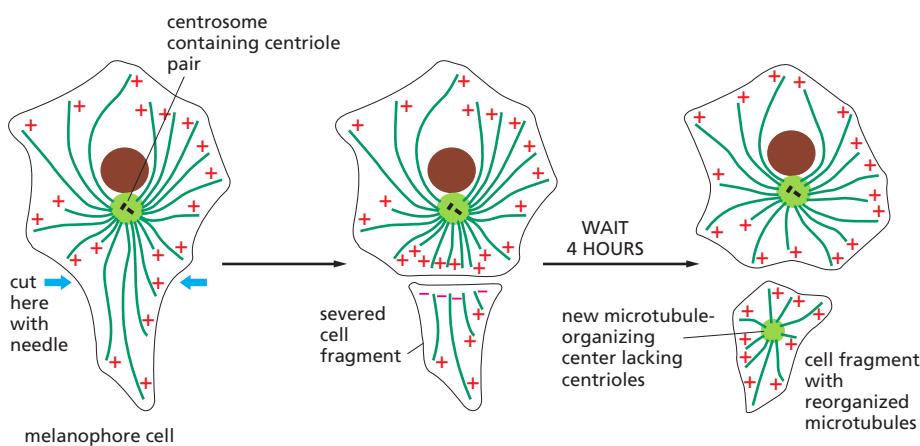


**Figure 16-47 The centrosome.** (A) The centrosome is the major MTOC of animal cells. Located in the cytoplasm next to the nucleus, it consists of an amorphous matrix of fibrous proteins to which the  $\gamma$ -tubulin ring complexes that nucleate microtubule growth are attached. This matrix is organized by a pair of centrioles, as described in the text. (B) A centrosome with attached microtubules. The minus end of each microtubule is embedded in the centrosome, having grown from a  $\gamma$ -tubulin ring complex, whereas the plus end of each microtubule is free in the cytoplasm. (C) In a reconstructed image of the MTOC from a *C. elegans* cell, a dense thicket of microtubules can be seen emanating from the centrosome. (C, from E.T. O'Toole et al., *J. Cell Biol.* 163:451–456, 2003. With permission from the authors.)

the centrosome acts as a device to survey the outlying regions of the cell and to position the centrosome at its center. Even in an isolated cell fragment lacking the centrosome, dynamic microtubules arrange themselves into a star-shaped array with the microtubule minus ends clustered at the center by minus-end-binding proteins (Figure 16-49). This ability of the microtubule cytoskeleton to find the center of the cell establishes a general coordinate system, which is then used to position many organelles within the cell. Highly differentiated cells with complex morphologies such as neurons, muscles, and epithelial cells must use additional measuring mechanisms to establish their more elaborate internal coordinate systems. Thus, for example, when an epithelial cell forms cell-cell junctions and becomes highly polarized, the microtubule minus ends move to a region near the



**Figure 16-48 A pair of centrioles in the centrosome.** (A) An electron micrograph of a thin section of an isolated centrosome showing the mother centriole with its distal appendages and the adjacent daughter centriole, which formed through a duplication event during S phase (see Figure 17-26). In the centrosome, the centriole pair is surrounded by a dense matrix of pericentriolar material from which microtubules nucleate. Centrioles also function as basal bodies to nucleate the formation of ciliary axonemes (see Figure 16-68). (B) Electron micrograph of a cross section through a centriole in the cortex of a protozoan. Each centriole is composed of nine sets of triplet microtubules arranged to form a cylinder. (C) Each triplet contains one complete microtubule (the A microtubule) fused to two incomplete microtubules (the B and C microtubules). (D) The centriolar protein SAS-6 forms a coiled-coil dimer. Nine SAS-6 dimers can self-associate to form a ring. Located at the hub of the centriole cartwheel structure, the SAS-6 ring is thought to generate the ninefold symmetry of the centriole. (A, from M. Paintrand, et al. *J. Struct. Biol.* 108:107, 1992. With permission from Elsevier; B, courtesy of Richard Linck; D, courtesy of Michel Steinmetz.)



**Figure 16–49** A microtubule array can find the center of a cell. After the arm of a fish pigment cell is cut off with a needle, the microtubules in the detached cell fragment reorganize so that their minus ends end up near the center of the fragment, buried in a new microtubule-organizing center.

apical plasma membrane. From this asymmetrical location, a microtubule array extends along the long axis of the cell, with plus ends directed toward the basal surface (see Figure 16–4).

### Microtubule-Binding Proteins Modulate Filament Dynamics and Organization

Microtubule polymerization dynamics are very different in cells than in solutions of pure tubulin. Microtubules in cells exhibit a much higher polymerization rate (typically 10–15  $\mu\text{m}/\text{min}$ , relative to about 1.5  $\mu\text{m}/\text{min}$  with purified tubulin at similar concentrations), a greater catastrophe frequency, and extended pauses in microtubule growth, a dynamic behavior rarely observed in pure tubulin solutions. These and other differences arise because microtubule dynamics inside the cell are governed by a variety of proteins that bind tubulin dimers or microtubules, as summarized in Panel 16–4.

Proteins that bind to microtubules are collectively called **microtubule-associated proteins**, or **MAPs**. Some MAPs can stabilize microtubules against disassembly. A subset of MAPs can also mediate the interaction of microtubules with other cell components. This subset is prominent in neurons, where stabilized microtubule bundles form the core of the axons and dendrites that extend from the cell body (Figure 16–50). These MAPs have at least one domain that binds to the microtubule surface and another that projects outward. The length of the projecting domain can determine how closely MAP-coated microtubules pack together, as demonstrated in cells engineered to overproduce different MAPs. Cells overexpressing MAP2, which has a long projecting domain, form bundles of stable microtubules that are kept widely spaced, while cells overexpressing tau, a MAP with a much shorter projecting domain, form bundles of more closely packed microtubules (Figure 16–51). MAPs are the targets of several protein kinases, and phosphorylation of a MAP can control both its activity and localization inside cells.

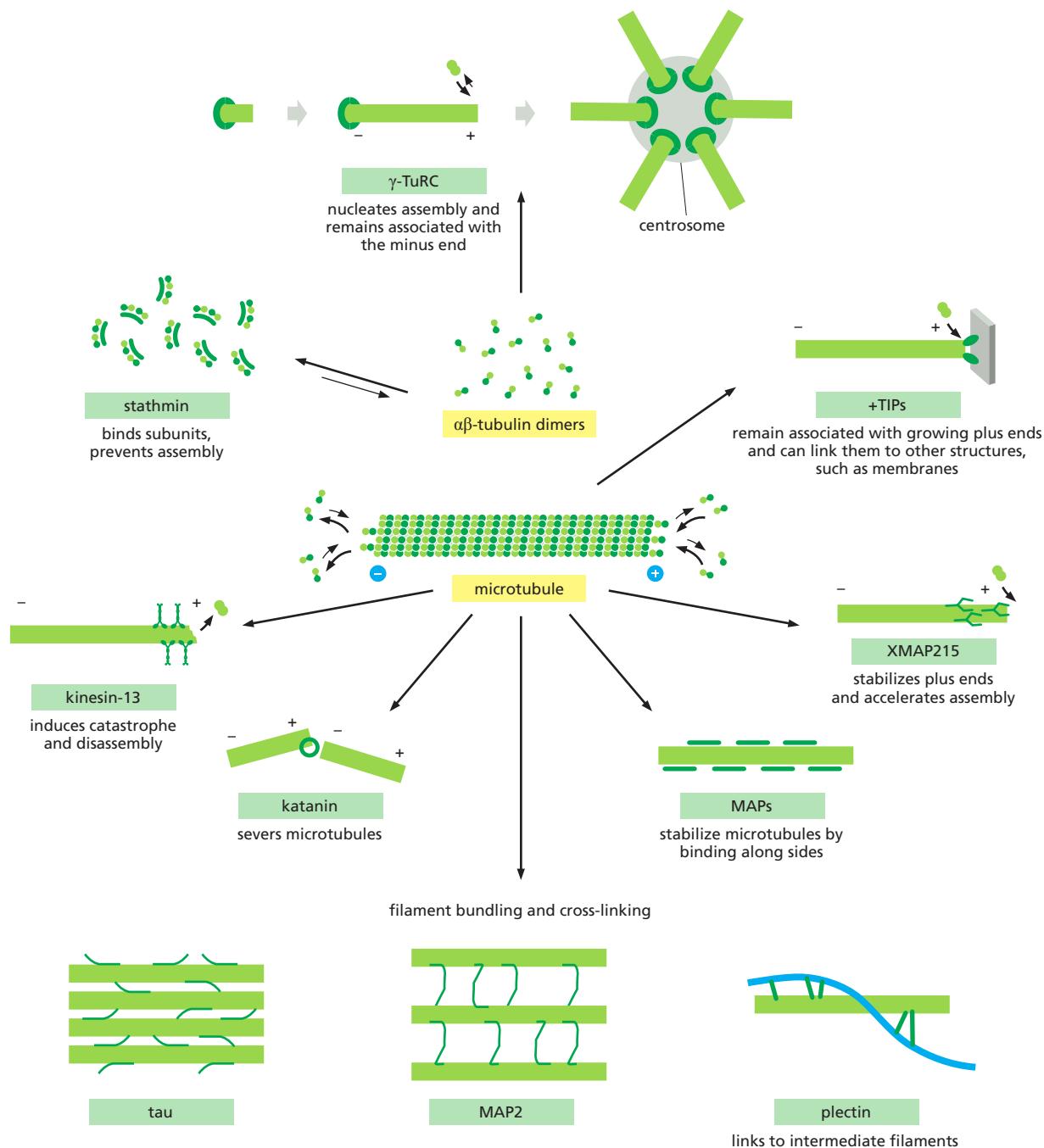
### Microtubule Plus-End-Binding Proteins Modulate Microtubule Dynamics and Attachments

Cells contain numerous proteins that bind the ends of microtubules and thereby influence microtubule stability and dynamics. These proteins can influence the

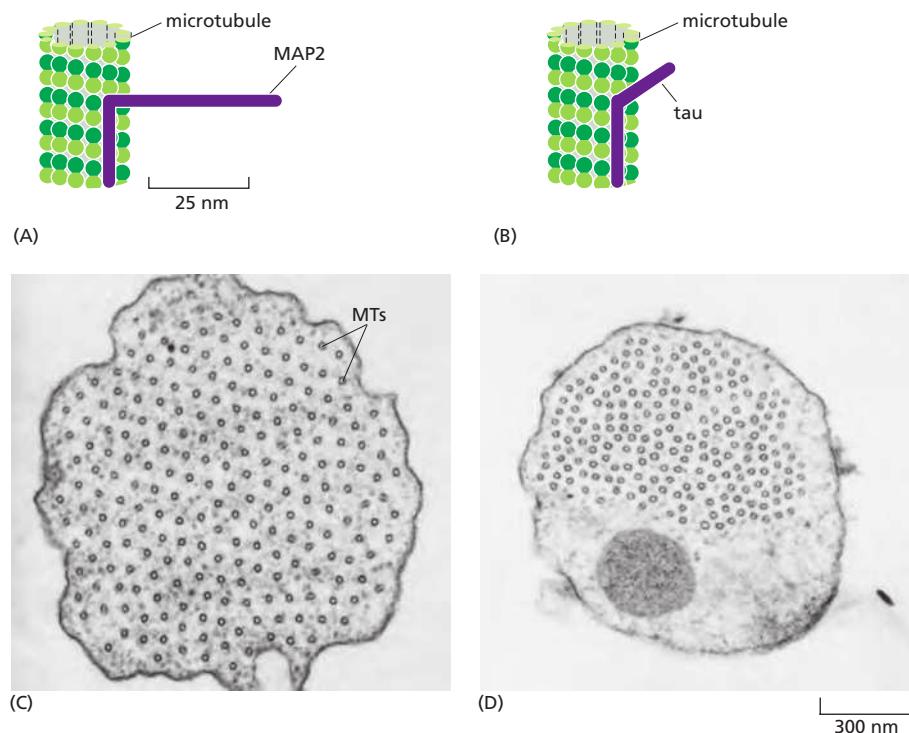
**Figure 16–50** Localization of MAPs in the axon and dendrites of a neuron. This immunofluorescence micrograph shows the distribution of the proteins tau (green) and MAP2 (orange) in a hippocampal neuron in culture. Whereas tau staining is confined to the axon (long and branched in this neuron), MAP2 staining is confined to the cell body and its dendrites. The antibody used here to detect tau binds only to unphosphorylated tau; phosphorylated tau is also present in dendrites. (Courtesy of James W. Mandell and Gary A. Banker.)



## MICROTUBULES

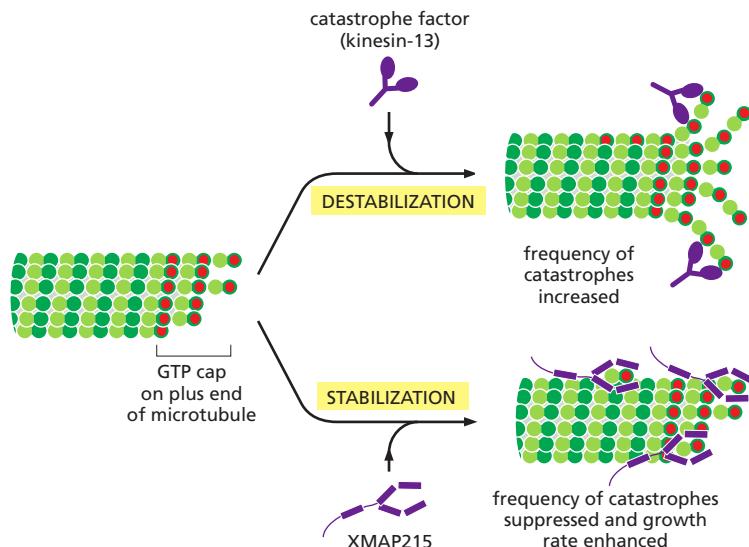


Some of the major accessory proteins of the microtubule cytoskeleton. Except for two classes of motor proteins, an example of each major type is shown. Each of these is discussed in the text. However, most cells contain more than a hundred different microtubule-binding proteins, and — as for the actin-associated proteins — it is likely that there are important types of microtubule-associated proteins that are not yet recognized.

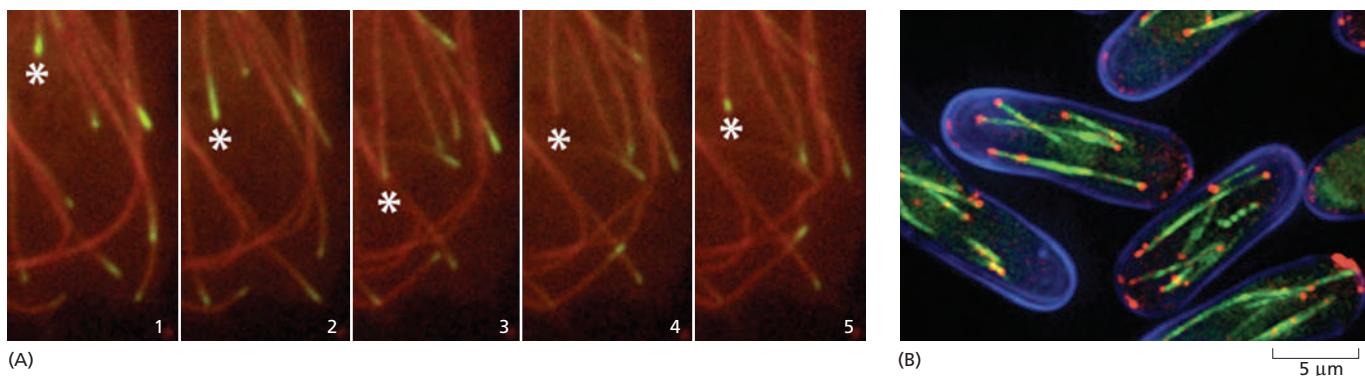


rate at which a microtubule switches from a growing to a shrinking state (the frequency of catastrophes) or from a shrinking to a growing state (the frequency of rescues). For example, members of a family of kinesin-related proteins known as *catastrophe factors* (or kinesin-13) bind to microtubule ends and appear to pry protofilaments apart, lowering the normal activation-energy barrier that prevents a microtubule from springing apart into the curved protofilaments that are characteristic of the shrinking state (Figure 16–52). Another protein, called Nezha or Patronin, protects microtubule minus ends from the effects of catastrophe factors.

While very few microtubule minus-end-binding proteins have been characterized, a large subset of MAPs has been identified that are enriched at microtubule plus ends. A particularly ubiquitous example is XMAP215, which has close homologs in organisms that range from yeast to humans. XMAP215 binds free tubulin subunits and delivers them to the plus end, thereby promoting microtubule polymerization and simultaneously counteracting catastrophe factor activity (see Figure 16–52). The phosphorylation of XMAP215 during mitosis inhibits



**Figure 16–52** The effects of proteins that bind to microtubule ends. The transition between microtubule growth and shrinkage is controlled in cells by a variety of proteins. Catastrophe factors such as kinesin-13, a member of the kinesin motor protein superfamily, bind to microtubule ends and pry them apart, thereby promoting depolymerization. On the other hand, a MAP such as XMAP215 stabilizes the end of a growing microtubule (XMAP stands for *Xenopus* microtubule-associated protein, and the number refers to its molecular mass in kilodaltons). XMAP215 binds tubulin dimers and delivers them to the microtubule plus end, thereby increasing the microtubule growth rate and suppressing catastrophes.



its activity and shifts the balance of its competition with catastrophe factors. This shift results in a tenfold increase in the dynamic instability of microtubules during mitosis, a transition that is critical for the efficient construction of the mitotic spindle (discussed in Chapter 17).

In many cells, the minus ends of microtubules are stabilized by association with a capping protein or the centrosome, or else they serve as microtubule depolymerization sites. The plus ends, in contrast, efficiently explore and probe the entire cell space. Microtubule-associated proteins called *plus-end tracking proteins* (+TIPs) accumulate at these active ends and appear to rocket around the cell as passengers at the ends of rapidly growing microtubules, dissociating from the ends when the microtubules begin to shrink (Figure 16–53).

The kinesin-related catastrophe factors and XMAP215 mentioned above behave as +TIPs and act to modulate the growth and shrinkage of the microtubule end to which they are attached. Other +TIPs control microtubule positioning by helping to capture and stabilize the growing microtubule end at specific cellular targets, such as the cell cortex or the kinetochore of a mitotic chromosome. EB1 and its relatives, small dimeric proteins that are highly conserved in animals, plants, and fungi, are key players in this process. EB1 proteins do not actively move toward plus ends, but rather recognize a structural feature of the growing plus end (see Figure 16–53). Several of the +TIPs depend on EB1 proteins for their plus-end accumulation and also interact with each other and with the microtubule lattice. By attaching to the plus end, these factors allow the cell to harness the energy of microtubule polymerization to generate pushing forces that can be used for positioning the spindle, chromosomes, or organelles.

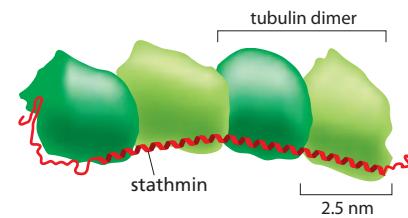
**Figure 16–53** +TIP proteins found at the growing plus ends of microtubules.

(A) Frames from a fluorescence time-lapse movie of the edge of a cell expressing fluorescently labeled tubulin that incorporates into microtubules (red) as well as the +TIP protein EB1 tagged with a different color (green). The same microtubule is marked (asterisk) in successive movie frames. When the microtubule is growing (frames 1, 2), EB1 is associated with the tip. When the microtubule undergoes a catastrophe and begins shrinking, EB1 is lost (frames 3, 4). The labeled EB1 is regained when growth of the microtubule is rescued (frame 5). See Movie 16.8. (B) In the fission yeast *Schizosaccharomyces pombe*, the plus ends of the microtubules (green) are associated with the homolog of EB1 (red) at the two poles of the rod-shaped cells. (A, courtesy of Anna Akhmanova and Ilya Grigoriev; B, courtesy of Takeshi Toda.)

## Tubulin-Sequestering and Microtubule-Severing Proteins Destabilize Microtubules

As it does with actin monomers, the cell sequesters unpolymerized tubulin subunits to maintain a pool of active subunits at a level near the critical concentration. One molecule of the small protein *stathmin* (also called Op18) binds to two tubulin heterodimers and prevents their addition to the ends of microtubules (Figure 16–54). Stathmin thus decreases the effective concentration of tubulin subunits that are available for polymerization (an action analogous to that of the drug colchicine), and enhances the likelihood that a growing microtubule will switch to the shrinking state. Phosphorylation of stathmin inhibits its binding to tubulin, and signals that cause stathmin phosphorylation can increase the rate of microtubule elongation and suppress dynamic instability. Stathmin has been implicated in the regulation of both cell proliferation and cell death. Interestingly, mice lacking stathmin develop normally but are less fearful than wild-type mice, reflecting a role for stathmin in neurons of the amygdala, where it is normally expressed at high levels.

Severing is another mechanism employed by the cell to destabilize microtubules. To sever a microtubule, thirteen longitudinal bonds must be broken, one for each protofilament. The protein *katanin*, named after the Japanese word for



**Figure 16–54** Sequestration of tubulin by stathmin. Structural studies with electron microscopy and crystallography suggest that the elongated stathmin protein binds along the side of two tubulin heterodimers. (Adapted from M.O. Steinmetz et al., *EMBO J.* 19:572–580, 2000. With permission from John Wiley and Sons.)

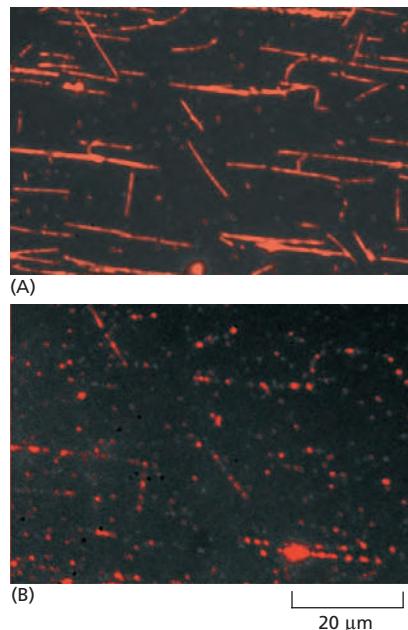
"sword," accomplishes this demanding task (Figure 16–55). Katanin is made up of two subunits: a smaller subunit that hydrolyzes ATP and performs the actual severing, and a larger one that directs katanin to the centrosome. Katanin releases microtubules from their attachment to a microtubule-organizing center and is thought to contribute to the rapid microtubule depolymerization observed at the poles of spindles during mitosis. It may also be involved in microtubule release and depolymerization in proliferating cells in interphase and in postmitotic cells such as neurons.

## Two Types of Motor Proteins Move Along Microtubules

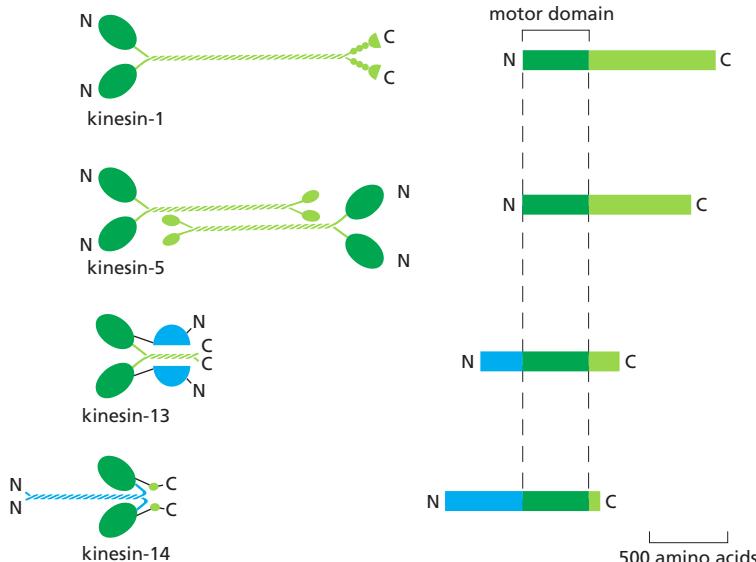
Like actin filaments, microtubules also use motor proteins to transport cargo and perform a variety of other functions within the cell. There are two major classes of microtubule-based motors, **kinesins** and **dyneins**. **Kinesin-1**, also called "conventional kinesin," was first purified from squid neurons, where it carries membrane-enclosed organelles away from the cell body toward the axon terminal by walking toward the plus end of microtubules. Kinesin-1 is similar to myosin II in having two heavy chains per active motor; these form two globular head motor domains that are held together by an elongated coiled-coil tail that is responsible for heavy-chain dimerization. One kinesin-1 light chain associates with each heavy chain through its tail domain and mediates cargo binding. Like myosin, kinesin is a member of a large protein superfamily, for which the motor domain is the common element (Figure 16–56). The yeast *Saccharomyces cerevisiae* has six distinct kinesins. The nematode *C. elegans* has 20 kinesins, and humans have 45.

There are at least fourteen distinct families in the kinesin superfamily. Most of them have the motor domain at the N-terminus of the heavy chain and walk toward the plus end of the microtubule. One family has the motor domain at the C-terminus and walks in the opposite direction, toward the minus end of the microtubule, while kinesin-13 has a central motor domain and does not walk at all, but uses the energy of ATP hydrolysis to depolymerize microtubule ends, as described above (see Figure 16–52). Some kinesin heavy chains are homodimers, and others are heterodimers. Most kinesins have a binding site in the tail for another microtubule; alternatively, they may link the motor to a membrane-enclosed organelle via a light chain or an adaptor protein. Many of the kinesin superfamily members have specific roles in mitotic spindle formation and in chromosome segregation during cell division.

In kinesin-1, instead of the rocking of a lever arm, small movements at the nucleotide-binding site regulate the docking and undocking of the motor head domain to a long linker region. This acts to throw the second head forward along



**Figure 16–55** Microtubule severing by katanin. Taxol stabilized, rhodamine-labeled microtubules were adsorbed on the surface of a glass slide, and purified katanin was added along with ATP. (A) There are a few breaks in the microtubules 30 seconds after the addition of katanin. (B) The same field 3 minutes after the addition of katanin. The filaments have been severed in many places, leaving a series of small fragments at the previous locations of the long microtubules. (From J.J. Hartman et al., *Cell* 93:277–287, 1998. With permission from Elsevier.)



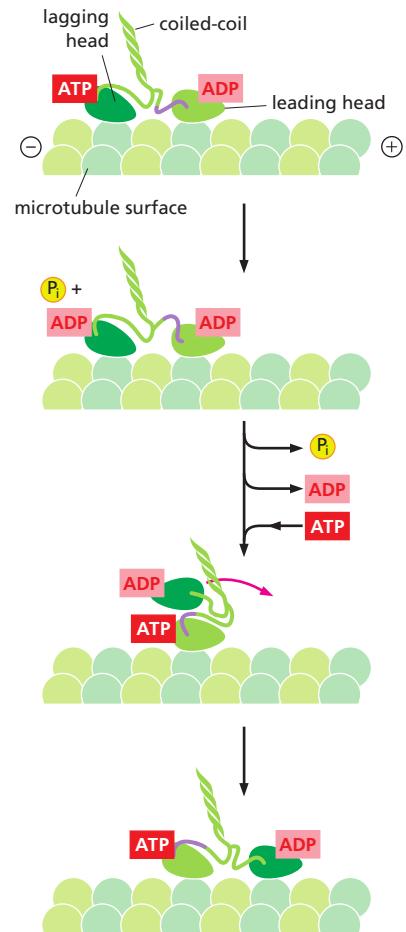
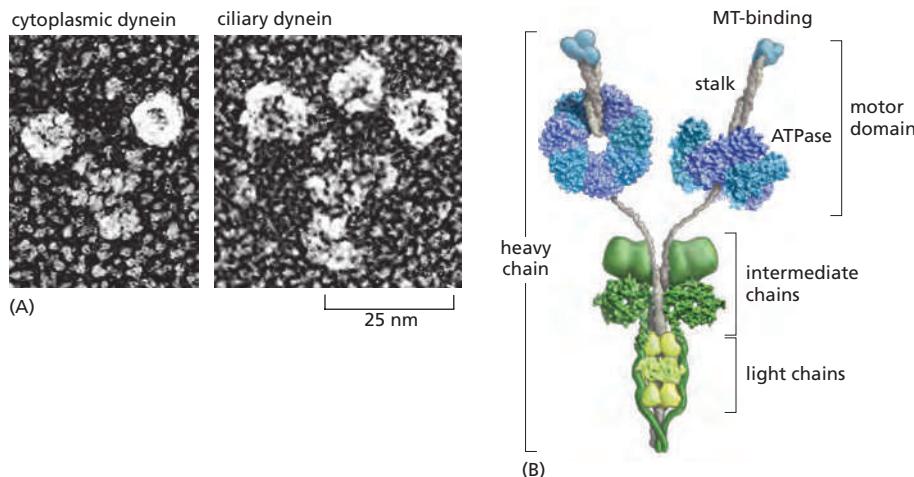
**Figure 16–56** Kinesin and kinesin-related proteins. Structures of four kinesin superfamily members. As in the myosin superfamily, only the motor domains are conserved. Kinesin-1 has the motor domain at the N-terminus of the heavy chain. The middle domain forms a long coiled-coil, mediating dimerization. The C-terminal domain forms a tail that attaches to cargo, such as a membrane-enclosed organelle. Kinesin-5 forms tetramers where two dimers associate by their tails. The bipolar kinesin-5 tetramer is able to slide two microtubules past each other, analogous to the activity of the bipolar thick filaments formed by myosin II. Kinesin-13 has its motor domain located in the middle of the heavy chain. It is a member of a family of kinesins that have lost typical motor activity and instead bind to microtubule ends to promote depolymerization (see Figure 16–52). Kinesin-14 is a C-terminal kinesin that includes the *Drosophila* protein Ncd and the yeast protein Kar3. These kinesins generally travel in the opposite direction from the majority of kinesins, toward the minus end instead of the plus end of a microtubule.

**Figure 16–57 The mechanochemical cycle of kinesin.** Kinesin-1 is a dimer of two nucleotide-binding motor domains (heads) that are connected through a long coiled-coil tail (see Figure 16–56). The two kinesin motor domains work in a coordinated manner; during a kinesin “step,” the rear head detaches from its tubulin binding site, passes the partner motor domain, and then rebinds to the next available tubulin binding site. Using this “hand-over-hand” motion, the kinesin dimer can move for long distances on the microtubule without completely letting go of its track.

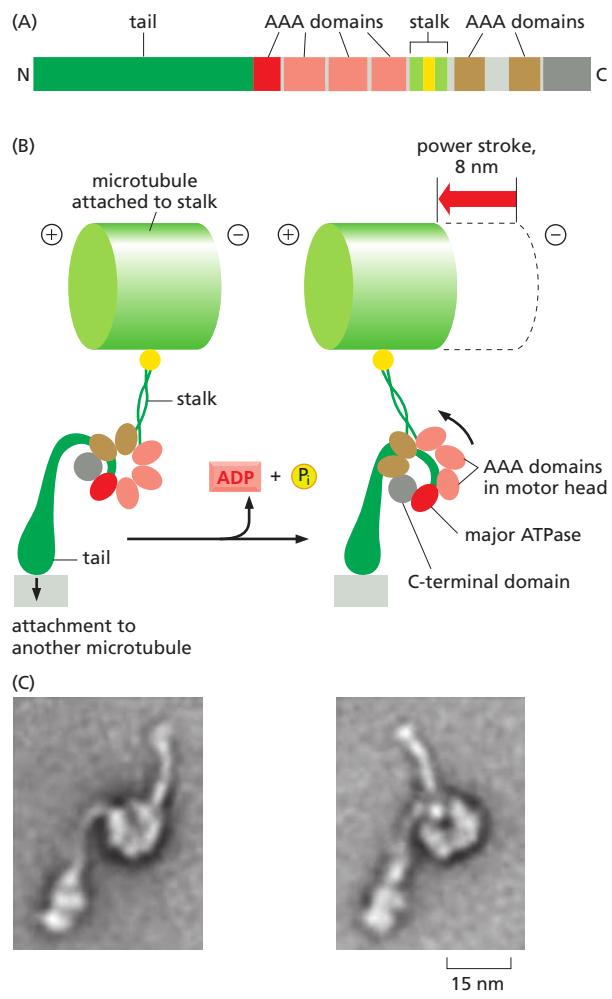
At the start of each step, one of the two kinesin motor domain heads, the rear or lagging head (dark green), is tightly bound to the microtubule and to ATP, while the front or leading head is loosely bound to the microtubule with ADP in its binding site. The forward displacement of the rear motor domain is driven by the dissociation of ADP and binding of ATP in the leading head (between panels 2 and 3 in this drawing). The binding of ATP to this motor domain causes a small peptide called the “neck linker” to shift from a rearward-pointing to a forward-pointing conformation (the neck linker is drawn here as a purple connecting line between the leading motor domain and the intertwined coiled-coil). This shift pulls the rear head forward, once it has detached from the microtubule with ADP bound [detachment requires ATP hydrolysis and phosphate ( $P_i$ ) release]. The kinesin molecule is now poised for the next step, which proceeds by an exact repeat of the process shown (Movie 16.9).

the protofilament to a binding site 8 nm closer to the microtubule plus end, which is the distance between tubulin dimers of a protofilament. The nucleotide-hydrolysis cycles in the two heads are closely coordinated, so that this cycle of linker docking and undocking allows the two-headed motor to move in a hand-over-hand (or head-over-head) stepwise manner (Figure 16–57).

The dyneins are a family of minus-end directed microtubule motors unrelated to the kinesins. They are composed of one, two, or three heavy chains (that include the motor domain) and a large and variable number of associated intermediate, light-intermediate, and light chains. The dynein family has two major branches (Figure 16–58). The first branch contains the *cytoplasmic dyneins*, which are homodimers of two heavy chains. Cytoplasmic dynein 1 is encoded by a single gene in almost all eukaryotic cells, but is missing from flowering plants and some algae. It is used for organelle and mRNA trafficking, for positioning the centrosome and nucleus during cell migration, and for construction of the microtubule spindle in mitosis and meiosis. Cytoplasmic dynein 2 is found only in eukaryotic organisms that have cilia and is used to transport material from the tip to the base of the cilia, a process called intraflagellar transport. *Axonemal dyneins* (also called *ciliary dyneins*) comprise the second branch and include monomers, heterodimers, and heterotrimers, with one, two, or three motor-containing heavy chains, respectively. They are highly specialized for the rapid and efficient sliding movements of microtubules that drive the beating of cilia and flagella (discussed later).



**Figure 16–58 Dyneins.** (A) Freeze-etch electron micrographs of a molecule of cytoplasmic dynein and a molecule of ciliary dynein (axonemal dynein). Like myosin II and kinesin-1, cytoplasmic dynein is a two-headed molecule. The ciliary dynein shown is from a protozoan and has three heads; ciliary dynein from animals has two heads. Note that the dynein head is very large compared with the head of either myosin or kinesin. (B) Schematic depiction of cytoplasmic dynein showing the two heavy chains (blue and gray) that contain domains for microtubule (MT) binding and ATP hydrolysis, connected by a long stalk. Bound to the heavy chain are multiple intermediate chains (dark green) and light chains (light green) that help to mediate many of dynein's functions. (A, courtesy of John Heuser; B, adapted from R. Vale, *Cell* 112:467–480, 2003. With permission from Cell Press.)



**Figure 16–59** The power stroke of dynein. (A) The organization of the domains in each dynein heavy chain. This is a huge polypeptide, containing nearly 4000 amino acids. The number of heavy chains in a dynein is equal to its number of motor heads. (B) Illustration of dynein c, a monomeric axonemal dynein found in the unicellular green alga *Chlamydomonas reinhardtii*. The large dynein motor head is a planar ring containing a C-terminal domain (gray) and six AAA domains, four of which retain ATP-binding sequences, but only one of which (dark red) has the major ATPase activity. Extending from the head are a long, coiled-coil stalk with the microtubule-binding site at the tip, and a tail that attaches to an adjacent microtubule in the axoneme. In the ATP-bound state, the stalk is detached from the microtubule, but ATP hydrolysis causes stalk–microtubule attachment (left). Subsequent release of ADP and phosphate ( $P_i$ ) then leads to a large conformational “power stroke” involving rotation of the head and stalk relative to the tail (right). Each cycle generates a step of about 8 nm, thereby contributing to flagellar beating (see Figure 16–65). In the case of cytoplasmic dynein, the tail is attached to a cargo such as a vesicle, and a single power stroke transports the cargo about 8-nm along the microtubule toward its minus end (see Figure 16–60). (C) Electron micrographs of purified monomeric dyneins in two different conformations representing different steps in the mechanochemical cycle. (C, from S.A. Burgess et al., *Nature* 421:715–718, 2003. With permission from Macmillan Publishers Ltd.)

Dyneins are the largest of the known molecular motors, and they are also among the fastest: axonemal dyneins attached to a glass slide can move microtubules at the rate of 14  $\mu\text{m/sec}$ . The dynein motor is structurally unrelated to myosins and kinesins, but still follows the general rule of coupling nucleotide hydrolysis to microtubule binding and unbinding as well as to a force-generating conformational change (Figure 16–59).

### Microtubules and Motors Move Organelles and Vesicles

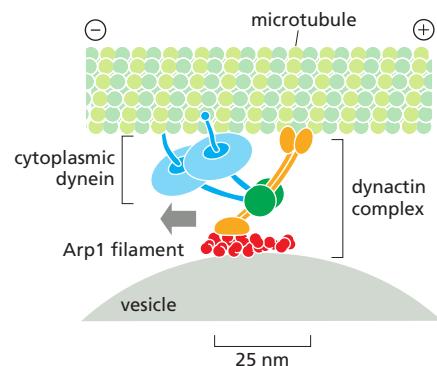
A major function of cytoskeletal motors in interphase cells is the transport and positioning of membrane-enclosed organelles (Movie 16.10). Kinesin was originally identified as the protein responsible for fast *anterograde axonal transport*, the rapid movement of mitochondria, secretory vesicle precursors, and various synapse components down the microtubule highways of the axon to the distant nerve terminals. Cytoplasmic dynein was identified as the motor responsible for transport in the opposite direction, *retrograde axonal transport*. Although organelles in most cells need not cover such long distances, their polarized transport is equally necessary. A typical microtubule array in an interphase cell is oriented with the minus ends near the center of the cell at the centrosome and the plus ends extending to the cell periphery. Thus, centripetal movements of organelles or vesicles toward the cell center require the action of minus-end directed cytoplasmic dynein motors, whereas centrifugal movements toward the periphery require plus-end directed kinesin motors. Interestingly, in animal cells, nearly all minus-end directed transport is driven by the single cytoplasmic dynein 1 motor, whereas 15 different kinesins are used for plus-end directed transport.

A clear example of the effect of microtubules and microtubule motors on the behavior of intracellular membranes is their role in organizing the endoplasmic reticulum (ER) and the Golgi apparatus. The network of ER membrane tubules aligns with microtubules and extends almost to the edge of the cell ([Movie 16.11](#)), whereas the Golgi apparatus is located near the centrosome. When cells are treated with a drug that depolymerizes microtubules, such as colchicine or nocodazole, the ER collapses to the center of the cell, while the Golgi apparatus fragments and disperses throughout the cytoplasm. *In vitro*, kinesins can tether ER-derived membranes to preformed microtubule tracks and walk toward the microtubule plus ends, dragging the ER membranes out into tubular protrusions and forming a membranous web that looks very much like the ER in cells. Conversely, dyneins are required for positioning the Golgi apparatus near the cell center of animal cells; they do this by moving Golgi vesicles along microtubule tracks toward the microtubules' minus ends at the centrosome.

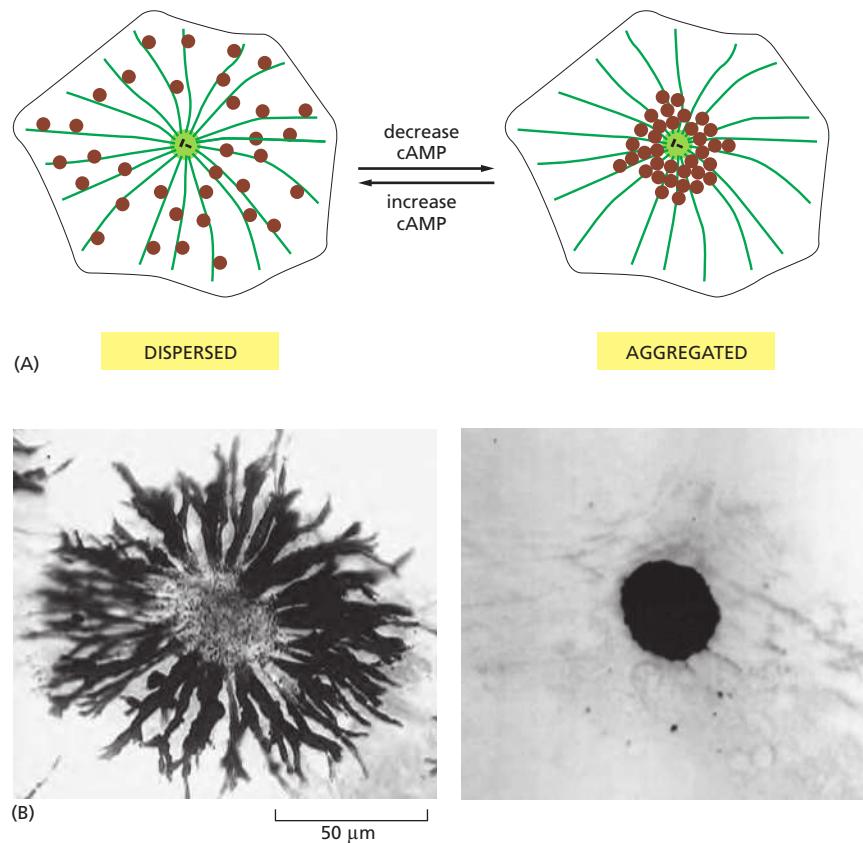
The different tails and their associated light chains on specific motor proteins allow the motors to attach to their appropriate organelle cargo. Membrane-associated motor receptors that are sorted to specific membrane-enclosed compartments interact directly or indirectly with the tails of the appropriate kinesin family members. Many viruses take advantage of microtubule motor-based transport during infection and use kinesin to move from their site of replication and assembly to the plasma membrane, from which they are poised to infect neighboring cells. An outer-membrane protein of *Vaccinia* virus, for example, contains an amino acid motif that mediates binding to kinesin-1 light chain and transport along microtubules to the plasma membrane. Interestingly, this motif is present in over 450 human proteins, one-third of which are associated with human diseases. Thus, kinesin transports a diverse set of cargoes involved in a wide range of important cellular functions.

For dynein, a large macromolecular assembly often mediates attachment to membranes. Cytoplasmic dynein, itself a huge protein complex, requires association with a second large protein complex called *dynactin* to translocate organelles effectively. The dynactin complex includes a short, actin-like filament that forms from the actin-related protein Arp1 (distinct from Arp2 and Arp3, the components of the Arp 2/3 complex involved in the nucleation of conventional actin filaments) ([Figure 16–60](#)). A number of other proteins also contribute to dynein cargo binding and motor regulation, and their function is especially important in neurons, where defects in microtubule-based transport have been linked to neurological diseases. A striking example is smooth brain, or lissencephaly, a human disorder in which cells fail to migrate to the cerebral cortex of the developing brain. One type of lissencephaly is caused by defects in Lis1, a dynein-binding protein required for nuclear migration in several species. In the normal brain, migration of the nucleus directs the developing neural cell body toward its correct position in the cortex. In the absence of Lis1, however, the nuclei of migrating neurons fail to attach to dynein, resulting in nuclear-migration defects. Dynein is required continuously for neuronal function, as mutations in a dynactin subunit or in the tail region of cytoplasmic dynein lead to neuronal degeneration in humans and mice. These effects are associated with decreased retrograde axonal transport and provide strong evidence for the importance of robust axonal transport in neuronal viability.

The cell can regulate the activity of motor proteins and thereby cause either a change in the positioning of its membrane-enclosed organelles or whole-cell movements. Fish melanocytes provide one of the most dramatic examples. These giant cells, which are responsible for rapid changes in skin coloration in several species of fish, contain large pigment granules that can alter their location in response to neuronal or hormonal stimulation ([Figure 16–61](#)). The pigment granules aggregate or disperse by moving along an extensive network of microtubules that are anchored at the centrosome by their minus ends. The tracking of individual pigment granules reveals that the inward movement is rapid and smooth, while the outward movement is jerky, with frequent backward steps. Both dynein and kinesin microtubule motors are associated with the pigment granules. The



**Figure 16–60** Dynactin mediates the attachment of dynein to a membrane-enclosed organelle. Dynein requires the presence of a large number of accessory proteins to associate with membrane-enclosed organelles. Dynactin is a large complex that includes components that bind weakly to microtubules, components that bind to dynein itself, and components that form a small, actin-like filament made of the actin-related protein Arp1.



**Figure 16–61** Regulated melanosome movements in fish pigment cells.

These giant cells, which are responsible for changes in skin coloration in several species of fish, contain large pigment granules, or melanosomes (brown). The melanosomes can change their location in the cell in response to a hormonal or neuronal stimulus. (A) Schematic view of a pigment cell, showing the dispersal and aggregation of melanosomes in response to an increase or decrease in intracellular cyclic AMP (cAMP), respectively. Both redistributions of melanosomes occur along microtubules. (B) Bright-field images of a single cell in a scale of an African cichlid fish, showing its melanosomes either dispersed throughout the cytoplasm (left) or aggregated in the center of the cell (right). (B, courtesy of Leah Haimo.)

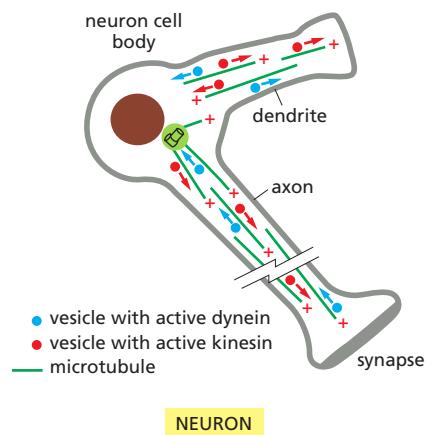
jerky outward movements may result from a tug-of-war between the two opposing microtubule motor proteins, with the stronger kinesin winning out overall. When intracellular cyclic AMP levels decrease, kinesin is inactivated, leaving dynein free to drag the pigment granules rapidly toward the cell center, changing the fish's color. In a similar way, the movement of other membrane organelles coated with particular motor proteins is controlled by a complex balance of competing signals that regulate both motor protein attachment and activity.

### Construction of Complex Microtubule Assemblies Requires Microtubule Dynamics and Motor Proteins

The construction of the mitotic spindle and the neuronal cytoskeleton are important and fascinating examples of the power of organization by teams of motor proteins interacting with dynamic cytoskeletal filaments. As described in Chapter 17, mitotic spindle assembly depends on reorganization of the interphase array of microtubules to form a bipolar array of microtubules, with their minus ends focused at the poles and their plus ends overlapping in the center or connecting to chromosomes. Spindle assembly depends on the coordinated actions of several motor proteins and other factors that modulate polymerization dynamics (see Figures 17–23 and 17–25).

Neurons also contain complex cytoskeletal structures. As they differentiate, neurons send out specialized processes that will either receive electrical signals (*dendrites*) or transmit electrical signals (*axons*) (see Figure 16–50). The beautiful and elaborate branching morphology of axons and dendrites enables neurons to form tremendously complex signaling networks, interacting with many other cells simultaneously and making possible the complicated behavior of the higher animals. Both axons and dendrites (collectively called *neurites*) are filled with bundles of microtubules that are critical to both their structure and their function.

In axons, all the microtubules are oriented in the same direction, with their minus end pointing back toward the cell body and their plus end pointing toward the axon terminals (Figure 16–62). The microtubules do not reach from the cell



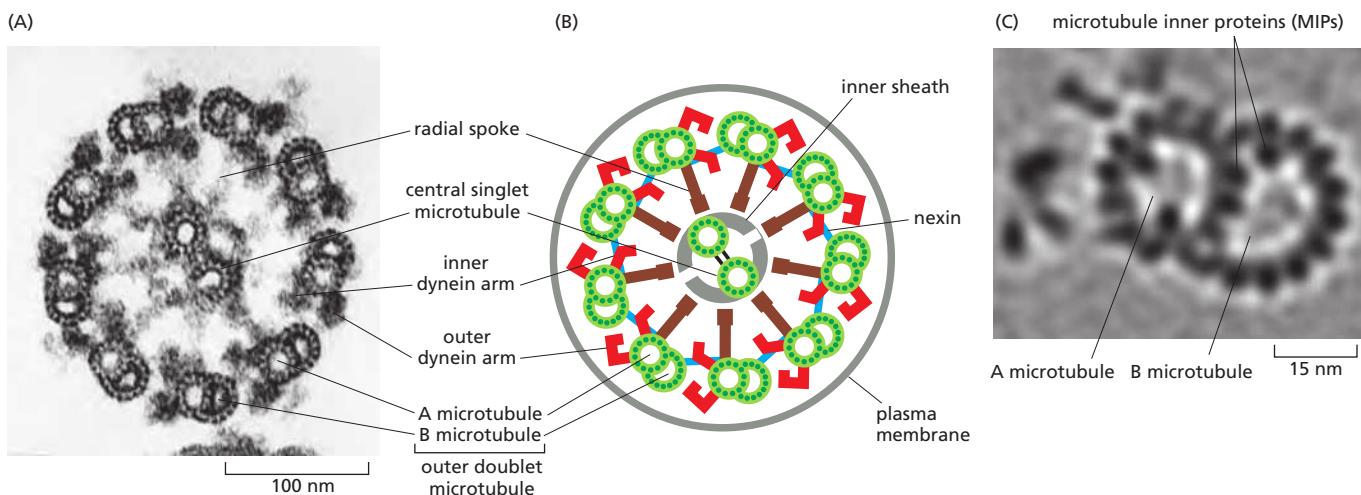
**Figure 16–62** Microtubule organization in a neuron. In a neuron, microtubule organization is complex. In the axon, all microtubules share the same polarity, with the plus ends pointing outward toward the axon terminus. No single microtubule stretches the entire length of the axon; instead, short overlapping segments of parallel microtubules make the tracks for fast axonal transport. In dendrites, the microtubules are of mixed polarity, with some plus ends pointing outward and some pointing inward. Vesicles can associate with both kinesin and dynein and move in either direction along the microtubules in axons and dendrites, depending on which motor is active.

body all the way to the axon terminals; each is typically only a few micrometers in length, but large numbers are staggered in an overlapping array. These aligned microtubule tracks act as a highway to transport specific proteins, protein-containing vesicles, and mRNAs to the axon terminals, where synapses are constructed and maintained. The longest axon in the human body reaches from the base of the spinal cord to the foot and is up to a meter in length. By comparison, dendrites are generally much shorter than axons. The microtubules in dendrites lie parallel to one another but their polarities are mixed, with some pointing their plus ends toward the dendrite tip, while others point back toward the cell body, reminiscent of the antiparallel microtubule array of the mitotic spindle.

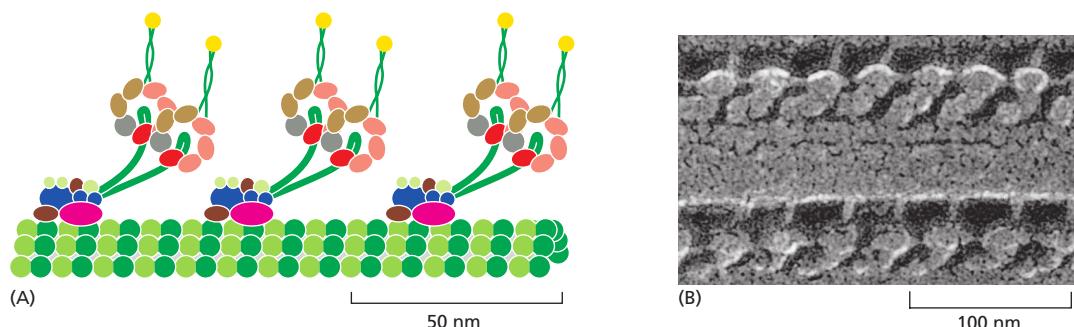
### Motile Cilia and Flagella Are Built from Microtubules and Dyneins

Just as myofibrils are highly specialized and efficient motility machines built from actin and myosin filaments, cilia and flagella are highly specialized and efficient motility structures built from microtubules and dynein. Both cilia and flagella are hairlike cell appendages that have a bundle of microtubules at their core. **Flagella** are found on sperm and many protozoa. By their undulating motion, they enable the cells to which they are attached to swim through liquid media. **Cilia** are organized in a similar fashion, but they beat with a whiplike motion that resembles the breaststroke in swimming. Ciliary beating can either propel single cells through a fluid (as in the swimming of the protozoan *Paramecium*) or can move fluid over the surface of a group of cells in a tissue. In the human body, huge numbers of cilia ( $10^9/\text{cm}^2$  or more) line our respiratory tract, sweeping layers of mucus, trapped particles of dust, and bacteria up to the mouth where they are swallowed and ultimately eliminated. Likewise, cilia along the oviduct help to sweep eggs toward the uterus.

The movement of a cilium or a flagellum is produced by the bending of its core, which is called the **axoneme**. The axoneme is composed of microtubules and their associated proteins, arranged in a distinctive and regular pattern. Nine special doublet microtubules (comprising one complete and one partial microtubule fused together so that they share a common tubule wall) are arranged in a ring around a pair of single microtubules (Figure 16–63). Almost all forms of motile eukaryotic flagella and cilia (from protozoans to humans) have this characteristic arrangement. The microtubules extend continuously for the length of the axoneme, which can be 10–200  $\mu\text{m}$ . At regular positions along the length of the microtubules, accessory proteins cross-link the microtubules together.



**Figure 16–63 The arrangement of microtubules in a flagellum or cilium.** (A) Electron micrograph of the flagellum of a green-alga cell (*Chlamydomonas*) shown in cross section, illustrating the distinctive “9 + 2” arrangement of microtubules. (B) Diagram of the parts of a flagellum or cilium. The various projections from the microtubules link the microtubules together and occur at regular intervals along the length of the axoneme. (C) High-resolution electron tomography image of an outer doublet microtubule showing structural details and features inside the microtubules called microtubule inner proteins (MIPs). (A, courtesy of Lewis Tilney; C, courtesy of Daniela Nicastro.)



Molecules of *axonemal dynein* form bridges between the neighboring doublet microtubules around the circumference of the axoneme (Figure 16–64). When the motor domain of this dynein is activated, the dynein molecules attached to one microtubule doublet (see Figure 16–59) attempt to walk along the adjacent microtubule doublet, tending to force the adjacent doublets to slide relative to one another, much as actin thin filaments slide during muscle contraction. However, the presence of other links between the microtubule doublets prevents this sliding, and the dynein force is instead converted into a bending motion (Figure 16–65).

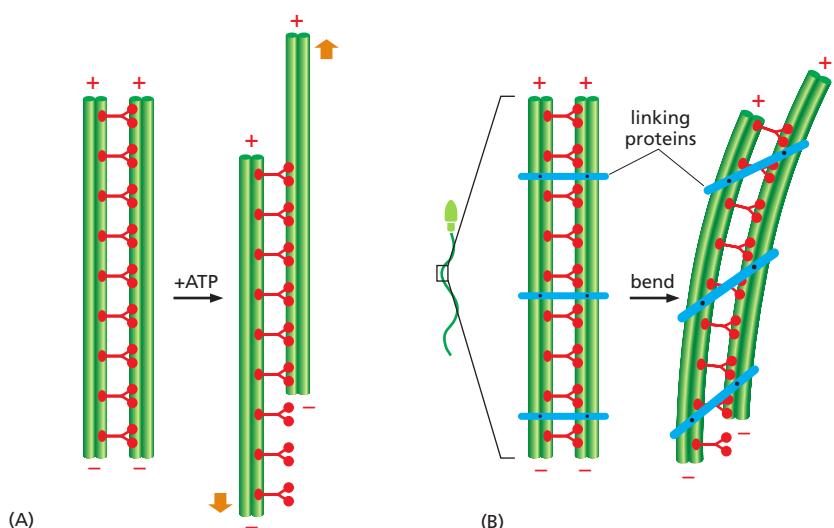
In humans, hereditary defects in axonemal dynein cause a condition called primary ciliary dyskinesia or Kartagener's syndrome. This syndrome is characterized by inversion of the normal asymmetry of internal organs (sinus inversus) due to disruption of fluid flow in the developing embryo, male sterility due to immotile sperm, and a high susceptibility to lung infections due to paralyzed cilia being unable to clear the respiratory tract of debris and bacteria.

Bacteria also swim using cell-surface structures called flagella, but these do not contain microtubules or dynein and do not wave or beat. Instead, *bacterial flagella* are long, rigid helical filaments, made up of repeating subunits of the protein flagellin. The flagella rotate like propellers, driven by a special rotary motor embedded in the bacterial cell wall. The use of the same name to denote these two very different types of swimming apparatus is an unfortunate historical accident.

### Primary Cilia Perform Important Signaling Functions in Animal Cells

Many cells possess a shorter, nonmotile counterpart of cilia and flagella called the *primary cilium*. Primary cilia can be viewed as specialized cellular compartments or organelles that perform a wide range of cellular functions, but share

**Figure 16–64 Ciliary dynein.** Ciliary (axonemal) dynein is a large protein assembly (nearly 2 million daltons) composed of 9–12 polypeptide chains, the largest of which is the heavy chain of more than 500,000 daltons. (A) The heavy chains form the major portion of the globular head and stem domains, and many of the smaller chains are clustered around the base of the stem. There are two heads in the outer dynein in metazoans (shown here), but three heads in protozoa, each formed from their own heavy chain. The tail of the molecule binds tightly to an A microtubule, while the large globular heads have an ATP-dependent binding site for a B microtubule (see Figure 16–63). When the heads hydrolyze their bound ATP, they move toward the minus end of the B microtubule, thereby producing a sliding force between the adjacent microtubule doublets in a cilium or flagellum (see Figure 16–59). (B) Freeze-etch electron micrograph of a cilium showing the dynein arms projecting at regular intervals from the doublet microtubules. (B, courtesy of John Heuser.)



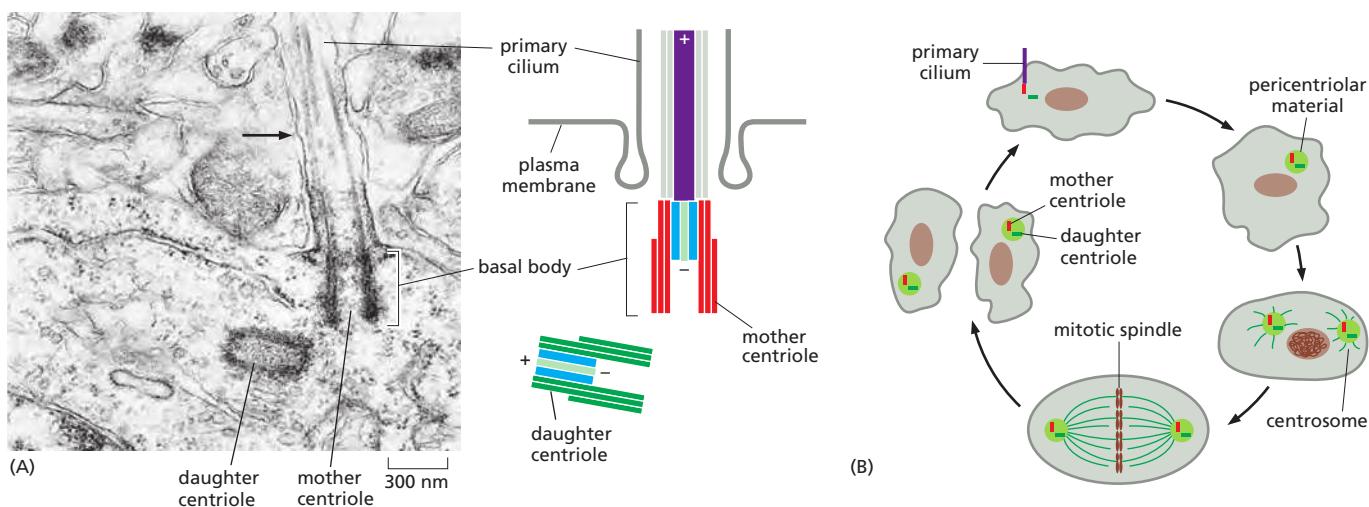
**Figure 16–65 The bending of an axoneme.** (A) When axonemes are exposed to the proteolytic enzyme trypsin, the linkages holding neighboring doublet microtubules together are broken. In this case, the addition of ATP allows the motor action of the dynein heads to slide one pair of doublet microtubules against the other pair. (B) In an intact axoneme (such as in a sperm), flexible protein links prevent the sliding of the doublet. The motor action therefore causes a bending motion, creating waves or beating motions.

many structural features with motile cilia. Both motile and nonmotile cilia are generated during interphase at plasma-membrane-associated structures called *basal bodies* that firmly root them at the cell surface. At the core of each basal body is a centriole, the same structure found embedded at the center of animal centrosomes, with nine groups of fused triplet microtubules arranged in a cartwheel (see Figure 16–48). Centrioles are multifunctional, contributing to assembly of the mitotic spindle in dividing cells but migrating to the plasma membrane of interphase cells to template the nucleation of the axoneme (**Figure 16–66**). Because no protein translation occurs in cilia, construction of the axoneme requires intraflagellar transport (IFT), a transport system discovered in the green algae *Chlamydomonas*. Analogous to the axon, motors move cargoes in both anterograde and retrograde directions, in this case driven by kinesin-2 and cytoplasmic dynein 2, respectively.

Primary cilia are found on the surface of almost all cell types, where they sense and respond to the exterior environment, functions best understood in the context of smell and sight. In the nasal epithelium, cilia protruding from dendrites of olfactory neurons are the site of both odorant reception and signal amplification. Similarly, the rod and cone cells of the vertebrate retina possess a primary cilium equipped with an expanded tip called the outer segment, which is specialized for converting light into a neural signal (see Figure 15–38). Maintenance of the outer segment requires continuous IFT-mediated transport of large quantities of lipids and proteins into the cilium, at rates of up to 2000 molecules per minute. The links between cilia function and the senses of sight and smell are underscored by Bardet-Biedl syndrome, a set of disorders associated with defects in IFT, the cilium, or the basal body. Patients with Bardet-Biedl syndrome cannot smell and suffer from retinal degeneration. Other characteristics of this multifaceted disorder include hearing loss, polycystic kidney disease, diabetes, obesity, and polydactyly, suggesting that primary cilia have functions in many aspects of human physiology.

## Summary

*Microtubules are stiff polymers of tubulin molecules. They assemble by addition of GTP-containing tubulin subunits to the free end of a microtubule, with one end (the plus end) growing faster than the other. Hydrolysis of the bound GTP takes place*



**Figure 16–66 Primary cilia.** (A) Electron micrograph and diagram of the basal body of a mouse neuron primary cilium. The axoneme of the primary cilium (black arrow) is nucleated by the mother centriole at the basal body, which localizes at the plasma membrane near the cell surface. (B) Centrioles function alternately as basal bodies and as the core of centrosomes. Before a cell enters the cell division cycle, the primary cilium is shed or resorbed. The centrioles recruit pericentriolar material and duplicate during S phase, generating two centrosomes, each of which contains a pair of centrioles. The centrosomes nucleate microtubules and localize to the poles of the mitotic spindle. Upon exit from mitosis, a primary cilium again grows from the mother centriole. (A, courtesy of Josef Spacek.)

after assembly and weakens the bonds that hold the microtubule together. Microtubules are dynamically unstable and liable to catastrophic disassembly, but they can be stabilized in cells by association with other structures. Microtubule-organizing centers such as centrosomes protect the minus ends of microtubules and continually nucleate the formation of new microtubules. Microtubule-associated proteins (MAPs) stabilize microtubules, and those that localize to the plus end (+TIPs) can alter the dynamic properties of the microtubule or mediate their interaction with other structures. Counteracting the stabilizing activity of MAPs are catastrophe factors, such as kinesin-13 proteins, that act to peel apart microtubule ends. Other kinesin family members as well as dynein use the energy of ATP hydrolysis to move unidirectionally along a microtubule. The motor dynein moves toward the minus end of microtubules, and its sliding of axonemal microtubules underlies the beating of cilia and flagella. Primary cilia are nonmotile sensory organs found on many cell types.

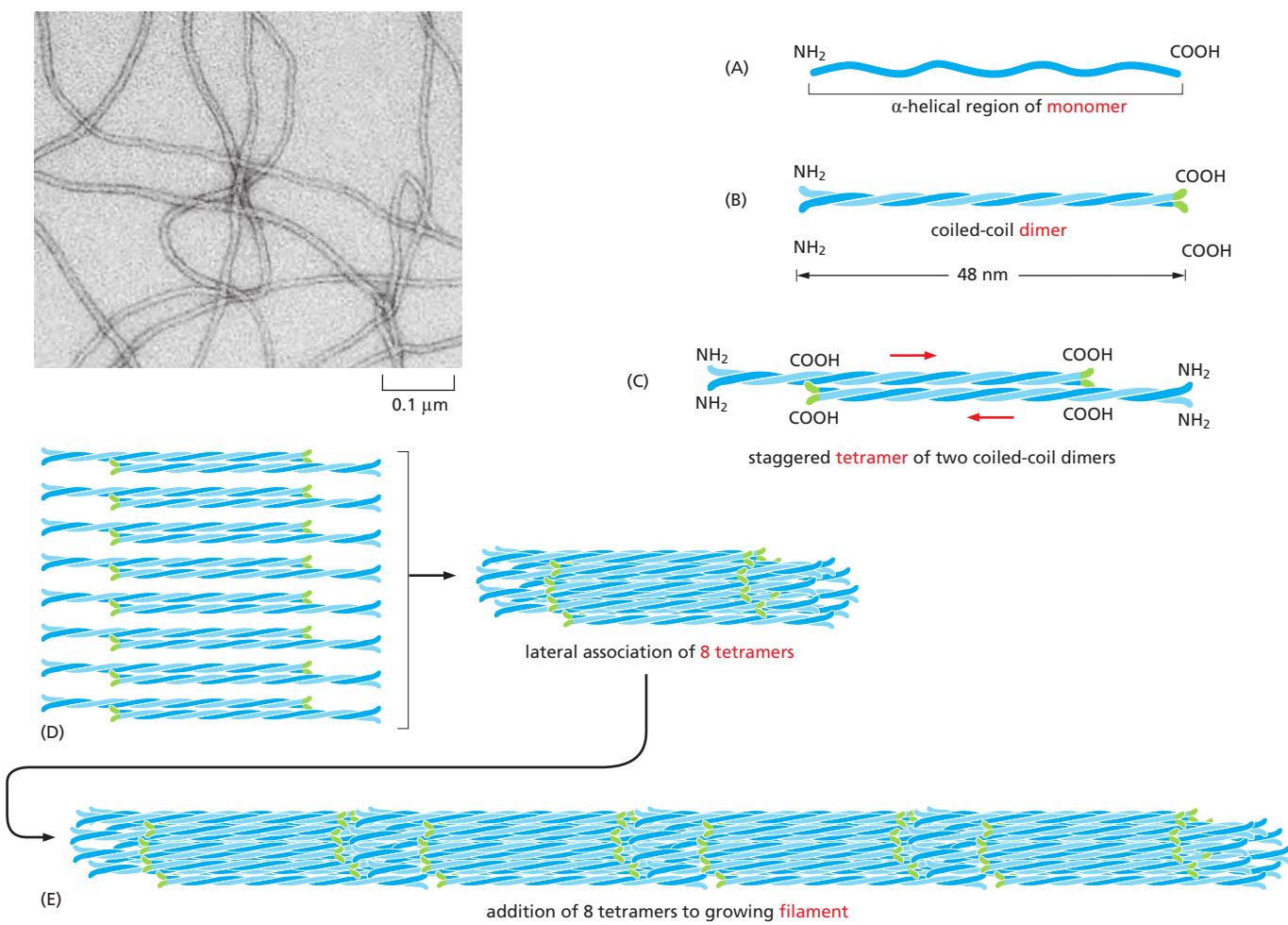
## INTERMEDIATE FILAMENTS AND SEPTINS

All eukaryotic cells contain actin and tubulin. But the third major type of cytoskeletal protein, the *intermediate filament*, forms a cytoplasmic filament only in some metazoans—including vertebrates, nematodes, and mollusks. Intermediate filaments are particularly prominent in the cytoplasm of cells that are subject to mechanical stress and are generally not found in animals that have rigid exoskeletons, such as arthropods and echinoderms. It seems that intermediate filaments impart mechanical strength to tissues for the squishier animals.

Cytoplasmic intermediate filaments are closely related to their ancestors, the much more prevalent *nuclear lamins*, which are found in many eukaryotes but missing from unicellular organisms. The nuclear lamins form a meshwork lining the inner membrane of the nuclear envelope, where they provide anchorage sites for chromosomes and nuclear pores. Several times during metazoan evolution, lamin genes have apparently duplicated, and the duplicates have evolved to produce ropelike, cytoplasmic intermediate filaments. In contrast to the highly conserved actins and tubulin isoforms that are encoded by a handful of genes, different families of intermediate filaments are much more diverse and are encoded by 70 different human genes with distinct, cell type-specific functions (**Table 16-2**).

**TABLE 16–2 Major Types of Intermediate Filament Proteins in Vertebrate Cells**

Types of intermediate filament	Component polypeptides	Location
Nuclear	Lamins A, B, and C	Nuclear lamina (inner lining of nuclear envelope)
Vimentin-like	Vimentin	Many cells of mesenchymal origin
	Desmin	Muscle
	Glial fibrillary acidic protein	Glial cells (astrocytes and some Schwann cells)
	Peripherin	Some neurons
Epithelial	Type I keratins (acidic)	Epithelial cells and their derivatives (e.g., hair and nails)
	Type II keratins (neutral/basic)	
Axonal	Neurofilament proteins (NF-L, NF-M, and NF-H)	Neurons



**Figure 16–67** A model of intermediate filament construction. The monomer shown in (A) pairs with another monomer to form a dimer (B), in which the conserved central rod domains are aligned in parallel and wound together into a coiled-coil. (C) Two dimers then line up side by side to form an antiparallel tetramer of four polypeptide chains. Dimers and tetramers are the soluble subunits of intermediate filaments. (D) Within each tetramer, the two dimers are offset with respect to one another, thereby allowing it to associate with another tetramer. (E) In the final 10-nm ropelike filament, tetramers are packed together in a helical array, which has 16 dimers (32 coiled-coils) in cross section. Half of these dimers are pointing in each direction. An electron micrograph of intermediate filaments is shown on the upper left ([Movie 16.12](#)). (Electron micrograph courtesy of Roy Quinlan.)

### Intermediate Filament Structure Depends on the Lateral Bundling and Twisting of Coiled-Coils

Although their amino- and carboxy-terminal domains differ, all intermediate filament family members are elongated proteins with a conserved central  $\alpha$ -helical domain containing 40 or so heptad repeat motifs that form an extended coiled-coil structure with another monomer (see Figure 3–9). A pair of parallel dimers then associates in an antiparallel fashion to form a staggered tetramer (Figure 16–67). Unlike actin or tubulin subunits, intermediate filament subunits do not contain a binding site for a nucleotide. Furthermore, since the tetrameric subunit is made up of two dimers pointing in opposite directions, its two ends are the same. The assembled intermediate filament therefore lacks the overall structural polarity that is critical for actin filaments and microtubules. The tetramers pack together laterally to form the filament, which includes eight parallel protofilaments made up of tetramers. Each individual intermediate filament therefore has a cross section of 32 individual  $\alpha$ -helical coils. This large number of polypeptides all lined up together, with the strong lateral hydrophobic interactions typical of coiled-coil proteins, gives intermediate filaments a ropelike character. They can be easily bent, with a persistence length of less than one micrometer (compared

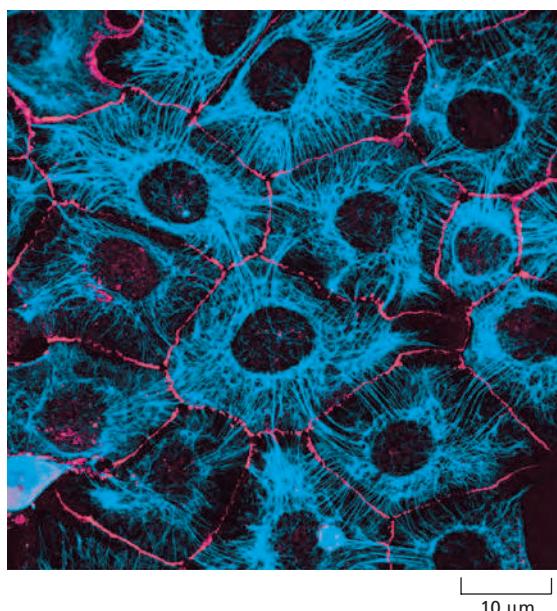
to several millimeters for microtubules and about ten micrometers for actin), but they are extremely difficult to break and can be stretched to over three times their length (see Figure 16–6).

Less is understood about the mechanism of assembly and disassembly of intermediate filaments than of actin filaments and microtubules. In pure protein solutions, intermediate filaments are extremely stable due to tight association of subunits, but some types of intermediate filaments, including *vimentin*, form highly dynamic structures in cells such as fibroblasts. Protein phosphorylation probably regulates their disassembly, in much the same way that phosphorylation regulates the disassembly of nuclear lamins in mitosis (see Figure 12–18). As evidence for rapid turnover, labeled subunits microinjected into tissue-culture cells incorporate into intermediate filaments within a few minutes. Remodeling of the intermediate filament network accompanies events requiring dynamic cellular reorganization, such as division, migration, and differentiation.

### Intermediate Filaments Impart Mechanical Stability to Animal Cells

**Keratins** are the most diverse intermediate filament family: there are about 20 found in different types of human epithelial cells and about 10 more that are specific to hair and nails; analysis of the human genome sequence has revealed that there are 54 distinct keratins. Every keratin filament is made up of an equal mixture of type I (acidic) and type II (neutral/basic) keratin proteins; these form a heterodimer filament subunit (see Figure 16–67). Cross-linked keratin networks held together by disulfide bonds can survive even the death of their cells, forming tough coverings for animals, as in the outer layer of skin and in hair, nails, claws, and scales. The diversity in keratins is clinically useful in the diagnosis of epithelial cancers (carcinomas), as the particular set of keratins expressed gives an indication of the epithelial tissue in which the cancer originated and thus can help to guide the choice of treatment.

A single epithelial cell may produce multiple types of keratins, and these copolymerize into a single network (Figure 16–68). Keratin filaments impart mechanical strength to epithelial tissues in part by anchoring the intermediate filaments at sites of cell-cell contact, called *desmosomes*, or cell-matrix contact, called *hemidesmosomes* (see Figure 16–4). We discuss these important adhesive structures in Chapter 19. Accessory proteins, such as *filaggrin*, bundle keratin filaments in differentiating cells of the epidermis to give the outermost layers of the



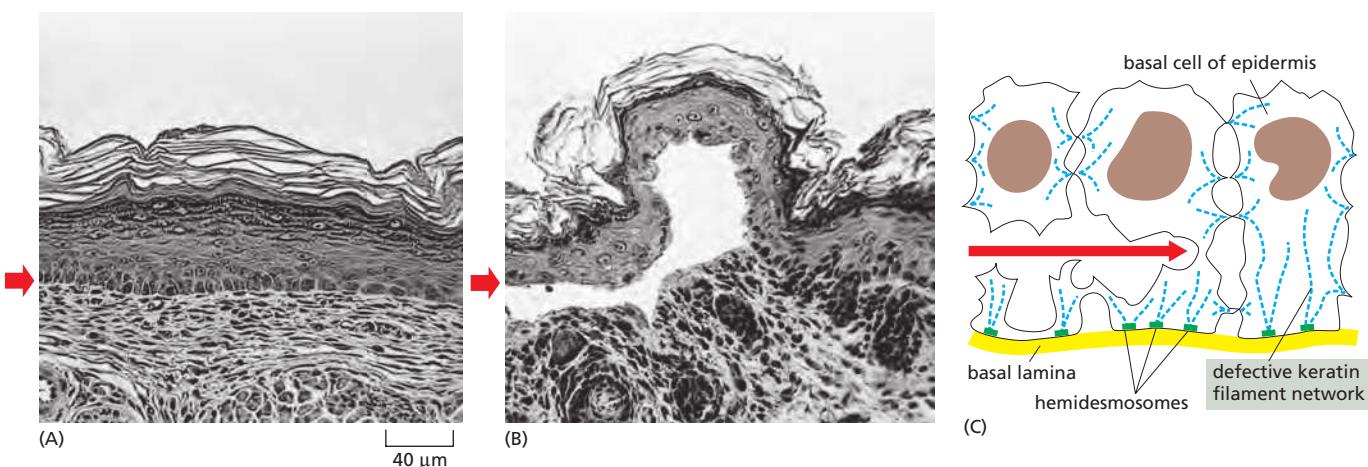
**Figure 16–68 Keratin filaments in epithelial cells.** Immunofluorescence micrograph of the network of keratin filaments (blue) in a sheet of epithelial cells in culture. The filaments in each cell are indirectly connected to those of its neighbors by desmosomes (discussed in Chapter 19). A second protein (red) has been stained to reveal the location of the cell boundaries. (Courtesy of Kathleen Green and Evangeline Amargo.)

skin their special toughness. Individuals with mutations in the gene encoding filaggrin are strongly predisposed to dry skin diseases such as eczema.

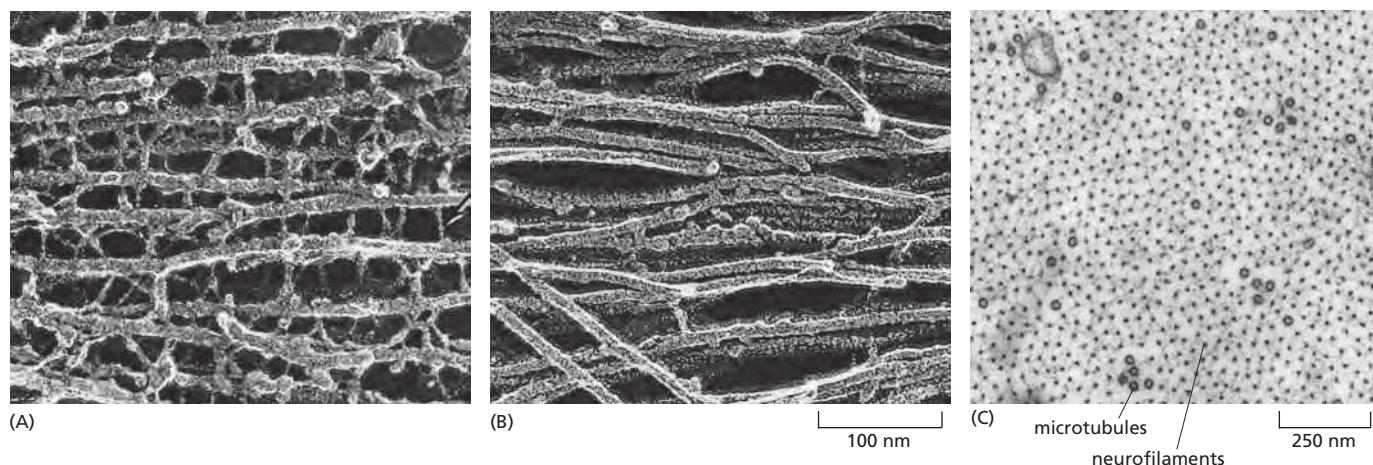
Mutations in keratin genes cause several human genetic diseases. For example, when defective keratins are expressed in the basal cell layer of the epidermis, they produce a disorder called *epidermolysis bullosa simplex*, in which the skin blisters in response to even very slight mechanical stress, which ruptures the basal cells (Figure 16-69). Other types of blistering diseases, including disorders of the mouth, esophageal lining, and the cornea of the eye, are caused by mutations in the different keratins whose expression is specific to those tissues. All of these maladies are typified by cell rupture as a consequence of mechanical trauma and a disorganization or clumping of the keratin filament cytoskeleton. Many of the specific mutations that cause these diseases alter the ends of the central rod domain, demonstrating the importance of this particular part of the protein for correct filament assembly.

Members of another family of intermediate filaments, called **neurofilaments**, are found in high concentrations along the axons of vertebrate neurons (Figure 16-70). Three types of neurofilament proteins (NF-L, NF-M, and NF-H) coassemble *in vivo*, forming heteropolymers. The NF-H and NF-M proteins have lengthy C-terminal tail domains that bind to neighboring filaments, generating aligned arrays with a uniform interfilament spacing. During axonal growth, new neurofilament subunits are incorporated all along the axon in a dynamic process that involves the addition of subunits along the filament length as well as the ends. After an axon has grown and connected with its target cell, the diameter of the axon may increase as much as fivefold. The level of neurofilament gene expression seems to directly control axonal diameter, which in turn influences how fast electrical signals travel down the axon. In addition, neurofilaments provide strength and stability to the long cell processes of neurons.

The neurodegenerative disease amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease) is associated with an accumulation and abnormal assembly of neurofilaments in motor neuron cell bodies and in the axon, aberrations that may interfere with normal axonal transport. The degeneration of the axons leads to muscle weakness and atrophy, which is usually fatal. The overexpression of human NF-L or NF-H in mice results in mice that have an ALS-like disease. However, a causative link between neurofilament pathology and ALS has not been firmly established.



**Figure 16-69 Blistering of the skin caused by a mutant keratin gene.** A mutant gene encoding a truncated keratin protein (lacking both the N- and C-terminal domains) was expressed in a transgenic mouse. The defective protein assembles with the normal keratins and thereby disrupts the keratin filament network in the basal cells of the skin. Light micrographs of cross sections of (A) normal and (B) mutant skin show that the blistering results from the rupturing of cells in the basal layer of the mutant epidermis (short red arrows). (C) A sketch of three cells in the basal layer of the mutant epidermis, as observed by electron microscopy. As indicated by the red arrow, the cells rupture between the nucleus and the hemidesmosomes (discussed in Chapter 19), which connect the keratin filaments to the underlying basal lamina. (From P.A. Coulombe et al., *J. Cell Biol.* 115:1661–1674, 1991. With permission from The Rockefeller University Press.)



The vimentin-like filaments are a third family of intermediate filaments. *Desmin*, a member of this family, is expressed in skeletal, cardiac, and smooth muscle, where it forms a scaffold around the Z disc of the sarcomere (see Figure 16–34). Mice lacking desmin show normal initial muscle development, but adults have various muscle-cell abnormalities, including misaligned muscle fibers. In humans, mutations in desmin are associated with various forms of muscular dystrophy and cardiac myopathy, illustrating the important role of desmin in stabilizing muscle fibers.

Besides their well-established role in maintaining the mechanical stability of the nucleus, it is becoming increasingly evident that one class of lamins, the A-type, together with many proteins of the nuclear envelope, are scaffolds for proteins that control myriad cellular processes including transcription, chromatin organization, and signal transduction. The majority of *laminopathies* are associated with mutant versions of lamin A and include tissue-specific diseases. Skeletal and cardiac abnormalities might be explained by a weakened nuclear envelope leading to cell damage and death, but laminopathies are also thought to arise from pathogenic and tissue-specific alterations in gene expression.

### Linker Proteins Connect Cytoskeletal Filaments and Bridge the Nuclear Envelope

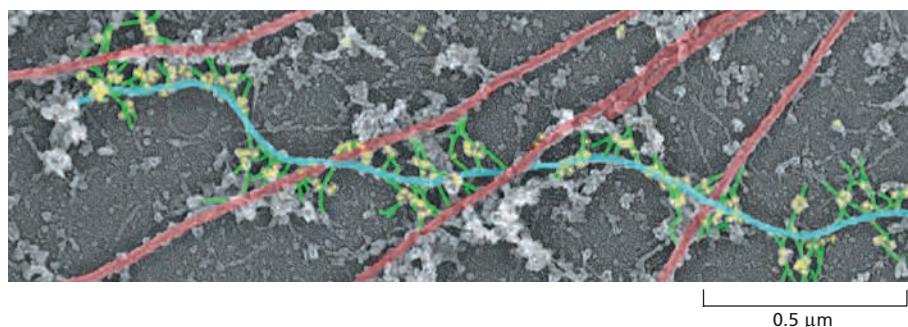
The intermediate filament network is linked to the rest of the cytoskeleton by members of a family of proteins called *plakins*. Plakins are large and modular, containing multiple domains that connect cytoskeletal filaments to each other and to junctional complexes. *Plectin* is a particularly interesting example. In addition to bundling intermediate filaments, it links the intermediate filaments to microtubules, actin filament bundles, and filaments of the motor protein myosin II; it also helps attach intermediate filament bundles to adhesive structures at the plasma membrane (Figure 16–71).

Plectin and other plakins can interact with protein complexes that connect the cytoskeleton to the nuclear interior. These complexes consist of SUN proteins

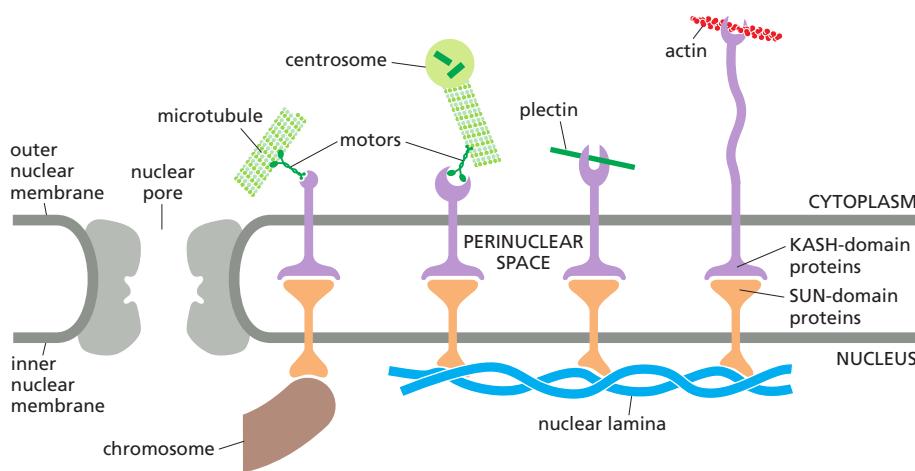
**Figure 16–70** Two types of intermediate filaments in cells of the nervous system.

(A) Freeze-etch electron microscopic image of neurofilaments in a nerve cell axon, showing the extensive cross-linking through protein cross-bridges—an arrangement believed to give this long cell process great tensile strength. The cross-bridges are formed by the long, nonhelical extensions at the C-terminus of the largest neurofilament protein (NF-H). (B) Freeze-etch image of glial filaments in glial cells, showing that these intermediate filaments are smooth and have few cross-bridges. (C) Conventional transmission electron micrograph of a cross section of an axon showing the regular side-to-side spacing of the neurofilaments, which greatly outnumber the microtubules.

(A and B, courtesy of Nobutaka Hirokawa; C, courtesy of John Hopkins.)



**Figure 16–71** Plectin cross-linking of diverse cytoskeletal elements. Plectin (green) is seen here making cross-links from intermediate filaments (blue) to microtubules (red). In this electron micrograph, the dots (yellow) are gold particles linked to anti-plectin antibodies. The entire actin filament network was removed to reveal these proteins. (From T.M. Svitkina et al., J. Cell Biol. 135:991–1007, 1996. With permission from The Rockefeller University Press.)



**Figure 16–72** SUN-KASH protein complexes connect the nucleus and cytoplasm through the nuclear envelope. The cytoplasmic cytoskeleton is linked across the nuclear envelope to the nuclear lamina or chromosomes through SUN and KASH proteins (orange and purple, respectively). The SUN and KASH domains of these proteins bind within the lumen of the nuclear envelope. From the inner nuclear envelope, SUN proteins connect to the nuclear lamina or chromosomes. KASH proteins in the outer nuclear envelope connect to the cytoplasmic cytoskeleton by binding microtubule motor proteins, actin filaments, or plectin.

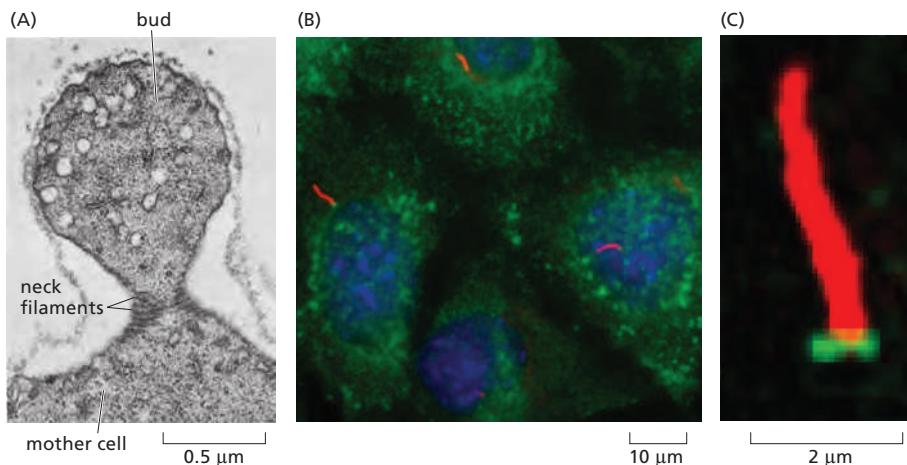
of the inner nuclear membrane and KASH proteins (also called nesprins) of the outer nuclear membrane (**Figure 16–72**). SUN and KASH proteins bind to each other within the lumen of the nuclear envelope, forming a bridge that connects the nuclear and cytoplasmic cytoskeletons. Inside the nucleus, the SUN proteins bind to the nuclear lamina or chromosomes, whereas in the cytoplasm, KASH proteins can bind directly to actin filaments and indirectly to microtubules and intermediate filaments through association with motor proteins and plakins, respectively. This linkage serves to mechanically couple the nucleus to the cytoskeleton and is involved in many cellular functions, including chromosome movements inside the nucleus during meiosis, nuclear and centrosome positioning, nuclear migration, and global cytoskeletal organization.

Mutations in the gene for plectin cause a devastating human disease that combines epidermolysis bullosa (caused by disruption of skin keratin filaments), muscular dystrophy (caused by disruption of desmin filaments), and neurodegeneration (caused by disruption of neurofilaments). Mice lacking a functional plectin gene die within a few days of birth, with blistered skin and abnormal skeletal and heart muscles. Thus, although plectin may not be necessary for the initial formation and assembly of intermediate filaments, its cross-linking action is required to provide cells with the strength they need to withstand the mechanical stresses inherent to vertebrate life.

### Septins Form Filaments That Regulate Cell Polarity

GTP-binding proteins called *septins* serve as an additional filament system in all eukaryotes except terrestrial plants. Septins assemble into nonpolar filaments that form rings and cagelike structures, which act as scaffolds to compartmentalize membranes into distinct domains, or recruit and organize the actin and microtubule cytoskeletons. First identified in budding yeast, septin filaments localize to the neck between a dividing yeast mother cell and its growing bud (**Figure 16–73A**). At this location, septins block the movement of proteins from one side of the bud neck to the other, thereby concentrating cell growth preferentially within the bud. Septins also recruit the actin-myosin machinery that forms the contractile ring required for cytokinesis. In animal cells, septins function in cell division, migration, and vesicle trafficking. In primary cilia, for example, a ring of septin filaments assembles at the base of the cilium and serves as a diffusion barrier at the plasma membrane, restricting the movement of membrane proteins and establishing a specific composition in the ciliary membrane (Figure 16–73B and C). Reduction of septin levels impairs primary cilium formation and signaling.

There are 7 septin genes in yeast and 13 in human, and septin proteins fall into four groups on the basis of sequence relationships. In a test tube, purified septins assemble into symmetrical hetero-hexamers or hetero-octamers that

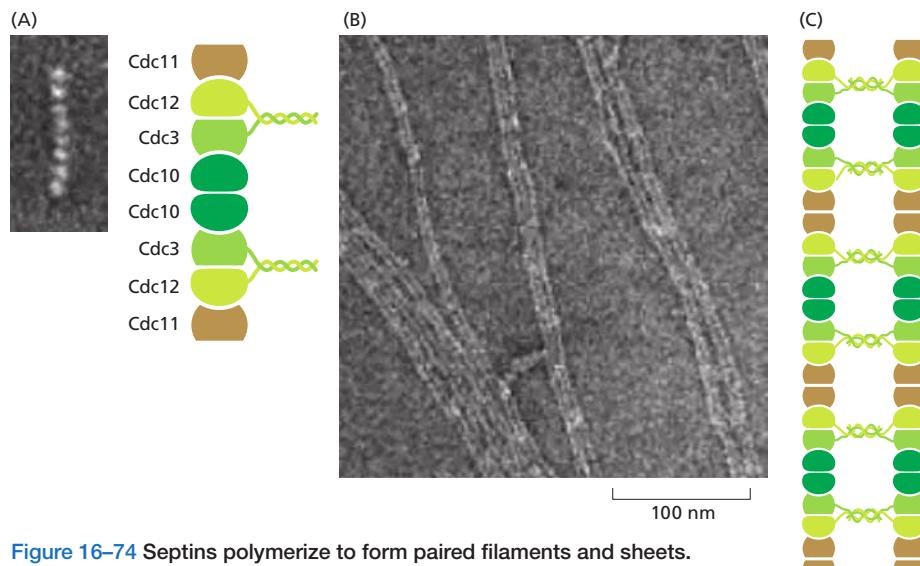


**Figure 16–73 Cell compartmentalization by septins.** (A) Septins form filaments in the neck region between a mother yeast cell and bud. (B) In this photomicrograph of human cultured cells, the DNA is stained blue and septins are labeled in green. The microtubules of primary cilia are labeled with an antibody that recognizes a modified (acetylated) form of tubulin (red) that is enriched in the axoneme. (C) A magnified image reveals a collar of septin at the base of the cilium. (A, from B. Byers and L. Goetsch, *J. Cell Biol.* 69:717–721, 1976. With permission from Rockefeller University Press. B and C, from Q. Hu et al., *Science* 329:436–439, 2010. With permission from AAAS.)

form nonpolar paired filaments (**Figure 16–74**). GTP binding is required for the folding of septin polypeptides, but the role of GTP hydrolysis in septin function is not understood. Septin structures assemble and disassemble inside cells, but they are not as dynamic as actin filaments and microtubules.

### Summary

Whereas tubulin and actin have been highly conserved in evolution, intermediate filament proteins are very diverse. There are many tissue-specific forms of intermediate filaments in the cytoplasm of animal cells, including keratin filaments in epithelial cells, neurofilaments in nerve cells, and desmin filaments in muscle cells. The primary function of these filaments is to provide mechanical strength. Septins comprise an additional system of filaments that organize compartments inside cells.



**Figure 16–74 Septins polymerize to form paired filaments and sheets.** (A) Electron micrograph of a septin rod assembled by combining two copies each of the four yeast septins illustrated at the right. The eight-subunit rod is nonpolar because the central pair of subunits (Cdc10) creates a symmetrical dimer. (B) Electron micrograph of paired septin filaments and sheets, assembled from purified septins in the presence of high salt concentrations. (C) Paired septin filaments may assemble by lateral association between filaments, mediated by coiled-coils formed between the paired C-terminal extensions of Cdc3 and Cdc12 that project from each filament. (Images and schematics adapted from A. Bertin et al., *Proc. Natl. Acad. Sci. USA* 105:8274–8279, 2008. With permission from the National Academy of Sciences.)

## CELL POLARIZATION AND MIGRATION

A central challenge in cell biology is to understand how multiple individual molecular components collaborate to produce complex cell behaviors. The process of cell migration, which we describe in this final section, relies on the coordinated deployment of the components and processes that we have explored in this chapter: the dynamic assembly and disassembly of cytoskeletal polymers, the regulation and modification of their structure by polymer-associated proteins, and the actions of motor proteins moving along the polymers or exerting tension against them. How does the cell coordinate all these activities to define its polarity and enable it to crawl?

### Many Cells Can Crawl Across a Solid Substratum

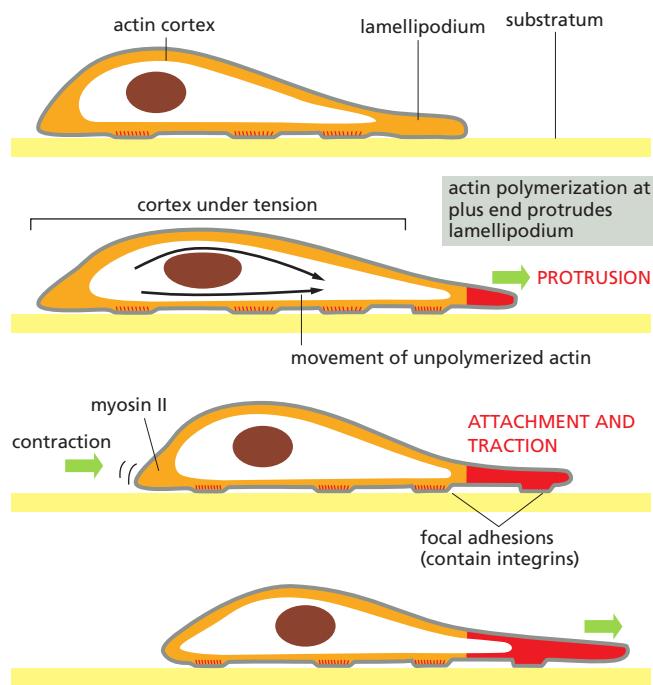
Many cells move by crawling over surfaces rather than by using cilia or flagella to swim. Predatory amoebae crawl continuously in search of food, and they can easily be observed to attack and devour smaller ciliates and flagellates in a drop of pond water (see Movie 1.4). In animals, almost all cell locomotion occurs by crawling, with the notable exception of swimming sperm. During embryogenesis, the structure of an animal is created by the migrations of individual cells to specific target locations and by the coordinated movements of whole epithelial sheets (discussed in Chapter 21). In vertebrates, *neural crest cells* are remarkable for their long-distance migrations from their site of origin in the neural tube to a variety of sites throughout the embryo (see Movie 21.5). Long-distance crawling is fundamental to the construction of the entire nervous system: it is in this way that the actin-rich growth cones at the advancing tips of developing axons travel to their eventual synaptic targets, guided by combinations of soluble signals and signals bound to cell surfaces and extracellular matrix along the way.

The adult animal also seethes with crawling cells. Macrophages and neutrophils crawl to sites of infection and engulf foreign invaders as a critical part of the innate immune response. Osteoclasts tunnel into bone, forming channels that are filled in by the osteoblasts that follow after them, in a continuous process of bone remodeling and renewal. Similarly, fibroblasts migrate through connective tissues, remodeling them where necessary and helping to rebuild damaged structures at sites of injury. In an ordered procession, the cells in the epithelial lining of the intestine travel up the sides of the intestinal villi, replacing absorptive cells lost at the tip of the villus. Unfortunately, cell crawling also has a role in many cancers, when cells in a primary tumor invade neighboring tissues and crawl into blood vessels or lymph vessels and then emerge at other sites in the body to form metastases.

Cell migration is a complex process that depends on the actin-rich cortex beneath the plasma membrane. Three distinct activities are involved: *protrusion*, in which the plasma membrane is pushed out at the front of the cell; *attachment*, in which the actin cytoskeleton connects across the plasma membrane to the substratum; and *traction*, in which the bulk of the trailing cytoplasm is drawn forward (Figure 16–75). In some crawling cells, such as keratocytes from the fish epidermis, these activities occur simultaneously, and the cells seem to glide forward smoothly without changing shape. In other cells, such as fibroblasts, these activities are more independent, and the locomotion is jerky and irregular.

### Actin Polymerization Drives Plasma Membrane Protrusion

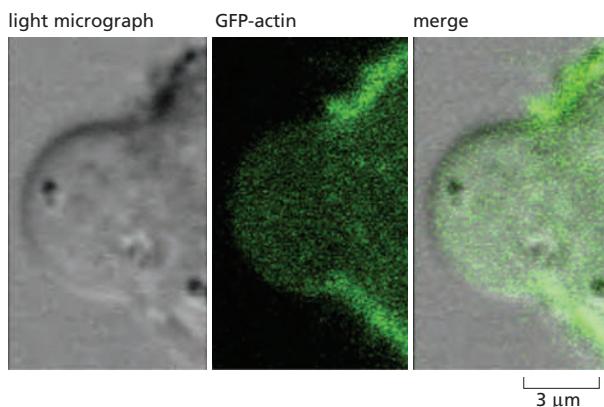
The first step in locomotion, protrusion of a leading edge, frequently relies on forces generated by actin polymerization pushing the plasma membrane outward. Different cell types generate different types of protrusive structures, including filopodia (also known as microspikes) and lamellipodia. These are filled with dense cores of filamentous actin, which excludes membrane-enclosed organelles. The structures differ primarily in the way in which the actin is organized by actin-cross-linking proteins (see Figure 16–22).



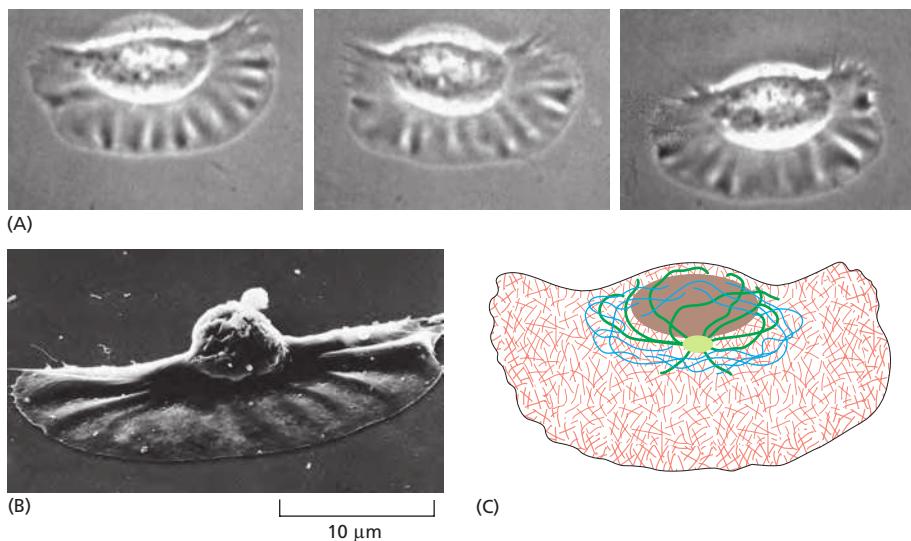
**Figure 16–75** A model of how forces generated in the actin-rich cortex move a cell forward. The actin-polymerization-dependent protrusion and firm attachment of a lamellipodium at the leading edge of the cell move the edge forward (green arrows at front) and stretch the actin cortex. Contraction at the rear of the cell propels the body of the cell forward (green arrow at back) to relax some of the tension (traction). New focal contacts are made at the front, and old ones are disassembled at the back as the cell crawls forward. The same cycle can be repeated, moving the cell forward in a stepwise fashion. Alternatively, all steps can be tightly coordinated, moving the cell forward smoothly. The newly polymerized cortical actin is shown in red.

**Filopodia**, formed by migrating growth cones of neurons and some types of fibroblasts, are essentially one-dimensional. They contain a core of long, bundled actin filaments, which are reminiscent of those in microvilli but longer and thinner, as well as more dynamic. **Lamellipodia**, formed by epithelial cells and fibroblasts, as well as by some neurons, are two-dimensional, sheetlike structures. They contain a cross-linked mesh of actin filaments, most of which lie in a plane parallel to the solid substratum. **Invadopodia** and related structures known as podosomes represent a third type of actin-rich protrusion. These extend in three dimensions and are important for cells to cross tissue barriers, as when a metastatic cancer cell invades the surrounding tissue. Invadopodia contain many of the same actin-regulatory components as filopodia and lamellipodia, and they also degrade the extracellular matrix, which requires the delivery of vesicles containing matrix-degrading proteases.

A distinct form of membrane protrusion called **blebbing** is often observed *in vivo* or when cells are cultured on a pliable extracellular matrix substratum. Blebs form when the plasma membrane detaches locally from the underlying actin cortex, thereby allowing cytoplasmic flow to push the membrane outward (Figure 16–76). Bleb formation also depends on hydrostatic pressure within the cell, which is generated by the contraction of actin and myosin assemblies. Once blebs have extended, actin filaments reassemble on the bleb membrane to form a new



**Figure 16–76** Membrane bleb induced by disruption of the actin cortex. On the left is a light micrograph showing a spherical membrane protrusion or bleb induced by laser ablation of a small region of the actin cortex. The cortex is labeled green in the middle image by expression of GFP-actin. (Courtesy of Ewa Paluch.)



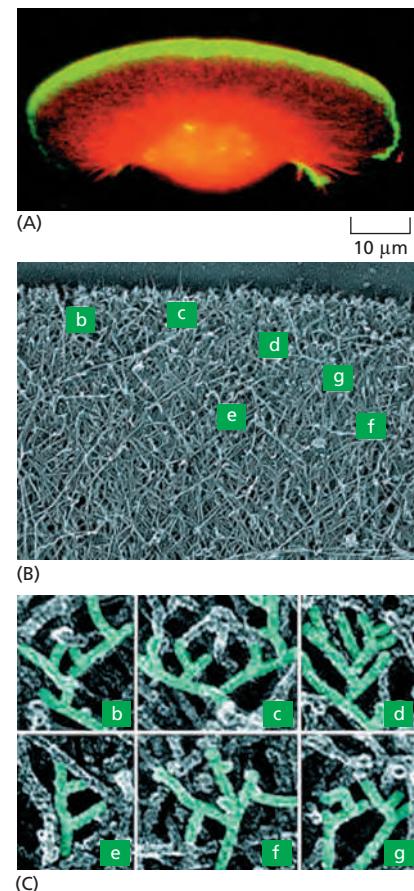
**Figure 16-77** Migratory keratocytes from a fish epidermis. (A) Light micrographs of a keratocyte in culture, taken about 15 seconds apart. This cell is moving at about 15  $\mu\text{m}/\text{min}$  ([Movie 16.13](#) and see Movie 1.1). (B) Keratocyte seen by scanning electron microscopy, showing its broad, flat lamellipodium and small cell body, including the nucleus, carried up above the substratum at the rear. (C) Distribution of cytoskeletal filaments in this cell. Actin filaments (red) fill the large lamellipodium and are responsible for the cell's rapid movement. Microtubules (green) and intermediate filaments (blue) are restricted to the regions close to the nucleus. (A and B, courtesy of Juliet Lee.)

actin cortex. Recruitment of myosin II and contraction of actin and myosin can then power retraction of membrane blebs. Alternatively, extension of new blebs from old ones can drive cell migration.

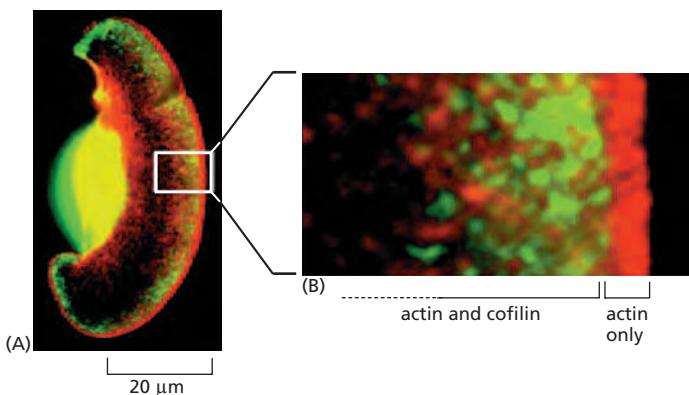
### Lamellipodia Contain All of the Machinery Required for Cell Motility

Lamellipodia have been particularly well studied in the epithelial cells of the epidermis of fish and frogs; these epithelial cells are known as *keratocytes* because of their abundant keratin filaments. These cells normally cover the animal by forming an epithelial sheet, and they are specialized to close wounds very rapidly, moving at rates of up to 30  $\mu\text{m}/\text{min}$ . When cultured as individual cells, keratocytes assume a distinctive shape with a very large lamellipodium and a small, trailing cell body that is not attached to the substratum ([Figure 16-77](#)). Fragments of this lamellipodium can be sliced off with a micropipette. Although the fragments generally lack microtubules and membrane-enclosed organelles, they continue to crawl normally, looking like tiny keratocytes.

The dynamic behavior of actin filaments in keratocyte lamellipodia can be studied by labeling a small patch of actin and examining its fate. This reveals that, while the lamellipodia crawl forward, the actin filaments remain stationary with respect to the substratum. The actin filaments in the meshwork are mostly oriented with their plus ends facing forward. The minus ends are frequently attached to the sides of other actin filaments by Arp 2/3 complexes (see Figure 16-16), helping to form the two-dimensional web ([Figure 16-78](#)). The web as a whole is undergoing treadmilling, assembling at the front and disassembling at the back, reminiscent of the treadmilling that occurs in individual actin filaments discussed previously (see Figure 16-14).



**Figure 16-78** Actin filament nucleation and web formation by the Arp 2/3 complex in lamellipodia. (A) A keratocyte with actin filaments labeled in red by fluorescent phalloidin and the Arp 2/3 complex labeled in green with an antibody against one of its subunits. The Arp 2/3 complex is highly concentrated near the front of the lamellipodium, where actin nucleation is most active. (B) Electron micrograph of a platinum-shadowed replica of the leading edge of a keratocyte, showing the dense actin filament meshwork. The labels denote areas enlarged in (C). (C) Close-up views of the marked regions of the actin web at the leading edge shown in (B). Numerous branched filaments can be seen, with the characteristic 70° angle formed when the Arp 2/3 complex nucleates a new actin filament off the side of a preexisting filament (see Figure 16-16). (From T. Svitkina and G. Borisy, *J. Cell Biol.* 145:1009–1026, 1999. With permission from the authors.)



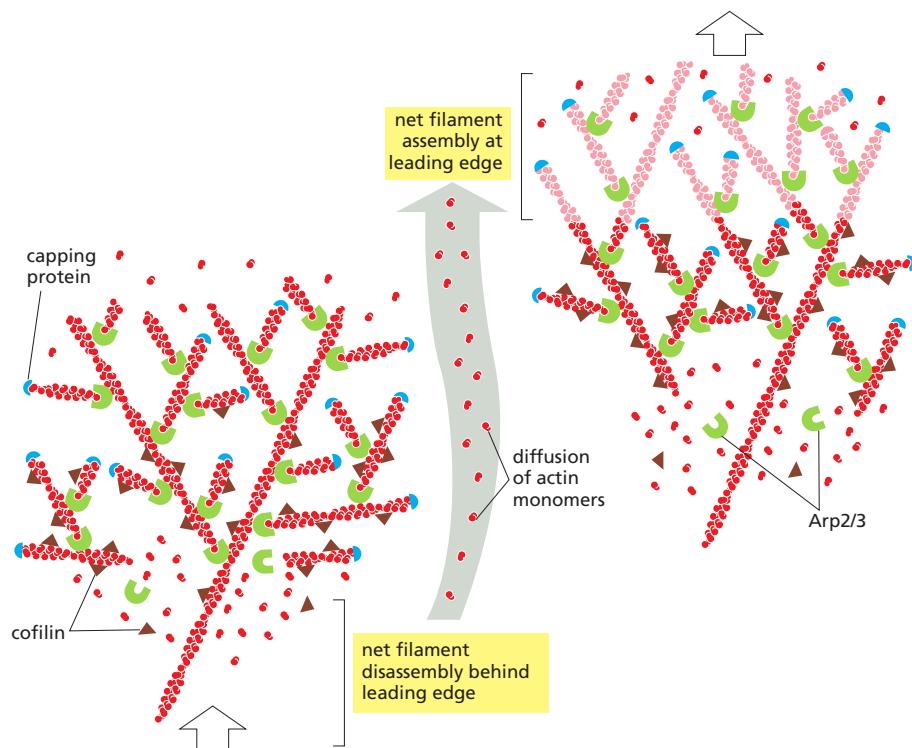
**Figure 16-79 Cofilin in lamellipodia.**

(A) A keratocyte with actin filaments labeled in red by fluorescent phalloidin, and cofilin labeled in green with a fluorescent antibody. Although the dense actin meshwork reaches all the way through the lamellipodium, cofilin is not found at the very leading edge. (B) Close-up view of the region marked with the white rectangle in (A). The actin filaments closest to the leading edge, which are also the ones that have formed most recently and that are most likely to contain ATP-actin (rather than ADP-actin), are generally not associated with cofilin. (From T. Svitkina and G. Borisy, *J. Cell Biol.* 145:1009–1026, 1999. With permission from the authors.)

Maintenance of unidirectional motion by lamellipodia is thought to require the cooperation and mechanical integration of several factors. Filament nucleation is localized at the leading edge, with new actin filament growth occurring primarily in that location to push the plasma membrane forward. Most filament depolymerization occurs at sites located well behind the leading edge. Because *cofilin* (see Figure 16-20) binds cooperatively and preferentially to actin filaments containing ADP-actin (the D form), the new T-form filaments generated at the leading edge should be resistant to depolymerization by cofilin (Figure 16-79). As the filaments age and ATP hydrolysis proceeds, cofilin can efficiently disassemble the older filaments. Thus, the delayed ATP hydrolysis by filamentous actin is thought to provide the basis for a mechanism that maintains an efficient, unidirectional treadmilling process in the lamellipodium (Figure 16-80); it also explains the intracellular movement of bacterial pathogens such as *Listeria* (see Figure 16-25).

### Myosin Contraction and Cell Adhesion Allow Cells to Pull Themselves Forward

Forces generated by actin filament polymerization at the front of a migrating cell are transmitted to the underlying substratum to drive cell motion. For the leading



**Figure 16-80 A model for protrusion of the actin meshwork at the leading edge.**

Two time points during advance of the lamellipodium are illustrated, with newly assembled structures at the later time point shown in a lighter color. Nucleation is mediated by the Arp 2/3 complex at the front. Newly nucleated actin filaments are attached to the sides of preexisting filaments, primarily at a 70° angle. Filaments elongate, pushing the plasma membrane forward because of some sort of anchorage of the array behind. At a steady rate, actin filament plus ends become capped. After newly polymerized actin subunits hydrolyze their bound ATP in the filament lattice, the filaments become susceptible to depolymerization by cofilin. This cycle causes a spatial separation between net filament assembly at the front and net filament disassembly at the rear, so that the actin filament network as a whole can move forward, even though the individual filaments within it remain stationary with respect to the substratum. Not all of the actin disassembles, however, and actin at the rear of the lamellipodium contributes to subsequent steps of migration together with myosin.

**Figure 16–81 Contribution of myosin II to polarized cell motility.**

(A) Myosin II bipolar filaments bind to actin filaments in the lamellipodial meshwork and cause network contraction. The myosin-driven reorientation of the actin filaments forms an actin bundle that recruits more myosin II and helps generate the contractile forces required for retraction of the trailing edge of the moving cell. (B) A fragment of the large lamellipodium of a keratocyte can be separated from the main cell body either by surgery with a micropipette or by treating the cell with certain drugs. Many of these fragments continue to move rapidly, with the same overall cytoskeletal organization as the intact keratocytes. Actin (blue) forms a protrusive meshwork at the front of the fragment. Myosin II (pink) is gathered into a band at the rear. (From A. Verkhovsky et al., *Curr. Biol.* 9:11–20, 1999. With permission from Elsevier.)

edge of a migrating cell to advance, protrusion of the membrane must be followed by adhesion to the substratum at the front. Conversely, in order for the cell body to follow, contraction must be coupled with de-adhesion at the rear of the cell. The processes contributing to migration are therefore tightly regulated in space and time, with actin polymerization, dynamic adhesions, and myosin contraction being employed to coordinate movement. Myosin II operates in at least two ways to assist cell migration. The first is by helping to connect the actin cytoskeleton to the substratum through integrin-mediated adhesions. Forces generated by both actin polymerization and myosin activity create tension at attachment sites, promoting their maturation into *focal adhesions*, which are dynamic assemblies of structural and signaling proteins that link the migrating cell to the extracellular matrix (see Figure 19–59). A second mechanism involves bipolar myosin II filaments, which associate with the actin filaments at the rear of the lamellipodium and pull them into a new orientation—from nearly perpendicular to the leading edge to almost parallel to the leading edge. This sarcomere-like contraction prevents protrusion, and it pinches in the sides of the locomoting lamellipodium, helping to gather in the sides of the cell as it moves forward (**Figure 16–81**).

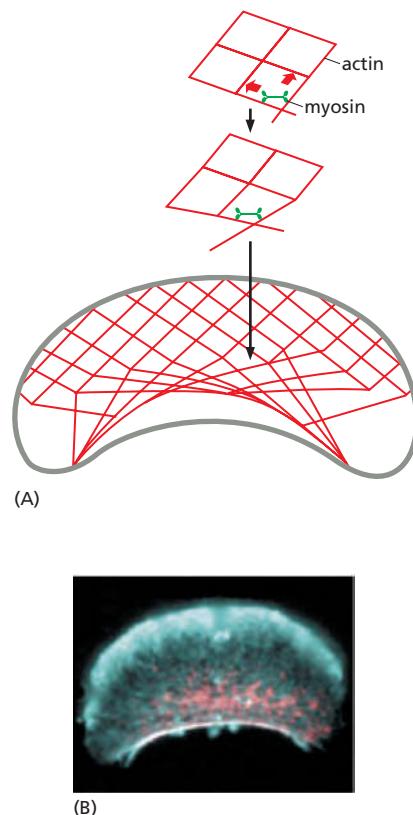
Actin-mediated protrusions can only push the leading edge of the cell forward if there are strong interactions between the actin network and the focal adhesions that link the cell to the substrate. When these interactions are disengaged, polymerization pressure at the leading edge and myosin-dependent contraction cause the actin network to slip back, resulting in a phenomenon known as retrograde flow (**Figure 16–82**).

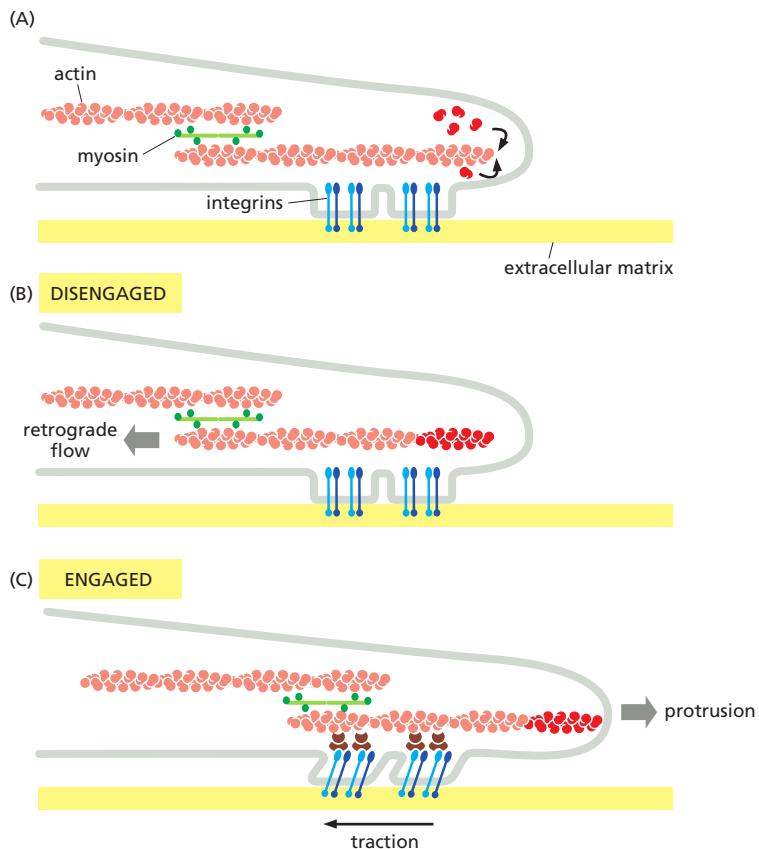
The traction forces generated by locomoting cells exert a significant pull on the substratum. By growing cells on a surface coated with tiny flexible posts, the force exerted on the substratum can be calculated by measuring the deflection of each post from its vertical position (**Figure 16–83**). In a living animal, most crawling cells move across a semiflexible substratum made of extracellular matrix, which can be deformed and rearranged by these cell forces. Conversely, mechanical tension or stretching applied externally to a cell will cause it to assemble stress fibers and focal adhesions, and become more contractile. Although poorly understood, this two-way mechanical interaction between cells and their physical environment is thought to help vertebrate tissues organize themselves.

### Cell Polarization Is Controlled by Members of the Rho Protein Family

Cell migration requires long-distance communication and coordination between one end of a cell and the other. During directed migration, it is important that the front end of the cell remain structurally and functionally distinct from the back end. In addition to driving local mechanical processes such as protrusion at the front and retraction at the rear, the cytoskeleton is responsible for coordinating cell shape, organization, and mechanical properties from one end of the cell to the other, a distance that is typically tens of micrometers for animal cells.

In many cases, including but not limited to cell migration, large-scale cytoskeletal coordination takes the form of the establishment of cell polarity, where a cell builds different structures with distinct molecular components at the front

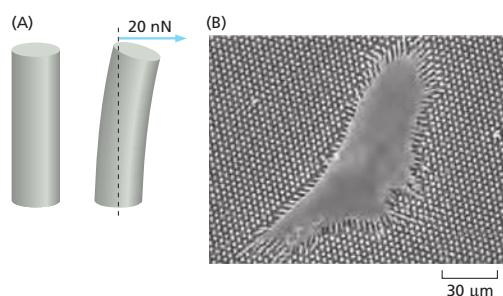




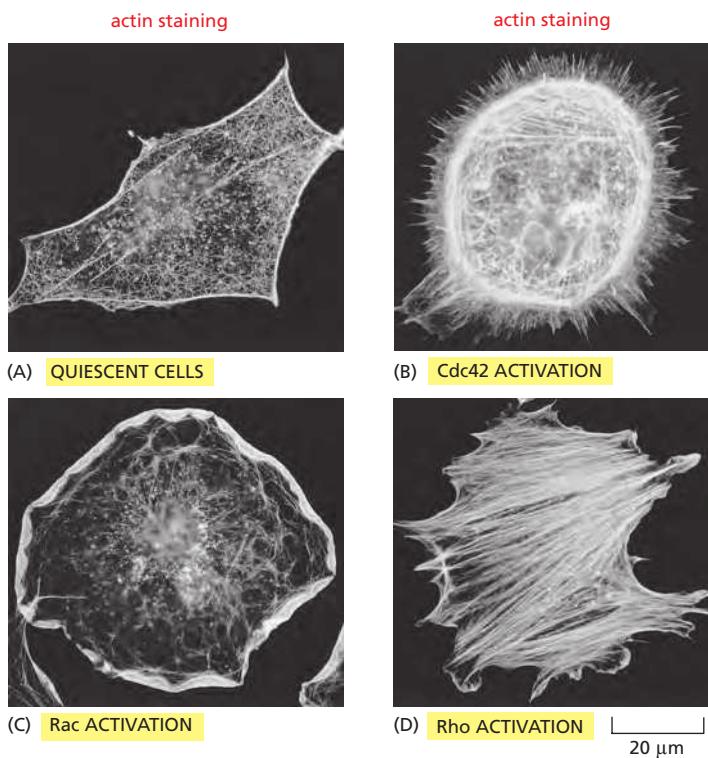
**Figure 16–82** Control of cell–substratum adhesion at the leading edge of a migrating cell. (A) Actin monomers assemble on the barbed end of actin filaments at the leading edge. Transmembrane integrin proteins (blue) help form focal adhesions that link the cell membrane to the substrate. (B) If there is no interaction between the actin filaments and focal adhesions, the actin filament is driven rearward by newly assembled actin. Myosin motors (green) also contribute to filament movement. (C) Interactions between actin-binding adaptor proteins (brown) and integrins link the actin cytoskeleton to the substratum. Myosin-mediated contractile forces are then transmitted through the focal adhesion to generate traction on the extracellular matrix, and new actin polymerization drives the leading edge forward in a protrusion.

versus the back, or at the top versus the bottom. Cell locomotion requires an initial polarization of the cell to set it off in a particular direction. Carefully controlled cell-polarization processes are also required for oriented cell divisions in tissues and for formation of a coherent, organized multicellular structure. Genetic studies in yeast, flies, and worms have provided most of our current understanding of the molecular basis of cell polarity. The mechanisms that generate cell polarity in vertebrates are only beginning to be explored. In all known cases, however, the cytoskeleton has a central role, and many of the molecular components have been evolutionarily conserved.

The establishment of many kinds of cell polarity depends on the local regulation of the actin cytoskeleton by external signals. Many of these signals seem to converge inside the cell on a group of closely related monomeric GTPases that are members of the **Rho protein family**—*Cdc42*, *Rac*, and *Rho*. Like other monomeric GTPases, the Rho proteins act as molecular switches that cycle between an active GTP-bound state and an inactive GDP-bound state (see Figure 3–66). Activation of *Cdc42* on the inner surface of the plasma membrane triggers actin polymerization and bundling to form filopodia. Activation of *Rac* promotes actin polymerization at the cell periphery, leading to the formation of sheetlike lamellipodial extensions. Activation of *Rho* promotes both the bundling of actin filaments with myosin II filaments into stress fibers and the clustering of integrins



**Figure 16–83** Traction forces exerted by a motile cell. (A) Tiny flexible pillars attached to the substratum bend in response to traction forces. (B) Scanning electron micrograph of a cell on a substratum coated with pillars that are 6.1  $\mu\text{m}$  in height. Pillar deflections are used to calculate force vectors corresponding to inward pulling forces on the underlying substratum. (Adapted from J. Fu et al., *Nat. Methods* 7:733–736, 2010. With permission from Macmillan Publishers.)



**Figure 16–84** The dramatic effects of Cdc42, Rac, and Rho on actin organization in fibroblasts. In each case, the actin filaments have been labeled with fluorescent phalloidin. (A) Serum-starved fibroblasts have actin filaments primarily in the cortex, and relatively few stress fibers. (B) Microinjection of a constitutively activated form of Cdc42 causes the protrusion of many long filopodia at the cell periphery. (C) Microinjection of a constitutively activated form of Rac, a closely related monomeric GTPase, causes the formation of an enormous lamellipodium that extends from the entire circumference of the cell. (D) Microinjection of a constitutively activated form of Rho causes the rapid assembly of many prominent stress fibers. (From A. Hall, *Science* 279:509–514, 1998. With permission from AAAS.)

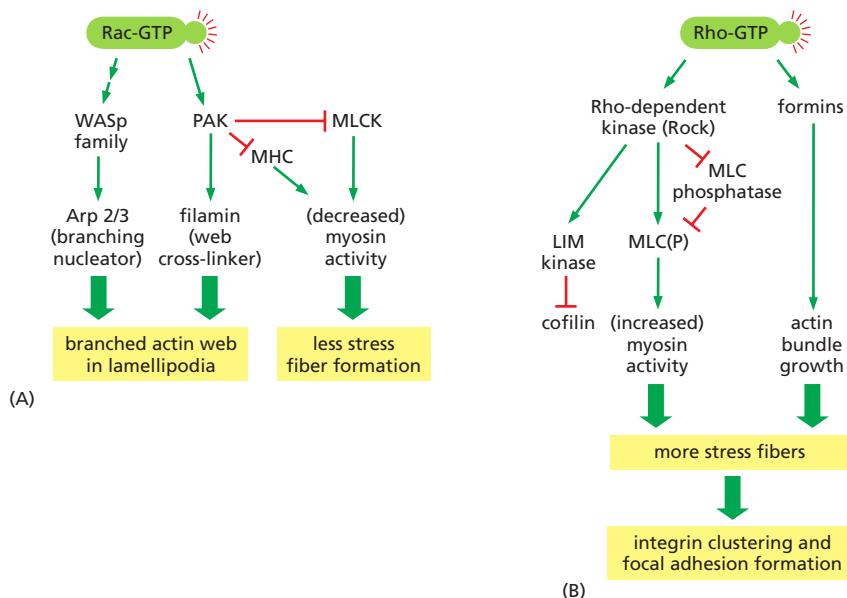
and associated proteins to form focal adhesions (Figure 16–84). These dramatic and complex structural changes occur because each of these three molecular switches has numerous downstream target proteins that affect actin organization and dynamics.

Some key targets of activated Cdc42 are members of the **WASp protein family**. Human patients deficient in WASp suffer from Wiskott-Aldrich Syndrome, a severe form of immunodeficiency in which immune system cells have abnormal actin-based motility and platelets do not form normally. Although WASp itself is expressed only in blood cells and immune system cells, other more ubiquitous versions enable activated Cdc42 to enhance actin polymerization in many cell types. WASp proteins can exist in an inactive folded conformation and an activated open conformation. Association with Cdc42-GTP stabilizes the open form of WASp, enabling it to bind to the Arp 2/3 complex and strongly enhance its actin-nucleating activity (see Figure 16–16). In this way, activation of Cdc42 increases actin nucleation.

Rac-GTP also activates WASp family members. Additionally, it activates the cross-linking activity of the gel-forming protein filamin and inhibits the contractile activity of the motor protein myosin II. It thereby stabilizes lamellipodia and inhibits the formation of contractile stress fibers (Figure 16–85A).

Rho-GTP has a very different set of targets. Instead of activating the Arp 2/3 complex to build actin networks, Rho-GTP turns on formin proteins to construct parallel actin bundles. At the same time, Rho-GTP activates a protein kinase that indirectly inhibits the activity of cofilin, leading to actin filament stabilization. The same protein kinase inhibits a phosphatase acting on myosin light chains (see Figure 16–39). The consequent increase in the net amount of myosin light chain phosphorylation increases the amount of contractile myosin motor protein activity in the cell, enhancing the formation of tension-dependent structures such as stress fibers (Figure 16–85B).

In some cell types, Rac-GTP activates Rho, usually at a rate that is slow compared to Rac's activation of the Arp 2/3 complex. This enables cells to use the Rac pathway to build a new actin structure while subsequently activating the Rho pathway to generate a contractility that builds up tension in this structure. This occurs, for example, during the formation and maturation of cell-cell contacts.



**Figure 16-85** The contrasting effects of Rac and Rho activation on actin organization. (A) Activation of the small GTPase Rac leads to alterations in actin accessory proteins that tend to favor the formation of actin networks, as in lamellipodia. Several different pathways contribute independently. Rac-GTP activates members of the WASp protein family, which in turn activate actin nucleation and branched web formation by the Arp 2/3 complex. In a parallel pathway, Rac-GTP activates a protein kinase, PAK, which has several targets including the web-forming cross-linker filamin, which is activated by phosphorylation, and the myosin light chain kinase (MLCK), which is inhibited by phosphorylation. Inhibition of MLCK results in decreased phosphorylation of the myosin regulatory light chain and leads to myosin II filament disassembly and a decrease in contractile activity. In some cells, PAK also directly inhibits myosin II activity by phosphorylation of the myosin heavy chain (MHC). (B) Activation of the related GTPase Rho leads to nucleation of actin filaments by formins and increases contraction by myosin II, promoting the formation of contractile actin bundles such as stress fibers. Activation of myosin II by Rho requires a Rho-dependent protein kinase called Rock. This kinase inhibits the phosphatase that removes the activating phosphate groups from myosin II light chains (MLC); it may also directly phosphorylate the myosin light chains in some cell types. Rock also activates other protein kinases, such as LIM kinase, which in turn contributes to the formation of stable contractile actin filament bundles by inhibiting the actin depolymerizing factor cofilin. A similar signaling pathway is important for forming the contractile ring necessary for cytokinesis (see Figure 17-44).

As we will explore in more detail below, the communication between the Rac and Rho pathways also facilitates maintenance of the large-scale differences between the cell front and the cell rear during migration.

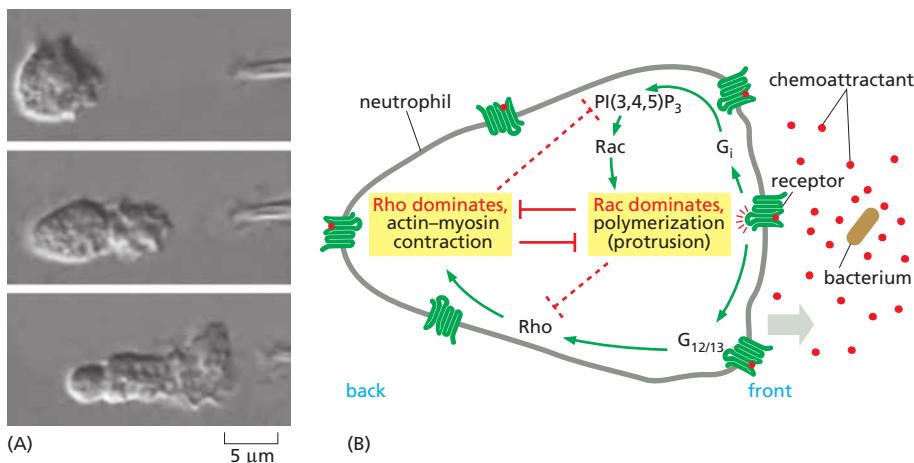
### Extracellular Signals Can Activate the Three Rho Protein Family Members

The activation of the monomeric GTPases Rho, Rac, and Cdc42 occurs through an exchange of GTP for a tightly bound GDP molecule, catalyzed by guanine nucleotide exchange factors (GEFs). Of the many GEFs that have been identified in the human genome, some are specific for an individual Rho family GTPase, whereas others seem to act on multiple family members. Different GEFs are restricted to specific tissues and even specific subcellular locations, and they are sensitive to distinct kinds of regulatory inputs. GEFs can be activated by extracellular cues through cell-surface receptors, or in response to intracellular signals. GEFs may also act as scaffolds that direct GTPases to downstream effectors. Interestingly, several of the Rho family GEFs associate with the growing ends of microtubules by binding to one of the +TIPs. This provides a connection between the dynamics of the microtubule cytoskeleton and the large-scale organization of the actin cytoskeleton; such a connection is important for the overall integration of cell shape and movement.

### External Signals Can Dictate the Direction of Cell Migration

**Chemotaxis** is the movement of a cell toward or away from a source of some diffusible chemical. These external signals act through Rho family proteins to set up large-scale cell polarity by influencing the organization of the cell motility apparatus. One well-studied example is the chemotactic movement of a class of white blood cells, called *neutrophils*, toward a source of bacterial infection. Receptor proteins on the surface of neutrophils enable them to detect very low concentrations of *N*-formylated peptides that are derived from bacterial proteins (only prokaryotes begin protein synthesis with *N*-formylmethionine). Using these receptors, neutrophils are guided to bacterial targets by their ability to detect a difference of only 1% in the concentration of these diffusible peptides on one side of the cell versus the other (Figure 16-86A).

In this case, and in the chemotaxis of *Dictyostelium* amoebae toward a source of cyclic AMP, binding of the chemoattractant to its G-protein-coupled receptor activates phosphoinositide 3-kinases (PI3Ks) (see Figure 15-52), which generate a signaling molecule [ $\text{PI}(3,4,5)\text{P}_3$ ] that in turn activates the Rac GTPase. Rac



then activates the Arp 2/3 complex leading to lamellipodial protrusion. Through an unknown mechanism, accumulation of the polarized actin web at the leading edge causes further local enhancement of PI3K activity in a positive feedback loop, strengthening the induction of protrusion. The PI(3,4,5)P<sub>3</sub> that activates Rac cannot diffuse far from its site of synthesis, since it is rapidly converted back into PI(4,5)P<sub>2</sub> by a constitutively active lipid phosphatase. At the same time, binding of the chemoattractant ligand to its receptor activates another signaling pathway that turns on Rho and enhances myosin-based contractility. The two processes directly inhibit each other, such that Rac activation dominates in the front of the cell and Rho activation dominates in the rear (Figure 16–86B). This enables the cell to maintain its functional polarity with protrusion at the leading edge and contraction at the back.

Nondiffusible chemical cues attached to the extracellular matrix or to the surface of cells can also influence the direction of cell migration. When these signals activate receptors, they can cause increased cell adhesion and directed actin polymerization. Most long-distance cell migrations in animals, including neural-crest-cell migration and the travels of neuronal growth cones, depend on a combination of diffusible and nondiffusible signals to steer the locomoting cells or growth cones to their proper destinations.

### Communication Among Cytoskeletal Elements Coordinates Whole-Cell Polarization and Locomotion

The interconnected cytoskeleton is crucial for cell migration. Although movement is driven primarily by actin polymerization and myosin contractility, septins and intermediate filaments also participate. For example, vimentin intermediate filament networks associate with integrins at focal adhesions, and vimentin-deficient fibroblasts display impaired mechanical stability, migration, and contractile capacity. Furthermore, disruption of linker proteins that connect different cytoskeletal elements, including several plakins and KASH proteins, leads to defects in cell polarization and migration. Thus, interactions among cytoplasmic filament systems, as well as mechanical linkage to the nucleus, are required for complex, whole-cell behaviors such as migration.

Cells also use microtubules to help organize persistent movement in a specific direction. In many locomoting cells, the position of the centrosome is influenced by the location of protrusive actin polymerization. Activation of receptors on the protruding front edge of a cell might locally activate dynein motor proteins that move the centrosome by pulling on its microtubules. Several effector proteins downstream of Rac and Rho modulate microtubule dynamics directly: for example, a protein kinase activated by Rac can phosphorylate (and thereby inhibit) the tubulin-binding protein stathmin (see Panel 16–4), thereby stabilizing microtubules.

**Figure 16–86 Neutrophil polarization and chemotaxis.** (A) The pipette tip at the right is leaking a small amount of the bacterial peptide formyl-Met-Leu-Phe, which is recognized by the human neutrophil as the product of a foreign invader. The neutrophil quickly extends a new lamellipodium toward the source of the chemoattractant peptide (top). It then extends this lamellipodium and polarizes its cytoskeleton so that contractile myosin II is located primarily at the rear, opposite the position of the lamellipodium (middle). Finally, the cell crawls toward the source of the peptide (bottom). If a real bacterium were the source of the peptide, rather than an investigator's pipette, the neutrophil would engulf the bacterium and destroy it (see also Figure 16–3 and Movie 16.14). (B) Binding of bacterial molecules to G-protein-coupled receptors on the neutrophil stimulates directed motility. These receptors are found all over the surface of the cell, but are more likely to be bound to the bacterial ligand at the front. Two distinct signaling pathways contribute to the cell's polarization. At the front of the cell, stimulation of the Rac pathway leads, via the trimeric G protein G<sub>i</sub>, to growth of protrusive actin networks. Second messengers within this pathway are short-lived, so protrusion is limited to the region of the cell closest to the stimulant. The same receptor also stimulates a second signaling pathway, via the trimeric G proteins G<sub>12</sub> and G<sub>13</sub>, that triggers the activation of Rho. The two pathways are mutually antagonistic. Since Rac-based protrusion is active at the front of the cell, Rho is activated only at the rear of the cell, stimulating contraction of the cell rear and assisting directed movement. (A, from O.D. Weiner et al., *Nat. Cell Biol.* 1:75–81, 1999. With permission from Macmillan Publishers Ltd.)

In turn, microtubules influence actin rearrangements and cell adhesion. The centrosome nucleates a large number of dynamic microtubules, and its repositioning means that the plus ends of many of these microtubules extend into the protrusive region of the cell. Direct interactions with microtubules help guide focal adhesion dynamics in migrating cells. Microtubules might also influence actin filament formation by delivering Rac-GEFs that bind to the +TIPs traveling on growing microtubule ends. Microtubules also transport cargoes to and from the focal adhesions, thereby affecting their signaling and disassembly. Thus, microtubules reinforce the polarity information that the actin cytoskeleton receives from the outside world, allowing a sensitive response to weak signals and enabling motility to persist in the same direction for a prolonged period.

## Summary

*Whole-cell movements and the large-scale shaping and structuring of cells require the coordinated activities of all three basic filament systems along with a large variety of cytoskeletal accessory proteins, including motor proteins. Cell crawling—a widespread behavior important in embryonic development and also in wound healing, tissue maintenance, and immune system function in the adult animal—is a prime example of such complex, coordinated cytoskeletal action. For a cell to crawl, it must generate and maintain an overall structural polarity, which is influenced by external cues. In addition, the cell must coordinate protrusion at the leading edge (by assembly of new actin filaments), adhesion of the newly protruded part of the cell to the substratum, and forces generated by molecular motors to bring the cell body forward.*

## WHAT WE DON'T KNOW

- How is the cell cortex regulated locally and globally to coordinate its activities at different places on the cell surface? What determines, for example, where filopodia form?
- How are actin-regulatory proteins controlled spatially in the cytoplasm to generate multiple distinct types of actin arrays in the same cell?
- Are there biologically important processes occurring inside a microtubule?
- How can we account for the fact that there are many different kinesins and myosins in the cytoplasm but only one dynein?
- Mutations in the nuclear lamin proteins cause a large number of diseases called laminopathies. What do we not understand about the nuclear lamina that could account for this fact?

## PROBLEMS

Which statements are true? Explain why or why not.

**16–1** The role of ATP hydrolysis in actin polymerization is similar to the role of GTP hydrolysis in tubulin polymerization: both serve to weaken the bonds in the polymer and thereby promote depolymerization.

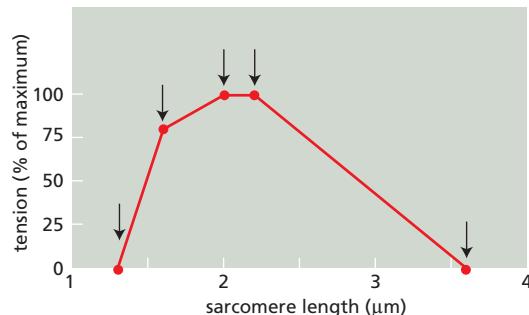
**16–2** Motor neurons trigger action potentials in muscle cell membranes that open voltage-sensitive  $\text{Ca}^{2+}$  channels in T tubules, allowing extracellular  $\text{Ca}^{2+}$  to enter the cytosol, bind to troponin C, and initiate rapid muscle contraction.

**16–3** In most animal cells, minus-end directed microtubule motors deliver their cargo to the periphery of the cell, whereas plus-end directed microtubule motors deliver their cargo to the interior of the cell.

Discuss the following problems.

**16–4** The concentration of actin in cells is 50–100 times greater than the critical concentration observed for pure actin in a test tube. How is this possible? What prevents the actin subunits in cells from polymerizing into filaments? Why is it advantageous to the cell to maintain such a large pool of actin subunits?

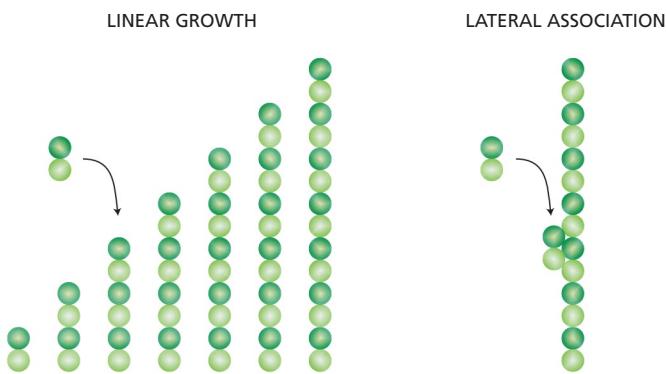
**16–5** Detailed measurements of sarcomere length and tension during isometric contraction in striated muscle provided crucial early support for the sliding-filament



**Figure Q16–1** Tension as a function of sarcomere length during isometric contraction (Problem 16–5).

model of muscle contraction. Based on your understanding of the sliding-filament model and the structure of a sarcomere, propose a molecular explanation for the relationship of tension to sarcomere length in the portions of **Figure Q16–1** marked I, II, III, and IV. (In this muscle, the length of the myosin filament is 1.6  $\mu\text{m}$ , and the lengths of the actin thin filaments that project from the Z discs are 1.0  $\mu\text{m}$ .)

**16–6** At 1.4 mg/mL pure tubulin, microtubules grow at a rate of about 2  $\mu\text{m}/\text{min}$ . At this growth rate, how many  $\alpha\beta$ -tubulin dimers (8 nm in length) are added to the ends of a microtubule each second?



**Figure Q16-2** Model for microtubule nucleation by pure  $\alpha\beta$ -tubulin dimers (Problem 16-7).

**16-7** A solution of pure  $\alpha\beta$ -tubulin dimers is thought to nucleate microtubules by forming a linear protofilament about seven dimers in length. At that point, the probabilities that the next  $\alpha\beta$ -dimer will bind laterally or to the end of the protofilament are about equal. The critical event for microtubule formation is thought to be the first lateral association (**Figure Q16-2**). How does lateral association promote the subsequent rapid formation of a microtubule?

**16-8** How does a centrosome “know” when it has found the center of the cell?

**16-9** The movements of single motor-protein molecules can be analyzed directly. Using polarized laser light, it is possible to create interference patterns that exert a centrally directed force, ranging from zero at the center to a few piconewtons at the periphery (about 200 nm from the center). Individual molecules that enter the interference pattern are rapidly pushed to the center, allowing them to be captured and moved at the experimenter’s discretion.

Using such “optical tweezers,” single kinesin molecules can be positioned on a microtubule that is fixed to a coverslip. Although a single kinesin molecule cannot be seen optically, it can be tagged with a silica bead and tracked indirectly by following the bead (**Figure Q16-3A**). In the absence of ATP, the kinesin molecule remains at the center of the interference pattern, but with ATP it moves toward the plus end of the microtubule. As kinesin moves along the microtubule, it encounters the force of the interference pattern, which simulates the load kinesin carries during its actual function in the cell. Moreover, the pressure against the silica bead counters the effects of Brownian (thermal) motion, so that the position of the bead more accurately reflects the position of the kinesin molecule on the microtubule.

A trace of the movements of a kinesin molecule along a microtubule is shown in **Figure Q16-3B**.

**A.** As shown in **Figure Q16-3B**, all movement of kinesin is in one direction (toward the plus end of the microtubule). What supplies the free energy needed to ensure a unidirectional movement along the microtubule?

**B.** What is the average rate of movement of kinesin along the microtubule?

**C.** What is the length of each step that a kinesin takes as it moves along a microtubule?

**D.** From other studies it is known that kinesin has two globular domains that can each bind to  $\beta$ -tubulin, and that kinesin moves along a single protofilament in a microtubule. In each protofilament, the  $\beta$ -tubulin subunit repeats at 8-nm intervals. Given the step length and the interval between  $\beta$ -tubulin subunits, how do you suppose a kinesin molecule moves along a microtubule?

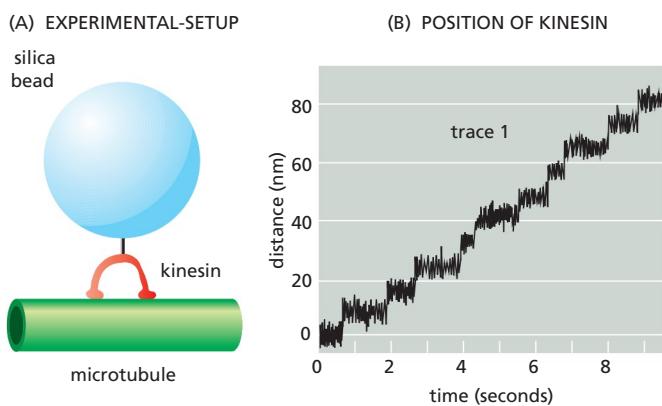
**E.** Is there anything in the data in **Figure Q16-3B** that tells you how many ATP molecules are hydrolyzed per step?

**16-10** A mitochondrion 1  $\mu\text{m}$  long can travel the 1 meter length of the axon from the spinal cord to the big toe in a day. The Olympic men’s freestyle swimming record for 200 meters is 1.75 minutes. In terms of body lengths per day, who is moving faster: the mitochondrion or the Olympic record holder? (Assume that the swimmer is 2 meters tall.)

**16-11** Cofilin preferentially binds to older actin filaments and promotes their disassembly. How does cofilin distinguish old filaments from new ones?

**16-12** Why is it that intermediate filaments have identical ends and lack polarity, whereas actin filaments and microtubules have two distinct ends with a defined polarity?

**16-13** How is the unidirectional motion of a lamellipodium maintained?



**Figure Q16-3** Movement of kinesin along a microtubule (Problem 16-9). (A) Experimental set-up, with kinesin linked to a silica bead, moving along a microtubule. (B) Position of kinesin (as visualized by the position of the silica bead) relative to the center of the interference pattern, as a function of time of movement along the microtubule. The jagged nature of the trace results from Brownian motion of the bead.

## REFERENCES

### General

- Bray D (2001) Cell Movements: From Molecules to Motility, 2nd ed. New York: Garland Science.
- Howard J (2001) Mechanics of Motor Proteins and the Cytoskeleton. Sunderland, MA: Sinauer.
- Kavallaris M (2012) Cytoskeleton and Human Disease. New York: Springer, Humana Press.

### Function and Origin of the Cytoskeleton

- Garner EC, Bernard R, Wang W et al. (2011) Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in *B. subtilis*. *Science* 333, 222–225.
- Garner EC, Campbell CS & Mullins RD (2004) Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* 306, 1021–1025.
- Hill TL & Kirschner MW (1982) Bioenergetics and kinetics of microtubule and actin filament assembly-disassembly. *Int. Rev. Cytol.* 78, 1–125.
- Jones LJ, Carballido-López R & Errington J (2001) Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* 104, 913–922.
- Luby-Phelps K (2000) Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *Int. Rev. Cytol.* 192, 189–221.
- Oosawa F & Asakura S (1975) Thermodynamics of the Polymerization of Protein, pp. 41–55 and 90–108. New York: Academic Press.
- Osawa M, Anderson DE & Erickson HP (2009) Curved FtsZ protofilaments generate bending forces on liposome membranes. *EMBO J.* 28, 3476–3484.
- Pauling L (1953) Aggregation of globular proteins. *Discuss. Faraday Soc.* 13, 170–176.
- Purcell EM (1977) Life at low Reynolds number. *Am. J. Phys.* 45, 3–11.
- Theriot JA (2013) Why are bacteria different from eukaryotes? *BMC Biol.* 11, 119.

### Actin and Actin-binding Proteins

- Fehon RG, McClatchey AI & Bretscher A (2010) Organizing the cell cortex: the role of ERM proteins. *Nat. Rev. Mol. Cell Biol.* 11, 276–287.
- Mullins RD, Heuser JA & Pollard TD (1998) The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. *Proc. Natl Acad. Sci. U.S.A.* 95, 6181–6186.
- Zigmond SH (2004) Formin-induced nucleation of actin filaments. *Curr. Opin. Cell Biol.* 16, 99–105.

### Myosin and Actin

- Cooke R (2004) The sliding filament model: 1972–2004. *J. Gen. Physiol.* 123, 643–656.
- Hammer JA 3rd & Sellers JR (2011) Walking to work: roles for class V myosins as cargo transporters. *Nat. Rev. Mol. Cell Biol.* 13, 13–26.
- Howard J (1997) Molecular motors: structural adaptations to cellular functions. *Nature* 389, 561–567.
- Rice S, Lin AW, Safer D et al. (1999) A structural change in the kinesin motor protein that drives motility. *Nature* 402, 778–784.
- Vikstrom KL & Leinwand LA (1996) Contractile protein mutations and heart disease. *Curr. Opin. Cell Biol.* 8, 97–105.
- Wells AL, Lin AW, Chen LQ et al. (1999) Myosin VI is an actin-based motor that moves backwards. *Nature* 401, 505–508.
- Yildiz A, Forkey JN, McKinney SA et al. (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* 300, 2061–2065.

### Microtubules

- Aldaz H, Rice LM, Stearns T & Agard DA (2005) Insights into microtubule nucleation from the crystal structure of human gamma-tubulin. *Nature* 435, 523–527.

- Brouhard GJ, Stear JH, Noetzel TL et al. (2008) XMAP215 is a processive microtubule polymerase. *Cell* 132, 79–88.
- Dogterom M & Yurke B (1997) Measurement of the force-velocity relation for growing microtubules. *Science* 278, 856–860.
- Doxsey S, McCollum D & Theurkauf W (2005) Centrosomes in cellular regulation. *Annu. Rev. Cell Dev. Biol.* 21, 411–434.
- Galjart N (2010) Plus-end-tracking proteins and their interactions at microtubule ends. *Curr. Biol.* 20, R528–R537.
- Hotani H & Horio T (1988) Dynamics of microtubules visualized by darkfield microscopy: treadmilling and dynamic instability. *Cell Motil. Cytoskeleton* 10, 229–236.
- Howard J, Hudspeth AJ & Vale RD (1989) Movement of microtubules by single kinesin molecules. *Nature* 342, 154–158.
- Kerssemakers JW, Munteanu EL, Laan L et al. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature* 442, 709–712.
- Kikkawa M (2013) Big steps toward understanding dynein. *J. Cell Biol.* 202, 15–23.
- Mitchison T & Kirschner M (1984) Dynamic instability of microtubule growth. *Nature* 312, 237–242.
- Reck-Peterson SL, Yildiz A, Carter AP et al. (2006) Single-molecule analysis of dynein processivity and stepping behavior. *Cell* 126, 335–348.
- Sharp DJ & Ross JL (2012) Microtubule-severing enzymes at the cutting edge. *J. Cell Sci.* 125, 2561–2569.
- Singla V & Reiter JF (2006) The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* 313, 629–633.
- Stearns T & Kirschner M (1994) *In vitro* reconstitution of centrosome assembly and function: the central role of gamma-tubulin. *Cell* 76, 623–637.
- Svoboda K, Schmidt CF, Schnapp BJ & Block SM (1993) Direct observation of kinesin stepping by optical trapping interferometry. *Nature* 365, 721–727.
- Verhey KJ, Kaul N & Soppina V (2011) Kinesin assembly and movement in cells. *Annu. Rev. Biophys.* 40, 267–288.

### Intermediate Filaments and Septins

- Helfand BT, Chang L & Goldman RD (2003) The dynamic and motile properties of intermediate filaments. *Annu. Rev. Cell Dev. Biol.* 19, 445–467.
- Isermann P & Lammerding J (2013) Nuclear mechanics and mechanotransduction in health and disease. *Curr. Biol.* 23, R1113–R1121.
- Saarikangas J & Barral Y (2011) The emerging functions of septins in metazoans. *EMBO Rep.* 12, 1118–1126.

### Cell Polarization and Migration

- Abercrombie M (1980) The crawling movement of metazoan cells. *Proc. R. Soc. Lond. B* 207, 129–147.
- Gardel ML, Schneider IC, Aratyn-Schaus Y & Waterman CM (2010) Mechanical integration of actin and adhesion dynamics in cell migration. *Annu. Rev. Cell Dev. Biol.* 26, 315–333.
- Lo CM, Wang HB, Dembo M & Wang YL (2000) Cell movement is guided by the rigidity of the substrate. *Biophys. J.* 79, 144–152.
- Madden K & Snyder M (1998) Cell polarity and morphogenesis in budding yeast. *Annu. Rev. Microbiol.* 52, 687–744.
- Parent CA & Devreotes PN (1999) A cell's sense of direction. *Science* 284, 765–770.
- Pollard TD & Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112, 453–465.
- Rafelski SM & Theriot JA (2004) Crawling toward a unified model of cell mobility: spatial and temporal regulation of actin dynamics. *Annu. Rev. Biochem.* 73, 209–239.
- Ridley A (2011) Life at the leading edge. *Cell* 145, 1012–1022.
- Vitriol EA & Zheng JQ (2012) Growth cone travel in space and time: the cellular ensemble of cytoskeleton, adhesion, and membrane. *Neuron* 73, 1068–1081.
- Weiner OD (2002) Regulation of cell polarity during eukaryotic chemotaxis: the chemotactic compass. *Curr. Opin. Cell Biol.* 14, 196–202.

# The Cell Cycle

CHAPTER

# 17

The only way to make a new cell is to duplicate a cell that already exists. This simple fact, first established in the middle of the nineteenth century, carries with it a profound message for the continuity of life. All living organisms, from the unicellular bacterium to the multicellular mammal, are products of repeated rounds of cell growth and division extending back in time to the beginnings of life on Earth over three billion years ago.

A cell reproduces by performing an orderly sequence of events in which it duplicates its contents and then divides in two. This cycle of duplication and division, known as the **cell cycle**, is the essential mechanism by which all living things reproduce. In unicellular species, such as bacteria and yeasts, each cell division produces a complete new organism. In multicellular species, long and complex sequences of cell divisions are required to produce a functioning organism. Even in the adult body, cell division is usually needed to replace cells that die. In fact, each of us must manufacture many millions of cells every second simply to survive: if all cell division were stopped—by exposure to a very large dose of x-rays, for example—we would die within a few days.

The details of the cell cycle vary from organism to organism and at different times in an organism's life. Certain characteristics, however, are universal. At a minimum, the cell must accomplish its most fundamental task: the passing on of its genetic information to the next generation of cells. To produce two genetically identical daughter cells, the DNA in each chromosome must first be faithfully replicated to produce two complete copies. The replicated chromosomes must then be accurately distributed (*segregated*) to the two daughter cells, so that each receives a copy of the entire genome (Figure 17-1). In addition to duplicating their genome, most cells also duplicate their other organelles and macromolecules; otherwise, daughter cells would get smaller with each division. To maintain their size, dividing cells must coordinate their growth (that is, their increase in cell mass) with their division.

This chapter describes the events of the cell cycle and how they are controlled and coordinated. We begin with a brief overview of the cell cycle. We then describe the *cell-cycle control system*, a complex network of regulatory proteins that triggers the different events of the cycle. We next consider in detail the major stages of the cell cycle, in which the chromosomes are duplicated and then segregated into the two daughter cells. Finally, we consider how extracellular signals govern the rates of cell growth and division and how these two processes are coordinated.

## OVERVIEW OF THE CELL CYCLE

The most basic function of the cell cycle is to duplicate the vast amount of DNA in the chromosomes and then segregate the copies into two genetically identical daughter cells. These processes define the two major phases of the cell cycle. Chromosome duplication occurs during *S phase* (S for DNA synthesis), which requires 10–12 hours and occupies about half of the cell-cycle time in a typical mammalian cell. After S phase, chromosome segregation and cell division occur in *M phase* (M for *mitosis*), which requires much less time (less than an hour in a mammalian cell). M phase comprises two major events: nuclear division, or *mitosis*, during

### IN THIS CHAPTER

OVERVIEW OF THE CELL CYCLE

THE CELL-CYCLE CONTROL SYSTEM

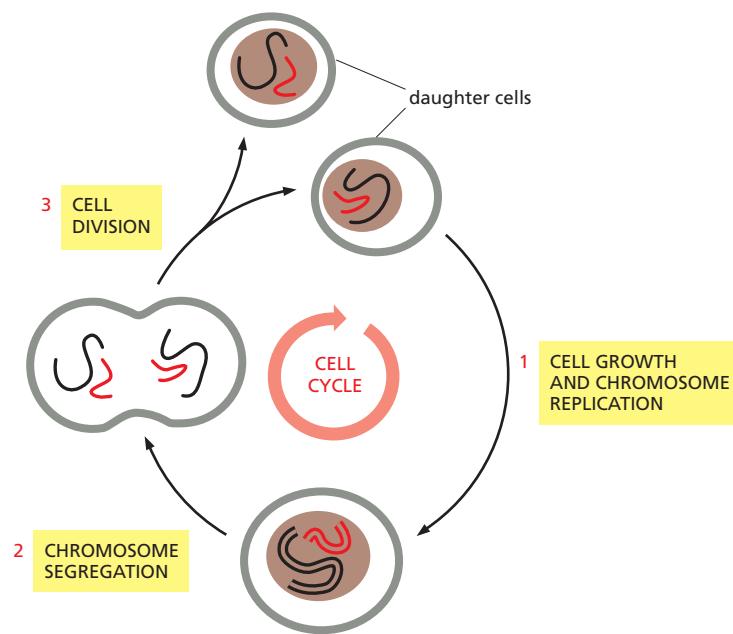
S PHASE

MITOSIS

CYTOKINESIS

MEIOSIS

CONTROL OF CELL DIVISION AND CELL GROWTH



**Figure 17-1** The cell cycle. The division of a hypothetical eukaryotic cell with two chromosomes (one red, and one black) is shown to illustrate how two genetically identical daughter cells are produced in each cycle. Each of the daughter cells will often continue to divide by going through additional cell cycles.

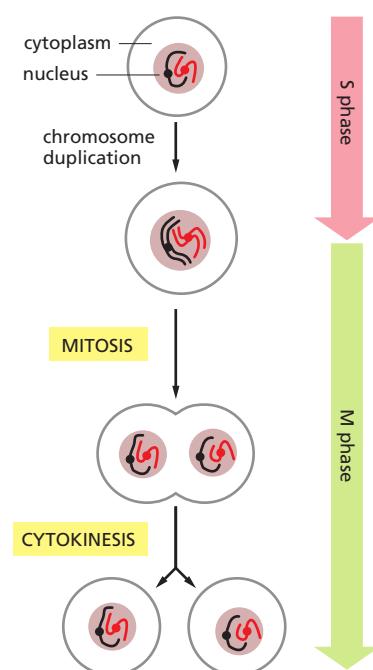
which the copied chromosomes are distributed into a pair of daughter nuclei; and cytoplasmic division, or *cytokinesis*, when the cell itself divides in two (Figure 17-2).

At the end of S phase, the DNA molecules in each pair of duplicated chromosomes are intertwined and held tightly together by specialized protein linkages. Early in mitosis at a stage called *prophase*, the two DNA molecules are gradually disentangled and condensed into pairs of rigid, compact rods called **sister chromatids**, which remain linked by *sister-chromatid cohesion*. When the nuclear envelope disassembles later in mitosis, the sister-chromatid pairs become attached to the *mitotic spindle*, a giant bipolar array of microtubules (discussed in Chapter 16). Sister chromatids are attached to opposite poles of the spindle and, eventually, align at the spindle equator in a stage called *metaphase*. The destruction of sister-chromatid cohesion at the start of *anaphase* separates the sister chromatids, which are pulled to opposite poles of the spindle. The spindle is then disassembled, and the segregated chromosomes are packaged into separate nuclei at *telophase*. Cytokinesis then cleaves the cell in two, so that each daughter cell inherits one of the two nuclei (Figure 17-3).

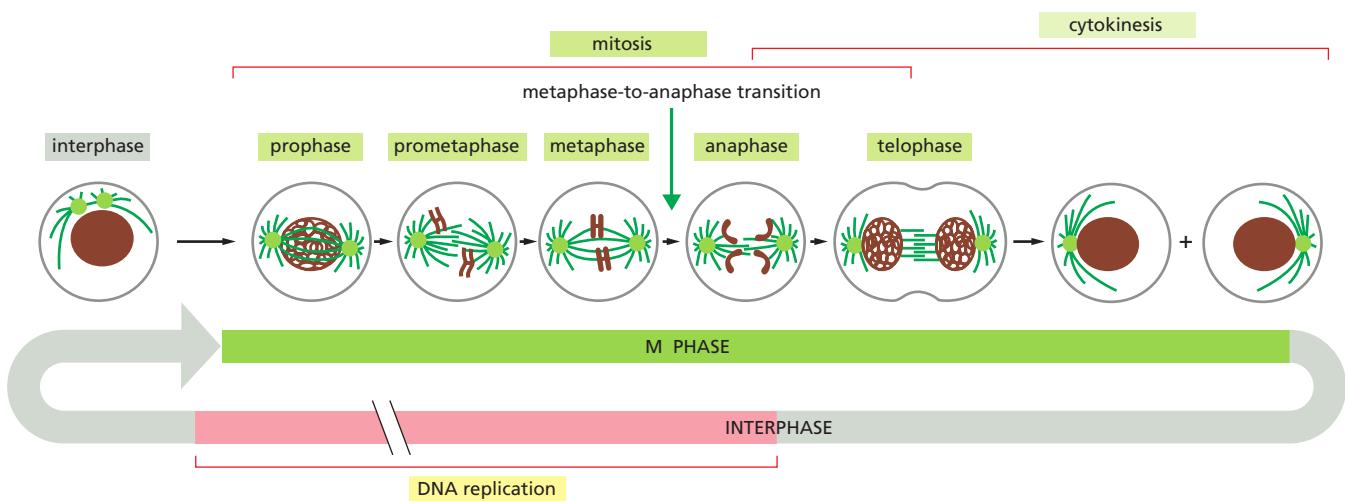
### The Eukaryotic Cell Cycle Usually Consists of Four Phases

Most cells require much more time to grow and double their mass of proteins and organelles than they require to duplicate their chromosomes and divide. Partly to allow time for growth, most cell cycles have *gap phases*—a **G<sub>1</sub> phase** between M phase and S phase and a **G<sub>2</sub> phase** between S phase and mitosis. Thus, the eukaryotic cell cycle is traditionally divided into four sequential phases: G<sub>1</sub>, S, G<sub>2</sub>, and M. G<sub>1</sub>, S, and G<sub>2</sub> together are called **interphase** (Figure 17-4, and see Figure 17-3). In a typical human cell proliferating in culture, interphase might occupy 23 hours of a 24-hour cycle, with 1 hour for M phase. Cell growth occurs throughout the cell cycle, except during mitosis.

The two gap phases are more than simple time delays to allow cell growth. They also provide time for the cell to monitor the internal and external environment



**Figure 17-2** The major events of the cell cycle. The major chromosomal events of the cell cycle occur in S phase, when the chromosomes are duplicated, and M phase, when the duplicated chromosomes are segregated into a pair of daughter nuclei (in mitosis), after which the cell itself divides into two (cytokinesis).

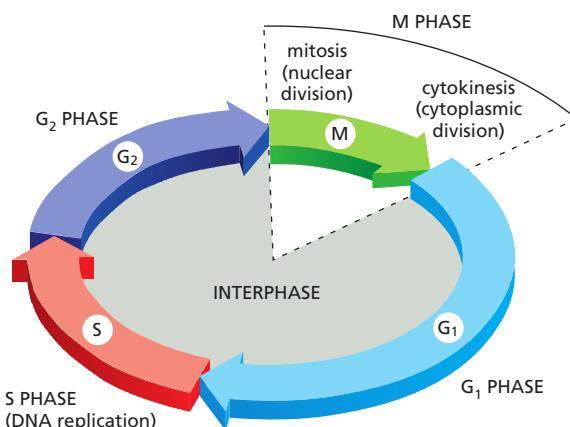


**Figure 17–3** The events of eukaryotic cell division as seen under a microscope. The easily visible processes of nuclear division (mitosis) and cell division (cytokinesis), collectively called M phase, typically occupy only a small fraction of the cell cycle. The other, much longer, part of the cycle is known as interphase, which includes S phase and the gap phases (discussed in text). The five stages of mitosis are shown: an abrupt change in the biochemical state of the cell occurs at the transition from metaphase to anaphase. A cell can pause in metaphase before this transition point, but once it passes this point, the cell carries on to the end of mitosis and through cytokinesis into interphase.

to ensure that conditions are suitable and preparations are complete before the cell commits itself to the major upheavals of S phase and mitosis. The G<sub>1</sub> phase is especially important in this respect. Its length can vary greatly depending on external conditions and extracellular signals from other cells. If extracellular conditions are unfavorable, for example, cells delay progress through G<sub>1</sub> and may even enter a specialized resting state known as G<sub>0</sub> (G zero), in which they can remain for days, weeks, or even years before resuming proliferation. Indeed, many cells remain permanently in G<sub>0</sub> until they or the organism dies. If extracellular conditions are favorable and signals to grow and divide are present, cells in early G<sub>1</sub> or G<sub>0</sub> progress through a commitment point near the end of G<sub>1</sub> known as **Start** (in yeasts) or the **restriction point** (in mammalian cells). We will use the term Start for both yeast and animal cells. After passing this point, cells are committed to DNA replication, even if the extracellular signals that stimulate cell growth and division are removed.

### Cell-Cycle Control Is Similar in All Eukaryotes

Some features of the cell cycle, including the time required to complete certain events, vary greatly from one cell type to another, even in the same organism. The basic organization of the cycle, however, is essentially the same in all eukaryotic



**Figure 17–4** The four phases of the cell cycle. In most cells, gap phases separate the major events of S phase and M phase. G<sub>1</sub> is the gap between M phase and S phase, while G<sub>2</sub> is the gap between S phase and M phase.

**Figure 17–5 Mammalian cells proliferating in culture.** The cells in this scanning electron micrograph are rat fibroblasts. Cells at the lower left have rounded up and are in mitosis. (Courtesy of Guenter Albrecht-Buehler.)

cells, and all eukaryotes appear to use similar machinery and control mechanisms to drive and regulate cell-cycle events. The proteins of the cell-cycle control system, for example, first appeared over a billion years ago. Remarkably, they have been so well conserved over the course of evolution that many of them function perfectly when transferred from a human cell to a yeast cell. We can therefore study the cell cycle and its regulation in a variety of organisms and use the findings from all of them to assemble a unified picture of how eukaryotic cells divide.

Several model organisms are used in the analysis of the eukaryotic cell cycle. The budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* are simple eukaryotes in which powerful molecular and genetic approaches can be used to identify and characterize the genes and proteins that govern the fundamental features of cell division. The early embryos of certain animals, particularly those of the frog *Xenopus laevis*, are excellent tools for biochemical dissection of cell-cycle control mechanisms, while the fruit fly *Drosophila melanogaster* is useful for the genetic analysis of mechanisms underlying the control and coordination of cell growth and division in multicellular organisms. Cultured human cells provide an excellent system for the molecular and microscopic exploration of the complex processes by which our own cells divide.



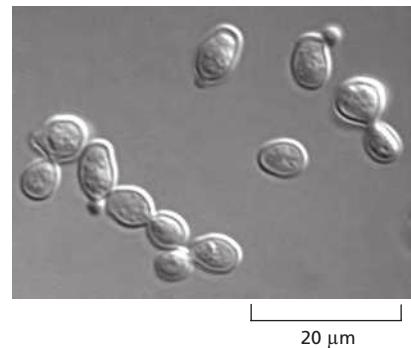
### Cell-Cycle Progression Can Be Studied in Various Ways

How can we tell what stage a cell has reached in the cell cycle? One way is simply to look at living cells with a microscope. A glance at a population of mammalian cells proliferating in culture reveals that a fraction of the cells have rounded up and are in mitosis (Figure 17–5). Others can be observed in the process of cytokinesis. Similarly, looking at budding yeast cells under a microscope is very useful, because the size of the bud provides an indication of cell-cycle stage (Figure 17–6). We can gain additional clues about cell-cycle position by staining cells with DNA-binding fluorescent dyes (which reveal the condensation of chromosomes in mitosis) or with antibodies that recognize specific cell components such as the microtubules (revealing the mitotic spindle). S-phase cells can be identified in the microscope by supplying them with visualizable molecules that are incorporated into newly synthesized DNA, such as the artificial thymidine analog bromodeoxyuridine (BrdU); cell nuclei that have incorporated BrdU are then revealed by staining with anti-BrdU antibodies (Figure 17–7).

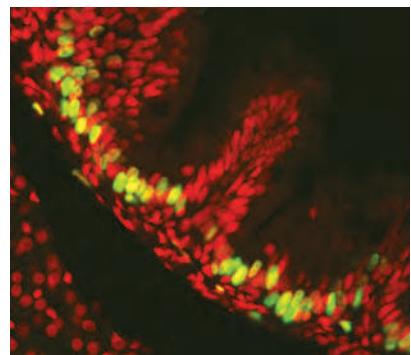
Typically, in a population of cultured mammalian cells that are all proliferating rapidly but asynchronously, about 30–40% will be in S phase at any instant and become labeled by a brief pulse of BrdU. From the proportion of cells in such a population that are labeled, we can estimate the duration of S phase as a fraction of the whole cell-cycle duration. Similarly, from the proportion of cells in mitosis (the *mitotic index*), we can estimate the duration of M phase.

Another way to assess the stage that a cell has reached in the cell cycle is by measuring its DNA content, which doubles during S phase. This approach is greatly facilitated by the use of fluorescent DNA-binding dyes and a *flow cytometer*, which allows the rapid and automatic analysis of large numbers of cells (Figure 17–8). We can use flow cytometry to determine the lengths of G<sub>1</sub>, S, and G<sub>2</sub> + M phases, by measuring DNA content in a synchronized cell population as it progresses through the cell cycle.

**Figure 17–6 The morphology of budding yeast cells.** In a normal population of proliferating yeast cells, buds vary in size according to the cell-cycle stage. Unbudded cells are in G<sub>1</sub>. Progression through the Start transition triggers formation of a tiny bud, which grows in size during the S and M phases until it is almost the size of the mother cell. (Courtesy of Jeff Ubersax.)



**Figure 17–7 Labeling S-phase cells.** An immunofluorescence micrograph of BrdU-labeled epithelial cells of the zebrafish gut. The fish was exposed to BrdU, after which the tissue was fixed and prepared for labeling with fluorescent anti-BrdU antibodies (green). All the cells are stained with a red fluorescent dye. (Courtesy of Cécile Crosnier.)



## Summary

Cell division usually begins with duplication of the cell's contents, followed by distribution of those contents into two daughter cells. Chromosome duplication occurs during S phase of the cell cycle, whereas most other cell components are duplicated continuously throughout the cycle. During M phase, the replicated chromosomes are segregated into individual nuclei (mitosis), and the cell then splits in two (cytokinesis). S phase and M phase are usually separated by gap phases called G<sub>1</sub> and G<sub>2</sub>, when various intracellular and extracellular signals regulate cell-cycle progression. Cell-cycle organization and control have been highly conserved during evolution, and studies in a wide range of systems have led to a unified view of eukaryotic cell-cycle control.

## THE CELL-CYCLE CONTROL SYSTEM

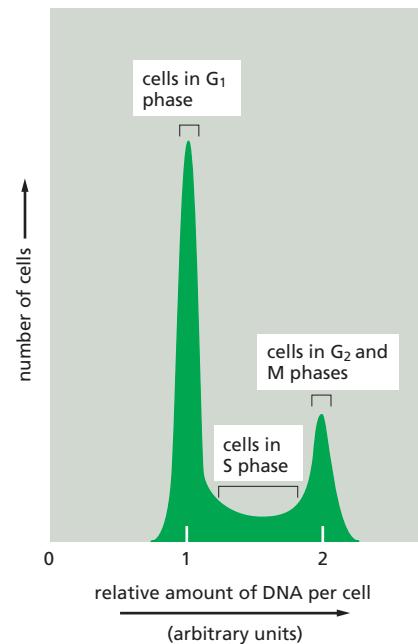
For many years, cell biologists watched the puppet show of DNA synthesis, mitosis, and cytokinesis but had no idea of what lay behind the curtain controlling these events. It was not even clear whether there was a separate control system, or whether the processes of DNA synthesis, mitosis, and cytokinesis somehow controlled themselves. A major breakthrough came in the late 1980s with the identification of the key proteins of the control system, along with the realization that they are distinct from the proteins that perform the processes of DNA replication, chromosome segregation, and so on.

In this section, we first consider the basic principles upon which the cell-cycle control system operates. We then discuss the protein components of the system and how they work together to time and coordinate the events of the cell cycle.

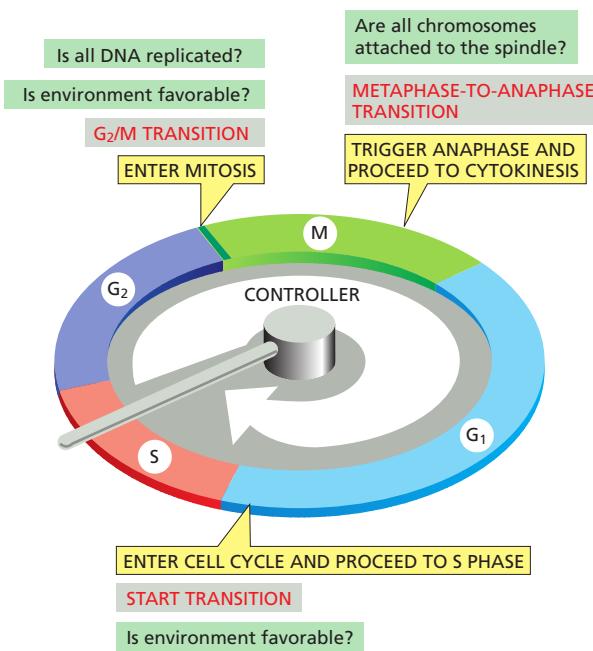
### The Cell-Cycle Control System Triggers the Major Events of the Cell Cycle

The **cell-cycle control system** operates much like a timer that triggers the events of the cell cycle in a set sequence (Figure 17–9). In its simplest form—as seen in the stripped-down cell cycles of early animal embryos, for example—the control system is rigidly programmed to provide a fixed amount of time for the completion of each cell-cycle event. The control system in these early embryonic divisions is independent of the events it controls, so that its timing mechanisms continue to operate even if those events fail. In most cells, however, the control system does respond to information received back from the processes it controls. If some malfunction prevents the successful completion of DNA synthesis, for example, signals are sent to the control system to delay progression to M phase. Such delays provide time for the machinery to be repaired and also prevent the disaster that might result if the cycle progressed prematurely to the next stage—and segregated incompletely replicated chromosomes, for example.

The cell-cycle control system is based on a connected series of biochemical switches, each of which initiates a specific cell-cycle event. This system of switches possesses many important features that increase the accuracy and reliability of cell-cycle progression. First, the switches are generally *binary* (on/off) and launch events in a complete, irreversible fashion. It would clearly be disastrous, for example, if events like chromosome condensation or nuclear-envelope breakdown were only partially initiated or started but not completed. Second, the cell-cycle control system is remarkably robust and reliable, partly because backup mechanisms and other features allow the system to operate effectively under a variety of conditions and even if some components fail. Finally, the control system is highly adaptable and can be modified to suit specific cell types or to respond to specific intracellular or extracellular signals.



**Figure 17–8 Analysis of DNA content with a flow cytometer.** This graph shows typical results obtained for a proliferating cell population when the DNA content of its individual cells is determined in a flow cytometer. (A flow cytometer, also called a fluorescence-activated cell sorter, or FACS, can also be used to sort cells according to their fluorescence—see Figure 8–2). The cells analyzed here were stained with a dye that becomes fluorescent when it binds to DNA, so that the amount of fluorescence is directly proportional to the amount of DNA in each cell. The cells fall into three categories: those that have an unreplicated complement of DNA and are therefore in G<sub>1</sub>, those that have a fully replicated complement of DNA (twice the G<sub>1</sub> DNA content) and are in G<sub>2</sub> or M phase, and those that have an intermediate amount of DNA and are in S phase. The distribution of cells indicates that there are greater numbers of cells in G<sub>1</sub> than in G<sub>2</sub> + M phase, showing that G<sub>1</sub> is longer than G<sub>2</sub> + M in this population.



**Figure 17–9** The control of the cell cycle. A cell-cycle control system triggers the essential processes of the cycle—such as DNA replication, mitosis, and cytokinesis. The control system is represented here as a central arm—the controller—that rotates clockwise, triggering essential processes when it reaches specific transitions on the outer dial (yellow boxes). Information about the completion of cell-cycle events, as well as signals from the environment, can cause the control system to arrest the cycle at these transitions.

In most eukaryotic cells, the cell-cycle control system governs cell-cycle progression at three major regulatory transitions (see Figure 17–9). The first is Start (or the restriction point) in late G<sub>1</sub>, where the cell commits to cell-cycle entry and chromosome duplication. The second is the G<sub>2</sub>/M transition, where the control system triggers the early mitotic events that lead to chromosome alignment on the mitotic spindle in metaphase. The third is the **metaphase-to-anaphase transition**, where the control system stimulates sister-chromatid separation, leading to the completion of mitosis and cytokinesis. The control system blocks progression through each of these transitions if it detects problems inside or outside the cell. If the control system senses problems in the completion of DNA replication, for example, it will hold the cell at the G<sub>2</sub>/M transition until those problems are solved. Similarly, if extracellular conditions are not appropriate for cell proliferation, the control system blocks progression through Start, thereby preventing cell division until conditions become favorable.

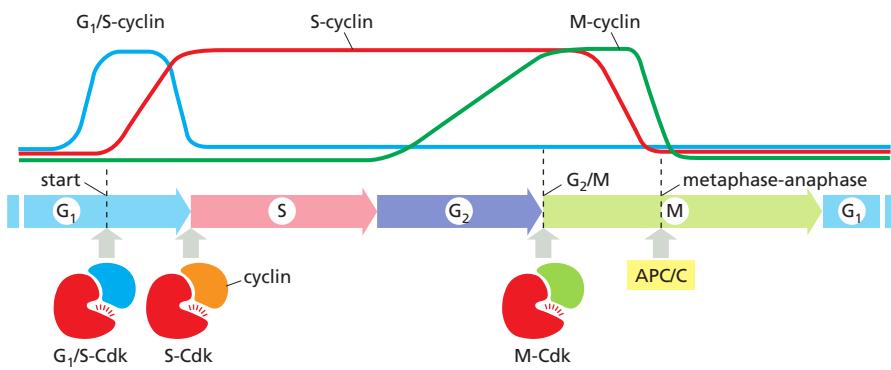
### The Cell-Cycle Control System Depends on Cyclically Activated Cyclin-Dependent Protein Kinases (Cdks)

Central components of the cell-cycle control system are members of a family of protein kinases known as **cyclin-dependent kinases (Cdks)**. The activities of these kinases rise and fall as the cell progresses through the cycle, leading to cyclical changes in the phosphorylation of intracellular proteins that initiate or regulate the major events of the cell cycle. An increase in Cdk activity at the G<sub>2</sub>/M transition, for example, increases the phosphorylation of proteins that control chromosome condensation, nuclear-envelope breakdown, spindle assembly, and other events that occur in early mitosis.

Cyclical changes in Cdk activity are controlled by a complex array of enzymes and other proteins. The most important of these Cdk regulators are proteins known as **cyclins**. Cdks, as their name implies, are dependent on cyclins for their activity: unless they are bound tightly to a cyclin, they have no protein kinase activity (Figure 17–10). Cyclins were originally named because they undergo a cycle of synthesis and degradation in each cell cycle. The levels of the Cdk proteins, by contrast, are constant. Cyclical changes in cyclin protein levels result in the cyclic assembly and activation of **cyclin-Cdk complexes** at specific stages of the cell cycle.



**Figure 17–10** Two key components of the cell-cycle control system. When cyclin forms a complex with Cdk, the protein kinase is activated to trigger specific cell-cycle events. Without cyclin, Cdk is inactive.



There are four classes of cyclins, each defined by the stage of the cell cycle at which they bind Cdks and function. All eukaryotic cells require three of these classes (Figure 17-11):

1. **G<sub>1</sub>/S-cyclins** activate Cdks in late G<sub>1</sub> and thereby help trigger progression through Start, resulting in a commitment to cell-cycle entry. Their levels fall in S phase.
2. **S-cyclins** bind Cdks soon after progression through Start and help stimulate chromosome duplication. S-cyclin levels remain elevated until mitosis, and these cyclins also contribute to the control of some early mitotic events.
3. **M-cyclins** activate Cdks that stimulate entry into mitosis at the G<sub>2</sub>/M transition. M-cyclin levels fall in mid-mitosis.

In most cells, a fourth class of cyclins, the **G<sub>1</sub>-cyclins**, helps govern the activities of the G<sub>1</sub>/S-cyclins, which control progression through Start in late G<sub>1</sub>.

In yeast cells, a single Cdk protein binds all classes of cyclins and triggers different cell-cycle events by changing cyclin partners at different stages of the cycle. In vertebrate cells, by contrast, there are four Cdks. Two interact with G<sub>1</sub>-cyclins, one with G<sub>1</sub>/S- and S-cyclins, and one with S- and M-cyclins. In this chapter, we simply refer to the different cyclin-Cdk complexes as **G<sub>1</sub>-Cdk**, **G<sub>1</sub>/S-Cdk**, **S-Cdk**, and **M-Cdk**. Table 17-1 lists the names of the individual Cdks and cyclins.

How do different cyclin-Cdk complexes trigger different cell-cycle events? The answer, at least in part, seems to be that the cyclin protein does not simply activate its Cdk partner but also directs it to specific target proteins. As a result, each cyclin-Cdk complex phosphorylates a different set of substrate proteins. The same cyclin-Cdk complex can also induce different effects at different times in the cycle, probably because the accessibility of some Cdk substrates changes during the cell cycle. Certain proteins that function in mitosis, for example, may become available for phosphorylation only in G<sub>2</sub>.

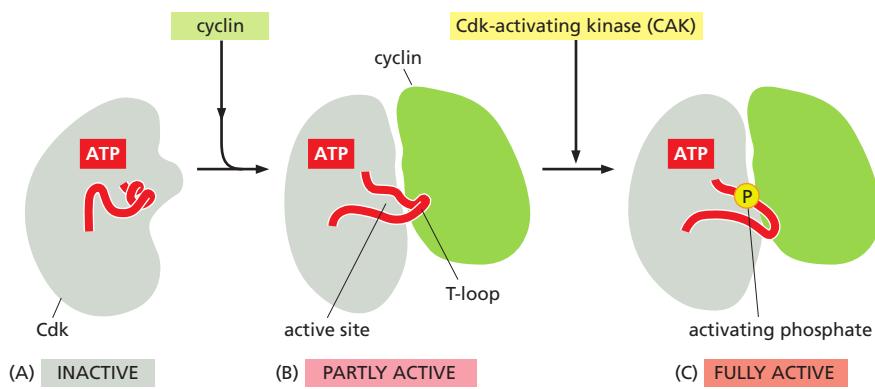
**Figure 17-11** Cyclin-Cdk complexes of the cell-cycle control system. The concentrations of the three major cyclin types oscillate during the cell cycle, while the concentrations of Cdks (not shown) do not change and exceed cyclin amounts. In late G<sub>1</sub>, rising G<sub>1</sub>/S-cyclin levels lead to the formation of G<sub>1</sub>/S-Cdk complexes that trigger progression through the Start transition. S-Cdk complexes form at the start of S phase and trigger DNA replication, as well as some early mitotic events. M-Cdk complexes form during G<sub>2</sub> but are held in an inactive state; they are activated at the end of G<sub>2</sub> and trigger entry into mitosis at the G<sub>2</sub>/M transition. A separate regulatory protein complex, the APC/C, initiates the metaphase-to-anaphase transition, as we discuss later.

**TABLE 17-1** The Major Cyclins and Cdks of Vertebrates and Budding Yeast

Cyclin-Cdk complex	Vertebrates		Budding yeast	
	Cyclin	Cdk partner	Cyclin	Cdk partner
G <sub>1</sub> -Cdk	Cyclin D*	Cdk4, Cdk6	Cln3	Cdk1**
G <sub>1</sub> /S-Cdk	Cyclin E	Cdk2	Cln1, 2	Cdk1
S-Cdk	Cyclin A	Cdk2, Cdk1**	Clb5, 6	Cdk1
M-Cdk	Cyclin B	Cdk1	Clb1, 2, 3, 4	Cdk1

\* There are three D cyclins in mammals (cyclins D1, D2, and D3).

\*\* The original name of Cdk1 was Cdc2 in both vertebrates and fission yeast, and Cdc28 in budding yeast.



**Figure 17–12** The structural basis of Cdk activation. These drawings are based on three-dimensional structures of human Cdk2 and cyclin A, as determined by x-ray crystallography. The location of the bound ATP is indicated. The enzyme is shown in three states. (A) In the inactive state, without cyclin bound, the active site is blocked by a region of the protein called the T-loop (red). (B) The binding of cyclin causes the T-loop to move out of the active site, resulting in partial activation of the Cdk2. (C) Phosphorylation of Cdk2 (by CAK) at a threonine residue in the T-loop further activates the enzyme by changing the shape of the T-loop, improving the ability of the enzyme to bind its protein substrates (Movie 17.1).

Studies of the three-dimensional structures of Cdk and cyclin proteins have revealed that, in the absence of cyclin, the active site in the Cdk protein is partly obscured by a protein loop, like a stone blocking the entrance to a cave (Figure 17–12A). Cyclin binding causes the loop to move away from the active site, resulting in partial activation of the Cdk enzyme (Figure 17–12B). Full activation of the cyclin–Cdk complex then occurs when a separate kinase, the **Cdk-activating kinase (CAK)**, phosphorylates an amino acid near the entrance of the Cdk active site. This causes a small conformational change that further increases the activity of the Cdk, allowing the kinase to phosphorylate its target proteins effectively and thereby induce specific cell-cycle events (Figure 17–12C).

### Cdk Activity Can Be Suppressed By Inhibitory Phosphorylation and Cdk Inhibitor Proteins (CKIs)

The rise and fall of cyclin levels is the primary determinant of Cdk activity during the cell cycle. Several additional mechanisms, however, help control Cdk activity at specific stages of the cycle.

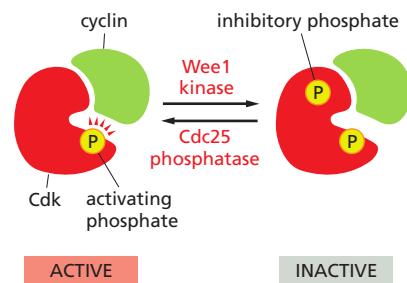
Phosphorylation at a pair of amino acids in the roof of the kinase active site inhibits the activity of a cyclin–Cdk complex. Phosphorylation of these sites by a protein kinase known as **Wee1** inhibits Cdk activity, while dephosphorylation of these sites by a phosphatase known as **Cdc25** increases Cdk activity (Figure 17–13). We will see later that this regulatory mechanism is particularly important in the control of M-Cdk activity at the onset of mitosis.

Binding of **Cdk inhibitor proteins (CKIs)** inactivates cyclin–Cdk complexes. The three-dimensional structure of a cyclin–Cdk–CKI complex reveals that CKI binding stimulates a large rearrangement in the structure of the Cdk active site, rendering it inactive (Figure 17–14). Cells use CKIs primarily to help govern the activities of G<sub>1</sub>/S- and S-Cdks early in the cell cycle.

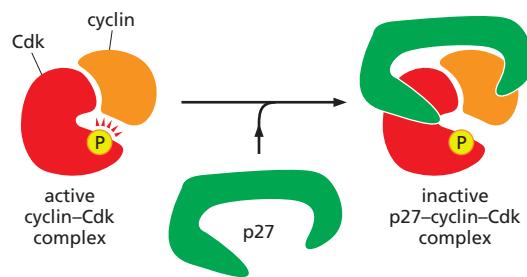
### Regulated Proteolysis Triggers the Metaphase-to-Anaphase Transition

Whereas activation of specific cyclin–Cdk complexes drives progression through the Start and G<sub>2</sub>/M transitions (see Figure 17–11), progression through the metaphase-to-anaphase transition is triggered not by protein phosphorylation but by protein destruction, leading to the final stages of cell division.

The key regulator of the metaphase-to-anaphase transition is the **anaphase-promoting complex, or cyclosome (APC/C)**, a member of the ubiquitin ligase family of enzymes. As discussed in Chapter 3, these enzymes are used in numerous cell processes to stimulate the proteolytic destruction of specific regulatory proteins. They polyubiquitylate specific target proteins, resulting in their destruction in proteasomes. Other ubiquitin ligases mark proteins for purposes other than destruction (discussed in Chapter 3).



**Figure 17–13** The regulation of Cdk activity by phosphorylation. The active cyclin–Cdk complex is turned off when the kinase Wee1 phosphorylates two closely spaced sites above the active site. Removal of these phosphates by the phosphatase Cdc25 activates the cyclin–Cdk complex. For simplicity, only one inhibitory phosphate is shown. CAK adds the activating phosphate, as shown in Figure 17–12.



**Figure 17–14** The inhibition of a cyclin–Cdk complex by a CKI. This drawing is based on the three-dimensional structure of the human cyclin A–Cdk2 complex bound to the CKI p27, as determined by x-ray crystallography. The p27 binds to both the cyclin and Cdk in the complex, distorting the active site of the Cdk. It also inserts into the ATP-binding site, further inhibiting the enzyme activity.

The APC/C catalyzes the ubiquitylation and destruction of two major types of proteins. The first is *securin*, which protects the protein linkages that hold sister-chromatid pairs together in early mitosis. Destruction of securin in metaphase activates a protease that separates the sisters and unleashes anaphase, as described later. The S- and M-cyclins are the second major targets of the APC/C. Destroying these cyclins inactivates most Cdks in the cell (see Figure 17–11). As a result, the many proteins phosphorylated by Cdks from S phase to early mitosis are dephosphorylated by various phosphatases in the anaphase cell. This dephosphorylation of Cdk targets is required for the completion of M phase, including the final steps in mitosis and then cytokinesis. Following its activation in mid-mitosis, the APC/C remains active in G<sub>1</sub> to provide a stable period of Cdk inactivity. When G<sub>1</sub>/S-Cdk is activated in late G<sub>1</sub>, the APC/C is turned off, thereby allowing cyclin accumulation in the next cell cycle.

The cell-cycle control system also uses another ubiquitin ligase called SCF (see Figure 3–71). It has many functions in the cell, but its major role in the cell cycle is to ubiquitylate certain CKI proteins in late G<sub>1</sub>, thereby helping to control the activation of S-Cdks and DNA replication. SCF is also responsible for the destruction of G<sub>1</sub>/S-cyclins in early S phase.

The APC/C and SCF are both large, multisubunit complexes with some related components (see Figure 3–71), but they are regulated differently. APC/C activity changes during the cell cycle, primarily as a result of changes in its association with an activating subunit—either **Cdc20** in mid-mitosis or **Cdh1** from late mitosis through early G<sub>1</sub>. These subunits help the APC/C recognize its target proteins (Figure 17–15A). SCF activity depends on substrate-binding subunits called F-box proteins. Unlike APC/C activity, however, SCF activity is constant during the cell cycle. Ubiquitylation by SCF is controlled instead by changes in the phosphorylation state of its target proteins, as F-box subunits recognize only specifically phosphorylated proteins (Figure 17–15B).

### Cell-Cycle Control Also Depends on Transcriptional Regulation

In the simple cell cycles of early animal embryos, gene transcription does not occur. Cell-cycle control depends exclusively on post-transcriptional mechanisms that involve the regulation of Cdks and ubiquitin ligases and their target proteins. In the more complex cell cycles of most cell types, however, transcriptional control provides an important additional level of regulation. Changes in cyclin gene transcription, for example, help control cyclin levels in most cells.

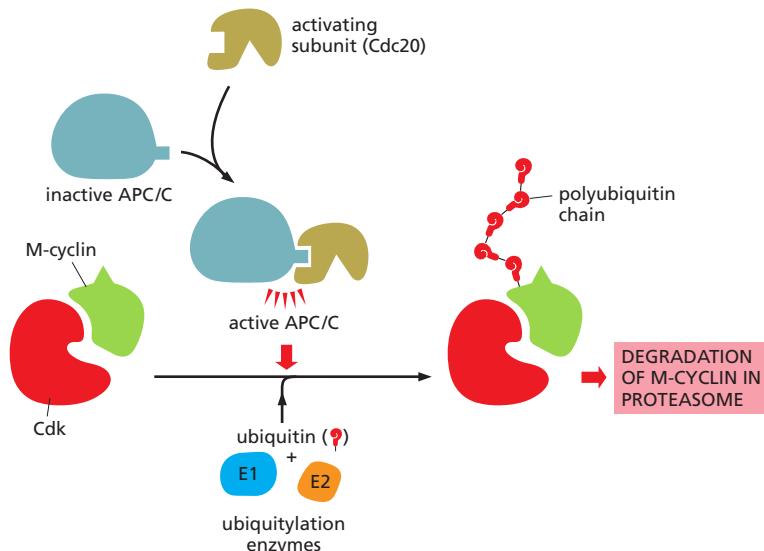
A variety of methods discussed in Chapter 8 have been used to analyze changes in the expression of all genes in the genome as the cell progresses through the cell cycle. The results of these studies are surprising. In budding yeast, for example, about 10% of the genes encode mRNAs whose levels oscillate during the cell cycle. Some of these genes encode proteins with known cell-cycle functions, but the functions of many others are unknown.

## The Cell-Cycle Control System Functions as a Network of Biochemical Switches

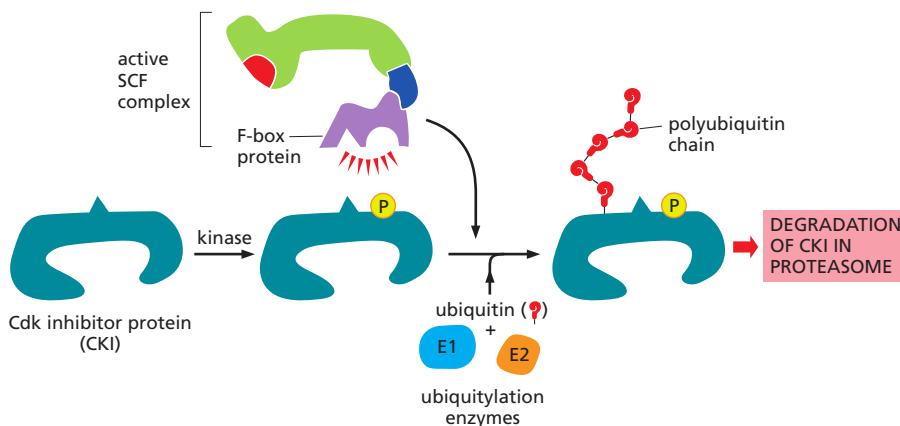
**Table 17–2** summarizes some of the major components of the cell-cycle control system. These proteins are functionally linked to form a robust network, which operates essentially autonomously to activate a series of biochemical switches, each of which triggers a specific cell-cycle event.

When conditions for cell proliferation are right, various external and internal signals stimulate the activation of G<sub>1</sub>-Cdk, which in turn stimulates the expression of genes encoding G<sub>1</sub>/S- and S-cyclins (Figure 17–16). The resulting activation of G<sub>1</sub>/S-Cdk then drives progression through the Start transition. By mechanisms we discuss later, G<sub>1</sub>/S-Cdks unleash a wave of S-Cdk activity, which initiates

(A) control of proteolysis by APC/C



(B) control of proteolysis by SCF

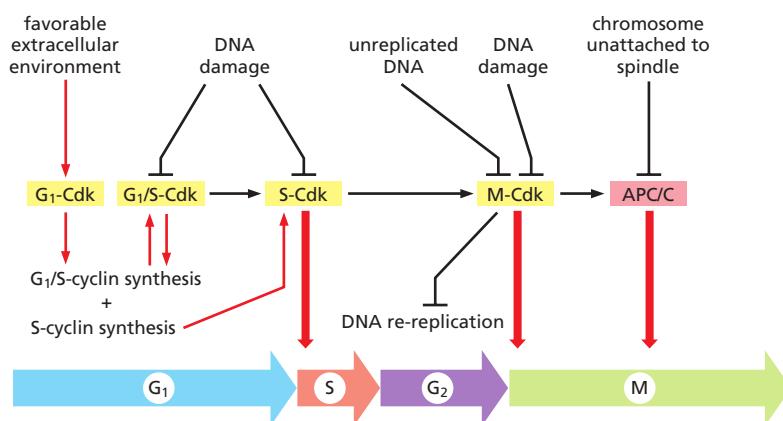


**Figure 17–15** The control of proteolysis by APC/C and SCF during the cell cycle. (A) The APC/C is activated in mitosis by association with Cdc20, which recognizes specific amino acid sequences on M-cyclin and other target proteins. With the help of two additional proteins called E1 and E2, the APC/C assembles polyubiquitin chains on the target protein. The polyubiquitylated target is then recognized and degraded in a proteasome. (B) The activity of the ubiquitin ligase SCF depends on substrate-binding subunits called F-box proteins, of which there are many different types. The phosphorylation of a target protein, such as the CKI shown, allows the target to be recognized by a specific F-box subunit.

**TABLE 17–2 Summary of the Major Cell Cycle Regulatory Proteins**

General name	Functions and comments
Protein kinases and protein phosphatases that modify Cdks	
Cdk-activating kinase (CAK)	Phosphorylates an activating site in Cdks
Wee1 kinase	Phosphorylates inhibitory sites in Cdks; primarily involved in suppressing Cdk1 activity before mitosis
Cdc25 phosphatase	Removes inhibitory phosphates from Cdks; three family members (Cdc25A, B, C) in mammals; primarily involved in controlling Cdk1 activation at the onset of mitosis
Cdk inhibitor proteins (CKIs)	
Sic1 (budding yeast)	Suppresses Cdk1 activity in G <sub>1</sub> ; phosphorylation by Cdk1 at the end of G <sub>1</sub> triggers its destruction
p27 (mammals)	Suppresses G <sub>1</sub> /S-Cdk and S-Cdk activities in G <sub>1</sub> ; helps cells withdraw from cell cycle when they terminally differentiate; phosphorylation by Cdk2 triggers its ubiquitylation by SCF
p21 (mammals)	Suppresses G <sub>1</sub> /S-Cdk and S-Cdk activities following DNA damage
p16 (mammals)	Suppresses G <sub>1</sub> -Cdk activity in G <sub>1</sub> ; frequently inactivated in cancer
Ubiquitin ligases and their activators	
APC/C	Catalyzes ubiquitylation of regulatory proteins involved primarily in exit from mitosis, including securin and S- and M-cyclins; regulated by association with activating subunits Cdc20 or Cdh1
Cdc20	APC/C-activating subunit in all cells; triggers initial activation of APC/C at metaphase-to-anaphase transition; stimulated by M-Cdk activity
Cdh1	APC/C-activating subunit that maintains APC/C activity after anaphase and throughout G <sub>1</sub> ; inhibited by Cdk activity
SCF	Catalyzes ubiquitylation of regulatory proteins involved in G <sub>1</sub> control, including some CKIs (Sic1 in budding yeast, p27 in mammals); phosphorylation of target protein usually required for this activity

chromosome duplication in S phase and also contributes to some early events of mitosis. M-Cdk activation then triggers progression through the G<sub>2</sub>/M transition and the events of early mitosis, leading to the alignment of sister-chromatid pairs at the equator of the mitotic spindle. Finally, the APC/C, together with its activator Cdc20, triggers the destruction of securin and cyclins, thereby unleashing sister-chromatid separation and segregation and the completion of mitosis. When mitosis is complete, multiple mechanisms collaborate to suppress Cdk activity, resulting in a stable G<sub>1</sub> period. We are now ready to discuss these cell-cycle stages in more detail, starting with S phase.



**Figure 17–16** An overview of the cell-cycle control system. The core of the cell-cycle control system consists of a series of cyclin-Cdk complexes (yellow). The activity of each complex is also influenced by various inhibitory mechanisms, which provide information about the extracellular environment, cell damage, and incomplete cell-cycle events (top). These inhibitory mechanisms are not present in all cell types; many are missing in early embryonic cell cycles, for example.

## Summary

The cell-cycle control system triggers the events of the cell cycle and ensures that they are properly timed and coordinated with each other. The control system responds to various intracellular and extracellular signals and arrests the cycle when the cell either fails to complete an essential cell-cycle process or encounters unfavorable environmental or intracellular conditions.

Central components of the control system are the cyclin-dependent protein kinases (Cdks), which depend on cyclin subunits for their activity. Oscillations in the activities of different cyclin-Cdk complexes control various cell-cycle events. Thus, activation of S-phase cyclin-Cdk complexes (S-Cdk) initiates S phase, whereas activation of M-phase cyclin-Cdk complexes (M-Cdk) triggers mitosis. The mechanisms that control the activities of cyclin-Cdk complexes include phosphorylation of the Cdk subunit, binding of Cdk inhibitor proteins (CKIs), proteolysis of cyclins, and changes in the transcription of genes encoding Cdk regulators. The cell-cycle control system also depends crucially on two additional enzyme complexes, the APC/C and SCF ubiquitin ligases, which catalyze the ubiquitylation and consequent destruction of specific regulatory proteins that control critical events in the cycle.

## S PHASE

The linear chromosomes of eukaryotic cells are vast and dynamic assemblies of DNA and protein, and their duplication is a complex process that takes up a major fraction of the cell cycle. Not only must the long DNA molecule of each chromosome be duplicated accurately—a remarkable feat in itself—but the protein packaging surrounding each region of that DNA must also be reproduced, ensuring that the daughter cells inherit all features of chromosome structure.

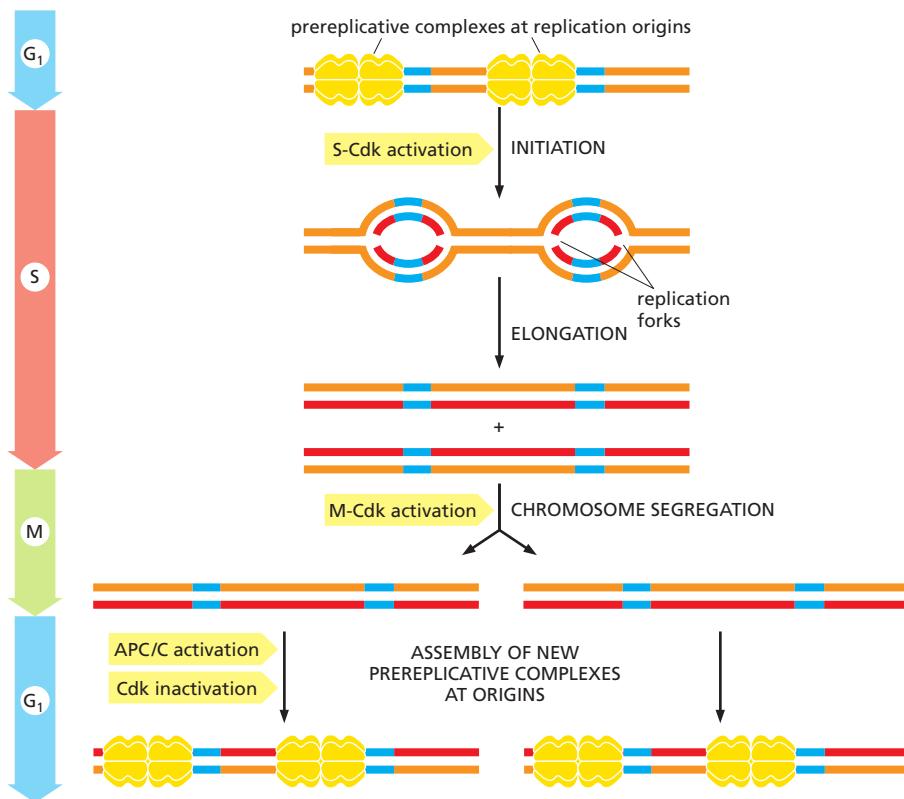
The central event of chromosome duplication—DNA replication—poses two problems for the cell. First, replication must occur with extreme accuracy to minimize the risk of mutations in the next cell generation. Second, every nucleotide in the genome must be copied once, and only once, to prevent the damaging effects of gene amplification. In Chapter 5, we discuss the sophisticated protein machinery that performs DNA replication with astonishing speed and accuracy. In this section, we consider the elegant mechanisms by which the cell-cycle control system initiates the replication process and, at the same time, prevents it from happening more than once per cycle.

### S-Cdk Initiates DNA Replication Once Per Cycle

DNA replication begins at *origins of replication*, which are scattered at numerous locations in every chromosome. During S phase, DNA replication is initiated at these origins when a *DNA helicase* unwinds the double helix and DNA replication enzymes are loaded onto the two single-stranded templates. This leads to the *elongation* phase of replication, when the replication machinery moves outward from the origin at two *replication forks* (discussed in Chapter 5).

To ensure that chromosome duplication occurs only once per cell cycle, the initiation phase of DNA replication is divided into two distinct steps that occur at different times in the cell cycle (**Figure 17–17**). The first step occurs in late mitosis and early G<sub>1</sub>, when a pair of inactive DNA helicases is loaded onto the replication origin, forming a large complex called the **prereplicative complex** or **preRC**. This step is sometimes called *licensing* of replication origins because initiation of DNA synthesis is permitted only at origins containing a preRC. The second step occurs in S phase, when the DNA helicases are activated, resulting in DNA unwinding and the initiation of DNA synthesis. Once a replication origin has been fired in this way, the two helicases move out from the origin with the replication forks, and that origin cannot be reused until a new preRC is assembled there at the end of mitosis. As a result, origins can be activated only once per cell cycle.

**Figure 17–18** illustrates some of the molecular details underlying the control of the two steps in the initiation of DNA replication. A key player is a large multiprotein complex called the **origin recognition complex** (ORC), which binds to



**Figure 17–17 Control of chromosome duplication.** Preparations for DNA replication begin in late mitosis and G<sub>1</sub>, when the DNA helicases are loaded by multiple proteins at the replication origin, forming the prereplicative complex (preRC). S-Cdk activation leads to activation of the DNA helicases, which unwind the DNA at origins to initiate DNA replication. Two replication forks move out from each origin until the entire chromosome is duplicated. Duplicated chromosomes are then segregated in M phase. S-Cdk activation in S phase also prevents assembly of new preRCs at any origin until the following G<sub>1</sub>—thereby ensuring that each origin is activated only once in each cell cycle.

replication origins throughout the cell cycle. In late mitosis and early G<sub>1</sub>, the proteins **Cdc6** and **Cdt1** collaborate with the ORC to load the inactive DNA helicases around the DNA next to the origin. The resulting large complex is the preRC, and the origin is now licensed for replication.

At the onset of S phase, S-Cdk triggers origin activation by phosphorylating specific initiator proteins, which then nucleate the assembly of a large protein complex that activates the DNA helicase and recruits the DNA synthesis machinery. Another protein kinase called DDK is also activated in S phase and helps drive origin activation by phosphorylating specific subunits of the DNA helicase.

At the same time as S-Cdk initiates DNA replication, several mechanisms prevent assembly of new preRCs. S-Cdk phosphorylates and thereby inhibits the ORC and Cdc6 proteins. Inactivation of the APC/C in late G<sub>1</sub> also helps turn off preRC assembly. In late mitosis and early G<sub>1</sub>, the APC/C triggers the destruction of a Cdt1 inhibitor called **geminin**, thereby allowing Cdt1 to be active. When the APC/C is turned off in late G<sub>1</sub>, geminin accumulates and inhibits the Cdt1 that is not associated with DNA. Also, the association of Cdt1 with a protein at active replication forks stimulates Cdt1 destruction. In these various ways, preRC formation is prevented from S phase to mitosis, thereby ensuring that each origin is fired only once per cell cycle. How, then, is the cell-cycle control system reset to allow replication in the next cell cycle? At the end of mitosis, APC/C activation leads to the inactivation of Cdks and the destruction of geminin. ORC and Cdc6 are dephosphorylated and Cdt1 is activated, allowing preRC assembly to prepare the cell for the next S phase.

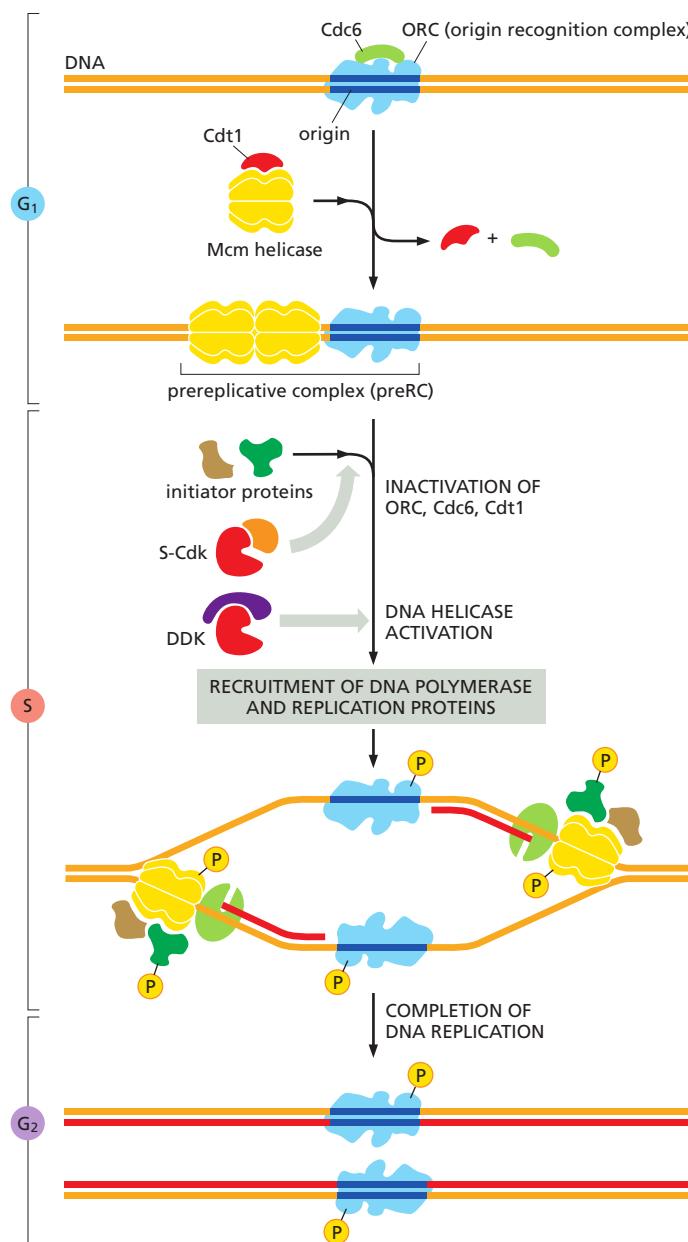
### Chromosome Duplication Requires Duplication of Chromatin Structure

The DNA of the chromosomes is extensively packaged in a variety of protein components, including histones and various regulatory proteins involved in the control of gene expression (discussed in Chapter 4). Thus, duplication of a chromosome is

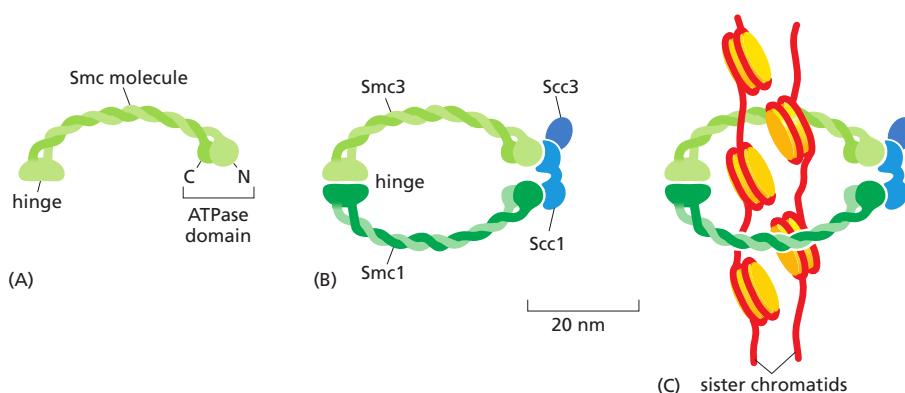
not simply a matter of replicating the DNA at its core but also requires the duplication of these chromatin proteins and their proper assembly on the DNA.

The production of chromatin proteins increases during S phase to provide the raw materials needed to package the newly synthesized DNA. Most importantly, S-Cdk's stimulate a large increase in the synthesis of the four histone subunits that form the histone octamers at the core of each nucleosome. These subunits are assembled into nucleosomes on the DNA by nucleosome assembly factors, which typically associate with the replication fork and distribute nucleosomes on both strands of the DNA as they emerge from the DNA synthesis machinery.

Chromatin packaging helps to control gene expression. In some parts of the chromosome, the chromatin is highly condensed and is called *heterochromatin*, whereas in other regions it has a more open structure and is called *euchromatin* (discussed in Chapter 4). These differences in chromatin structure depend on a variety of mechanisms, including modification of histone tails and the presence of non-histone proteins. Because these differences are important in gene regulation,



**Figure 17–18 Control of the initiation of DNA replication.** The replication origin is bound by the ORC throughout the cell cycle. In early G<sub>1</sub>, Cdc6 associates with the ORC, and these proteins bind the DNA helicase, which contains six closely related subunits called Mcm proteins. The helicase also associates with a protein called Cdt1. Using energy provided by ATP hydrolysis, the ORC and Cdc6 proteins load two copies of the DNA helicase, in an inactive form, around the DNA next to the origin, thereby forming the prereplicative complex (preRC). At the onset of S phase, S-Cdk stimulates the assembly of several initiator proteins on each DNA helicase, while another protein kinase, DDK, phosphorylates subunits of the DNA helicase. As a result, the DNA helicases are activated and unwind the DNA. DNA polymerase and other replication proteins are recruited to the origin, and DNA replication begins. The ORC is displaced by the replication machinery and then rebinds. S-Cdk and other mechanisms also inactivate the preRC components ORC, Cdc6, and Cdt1, thereby preventing formation of new preRCs at the origins until the end of mitosis (see text).



**Figure 17–19 Cohesin.** Cohesin is a protein complex with four subunits. (A) Two subunits, Smc1 and Smc3, are coiled-coil proteins with an ATPase domain at one end; (B) two additional subunits, Scc1 and Scc3, connect the ATPase head domains, forming a ring structure that may encircle the sister chromatids as shown in (C). The ATPase domains are required for cohesin loading on the DNA.

it is crucial that chromatin structure, like the DNA within, is reproduced accurately during S phase. How chromatin structure is reproduced is not well understood, however. During DNA synthesis, histone-modifying enzymes and various non-histone proteins are probably deposited onto the two new DNA strands as they emerge from the replication fork, and these proteins are thought to reproduce the local chromatin structure of the parent chromosome (see Figure 4–45).

### Cohesins Hold Sister Chromatids Together

At the end of S phase, each replicated chromosome consists of a pair of identical sister chromatids glued together along their length. This sister-chromatid cohesion sets the stage for a successful mitosis because it greatly facilitates the attachment of the two sister chromatids to opposite poles of the mitotic spindle. Imagine how difficult it would be to achieve this bipolar attachment if sister chromatids were allowed to drift apart after S phase. Indeed, defects in sister-chromatid cohesion—in yeast mutants, for example—lead inevitably to major errors in chromosome segregation.

Sister-chromatid cohesion depends on a large protein complex called **cohesin**, which is deposited at many locations along the length of each sister chromatid as the DNA is replicated in S phase. Two of the subunits of cohesin are members of a large family of proteins called *SMC proteins* (for Structural Maintenance of Chromosomes). Cohesin forms giant ringlike structures, and it has been proposed that these surround the two sister chromatids (Figure 17–19).

Sister-chromatid cohesion also results, at least in part, from *DNA catenation*, the intertwining of sister DNA molecules that occurs when two replication forks meet during DNA synthesis. The enzyme topoisomerase II gradually disentangles the catenated sister DNAs between S phase and early mitosis by cutting one DNA molecule, passing the other through the break, and then resealing the cut DNA (see Figure 5–22). Once the catenation has been removed, sister-chromatid cohesion depends primarily on cohesin complexes. The sudden and synchronous loss of sister cohesion at the metaphase-to-anaphase transition therefore depends primarily on disruption of these complexes, as we describe later.

### Summary

*Duplication of the chromosomes in S phase involves the accurate replication of the entire DNA molecule in each chromosome, as well as the duplication of the chromatin proteins that associate with the DNA and govern various aspects of chromosome function. Chromosome duplication is triggered by the activation of S-Cdk, which activates proteins that unwind the DNA and initiate its replication at replication origins. Once a replication origin is activated, S-Cdk also inhibits proteins that are required to allow that origin to initiate DNA replication again. Thus, each origin is fired once and only once in each S phase and cannot be reused until the next cell cycle.*

## MITOSIS

Following the completion of S phase and transition through G<sub>2</sub>, the cell undergoes the dramatic upheaval of M phase. This begins with mitosis, during which the sister chromatids are separated and distributed (*segregated*) to a pair of identical daughter nuclei, each with its own copy of the genome. Mitosis is traditionally divided into five stages—*prophase, prometaphase, metaphase, anaphase, and telophase*—defined primarily on the basis of chromosome behavior as seen in a microscope. As mitosis is completed, the second major event of M phase—cytokinesis—divides the cell into two halves, each with an identical nucleus. **Panel 17-1** summarizes the major events of M phase ([Movie 17.2](#), [Movie 17.3](#), [Movie 17.4](#), and [Movie 17.5](#)).

From a regulatory point of view, mitosis can be divided into two major parts, each governed by distinct components of the cell-cycle control system. First, an abrupt increase in M-Cdk activity at the G<sub>2</sub>/M transition triggers the events of early mitosis (prophase, prometaphase, and metaphase). During this period, M-Cdk and several other mitotic protein kinases phosphorylate a variety of proteins, leading to the assembly of the mitotic spindle and its attachment to the sister-chromatid pairs. The second major part of mitosis begins at the metaphase-to-anaphase transition, when the APC/C triggers the destruction of securin, liberating a protease that cleaves cohesin and thereby initiates separation of the sister chromatids. The APC/C also promotes the destruction of cyclins, which leads to Cdk inactivation and the dephosphorylation of Cdk targets, which is required for all events of late M phase, including the completion of anaphase, the disassembly of the mitotic spindle, and the division of the cell by cytokinesis.

In this section, we describe the key mechanical events of mitosis and how M-Cdk and the APC/C orchestrate them.

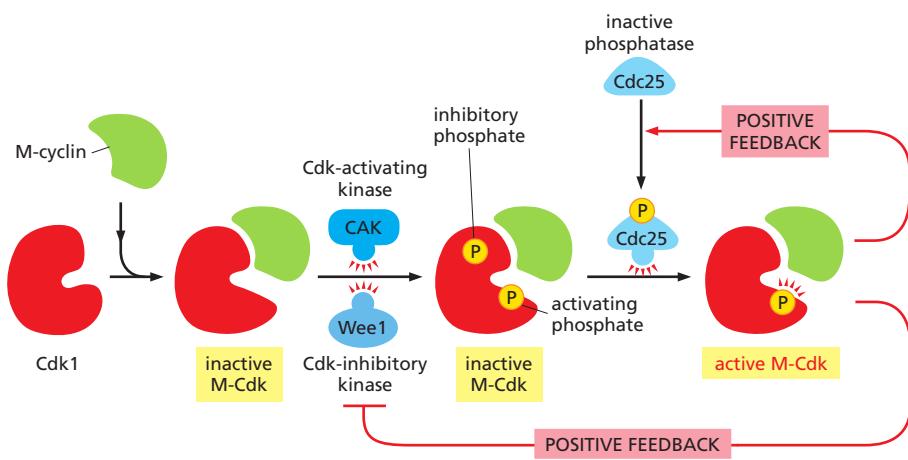
### M-Cdk Drives Entry Into Mitosis

One of the most remarkable features of cell-cycle control is that a single protein kinase, M-Cdk, brings about all of the diverse and complex cell rearrangements that occur in the early stages of mitosis. At a minimum, M-Cdk must induce the assembly of the mitotic spindle and ensure that each sister chromatid in a pair is attached to the opposite pole of the spindle. It also triggers *chromosome condensation*, the large-scale reorganization of the intertwined sister chromatids into compact, rodlike structures. In animal cells, M-Cdk also promotes the breakdown of the nuclear envelope and rearrangements of the actin cytoskeleton and the Golgi apparatus. Each of these processes is thought to be initiated when M-Cdk phosphorylates specific proteins involved in the process, although most of these proteins have not yet been identified.

M-Cdk does not act alone to phosphorylate key proteins involved in early mitosis. Two additional families of protein kinases, the *Polo-like kinases* and the *Aurora kinases*, also make important contributions to the control of early mitotic events. The Polo-like kinase Plk, for example, is required for the normal assembly of a bipolar mitotic spindle, in part because it phosphorylates proteins involved in separation of the spindle poles early in mitosis. The Aurora kinase Aurora-A also helps control proteins that govern the assembly and stability of the spindle, whereas Aurora-B controls attachment of sister chromatids to the spindle, as we discuss later.

### Dephosphorylation Activates M-Cdk at the Onset of Mitosis

M-Cdk activation begins with the accumulation of M-cyclin (cyclin B in vertebrate cells; see Table 17-1). In embryonic cell cycles, the synthesis of M-cyclin is constant throughout the cell cycle, and M-cyclin accumulation results from the high stability of the protein in interphase. In most cell types, however, M-cyclin synthesis increases during G<sub>2</sub> and M, owing primarily to an increase in M-cyclin gene transcription. The increase in M-cyclin protein leads to a corresponding accumulation of M-Cdk (the complex of Cdk1 and M-cyclin) as the cell approaches



mitosis. Although the Cdk in these complexes is phosphorylated at an activating site by the Cdk-activating kinase (CAK), as discussed earlier, the protein kinase Wee1 holds it in an inactive state by inhibitory phosphorylation at two neighboring sites (see Figure 17–13). Thus, by the time the cell reaches the end of G<sub>2</sub>, it contains an abundant stockpile of M-Cdk that is primed and ready to act but is suppressed by phosphates that block the active site of the kinase.

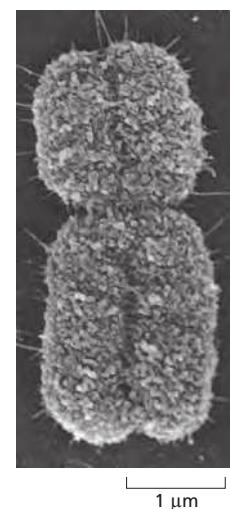
What, then, triggers the activation of the M-Cdk stockpile? The crucial event is the activation of the protein phosphatase Cdc25, which removes the inhibitory phosphates that restrain M-Cdk (Figure 17–20). At the same time, the inhibitory activity of the kinase Wee1 is suppressed, further ensuring that M-Cdk activity increases. The mechanisms that unleash Cdc25 activity in early mitosis are not well understood. One possibility is that the S-Cdks that are active in G<sub>2</sub> and early prophase stimulate Cdc25.

Interestingly, Cdc25 can also be activated, at least in part, by its target, M-Cdk. M-Cdk may also inhibit the inhibitory kinase Wee1. The ability of M-Cdk to activate its own activator (Cdc25) and inhibit its own inhibitor (Wee1) suggests that M-Cdk activation in mitosis involves positive feedback loops (see Figure 17–20). According to this attractive model, the partial activation of Cdc25 (perhaps by S-Cdk) leads to the partial activation of a subpopulation of M-Cdk complexes, which then phosphorylate Cdc25 and Wee1 molecules. This leads to more M-Cdk activation, and so on. Such a mechanism would quickly promote the activation of all M-Cdk complexes in the cell. As mentioned earlier, similar molecular switches operate at various points in the cell cycle to promote the abrupt and complete transition from one cell-cycle state to the next.

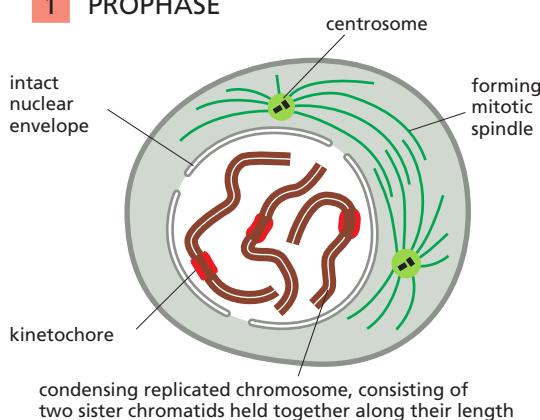
### Condensin Helps Configure Duplicated Chromosomes for Separation

At the end of S phase, the immensely long DNA molecules of the sister chromatids are tangled in a mass of partially catenated DNA and proteins. Any attempt to pull the sisters apart in this state would undoubtedly lead to breaks in the chromosomes. To avoid this disaster, the cell devotes a great deal of energy in early mitosis to gradually reorganizing the sister chromatids into relatively short, distinct structures that can be pulled apart more easily in anaphase. These chromosomal changes involve two processes: *chromosome condensation*, in which the chromatids are dramatically compacted; and *sister-chromatid resolution*, whereby the two sisters are resolved into distinct, separable units (Figure 17–21). Resolution results from the decatenation of the sister DNAs, accompanied by the partial removal of cohesin molecules along the chromosome arms. As a result, when the cell reaches metaphase, the sister chromatids appear in the microscope as compact, rodlike structures that are joined tightly at their centromeric regions and only loosely along their arms.

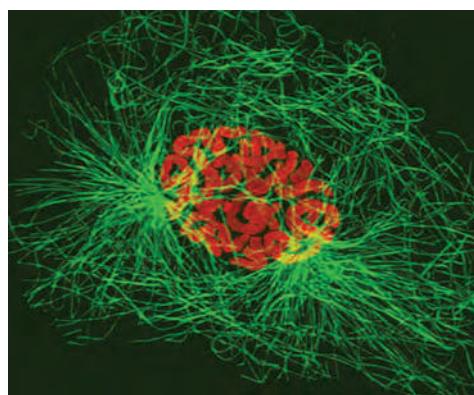
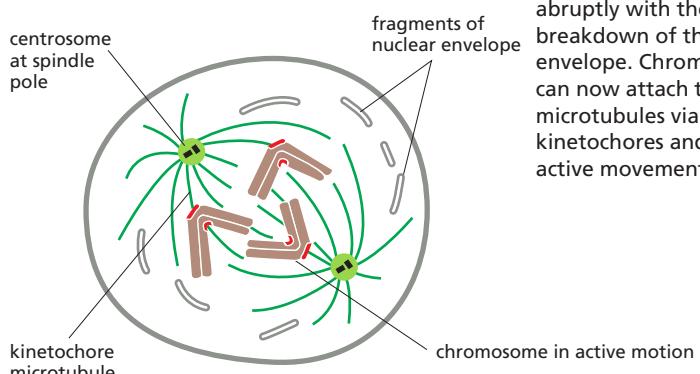
**Figure 17–20** The activation of M-Cdk. Cdk1 associates with M-cyclin as the levels of M-cyclin gradually rise. The resulting M-Cdk complex is phosphorylated on an activating site by the Cdk-activating kinase (CAK) and on a pair of inhibitory sites by the Wee1 kinase. The resulting inactive M-Cdk complex is then activated at the end of G<sub>2</sub> by the phosphatase Cdc25. Cdc25 is further stimulated by active M-Cdk, resulting in positive feedback. This feedback is enhanced by the ability of M-Cdk to inhibit Wee1.



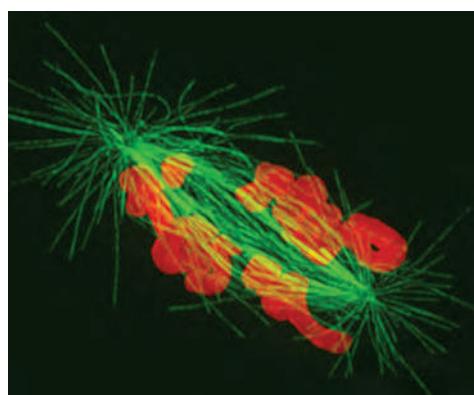
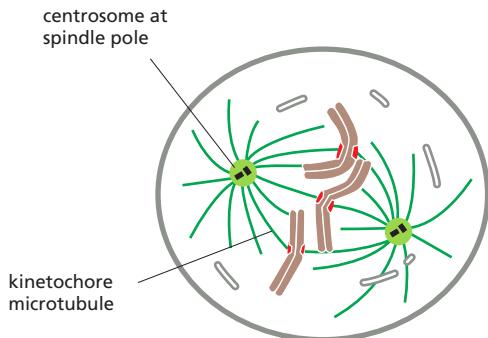
**Figure 17–21** The mitotic chromosome. Scanning electron micrograph of a human mitotic chromosome, consisting of two sister chromatids joined along their length. The constricted regions are the centromeres. (Courtesy of Terry D. Allen.)

**1 PROPHASE**

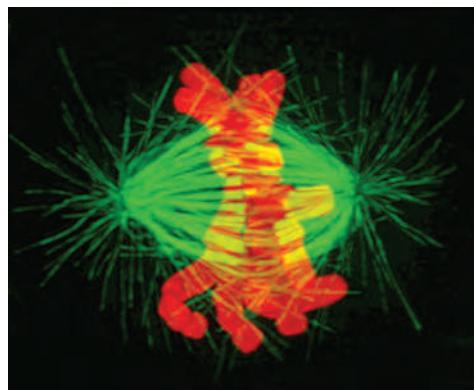
At **prophase**, the replicated chromosomes, each consisting of two closely associated sister chromatids, condense. Outside the nucleus, the mitotic spindle assembles between the two centrosomes, which have replicated and moved apart. For simplicity, only three chromosomes are shown. In diploid cells, there would be two copies of each chromosome present. In the fluorescence micrograph, chromosomes are stained orange and microtubules are green.

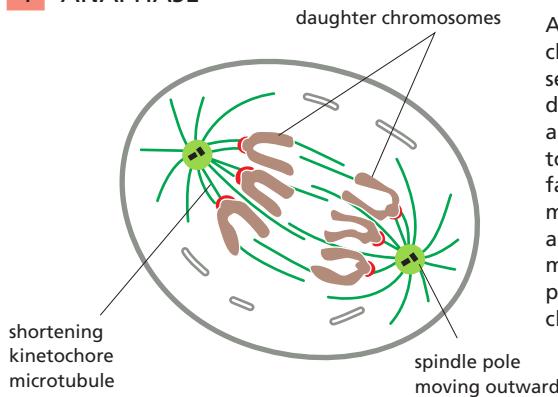
**2 PROMETAPHASE**

**Prometaphase** starts abruptly with the breakdown of the nuclear envelope. Chromosomes can now attach to spindle microtubules via their kinetochores and undergo active movement.

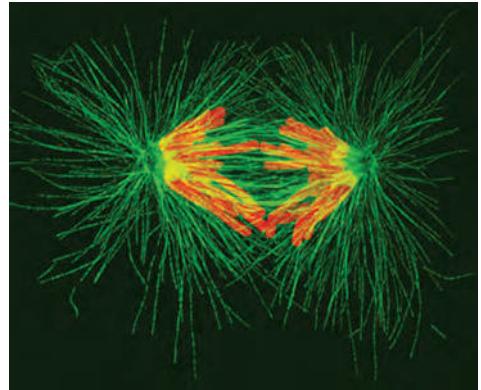
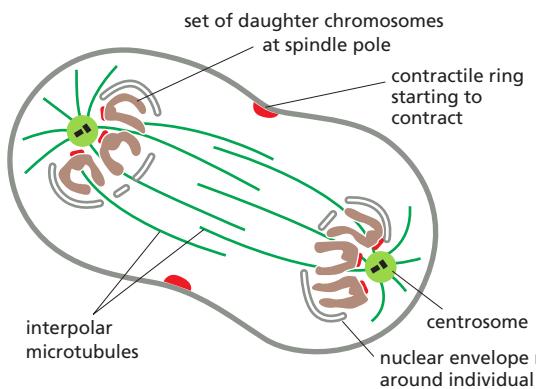
**3 METAPHASE**

At **metaphase**, the chromosomes are aligned at the equator of the spindle, midway between the spindle poles. The kinetochore microtubules attach sister chromatids to opposite poles of the spindle.

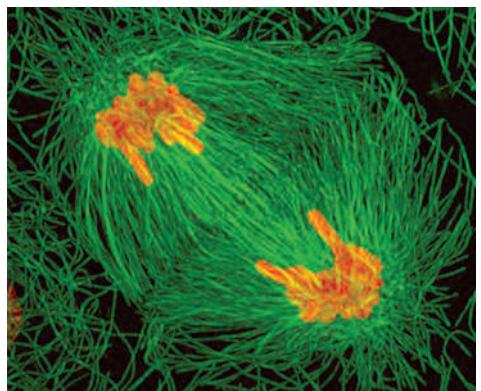
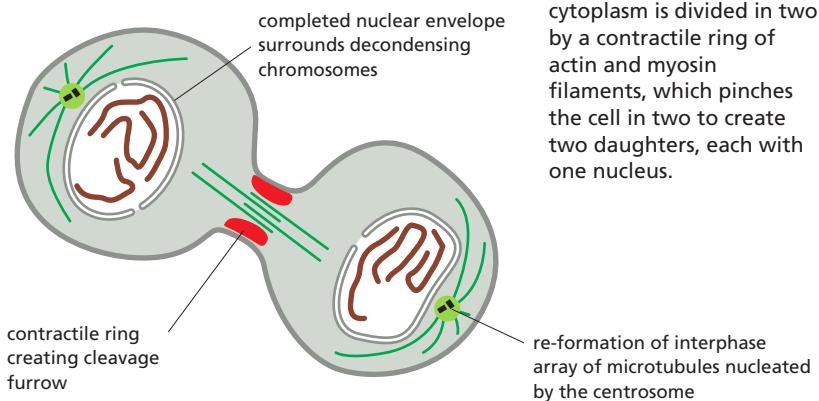


**4 ANAPHASE**

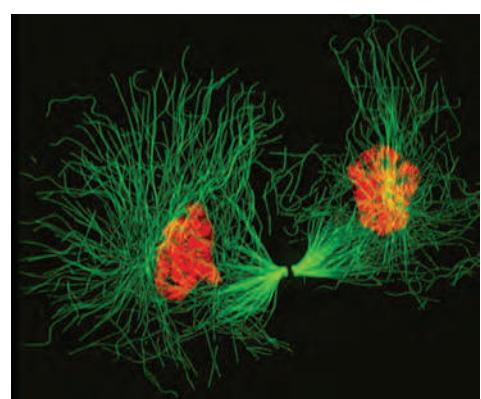
At **anaphase**, the sister chromatids synchronously separate to form two daughter chromosomes, and each is pulled slowly toward the spindle pole it faces. The kinetochore microtubules get shorter, and the spindle poles also move apart; both processes contribute to chromosome segregation.

**5 TELOPHASE**

During **telophase**, the two sets of daughter chromosomes arrive at the poles of the spindle and decondense. A new nuclear envelope reassembles around each set, completing the formation of two nuclei and marking the end of mitosis. The division of the cytoplasm begins with contraction of the contractile ring.

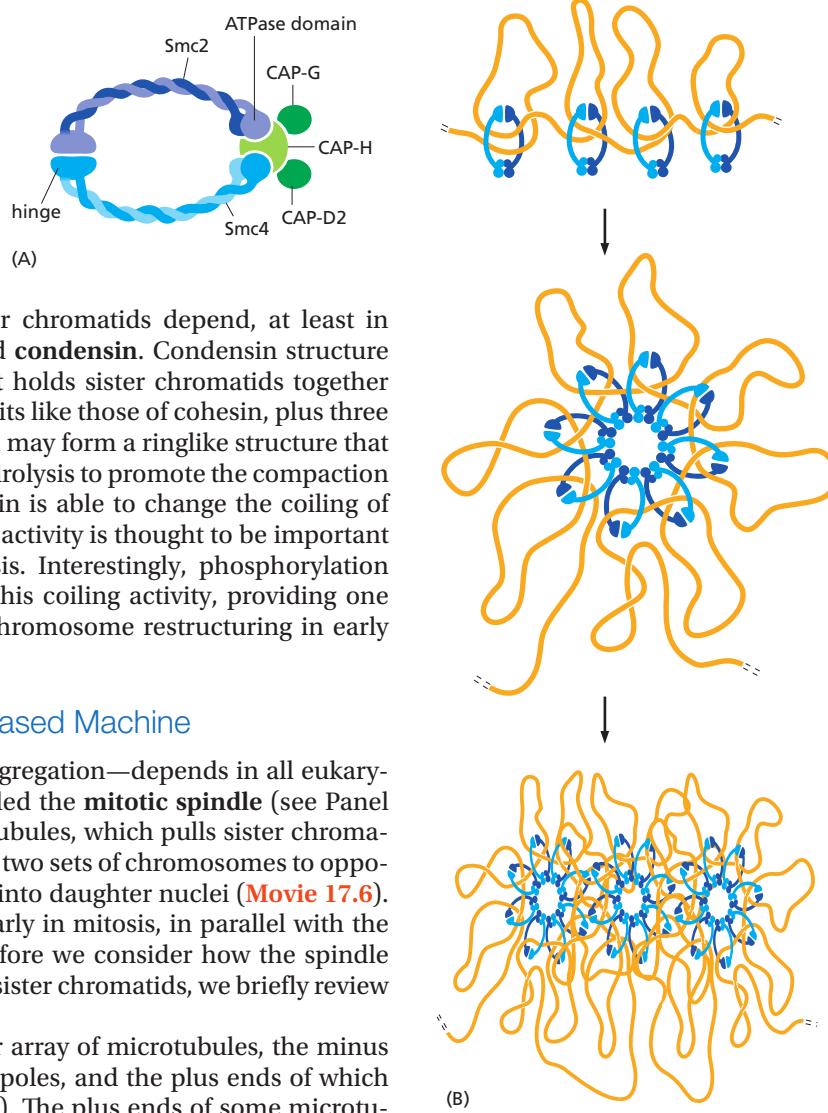
**6 CYTOKINESIS**

During **cytokinesis**, the cytoplasm is divided in two by a contractile ring of actin and myosin filaments, which pinches the cell in two to create two daughters, each with one nucleus.



(Micrographs courtesy of Julie Canman and Ted Salmon.)

**Figure 17–22 Condensin.** (A) Condensin is a five-subunit protein complex that resembles cohesin (see Figure 17–19). The ATPase head domains of its two major subunits, Smc2 and Smc4, are held together by three additional subunits. (B) It is not clear how condensin catalyzes the restructuring and compaction of chromosome DNA, but it may form a ring structure that encircles loops of DNA within each sister chromatid.



The condensation and resolution of sister chromatids depend, at least in part, on a five-subunit protein complex called **condensin**. Condensin structure is related to that of the cohesin complex that holds sister chromatids together (see Figure 17–19). It contains two SMC subunits like those of cohesin, plus three non-SMC subunits (Figure 17–22). Condensin may form a ringlike structure that somehow uses the energy provided by ATP hydrolysis to promote the compaction and resolution of sister chromatids. Condensin is able to change the coiling of DNA molecules in a test tube, and this coiling activity is thought to be important for chromosome condensation during mitosis. Interestingly, phosphorylation of condensin subunits by M-Cdk stimulates this coiling activity, providing one mechanism by which M-Cdk may promote chromosome restructuring in early mitosis.

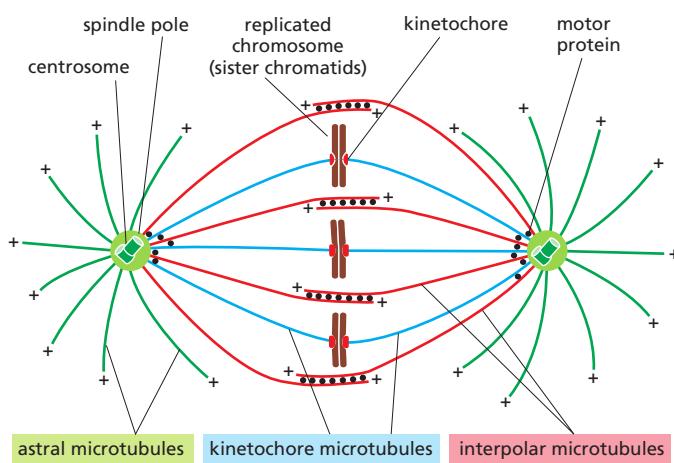
### The Mitotic Spindle Is a Microtubule-Based Machine

The central event of mitosis—chromosome segregation—depends in all eukaryotes on a complex and beautiful machine called the **mitotic spindle** (see Panel 17–1). The spindle is a bipolar array of microtubules, which pulls sister chromatids apart in anaphase, thereby segregating the two sets of chromosomes to opposite ends of the cell, where they are packaged into daughter nuclei (Movie 17.6). M-Cdk triggers the assembly of the spindle early in mitosis, in parallel with the chromosome restructuring just described. Before we consider how the spindle assembles and how its microtubules attach to sister chromatids, we briefly review the basic features of spindle structure.

The core of the mitotic spindle is a bipolar array of microtubules, the minus ends of which are focused at the two spindle poles, and the plus ends of which radiate outward from the poles (Figure 17–23). The plus ends of some microtubules—called the **interpolar microtubules**—overlap with the plus ends of microtubules from the other pole, resulting in an antiparallel array in the spindle mid-zone. The plus ends of other microtubules—the **kinetochore microtubules**—are attached to sister-chromatid pairs at large protein structures called *kinetochores*, which are located at the *centromere* of each sister chromatid. Finally, many spindles also contain **astral microtubules** that radiate outward from the poles and contact the cell cortex, helping to position the spindle in the cell.

In most somatic animal cells, each spindle pole is focused at a protein organelle called the **centrosome** (see Figures 16–47 and 16–48). Each centrosome consists of a cloud of amorphous material (called the *pericentriolar matrix*) that surrounds a pair of *centrioles* (Figure 17–24). The pericentriolar matrix nucleates a radial array of microtubules, with their fast-growing plus ends projecting outward and their minus ends associated with the centrosome. The matrix contains a variety of proteins, including microtubule-dependent motor proteins, coiled-coil proteins that link the motors to the centrosome, structural proteins, and components of the cell-cycle control system. Most important, it contains *γ-tubulin ring complexes*, which are the components mainly responsible for nucleating microtubules (see Figure 16–46).

Some cells—notably the cells of higher plants and the oocytes of many vertebrates—do not have centrosomes, and microtubule-dependent motor proteins and other proteins associate with microtubule minus ends to organize and focus the spindle poles.

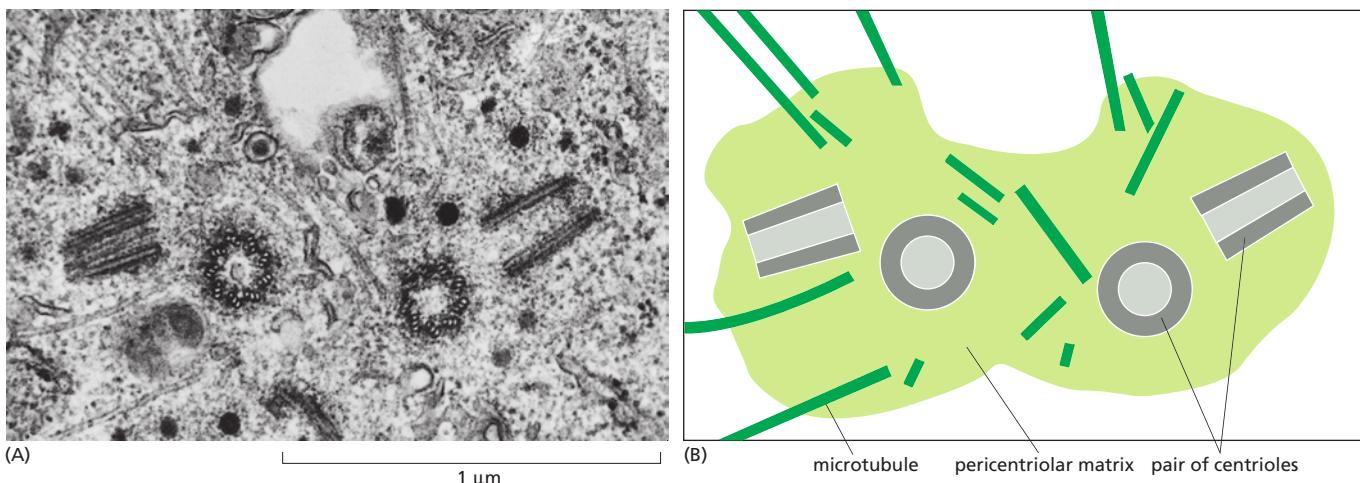


**Figure 17–23** The metaphase mitotic spindle in an animal cell. The plus ends of the microtubules project away from the spindle pole, while the minus ends are anchored at the spindle poles, which in this example are organized by centrosomes. Kinetochore microtubules connect the spindle poles with the kinetochores of sister chromatids, while interpolar microtubules from the two poles interdigitate at the spindle equator. Astral microtubules radiate out from the poles into the cytoplasm.

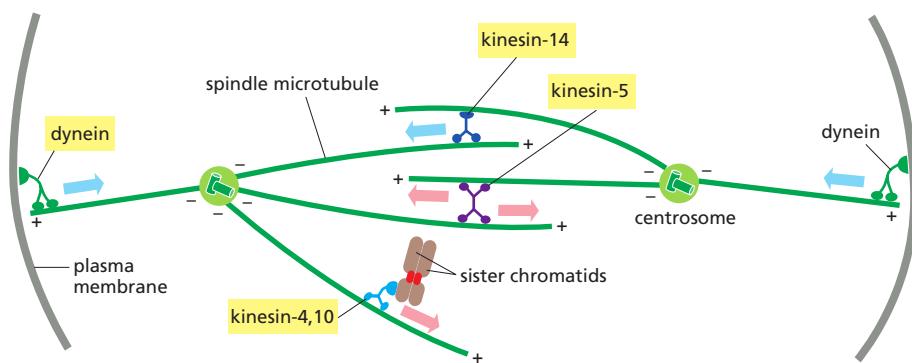
### Microtubule-Dependent Motor Proteins Govern Spindle Assembly and Function

The function of the mitotic spindle depends on numerous microtubule-dependent motor proteins. As discussed in Chapter 16, these proteins belong to two families—the kinesin-related proteins, which usually move toward the plus end of microtubules, and dyneins, which move toward the minus end. In the mitotic spindle, these motor proteins generally operate at or near the ends of the microtubules. Four major types of motor proteins—*kinesin-5*, *kinesin-14*, *kinesins-4/10*, and *dynein*—are particularly important in spindle assembly and function (Figure 17–25).

Kinesin-5 proteins contain two motor domains that interact with the plus ends of antiparallel microtubules in the spindle midzone. Because the two motor domains move toward the plus ends of the microtubules, they slide the two antiparallel microtubules past each other toward the spindle poles, pushing the poles apart. Kinesin-14 proteins, by contrast, are minus-end directed motors with a single motor domain and other domains that can interact with a neighboring microtubule. They can cross-link antiparallel interpolar microtubules at the spindle midzone and tend to pull the poles together. Kinesin-4 and kinesin-10



**Figure 17–24** The centrosome. (A) Electron micrograph of an S-phase mammalian cell in culture, showing a duplicated centrosome. Each centrosome contains a pair of centrioles; although the centrioles have duplicated, they remain together in a single complex, as shown in the drawing of the micrograph in (B). One centriole of each centriole pair has been cut in cross section, while the other is cut in longitudinal section, indicating that the two members of each pair are aligned at right angles to each other. The two halves of the replicated centrosome, each consisting of a centriole pair surrounded by pericentriolar matrix, will split and migrate apart to initiate the formation of the two poles of the mitotic spindle when the cell enters M phase. (A, from M. McGill, D.P. Highfield, T.M. Monahan, and B.R. Brinkley, *J. Ultrastruct. Res.* 57:43–53, 1976. With permission from Academic Press.)



**Figure 17–25 Major motor proteins of the spindle.** Four major classes of microtubule-dependent motor proteins (yellow boxes) contribute to spindle assembly and function (see text). The colored arrows indicate the direction of motor protein movement along a microtubule—blue toward the minus end and red toward the plus end.

proteins, also called *chromokinesins*, are plus-end directed motors that associate with chromosome arms and push the attached chromosome away from the pole (or the pole away from the chromosome). Finally, dyneins are minus-end directed motors that, together with associated proteins, organize microtubules at various locations in the cell. They link the plus ends of astral microtubules to components of the actin cytoskeleton at the cell cortex, for example; by moving toward the minus end of the microtubules, the dynein motors pull the spindle poles toward the cell cortex and away from each other.

### Multiple Mechanisms Collaborate in the Assembly of a Bipolar Mitotic Spindle

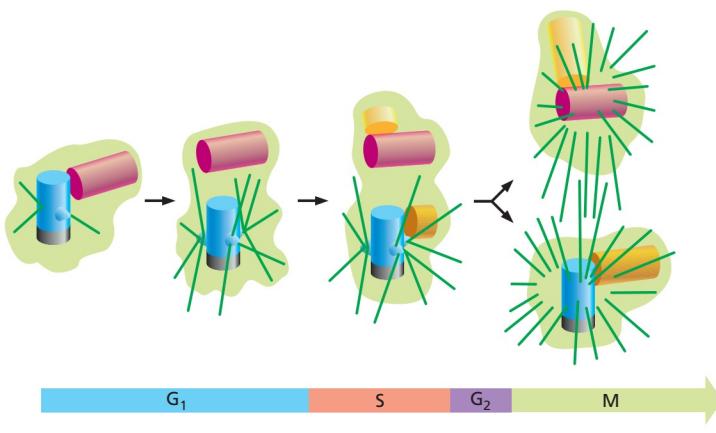
The mitotic spindle must have two poles if it is to pull the two sets of sister chromatids to opposite ends of the cell in anaphase. In most animal cells, several mechanisms ensure the bipolarity of the spindle. One depends on centrosomes. A typical animal cell enters mitosis with a pair of centrosomes, each of which nucleates a radial array of microtubules. The two centrosomes provide prefabricated spindle poles that greatly facilitate bipolar spindle assembly. The other mechanisms depend on the ability of mitotic chromosomes to nucleate and stabilize microtubules and on the ability of motor proteins to organize microtubules into a bipolar array. These “self-organization” mechanisms can produce a bipolar spindle even in cells lacking centrosomes.

We now describe the steps of spindle assembly, beginning with centrosome-dependent assembly in early mitosis. We then consider the self-organization mechanisms that do not require centrosomes and become particularly important after nuclear-envelope breakdown.

### Centrosome Duplication Occurs Early in the Cell Cycle

Most animal cells contain a single centrosome that nucleates most of the cell's cytoplasmic microtubules. The centrosome duplicates when the cell enters the cell cycle, so that by the time the cell reaches mitosis there are two centrosomes. Centrosome duplication begins at about the same time as the cell enters S phase. The G<sub>1</sub>/S-Cdk (a complex of cyclin E and Cdk2 in animal cells; see Table 17–1) that triggers cell-cycle entry also helps initiate centrosome duplication. The two centrioles in the centrosome separate, and each nucleates the formation of a single new centriole, resulting in two centriole pairs within an enlarged pericentriolar matrix (Figure 17–26). This centrosome pair remains together on one side of the nucleus until the cell enters mitosis.

There are interesting parallels between centrosome duplication and chromosome duplication. Both use a semiconservative mechanism of duplication, in which the two halves separate and serve as templates for construction of a new half. Centrosomes, like chromosomes, must replicate once and only once per cell cycle, to ensure that the cell enters mitosis with only two copies: an incorrect number of centrosomes could lead to defects in spindle assembly and thus errors in chromosome segregation.



**Figure 17–26 Centriole replication.** The centrosome consists of a centriole pair and associated pericentriolar matrix (green). At a certain point in G<sub>1</sub>, the two centrioles of the pair separate by a few micrometers. During S phase, a daughter centriole begins to grow near the base of each mother centriole and at a right angle to it. The elongation of the daughter centriole is usually completed in G<sub>2</sub>. The two centriole pairs remain close together in a single centrosomal complex until the beginning of M phase, when the complex splits in two and the two daughter centrosomes begin to separate. Each centrosome now nucleates its own radial array of microtubules (called an aster), mainly from the mother centriole.

The mechanisms that limit centrosome duplication to once per cell cycle are uncertain. In many cell types, experimental inhibition of DNA synthesis blocks centrosome duplication, providing one mechanism by which centrosome number is kept in check. Other cell types, however, including those in the early embryos of flies, sea urchins, and frogs, do not have such a mechanism and centrosome duplication continues if chromosome duplication is blocked. It is not known how such cells limit centrosome duplication to once per cell cycle.

### M-Cdk Initiates Spindle Assembly in Prophase

Spindle assembly begins in early mitosis, when the two centrosomes move apart along the nuclear envelope, pulled by dynein motor proteins that link astral microtubules to the cell cortex (see Figure 17–25). The plus ends of the microtubules between the centrosomes interdigitate to form the interpolar microtubules, and kinesin-5 motor proteins associate with these microtubules and push the centrosomes apart (see Figure 17–25). Also in early mitosis, the number of  $\gamma$ -tubulin ring complexes in each centrosome increases greatly, increasing the ability of the centrosomes to nucleate new microtubules, a process called *centrosome maturation*.

The balance of opposing forces generated by different types of motor proteins determines the final length of the spindle. Dynein and kinesin-5 motors generally promote centrosome separation and increase spindle length. Kinesin-14 proteins do the opposite: they tend to pull the poles together (see Figure 17–25). It is not clear how the cell regulates the balance of opposing forces to generate the appropriate spindle length.

M-Cdk and other mitotic protein kinases are required for centrosome separation and maturation. M-Cdk and Aurora-A phosphorylate kinesin-5 motors and stimulate them to drive centrosome separation. Aurora-A and Plk also phosphorylate components of the centrosome and thereby promote its maturation.

### The Completion of Spindle Assembly in Animal Cells Requires Nuclear-Envelope Breakdown

The centrosomes and microtubules of animal cells are located in the cytoplasm, separated from the chromosomes by the double-membrane barrier of the nuclear envelope (discussed in Chapter 12). Clearly, the attachment of sister-chromatid pairs to the spindle requires the removal of this barrier. In addition, many of the motor proteins and microtubule regulators that promote spindle assembly are associated with the chromosomes inside the nucleus, and they require nuclear-envelope breakdown to carry out their functions.

Nuclear-envelope breakdown is a complex, multistep process, which is thought to begin when M-Cdk phosphorylates several subunits of the nuclear pore complexes in the nuclear envelope. This phosphorylation initiates the disassembly of nuclear pore complexes and their dissociation from the envelope. M-Cdk

also phosphorylates components of the nuclear lamina, the structural framework beneath the envelope. The phosphorylation of these lamina components and of several inner-nuclear-envelope proteins leads to disassembly of the nuclear lamina and the breakdown of the envelope membranes into small vesicles.

### Microtubule Instability Increases Greatly in Mitosis

Most animal cells in interphase contain a cytoplasmic array of microtubules radiating out from the single centrosome. As discussed in Chapter 16, the microtubules of this interphase array are in a state of *dynamic instability*, in which individual microtubules are either growing or shrinking and stochastically switch between the two states. The switch from growth to shrinkage is called a *catastrophe*, and the switch from shrinkage to growth is called a *rescue*. New microtubules are continually being created to balance the loss of those that disappear completely by depolymerization.

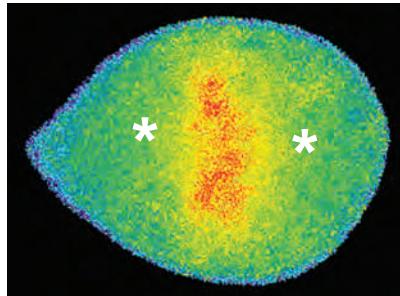
Entry into mitosis signals an abrupt change in the cell's microtubules. The interphase array of few, long microtubules radiating from the single centrosome is converted to a larger number of shorter and more dynamic microtubules emanating from both centrosomes. During prophase, and particularly in prometaphase and metaphase (see Panel 17–1), the half-life of microtubules decreases dramatically. This increase in microtubule instability, coupled with the increased ability of centrosomes to nucleate microtubules as mentioned earlier, results in remarkably dense and dynamic arrays of spindle microtubules that are ideally suited for capturing sister chromatids.

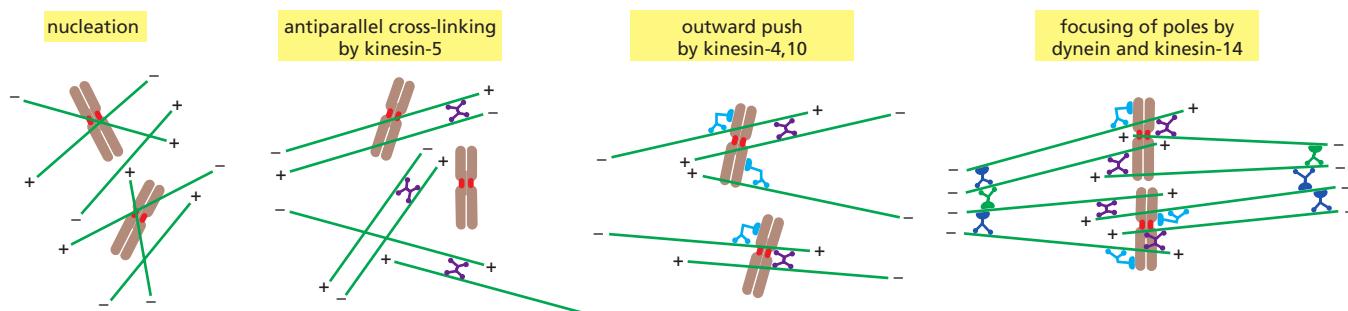
Microtubule dynamics are controlled in the cell by a variety of regulatory proteins, including microtubule-associated proteins (MAPs) that promote stability and catastrophe factors that destabilize microtubule plus ends. Changes in the activities of these regulatory proteins are responsible for the changes in microtubule dynamics that occur during mitosis. Many of these changes result from phosphorylation of specific proteins by M-Cdk and other mitotic protein kinases.

### Mitotic Chromosomes Promote Bipolar Spindle Assembly

Chromosomes are not just passive passengers in the process of spindle assembly. By creating a local environment that favors both microtubule nucleation and microtubule stabilization, they play an active part in spindle formation. The influence of the chromosomes can be demonstrated by using a fine glass needle to reposition them after the spindle has formed. For some cells in metaphase, if a single chromosome is tugged out of alignment, a mass of new spindle microtubules rapidly appears around the newly positioned chromosome, while the spindle microtubules at the chromosome's former position depolymerize. This property of the chromosomes seems to depend, at least in part, on a guanine nucleotide exchange factor (GEF) that is bound to chromatin; the GEF stimulates a small GTPase in the cytosol called *Ran* to bind GTP in place of GDP. The activated Ran-GTP, which is also involved in nuclear transport (discussed in Chapter 12), releases microtubule-stabilizing proteins from protein complexes in the cytosol, thereby stimulating the local nucleation and stabilization of microtubules around chromosomes (Figure 17–27). Local microtubule stabilization is also promoted by the protein kinase Aurora-B, which associates with mitotic chromosomes.

**Figure 17–27 Activation of the GTPase Ran around mitotic chromosomes.** The Ran protein, like other members of the small GTPase family (discussed in Chapter 15), can exist in two conformations depending on whether it is bound to GDP (inactive state) or GTP (active state). The localization of active Ran in mitosis was determined using a protein that emits fluorescence at a specific wavelength when it is activated by Ran-GTP. In the metaphase human cell shown here, Ran activity (yellow and red) is highest around the chromosomes, between the poles of the mitotic spindle (indicated by asterisks). (From P. Kaláb et al., *Nature* 440:697–701, 2006. With permission from Macmillan Publishers Ltd.)





**Figure 17–28** Spindle self-organization by motor proteins. Mitotic chromosomes stimulate the local activation of proteins that nucleate and promote the formation of microtubules in the vicinity of the chromosomes. Kinesin-5 motor proteins (see Figure 17–25) organize these microtubules into antiparallel bundles, while plus-end directed kinesins-4 and 10 link the microtubules to chromosome arms and push minus ends away from the chromosomes. Dynein and kinesin-14 motors, together with numerous other proteins, focus these minus ends into a pair of spindle poles.

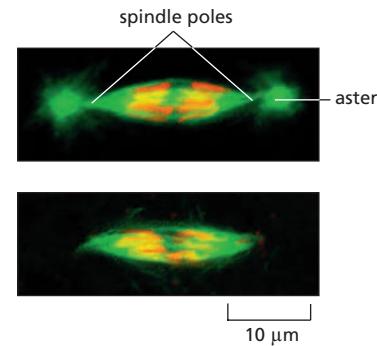
The ability of chromosomes to stabilize and organize microtubules enables cells to form bipolar spindles in the absence of centrosomes. Acentrosomal spindle assembly is thought to begin with the formation of microtubules around the chromosomes. Various motor proteins then organize the microtubules into a bipolar spindle, as illustrated in **Figure 17–28**.

Cells that normally lack centrosomes, such as those of higher plants and many animal oocytes, use this chromosome-based self-organization process to form spindles. It is also the process used to assemble spindles in certain animal embryos that have been induced to develop from eggs without fertilization (that is, *parthenogenetically*); as the sperm normally provides the centrosome when it fertilizes an egg, the mitotic spindles in these parthenogenetic embryos develop without centrosomes (**Figure 17–29**). Even in cells that normally contain centrosomes, the chromosomes help organize the spindle microtubules and, with the help of various motor proteins, can promote the assembly of a bipolar mitotic spindle if the centrosomes are removed. Although the resulting acentrosomal spindle can segregate chromosomes normally, it lacks astral microtubules, which are responsible for positioning the spindle in animal cells; as a result, the spindle is often mispositioned in the cell.

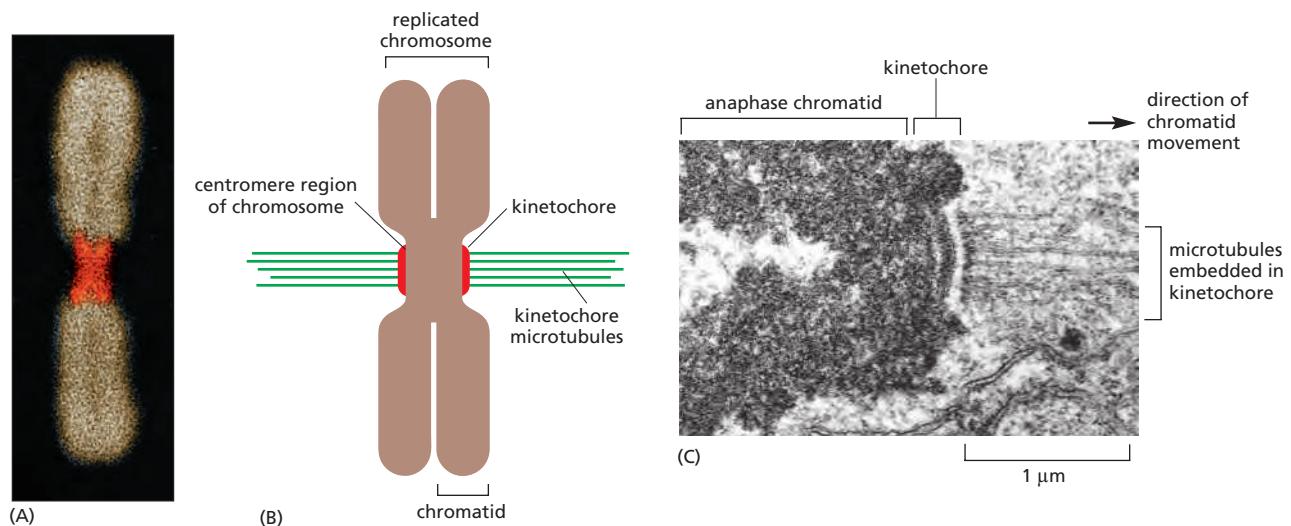
### Kinetochores Attach Sister Chromatids to the Spindle

Following the assembly of a bipolar microtubule array, the second major step in spindle formation is the attachment of the array to the sister-chromatid pairs. Spindle microtubules become attached to each chromatid at its **kinetochore**, a giant, multilayered protein structure that is built at the centromeric region of the chromatid (**Figure 17–30**; also see Chapter 4). In metaphase, the plus ends of kinetochore microtubules are embedded head-on in specialized microtubule-attachment sites within the outer region of the kinetochore, furthest from the DNA. The kinetochore of an animal cell can bind 10–40 microtubules, whereas a budding yeast kinetochore can bind only one. Attachment of each microtubule depends on multiple copies of a rod-shaped protein complex called the Ndc80 complex, which is anchored in the kinetochore at one end and interacts with the sides of the microtubule at the other, thereby linking the microtubule to the kinetochore while still allowing the addition or removal of tubulin subunits at this end (**Figure 17–31**). Regulation of plus-end polymerization and depolymerization at the kinetochore is critical for the control of chromosome movement on the spindle, as we discuss later.

Kinetochore attachment to the spindle occurs by a complex sequence of events. At the end of prophase in animal cells, the centrosomes of the growing spindle generally lie on opposite sides of the nuclear envelope. Thus, when the envelope breaks down, the sister-chromatid pairs are bombarded by microtubule



**Figure 17–29** Bipolar spindle assembly without centrosomes in parthenogenetic embryos of the insect *Sciara* (or fungus gnat). The microtubules are stained green, the chromosomes red. The top fluorescence micrograph shows a normal spindle formed with centrosomes in a normally fertilized *Sciara* embryo. The bottom micrograph shows a spindle formed without centrosomes in an embryo that initiated development without fertilization. Note that the spindle with centrosomes has an aster at each pole of the spindle, whereas the spindle formed without centrosomes does not. Both types of spindles are able to segregate the replicated chromosomes. (From B. de Saint Phalle and W. Sullivan, *J. Cell Biol.* 141:1383–1391, 1998. With permission from The Rockefeller University Press.)



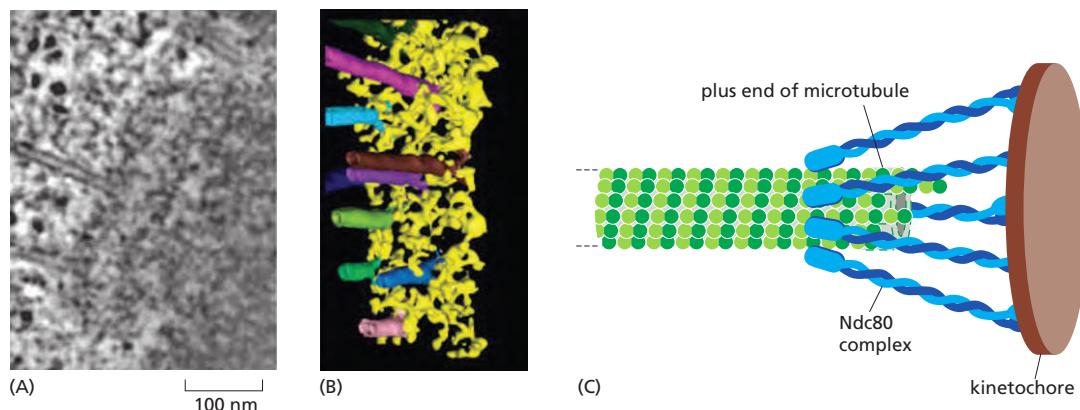
plus ends coming from two directions. However, the kinetochores do not instantly achieve the correct ‘end-on’ microtubule attachment to both spindle poles. Instead, detailed studies with light and electron microscopy show that most initial attachments are unstable *lateral* attachments, in which a kinetochore attaches to the side of a passing microtubule, with assistance from kinesin motor proteins in the outer kinetochore. Soon, however, the dynamic microtubule plus ends capture the kinetochores in the correct end-on orientation (Figure 17–32).

Another attachment mechanism also plays a part, particularly in the absence of centrosomes. Careful microscopic analysis suggests that short microtubules in the vicinity of the chromosomes become embedded in the plus-end-binding sites of the kinetochore. Polymerization at these plus ends then results in growth of the microtubules away from the kinetochore. The minus ends of these kinetochore microtubules are eventually cross-linked to other minus ends and focused by motor proteins at the spindle pole (see Figure 17–28).

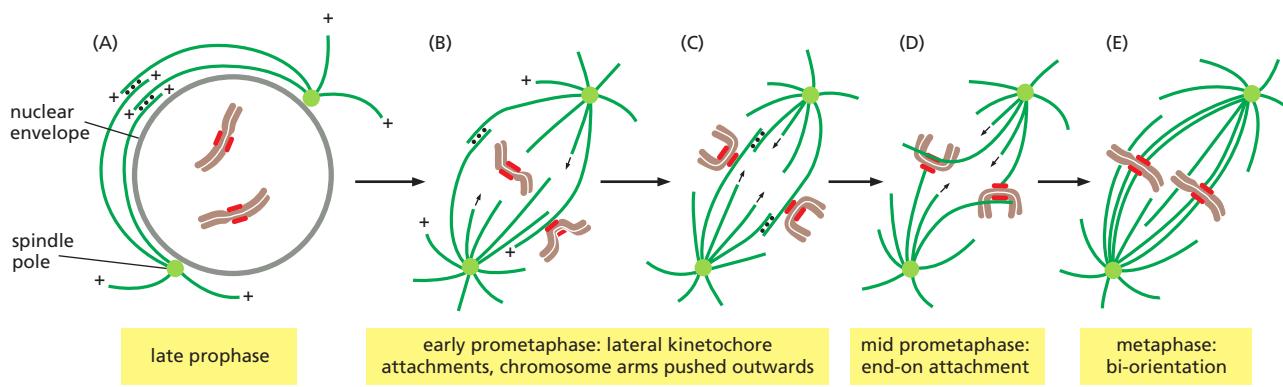
### Bi-orientation Is Achieved by Trial and Error

The success of mitosis demands that sister chromatids in a pair attach to opposite poles of the mitotic spindle, so that they move to opposite ends of the cell when

**Figure 17–30** The kinetochore.  
 (A) A fluorescence micrograph of a metaphase chromosome stained with a DNA-binding fluorescent dye and with human autoantibodies that react with specific kinetochore proteins. The two kinetochores, one associated with each sister chromatid, are stained red.  
 (B) A drawing of a metaphase chromosome showing its two sister chromatids attached to the plus ends of kinetochore microtubules. Each kinetochore forms a plaque on the surface of the centromere.  
 (C) Electron micrograph of an anaphase chromatid with microtubules attached to its kinetochore. While most kinetochores have a trilaminar structure, the one shown here (from a green alga) has an unusually complex structure with additional layers.  
 (A, courtesy of B.R. Brinkley; C, from J.D. Pickett-Heaps and L.C. Fowke, *Aust. J. Biol. Sci.* 23:71–92, 1970. With permission from CSIRO.)



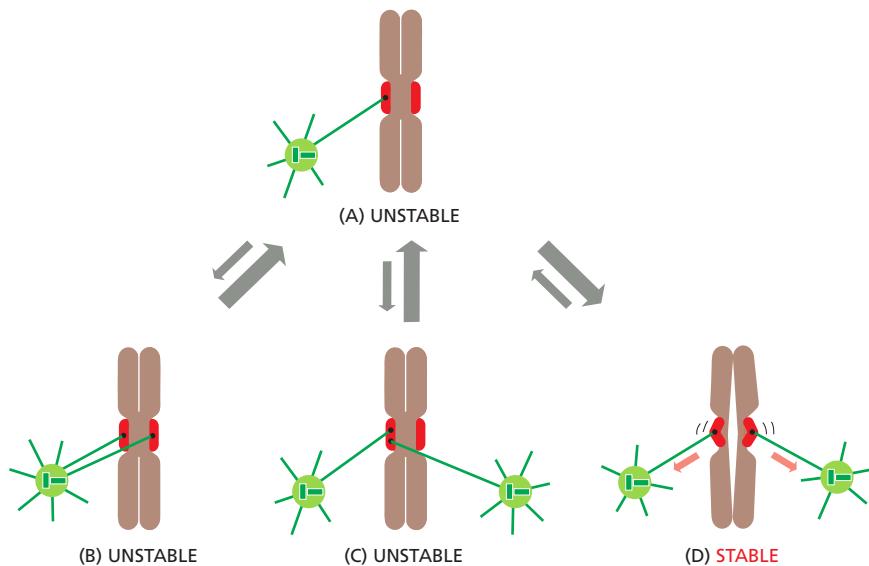
**Figure 17–31** Microtubule attachment sites in the kinetochore. (A) In this electron micrograph of a mammalian kinetochore, the chromosome is on the right, and the plus ends of multiple microtubules are embedded in the outer kinetochore on the left. (B) Electron tomography (discussed in Chapter 9) was used to construct a low-resolution three-dimensional image of the outer kinetochore in (A). Several microtubules (*in multiple colors*) are embedded in fibrous material of the kinetochore, which is thought to be composed of the Ndc80 complex and other proteins. (C) Each microtubule is attached to the kinetochore by interactions with multiple copies of the Ndc80 complex (*blue*). This complex binds to the sides of the microtubule near its plus end, allowing polymerization and depolymerization to occur while the microtubule remains attached to the kinetochore.  
 (A and B, from Y. Dong et al., *Nature Cell Biol.* 9:516–522, 2007. With permission from Macmillan Publishers Ltd.)



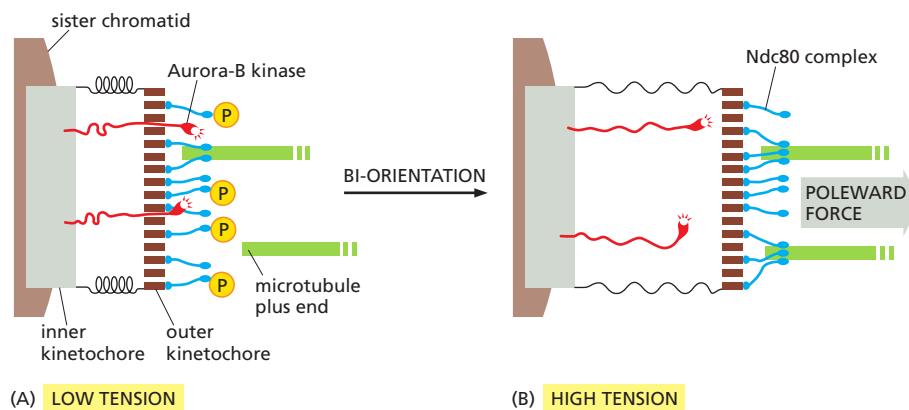
**Figure 17-32 Chromosome attachment to the mitotic spindle in animal cells.** (A) In late prophase of most animal cells, the mitotic spindle poles have moved to opposite sides of the nuclear envelope, with an array of overlapping microtubules between them. (B) Following nuclear envelope breakdown, the sister-chromatid pairs are exposed to the large number of dynamic plus ends of microtubules radiating from the spindle poles. In most cases, the kinetochores are first attached to the sides of these microtubules, while at the same time the arms of the chromosomes are pushed outward from the spindle interior, preventing the arms from blocking microtubule access to the kinetochores. (C) Eventually, the laterally-attached sister chromatids are arranged in a ring around the outside of the spindle. Most of the microtubules are concentrated in this ring, so that the spindle is relatively hollow inside. (D) Dynamic microtubule plus ends eventually encounter the kinetochores in an end-on orientation and are captured and stabilized. (E) Stable end-on attachment to both poles results in *bi-orientation*. Additional microtubules are attached to the kinetochore, resulting in a *kinetochore fiber* containing 10–40 microtubules.

they separate in anaphase. How is this mode of attachment, called **bi-orientation**, achieved? What prevents the attachment of both kinetochores to the same spindle pole or the attachment of one kinetochore to both spindle poles? Part of the answer is that sister kinetochores are constructed in a back-to-back orientation that reduces the likelihood that both kinetochores can face the same spindle pole. Nevertheless, incorrect attachments do occur, and elegant regulatory mechanisms have evolved to correct them.

Incorrect attachments are corrected by a system of trial and error that is based on a simple principle: incorrect attachments are highly unstable and do not last, whereas correct attachments become locked in place. How does the kinetochore sense a correct attachment? The answer appears to be tension (Figure 17-33). When a sister-chromatid pair is properly bi-oriented on the spindle, the two kinetochores are pulled in opposite directions by strong poleward forces. Sister-chromatid cohesion resists these poleward forces, creating high levels of tension within the kinetochores. When chromosomes are incorrectly attached—when both sister chromatids are attached to the same spindle pole, for example—tension is low



**Figure 17-33 Alternative forms of kinetochore attachment to the spindle poles.** (A) Initially, a single microtubule from a spindle pole binds to one kinetochore in a sister-chromatid pair. Additional microtubules can then bind to the chromosome in various ways. (B) A microtubule from the same spindle pole can attach to the other sister kinetochore, or (C) microtubules from both spindle poles can attach to the same kinetochore. These incorrect attachments are unstable, however, so that one of the two microtubules tends to dissociate. (D) When a microtubule from the opposite pole binds to the second kinetochore, the sister kinetochores are thought to sense tension across their microtubule-binding sites. This triggers an increase in microtubule binding affinity, thereby locking the correct attachment in place.



**Figure 17–34** How tension might increase microtubule attachment to the kinetochore. These diagrams illustrate one speculative mechanism by which bi-orientation might increase microtubule attachment to the kinetochore. A single kinetochore is shown for clarity; the spindle pole is on the right. (A) When a sister-chromatid pair is unattached to the spindle or attached to just one spindle pole, there is little tension between the outer and inner kinetochores. The protein kinase Aurora-B is tethered to the inner kinetochore and phosphorylates the microtubule attachment sites, including the Ndc80 complex (blue), in the outer kinetochore as shown, thereby reducing the affinity of microtubule binding. Microtubules therefore associate and dissociate rapidly, and attachment is unstable. (B) When bi-orientation is achieved, the forces pulling the kinetochore toward the spindle pole are resisted by forces pulling the other sister kinetochore toward the opposite pole, and the resulting tension pulls the outer kinetochore away from the inner kinetochore. As a result, Aurora-B is unable to reach the outer kinetochore, and microtubule attachment sites are not phosphorylated. Microtubule binding affinity is therefore increased, resulting in the stable attachment of multiple microtubules to both kinetochores. The dephosphorylation of outer kinetochore proteins depends on a phosphatase that is not shown here.

and the kinetochore sends an inhibitory signal that loosens the grip of its microtubule attachment site, allowing detachment to occur. When bi-orientation occurs, the high tension at the kinetochore shuts off the inhibitory signal, strengthening microtubule attachment. In animal cells, tension not only increases the affinity of the attachment site but also leads to the attachment of additional microtubules to the kinetochore. This results in the formation of a thick *kinetochore fiber* composed of multiple microtubules.

The tension-sensing mechanism depends on the protein kinase Aurora-B, which is associated with the kinetochore and is thought to generate the inhibitory signal that reduces the strength of microtubule attachment in the absence of tension. It phosphorylates several components of the microtubule attachment site, including the Ndc80 complex, decreasing the site's affinity for a microtubule plus end. When bi-orientation occurs, the resulting tension somehow reduces phosphorylation by Aurora-B, thereby increasing the affinity of the attachment site (Figure 17–34).

Following their attachment to the two spindle poles, the chromosomes are tugged back and forth, eventually assuming a position equidistant between the two poles, a position called the **metaphase plate**. In vertebrate cells, the chromosomes then oscillate gently at the metaphase plate, awaiting the signal for the sister chromatids to separate. The signal is produced, with a predictable lag time, after the bi-oriented attachment of the last of the chromosomes.

### Multiple Forces Act on Chromosomes in the Spindle

Multiple mechanisms generate the forces that move chromosomes back and forth after they are attached to the spindle, and produce the tension that is so important for the stabilization of correct attachments. In anaphase, similar forces pull the separated chromatids to opposite ends of the spindle. Three major spindle forces are particularly critical, although their strength and importance vary at different stages of mitosis.

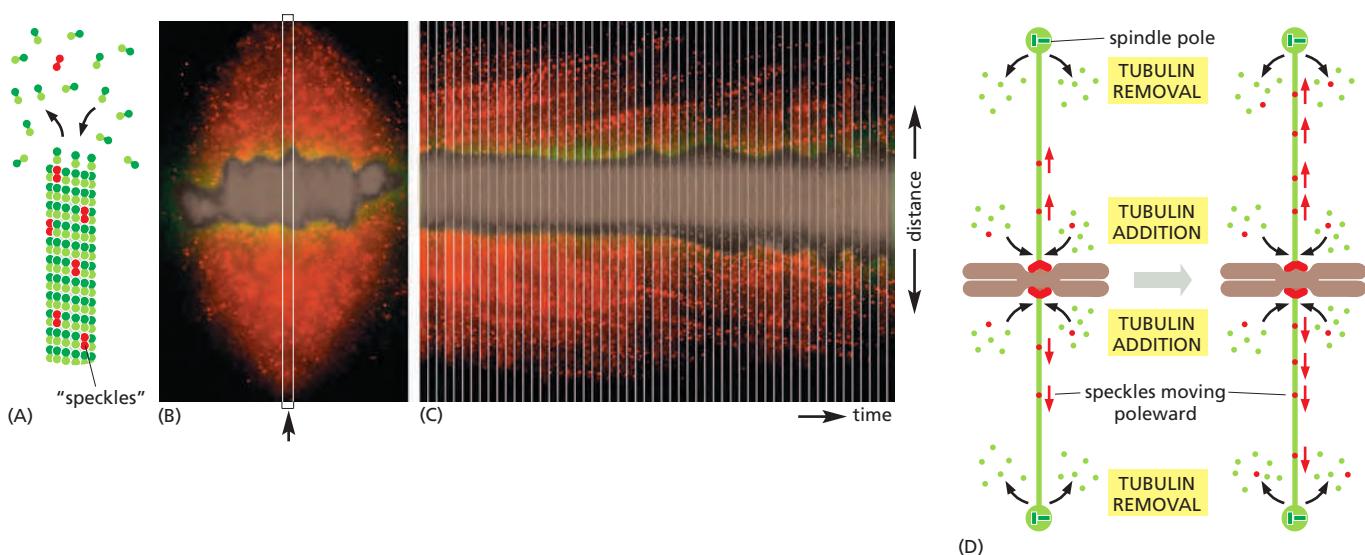
The first major force pulls the kinetochore and its associated chromatid along the kinetochore microtubule toward the spindle pole. It is produced by proteins at the kinetochore itself. By an uncertain mechanism, depolymerization at the plus end of the microtubule generates a force that pulls the kinetochore poleward. This force pulls on chromosomes during prometaphase and metaphase but is particularly important for moving sister chromatids toward the poles after they separate in anaphase. Interestingly, this kinetochore-generated poleward force does not require ATP or motor proteins. This might seem implausible at first, but it has been shown that purified kinetochores in a test tube, with no ATP present, can remain attached to depolymerizing microtubules and thereby move. The energy that drives the movement is stored in the microtubule and is released when the microtubule depolymerizes; it ultimately comes from the hydrolysis of GTP that occurs after a tubulin subunit adds to the end of a microtubule (discussed in Chapter 16).

How does plus-end depolymerization drive the kinetochore toward the pole? As we discussed earlier (see Figure 17–31C), Ndc80 complexes in the kinetochore make multiple low-affinity attachments along the side of the microtubule. Because the attachments are constantly breaking and re-forming at new sites, the kinetochore remains attached to a microtubule even as the microtubule depolymerizes. In principle, this could move the kinetochore toward the spindle pole.

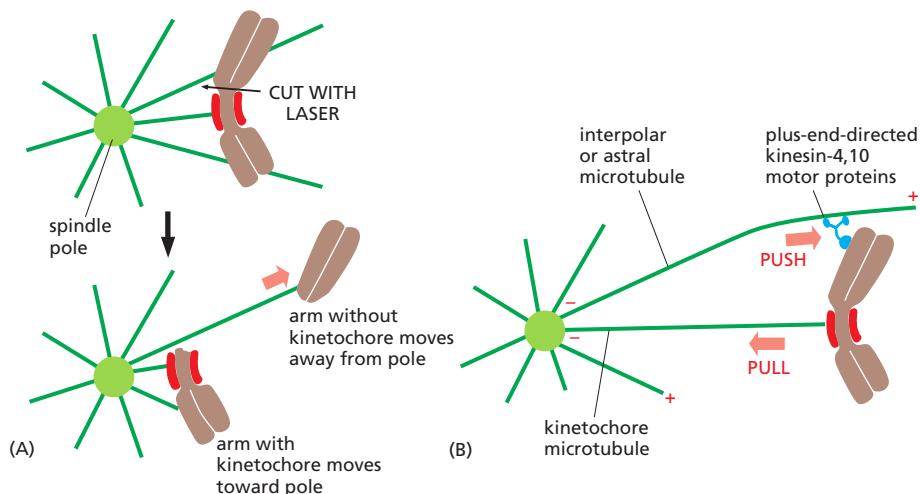
A second poleward force is provided in some cell types by **microtubule flux**, whereby the microtubules themselves are pulled toward the spindle poles and dismantled at their minus ends. The mechanism underlying this poleward movement is not clear, although it might depend on forces generated by motor proteins and minus-end depolymerization at the spindle pole. In metaphase, the addition of new tubulin at the plus end of a microtubule compensates for the loss of tubulin at the minus end, so that microtubule length remains constant despite the movement of microtubules toward the spindle pole (Figure 17–35). Any kinetochore that is attached to a microtubule undergoing such flux experiences a poleward force, which contributes to the generation of tension at the kinetochore in metaphase. Together with the kinetochore-based forces discussed above, flux also contributes to the poleward forces that move sister chromatids after they separate in anaphase.

A third force acting on chromosomes is the *polar ejection force*, or *polar wind*. Plus-end directed kinesin-4 and 10 motors on chromosome arms interact with interpolar microtubules and transport the chromosomes away from the spindle poles (see Figure 17–25). This force is particularly important in prometaphase and metaphase, when it helps push chromosome arms out from the spindle. This force might also help align the sister-chromatid pairs at the metaphase plate (Figure 17–36).

One of the most striking aspects of mitosis in vertebrate cells is the continuous oscillatory movement of the chromosomes in prometaphase and metaphase. When studied by video microscopy in newt lung cells, the movements are seen to switch between two states—a poleward state, when the chromosomes are pulled toward the pole, and an away-from-the-pole, or neutral, state, when the poleward forces are turned off and the polar ejection force pushes the chromosomes away



**Figure 17–35 Microtubule flux in the metaphase spindle.** (A) To observe microtubule flux, a very small amount of fluorescent tubulin is injected into living cells so that individual microtubules form with a very small proportion of fluorescent tubulin. Such microtubules have a speckled appearance when viewed by fluorescence microscopy. (B) Fluorescence micrograph of a mitotic spindle in a living newt lung epithelial cell. The chromosomes are colored brown, and the tubulin speckles are red. (C) The movement of individual speckles can be followed by time-lapse video microscopy. Images of the thin vertical boxed region (arrow) in (B), taken at sequential times, show that individual speckles move toward the poles at a rate of about  $0.75 \mu\text{m}/\text{min}$ , indicating that the microtubules are moving poleward. (D) Microtubule length in the metaphase spindle does not change significantly because new tubulin subunits are added at the microtubule plus end at the same rate as tubulin subunits are removed from the minus end. (B and C, from T.J. Mitchison and E.D. Salmon, *Nat. Cell Biol.* 3:E17–21, 2001. With permission from Macmillan Publishers Ltd.)



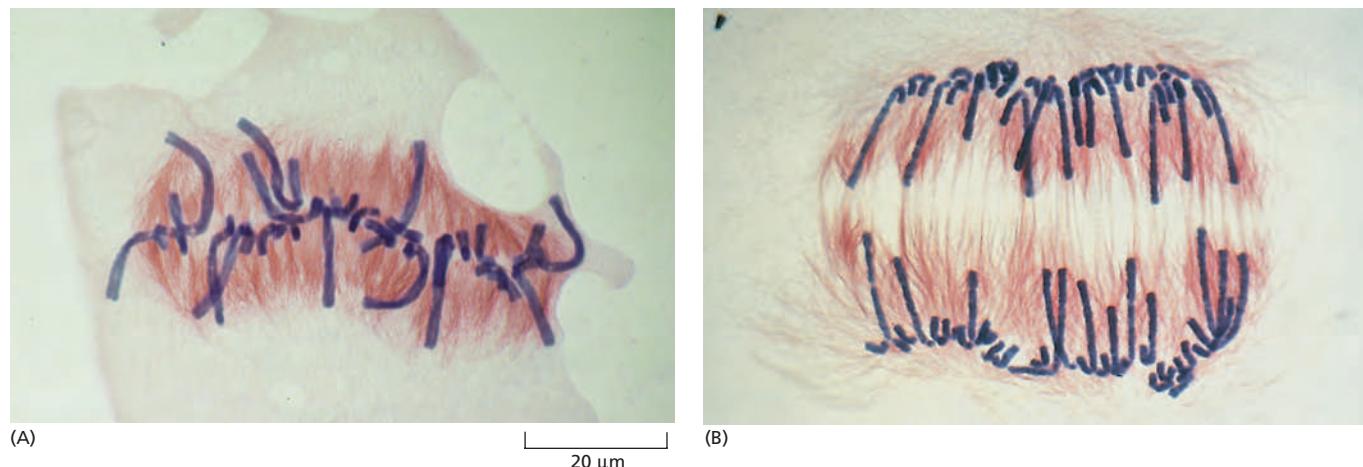
**Figure 17–36** How opposing forces may drive chromosomes to the metaphase plate. (A) Evidence for a polar ejection force that pushes chromosomes away from the spindle poles toward the spindle equator. In this experiment, a laser beam severs a prometaphase chromosome that is attached to a single pole by a kinetochore microtubule. The part of the severed chromosome without a kinetochore is pushed rapidly away from the pole, whereas the part with the kinetochore moves toward the pole, reflecting a decreased repulsion. (B) A model of how two opposing forces may cooperate to move chromosomes to the metaphase plate. Plus-end-directed motor proteins (kinesin-4 and kinesin-10) on the chromosome arms are thought to interact with microtubules to generate the polar ejection force, which pushes chromosomes toward the spindle equator (see Figure 17–25). Poleward forces generated by depolymerization at the kinetochore, together with microtubule flux, are thought to pull chromosomes toward the pole.

from the pole. The switch between the two states may depend on the degree of tension in the kinetochore. It has been proposed, for example, that, as chromosomes move toward a spindle pole, an increasing polar ejection force generates tension in the kinetochore nearest the pole, triggering a switch to the away-from-the-pole state and gradually resulting in the accumulation of chromosomes at the equator of the spindle.

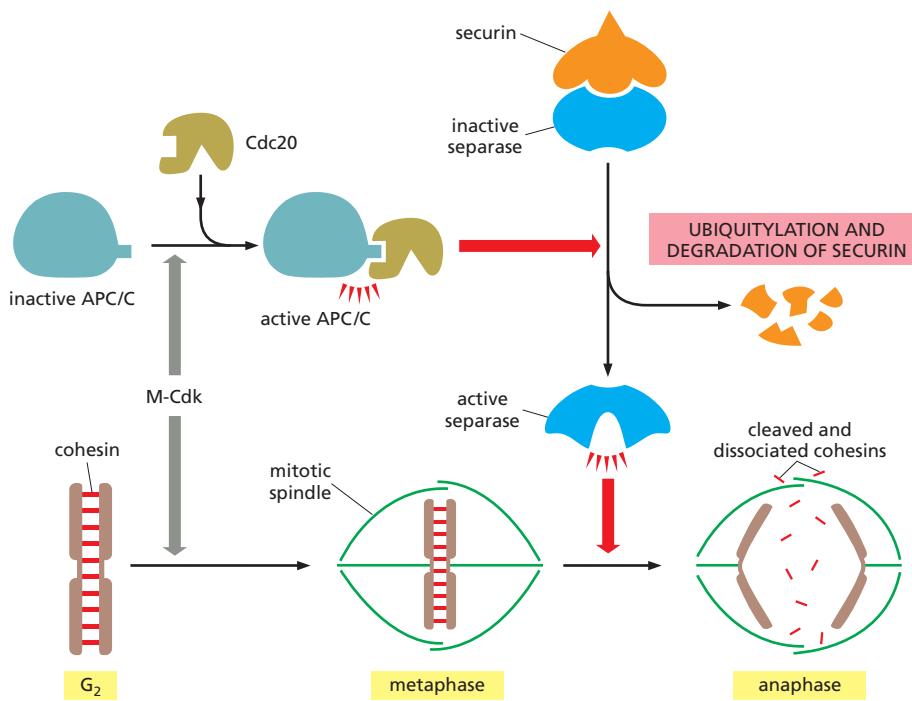
### The APC/C Triggers Sister-Chromatid Separation and the Completion of Mitosis

After M-Cdk has triggered the complex processes leading up to metaphase, the cell cycle reaches its climax with the separation of the sister chromatids at the metaphase-to-anaphase transition (Figure 17–37). Although M-Cdk activity sets the stage for this event, the anaphase-promoting complex (APC/C) discussed earlier throws the switch that initiates sister-chromatid separation by ubiquitylating several mitotic regulatory proteins and thereby triggering their destruction (see Figure 17–15A).

During metaphase, cohesins holding the sister chromatids together resist the poleward forces that pull the sister chromatids apart. Anaphase begins with the sudden loss of sister-chromatid cohesion, which allows the sisters to separate and move to opposite poles of the spindle. The APC/C initiates the process by targeting the inhibitory protein **securin** for destruction. Before anaphase, securin



**Figure 17–37** Sister-chromatid separation at anaphase. In the transition from metaphase (A) to anaphase (B), sister chromatids suddenly and synchronously separate and move toward opposite poles of the mitotic spindle—as shown in these light micrographs of *Haemanthus* (lily) endosperm cells that were stained with gold-labeled antibodies against tubulin. (Courtesy of Andrew Bajer.)



**Figure 17–38** The initiation of sister-chromatid separation by the APC/C.

The activation of APC/C by Cdc20 leads to the ubiquitylation and destruction of securin, which normally holds separase in an inactive state. The destruction of securin allows separase to cleave Scc1, a subunit of the cohesin complex holding the sister chromatids together (see Figure 17–19). The pulling forces of the mitotic spindle then pull the sister chromatids apart. In animal cells, phosphorylation by Cdks also inhibits separase (not shown). Thus, Cdk inactivation in anaphase (resulting from cyclin destruction) also promotes separase activation by allowing its dephosphorylation.

binds to and inhibits the activity of a protease called **separase**. The destruction of securin at the end of metaphase releases separase, which is then free to cleave one of the subunits of cohesin. The cohesins fall away, and the sister chromatids separate (**Figure 17–38**).

In addition to securin, the APC/C also targets the S- and M-cyclins for destruction, leading to the loss of most Cdk activity in anaphase. Cdk inactivation allows phosphatases to dephosphorylate the many Cdk target substrates in the cell, as required for the completion of mitosis and cytokinesis.

If the APC/C triggers anaphase, what activates the APC/C? The answer is only partly known. As mentioned earlier, APC/C activation requires binding to the protein Cdc20 (see Figure 17–15A). At least two processes regulate Cdc20 and its association with the APC/C. First, Cdc20 synthesis increases as the cell approaches mitosis, owing to an increase in the transcription of its gene. Second, phosphorylation of the APC/C helps Cdc20 bind to the APC/C, thereby helping to create an active complex. Among the kinases that phosphorylate and thus activate the APC/C is M-Cdk. Thus, M-Cdk not only triggers the early mitotic events leading up to metaphase, but it also sets the stage for progression into anaphase. The ability of M-Cdk to promote Cdc20-APC/C activity creates a negative feedback loop: M-Cdk sets in motion a regulatory process that leads to cyclin destruction and thus its own inactivation.

### Unattached Chromosomes Block Sister-Chromatid Separation: The Spindle Assembly Checkpoint

Drugs that destabilize microtubules, such as colchicine or vinblastine (discussed in Chapter 16), arrest cells in mitosis for hours or even days. This observation led to the identification of a **spindle assembly checkpoint** mechanism that is activated by the drug treatment and blocks progression through the metaphase-to-anaphase transition. The checkpoint mechanism ensures that cells do not enter anaphase until all chromosomes are correctly bi-oriented on the mitotic spindle.

The spindle assembly checkpoint depends on a sensor mechanism that monitors the strength of microtubule attachment at the kinetochore, possibly by sensing tension as described earlier (see Figure 17–34). Any kinetochore that is not properly attached to the spindle sends out a diffusible negative signal that blocks Cdc20-APC/C activation throughout the cell and thus blocks the

metaphase-to-anaphase transition. When the last sister-chromatid pair is properly bi-oriented, this block is removed, allowing sister-chromatid separation to occur.

The negative checkpoint signal depends on several proteins, including *Mad2*, which are recruited to unattached kinetochores (Figure 17–39). Detailed structural analyses of *Mad2* suggest that the unattached kinetochore acts like an enzyme that catalyzes a change in the conformation of *Mad2*, so that *Mad2*, together with other proteins, can bind and inhibit Cdc20-APC/C.

In mammalian somatic cells, the spindle assembly checkpoint determines the normal timing of anaphase. The destruction of securin in these cells begins moments after the last sister-chromatid pair becomes bi-oriented on the spindle, and anaphase begins about 20 minutes later. Experimental inhibition of the checkpoint mechanism causes premature sister-chromatid separation and anaphase. Surprisingly, the normal timing of anaphase does not depend on the spindle assembly checkpoint in some cells, such as yeasts and the cells of early frog and fly embryos. Other mechanisms, as yet unknown, must determine the timing of anaphase in these cells.

### Chromosomes Segregate in Anaphase A and B

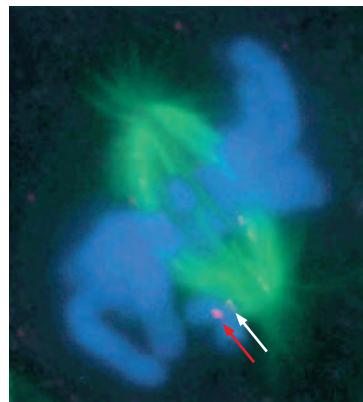
The sudden loss of sister-chromatid cohesion at the onset of anaphase leads to sister-chromatid separation, which allows the forces of the mitotic spindle to pull the sisters to opposite poles of the cell—called *chromosome segregation*. The chromosomes move by two independent and overlapping processes. The first, **anaphase A**, is the initial poleward movement of the chromosomes, which is accompanied by shortening of the kinetochore microtubules. The second, **anaphase B**, is the separation of the spindle poles themselves, which begins after the sister chromatids have separated and the daughter chromosomes have moved some distance apart (Figure 17–40).

Chromosome movement in anaphase A depends on a combination of the two major poleward forces described earlier. The first is the force generated by microtubule depolymerization at the kinetochore, which results in the loss of tubulin subunits at the plus end as the kinetochore moves toward the pole. The second is provided by microtubule flux, which is the poleward movement of the microtubules toward the spindle pole, where minus-end depolymerization occurs. The relative importance of these two forces during anaphase varies in different cell types: in embryonic cells, chromosome movement depends mainly on microtubule flux, for example, whereas movement in yeast and vertebrate somatic cells results primarily from forces generated at the kinetochore.

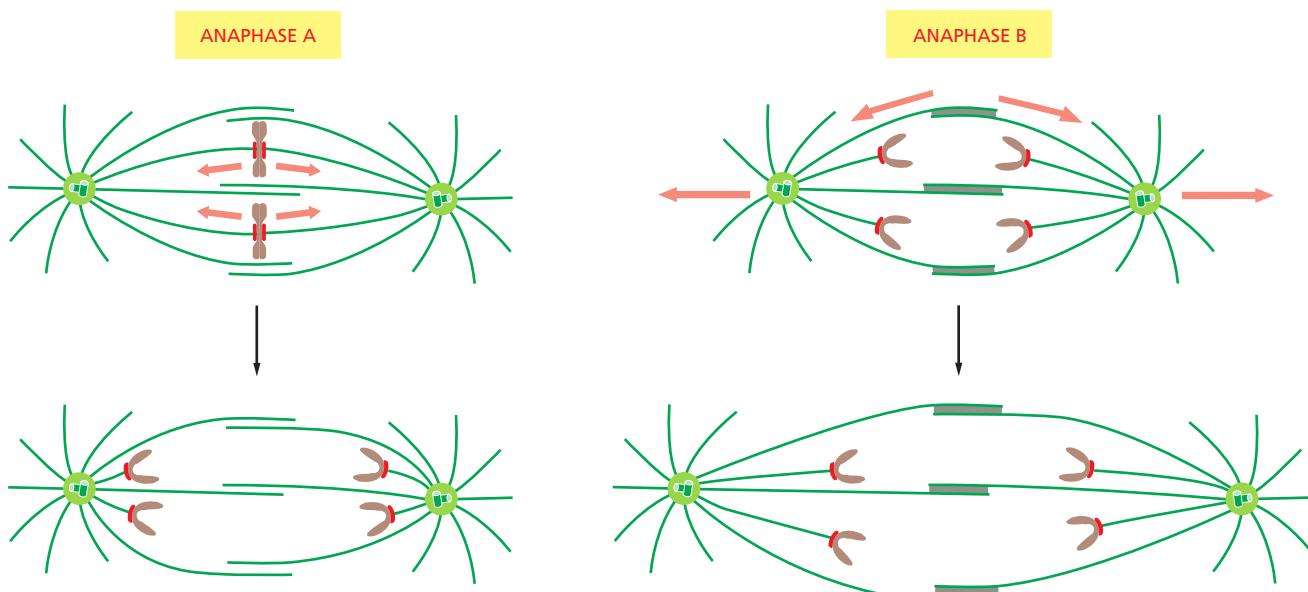
Spindle-pole separation during anaphase B depends on motor-driven mechanisms similar to those that separate the two centrosomes in early mitosis. Plus-end directed kinesin-5 motor proteins, which cross-link the overlapping plus ends of the interpolar microtubules, push the poles apart. In addition, dynein motors that anchor astral microtubule plus ends to the cell cortex pull the poles apart (see Figure 17–25).

Although sister-chromatid separation initiates the chromosome movements of anaphase A, other mechanisms also ensure correct chromosome movements in anaphase A and spindle elongation in anaphase B. Most importantly, the completion of a normal anaphase depends on the dephosphorylation of Cdk substrates, which in most cells results from the APC/C-dependent destruction of cyclins. If M-cyclin destruction is prevented—by the production of a mutant form that is not recognized by the APC/C, for example—sister-chromatid separation generally occurs, but the chromosome movements and microtubule behavior of anaphase are abnormal.

The relative contributions of anaphase A and anaphase B to chromosome segregation vary greatly, depending on the cell type. In mammalian cells, anaphase B begins shortly after anaphase A and stops when the spindle is about twice its metaphase length; in contrast, the spindles of yeasts and certain protozoa primarily use anaphase B to separate the chromosomes at anaphase, and their spindles elongate to up to 15 times their metaphase length.



**Figure 17–39** *Mad2* protein on unattached kinetochores. This fluorescence micrograph shows a mammalian cell in prometaphase, with the mitotic spindle in green and the sister chromatids in blue. One sister-chromatid pair is attached to only one pole of the spindle. Staining with anti-*Mad2* antibodies indicates that *Mad2* is bound to the kinetochore of the unattached sister chromatid (red dot, indicated by red arrow). A small amount of *Mad2* is associated with the kinetochore of the sister chromatid that is attached to the spindle pole (pale dot, indicated by white arrow). (From J.C. Waters et al., *J. Cell Biol.* 141:1181–1191, 1998. With permission from the authors.)



### Segregated Chromosomes Are Packaged in Daughter Nuclei at Telophase

By the end of anaphase, the daughter chromosomes have segregated into two equal groups at opposite ends of the cell. In **telophase**, the final stage of mitosis, the two sets of chromosomes are packaged into a pair of daughter nuclei. The first major event of telophase is the disassembly of the mitotic spindle, followed by the re-formation of the nuclear envelope. Initially, nuclear membrane fragments associate with the surface of individual chromosomes. These membrane fragments fuse to partly enclose clusters of chromosomes and then coalesce to reform the complete nuclear envelope. Nuclear pore complexes are incorporated into the envelope, the nuclear lamina re-forms, and the envelope once again becomes continuous with the endoplasmic reticulum. Once the nuclear envelope has re-formed, the pore complexes pump in nuclear proteins, the nucleus expands, and the mitotic chromosomes are reorganized into their interphase state, allowing gene transcription to resume. A new nucleus has been created, and mitosis is complete. All that remains is for the cell to complete its division into two.

We saw earlier that phosphorylation of various proteins by M-Cdk promotes spindle assembly, chromosome condensation, and nuclear-envelope breakdown in early mitosis. It is thus not surprising that the dephosphorylation of these same proteins is required for spindle disassembly and the re-formation of daughter nuclei in telophase. In principle, these dephosphorylations and the completion of mitosis could be triggered by the inactivation of Cdks, the activation of phosphatases, or both. Although Cdk inactivation—resulting primarily from cyclin destruction—is mainly responsible in most cells, some cells also rely on activation of phosphatases. In budding yeast, for example, the completion of mitosis depends on the activation of a phosphatase called *Cdc14*, which dephosphorylates a subset of Cdk substrates involved in anaphase and telophase.

**Figure 17–40** The two processes of anaphase in mammalian cells. Separated sister chromatids move toward the poles in anaphase A. In anaphase B, the two spindle poles move apart.

### Summary

*M-Cdk triggers the events of early mitosis, including chromosome condensation, assembly of the mitotic spindle, and bipolar attachment of the sister-chromatid pairs to microtubules of the spindle. Spindle formation in animal cells depends largely on the ability of mitotic chromosomes to stimulate local microtubule nucleation and stability, as well as on the ability of motor proteins to organize microtubules into a bipolar array. Many cells also use centrosomes to facilitate spindle assembly. Anaphase is triggered by the APC/C, which stimulates the destruction of*

the proteins that hold the sister chromatids together. APC/C also promotes cyclin destruction and thus the inactivation of M-Cdk. The resulting dephosphorylation of Cdk targets is required for the events that complete mitosis, including the disassembly of the spindle and the re-formation of the nuclear envelope.

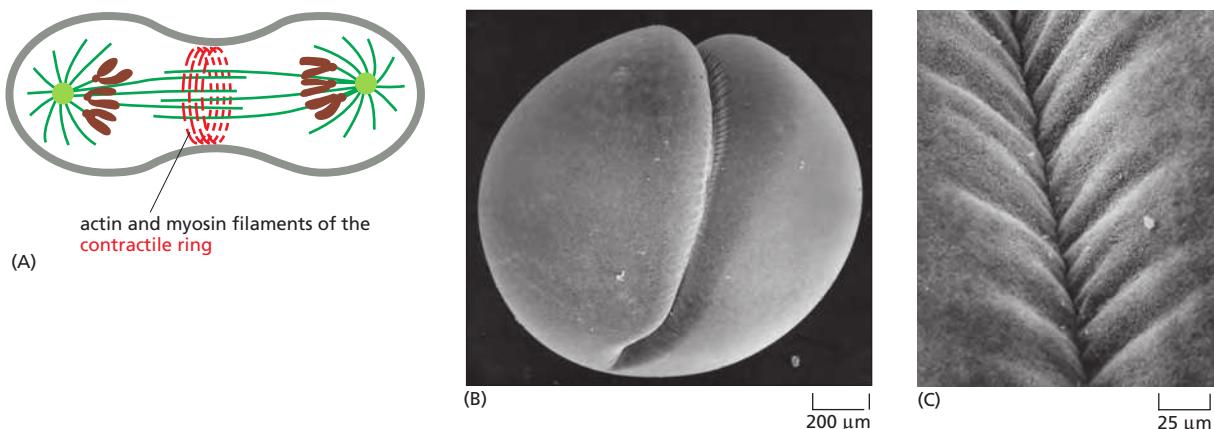
## CYTOKINESIS

The final step in the cell cycle is **cytokinesis**, the division of the cytoplasm in two. In most cells, cytokinesis follows every mitosis, although some cells, such as early *Drosophila* embryos and some mammalian hepatocytes and heart muscle cells, undergo mitosis without cytokinesis and thereby acquire multiple nuclei. In most animal cells, cytokinesis begins in anaphase and ends shortly after the completion of mitosis in telophase.

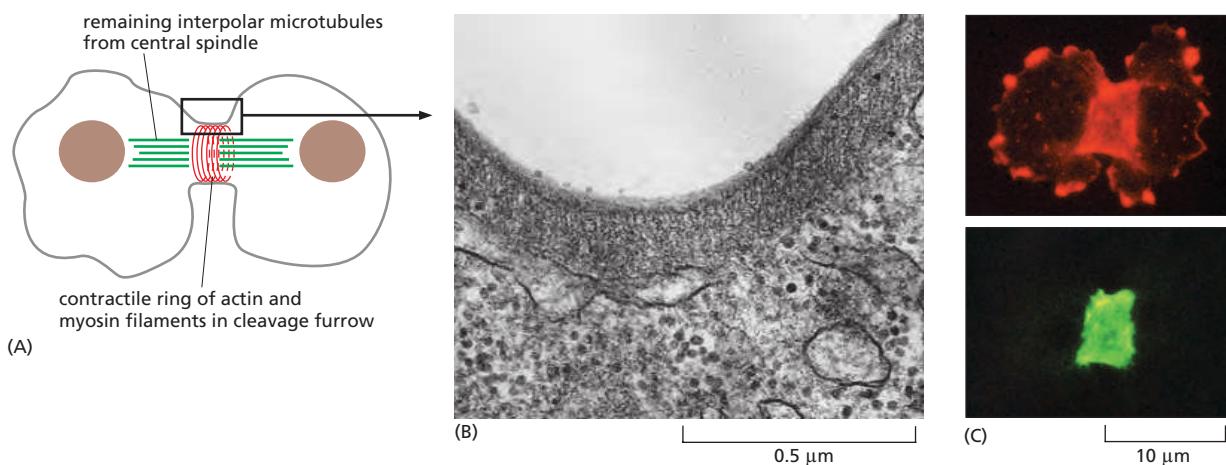
The first visible change of cytokinesis in an animal cell is the sudden appearance of a pucker, or *cleavage furrow*, on the cell surface. The furrow rapidly deepens and spreads around the cell until it completely divides the cell in two. The structure underlying this process is the **contractile ring**—a dynamic assembly composed of actin filaments, myosin II filaments, and many structural and regulatory proteins. During anaphase, the ring assembles just beneath the plasma membrane (Figure 17–41; see also Panel 17–1). The ring gradually contracts, and, at the same time, fusion of intracellular vesicles with the plasma membrane inserts new membrane adjacent to the ring. This addition of membrane compensates for the increase in surface area that accompanies cytoplasmic division. When ring contraction is completed, membrane insertion and fusion seal the gap between the daughter cells.

### Actin and Myosin II in the Contractile Ring Generate the Force for Cytokinesis

In interphase cells, actin and myosin II filaments form a cortical network underlying the plasma membrane. In some cells, they also form large cytoplasmic bundles called *stress fibers* (discussed in Chapter 16). As cells enter mitosis, these arrays of actin and myosin disassemble; much of the actin reorganizes, and myosin II filaments are released. As the sister chromatids separate in anaphase, actin and myosin II begin to accumulate in the rapidly assembling **contractile ring** (Figure 17–42), which also contains numerous other proteins that provide structural support or assist in ring assembly. Assembly of the contractile ring results in part from the local formation of new actin filaments, which depends on *formin* proteins that



**Figure 17–41 Cytokinesis.** (A) The actin–myosin bundles of the contractile ring are oriented as shown, so that their contraction pulls the membrane inward. (B) In this low-magnification scanning electron micrograph of a cleaving frog egg, the cleavage furrow is especially prominent, as the cell is unusually large. The furrowing of the cell membrane is caused by the activity of the contractile ring underneath it. (C) The surface of a furrow at higher magnification. (B and C, from H.W. Beams and R.G. Kessel, *Am. Sci.* 64:279–290, 1976. With permission from Sigma Xi.)



nucleate the assembly of parallel arrays of linear, unbranched actin filaments (discussed in Chapter 16). After anaphase, the overlapping arrays of actin and myosin II filaments contract to generate the force that divides the cytoplasm in two. Once contraction begins, the ring exerts a force large enough to bend a fine glass needle that is inserted in its path. As the ring constricts, it maintains the same thickness, suggesting that its total volume and the number of filaments it contains decrease steadily. Moreover, unlike actin in muscle, the actin filaments in the ring are highly dynamic, and their arrangement changes continually during cytokinesis.

The contractile ring is finally dispensed with altogether when cleavage ends, as the plasma membrane of the cleavage furrow narrows to form the **midbody**. The midbody persists as a tether between the two daughter cells and contains the remains of the *central spindle*, a large protein structure derived from the anti-parallel interpolar microtubules of the spindle midzone, packed tightly together within a dense matrix material (Figure 17-43). After the daughter cells separate completely, some of the components of the residual midbody often remain on the inside of the plasma membrane of each cell, where they may serve as a mark on the cortex that helps to orient the spindle in the subsequent cell division.

### Local Activation of RhoA Triggers Assembly and Contraction of the Contractile Ring

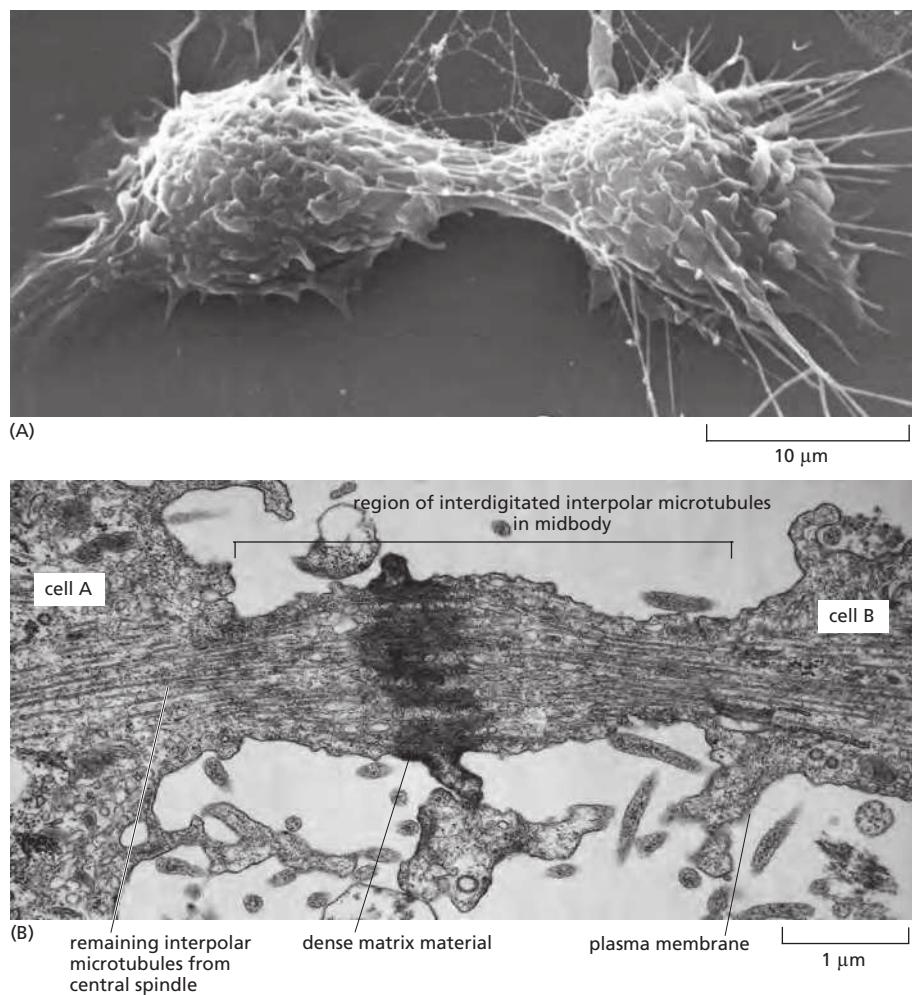
*RhoA*, a small GTPase of the Ras superfamily (see Table 15-5), controls the assembly and function of the contractile ring at the site of cleavage. RhoA is activated at the cell cortex at the future division site, where it promotes actin filament formation, myosin II assembly, and ring contraction. It stimulates actin filament formation by activating formins, and it promotes myosin II assembly and contractions by activating multiple protein kinases, including the Rho-activated kinase Rock (Figure 17-44). These kinases phosphorylate the regulatory myosin light chain, a subunit of myosin II, thereby stimulating bipolar myosin II filament formation and motor activity.

RhoA is thought to be activated by a guanine nucleotide exchange factor (Rho-GEF), which is found at the cell cortex at the future division site and stimulates the release of GDP and binding of GTP to RhoA (see Figure 17-44). We know little about how the RhoGEF is localized or activated at the division site, although the microtubules of the anaphase spindle seem to be involved, as we discuss next.

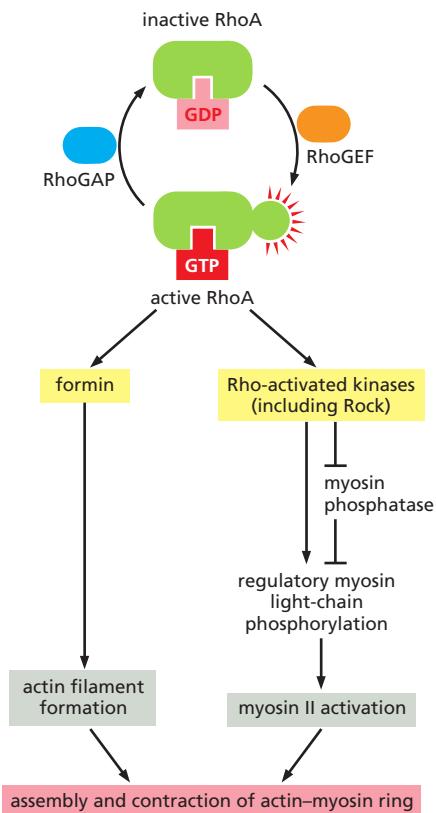
### The Microtubules of the Mitotic Spindle Determine the Plane of Animal Cell Division

The central problem in cytokinesis is how to ensure that division occurs at the right time and in the right place. Cytokinesis must occur only after the two sets of chromosomes are fully segregated from each other, and the site of division must

**Figure 17-42** The contractile ring.  
 (A) A drawing of the cleavage furrow in a dividing cell. (B) An electron micrograph of the ingrowing edge of a cleavage furrow of a dividing animal cell. (C) Fluorescence micrographs of a dividing slime mold amoeba stained for actin (red) and myosin II (green). Whereas all of the visible myosin II has redistributed to the contractile ring, only some of the actin has done so; the rest remains in the cortex of the nascent daughter cells. (B, from H.W. Beams and R.G. Kessel, *Am. Sci.* 64:279–290, 1976. With permission from Sigma Xi; C, courtesy of Yoshio Fukui.)



**Figure 17–43 The midbody.** (A) A scanning electron micrograph of a cultured animal cell dividing; the midbody still joins the two daughter cells. (B) A conventional electron micrograph of the midbody of a dividing animal cell. Cleavage is almost complete, but the daughter cells remain attached by this thin strand of cytoplasm containing the remains of the central spindle. (A, courtesy of Guenter Albrecht-Buehler; B, courtesy of J.M. Mullins.)

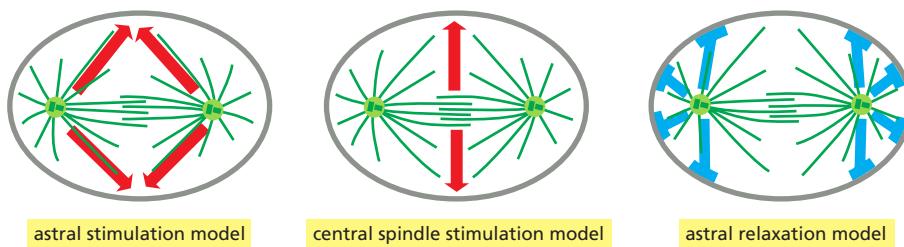


**Figure 17–44 Regulation of the contractile ring by the GTPase RhoA.** Like other Rho family GTPases, RhoA is activated by a RhoGEF protein and inactivated by a Rho GTPase-activating protein (RhoGAP). The active GTP-bound form of RhoA is focused at the future cleavage site. By binding formins, activated RhoA promotes the assembly of actin filaments in the contractile ring. By activating Rho-activated protein kinases, such as Rock, it stimulates myosin II filament formation and activity, thereby promoting contraction of the ring.

be placed between the two sets of daughter chromosomes, thereby ensuring that each daughter cell receives a complete set. The correct timing and positioning of cytokinesis in animal cells are achieved by mechanisms that depend on the mitotic spindle. During anaphase, the spindle generates signals that initiate furrow formation at a position midway between the spindle poles, thereby ensuring that division occurs between the two sets of separated chromosomes. Because these signals originate in the anaphase spindle, this mechanism also contributes to the correct timing of cytokinesis in late mitosis. Cytokinesis also occurs at the correct time because dephosphorylation of Cdk substrates, which depends on cyclin destruction in metaphase and anaphase, initiates cytokinesis. We now describe these regulatory mechanisms in more detail, with an emphasis on cytokinesis in animal cells.

Studies of the fertilized eggs of marine invertebrates first revealed the importance of spindle microtubules in determining the placement of the contractile ring. After fertilization, these embryos cleave rapidly without intervening periods of growth. In this way, the original egg is progressively divided into smaller and smaller cells. Because the cytoplasm is clear, the spindle can be observed in real time with a microscope. If the spindle is tugged into a new position with a fine glass needle in early anaphase, the incipient cleavage furrow disappears, and a new one develops in accord with the new spindle site—supporting the idea that signals generated by the spindle induce local furrow formation.

How does the mitotic spindle specify the site of division? Three general mechanisms have been proposed, and most cells appear to employ a combination of these (Figure 17–45). The first is termed the *astral stimulation model*, in which the



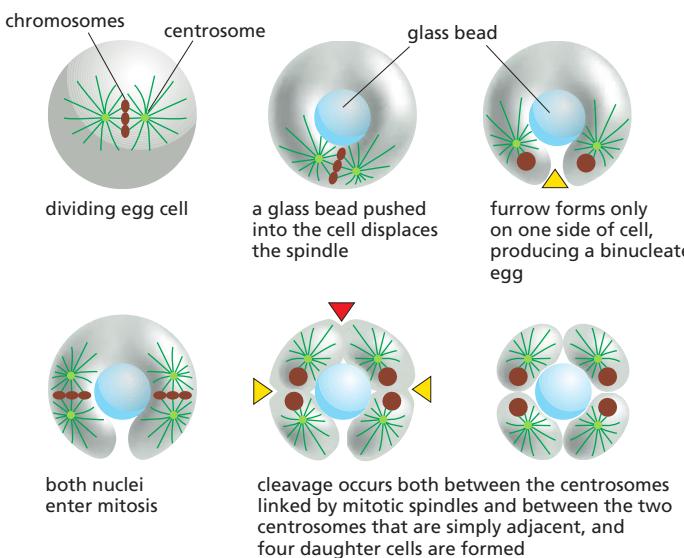
**Figure 17–45** Three current models of how the microtubules of the anaphase spindle generate signals that influence the positioning of the contractile ring. No single model explains all the observations, and furrow positioning is probably determined by a combination of these mechanisms, with the importance of the different mechanisms varying in different organisms. See text for details.

astral microtubules carry furrow-inducing signals, which are somehow focused in a ring on the cell cortex, halfway between the spindle poles. Evidence for this model comes from ingenious experiments in large embryonic cells, which demonstrate that a cleavage furrow forms midway between two asters, even when the two centrosomes nucleating the asters are not connected to each other by a mitotic spindle (Figure 17–46).

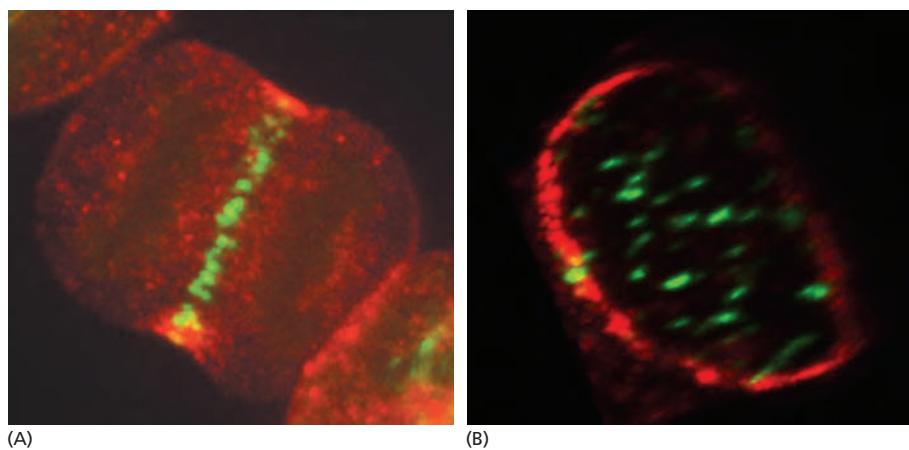
A second possibility, called the *central spindle stimulation model*, is that the spindle midzone, or central spindle, generates a furrow-inducing signal that specifies the site of furrow formation at the cell cortex (see Figure 17–45). The overlapping interpolar microtubules of the central spindle associate with numerous signaling proteins, including proteins that may stimulate RhoA (Figure 17–47). Defects in the functions of these proteins (in *Drosophila* mutants, for example) result in failure of cytokinesis.

A third model proposes that, in some cell types, the astral microtubules promote the local relaxation of actin-myosin bundles at the cell cortex. According to this *astral relaxation model*, the cortical relaxation is minimal at the spindle equator, thus promoting cortical contraction at that site (see Figure 17–45). In the early embryos of *Caenorhabditis elegans*, for example, treatments that result in the loss of astral microtubules lead to increased contractile activity throughout the cell cortex, consistent with this model.

In some cell types, the site of ring assembly is chosen before mitosis. In budding yeasts, for example, a ring of proteins called *septins* assembles in late G<sub>1</sub> at the future division site. The septins are thought to form a scaffold onto which other components of the contractile ring, including myosin II, assemble. In plant cells, an organized band of microtubules and actin filaments, called the **preprophase band**, assembles just before mitosis and marks the site where the cell wall will assemble and divide the cell in two, as we now discuss.



**Figure 17–46** An experiment demonstrating the influence of the position of microtubule asters on the subsequent plane of cleavage in a large egg cell. If the mitotic spindle is mechanically pushed to one side of the cell with a glass bead, the membrane furrowing is incomplete, failing to occur on the opposite side of the cell. Subsequent cleavages occur not only at the midzone of each of the two subsequent mitotic spindles (yellow arrowheads), but also between the two adjacent asters that are not linked by a mitotic spindle—but in this abnormal cell share the same cytoplasm (red arrowhead). Apparently, the contractile ring that produces the cleavage furrow in these cells always forms in the region midway between two asters, suggesting that the asters somehow alter the adjacent region of cell cortex to induce furrow formation between them.

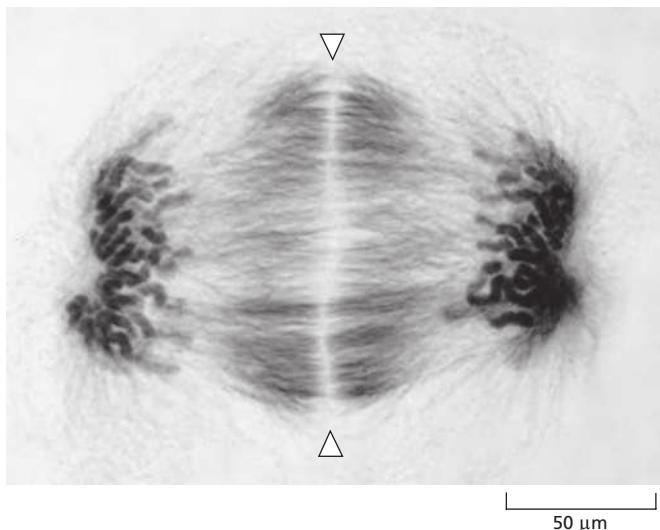


**Figure 17–47** Localization of cytokinesis regulators at the central spindle of the human cell. (A) At center is a cultured human cell at the beginning of cytokinesis, showing the locations of the GTPase RhoA (red) and a protein called Cyk4 (green), which is one of several regulatory proteins that form complexes at the overlapping plus ends of interpolar microtubules. These proteins are thought to generate signals that help control RhoA activity at the cell cortex. (B) When the same three-dimensional image is viewed in the plane of the contractile ring, as shown here, RhoA (red) is seen as a ring beneath the cell surface, while the central spindle protein Cyk4 (green) is associated with microtubule bundles scattered throughout the equatorial plane of the cell. (Courtesy of Alisa Piekny and Michael Glotzer.)

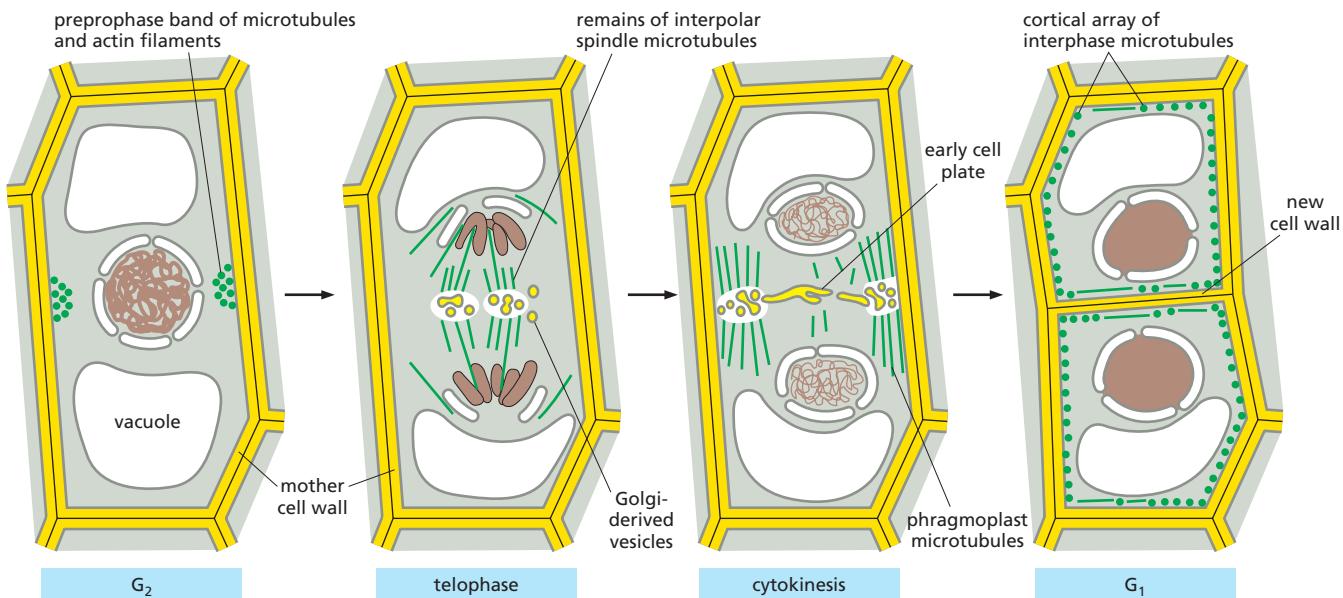
### The Phragmoplast Guides Cytokinesis in Higher Plants

In most animal cells, the inward movement of the cleavage furrow depends on an increase in the surface area of the plasma membrane. New membrane is added at the inner edge of the cleavage furrow and is generally provided by small membrane vesicles that are transported on microtubules from the Golgi apparatus to the furrow.

Membrane deposition is particularly important for cytokinesis in higher-plant cells. These cells are enclosed by a semirigid *cell wall*. Rather than a contractile ring dividing the cytoplasm from the outside in, the cytoplasm of the plant cell is partitioned from the inside out by the construction of a new cell wall, called the **cell plate**, between the two daughter nuclei (Figure 17–48). The assembly of the cell plate begins in late anaphase and is guided by a structure called the **phragmoplast**, which contains microtubules derived from the mitotic spindle. Motor proteins transport small vesicles along these microtubules from the Golgi apparatus to the cell center. These vesicles, filled with polysaccharide and glycoproteins required for the synthesis of the new cell wall, fuse to form a disc-like, membrane-enclosed structure called the *early cell plate*. The plate expands outward by further



**Figure 17–48** Cytokinesis in a plant cell in telophase. In this light micrograph, the early cell plate (between the two arrowheads) has formed in a plane perpendicular to the plane of the page. The microtubules of the spindle are stained with gold-labeled antibodies against tubulin, and the DNA in the two sets of daughter chromosomes is stained with a fluorescent dye. Note that there are no astral microtubules, because there are no centrosomes in higher-plant cells. (Courtesy of Andrew Bajer.)



vesicle fusion until it reaches the plasma membrane and the original cell wall and divides the cell in two. Later, cellulose microfibrils are laid down within the matrix of the cell plate to complete the construction of the new cell wall (**Figure 17–49**).

### Membrane-Enclosed Organelles Must Be Distributed to Daughter Cells During Cytokinesis

The process of mitosis ensures that each daughter cell receives a full complement of chromosomes. When a eukaryotic cell divides, however, each daughter cell must also inherit all of the other essential cell components, including the membrane-enclosed organelles. As discussed in Chapter 12, organelles such as mitochondria and chloroplasts cannot be assembled *de novo* from their individual components; they can arise only by the growth and division of the preexisting organelles. Similarly, cells cannot make a new endoplasmic reticulum (ER) unless some part of it is already present.

How, then, do the various membrane-enclosed organelles segregate when a cell divides? Organelles such as mitochondria and chloroplasts are usually present in large enough numbers to be safely inherited if, on average, their numbers roughly double once each cycle. The ER in interphase cells is continuous with the nuclear envelope and is organized by the microtubule cytoskeleton. Upon entry into M phase, the reorganization of the microtubules and breakdown of the nuclear envelope releases the ER. In most cells, the ER remains largely intact and is cut in two during cytokinesis. The Golgi apparatus is reorganized and fragmented during mitosis. Golgi fragments associate with the spindle poles and are thereby distributed to opposite ends of the spindle, ensuring that each daughter cell inherits the materials needed to reconstruct the Golgi in telophase.

**Figure 17–49** The special features of cytokinesis in a higher-plant cell. The division plane is established before M phase by a band of microtubules and actin filaments (the preprophase band) at the cell cortex. At the beginning of telophase, after the chromosomes have segregated, a new cell wall starts to assemble inside the cell at the equator of the old spindle. The interpolar microtubules of the mitotic spindle remaining at telophase form the phragmoplast. The plus ends of these microtubules no longer overlap but end at the cell equator. Golgi-derived vesicles, filled with cell-wall material, are transported along these microtubules and fuse to form the new cell wall, which grows outward to reach the plasma membrane and original cell wall. The plasma membrane and the membrane surrounding the new cell wall fuse, separating the two daughter cells.

### Some Cells Reposition Their Spindle to Divide Asymmetrically

Most animal cells divide symmetrically: the contractile ring forms around the equator of the parent cell, producing two daughter cells of equal size and with the same components. This symmetry results from the placement of the mitotic spindle, which in most cases tends to center itself in the cytoplasm. Astral microtubules and motor proteins that either push or pull on these microtubules contribute to the centering process.

There are many instances in development, however, when cells divide asymmetrically to produce two cells that differ in size, in the cytoplasmic contents they inherit, or in both. Usually, the two different daughter cells are destined to develop

along different pathways. To create daughter cells with different fates in this way, the mother cell must first segregate certain components (called *cell fate determinants*) to one side of the cell and then position the plane of division so that the appropriate daughter cell inherits these components (Figure 17–50). To position the plane of division asymmetrically, the spindle has to be moved in a controlled manner within the dividing cell. It seems likely that changes in local regions of the cell cortex direct such spindle movements and that motor proteins localized there pull one of the spindle poles, via its astral microtubules, to the appropriate region. Genetic analyses in *C. elegans* and *Drosophila* have identified some of the proteins required for such asymmetric divisions, and some of these proteins seem to have a similar role in vertebrates.

### Mitosis Can Occur Without Cytokinesis

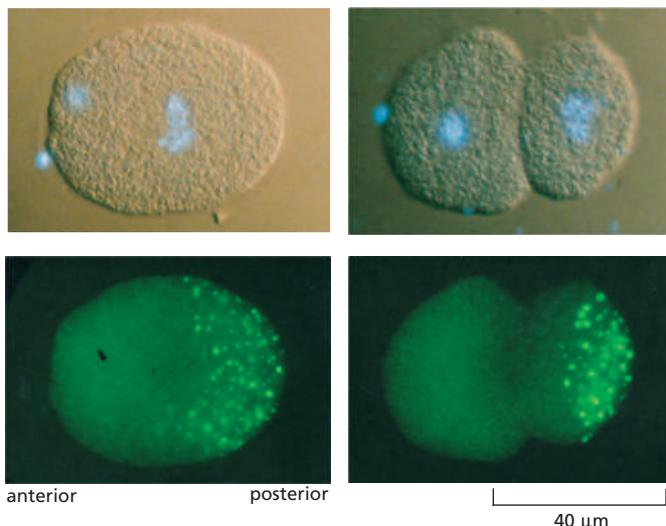
Although nuclear division is usually followed by cytoplasmic division, there are exceptions. Some cells undergo multiple rounds of nuclear division without intervening cytoplasmic division. In the early *Drosophila* embryo, for example, the first 13 rounds of nuclear division occur without cytoplasmic division, resulting in the formation of a single large cell containing several thousand nuclei, arranged in a monolayer near the surface. A cell in which multiple nuclei share the same cytoplasm is called a **syncytium**. This arrangement greatly speeds up early development, as the cells do not have to take the time to go through all the steps of cytokinesis for each division. After these rapid nuclear divisions, membranes are created around each nucleus in one round of coordinated cytokinesis called *cellularization*. The plasma membrane extends inward and, with the help of an actin-myosin ring, pinches off to enclose each nucleus (Figure 17–51).

Nuclear division without cytokinesis also occurs in some types of mammalian cells. Megakaryocytes, which produce blood platelets, and some hepatocytes and heart muscle cells, for example, become multinucleated in this way.

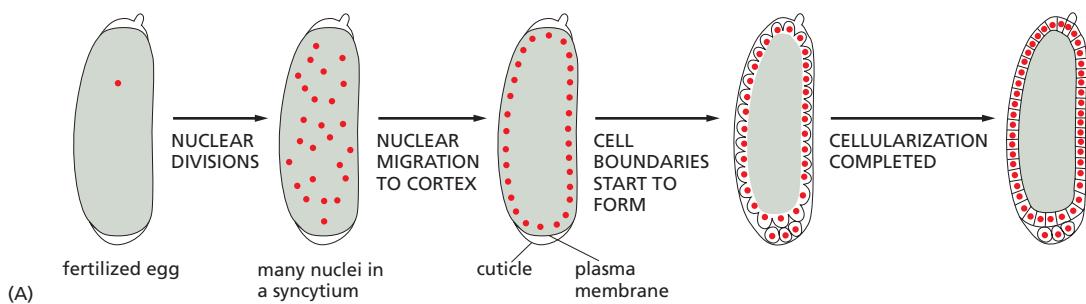
After cytokinesis, most cells enter G<sub>1</sub>, in which Cdks are mostly inactive. We end this section by discussing how this state is achieved at the end of M phase.

### The G<sub>1</sub> Phase Is a Stable State of Cdk Inactivity

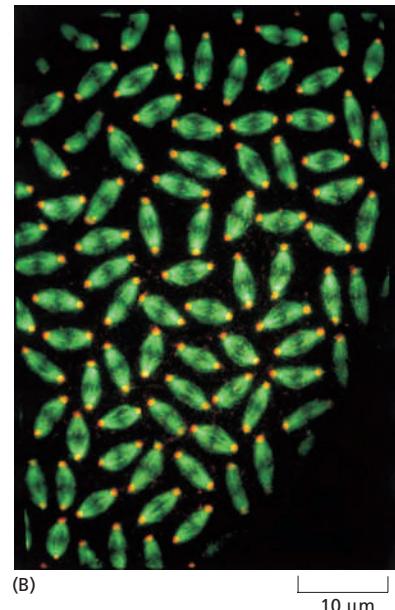
A key regulatory event in late M phase is the inactivation of Cdks, which is driven primarily by APC/C-dependent cyclin destruction. As described earlier, the inactivation of Cdks in late M phase has many functions: it triggers the events of late mitosis, promotes cytokinesis, and enables the synthesis of prereplicative complexes at DNA replication origins. It also provides a mechanism for resetting the



**Figure 17–50** An asymmetric cell division segregating cytoplasmic components to only one daughter cell. These light micrographs illustrate the controlled asymmetric segregation of specific cytoplasmic components to one daughter cell during the first division of a fertilized egg of the nematode *C. elegans*. The fertilized egg is shown in the left micrographs and the two daughter cells in the right micrographs. The cells above have been stained with a blue, DNA-binding, fluorescent dye to show the nucleus (and polar bodies); they are viewed by both differential-interference-contrast and fluorescence microscopy. The cells below are the same cells stained with an antibody against P-granules and viewed by fluorescence microscopy. These small granules are made of RNA and proteins and determine which cells become germ cells. They are distributed randomly throughout the cytoplasm of the unfertilized egg (not shown) but become segregated to the posterior pole of the fertilized egg. The cleavage plane is oriented to ensure that only the posterior daughter cell receives the P-granules when the egg divides. The same segregation process is repeated in several subsequent cell divisions, so that the P-granules end up only in cells that give rise to eggs and sperm. (Courtesy of Susan Strome.)



**Figure 17–51 Mitosis without cytokinesis in the early *Drosophila* embryo.** (A) The first 13 nuclear divisions occur synchronously and without cytoplasmic division to create a large syncytium. Most of the nuclei migrate to the cortex, and the plasma membrane extends inward and pinches off to surround each nucleus to form individual cells in a process called cellularization. (B) Fluorescence micrograph of multiple mitotic spindles in a *Drosophila* embryo before cellularization. The microtubules are stained green and the centrosomes red. Note that all the nuclei go through the cycle synchronously; here, they are all in metaphase, with the unlabeled chromosomes seen as a dark band at the spindle equator. (B, courtesy of Kristina Yu and William Sullivan.)

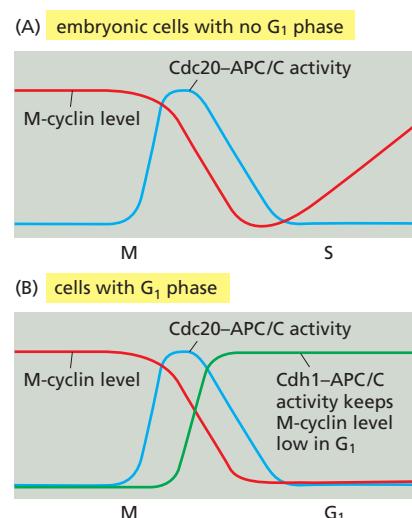


cell-cycle control system to a state of Cdk inactivity as the cell prepares to enter a new cell cycle. In most cells, this state of Cdk inactivity generates a G<sub>1</sub> gap phase, during which the cell grows and monitors its environment before committing to a new cell cycle.

In early animal embryos, the inactivation of M-Cdk in late mitosis is due almost entirely to the action of Cdc20-APC/C, discussed earlier. Recall, however, that M-Cdk stimulates Cdc20-APC/C activity. Thus, the destruction of M-cyclin in late mitosis soon leads to the inactivation of all APC/C activity in an embryonic cell. This APC/C inactivation immediately after mitosis is especially useful in rapid embryonic cell cycles, as it allows the cell to quickly begin accumulating new M-cyclin for the next cycle (Figure 17–52A).

Rapid cyclin accumulation immediately after mitosis is not useful, however, for cells in which a G<sub>1</sub> phase is needed to allow control of entry into the next cell cycle. These cells employ several mechanisms to prevent Cdk reactivation after mitosis. One mechanism uses another APC/C-activating protein called Cdh1, mentioned earlier as a close relative of Cdc20 (see Table 17–2). Although both Cdh1 and Cdc20 bind to and activate the APC/C, they differ in one important respect. Whereas M-Cdk activates the Cdc20-APC/C complex, it inhibits the Cdh1-APC/C complex by directly phosphorylating Cdh1. As a result of this relationship, Cdh1-APC/C activity increases in late mitosis after the Cdc20-APC/C complex has initiated the destruction of M-cyclin. M-cyclin destruction therefore continues after mitosis: although Cdc20-APC/C activity has declined, Cdh1-APC/C activity is high (Figure 17–52B).

A second mechanism that suppresses Cdk activity in G<sub>1</sub> depends on the increased production of CKIs, the Cdk inhibitor proteins discussed earlier. Budding yeast cells, in which this mechanism is best understood, contain a CKI protein called Sic1, which binds to and inactivates M-Cdk in late mitosis and G<sub>1</sub> (see



**Figure 17–52 The creation of a G<sub>1</sub> phase by stable Cdk inhibition after mitosis.** (A) In early embryonic cell cycles, Cdc20-APC/C activity rises at the end of metaphase, triggering M-cyclin destruction. Because M-Cdk activity stimulates Cdc20-APC/C activity, the loss of M-cyclin leads to APC/C inactivation after mitosis, which allows M-cyclins to begin accumulating again. (B) In cells that have a G<sub>1</sub> phase, the drop in M-Cdk activity in late mitosis leads to the activation of Cdh1-APC/C (as well as to the accumulation of Cdk inhibitor proteins; not shown). This ensures a continued suppression of Cdk activity after mitosis, as required for a G<sub>1</sub> phase.

Table 17–2). Like Cdh1, Sic1 is inhibited by M-Cdk, which phosphorylates Sic1 during mitosis and thereby promotes its ubiquitylation by SCF. Thus, Sic1 and M-Cdk, like Cdh1 and M-Cdk, inhibit each other. As a result, the decline in M-Cdk activity that occurs in late mitosis causes the Sic1 protein to accumulate, and this CKI helps keep M-Cdk activity low after mitosis. A CKI protein called *p27* (see Figure 17–14) may serve similar functions in animal cells.

In most cells, decreased transcription of M-cyclin genes also inactivates M-Cdks in late mitosis. In budding yeast, for example, M-Cdk promotes the expression of these genes, resulting in a positive feedback loop. This loop is turned off as cells exit from mitosis: the inactivation of M-Cdk by Cdh1 and Sic1 leads to decreased M-cyclin gene transcription and thus decreased M-cyclin synthesis. Gene regulatory proteins that promote the expression of G<sub>1</sub>/S- and S-cyclins are also inhibited during G<sub>1</sub>.

Thus, Cdh1-APC/C activation, CKI accumulation, and decreased cyclin gene expression act together to ensure that the early G<sub>1</sub> phase is a time when essentially all Cdk activity is suppressed. As in many other aspects of cell-cycle control, the use of multiple regulatory mechanisms allows the system to operate with reasonable efficiency even if one mechanism fails. So how does the cell escape from this stable G<sub>1</sub> state to initiate a new cell cycle? The answer is that G<sub>1</sub>/S-Cdk activity, which rises in late G<sub>1</sub>, releases all the braking mechanisms that suppress Cdk activity, as we describe later, in the last section of this chapter.

## Summary

*After mitosis completes the formation of a pair of daughter nuclei, cytokinesis finishes the cell cycle by dividing the cell itself. Cytokinesis depends on a ring of actin and myosin filaments that contracts in late mitosis at a site midway between the segregated chromosomes. In animal cells, the positioning of the contractile ring is determined by signals emanating from the microtubules of the anaphase spindle. Dephosphorylation of Cdk targets, which results from Cdk inactivation in anaphase, triggers cytokinesis at the correct time after anaphase. After cytokinesis, the cell enters a stable G<sub>1</sub> state of low Cdk activity, where it awaits signals to enter a new cell cycle.*

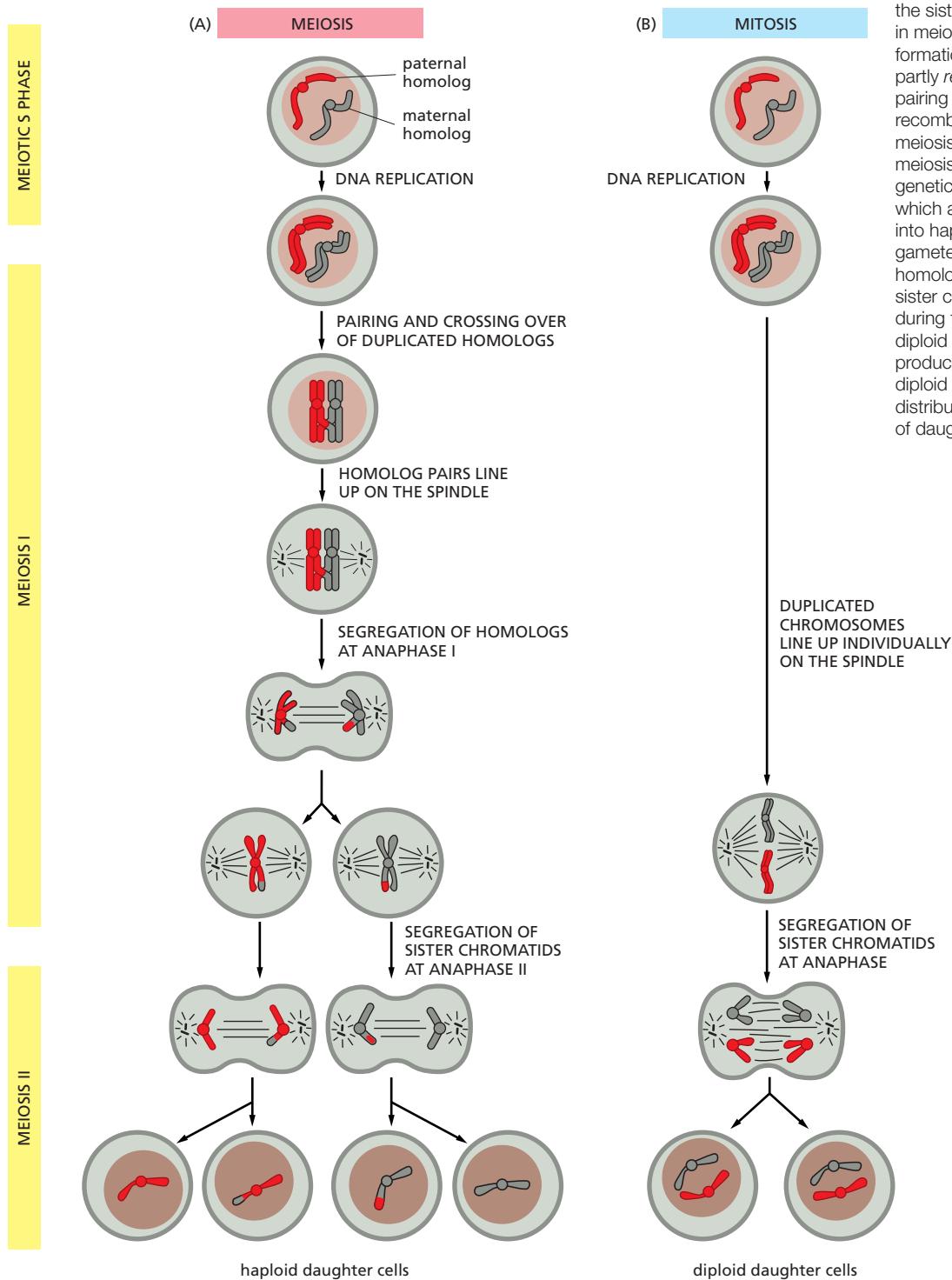
## MEIOSIS

Most eukaryotic organisms reproduce sexually: the genomes of two parents mix to generate offspring that are genetically distinct from either parent. The cells of these organisms are generally *diploid*: that is, they contain two slightly different copies, or *homologs*, of each chromosome, one from each parent. Sexual reproduction depends on a specialized nuclear division process called *meiosis*, which produces *haploid* cells carrying only a single copy of each chromosome. In many organisms, the haploid cells differentiate into specialized reproductive cells called *gametes*—eggs and sperm in most species. In these species, the reproductive cycle ends when a sperm and egg fuse to form a diploid *zygote*, which has the potential to form a new individual. In this section, we consider the basic mechanisms and regulation of meiosis, with an emphasis on how they compare with those of mitosis.

### Meiosis Includes Two Rounds of Chromosome Segregation

Meiosis reduces the chromosome number by half using many of the same molecular machines and control systems that operate in mitosis. As in the mitotic cell cycle, the cell begins the meiotic program by duplicating its chromosomes in meiotic S phase, resulting in pairs of sister chromatids that are tightly linked along their entire lengths by cohesin complexes. Unlike mitosis, however, two successive rounds of chromosome segregation then occur (Figure 17–53). The first of these divisions (**meiosis I**) solves the problem, unique to meiosis, of segregating the homologs. The duplicated paternal and maternal homologs pair up alongside

each other and become physically linked by the process of genetic recombination. These pairs of homologs, each containing a pair of sister chromatids, then line up on the first meiotic spindle. In the first meiotic anaphase, duplicated homologs rather than sister chromatids are pulled apart and segregated into the two daughter nuclei. Only in the second division (**meiosis II**), which occurs without further DNA replication, are the sister chromatids pulled apart and segregated (as in mitosis) to produce haploid daughter nuclei. In this way, each diploid nucleus that enters meiosis produces four haploid nuclei, each of which contains either the maternal or paternal copy of each chromosome, but not both (Movie 17.7).



**Figure 17–53 Comparison of meiosis and mitosis.** For clarity, only one pair of homologous chromosomes (homologs) is shown. (A) Meiosis is a form of nuclear division in which a single round of chromosome duplication (meiotic S phase) is followed by two rounds of chromosome segregation. The duplicated homologs, each consisting of tightly bound sister chromatids, pair up and are segregated into different daughter nuclei in meiosis I; the sister chromatids are segregated in meiosis II. As indicated by the formation of chromosomes that are partly red and partly gray, homolog pairing in meiosis leads to genetic recombination (crossing-over) during meiosis I. Each diploid cell that enters meiosis therefore produces four genetically different haploid nuclei, which are distributed by cytokinesis into haploid cells that differentiate into gametes. (B) In mitosis, by contrast, homologs do not pair up, and the sister chromatids are segregated during the single division. Thus, each diploid cell that divides by mitosis produces two genetically identical diploid daughter nuclei, which are distributed by cytokinesis into a pair of daughter cells.

## Duplicated Homologs Pair During Meiotic Prophase

During mitosis in most organisms, homologous chromosomes behave independently of each other. During meiosis I, however, it is crucial that homologs recognize each other and associate physically in order for the maternal and paternal homologs to be bi-oriented on the first meiotic spindle. Special mechanisms mediate these interactions.

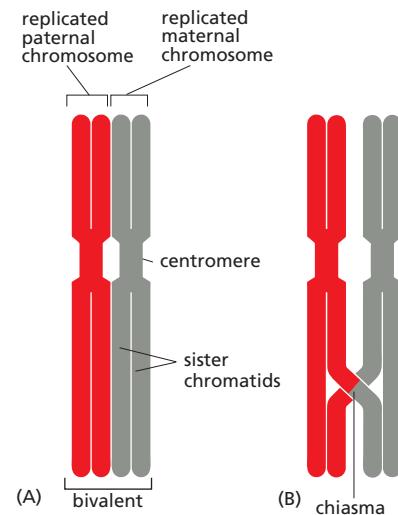
The gradual juxtaposition of homologs occurs during a prolonged period called meiotic prophase (or prophase I), which can take hours in yeasts, days in mice, and weeks in higher plants. Like their mitotic counterparts, duplicated meiotic prophase chromosomes first appear as long threadlike structures, in which the sister chromatids are so tightly glued together that they appear as one. It is during early prophase I that the homologs begin to associate along their length in a process called **pairing**, which, in some organisms at least, begins with interactions between complementary DNA sequences (called *pairing sites*) in the two homologs. As prophase progresses, the homologs become more closely juxtaposed, forming a four-chromatid structure called a **bivalent** (Figure 17–54A). In most species, homolog pairs are then locked together by homologous recombination: DNA double-strand breaks are formed at several locations in each sister chromatid, resulting in large numbers of DNA recombination events between the homologs (as described in Chapter 5). Some of these events lead to reciprocal DNA exchanges called *crossovers*, where the DNA of a chromatid crosses over to become continuous with the DNA of a homologous chromatid (Figure 17–54B; also see Figure 5–54).

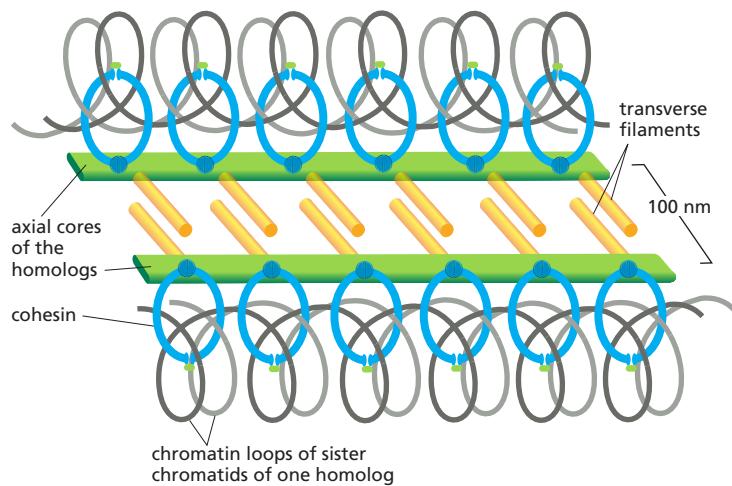
## Homolog Pairing Culminates in the Formation of a Synaptonemal Complex

The paired homologs are brought into close juxtaposition, with their structural axes (*axial cores*) about 400 nm apart, by a mechanism that depends in most species on the double-strand DNA breaks that occur in sister chromatids. What pulls the axes together? One possibility is that the large protein machine, called a *recombination complex*, which assembles on a double-strand break in a chromatid, binds the matching DNA sequence in the nearby homolog and helps reel in this partner. This so-called *presynaptic alignment* of the homologs is followed by *synapsis*, in which the axial core of a homolog becomes tightly linked to the axial core of its partner by a closely packed array of *transverse filaments* to create a **synaptonemal complex**, which bridges the gap, now only 100 nm, between the homologs (Figure 17–55). Although crossing-over begins before the synaptonemal complex assembles, the final steps occur while the DNA is held in the complex.

The morphological changes that occur during homolog pairing are the basis for dividing meiotic prophase into five sequential stages—leptotene, zygotene, pachytene, diplotene, and diakinesis (Figure 17–56). Prophase starts with *leptotene*, when homologs condense and pair and genetic recombination begins. At *zygotene*, the synaptonemal complex begins to assemble at sites where the homologs are closely associated and recombination events are occurring. At *pachytene*, the assembly process is complete, and the homologs are synapsed

**Figure 17–54 Homolog pairing and crossing-over.** (A) The structure formed by two closely aligned duplicated homologs is called a **bivalent**. As in mitosis, the sister chromatids in each homolog are tightly connected along their entire lengths, as well as at their centromeres. At this stage, the homologs are usually joined by a protein complex called the *synaptonemal complex* (not shown; see Figure 17–55). (B) A later-stage bivalent in which a single crossover has occurred between nonsister chromatids. It is only when the synaptonemal complex disassembles and the paired homologs separate a little at the end of prophase I, as shown, that the crossover is seen microscopically as a thin connection between the homologs called a *chiasma*.

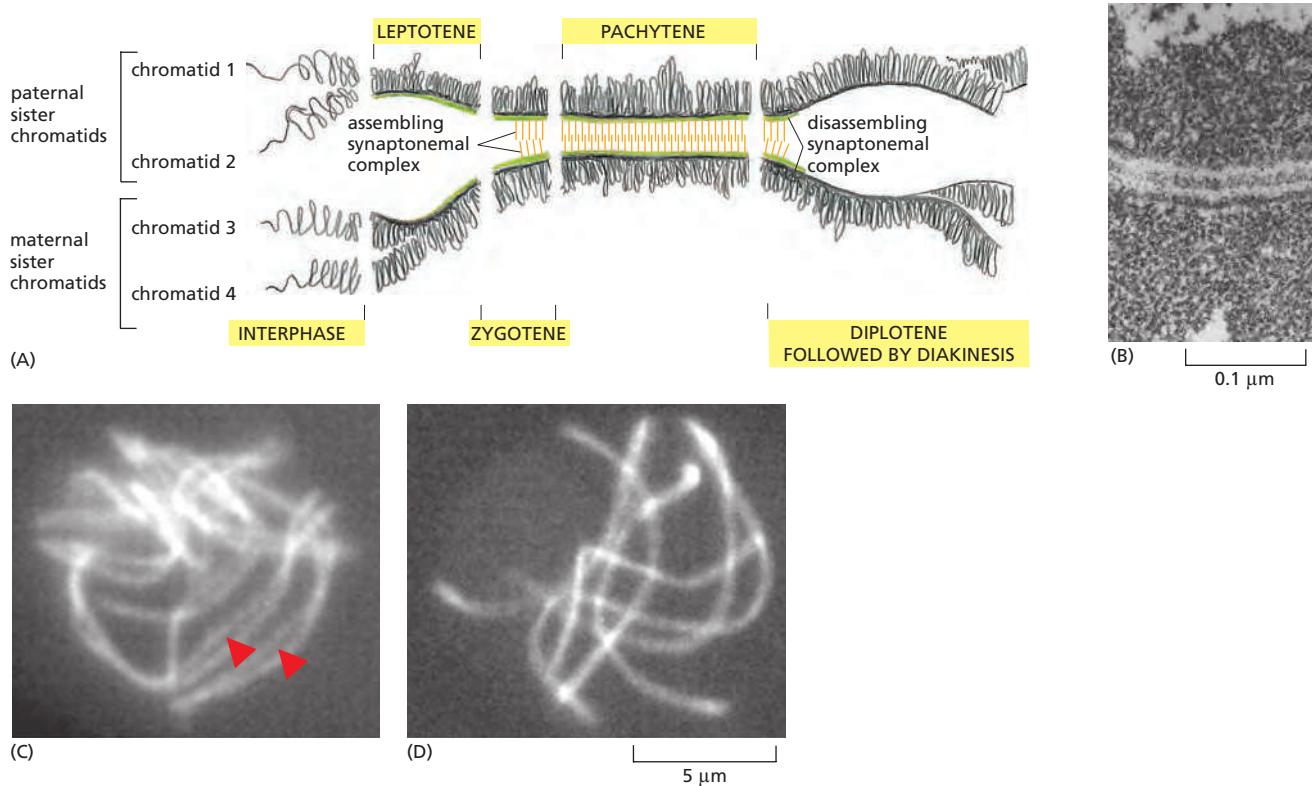




**Figure 17–55** Simplified schematic drawing of a synaptonemal complex.

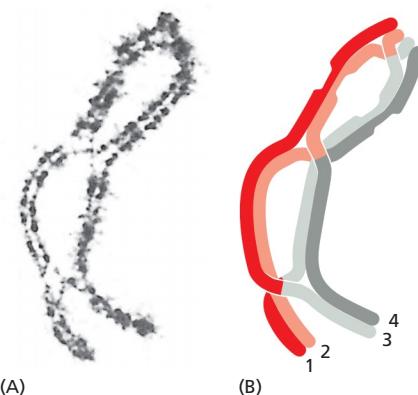
Each homolog is organized around a protein axial core, and the synaptonemal complex forms when these homolog axes are linked by rod-shaped transverse filaments. The axial core of each homolog also interacts with the cohesin complexes that hold the sister chromatids together (see Figure 9–35). (Modified from K. Nasmyth, *Annu. Rev. Genet.* 35:673–745, 2001.)

along their entire lengths (see Figure 9–35). The pachytene stage can persist for days or longer, until desynapsis begins at *diplotene* with the disassembly of the synaptonemal complexes and the concomitant condensation and shortening of the chromosomes. It is only at this stage, after the complexes have disassembled, that the individual crossover events between nonsister chromatids can be seen as inter-homolog connections called **chiasmata** (singular **chiasma**), which now play a crucial part in holding the compact homologs together (Figure 17–57). The homologs are now ready to begin the process of segregation.



**Figure 17–56** Homolog synapsis and desynthesis during the different stages of prophase I. (A) A single bivalent is shown schematically. At leptotene, the two sister chromatids coalesce, and their chromatin loops extend out from a common axial core. Assembly of the synaptonemal complex begins in early zygotene and is complete in pachytene. The complex disassembles in diplotene. (B) An electron micrograph of a synaptonemal complex from a meiotic cell at pachytene in a lily flower. (C and D) Immunofluorescence micrographs of prophase I cells of the fungus *Sordaria*. Partially synapsed bivalents are shown in (C) and fully synapsed bivalents are shown in (D). Red arrowheads in (C) point to regions where synapsis is still incomplete. (B, courtesy of Brian Wells; C and D, from A. Storlazzi et al., *Genes Dev.* 17:2675–2687, 2003. With permission from Cold Spring Harbor Laboratory Press.)

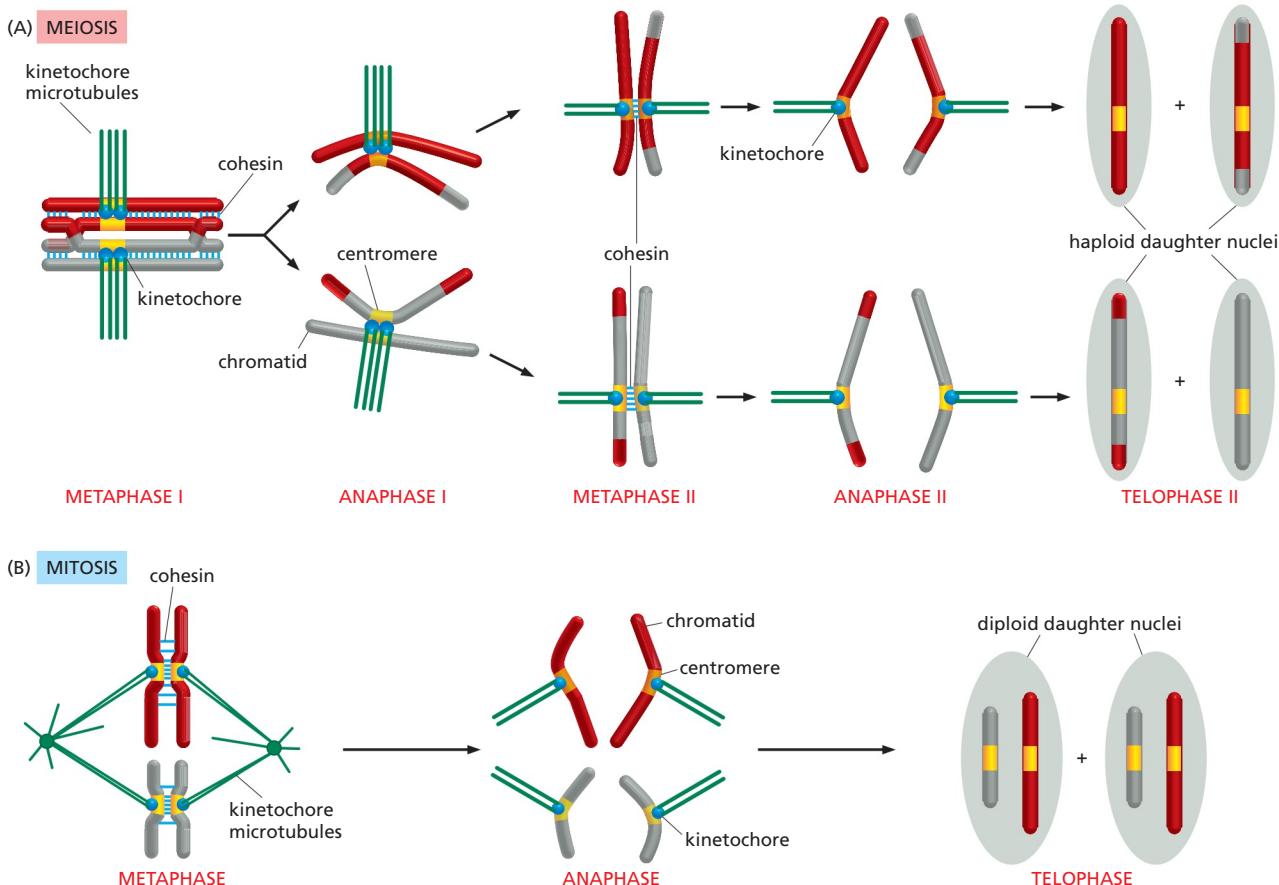
**Figure 17–57 A bivalent with three chiasmata resulting from three crossover events.** (A) Light micrograph of a grasshopper bivalent. (B) Drawing showing the arrangement of the crossovers in (A). Note that chromatid 1 has undergone an exchange with chromatid 3, and chromatid 2 has undergone exchanges with chromatids 3 and 4. Note also how the combination of the chiasmata and the tight attachment of the sister chromatid arms to each other (mediated by cohesin complexes) holds the two homologs together after the synaptonemal complex has disassembled; if either the chiasmata or the sister-chromatid cohesion failed to form, the homologs would come apart at this stage and not be segregated properly in meiosis I. (A, courtesy of Bernard John.)



### Homolog Segregation Depends on Several Unique Features of Meiosis I

A fundamental difference between meiosis I and mitosis (and meiosis II) is that in meiosis I homologs rather than sister chromatids separate and then segregate (see Figure 17–53). This difference depends on three features of meiosis I that distinguish it from mitosis (**Figure 17–58**).

First, both sister kinetochores in a homolog must attach stably to the same spindle pole. This type of attachment is normally avoided during mitosis (see Figure 17–33). In meiosis I, however, the two sister kinetochores are fused into a single microtubule-binding unit that attaches to just one pole (see Figure 17–58A). The fusion of sister kinetochores is achieved by a complex of proteins that is



**Figure 17–58 Comparison of chromosome behavior in meiosis I, meiosis II, and mitosis.** Chromosomes behave similarly in mitosis and meiosis II, but they behave very differently in meiosis I. (A) In meiosis I, the two sister kinetochores are located side-by-side on each homolog and attach to microtubules from the same spindle pole. The proteolytic cleavage of cohesin along the sister-chromatid arms unglues the arms and resolves the crossovers, allowing the duplicated homologs to separate at anaphase I, while the residual cohesin at the centromeres keeps the sisters together. Cleavage of centromeric cohesin allows the sister chromatids to separate at anaphase II. (B) In mitosis, by contrast, the two sister kinetochores attach to microtubules from different spindle poles, and the two sister chromatids come apart at the start of anaphase and segregate into separate daughter nuclei.

localized at the kinetochores in meiosis I, but we do not know in any detail how these proteins work. They are removed from kinetochores after meiosis I, so that in meiosis II the sister-chromatid pairs can be bi-oriented on the spindle as they are in mitosis.

Second, crossovers generate a strong physical linkage between homologs, allowing their bi-orientation at the equator of the spindle—much like cohesion between sister chromatids is important for their bi-orientation in mitosis (and meiosis II). Crossovers hold homolog pairs together only because the arms of the sister chromatids are connected by sister-chromatid cohesion (see Figure 17–58A).

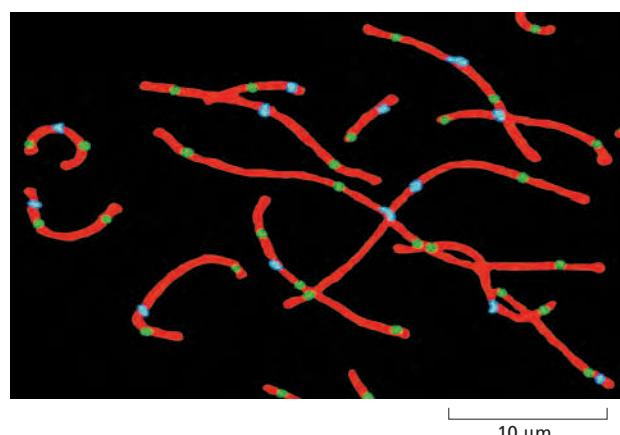
Third, cohesion is removed in anaphase I only from chromosome arms and not from the regions near the centromeres, where the kinetochores are located. The loss of arm cohesion triggers homolog separation at the onset of anaphase I. This process depends on APC/C activation, which leads to securin destruction, separase activation, and cohesin cleavage along the arms (see Figure 17–38).

Cohesins near the centromeres are protected from separase in meiosis I by a kinetochore-associated protein called *shugoshin* (from the Japanese word for “guardian spirit”). Shugoshin acts by recruiting a protein phosphatase that removes phosphates from centromeric cohesins. Cohesin phosphorylation is normally required for separase to cleave cohesin; thus, removal of this phosphorylation near the centromere prevents cohesin cleavage. Sister-chromatid pairs therefore remain linked through meiosis I, allowing their correct bi-orientation on the spindle in meiosis II. Shugoshin is inactivated after meiosis I. At the onset of anaphase II, APC/C activation triggers centromeric cohesin cleavage and sister-chromatid separation—much as it does in mitosis. Following anaphase II, nuclear envelopes form around the chromosomes to produce four haploid nuclei, after which cytokinesis and other differentiation processes lead to the production of haploid gametes.

### Crossing-Over Is Highly Regulated

Crossing-over has two distinct functions in meiosis: it helps hold homologs together so that they are properly segregated to the two daughter nuclei produced by meiosis I, and it contributes to the genetic diversification of the gametes that are eventually produced. As might be expected, therefore, crossing-over is highly regulated: the number and location of double-strand breaks along each chromosome is controlled, as is the likelihood that a break will be converted into a crossover. On average, the result of this regulation is that each pair of human homologs is linked by about two or three crossovers (Figure 17–59).

Although the double-strand breaks that occur in meiosis I can be located almost anywhere along the chromosome, they are not distributed uniformly: they cluster at “hot spots,” where the DNA is accessible, and occur only rarely in “cold spots,” such as the heterochromatin regions around centromeres and telomeres.



**Figure 17–59** Crossovers between homologs in the human testis. In these immunofluorescence micrographs, antibodies have been used to stain the synaptonemal complexes (red), the centromeres (blue), and the sites of crossing-over (green). Note that all of the bivalents have at least one crossover and none have more than four. (Modified from A. Lynn et al., *Science* 296:2222–2225, 2002. With permission from AAAS.)

At least two kinds of regulation influence the location and number of crossovers that form, neither of which is well understood. Both operate before the synaptonemal complex assembles. One ensures that at least one crossover forms between the members of each homolog pair, as is necessary for normal homolog segregation in meiosis I. In the other, called *crossover interference*, the presence of one crossover event inhibits another from forming close by, perhaps by locally depleting proteins required for converting a double-strand DNA break into a stable crossover.

### Meiosis Frequently Goes Wrong

The sorting of chromosomes that takes place during meiosis is a remarkable feat of intracellular bookkeeping. In humans, each meiosis requires that the starting cell keep track of 92 chromatids (46 chromosomes, each of which has duplicated), distributing one complete set of each type of autosome to each of the four haploid progeny. Not surprisingly, mistakes can occur in allocating the chromosomes during this elaborate process. Mistakes are especially common in human female meiosis, which arrests for years after diplotene: meiosis I is completed only at *ovulation*, and meiosis II only after the egg is fertilized. Indeed, such chromosome segregation errors during egg development are the most common cause of both spontaneous abortion (miscarriage) and mental retardation in humans.

When homologs fail to separate properly—a phenomenon called **nondisjunction**—the result is that some of the resulting haploid gametes lack a particular chromosome, while others have more than one copy of it. Upon fertilization, these gametes form abnormal embryos, most of which die. Some survive, however. *Down syndrome* in humans, for example, which is the leading cause of mental retardation, is caused by an extra copy of chromosome 21, usually resulting from nondisjunction during meiosis I in the female ovary. Segregation errors during meiosis I increase greatly with advancing maternal age.

### Summary

*Haploid gametes are produced by meiosis, in which a diploid nucleus undergoes two successive cell divisions after one round of DNA replication. Meiosis is dominated by a prolonged prophase. At the start of prophase, the chromosomes have replicated and consist of two tightly joined sister chromatids. Homologous chromosomes then pair up and become progressively more closely juxtaposed as prophase proceeds. The tightly aligned homologs undergo genetic recombination, forming crossovers that help hold each pair of homologs together during metaphase I. Meiosis-specific, kinetochore-associated proteins help ensure that both sister chromatids in a homolog attach to the same spindle pole; other kinetochore-associated proteins ensure that the homologs remain connected at their centromeres during anaphase I, so that homologs rather than sister chromatids are segregated in meiosis I. After meiosis I, meiosis II follows rapidly, without DNA replication, in a process that resembles mitosis, in that sister chromatids are pulled apart at anaphase.*

## CONTROL OF CELL DIVISION AND CELL GROWTH

A fertilized mouse egg and a fertilized human egg are similar in size, yet they produce animals of very different sizes. What factors in the control of cell behavior in humans and mice are responsible for these size differences? The same fundamental question can be asked for each organ and tissue in an animal's body. What factors determine the length of an elephant's trunk or the size of its brain or its liver? These questions are largely unanswered, but it is nevertheless possible to say what the ingredients of an answer must be.

The size of an organ or organism depends on its total cell mass, which depends on both the total number of cells and their size. Cell number, in turn, depends on the amounts of cell division and cell death. Organ and body size are therefore determined by three fundamental processes: cell growth, cell division, and cell

survival. Each is tightly regulated—both by intracellular programs and by extracellular signal molecules that control these programs.

The extracellular signal molecules that regulate cell growth, division, and survival are generally soluble secreted proteins, proteins bound to the surface of cells, or components of the extracellular matrix. They can be divided operationally into three major classes:

1. *Mitogens*, which stimulate cell division, primarily by triggering a wave of G<sub>1</sub>/S-Cdk activity that relieves intracellular negative controls that otherwise block progress through the cell cycle.
2. *Growth factors*, which stimulate cell growth (an increase in cell mass) by promoting the synthesis of proteins and other macromolecules and by inhibiting their degradation.
3. *Survival factors*, which promote cell survival by suppressing the form of programmed cell death known as *apoptosis*.

Many extracellular signal molecules promote all of these processes, while others promote one or two of them. Indeed, the term *growth factor* is often used inappropriately to describe a factor that has any of these activities. Even worse, the term *cell growth* is often used to mean an increase in cell number, or *cell proliferation*.

In addition to these three classes of stimulating signals, there are extracellular signal molecules that suppress cell proliferation, cell growth, or both; in general, less is known about them. There are also extracellular signal molecules that activate apoptosis.

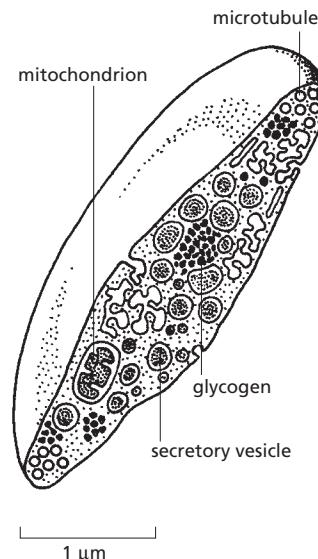
In this section, we focus primarily on how mitogens and other factors, such as DNA damage, control the rate of cell division. We then turn to the important but poorly understood problem of how a proliferating cell coordinates its growth with cell division so as to maintain its appropriate size. We discuss the control of cell survival and cell death by apoptosis in Chapter 18.

## Mitogens Stimulate Cell Division

Unicellular organisms tend to grow and divide as fast as they can, and their rate of proliferation depends largely on the availability of nutrients in the environment. The cells of a multicellular organism, however, divide only when the organism needs more cells. Thus, for an animal cell to proliferate, it must receive stimulatory extracellular signals, in the form of **mitogens**, from other cells, usually its neighbors. Mitogens overcome intracellular braking mechanisms that block progress through the cell cycle.

One of the first mitogens to be identified was *platelet-derived growth factor* (PDGF), and it is typical of many others discovered since. The path to its isolation began with the observation that fibroblasts in a culture dish proliferate when provided with *serum* but not when provided with *plasma*. Plasma is prepared by removing the cells from blood without allowing clotting to occur; serum is prepared by allowing blood to clot and taking the cell-free liquid that remains. When blood clots, platelets incorporated in the clot are stimulated to release the contents of their secretory vesicles (Figure 17–60). The superior ability of serum to support cell proliferation suggested that platelets contain one or more mitogens. This hypothesis was confirmed by showing that extracts of platelets could serve instead of serum to stimulate fibroblast proliferation. The crucial factor in the extracts was shown to be a protein, which was subsequently purified and named PDGF. In the body, PDGF liberated from blood clots helps stimulate cell division during wound healing.

PDGF is only one of over 50 animal proteins that are known to act as mitogens. Most of these proteins have a broad specificity. PDGF, for example, can stimulate many types of cells to divide, including fibroblasts, smooth muscle cells, and neuroglial cells. Similarly, *epidermal growth factor* (EGF) acts not only on epidermal cells but also on many other cell types, including both epithelial and nonepithelial cells. Some mitogens, however, have a narrow specificity; *erythropoietin*, for



**Figure 17–60 A platelet.** Platelets are miniature cells without a nucleus. They circulate in the blood and help stimulate blood clotting at sites of tissue damage, thereby preventing excessive bleeding. They also release various factors that stimulate wound healing. The platelet shown here has been cut in half to show its secretory vesicles, some of which contain platelet-derived growth factor (PDGF).

example, only induces the proliferation of red blood cell precursors. Many mitogens, including PDGF, also have actions other than the stimulation of cell division: they can stimulate cell growth, survival, differentiation, or migration, depending on the circumstances and the cell type.

In some tissues, inhibitory extracellular signal proteins oppose the positive regulators and thereby inhibit organ growth. The best-understood inhibitory signal proteins are transforming growth factor- $\beta$  (TGF $\beta$ ) and its relatives. TGF $\beta$  inhibits the proliferation of several cell types, mainly by blocking cell-cycle progression in G<sub>1</sub>.

### Cells Can Enter a Specialized Nondividing State

In the absence of a mitogenic signal to proliferate, Cdk inhibition in G<sub>1</sub> is maintained by the multiple mechanisms discussed earlier, and progression into a new cell cycle is blocked. In some cases, cells partly disassemble their cell-cycle control system and withdraw from the cycle to a specialized nondividing state called G<sub>0</sub>.

Most cells in our body are in G<sub>0</sub>, but the molecular basis and reversibility of this state vary in different cell types. Most of our neurons and skeletal muscle cells, for example, are in a *terminally differentiated* G<sub>0</sub> state, in which their cell-cycle control system is completely dismantled: the expression of the genes encoding various Cdks and cyclins is permanently turned off, and cell division rarely occurs. Some cell types withdraw from the cell cycle only transiently and retain the ability to reassemble the cell-cycle control system quickly and re-enter the cycle. Most liver cells, for example, are in G<sub>0</sub>, but they can be stimulated to divide if the liver is damaged. Still other types of cells, including fibroblasts and some lymphocytes, withdraw from and re-enter the cell cycle repeatedly throughout their lifetime.

Almost all the variation in cell-cycle length in the adult body occurs during the time the cell spends in G<sub>1</sub> or G<sub>0</sub>. By contrast, the time a cell takes to progress from the beginning of S phase through mitosis is usually brief (typically 12–24 hours in mammals) and relatively constant, regardless of the interval from one division to the next.

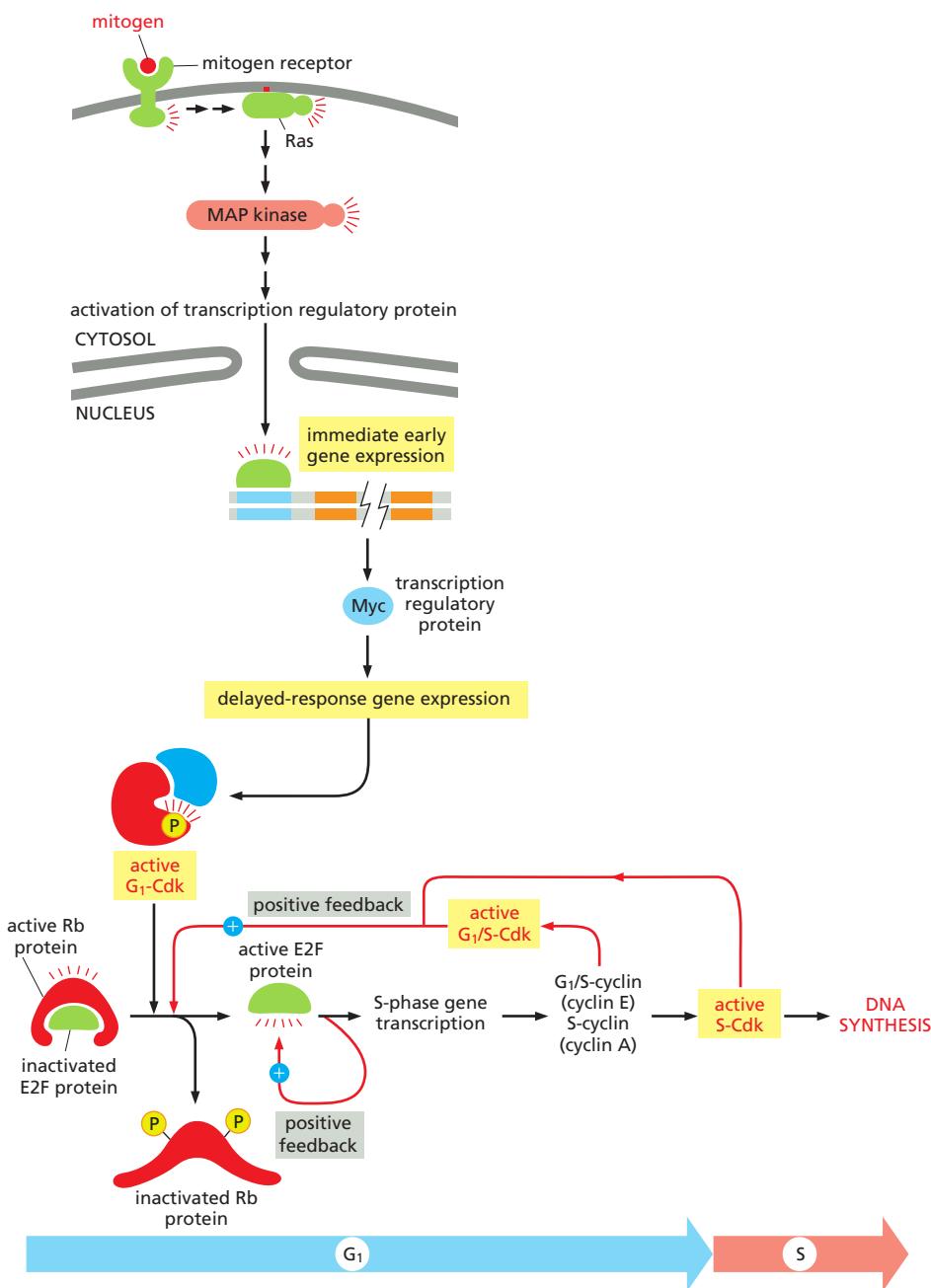
### Mitogens Stimulate G<sub>1</sub>-Cdk and G<sub>1</sub>/S-Cdk Activities

For the vast majority of animal cells, mitogens control the rate of cell division by acting in the G<sub>1</sub> phase of the cell cycle. As discussed earlier, multiple mechanisms act during G<sub>1</sub> to suppress Cdk activity. Mitogens release these brakes on Cdk activity, thereby allowing entry into a new cell cycle.

As we discuss in Chapter 15, mitogens interact with cell-surface receptors to trigger multiple intracellular signaling pathways. One major pathway acts through the monomeric GTPase **Ras**, which leads to the activation of a *mitogen-activated protein kinase (MAP kinase) cascade* (see Figure 15–49). This leads to an increase in the production of transcription regulatory proteins, including **Myc**. Myc is thought to promote cell-cycle entry by several mechanisms, one of which is to increase the expression of genes encoding G<sub>1</sub> cyclins (D cyclins), thereby increasing G<sub>1</sub>-Cdk (cyclin D-Cdk4) activity. Myc also has a major role in stimulating the transcription of genes that increase cell growth.

The key function of G<sub>1</sub>-Cdk complexes in animal cells is to activate a group of gene regulatory factors called the **E2F proteins**, which bind to specific DNA sequences in the promoters of a wide variety of genes that encode proteins required for S-phase entry, including G<sub>1</sub>/S-cyclins, S-cyclins, and proteins involved in DNA synthesis and chromosome duplication. In the absence of mitogenic stimulation, E2F-dependent gene expression is inhibited by an interaction between E2F and members of the **retinoblastoma protein (Rb)** family. When cells are stimulated to divide by mitogens, active G<sub>1</sub>-Cdk accumulates and phosphorylates Rb family members, reducing their binding to E2F. The liberated E2F proteins then activate expression of their target genes (**Figure 17–61**).

This transcriptional control system, like so many other control systems that regulate the cell cycle, includes feedback loops that ensure that entry into the



**Figure 17–61** Mitogen stimulation of cell-cycle entry. As discussed in Chapter 15, mitogens bind to cell-surface receptors to initiate intracellular signaling pathways. One of the major pathways involves activation of the small GTPase Ras, which activates a MAP kinase cascade, leading to increased expression of numerous *immediate early* genes, including the gene encoding the transcription regulatory protein Myc. Myc increases the expression of many *delayed-response* genes, including some that lead to increased G<sub>1</sub>-Cdk activity (cyclin D–Cdk4), which triggers the phosphorylation of members of the Rb family of proteins. This inactivates the Rb proteins, freeing the gene regulatory protein E2F to activate the transcription of G<sub>1</sub>/S genes, including the genes for a G<sub>1</sub>/S-cyclin (cyclin E) and S-cyclin (cyclin A). The resulting G<sub>1</sub>/S-Cdk and S-Cdk activities further enhance Rb protein phosphorylation, forming a positive feedback loop. E2F proteins also stimulate the transcription of their own genes, forming another positive feedback loop.

cell cycle is complete and irreversible. The liberated E2F proteins, for example, increase the transcription of their own genes. In addition, E2F-dependent transcription of G<sub>1</sub>/S-cyclin (cyclin E) and S-cyclin (cyclin A) genes leads to increased G<sub>1</sub>/S-Cdk and S-Cdk activities, which in turn increase Rb protein phosphorylation and promote further E2F release (see Figure 17–61).

The central member of the Rb family, the Rb protein itself, was identified originally through studies of an inherited form of eye cancer in children, known as *retinoblastoma* (discussed in Chapter 20). The loss of both copies of the *Rb* gene leads to excessive proliferation of some cells in the developing retina, suggesting that the Rb protein is particularly important for restraining cell division in this tissue. The complete loss of Rb does not immediately cause increased proliferation of retinal or other types of cells, in part because Cdh1 and CKIs also help inhibit progression through G<sub>1</sub> and in part because other cell types contain Rb-related proteins that provide backup support in the absence of Rb. It is also likely that other proteins, unrelated to Rb, help to regulate the activity of E2F.

Additional layers of control promote an overwhelming increase in S-Cdk activity at the beginning of S phase. We mentioned earlier that the APC/C activator Cdh1 suppresses cyclin levels after mitosis. In animal cells, however, G<sub>1</sub>- and G<sub>1</sub>/S-cyclins are resistant to Cdh1-APC/C and can therefore act unopposed by the APC/C to promote Rb protein phosphorylation and E2F-dependent gene expression. S-cyclin, by contrast, is not resistant, and its level is initially restrained by Cdh1-APC/C activity. However, G<sub>1</sub>/S-Cdk also phosphorylates and inactivates Cdh1-APC/C, thereby allowing the accumulation of S-cyclin, further promoting S-Cdk activation. G<sub>1</sub>/S-Cdk also inactivates CKI proteins that suppress S-Cdk activity. The overall effect of all these interactions is the rapid and complete activation of the S-Cdk complexes required for S-phase initiation.

### DNA Damage Blocks Cell Division: The DNA Damage Response

Progression through the cell cycle, and thus the rate of cell proliferation, is controlled not only by extracellular mitogens but also by other extracellular and intracellular signals. One of the most important influences is DNA damage, which can occur as a result of spontaneous chemical reactions in DNA, errors in DNA replication, or exposure to radiation or certain chemicals (discussed in Chapter 5). It is essential that damaged chromosomes are repaired before attempting to duplicate or segregate them. The cell-cycle control system can readily detect DNA damage and arrest the cycle at either of two transitions—one at Start, which prevents entry into the cell cycle and into S phase, and one at the G<sub>2</sub>/M transition, which prevents entry into mitosis (see Figure 17–16).

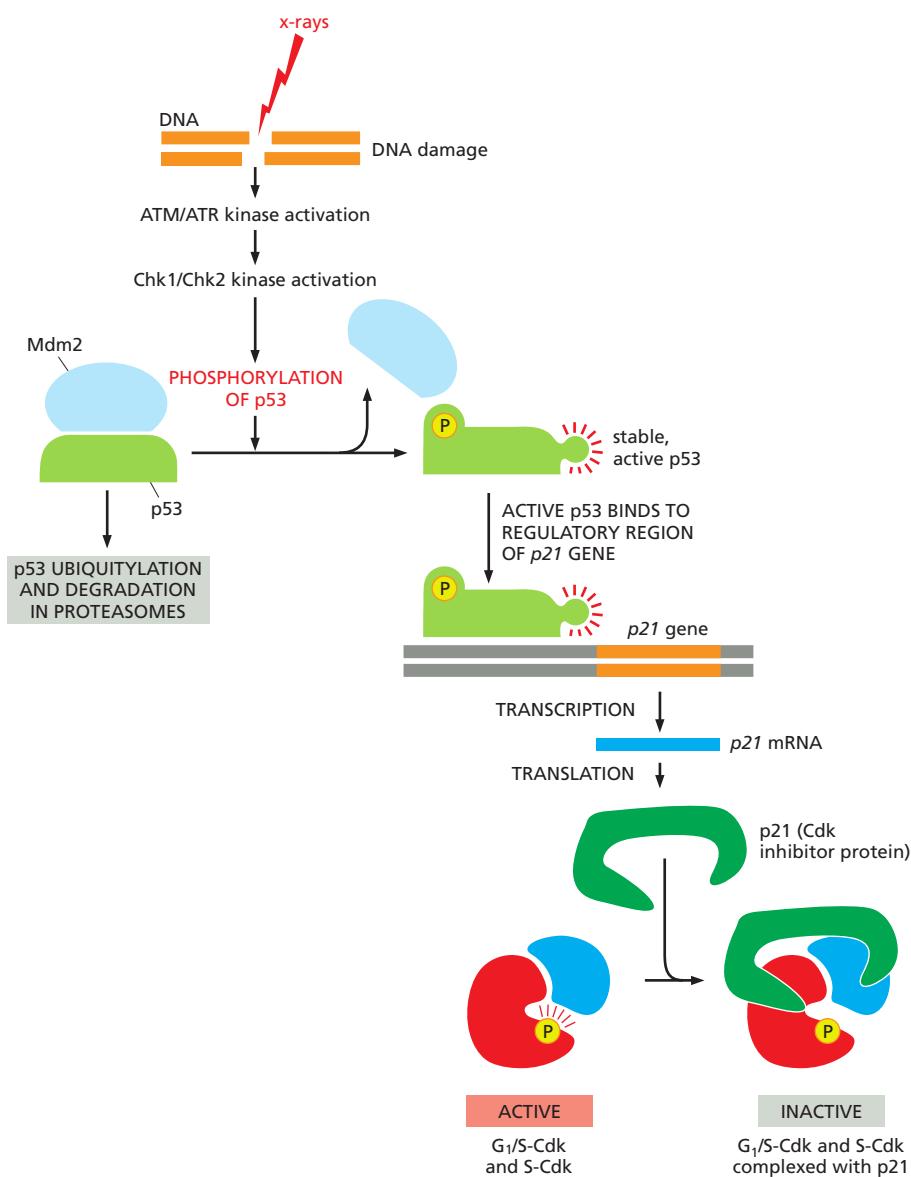
DNA damage initiates a signaling pathway by activating one of a pair of related protein kinases called ATM and ATR, which associate with the site of damage and phosphorylate various target proteins, including two other protein kinases called Chk1 and Chk2. These various kinases phosphorylate other target proteins that lead to cell-cycle arrest. A major target is the gene regulatory protein p53, which stimulates transcription of the gene encoding p21, a CKI protein; p21 binds to G<sub>1</sub>/S-Cdk and S-Cdk complexes and inhibits their activities, thereby helping to block entry into the cell cycle (Figure 17–62 and Movie 17.8).

DNA damage activates p53 by an indirect mechanism. In undamaged cells, p53 is highly unstable and is present at very low concentrations. This is largely because it interacts with another protein, Mdm2, which acts as a ubiquitin ligase that targets p53 for destruction by proteasomes. Phosphorylation of p53 after DNA damage reduces its binding to Mdm2. This decreases p53 degradation, which results in a marked increase in p53 concentration in the cell. In addition, the decreased binding to Mdm2 enhances the ability of p53 to stimulate gene transcription (see Figure 17–62).

The protein kinases Chk1 and Chk2 also block cell-cycle progression by phosphorylating members of the Cdc25 family of protein phosphatases, thereby inhibiting their function. As described earlier, these phosphatases are particularly important in the activation of M-Cdk at the beginning of mitosis (see Figure 17–20). Chk1 and Chk2 phosphorylate Cdc25 at inhibitory sites that are distinct from the phosphorylation sites that stimulate Cdc25 activity. The inhibition of Cdc25 activity by DNA damage helps block entry into mitosis (see Figure 17–16).

The DNA damage response can also be activated by problems that arise when a replication fork fails during DNA replication. When nucleotides are depleted, for example, replication forks stall during the elongation phase of DNA synthesis. To prevent the cell from attempting to segregate partially replicated chromosomes, the same mechanisms that respond to DNA damage detect the stalled replication forks and block entry into mitosis until the problems are resolved.

A low level of DNA damage occurs in the normal life of any cell, and this damage accumulates in the cell's progeny if the DNA damage response is not functioning. Over the long term, the accumulation of genetic damage in cells lacking the DNA damage response leads to an increased frequency of cancer-promoting mutations. Indeed, mutations in the p53 gene occur in at least half of all human cancers (discussed in Chapter 20). This loss of p53 function allows the cancer cell



**Figure 17–62** How DNA damage arrests the cell cycle in G<sub>1</sub>. When DNA is damaged, various protein kinases are recruited to the site of damage and initiate a signaling pathway that causes cell-cycle arrest. The first kinase at the damage site is either ATM or ATR, depending on the type of damage. Additional protein kinases, called Chk1 and Chk2, are then recruited and activated, resulting in the phosphorylation of the transcription regulatory protein p53. Mdm2 normally binds to p53 and promotes its ubiquitylation and destruction in proteasomes. Phosphorylation of p53 blocks its binding to Mdm2; as a result, p53 accumulates to high levels and stimulates transcription of numerous genes, including the gene that encodes the CKI protein p21. The p21 binds and inactivates G<sub>1</sub>/S-Cdk and S-Cdk complexes, arresting the cell in G<sub>1</sub>. In some cases, DNA damage also induces either the phosphorylation of Mdm2 or a decrease in Mdm2 production, which causes a further increase in p53 (not shown).

to accumulate mutations more readily. Similarly, a rare genetic disease known as *ataxia telangiectasia* is caused by a defect in ATM, one of the protein kinases that are activated in response to x-ray-induced DNA damage; patients with this disease are very sensitive to x-rays and suffer from increased rates of cancer.

What happens if DNA damage is so severe that repair is not possible? The answer differs in different organisms. Unicellular organisms such as budding yeast arrest their cell cycle to try to repair the damage, but the cycle resumes even if the repair cannot be completed. For a single-celled organism, life with mutations is apparently better than no life at all. In multicellular organisms, however, the health of the organism takes precedence over the life of an individual cell. Cells that divide with severe DNA damage threaten the life of the organism, since genetic damage can often lead to cancer and other diseases. Thus, animal cells with severe DNA damage do not attempt to continue division, but instead commit suicide by undergoing apoptosis. Thus, unless the DNA damage is repaired, the DNA damage response can lead to either cell-cycle arrest or cell death. DNA damage-induced apoptosis often depends on the activation of p53. Indeed, it is this apoptosis-promoting function of p53 that is apparently most important in protecting us against cancer.

## Many Human Cells Have a Built-In Limitation on the Number of Times They Can Divide

Many human cells divide a limited number of times before they stop and undergo a permanent cell-cycle arrest. Fibroblasts taken from normal human tissue, for example, go through only about 25–50 population doublings when cultured in a standard mitogenic medium. Toward the end of this time, proliferation slows down and finally halts, and the cells enter a nondividing state from which they never recover. This phenomenon is called **replicative cell senescence**.

Reactive cell senescence in human fibroblasts seems to be caused by changes in the structure of the **telomeres**, the repetitive DNA sequences and associated proteins at the ends of chromosomes. As discussed in Chapter 5, when a cell divides, telomeric DNA sequences are not replicated in the same manner as the rest of the genome but instead are synthesized by the enzyme **telomerase**. Telomerase also promotes the formation of protein cap structures that protect the chromosome ends. Because human fibroblasts, and many other human somatic cells, do not produce telomerase, their telomeres become shorter with every cell division, and their protective protein caps progressively deteriorate. Eventually, the exposed chromosome ends are sensed as DNA damage, which activates a p53-dependent cell-cycle arrest (see Figure 17–62). Rodent cells, by contrast, maintain telomerase activity when they proliferate in culture and therefore do not have such a telomere-dependent mechanism for limiting proliferation. The forced expression of telomerase in normal human fibroblasts, using genetic engineering techniques, blocks this form of senescence. Unfortunately, most cancer cells have regained the ability to produce telomerase and therefore maintain telomere function as they proliferate; as a result, they do not undergo replicative cell senescence.

## Abnormal Proliferation Signals Cause Cell-Cycle Arrest or Apoptosis, Except in Cancer Cells

Many of the components of mitogenic signaling pathways are encoded by genes that were originally identified as cancer-promoting genes, because mutations in them contribute to the development of cancer. The mutation of a single amino acid in the small GTPase Ras, for example, causes the protein to become permanently overactive, leading to constant stimulation of Ras-dependent signaling pathways, even in the absence of mitogenic stimulation. Similarly, mutations that cause an overexpression of Myc stimulate excessive cell growth and proliferation and thereby promote the development of cancer (discussed in Chapter 20).

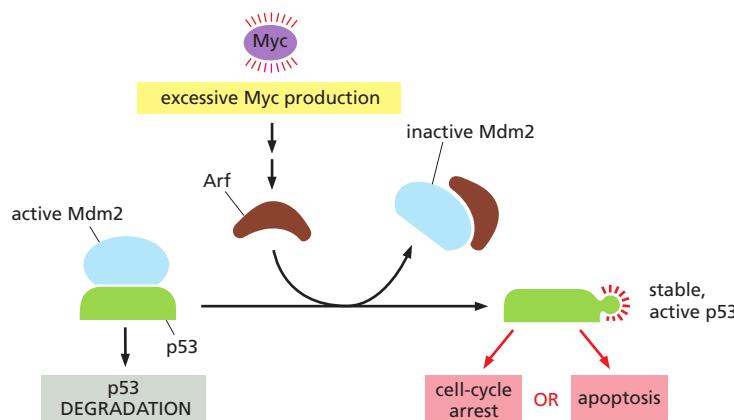
Surprisingly, however, when a hyperactivated form of Ras or Myc is experimentally overproduced in most normal cells, the result is not excessive proliferation but the opposite: the cells undergo either permanent cell-cycle arrest or apoptosis. The normal cell seems able to detect abnormal mitogenic stimulation, and it responds by preventing further division. Such responses help prevent the survival and proliferation of cells with various cancer-promoting mutations.

Although it is not known how a cell detects excessive mitogenic stimulation, such stimulation often leads to the production of a cell-cycle inhibitor protein called *Arf*, which binds and inhibits Mdm2. As discussed earlier, Mdm2 normally promotes p53 degradation. Activation of Arf therefore causes p53 levels to increase, inducing either cell-cycle arrest or apoptosis (Figure 17–63).

How do cancer cells ever arise if these mechanisms block the division or survival of mutant cells with overactive proliferation signals? The answer is that the protective system is often inactivated in cancer cells by mutations in the genes that encode essential components of the blocking mechanisms, such as Arf or p53 or the proteins that help activate them.

## Cell Proliferation is Accompanied by Cell Growth

If cells proliferated without growing, they would get progressively smaller and there would be no net increase in total cell mass. In most proliferating cell populations, therefore, cell growth accompanies cell division. In single-celled organisms

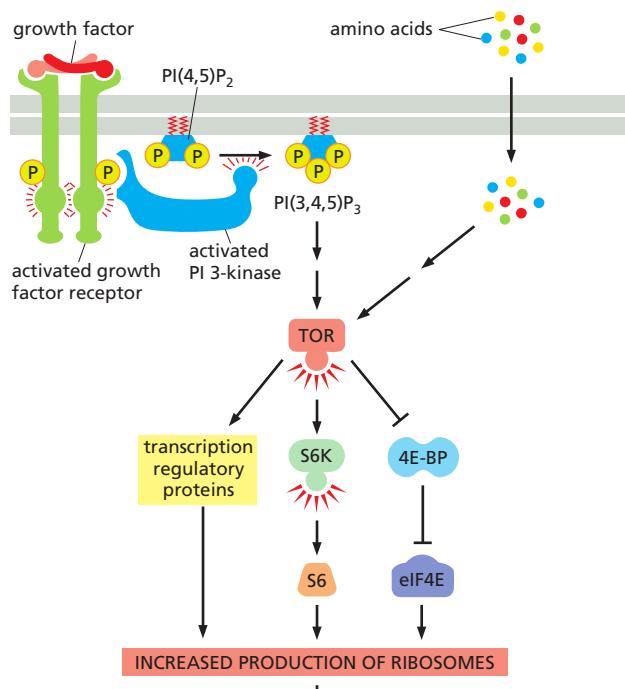


**Figure 17–63 Cell-cycle arrest or apoptosis induced by excessive stimulation of mitogenic pathways.**

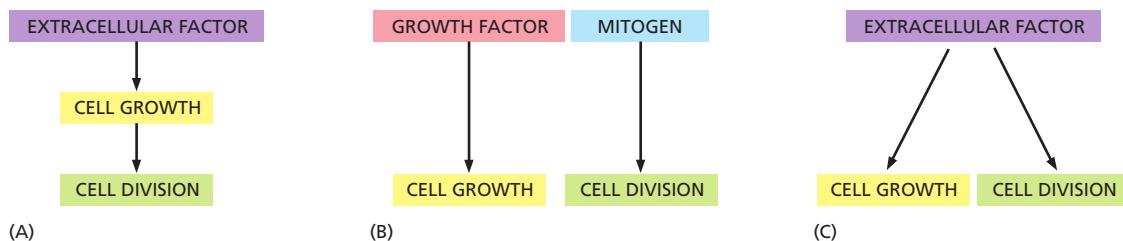
Abnormally high levels of Myc cause the activation of Arf, which binds and inhibits Mdm2 and thereby increases p53 levels (see Figure 17–62). Depending on the cell type and extracellular conditions, p53 then causes either cell-cycle arrest or apoptosis.

such as yeasts, both cell growth and cell division require only nutrients. In animals, by contrast, both cell growth and cell proliferation depend on extracellular signal molecules, produced by other cells, which we call **growth factors** and mitogens, respectively.

Like mitogens, the extracellular growth factors that stimulate animal cell growth bind to receptors on the cell surface and activate intracellular signaling pathways. These pathways stimulate the accumulation of proteins and other macromolecules, and they do so by both increasing their rate of synthesis and decreasing their rate of degradation. They also trigger increased uptake of nutrients and production of the ATP required to fuel the increased protein synthesis. One of the most important intracellular signaling pathways activated by growth factor receptors involves the enzyme phosphoinositide 3-kinase (*PI 3-kinase*), which adds a phosphate from ATP to the 3' position of inositol phospholipids in the plasma membrane (discussed in Chapter 15). The activation of PI 3-kinase leads to the activation of a kinase called *TOR*, which lies at the heart of cell growth regulatory pathways in all eukaryotes. TOR activates many targets in the cell that stimulate metabolic processes, including protein synthesis. One target is a protein kinase called *S6 kinase* (*S6K*), which phosphorylates ribosomal protein S6, increasing the ability of ribosomes to translate a subset of mRNAs that mostly encode ribosomal components. TOR also indirectly activates a translation initiation factor called *eIF4E* and directly activates transcription regulators that promote the increased expression of genes encoding ribosomal subunits (Figure 17–64).



**Figure 17–64 Stimulation of cell growth by extracellular growth factors and nutrients.** The occupation of cell-surface receptors by growth factors leads to the activation of PI 3-kinase, which promotes protein synthesis through a complex signaling pathway that leads to the activation of the protein kinase TOR; extracellular nutrients such as amino acids also help activate TOR. TOR phosphorylates multiple proteins to stimulate protein synthesis, as shown; it also inhibits protein degradation (not shown). Growth factors also stimulate increased production of the transcription regulatory protein Myc (not shown), which activates the transcription of various genes that promote cell metabolism and growth. 4E-BP is an inhibitor of the translation initiation factor eIF4E. PI(4,5)P<sub>2</sub>, phosphatidylinositol 4,5-bisphosphate; PI(3,4,5)P<sub>3</sub>, phosphatidylinositol 3,4,5-trisphosphate.



**Figure 17–65** Potential mechanisms for coordinating cell growth and division. In proliferating cells, cell size is maintained by mechanisms that coordinate rates of cell division and cell growth. Numerous alternative coupling mechanisms are thought to exist, and different cell types appear to employ different combinations of these mechanisms. (A) In many cell types—particularly yeast—the rate of cell division is governed by the rate of cell growth, so that division occurs only when growth rate achieves some minimal threshold; in yeasts, it is mainly the levels of extracellular nutrients that regulate the rate of cell growth and thereby the rate of cell division. (B) In some animal cell types, growth and division can each be controlled by separate extracellular factors (growth factors and mitogens, respectively), and cell size depends on the relative levels of the two types of factors. (C) Some extracellular factors can stimulate both cell growth and cell division by simultaneously activating signaling pathways that promote growth and other pathways that promote cell-cycle progression.

## Proliferating Cells Usually Coordinate Their Growth and Division

For proliferating cells to maintain a constant size, they must coordinate their growth with cell division to ensure that cell size doubles with each division: if cells grow too slowly, they will get smaller with each division, and if they grow too fast, they will get larger with each division. It is not clear how cells achieve this coordination, but it is likely to involve multiple mechanisms that vary in different organisms and even in different cell types of the same organism (Figure 17–65).

Animal cell growth and division are not always coordinated, however. In many cases, they are completely uncoupled to allow growth without division or division without growth. Muscle cells and nerve cells, for example, can grow dramatically after they have permanently withdrawn from the cell cycle. Similarly, the eggs of many animals grow to an extremely large size without dividing; after fertilization, however, this relationship is reversed, and many rounds of division occur without growth.

Compared to cell division, there has been surprisingly little study of how cell size is controlled in animals. As a result, it remains a mystery how cell size is determined and why different cell types in the same animal grow to be so different in size. One of the best-understood cases in mammals is the adult *sympathetic neuron*, which has permanently withdrawn from the cell cycle. Its size depends on the amount of *nerve growth factor* (NGF) secreted by the target cells it innervates; the greater the amount of NGF the neuron has access to, the larger it becomes. It seems likely that the genes a cell expresses set limits on the size it can be, while extracellular signal molecules and nutrients regulate the size within these limits. The challenge is to identify the relevant genes and signal molecules for each cell type.

## Summary

In multicellular animals, cell size, cell division, and cell survival are carefully controlled to ensure that the organism and its organs achieve and maintain an appropriate size. Mitogens stimulate the rate of cell division by removing intracellular molecular brakes that restrain cell-cycle progression in  $G_1$ . Growth factors promote cell growth (an increase in cell mass) by stimulating the synthesis and inhibiting the degradation of macromolecules. To maintain a constant cell size, proliferating cells employ multiple mechanisms to ensure that cell growth is coordinated with cell division.

## WHAT WE DON'T KNOW

- Progression through the cell cycle depends on the phosphorylation of hundreds of different proteins by cyclin–Cdk complexes. What are the molecular mechanisms ensuring that these proteins are phosphorylated at precisely the right time and place?
- During S phase, how are histones and their modifying enzymes controlled to replicate chromatin structure on the duplicated DNA?
- What is the structural basis of chromosome condensation, and how is the process stimulated during mitosis?
- What are the mechanisms by which microtubule attachment and tension are sensed at the kinetochore by the components of the spindle assembly checkpoint?
- How is cell growth coordinated with cell division to ensure that cell size remains constant?

## PROBLEMS

Which statements are true? Explain why or why not.

**17–1** Since there are about  $10^{13}$  cells in an adult human, and about  $10^{10}$  cells die and are replaced each day, we become new people every three years.

**17–2** In order for proliferating cells to maintain a relatively constant size, the length of the cell cycle must match the time it takes for the cell to double in size.

**17–3** While other proteins come and go during the cell cycle, the proteins of the origin recognition complex remain bound to the DNA throughout.

**17–4** Chromosomes are positioned on the metaphase plate by equal and opposite forces that pull them toward the two poles of the spindle.

**17–5** Meiosis segregates the paternal homologs into sperm and the maternal homologs into eggs.

**17–6** If we could turn on telomerase activity in all our cells, we could prevent aging.

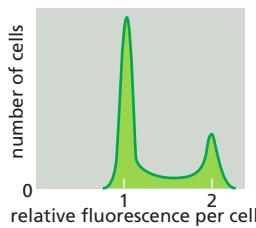
Discuss the following problems.

**17–7** Many cell-cycle genes from human cells function perfectly well when expressed in yeast cells. Why do you suppose that is considered remarkable? After all, many human genes encoding enzymes for metabolic reactions also function in yeast, and no one thinks that is remarkable.

**17–8** Hoechst 33342 is a membrane-permeant dye that fluoresces when it binds to DNA. When a population of cells is incubated briefly with Hoechst dye and then sorted in a flow cytometer, which measures the fluorescence of each cell, the cells display various levels of fluorescence as shown in **Figure Q17–1**.

**A.** Which cells in Figure Q17–1 are in the G<sub>1</sub>, S, G<sub>2</sub>, and M phases of the cell cycle? Explain the basis for your answer.

**B.** Sketch the sorting distributions you would expect for cells that were treated with inhibitors that block the cell cycle in the G<sub>1</sub>, S, or M phase. Explain your reasoning.



**Figure Q17–1** Analysis of Hoechst 33342 fluorescence in a population of cells sorted in a flow cytometer (Problem 17–8).

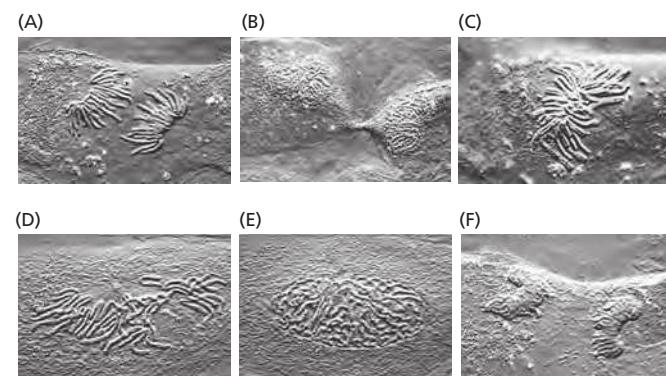
**17–9** The yeast cohesin subunit Scc1, which is essential for sister-chromatid cohesion, can be artificially regulated for expression at any point in the cell cycle. If expression is turned on at the beginning of S phase, all the cells divide satisfactorily and survive. By contrast, if Scc1 expression is turned on only after S phase is completed,

the cells fail to divide and they die, even though Scc1 accumulates in the nucleus and interacts efficiently with chromosomes. Why do you suppose that cohesin must be present during S phase for cells to divide normally?

**17–10** High doses of caffeine interfere with the DNA damage response in mammalian cells. Why then do you suppose the Surgeon General has not yet issued an appropriate warning to heavy coffee and cola drinkers? A typical cup of coffee (150 mL) contains 100 mg of caffeine (196 g/mole). How many cups of coffee would you have to drink to reach the dose (10 mM) required to interfere with the DNA damage response? (A typical adult contains about 40 liters of water.)

**17–11** How many kinetochores are there in a human cell at mitosis?

**17–12** A living cell from the lung epithelium of a newt is shown at different stages in M phase in **Figure Q17–2**. Order these light micrographs into the correct sequence and identify the stage in M phase that each represents.



**Figure Q17–2** Light micrographs of a single cell at different stages of M phase (Problem 17–12). (Courtesy of Conly L. Rieder.)

**17–13** Down syndrome (trisomy 21) and Edwards syndrome (trisomy 18) are the most common autosomal trisomies seen in human infants. Does this fact mean that these chromosomes are the most difficult to segregate properly during meiosis?

**17–14** The human genome consists of 23 pairs of chromosomes (22 pairs of autosomes and one pair of sex chromosomes). During meiosis, the maternal and paternal sets of homologs pair, and then are separated into gametes, so that each contains 23 chromosomes. If you assume that the chromosomes in the paired homologs are randomly assorted to daughter cells, how many potential combinations of paternal and maternal homologs can be generated during meiosis? (For the purposes of this calculation, assume that no recombination occurs.)

## REFERENCES

### Overview of the Cell Cycle

- Morgan DO (2007) *The Cell Cycle: Principles of Control*. London: New Science Press.
- Murray AW & Hunt T (1993) *The Cell Cycle: An Introduction*. New York: WH Freeman and Co.

### The Cell-Cycle Control System

- Evans T, Rosenthal ET, Youngblom J et al. (1983) Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* 33, 389–396.
- Hartwell LH, Culotti J, Pringle JR et al (1974) Genetic control of the cell division cycle in yeast. *Science* 183, 46–51.
- Holt LJ, Tuch BB, Villen J et al. (2009) Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 325, 1682–1686.
- Nurse P, Thuriaux P & Nasmyth K (1976) Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* 146, 167–178.
- Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and CIP and Ink4 inhibitors. *J. Mol. Biol.* 287, 821–828.
- Primorac I & Musacchio A (2013) Panta rheo: The APC/C at steady state. *J. Cell Biol.* 201, 177–189.
- Wittenberg C & Reed SI (2005) Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* 24, 2746–2755.

### S Phase

- Arias EE & Walter JC (2007) Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev.* 21, 497–518.
- Bell SP & Kaguni JM (2013) Helicase loading at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* 5, a010124.
- Groth A, Rocha W, Verreault A et al. (2007) Chromatin challenges during DNA replication and repair. *Cell* 128, 721–733.
- Masai H, Matsumoto S, You Z et al. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* 79, 89–130.
- Siddiqui K, On KF & Diffley JF (2013) Regulating DNA replication in eukarya. *Cold Spring Harb. Perspect. Biol.* 5, a012930.
- Tanaka S & Araki H (2013) Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* 5, a010371.

### Mitosis

- Alushin G & Nogales E (2011) Visualizing kinetochore architecture. *Curr. Opin. Struct. Biol.* 21, 661–669.
- Cuylen S & Haering CH (2011) Deciphering condensin action during chromosome segregation. *Trends Cell Biol.* 21, 552–559.
- Gonczy P (2012) Towards a molecular architecture of centriole assembly. *Nat. Rev. Mol. Cell Biol.* 13, 425–435.
- Hirano T (2012) Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev.* 26, 1659–1678.
- Joglekar AP, Bloom KS & Salmon ED (2010) Mechanisms of force generation by end-on kinetochore-microtubule attachments. *Curr. Opin. Cell Biol.* 22, 57–67.
- Lampson MA & Cheeseman IM (2011) Sensing centromere tension: Aurora B and the regulation of kinetochore function. *Trends Cell Biol.* 21, 133–140.
- Magidson V, O'Connell CB, Loncarek J et al. (2011) The spatial arrangement of chromosomes during prometaphase facilitates spindle assembly. *Cell* 146, 555–567.
- Musacchio A & Salmon ED (2007) The spindle-assembly checkpoint in space and time. *Nat. Rev. Mol. Cell Biol.* 8, 379–393.
- Nasmyth K & Haering CH (2009) Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* 43, 525–558.

Nigg EA & Stearns T (2011) The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. *Nat. Cell Biol.* 13, 1154–1160.

Rago F & Cheeseman IM (2013) The functions and consequences of force at kinetochores. *J. Cell Biol.* 200, 557–565.

Wadsworth P & Khodjakov A (2004) E pluribus unum: towards a universal mechanism for spindle assembly. *Trends Cell Biol.* 14, 413–419.

Walczak CE, Cai S & Khodjakov A (2010) Mechanisms of chromosome behaviour during mitosis. *Nat. Rev. Mol. Cell Biol.* 11, 91–102.

### Cytokinesis

Fededa JP & Gerlich DW (2012) Molecular control of animal cell cytokinesis. *Nat. Cell Biol.* 14, 440–447.

Green RA, Paluch E & Oegema K (2012) Cytokinesis in animal cells. *Annu. Rev. Cell Dev. Biol.* 28, 29–58.

Jurgens G (2005) Plant cytokinesis: fission by fusion. *Trends Cell Biol.* 15, 277–283.

Oliferenko S, Chew TG & Balasubramanian MK (2009) Positioning cytokinesis. *Genes Dev.* 23, 660–674.

Pollard TD (2010) Mechanics of cytokinesis in eukaryotes. *Curr. Opin. Cell Biol.* 22, 50–56.

Rappaport R (1986) Establishment of the mechanism of cytokinesis in animal cells. *Int. Rev. Cytol.* 105, 245–281.

Schiel JA & Prekeris R (2013) Membrane dynamics during cytokinesis. *Curr. Opin. Cell Biol.* 25, 92–98.

### Meiosis

Bhalla N & Dernburg AF (2008) Prelude to a division. *Annu. Rev. Cell Dev. Biol.* 24, 397–424.

Gerton JL & Hawley RS (2005) Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nat. Rev. Genet.* 6, 477–487.

Hall H, Hunt P & Hassold T (2006) Meiosis and sex chromosome aneuploidy: how meiotic errors cause aneuploidy; how aneuploidy causes meiotic errors. *Curr. Opin. Genet. Dev.* 16, 323–329.

Jordan P (2006) Initiation of homologous chromosome pairing during meiosis. *Biochem. Soc. Trans.* 34, 545–549.

Lake CM & Hawley RS (2012) The molecular control of meiotic chromosomal behavior: events in early meiotic prophase in *Drosophila* oocytes. *Annu. Rev. Physiol.* 74, 425–51.

Watanabe Y (2012) Geometry and force behind kinetochore orientation: lessons from meiosis. *Nat. Rev. Mol. Cell Biol.* 13, 370–382.

### Control of Cell Division and Cell Growth

Adhikary S & Eilers M (2005) Transcriptional regulation and transformation by Myc proteins. *Nat. Rev. Mol. Cell Biol.* 6, 635–645.

Bertoli C, Skotheim JM & de Bruin RA (2013) Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* 14, 518–528.

Dick FA & Rubin SM (2013) Molecular mechanisms underlying RB protein function. *Nat. Rev. Mol. Cell Biol.* 14, 297–306.

Jackson SP & Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461, 1071–1078.

Jorgensen P & Tyers M (2004) How cells coordinate growth and division. *Curr. Biol.* 14, R1014–R1027.

Shimobayashi M & Hall MN (2014) Making new contacts: the mTOR network in metabolism and signalling crosstalk. *Nat. Rev. Mol. Cell Biol.* 15, 155–162.

Turner JJ, Ewald JC & Skotheim JM (2012) Cell size control in yeast. *Curr. Biol.* 22, R350–R359.

van den Heuvel S & Dyson NJ (2008) Conserved functions of the pRB and E2F families. *Nat. Rev. Mol. Cell Biol.* 9, 713–724.

Vousden KH & Lu X (2002) Live or let die: the cell's response to p53. *Nat. Rev. Cancer* 2, 594–604.

Zoncu R, Efeyan A & Sabatini DM (2011) mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat. Rev. Mol. Cell Biol.* 12, 21–35.

# Cell Death

CHAPTER

# 18

The growth, development, and maintenance of multicellular organisms depend not only on the production of cells but also on mechanisms to destroy them. The maintenance of tissue size, for example, requires that cells die at the same rate as they are produced. During development, carefully orchestrated patterns of cell death help determine the size and shape of limbs and other tissues. Cells also die when they become damaged or infected, ensuring that they are removed before they threaten the health of the organism. In these and most other cases, cell death is not a random process but occurs by a programmed sequence of molecular events, in which the cell systematically destroys itself from within and is then eaten by other cells, leaving no trace. In most cases, this **programmed cell death** occurs by a process called **apoptosis**—from the Greek word meaning “falling off,” as leaves from a tree.

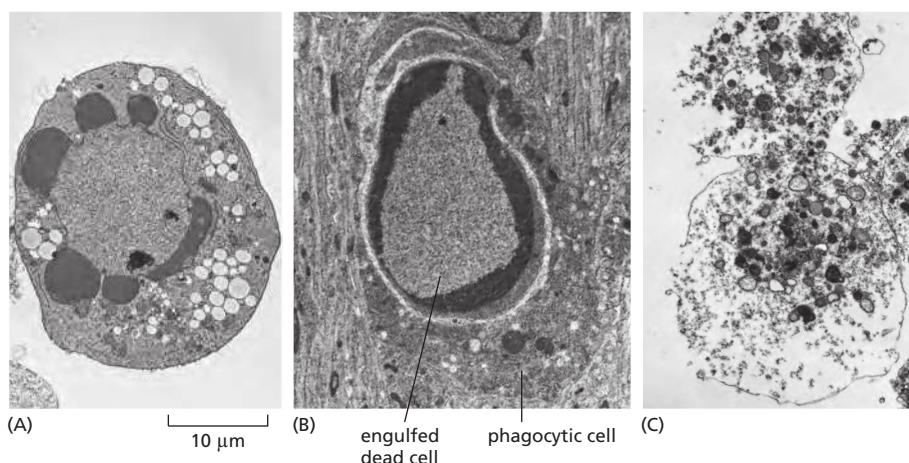
Cells dying by apoptosis undergo characteristic morphological changes. They shrink and condense, the cytoskeleton collapses, the nuclear envelope disassembles, and the nuclear chromatin condenses and breaks up into fragments (**Figure 18–1A**). The cell surface often bulges outward and, if the cell is large, it breaks up into membrane-enclosed fragments called *apoptotic bodies*. The surface of the cell or apoptotic bodies becomes chemically altered, so that a neighboring cell or a macrophage (a specialized phagocytic cell, discussed in Chapter 22) rapidly engulfs them, before they can spill their contents (Figure 18–1B). In this way, the cell dies neatly and is rapidly cleared away, without causing a damaging inflammatory response. Because the cells are eaten and digested so quickly, there are usually few dead cells to be seen, even when large numbers of cells have died by apoptosis. This is probably why biologists overlooked apoptosis for many years and still might underestimate its extent.

In contrast to apoptosis, animal cells that die in response to an acute insult, such as trauma or a lack of blood supply, usually do so by a process called *cell necrosis*. Necrotic cells swell and burst, spilling their contents over their neighbors and eliciting an inflammatory response (Figure 18–1C). In most cases, necrosis is likely to be caused by energy depletion, which leads to metabolic defects and loss of the ionic gradients that normally exist across the cell membrane. One form of necrosis, called *necroptosis*, is a form of programmed cell death that is triggered by a specific regulatory signal from other cells, although we are only just beginning to understand the underlying mechanisms.

Some form of programmed cell death occurs in many organisms, but apoptosis is found primarily in animals. This chapter focuses on the major functions of apoptosis, its mechanism and regulation, and how excessive or insufficient apoptosis can contribute to human disease.

## Apoptosis Eliminates Unwanted Cells

The amount of apoptotic cell death that occurs in developing and adult animal tissues is astonishing. In the developing vertebrate nervous system, for example, more than half of many types of nerve cells normally die soon after they are formed. It seems remarkably wasteful for so many cells to die, especially as the vast majority are perfectly healthy at the time they kill themselves. What purposes does this massive cell death serve?



**Figure 18-1 Two distinct forms of cell death.** These electron micrographs show cells that have died by apoptosis (A and B) or by necrosis (C). The cells in (A) and (C) died in a culture dish, whereas the cell in (B) died in a developing tissue and has been engulfed by a phagocytic cell. Note that the cells in (A) and (B) have condensed but seem relatively intact, whereas the cell in (C) seems to have exploded. The large vacuoles visible in the cytoplasm of the cell in (A) are a variable feature of apoptosis. (Courtesy of Julia Burne.)

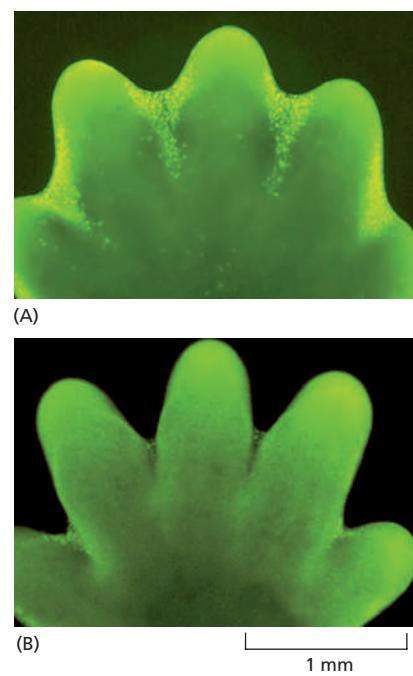
In some cases, the answer is clear. Cell death helps sculpt hands and feet during embryonic development: they start out as spade-like structures, and the individual digits separate only as the cells between them die, as illustrated for a mouse paw in **Figure 18–2**. In other cases, cells die when the structure they form is no longer needed. When a tadpole changes into a frog at metamorphosis, the cells in the tail die, and the tail, which is not needed in the frog, disappears. Apoptosis also functions as a quality-control process in development, eliminating cells that are abnormal, misplaced, nonfunctional, or potentially dangerous to the animal. Striking examples occur in the vertebrate adaptive immune system, where apoptosis eliminates developing T and B lymphocytes that either fail to produce potentially useful antigen-specific receptors or produce self-reactive receptors that make the cells potentially dangerous (discussed in Chapter 24); it also eliminates most of the lymphocytes activated by an infection, after they have helped destroy the responsible microbes.

In adult tissues that are neither growing nor shrinking, cell death and cell division must be tightly regulated to ensure that they are exactly in balance. If part of the liver is removed in an adult rat, for example, liver cell proliferation increases to make up the loss. Conversely, if a rat is treated with the drug phenobarbital—which stimulates liver cell division (and thereby liver enlargement)—and then the phenobarbital treatment is stopped, apoptosis in the liver greatly increases until the liver has returned to its original size, usually within a week or so. Thus, the liver is kept at a constant size through the regulation of both the cell death rate and the cell birth rate. The control mechanisms responsible for such regulation are largely unknown.

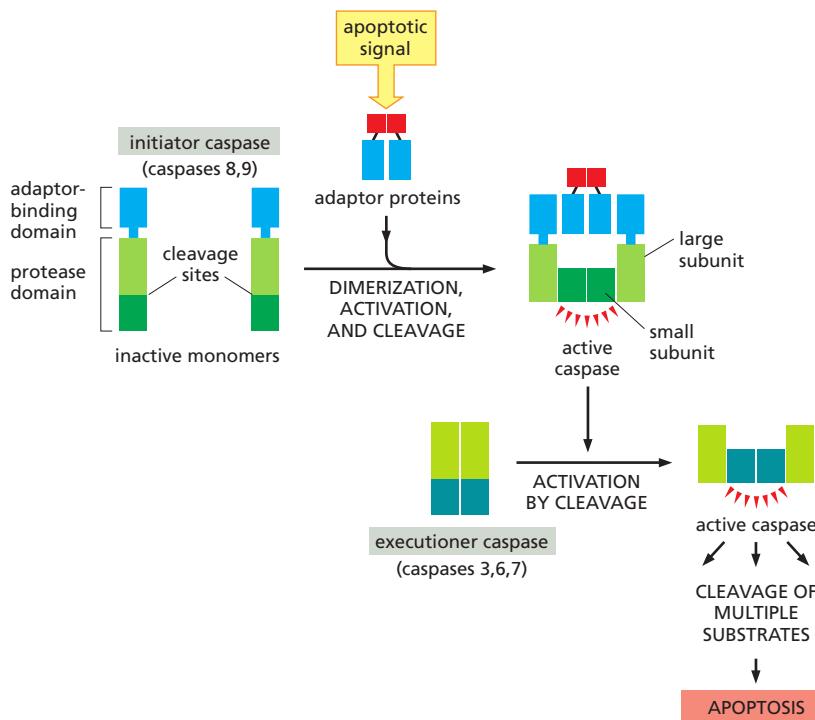
Animal cells can recognize damage in their various organelles and, if the damage is great enough, they can kill themselves by undergoing apoptosis. An important example is DNA damage, which can produce cancer-promoting mutations if not repaired. Cells have various ways of detecting DNA damage, and undergo apoptosis if they cannot repair it.

### Apoptosis Depends on an Intracellular Proteolytic Cascade That Is Mediated by Caspases

Apoptosis is triggered by members of a family of specialized intracellular proteases, which cleave specific sequences in numerous proteins inside the cell, thereby bringing about the dramatic changes that lead to cell death and engulfment. These proteases have a cysteine at their active site and cleave their target proteins at specific aspartic acids; they are therefore called **caspases** (c for cysteine and asp for aspartic acid). Caspases are synthesized in the cell as inactive precursors and are activated only during apoptosis. There are two major classes of apoptotic caspases: *initiator* caspases and *executioner* caspases.



**Figure 18–2 Sculpting the digits in the developing mouse paw by apoptosis.** (A) The paw in this mouse fetus has been stained with a dye that specifically labels cells that have undergone apoptosis. The apoptotic cells appear as *bright green dots* between the developing digits. (B) The interdigital cell death has eliminated the tissue between the developing digits, as seen one day later, when there are very few apoptotic cells. (From W. Wood et al., *Development* 127:5245–5252, 2000. With permission from The Company of Biologists.)



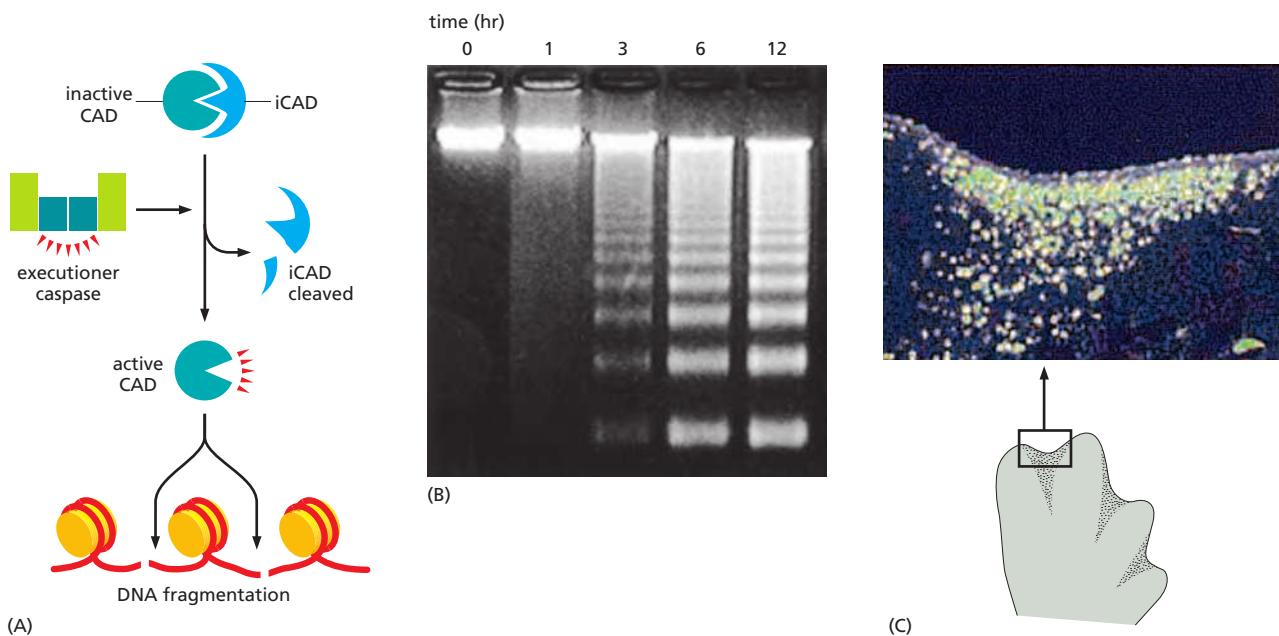
**Figure 18–3 Caspase activation during apoptosis.** An initiator caspase contains a protease domain in its carboxy-terminal region and a small protein interaction domain near its amino terminus. It is initially made in an inactive, monomeric form, sometimes called procaspase. Apoptotic signals trigger the assembly of adaptor proteins carrying multiple binding sites for the caspase amino-terminal domain. Upon binding to the adaptor proteins, the initiator caspases dimerize and are thereby activated, leading to cleavage of a specific site in their protease domains. Each protease domain is then rearranged into a large and small subunit. In some cases (not shown), the adaptor-binding domain of the initiator caspase is also cleaved (see Figure 18–5). Executioner caspases are initially formed as inactive dimers. Upon cleavage at a site in the protease domain by an initiator caspase, the executioner caspase dimer undergoes an activating conformational change. The executioner caspases then cleave a variety of key proteins, leading to the controlled death of the cell.

**Initiator caspases**, as their name implies, begin the apoptotic process. They normally exist as inactive, soluble monomers in the cytosol. An apoptotic signal triggers the assembly of large protein platforms that bring multiple initiator caspases together into large complexes. Within these complexes, pairs of caspases associate to form dimers, resulting in protease activation (Figure 18–3). Each caspase in the dimer then cleaves its partner at a specific site in the protease domain, which stabilizes the active complex and is required for the proper function of the enzyme in the cell.

The major function of the initiator caspases is to activate the **executioner caspases**. These normally exist as inactive dimers. When they are cleaved by an initiator caspase at a site in the protease domain, the active site is rearranged from an inactive to an active conformation. One initiator caspase complex can activate many executioner caspases, resulting in an amplifying proteolytic cascade. Once activated, executioner caspases catalyze the widespread protein cleavage events that kill the cell.

Various experimental approaches have led to the identification of over a thousand proteins that are cleaved by caspases during apoptosis. Only a few of these proteins have been studied in any detail. These include the nuclear lamins, the cleavage of which causes the irreversible breakdown of the nuclear lamina (discussed in Chapter 12). Another target is a protein that normally holds a DNA-degrading endonuclease in an inactive form; its cleavage frees the endonuclease to cut up the DNA in the cell nucleus (Figure 18–4). Other target proteins include components of the cytoskeleton and cell–cell adhesion proteins that attach cells to their neighbors; the cleavage of these proteins helps the apoptotic cell to round up and detach from its neighbors, making it easier for a neighboring cell to engulf it, or, in the case of an epithelial cell, for the neighbors to extrude the apoptotic cell from the cell sheet. The caspase cascade is not only destructive and self-amplifying but also irreversible, so that once a cell starts out along the path to destruction, it cannot turn back.

How is the initiator caspase first activated in response to an apoptotic signal? The two best-understood activation mechanisms in mammalian cells are called the *extrinsic pathway* and the *intrinsic, or mitochondrial, pathway*. Each uses its own initiator caspase and activation system, as we now discuss.

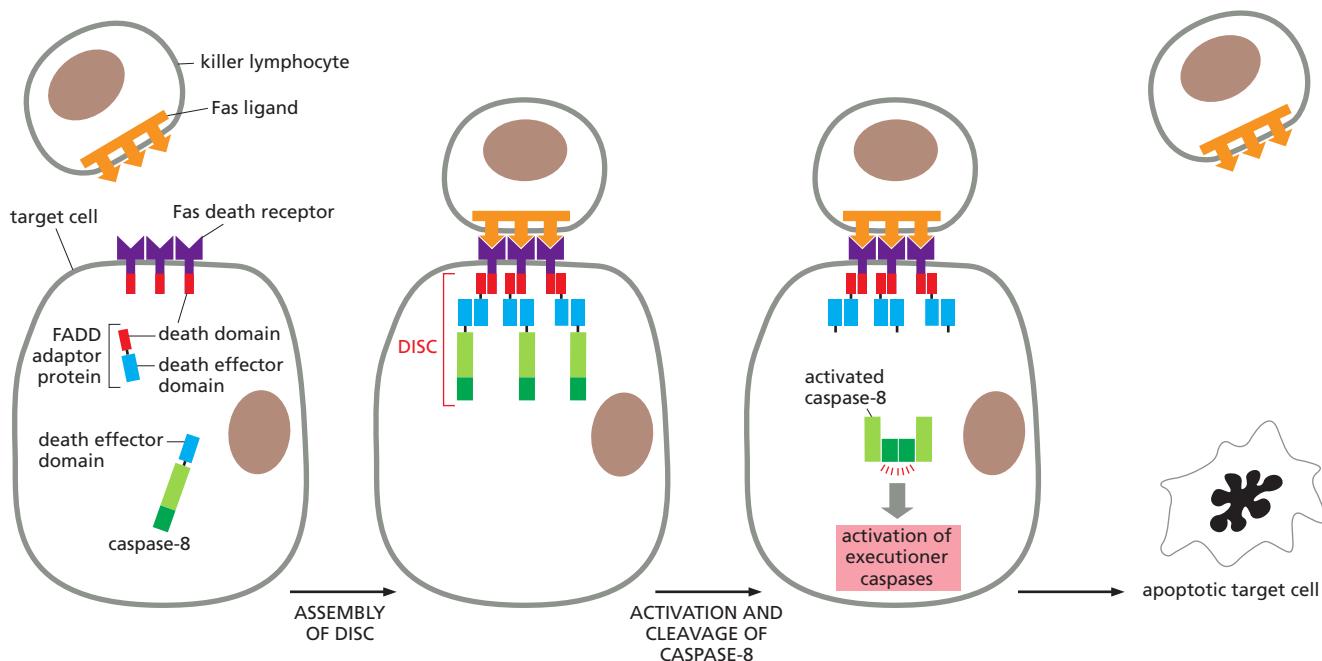


**Figure 18–4 DNA fragmentation during apoptosis.** (A) In healthy cells, the endonuclease CAD associates with its inhibitor, iCAD. Activation of executioner caspases in the cell leads to cleavage of iCAD, which unleashes the nuclease. Activated CAD cuts the chromosomal DNA between nucleosomes, resulting in the production of DNA fragments that form a ladder pattern (see B) upon gel electrophoresis. (B) Mouse thymus lymphocytes were treated with an antibody against the cell-surface death receptor Fas (discussed in the text), inducing the cells to undergo apoptosis. DNA was extracted at the times indicated above the figure, and the fragments were separated by size by electrophoresis in an agarose gel and stained with ethidium bromide. Because the cleavages occur in the linker regions between nucleosomes, the fragments separate into a characteristic ladder pattern on these gels. Note that in gel electrophoresis, smaller molecules are more widely separated in the lower part of the gel, so that removal of a single nucleosome has a greater apparent effect on their gel mobility. (C) Apoptotic nuclei can be detected using a technique that adds a fluorescent label to DNA ends. In the image shown here, this technique was used in a tissue section of a developing chick leg bud; this cross section through the skin and underlying tissue is from a region between two developing digits, as indicated in the underlying drawing. The procedure is called the TUNEL (*TdT*-mediated dUTP nick end labeling) technique because the enzyme terminal deoxynucleotidyl transferase (*TdT*) adds chains of labeled deoxynucleotide (dUTP) to the 3'-OH ends of DNA fragments. The presence of large numbers of DNA fragments therefore results in bright fluorescent dots in apoptotic cells. (B, from D. McIlroy et al., *Genes Dev.* 14:549–558, 2000. With permission from Cold Spring Harbor Laboratory Press; C, from V. Zuzarte-Luis and J.M. Hurlé, *Int. J. Dev. Biol.* 46:871–876, 2002. With permission from UBC Press.)

### Cell-Surface Death Receptors Activate the Extrinsic Pathway of Apoptosis

Extracellular signal proteins binding to cell-surface **death receptors** trigger the **extrinsic pathway** of apoptosis. Death receptors are transmembrane proteins containing an extracellular ligand-binding domain, a single transmembrane domain, and an intracellular *death domain*, which is required for the receptors to activate the apoptotic program. The receptors are homotrimers and belong to the *tumor necrosis factor (TNF) receptor* family, which includes a receptor for TNF itself and the *Fas* death receptor. The ligands that activate the death receptors are also homotrimers; they are structurally related to one another and belong to the *TNF family* of signal proteins.

A well-understood example of how death receptors trigger the extrinsic pathway of apoptosis is the activation of *Fas* on the surface of a target cell by *Fas ligand* on the surface of a killer (cytotoxic) lymphocyte. When activated by the binding of *Fas ligand*, the death domains on the cytosolic tails of the *Fas* death receptors bind intracellular adaptor proteins, which in turn bind initiator caspases (primarily caspase-8), forming a **death-inducing signaling complex (DISC)**. Once dimerized and activated in the DISC, the initiator caspases cleave their partners and then activate downstream executioner caspases to induce apoptosis (Figure 18–5). In some cells, the extrinsic pathway recruits the intrinsic apoptotic pathway to amplify the caspase cascade and kill the cell.



Many cells produce inhibitory proteins that act to restrain the extrinsic pathway. For example, some cells produce the protein *FLIP*, which resembles an initiator caspase but has no protease activity because it lacks the key cysteine in its active site. *FLIP* dimerizes with caspase-8 in the DISC; although caspase-8 appears to be active in these heterodimers, it is not cleaved at the site required for its stable activation, and the apoptotic signal is blocked. Such inhibitory mechanisms help prevent the inappropriate activation of the extrinsic pathway of apoptosis.

### The Intrinsic Pathway of Apoptosis Depends on Mitochondria

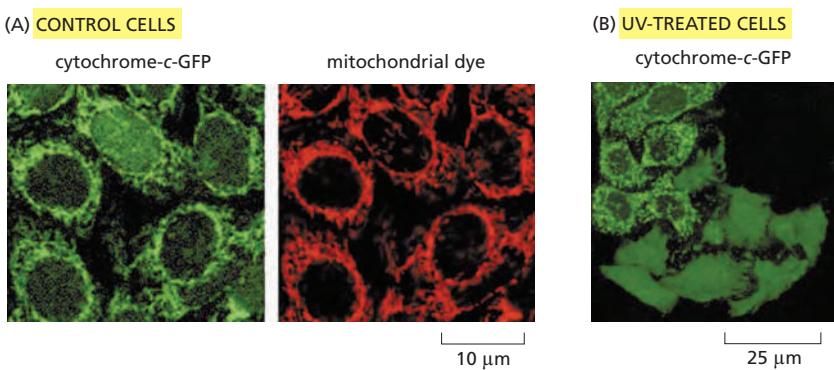
Cells can also activate their apoptosis program from inside the cell, often in response to stresses, such as DNA damage, or in response to developmental signals. In vertebrate cells, these responses are governed by the **intrinsic, or mitochondrial, pathway of apoptosis**, which depends on the release into the cytosol of mitochondrial proteins that normally reside in the intermembrane space of these organelles (see Figure 12–19). Some of the released proteins activate a caspase proteolytic cascade in the cytoplasm, leading to apoptosis.

A key protein in the intrinsic pathway is **cytochrome *c***, a water-soluble component of the mitochondrial electron-transport chain. When released into the cytosol (Figure 18–6), it takes on a new function: it binds to an adaptor protein called **Apaf1** (*apoptotic protease activating factor-1*), causing the Apaf1 to oligomerize into a wheel-like heptamer called an **apoptosome**. The Apaf1 proteins in the apoptosome then recruit initiator caspase-9 proteins, which are thought to be activated by proximity in the apoptosome, just as caspase-8 is activated in the DISC. The activated caspase-9 molecules then activate downstream executioner caspases to induce apoptosis (Figure 18–7).

### Bcl2 Proteins Regulate the Intrinsic Pathway of Apoptosis

The intrinsic pathway of apoptosis is tightly regulated to ensure that cells kill themselves only when it is appropriate. A major class of intracellular regulators of the intrinsic pathway is the **Bcl2 family** of proteins, which, like the caspase family, has been conserved in evolution from worms to humans; a human Bcl2 protein, for example, can suppress apoptosis when expressed in the worm *Caenorhabditis elegans*.

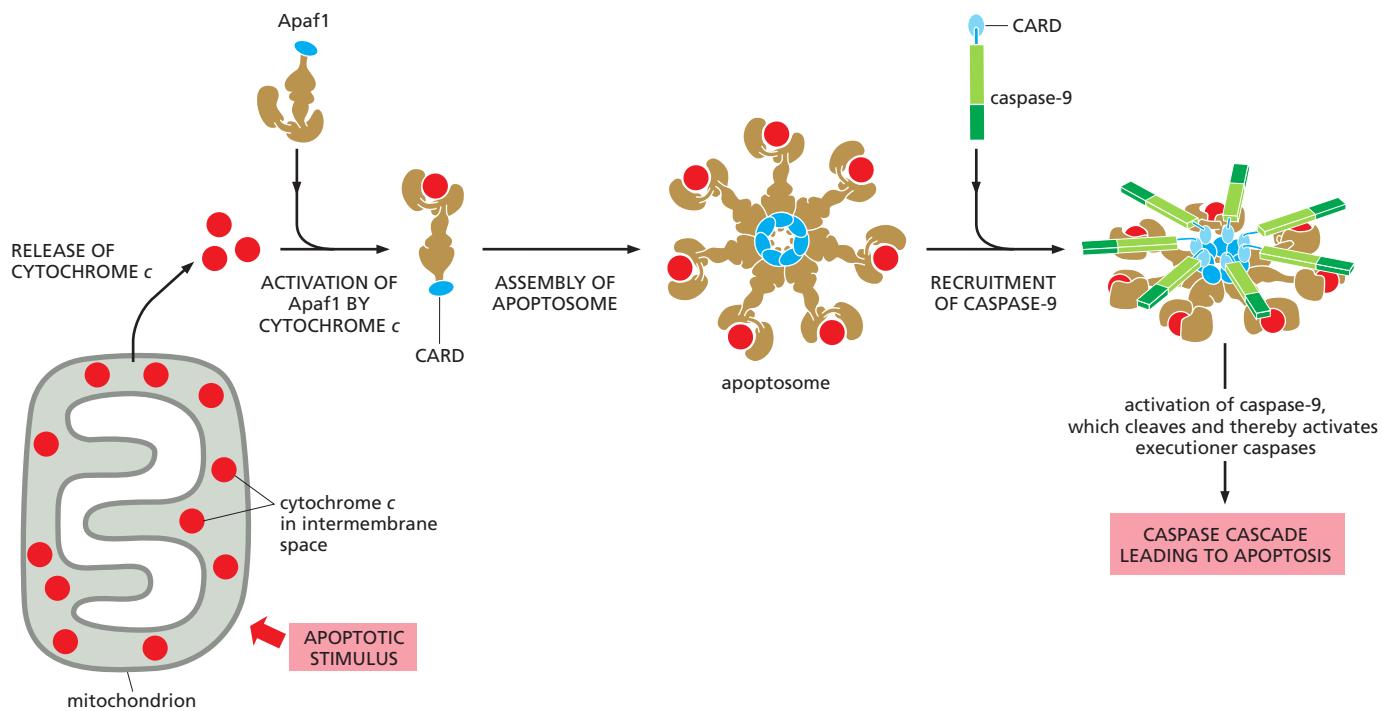
**Figure 18–5** The extrinsic pathway of apoptosis activated through Fas death receptors. Trimeric Fas ligands on the surface of a killer lymphocyte interact with trimeric Fas receptors on the surface of the target cell, leading to clustering of several ligand-bound receptor trimers (only one trimer is shown here for clarity). Receptor clustering activates death domains on the receptor tails, which interact with similar domains on the adaptor protein FADD (FADD stands for Fas-associated death domain). Each FADD protein then recruits an initiator caspase (caspase-8) via a death effector domain on both FADD and the caspase, forming a death-inducing signaling complex (DISC). Within the DISC, two adjacent initiator caspases interact and cleave one another to form an activated protease dimer, which then cleaves itself in the region linking the protease to the death effector domain. This stabilizes and releases the active caspase dimer into the cytosol, where it activates executioner caspases by cleaving them.



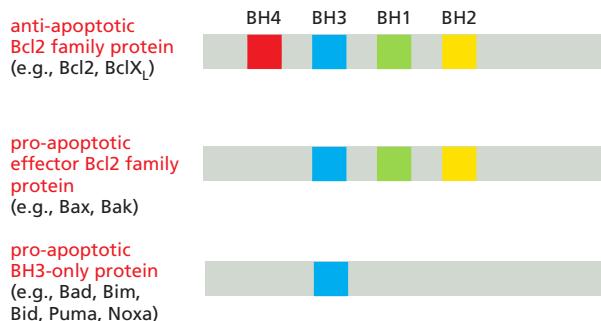
Mammalian Bcl2 family proteins regulate the intrinsic pathway of apoptosis mainly by controlling the release of cytochrome *c* and other intermembrane mitochondrial proteins into the cytosol. Some Bcl2 family proteins are *pro-apoptotic* and promote apoptosis by enhancing the release, whereas others are *anti-apoptotic* and inhibit apoptosis by blocking the release. The pro-apoptotic and anti-apoptotic proteins can bind to each other in various combinations to form heterodimers in which the two proteins inhibit each other's function. The balance between the activities of these two functional classes of Bcl2 family proteins largely determines whether a mammalian cell lives or dies by the intrinsic pathway of apoptosis.

As illustrated in Figure 18–8, the anti-apoptotic Bcl2 family proteins, including *Bcl2* itself (the founding member of the Bcl2 family) and *BclX<sub>L</sub>*, share four distinctive *Bcl2 homology (BH)* domains (BH1–4). The pro-apoptotic Bcl2 family proteins

**Figure 18–6 Release of cytochrome *c* from mitochondria in the intrinsic pathway of apoptosis.** Fluorescence micrographs of human cancer cells in culture. (A) The control cells were transfected with a gene encoding a fusion protein consisting of cytochrome *c* linked to green fluorescent protein (cytochrome-c-GFP); they were also treated with a red dye that accumulates in mitochondria. The overlapping distribution of the green and red indicates that the cytochrome-c-GFP is located in mitochondria. (B) Cells expressing cytochrome-c-GFP were irradiated with ultraviolet (UV) light to induce the intrinsic pathway of apoptosis and were photographed 5 hours later. The six cells in the bottom half of this micrograph have released their cytochrome *c* from mitochondria into the cytosol, whereas the cells in the upper half of the micrograph have not yet done so (Movie 18.1). (From J.C. Goldstein et al., *Nat. Cell Biol.* 2:156–162, 2000. With permission from Macmillan Publishers Ltd.)



**Figure 18–7 The intrinsic pathway of apoptosis.** Intracellular apoptotic stimuli cause mitochondria to release cytochrome *c*, which interacts with Apaf1. The binding of cytochrome *c* causes Apaf1 to unfold partly, exposing a domain that interacts with the same domain in other activated Apaf1 molecules. Seven activated Apaf1 proteins form a large ring complex called the apoptosome. Each Apaf1 protein contains a caspase recruitment domain (CARD), and these are clustered above the central hub of the apoptosome. The CARDs bind similar domains in multiple caspase-9 molecules, which are thereby recruited into the apoptosome and activated. The mechanism of caspase-9 activation is not clear: it probably results from dimerization and cleavage of adjacent caspase-9 proteins, but it might also depend on interactions between caspase-9 and Apaf1. Once activated, caspase-9 cleaves and thereby activates downstream executioner caspases. Note that the CARD is related in structure and function to the death effector domain of caspase-8 (see Figure 18–5). Some scientists use the term “apoptosome” to refer to the complex containing caspase-9.



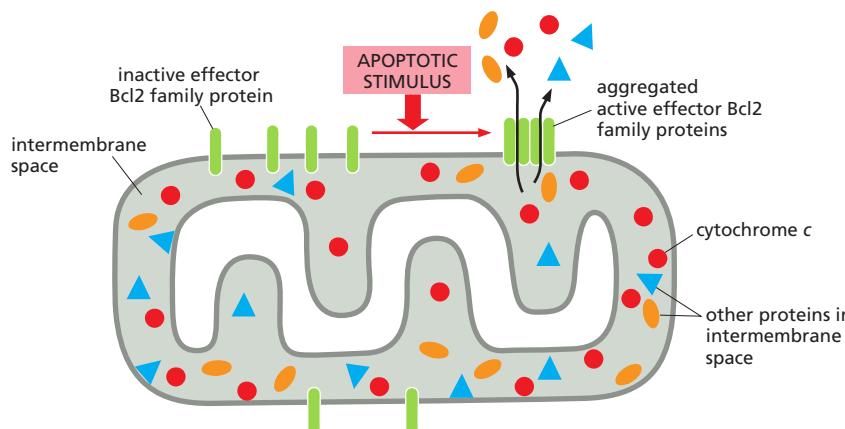
**Figure 18-8** The three classes of Bcl2 family proteins. Note that the BH3 domain is the only BH domain shared by all Bcl2 family members; it mediates the direct interactions between pro-apoptotic and anti-apoptotic family members.

consist of two subfamilies—the **effector Bcl2 family proteins** and the **BH3-only proteins**. The main effector proteins are *Bax* and *Bak*, which are structurally similar to Bcl2 but lack the BH4 domain. The BH3-only proteins share sequence homology with Bcl2 in only the BH3 domain.

When an apoptotic stimulus triggers the intrinsic pathway, the pro-apoptotic **effector Bcl2 family proteins** become activated and aggregate to form oligomers in the mitochondrial outer membrane, inducing the release of cytochrome *c* and other intermembrane proteins by an unknown mechanism (Figure 18-9). In mammalian cells, **Bax** and **Bak** are the main effector Bcl2 family proteins, and at least one of them is required for the intrinsic pathway of apoptosis to operate: mutant mouse cells that lack both proteins are resistant to all pro-apoptotic signals that normally activate this pathway. Whereas Bak is bound to the mitochondrial outer membrane even in the absence of an apoptotic signal, Bax is mainly located in the cytosol and translocates to the mitochondria only after an apoptotic signal activates it. As we discuss below, the activation of Bax and Bak usually depends on activated pro-apoptotic BH3-only proteins.

The **anti-apoptotic Bcl2 family proteins** such as Bcl2 itself and BclX<sub>L</sub> are also located on the cytosolic surface of the outer mitochondrial membrane, where they help prevent inappropriate release of intermembrane proteins. The anti-apoptotic Bcl2 family proteins inhibit apoptosis mainly by binding to and inhibiting pro-apoptotic Bcl2 family proteins—either on the mitochondrial membrane or in the cytosol. On the outer mitochondrial membrane, for example, they bind to Bak and prevent it from oligomerizing, thereby inhibiting the release of cytochrome *c* and other intermembrane proteins. There are at least five mammalian anti-apoptotic Bcl2 family proteins, and every mammalian cell requires at least one to survive. Moreover, a number of these proteins must be inhibited for the intrinsic pathway to induce apoptosis; the BH3-only proteins mediate the inhibition.

The **BH3-only proteins** are the largest subclass of Bcl2 family proteins. The cell either produces or activates them in response to an apoptotic stimulus, and they are thought to promote apoptosis mainly by inhibiting anti-apoptotic Bcl2 family

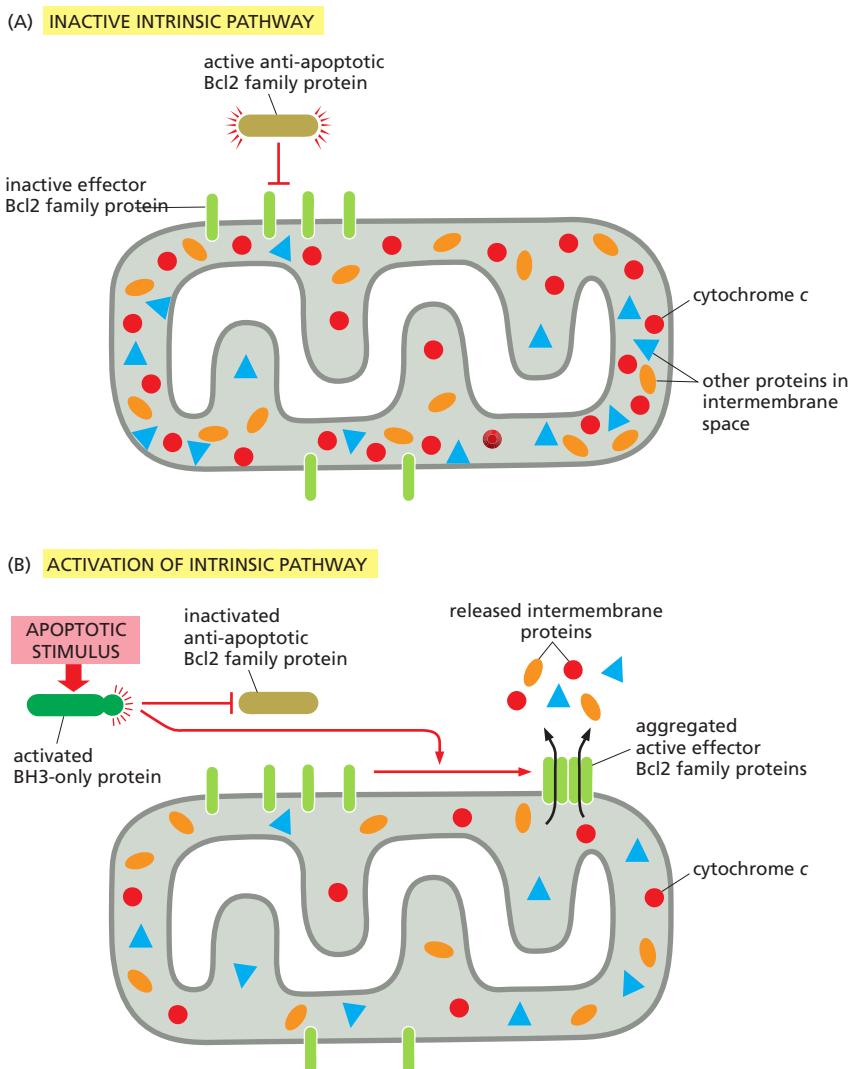


**Figure 18-9** The role of pro-apoptotic effector Bcl2 family proteins (mainly Bax and Bak) in the release of mitochondrial intermembrane proteins in the intrinsic pathway of apoptosis. When activated by an apoptotic stimulus, the effector Bcl2 family proteins aggregate on the outer mitochondrial membrane and release cytochrome *c* and other proteins from the intermembrane space into the cytosol by an unknown mechanism.

proteins. Their BH3 domain binds to a long hydrophobic groove on anti-apoptotic Bcl2 family proteins, neutralizing their activity. This binding and inhibition enables the aggregation of Bax and Bak on the surface of mitochondria, which triggers the release of the intermembrane mitochondrial proteins that induce apoptosis (Figure 18–10). Some BH3-only proteins may bind directly to Bax and Bak to help stimulate their aggregation.

BH3-only proteins provide the crucial link between apoptotic stimuli and the intrinsic pathway of apoptosis, with different stimuli activating different BH3-only proteins. Some extracellular survival signals, for example, block apoptosis by inhibiting the synthesis or activity of certain BH3-only proteins (see Figure 18–12B). Similarly, in response to DNA damage that cannot be repaired, the tumor suppressor protein p53 accumulates (discussed in Chapters 17 and 20) and activates the transcription of genes that encode the BH3-only proteins *Puma* and *Noxa*. These BH3-only proteins then trigger the intrinsic pathway, thereby eliminating a potentially dangerous cell that could otherwise become cancerous.

As mentioned earlier, in some cells the extrinsic apoptotic pathway recruits the intrinsic pathway to amplify the caspase cascade to kill the cell. The BH3-only protein *Bid* is the link between the two pathways. *Bid* is normally inactive. However, when death receptors activate the extrinsic pathway in some cells, the initiator caspase, caspase-8, cleaves *Bid*, producing an active form of *Bid* that translocates to the outer mitochondrial membrane and inhibits anti-apoptotic Bcl2 family proteins, thereby amplifying the death signal.



**Figure 18–10** How pro-apoptotic BH3-only and anti-apoptotic Bcl2 family proteins regulate the intrinsic pathway of apoptosis. (A) In the absence of an apoptotic stimulus, anti-apoptotic Bcl2 family proteins bind to and inhibit the effector Bcl2 family proteins on the mitochondrial outer membrane (and in the cytosol—not shown). (B) In the presence of an apoptotic stimulus, BH3-only proteins are activated and bind to the anti-apoptotic Bcl2 family proteins so that they can no longer inhibit the effector Bcl2 family proteins; the latter then become activated, aggregate in the outer mitochondrial membrane, and promote the release of intermembrane mitochondrial proteins into the cytosol. Some activated BH3-only proteins may stimulate mitochondrial protein release more directly by binding to and activating the effector Bcl2 family proteins. Although not shown, the anti-apoptotic Bcl2 family proteins are bound to the mitochondrial surface.

## IAPs Help Control Caspases

Because activation of a caspase cascade leads to certain death, the cell employs multiple robust mechanisms to ensure that these proteases are activated only when appropriate. One line of defense is provided by a family of proteins called **inhibitors of apoptosis (IAPs)**. These proteins were first identified in certain insect viruses (baculoviruses), which encode IAP proteins to prevent a host cell that is infected by the virus from killing itself by apoptosis. It is now known that most animal cells also make IAP proteins.

All IAPs have one or more BIR (baculovirus IAP repeat) domains, which enable them to bind to and inhibit activated caspases. Some IAPs also polyubiquitylate caspases, marking the caspases for destruction by proteasomes. In this way, the IAPs set an inhibitory threshold that caspases must overcome to trigger apoptosis.

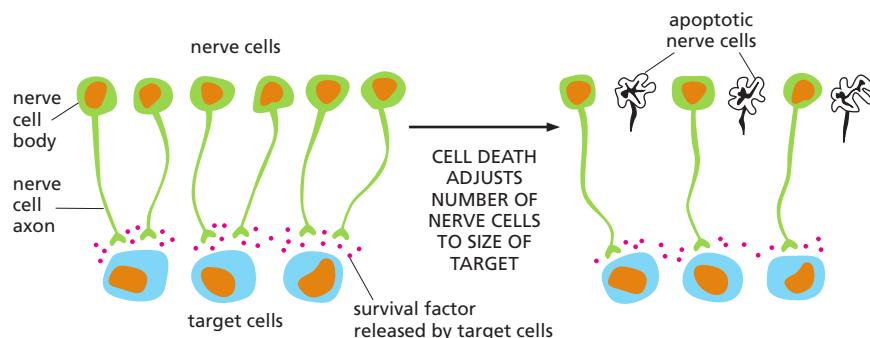
In *Drosophila* at least, the inhibitory barrier provided by IAPs can be neutralized by anti-IAP proteins, which are produced in response to various apoptotic stimuli. There are numerous anti-IAPs in flies, including *Reaper*, *Grim*, and *Hid*, and their only structural similarity is their short, N-terminal, IAP-binding motif, which binds to the BIR domain of IAPs, preventing the domain from binding to a caspase. Deletion of the three genes encoding *Reaper*, *Grim*, and *Hid* blocks apoptosis in flies. Conversely, inactivation of one of the two genes that encode IAPs in *Drosophila* causes all of the cells in the developing fly embryo to undergo apoptosis. Clearly, the balance between IAPs and anti-IAPs is tightly regulated and is crucial for controlling apoptosis in the fly.

The role of mammalian IAP and anti-IAP proteins in apoptosis is less clear. Anti-IAPs are released from the mitochondrial intermembrane space when the intrinsic pathway of apoptosis is activated, blocking IAPs in the cytosol and thereby promoting apoptosis. However, mice appear to develop normally if they are missing either the major mammalian IAP (called XIAP) or the two known mammalian anti-IAPs (called Smac/Diablo and Omi). Worms do not even contain a caspase-inhibiting IAP protein. Apparently, the tight control of caspase activity is achieved by different mechanisms in different animals.

## Extracellular Survival Factors Inhibit Apoptosis in Various Ways

Intercellular signals regulate most activities of animal cells, including apoptosis. These extracellular signals are part of the normal “social” controls that ensure that individual cells behave for the good of the organism as a whole—in this case, by surviving when they are needed and killing themselves when they are not. Some extracellular signal molecules stimulate apoptosis, whereas others inhibit it. We have discussed signal proteins such as Fas ligand that activate death receptors and thereby trigger the extrinsic pathway of apoptosis. Other extracellular signal molecules that stimulate apoptosis are especially important during vertebrate development: a surge of thyroid hormone in the bloodstream, for example, signals cells in the tadpole tail to undergo apoptosis at metamorphosis. In mice, locally produced signal proteins stimulate cells between developing fingers and toes to kill themselves (see Figure 18–2). Here, however, we focus on extracellular signal molecules that inhibit apoptosis, which are collectively called **survival factors**.

Most animal cells require continuous signaling from other cells to avoid apoptosis. This surprising arrangement apparently helps ensure that cells survive only when and where they are needed. Nerve cells, for example, are produced in excess in the developing nervous system and then compete for limited amounts of survival factors that are secreted by the target cells that they normally connect to (see Figure 21–81). Nerve cells that receive enough survival signals live, while the others die. In this way, the number of surviving neurons is automatically adjusted so that it is appropriate for the number of target cells they connect with (Figure 18–11). A similar competition for limited amounts of survival factors produced by neighboring cells is thought to control cell numbers in other tissues, both during development and in adulthood.

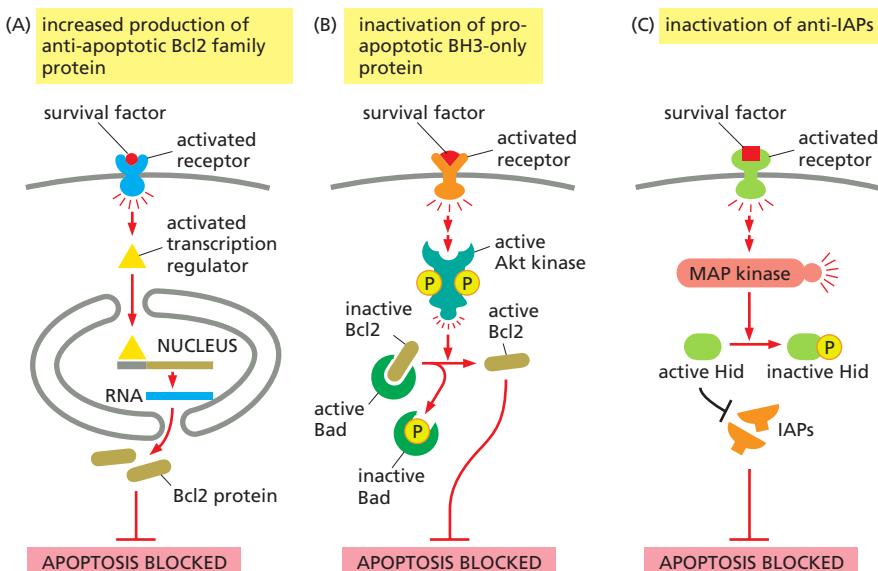


**Figure 18–11** The role of survival factors and cell death in adjusting the number of developing nerve cells to the amount of target tissue. More nerve cells are produced than can be supported by the limited amount of survival factors released by the target cells. Therefore, some nerve cells receive an insufficient amount of survival factors to avoid apoptosis. This strategy of overproduction followed by culling helps ensure that all target cells are contacted by nerve cells and that the extra nerve cells are automatically eliminated.

Survival factors usually bind to cell-surface receptors, which activate intracellular signaling pathways that suppress the apoptotic program, often by regulating members of the Bcl2 family of proteins. Some survival factors, for example, stimulate the synthesis of anti-apoptotic Bcl2 family proteins such as Bcl2 itself or BclXL (Figure 18–12A). Others act by inhibiting the function of pro-apoptotic BH3-only proteins such as Bad (Figure 18–12B). In *Drosophila*, some survival factors act by phosphorylating and inactivating anti-IAP proteins such as Hid, thereby enabling IAP proteins to suppress apoptosis (Figure 18–12C). Some developing neurons, like those illustrated in Figure 18–11, use an ingenious alternative approach: survival-factor receptors stimulate apoptosis—by an unknown mechanism—when they are not occupied, and then stop promoting death when survival factor binds. The end result in all these cases is the same: cell survival depends on survival factor binding.

### Phagocytes Remove the Apoptotic Cell

Apoptotic cell death is a remarkably tidy process: the apoptotic cell and its fragments do not break open and release their contents, but instead remain intact as they are efficiently eaten—or *phagocytosed*—by neighboring cells, leaving no trace and therefore triggering no inflammatory response (see Figure 18–1B and Movie 13.5). This engulfment process depends on chemical changes on the surface of the apoptotic cell, which displays signals that recruit phagocytic cells. An especially important change occurs in the distribution of the negatively charged phospholipid *phosphatidylserine* on the cell surface. This phospholipid is normally located exclusively in the inner leaflet of the lipid bilayer of the plasma membrane (see Figure 10–15), but it flips to the outer leaflet in apoptotic cells. The



**Figure 18–12** Three ways that extracellular survival factors can inhibit apoptosis. (A) Some survival factors suppress apoptosis by stimulating the transcription of genes that encode anti-apoptotic Bcl2 family proteins such as Bcl2 itself or BclXL. (B) Many others activate the serine/threonine protein kinase Akt, which, among many other targets, phosphorylates and inactivates the pro-apoptotic BH3-only protein Bad (see Figure 15–53). When not phosphorylated, Bad promotes apoptosis by binding to and inhibiting Bcl2; once phosphorylated, Bad dissociates, freeing Bcl2 to suppress apoptosis. Akt also suppresses apoptosis by phosphorylating and inactivating transcription regulatory proteins that stimulate the transcription of genes encoding proteins that promote apoptosis (not shown). (C) In *Drosophila*, some survival factors inhibit apoptosis by stimulating the phosphorylation of the anti-IAP protein Hid. When not phosphorylated, Hid promotes cell death by inhibiting IAPs. Once phosphorylated, Hid no longer inhibits IAPs, which become active and block apoptosis. MAP kinase, mitogen-activated protein kinase.

underlying mechanism is poorly understood, but the external exposure of phosphatidylserine is likely to depend on caspase cleavage of some protein involved in phospholipid distribution in the membrane. A variety of soluble “bridging” proteins interact with the exposed phosphatidylserine on the apoptotic cell. These bridging proteins also interact with specific receptors on the surface of a neighboring cell or macrophage, triggering cytoskeletal and other changes that initiate the engulfment process.

Macrophages do not phagocytose healthy cells in the animal—despite the fact that healthy cells normally expose some phosphatidylserine on their surfaces. Healthy cells express signal proteins on their surface that interact with inhibitory receptors on macrophages that block phagocytosis. Thus, in addition to expressing cell-surface signals such as phosphatidylserine that stimulate phagocytosis, apoptotic cells must lose or inactivate these “don’t eat me” signals that block phagocytosis.

### [Either Excessive or Insufficient Apoptosis Can Contribute to Disease](#)

There are many human disorders in which excessive numbers of cells undergo apoptosis and thereby contribute to tissue damage. Among the most dramatic examples are heart attacks and strokes. In these acute conditions, many cells die by necrosis as a result of ischemia (inadequate blood supply), but some of the less affected cells die by apoptosis. It is hoped that, in the future, drugs that block apoptosis—such as specific caspase inhibitors—will prove useful in saving such cells.

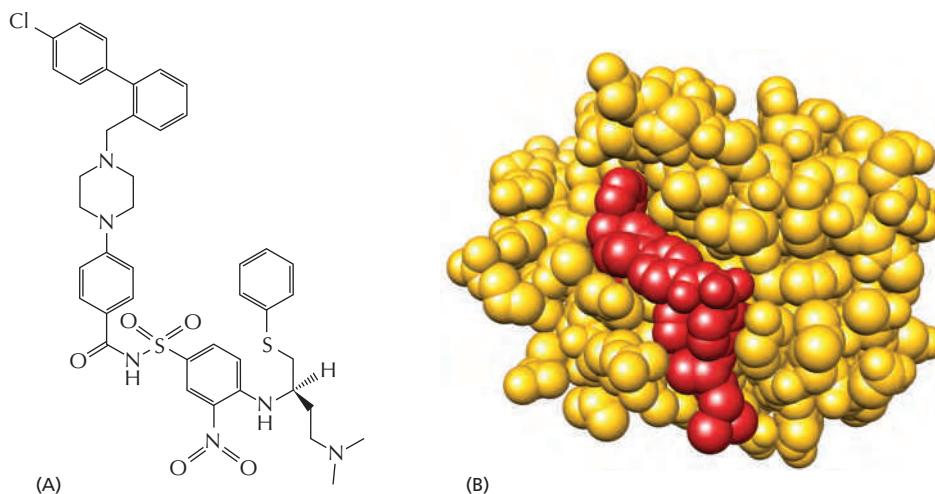
There are other conditions where too few cells die by apoptosis. Mutations in mice and humans, for example, that inactivate the genes that encode the Fas death receptor or the Fas ligand prevent the normal death of some lymphocytes, causing these cells to accumulate in excessive numbers in the spleen and lymph glands. In many cases, this leads to autoimmune disease, in which the lymphocytes react against the individual’s own tissues.

Decreased apoptosis also makes an important contribution to many tumors, as cancer cells often regulate their apoptotic program abnormally. The *Bcl2* gene, for example, was first identified in a common form of lymphocyte cancer in humans, where a chromosome translocation causes excessive production of the Bcl2 protein; indeed, Bcl2 gets its name from this *B cell lymphoma*. The high level of Bcl2 protein in the lymphocytes that carry the translocation promotes the development of cancer by inhibiting apoptosis, thereby prolonging lymphocyte survival and increasing their number; it also decreases the cells’ sensitivity to anticancer drugs, which commonly work by causing cancer cells to undergo apoptosis.

Similarly, the gene encoding the tumor suppressor protein p53 is mutated in about 50% of human cancers so that it no longer promotes apoptosis or cell-cycle arrest in response to DNA damage. The lack of p53 function therefore enables the cancer cells to survive and proliferate even when their DNA is damaged; in this way, the cells accumulate more mutations, some of which make the cancer more malignant (discussed in Chapter 20). As many anticancer drugs induce apoptosis (and cell-cycle arrest) by a p53-dependent mechanism (discussed in Chapters 17 and 20), the loss of p53 function also makes cancer cells less sensitive to these drugs.

If decreased apoptosis contributes to many cancers, then we might be able to treat those cancers with drugs that stimulate apoptosis. This line of thinking has recently led to the development of small chemicals that interfere with the function of anti-apoptotic Bcl2 family proteins such as Bcl2 and BclX<sub>L</sub>. These chemicals bind with high affinity to the hydrophobic groove on anti-apoptotic Bcl2 family proteins, blocking their function in essentially the same way that BH3-only proteins do ([Figure 18–13](#)). The intrinsic pathway of apoptosis is thereby stimulated, which in certain tumors increases the amount of cell death.

Most human cancers arise in epithelial tissues such as those in the lung, intestinal tract, breast, and prostate. Such cancer cells display many abnormalities in



**Figure 18–13** How the chemical ABT-737 inhibits anti-apoptotic Bcl2 family proteins. As shown in Figure 18–10B, an apoptotic signal results in activation of BH3-only proteins, which interact with a long hydrophobic groove in anti-apoptotic Bcl2 family proteins, thereby preventing them from blocking apoptosis. Using the crystal structure of the groove, the drug shown in (A), called ABT-737, was designed and synthesized to bind tightly in the groove, as shown for the anti-apoptotic Bcl2 family protein, Bcl $X_L$ , in (B). By inhibiting the activity of these proteins, the drug promotes apoptosis in any cell that depends on them for survival. (PDB code: 2YXJ.)

their behavior, including a decreased ability to adhere to the extracellular matrix and to one another at specialized cell-cell junctions. In the next chapter, we discuss the remarkable structures and functions of the extracellular matrix and cell junctions.

## Summary

*Animal cells can activate an intracellular death program and kill themselves in a controlled way when they are irreversibly damaged, no longer needed, or are a threat to the organism. In most cases, these deaths occur by apoptosis: the cells shrink, condense, and frequently fragment, and neighboring cells or macrophages rapidly phagocytose the cells or fragments before there is any leakage of cytoplasmic contents. Apoptosis is mediated by proteolytic enzymes called caspases, which cleave specific intracellular proteins to help kill the cell. Caspases are present in all nucleated animal cells as inactive precursors. Initiator caspases are activated when brought into proximity in activation complexes: once activated, they cleave and thereby activate downstream executioner caspases, which then cleave various target proteins in the cell, producing an amplifying, irreversible proteolytic cascade.*

*Cells use at least two distinct pathways to activate initiator caspases and trigger a caspase cascade leading to apoptosis: the extrinsic pathway is activated by extracellular ligands binding to cell-surface death receptors; the intrinsic pathway is activated by intracellular signals generated when cells are stressed. Each pathway uses its own initiator caspases, which are activated in distinct activation complexes: in the extrinsic pathway, the death receptors recruit caspase-8 via adaptor proteins to form the DISC; in the intrinsic pathway, cytochrome c released from the intermembrane space of mitochondria activates Apaf1, which assembles into an apoptosome and recruits and activates caspase-9.*

*Intracellular Bcl2family proteins and IAP proteins tightly regulate the apoptotic program to ensure that cells kill themselves only when it benefits the animal. Both anti-apoptotic and pro-apoptotic Bcl2 family proteins regulate the intrinsic pathway by controlling the release of mitochondrial intermembrane proteins, while IAP proteins inhibit activated caspases and promote their degradation.*

## WHAT WE DON'T KNOW

- How many forms of programmed cell death exist? What are the underlying mechanisms and benefits of each?
- Thousands of caspase substrates have been identified. Which ones are the critical proteins that must be cleaved to trigger the major cell remodeling events underlying apoptosis?
- How did the intrinsic pathway of apoptosis evolve, and what is the advantage of having mitochondria play such a central role in regulating apoptosis?
- How are “don’t eat me” signals eliminated or inactivated during apoptosis to allow the cells to be phagocytosed?

## PROBLEMS

Which statements are true? Explain why or why not.

**18–1** In normal adult tissues, cell death usually balances cell division.

**18–2** Mammalian cells that do not have cytochrome *c* should be resistant to apoptosis induced by DNA damage.

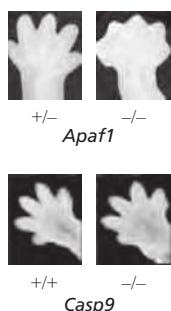
Discuss the following problems.

**18–3** One important role of Fas and Fas ligand is to mediate the elimination of tumor cells by killer lymphocytes. In a study of 35 primary lung and colon tumors, half the tumors were found to have amplified and overexpressed a gene for a secreted protein that binds to Fas ligand. How do you suppose that overexpression of this protein might contribute to the survival of these tumor cells? Explain your reasoning.

**18–4** Development of the nematode *Caenorhabditis elegans* generates exactly 959 somatic cells; it also produces an additional 131 cells that are later eliminated by apoptosis. Classical genetic experiments in *C. elegans* isolated mutants that led to the identification of the first genes involved in apoptosis. Of the many mutations affecting apoptosis in the nematode, none have ever been found in the gene for cytochrome *c*. Why do you suppose that such a central effector molecule in apoptosis was not found in the many genetic screens for “death” genes that have been carried out in *C. elegans*?

**18–5** Imagine that you could microinject cytochrome *c* into the cytosol of wild-type mammalian cells and of cells that were doubly defective for Bax and Bak. Would you expect one, both, or neither type of cell to undergo apoptosis? Explain your reasoning.

**18–6** In contrast to their similar brain abnormalities, newborn mice deficient in Apaf1 or caspase-9 have distinctive abnormalities in their paws. Apaf1-deficient mice fail to eliminate the webs between their developing digits, whereas caspase-9-deficient mice have normally formed digits (**Figure Q18–1**). If Apaf1 and caspase-9 function in the same apoptotic pathway, how is it possible for these deficient mice to differ in web-cell apoptosis?

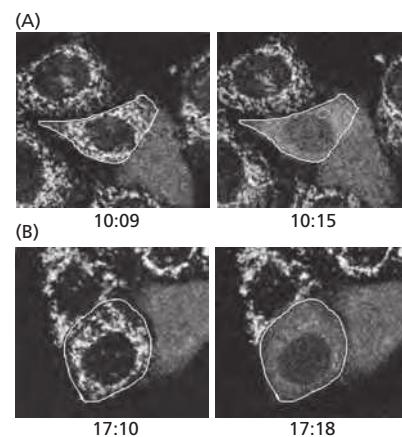


**Figure Q18–1** Appearance of paws in *Apaf1*<sup>−/−</sup> and *Casp9*<sup>−/−</sup> newborn mice relative to normal newborn mice (Problem 18–6). (From H. Yoshida et al., *Cell* 94:739–750, 1998. With permission from Elsevier.)

**18–7** When human cancer cells are exposed to ultraviolet (UV) light at 90 mJ/cm<sup>2</sup>, most of the cells undergo apoptosis within 24 hours. Release of cytochrome *c* from mitochondria can be detected as early as 6 hours after exposure of a population of cells to UV light, and it continues to increase for more than 10 hours thereafter. Does this mean that individual cells slowly release their cytochrome *c* over this time period? Or, alternatively, do individual cells release their cytochrome *c* rapidly but with different cells being triggered over the longer time period?

To answer this fundamental question, you have fused the gene for green fluorescent protein (GFP) to the gene for cytochrome *c*, so that you can observe the behavior of individual cells by confocal fluorescence microscopy. In cells that are expressing the cytochrome *c*-GFP fusion, fluorescence shows the punctate pattern typical of mitochondrial proteins. You then irradiate these cells with UV light and observe individual cells for changes in the punctate pattern. Two such cells (outlined in white) are shown in **Figure Q18–2A** and B. Release of cytochrome *c*-GFP is detected as a change from a punctate to a diffuse pattern of fluorescence. Times after UV exposure are indicated as hours:minutes below the individual panels.

Which model for cytochrome *c* release do these observations support? Explain your reasoning.



**Figure Q18–2** Time-lapse video fluorescence microscopic analysis of cytochrome *c*-GFP release from mitochondria of individual cells (Problem 18–7). (A) Cells observed for 6 minutes, 10 hours after UV irradiation. (B) Cells observed for 8 minutes, 17 hours after UV irradiation. One cell in (A) and one in (B), each outlined in white, have released their cytochrome *c*-GFP during the time frame of the observation, which is shown as hours:minutes below each panel. (From J.C. Goldstein et al., *Nat. Cell Biol.* 2:156–162, 2000. With permission from Macmillan Publishers Ltd.)

**18–8** Fas ligand is a trimeric, extracellular protein that binds to its receptor, Fas, which is composed of three identical transmembrane subunits (**Figure Q18–3**). The binding of Fas ligand alters the conformation of Fas so that it binds an adaptor protein, which then recruits and activates caspase-8, triggering a caspase cascade that leads to cell death. In humans, the autoimmune lymphoproliferative syndrome (ALPS) is associated with dominant mutations in Fas that include point mutations and C-terminal

truncations. In individuals that are heterozygous for such mutations, lymphocytes do not die at their normal rate and accumulate in abnormally large numbers, causing a variety of clinical problems. In contrast to these patients, individuals that are heterozygous for mutations that eliminate Fas expression entirely have no clinical symptoms.

**A.** Assuming that the normal and dominant forms of Fas are expressed to the same level and bind Fas ligand equally, what fraction of Fas–Fas ligand complexes on a lymphocyte from a heterozygous ALPS patient would be expected to be composed entirely of normal Fas subunits?

**B.** In an individual heterozygous for a mutation that eliminates Fas expression, what fraction of Fas–Fas ligand complexes would be expected to be composed entirely of normal Fas subunits?

**C.** Why are the Fas mutations that are associated with ALPS dominant, while those that eliminate expression of Fas are recessive?

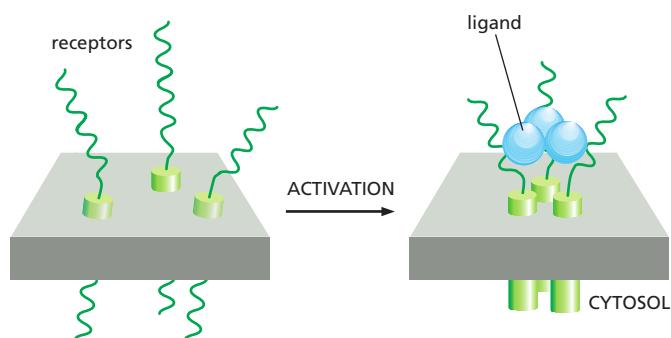


Figure Q18–3 The binding of trimeric Fas ligand to Fas (Problem 18–8).

## REFERENCES

- Crawford ED & Wells JA (2011) Caspase substrates and cellular remodeling. *Annu. Rev. Biochem.* 80, 1055–1087.
- Czabotar PE, Lessene G, Strasser A et al. (2014) Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* 15, 49–63.
- Danial NN & Korsmeyer SJ (2004) Cell death: critical control points. *Cell* 116, 205–219.
- Elliott MR & Ravichandran KS (2010) Clearance of apoptotic cells: implications in health and disease. *J. Cell Biol.* 189, 1059–1070.
- Ellis RE, Yuan JY & Horvitz RA (1991) Mechanisms and functions of cell death. *Annu. Rev. Cell Biol.* 7, 663–698.
- Fadok VA & Henson PM (2003) Apoptosis: giving phosphatidylserine recognition an assist—with a twist. *Curr. Biol.* 13, R655–R657.
- Green DR (2011) Means to an End: Apoptosis and Other Cell Death Mechanisms. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Jacobson MD, Weil M & Raff MC (1997) Programmed cell death in animal development. *Cell* 88, 347–354.
- Jiang X & Wang X (2004) Cytochrome C-mediated apoptosis. *Annu. Rev. Biochem.* 73, 87–106.
- Kerr JF, Wyllie AH & Currie AR (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Brit. J. Cancer* 26, 239–257.
- Kumar S (2007) Caspase function in programmed cell death. *Cell Death Differ.* 14, 32–43.
- Lavrik I, Golks A & Krammer PH (2005) Death receptor signaling. *J. Cell Sci.* 118, 265–267.
- Lessene G, Czabotar PE & Colman PM (2008) BCL-2 family antagonists for cancer therapy. *Nat. Rev. Drug Discov.* 7, 989–1000.
- Mace PD & Riedl SJ (2010) Molecular cell death platforms and assemblies. *Curr. Opin. Cell Biol.* 22, 828–836.
- Nagata S (2005) DNA degradation in development and programmed cell death. *Annu. Rev. Immunol.* 23, 853–875.
- Raff MC (1999) Cell suicide for beginners. *Nature* 396, 119–122.
- Tait SW & Green DR (2013) Mitochondrial regulation of cell death. *Cold Spring Harb. Perspect. Biol.* 5, a008706.
- Vanden Berghe T, Linkermann A, Jouan-Lanhouet S et al. (2014) Regulated necrosis: the expanding network of non-apoptotic cell death pathways. *Nat. Rev. Mol. Cell Biol.* 15, 135–147.
- Vousden KH (2005) Apoptosis. p53 and PUMA: a deadly duo. *Science* 309, 1685–1686.
- Willis SN & Adams JM (2005) Life in the balance: how BH3-only proteins induce apoptosis. *Curr. Opin. Cell Biol.* 17, 617–625.
- Yuan S & Akey CW (2013) Apoptosome structure, assembly, and procaspase activation. *Structure* 21, 501–515.